

Detecting Bubbles in the USD-JPY Exchange Rate by  
Sequential Monte Carlo Methods

by

John Rizcallah

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree

MASTER OF SCIENCE

Major Subject: Mathematics

West Texas A&M University

Canyon, Texas

December, 2020

## ABSTRACT

This paper uses recently developed Bayesian techniques in the analysis of stochastic bubbles in the USD-JPY exchange rate. After the fundamental value of the price series is removed, the exchange rate is subject to two regimes. The first regime follows a mean-reverting process around a long-term moving average. The second regime is an autoregressive process with an explosive root. The *SMC*<sup>2</sup> particle filter jointly estimates the hidden state and model parameters in real time. This method can readily deal with changes in market behavior and provides a measure of parameter uncertainty. Significant evidence of bubbles in the USD-JPY exchange rate were found. Furthermore, two trading strategies are devised and tested. Both strategies produce higher Sharpe ratios than a directional-trading benchmark.

## ACKNOWLEDGEMENTS

“It is the supreme art of the teacher to awaken joy in creative expression and knowledge.”

-Albert Einstein

I would like to thank Dr. Pamela Lockwood-Cooke, Dr. Daniel Seth, and Dr. Yong Yang, among the other artists in the math department, for fostering the joy of learning in myself and many others. While this thesis may determine my graduate degree, it is the thousands of hours spent engrossed in learning, both in class and out, that most affect my character. I hope they can recognize the influence they’ve had on me, and that they can be proud of yet another student.

I would also like to thank Gabi Castro for her immense patience and support through the long, stressful, and cranky research process. The late dinners, walks in the park, and shared bottles of wine kept me sane.

Approved:

[Chairman, Thesis Committee]

[Date]

[Member, Thesis Committee]

[Date]

[Member, Thesis Committee]

[Date]

[Department Head/Direct Supervisor]

[Date]

[Dean, Academic College]

[Date]

[Dean, Graduate School]

[Date]

# Contents

<b>1. Introduction</b>	<b>1</b>
<b>II. Literature Review</b>	<b>3</b>
Economic Literature . . . . .	4
Particle-Filter Literature . . . . .	10
<b>III. Methodology</b>	<b>16</b>
Fundamental Model for the Exchange Rate . . . . .	16
Two Time Series . . . . .	19
Hidden Markov Model . . . . .	21
Parameters . . . . .	22
Auxiliary Variables . . . . .	24
State-Particle Filter . . . . .	24
$\theta$ -Particle Filter . . . . .	28
Nested Particle Filter . . . . .	33
Bubble-Stamping Loss Function . . . . .	33
Simulated Data . . . . .	35
Data . . . . .	41
<b>IV. Results</b>	<b>44</b>
Results from Simulated Data . . . . .	44
Empirical Results . . . . .	47
<b>V. Trading Strategy</b>	<b>53</b>
<b>VI. References</b>	<b>59</b>

## List of Figures

1	Position of State-Particles at time 1 and time 160 . . . . .	28
2	Nested Algorithm Flowchart . . . . .	34
3	Simulated Data Series . . . . .	37
4	True and Stamped Bubbles . . . . .	45
5	Histogram of prior (white) and posterior (black) distributions of parameters . . . . .	46
6	ESS on Simulated Data . . . . .	47
7	Stamped Bubbles in the Bubble Component of the Exchange Rate . . . . .	49
8	Stamped Bubbles in the Bubble Component of the Exchange Rate . . . . .	50
9	Effective Sample Size in analysis of the bubble component, over time . . . . .	51
10	Histograms of the Distribution of Parameters in the Real Data . . . . .	52
11	Equity Growth in the Simple Directional Strategy . . . . .	55
12	Equity plots for each of the Simple Bubble Strategies . . . . .	56
13	Equity plots for each of the Directional Bubble Strategies . . . . .	57

## List of Tables

1	Table of States . . . . .	22
2	Simulated Data States . . . . .	37
3	Posterior Mean and Standard Deviation of Parameters . . . . .	48
4	Trade Statistics for Each Strategy . . . . .	54

# 1. Introduction

Since the global switch to floating exchange rates in the early 1970s, many econometric researchers have looked for evidence of speculative bubbles in currency markets. The literature is inconclusive; some researchers find evidence of speculative bubbles while others report no evidence. Recent developments in time series analysis have produced powerful bubble-detection methods, leading to a new wave of bubble-stamping research. Most of the new research focuses on stock markets, which are known to have periods of rampant speculation. Among the newly-discovered methods of bubble-detection is the use of Particle Filters to identify bubbles in real time. The following paper takes this recent and powerful method of analyzing markets and applies it to foreign exchange rates.

The research on bubble-detection in stock markets has included the application of quantitative bubble-stamping methods in trading strategies. Such research is lacking for currencies. If exchange rates are floating, then they are susceptible to speculation, and so bubble-trading strategies may be feasible. This research investigates the existence of speculative bubbles, then attempts to devise an algorithmic trading strategy based on that evidence.

The detection of speculative bubbles is important to central banks as well; the *raison d'être* of central banks is to regulate the value of the domestic currency, both to benefit international trade and to control inflation. As currency trading is buying one currency while selling another, explosive behavior in an exchange rate can have drastic effects on international trade agreements and may even lead to hyperinflation. The impact of exchange rate bubbles on global industry is also of importance. Many businesses have interests in countries with different currencies; the trend in globalization makes this research ever more meaningful.

This paper is organized as follows:

- **Chapter 2** is a literature review featuring in-depth discussion of the evidence for and against the existence of speculative bubbles in exchange rates, as well as the economic and mathematical basis for the proposed bubble-detection model.
- **Chapter 3** is an explanation of the mathematical and economic methods used in the detection model and the creation of the simulated test-data series.
- **Chapter 4** includes a presentation of the results of the particle-filter on both real and simulated data.
- **Chapter 5** contains a description and evaluation of a real-time FOREX trading strategy based on the bubble-detection model.

## II. Literature Review

The econometric literature surrounding bubbles, both in theoretical and empirical bubble-stamping methods, is vast and deep, going back at least to the early 1980s and covering every asset class and many economic indicators. The literature is mixed; while most economists agree that bubbles in stock prices and commodities exist, the evidence for bubbles in exchange rates is much less clear. While testing for exchange-rate bubbles was a hot topic in the late 20th century, the idea seems to have cooled off significantly since then. New advances in bubble-stamping techniques and recent leaps forward in computing power have made bubble-detection once again an area of interest to econometricians, and searching for bubbles in exchange rates should be part of that research.

Since the abandonment of the Bretton-Woods system, which kept exchange rates tied to the price of gold through the U.S. dollar, exchange rates have been controlled by currency markets. Currency is bought and sold, like a commodity or a stock; currency can be thought of as an asset. It makes intuitive sense, then, to expect currencies to act like other assets.

The major difference between a currency and any other asset is the role of central banks. The stability of the value of money is important to the economic stability of the country in which that currency is used, and so the main goal of central banks is to stabilize the value of their currency. The relative strength of a currency is connected to the domestic interest rate, and the manipulation of interest rates is a powerful tool of central banks. The presence of an entity with the power and mission to keep the exchange rate within a certain interval might make speculative bubbles rare or perhaps even impossible.

This chapter proceeds with a discussion of fundamental-pricing models for currencies and research in which those models were used to search for bubbles in currency prices. Then, the discussion will turn toward technical discussion of Sequential Monte Carlo (SMC) methods, also known as particle filters.

## **Economic Literature**

An early example of a possible bubble in currency value is in the Weimar Republic in the early 1920s. Flood and Garber (1980) investigated this period using the classical Cagan model of currency value fundamentals. They wrote, “The possibility of a market’s launching itself into a price bubble exists when the expected rate of market price change is an important factor determining current market price” (Flood & Garber, 1980), implying that anticipations influence market fundamentals. The study found evidence of a hyperinflationary bubble, but the authors are careful to say that this evidence is subject to model error. However, “if bubbles actually exist, then explaining both their paths and their terminations is important, and dynamic models which ignore bubbles are clearly inadequate” (Flood & Garber, 1980).

Flood and Garber and Scott (1984) compared price levels across Germany, Poland, and Hungary during the early 1920s in an investigation of possible simultaneous bubbles across the three closely-linked countries. In this study, they found no evidence of hyperinflation in Poland and no evidence of a bubble common to any combination of the three countries (Flood et al, 1984).

Okina (1984) provides an overview of several exchange-rate value models that became popular in bubble-stamping research. While none of the theoretical models have been shown to be empirically accurate, the common explanation of bubbles is as significant deviations from the fundamental value. Okina’s (1984) paper outlines what are known as the Monetary Model, the Overshooting Monetary Model, and the Portfolio Balance Model. The focus in this thesis is on the Monetary Model, which is explained in Chapter 3. Okina (1984) also outlines three mathematical forms of bubbles: Stochastic bubbles with no regeneration; stochastic bubbles with regeneration; and deterministic bubbles.

The study by Charemza (1996) is based on the theory that the fundamental relative value of a currency is proportional to the expected value of the future relative price of that currency plus monthly domestic inflation plus some measure of uncertainty.

$$p_t = \varphi_0 E_t(p_{t+1} + \varphi_1 \pi_{t+1} + \varphi_2 v_{t+1}) + e_t$$

There is weak evidence that there was a speculative bubble in Poland in the late 1980s, and that the collapse of this bubble caused exchange-rate instability in the Zloty (Charemza, 1996).

An interesting perspective, taken by Kiman, Ricciotti, and Topol (2007), for analyzing exchange rates is to consider the rate as entirely set by investors with a profit motive. They simulate foreign and domestic traders as either chartists or fundamental-value traders. The model was flexible enough to adequately describe complex market fluctuations, but is heavily affected by exogenous parameters and thus prone to specification error (Kiman et al., 2007).

Many of the popular fundamental-price models for exchange rates are compared in Macerinskiene and Balciunas (2013), who list the advantages and disadvantages of each from a theoretical perspective. It is widely agreed that purchasing power parity (PPP) does not hold, at least in the short term, and so any models that rely on the assumptions of PPP must necessarily be inaccurate in the short term. Similarly, the assumption of interest-rate parity (IRP) is thought to be problematic due to the heterogeneity of risk attitudes among investors and models that rely on IRP are also expected to be inaccurate. Unfortunately, most available fundamental-price models for exchange rates rely on at least one of these two assumptions. The compared model that is closest to the Monetary model used in this thesis is the Flexible Price Monetary Model. The major advantage to this model is that it is one of the few that include the impact of monetary policy on exchange rates (Macerinskiene & Balciunas, 2013). The drawback to this class of models is that they rely on both PPP and IRP, and thus do not explain all exchange rate movements. However, this paper supposes that the unexplained movements are the product of speculative bubbles that are additively separate from the fundamental rate, and so a lack of fit with the fundamental model is expected.

The Portfolio Balance Model from Okina (1984) is used by Woo (1987). The stochastic-bubble model accounts for fiscal and monetary policy-making that attempts to keep inflation within

a target zone. This leads to the conclusion that non-fundamental values should be either stationary or explosive. The paper shows that the model fits the DEU/USD and FRA/USD exchange rates quite well, but additional assumptions were needed to fit the model to the JPY/USD data. There was strong evidence of bubbles in each of these; the longest bubble period was from June to October 1978, in DEU/USD (Woo, 1987).

Wu (1995) also tested DEU/USD and JPY/USD for bubbles during the same period, but found no evidence. The fundamental value in the exchange rate is determined by a monetary model, rather than the portfolio balance model. After finding no evidence of bubbles, the author speculates that the Kalman filter model may suffer from accumulated error and parameter uncertainty; however, Wu (1995) also argues that accumulated error would increase the standard error and make the coefficient of interest appear even less statistically significant.

The Kalman filter approach is used regularly in finance, and research featuring its application to econometric time series is abundant. It is used in both Al-Anaswah and Wilfling (2011) and Chen and Yan (2011) to search for bubbles in stock markets. Both studies look for bubbles in the U.S. stock market; Al-Anaswah and Wilfling (2011) also test the Brazilian stock market; Chen and Yan (2011) also test the Shanghai stock exchange. The difference between these two Kalman-filtering approaches and that used by Wu (1995) is the inclusion of Markov-switching.

The empirical analysis of Al-Anaswah and Wilfling (2011) finds strong evidence of bubbles. They also find that the probability of bubble survival is significantly higher than the probability of bubble collapse, indicating that bubbles are expected to continue. The Kalman-filter with Markov-switching method is able to correctly identify the Great Depression and the recession at the beginning of the second world war, but it smoothes over Black Monday, the dot-com bubble, and the oil crisis. The study also reports evidence of the Brazilian stock market crash in the late 1990s and a crash in the Asian stock market as shown in Japanese and Malaysian stock indices (Al-Anaswah & Wilfling, 2011).

Chen and Yan (2011) report similar findings in the U.S. stock market. They also found significant evidence of bubbles in China from the summer of 1992 to the summer of 1993, in October

1994, through the second half of 1999, and toward the end of 2008 (Chen & Yan, 2011). It is interesting to note that much of the globe was in a recession in 2008, the Subprime Mortgage Crisis, due to the collapse of an enormous housing bubble in the United States.

The Kalman filter was used again to detect bubbles in the U.S. stock market in the master's thesis by Hrabovska (2017). The same technique used by Chen and Yan (2011) and Al-Anaswah and Wilfling (2011) is applied not only to the S&P 500 index but also to 17 large American companies. The analysis finds that the S&P 500 has been constantly overvalued since the 1980s; the bubble component has been positive for the last 40 years. Stock prices of individual companies, however, display much more noise. The only two crashes at the company level were at Wells Fargo and Abbott Laboratories in 2012 and 2004, respectively (Hrabovska, 2017).

More recently, the PSY bubble-stamping method, developed by Phillips, Shi, and Yu (2011), has received a lot of attention from economists of all kinds. The method is a recursive application of the Augmented Dickey-Fuller test that uses the supremum of ADF test statistics over each time period to determine if there is a bubble in that period. This is the test used by Bettendorf and Chen (2013) to analyze bubbles in the Sterling-Dollar exchange rate. This study revealed evidence of a bubble in 1976, corresponding to the well-known Sterling Crisis, and in 1985, when the U.S. dollar was appreciating quickly against many currencies.

The PSY method has been used to detect negative financial bubbles as well, where the asset price is below the fundamental value, in oil companies in 2014/15 by Fantazzini (2016). The stock prices of West Texas Intermediate and Brent Crude were analyzed using both PSY and the log-periodic power law (LPPL) devised by Yan et al. (2012). When daily prices from January 2013 to April 2015 were examined, both methods stamped bubbles in both time series. The analysis was then applied to weekly data from 2002 through 2015, and the same bubbles were stamped again, further emphasizing the evidence of negative bubbles. However, the author notes that PSY results are quite sensitive to the user-input minimum bubble length (Fantazzini, 2016).

Milunovich, Shi, and Tan (2019) brought PSY into the world of trading, using the bubble-detection method to devise a bubble-trading strategy. To implement a bubble-based portfolio

management technique, the authors focus on sector indices which display less statistical noise than individual stocks. The use of weekly, rather than daily, data allows the strategy to avoid stochastic jumps and the associated discontinuity. The basic strategy is that a trader buys the sector index when a bubble is detected, sells the index when a collapse is detected, and otherwise buys short-term treasury bills. In addition to considering the PSY bubble-detection indicator alone, Milunovich et al. consider a modified indicator which requires the market index to have a positive most recent change as well. They call this indicator PSY-MBI. They compare the PSY-Base and PSY-MBI strategies to a simple directional strategy which executes trades according to the direction of the most recent return and benchmark all strategies to the buy-and-hold strategy. “The buy-and-hold strategy is difficult to surpass, especially when transaction costs are taken into consideration, and therefore serves as a powerful benchmark” (Milunovich, Shi, & Tan, 2019). Round-trip transaction costs of 0.5% are assumed.

The PSY-MBI far outperforms the others. With the optimal 99% confidence level, the final wealth of the PSY-MBI strategy is over thrice the final wealth of PSY-Base, and the Sharpe ratio is approximately 2 percentage points higher. Both PSY strategies outperform buy-and-hold, and the directional strategy loses all equity. Bubbles are detected in all eleven sectors (Milunovich et al., 2019).

The final bubble-stamping strategy to be discussed, and the strategy used in this research, is particle-filter analysis. The methodology in this thesis closely resembles the methodology of Fulop and Yu (2017), who compare the  $SMC^2$  particle filter with the PSY method for bubble detection in the S&P 500. The filter itself is described in detail in Chapter 3, and will be discussed in detail later in this chapter. The next few paragraphs discuss the econometric results of the implementation.

The time series used in Fulop and Yu (2017) is the price-dividend ratio of the S&P 500 as collected by Robert Shiller. According to the present-value model of fundamentals, if the log-dividends  $d_t$  is an  $I(1)$  process, then the fundamentals  $f_t$  must be  $I(1)$  as well. This has two implications:

- If there is no bubble, then the log-price function  $p_t = f_t + b_t$  implies  $p_t = f_t$ , and so the price must be  $I(1)$ .

- If there is an explosive bubble, then  $p_t - f_t = b_t$  implies that  $p_t$  must be explosive, and so must  $p_t - d_t$ , which is the log-price-dividend ratio.

The reliability of the algorithm was tested by Monte Carlo methods; 100 datasets were simulated from the proposed regime-switching model, each with 1698 observations. Then, the particle filter was run on these simulated data series with  $M = 2048$   $\theta$ -particles and  $N = 128$  state-particles (Fulop & Yu, 2017). The results of the test showed significant tightening of the confidence intervals for parameters, with all true parameters falling within the 90% confidence interval of the posterior mean (Fulop & Yu, 2017).

Fulop and Yu (2017) used a loss function to determine when bubbles should be stamped. The loss function is the ratio of the bubble-probability to the non-bubble probability, and if this ratio surpasses a chosen level  $\zeta$  the period is stamped. In the Monte Carlo analysis,  $\zeta = 1$  led to overestimation of bubble periods and the particle filter performed worse than PSY. However, with  $\zeta = 2$ , the  $SMC^2$  method stamped closest to the known number of bubbles despite slightly underestimating the number of bubbles. Neither PSY nor  $SMC^2$  accurately stamped the correct total bubble length, and both methods underestimated the average bubble duration (Fulop and Yu, 2017).

On the real S&P 500 data, the confidence intervals for the parameters tightened over the course of analysis. The algorithm returned high estimates for the probability of regime persistence, with the expected length of a normal regime near 147 months and the expected length of the bubble regime at 31 months. The Markov-switching probabilities for volatility regimes were similar. With  $\zeta = 2$ , a total of 24 bubbles were found in the S&P 500 from 1874-2015, with an average bubble length of 9.7 months. These included the banking crisis in October 1907, the crash which started the Great Depression in September 1929, Black Monday in October 1987, the DotCom crash in March 2000, and the subprime mortgage crisis in 2008. Importantly, the particle filter is prone to interpret post-crash rebounds as new bubbles (Fulop and Yu, 2017).

## Particle-Filter Literature

Particle filters, also called Sequential Monte Carlo (SMC) methods, are sampling-based techniques for approximating posterior distributions, often in the context of Hidden Markov Models (HMMs). SMC methods can approximate posterior distributions in far more complex HMMs (Doucet, de Freitas, & Gordon, 2001). “All [particle filters] approach the filtering problem from a sampling perspective, with the aim being to generate a random sample from the true posterior distribution” (Fearnhead, 1998). A very basic SMC algorithm, given by Doucet, de Freitas, and Gordon (2001), is comprised of these two steps:

- Assume that  $y_t$  is the observed measurement at time  $t$  and  $x_t$  is the hidden state at time  $t$ .
- *Prediction Step:*

$$P(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}$$

- *Updating Step:*

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

The integrals involved are generally intractable and have to be approximated (Fearnhead, 1998; Andrieu, Doucet, & Punsakaya, 2001). There are many similar algorithms, and a few will be discussed here.

The main benefit of SMC rather than traditional Monte Carlo (MC) methods is that new observations can be added without the need to recalculate the entire approximation; they are less computationally expensive than batch methods when continuous measures are taken. The downside, however, is that error accumulates with each additional measurement. Therefore, SMC methods are expected to degenerate over time. Andrieu, Doucet, and Punsakaya (2001) conclude that the number of particles must grow as more measurements are observed, but Fearnhead (1998), in his doctoral dissertation, argues that this degeneration can be postponed, if not avoided, by the purposeful selection of investigated states. “A few well chosen points will give a better estimate than a much larger number of points chosen at random” (Fearnhead, 1998). In fact, when there are only a small number of possible values for the hidden state, the best estimate can be found by propagating the particles through every possible state. If there are  $S$  possible states, then this method would increase

the number of particles by a factor of  $S$  at each measurement and the algorithm would quickly become too computationally expensive for even the most powerful computers; therefore, the original number of particles  $N$  should be resampled from the  $SN$  particles according to normalized weights after posterior probabilities are calculated at each timestep (Fearnhead, 1998; Fulop & Yu, 2017). This is known as the Discrete Particle Filter (DPF).

Commonly, the probability distributions required for the update step are either not known or are too complex to be sampled from efficiently. The usual way to overcome this problem is with the use of Importance Sampling (IS). In IS, a simpler distribution is chosen to represent the true target and samples are selected from this new distribution, then weighted according to the target distribution. For this to work, every possible non-zero value under the target distribution must have a non-zero probability under the importance distribution. While particle filters using IS have been proven convergent, approximations based on importance sampling are biased, and get more and more skewed over time (Doucet et al., 2001). For Sequential Importance Sampling (SIS) methods to be efficient, particle weights should be close to uniform. When weights are skewed, particles with high weights are resampled too often and the sample will be depleted; however, the sample can be regenerated using Markov-chain Monte Carlo (MCMC) methods (Andrieu et al., 2001). Berquini and Gilks (2001) describe a technique for regenerating the set of particles using the Metropolis-Hastings (MH) algorithm, a common MCMC method. An occasional MH update, reviewing the entirety of the time series up to time  $t$ , redisperses the particles and allows them to adapt to evolution in the target distribution (Berquini & Gilks, 2001). This can allow an otherwise degenerative particle filter to become ergodic (Andrieu et al., 2001), especially when the hidden state involves multiple models, like the system in this thesis.

“In applications involving multiple models, the particles are scattered in a complex parameter space, which is the union of model-specific sub-spaces. In this context, a merit of the hybrid particle filters ... is that they allow particles to jump from one model subspace to another. This may be useful to recover from occasional depletion of some subspaces, and moreover, in those applications involving nested models, to take advantage

of relationships between parameters of different subspaces” (Berquini & Gilks, 2001).

When a system being analyzed is non-trivially dependent on the set of parameters,  $\theta$ , the asymptotic behavior of SMC methods can vary widely (Andrieu et al., 2001). Methods for accurately estimating parameters can have a large effect on the outcome of the particle filter as a whole. Until 2010, most of the literature relied on Maximum Likelihood Estimates (MLE) of the parameters, as it is computationally easier and more natural than Bayesian estimation (Svensson & Schon, 2016). However, the accuracy of MLE depends on the sample size from which estimates are calculated, and so sequential estimation is typically inaccurate or impossible.

A major breakthrough that had a tremendous impact on parameter-estimation techniques in particle filters is the work of Andrieu, Doucet, and Holenstein (2010) in which they describe a sequential adaptation of MCMC methods, dubbed Particle Markov-chain Monte Carlo (PMCMC) methods. “[PMCMC] methods rely on a non-trivial and non-standard combination of MCMC and SMC methods which takes advantage of the strength of its two components” (Andrieu et al., 2010). They offer particle-filter versions of the Marginal Metropolis-Hastings sampler, the Independent Metropolis-Hastings sampler, and the Gibbs sampler. Of importance to this paper is the Particle Marginal Metropolis-Hastings (PMMH) sampler, and it is the only PMCMC method that will be discussed here.

The intuition behind the PMMH sampler is to use the Bayesian decomposition  $p(\theta, x_{1:T}|y_{1:T}) = p(\theta|y_{1:T})p_{\theta}(x_{1:T}|y_{1:T})$  to draw a sample from the joint distribution of states and parameters. The traditional MH sampler compares the likelihood of two importance-sampled sets of states and accepts the most likely, repeating until convergence. The PMMH sampler compares the likelihoods of two particles up to time  $T$  under the importance distribution  $q$ :

$$q(\{\theta^*, x_{1:T}^*\}|\{\theta, x_{1:T}\}) = q(\theta^*|\theta)p_{\theta^*}(x_{1:T}^*|y_{1:T})$$

When  $x_{1:T}^*$  is perfectly adapted to  $\theta^*$ , the only degree of freedom of the proposed distribution that affects performance is the choice of  $q(\theta^*|\theta)$  (Andrieu et al., 2010). The acceptance distribution for the MH part of the PMMH sampler is the ratio of marginal-likelihoods of the particles times the

importance likelihoods:

$$\frac{p(\theta^*, x_{1:T}^* | y_{1:T})q(\{\theta, x_{1:T}\} | \{\theta^*, x_{1:T}^*\})}{p(\theta, x_{1:T} | y_{1:T})q(\{\theta^*, x_{1:T}^*\} | \{\theta, x_{1:T}\})}$$

This complex ratio can be decomposed and simplified to:

$$\frac{p_{\theta^*}(y_{1:T})p(\theta^*)q(\theta | \theta^*)}{p_{\theta}(y_{1:T})p(\theta)q(\theta^* | \theta)}$$

And when the importance sampling distribution  $q$  is symmetric, this can be further simplified to:

$$\frac{p_{\theta^*}(y_{1:T})p(\theta^*)}{p_{\theta}(y_{1:T})p(\theta)}$$

Which is just the marginal likelihood of  $y_{1:T}$  under each set of parameters and the prior marginal likelihood of that set of parameters. This final form of the MH acceptance ratio shows that the target of the PMMH sampler is in fact  $p(\theta | y_{1:T}) \propto p_{\theta}(y_{1:T})p(\theta)$ , and so the algorithm will converge to the marginal distribution of  $\theta$  just like a traditional MH sampler would (Andrieu et al., 2010). The particle part of PMMH is the use of particle-filter approximations for  $p_{\theta}(y_{1:T})$  in the acceptance ratio. The algorithm for this sampler is shown in detail in Chapter 3.

PMCMC methods create unbiased estimates of the target distribution and are ergodic under weak assumptions (Andrieu et al, 2010), but can be computationally expensive and are not suitable to on-line analysis. They also have an initial transient period where estimates vary widely and are typically quite inaccurate. One thought on how to avoid this problem is to discard the first  $n$  estimates; this is not a solution, as there is no way to tell how long the “burn-in” period will be and therefore no way to tell if too many or too few of the estimates have been discarded (Svensson & Schon, 2016). Causing further problems, once PMCMC algorithms get past the initial transience they often get stuck; they explore one small area of the possible space and reject all proposals (Svensson & Schon, 2016).

Despite the drawbacks of the PMMH sampler, it is this algorithm that makes sequential joint Monte Carlo estimation of states and parameters possible in the algorithm known as  $SMC^2$ , concurrently devised by Fulop and Li (2013) and Chopin, Jacob, and Papaspiliopoulos (2013). The name  $SMC^2$  is used by Chopin et al. but not by Fulop & Li; the algorithm requires both a

parameter-learning filter and a nested state-learning particle filter, and that is the inspiration for the name.

Fulop and Li (2013) call the algorithm “a marginalized resample-move approach,” which is an appropriate and descriptive name. In order to update the Bayesian probability assigned to each  $\theta$ -particle, dependence between parameters and states is broken by marginalizing out the states. Running a state-particle filter creates a set of auxiliary random variables; for instance, the particle filter in this thesis creates variables that represent the length of time since the last regime change, which are samples from a geometric distribution, and it is these variables that are used to reweight the  $\theta$ -particles.

The resample-move portion of  $SMC^2$  is quite similar to the resample-move sample-rejuvenation technique proposed by Berquini & Gilks (2001). Berquini and Gilks (2001) recommend a MH sampler as a way to replenish a particle filter with degenerating sample size, and in  $SMC^2$  this idea is done sequentially via PMMH (Chopin et al., 2013; Fulop & Li, 2013; Andrieu et al., 2010). As the purpose of the resample-move step is to replenish the ESS, it is not needed at every timestep. Rather, it is recommended only when  $ESS < B$ , where  $B$  is set by the user. Fulop and Li (2013) recommend  $B = \frac{M}{2}$ , or half the total sample size, as it balances the need for continued parameter-learning with the computationally demanding PMMH update.

Like other SMC algorithms,  $SMC^2$  has a tendency to degenerate over a large number of observations. This degeneracy is slowed by the resample-move step; the PMMH update is needed less often over time, but it also becomes much more demanding. Chopin et al. (2013) discuss the computation cost at length. “The  $SMC^2$  algorithm is memory intensive: up to iteration  $t$ ,  $O(tMN)$  variables have been generated and potentially must be carried forward to the next iteration” (Chopin et al., 2013). Over time, the computation cost approaches  $O(Nt^2)$ ; however, only the most recent particles need to be kept and so, under specific strict conditions, the computation cost can be held to  $O(MN)$  (Chopin et al., 2013).

Svensson and Schon (2016) compare  $SMC^2$  to PMMH. If  $K$  is the number of iterations,  $T$  is the number of observations, and  $N_x$  is the number of particles, then the PMMH algorithm is

governed by  $O(KTN_x)$ , but  $N_x$  can be optimally chosen so that the computation cost does not vary with  $T$ , in which case the cost is governed by  $O(KT^2)$  (Svensson & Schon, 2016). Alternatively, the computation cost of  $SMC^2$  is governed by  $O(KT^2N_xN_\theta)$ , or with  $N_x$  chosen to keep the acceptance rate invariant with  $T$ ,  $O(KT^3N_\theta)$ . Thus, although  $SMC^2$  can be used for on-line estimation, its computation cost grows prohibitively fast with  $T$  and so is typically not usable in situations with real-time requirements. In general, PMMH may be computationally advantageous with large  $T$ , while the two methods are comparable when  $T$  is not large; the advantage of  $SMC^2$  is that it offers far more flexibility in tuning (Svensson & Schon, 2016).

Chopin et al. (2013) also compare  $SMC^2$  to other filtering algorithms, including PMMH. They first argue that “even in batch estimation,  $SMC^2$  may offer several advantages over PMCMC, in the same way that SMC approaches may be advantageous over MCMC methods” (Chopin et al., 2013), and post-testing conclude that “ $SMC^2$  was shown typically to outperform competing algorithms, whether in sequential scenarios (where data points are obtained sequentially) or in batch scenarios” (Chopin et al., 2013). Specifically,  $SMC^2$  had smaller Monte Carlo error than other tested filters when computation cost was held equal.

Fulop and Li (2013), in their paper proposing  $SMC^2$ , run a Monte Carlo test for the accuracy of the algorithm. They created a set of 100 simulated time series from a linear Gaussian distribution (the exact states and parameters could be found by a Kalman filter) and ran  $SMC^2$  on the simulated data with a varying number of state-particles and  $\theta$ -particles. With only 100 state-particles and 500  $\theta$ -particles, the algorithm obtained similar accuracy (measured by RMSE) to an MCMC algorithm with 10,000 iterations (of which the first 4,000 were discarded for the “burn-in” period). The algorithm is accurate even when with stochastic jumps and stochastic volatility in the dataset (Fulop & Li, 2013).

The  $SMC^2$  algorithm, with Markov-switching, was chosen for this thesis because of its flexibility and usability for on-line analysis. In the words of Chopin et al. (2013), “ $SMC^2$  is particularly well suited to tackle several of the challenges that arise in the probabilistic modelling of financial time series.”

### III. Methodology

#### Fundamental Model for the Exchange Rate

The time series of exchange rates cannot be input into the model directly; there may be legitimate, non-speculative reasons for the exchange rate to behave explosively. Thus, the fundamental value of the underlying asset must be removed from the exchange rate first. There are many ways to derive a “fundamental” value, and in this paper the *Small Country Monetary Model* is selected due to the ease of collecting the required data. This model was first proposed by Michael Mussa in 1980, and is described in detail in Okina (1984).

The biggest shortcoming of this model is that it relies on the assumption of Purchasing Power Parity (PPP), which is not empirically valid. Exchange rates often deviate from PPP significantly and for extended periods. However, when discussing the fundamental relative-value of currencies, PPP makes intuitive sense: Two currencies should have the same purchasing value, especially considering the near-complete internationalization of most industries. The mathematical representation of purchasing power parity is

$$P(t) = X(t)P^*(t)$$

wherein  $P(t)$  is the domestic price level at time  $t$ ,  $X(t)$  is the exchange rate at time  $t$ , and  $P^*(t)$  is the foreign price level at time  $t$ . Natural logarithms are used for the remainder of this section, to allow algebraic manipulation of the variables; this is shown by lower-case letters.

$$p(t) = x(t) + p^*(t) \tag{1}$$

$$i(t) - i^*(t) = E_t[x(t+1)] - x(t) \tag{2}$$

$$m(t) = p(t) + ay(t) - ki(t) \tag{3}$$

Equation 2 is the interest parity condition, which states that the ratio of interest rates between two countries should be equal to the ratio of the expected future exchange rate to the present exchange rate. This comes from the idea that an investor can purchase foreign currency in order to gain the foreign interest rate on his or her holdings, if that interest rate is higher than the domestic rate. The idea is used regularly by corporations to hedge investments in foreign countries.

- $i(t)$  is the domestic interest rate at time  $t$
- $i^*(t)$  is the foreign interest rate at time  $t$
- $E_t[x(t + 1)]$  is the expected future value of the exchange rate, given all of the information available at time  $t$

Equation 3 is the money market equilibrium condition, which describes the expected money supply. The money supply is controlled by central banks, and is adjusted to encourage or discourage spending and inflation. Essentially, this equation says that money supply should be determined by the price level, times some proportion of income, divided by the interest rate times the interest rate semi-elasticity.

- $m(t)$  is the money supply at time  $t$
- $y(t)$  is real income at time  $t$
- $k$  is the interest rate semi-elasticity

The derivation of the fundamental value of the exchange rate can now proceed. Rearrange Equation 2;

$$i(t) = E_t[x(t + 1)] - x(t) + i^*(t)$$

Substitute this into Equation 3;

$$m(t) = p(t) + ay(t) - k [E_t[x(t + 1)] - x(t) + i^*(t)]$$

Now rearrange this equation and distribute  $k$ ;

$$m(t) - p(t) - ay(t) = -kE_t[x(t + 1)] + kx(t) - ki^*(t)$$

Get  $x(t)$  alone on one side;

$$m(t) - p(t) - ay(t) + ki^*(t) + kE_t[x(t + 1)] = kx(t)$$

Now, rearrange Equation 1;

$$x(t) = p(t) - p^*(t)$$

And add this to both sides,  $p(t) - p^*(t)$  on the left,  $x(t)$  on the right;

$$m(t) - p^*(t) - ay(t) + ki^*(t) + kE_t[x(t+1)] = (k+1)x(t)$$

Divide by  $(k+1)$  to get  $x(t)$  alone;

$$x(t) = \frac{k}{k+1}E_t[x(t+1)] + \frac{1}{k+1}[m(t) - p^*(t) - ay(t) + ki^*(t)]$$

Finally, let  $z(t) = m(t) - p^*(t) - ay(t) + ki^*(t)$  and rewrite this equation as;

$$x(t) = \frac{k}{k+1}E_t[x(t+1)] + \frac{1}{k+1}z(t) \tag{4}$$

This function falls into the Rational Expectations Asset Pricing Model

$$x(t) = aE_t[x(t+1)] + bz(t)$$

wherein  $z(t)$  is the fundamental value of the asset, with  $a = \frac{k}{k+1}$  and  $b = \frac{1}{k+1}$ . The idea that  $z(t)$  is the fundamental value of this asset will be shown again. According to Okina (1984), the fundamental value can be found by the series

$$\bar{x} = b \sum_{i=0}^{\infty} a^i E_t[z(t+i)]$$

With  $a$  and  $b$  from Equation 4;

$$\bar{x} = \frac{1}{k+1} \sum_{i=0}^{\infty} \left(\frac{k}{k+1}\right)^i E_t[z(t+i)]$$

Suppose  $z(t)$  has the martingale property, such that  $E_t[z(t+i)] = z(t)$ , or that things continue as they are;

$$\bar{x} = \frac{1}{k+1} \sum_{i=0}^{\infty} \left(\frac{k}{k+1}\right)^i z(t)$$

Finally, include the convention  $0 < k < 1$  to find the geometric series;

$$\bar{x} = \frac{z(t)}{k+1} \sum_{i=0}^{\infty} \left(\frac{k}{k+1}\right)^i$$

The solution to the sum is;

$$\begin{aligned}\sum_{i=0}^{\infty} \left(\frac{k}{k+1}\right)^i &= \left(\frac{1}{1 - \frac{k}{k+1}}\right) \left(\frac{k+1}{k+1}\right) \\ &= \frac{k+1}{k+1-k} \\ &= k+1\end{aligned}$$

Thus, the final solution for the fundamental value is;

$$\bar{x} = \frac{z(t)}{k+1}(k+1) = z(t)$$

The settled fundamental value of the asset which must be removed from the exchange rate is

$$z(t) = m(t) - ay(t) - p^*(t) + ki^*(t).$$

## Two Time Series

In a Dickey-Fuller test, the stationarity of a time series is determined by the presence of a unit root. For example, in the following AR(1) model:

$$y_t = \rho y_{t-1} + u_t$$

Taking the first difference:

$$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + u_t$$

A unit root, meaning  $\rho = 1$ , would leave only the noise term on the right side, making the time series a random walk:

$$\Delta y_t = u_t$$

However, if the coefficient  $\rho$  is not equal to 1, the series would still have a  $y_{t-1}$  term on the right, indicating that the change from one time-step to the next is dependent on the value of the series at the previous time step. A coefficient less than one would leave a negative  $y_{t-1}$  term and is indicative of a stationary time series. A coefficient greater than one leaves a positive term and indicates explosive growth. In the literature, there are two main “types” of asset price movement, trended or mean-reverting (Rockefeller, 2014). While unit root tests can be altered to allow for trend-stationarity,

in the context of this paper the time series is classified as either mean-reverting or explosive. These two models of series movement correspond to either the “normal” or the “bubble” regime, and the probability of falling into one of these two regimes is the main focus of the state-particle filter. The time series being analyzed is assumed to follow one of the two following models.

In the mean-reverting regime:

$$x_t = \alpha_t(1 - \beta_1) + \beta_1 x_{t-1} + \sigma_t \epsilon_t, \quad \beta_1 \leq 1 \quad (5)$$

In the bubble regime:

$$x_t = \beta_2 x_{t-1} + \sigma_t \epsilon_t, \quad \beta_2 > 1 \quad (6)$$

These are the same process equations used by (Fulop & Yu, 2017). The parameters are defined as follows:

- $\alpha_t$  is the long-term mean to which the time series reverts. For the model in this paper, a moving average is used. The window of the moving average is chosen by the researcher; it should be long enough to smooth out the noise, but short enough to capture the natural movements of the time series. Some common moving-average windows among technical traders are 20, 50, 100, and 200 time-steps.
- $\beta_1$  is the coefficient of the previous value of the time series ( $\rho$  in the Dickey-Fuller example at the beginning of this section). For a mean-reverting time series, this value must be less than or equal to one. In the proposed model, this value is determined by the  $\theta$ -particle filter described later in this chapter.
- $\sigma_t$  is the standard deviation of the noise term. Like the regime, this is allowed to take two values, either a “high-volatility” or “low-volatility” value, and is present in the state-particle filter. The two values are determined by the  $\theta$ -particle filter described later in this chapter.
- $\epsilon_t$  is NID(0,1), and represents the random noise present in the time series
- $\beta_2$  is the coefficient of the previous value of the time series for the bubble regime. Since the bubble is defined by its explosivity,  $\beta_2$  must be greater than one. This value is determined by the  $\theta$ -particle filter described later in this chapter.

## Hidden Markov Model

The proposed model takes advantage of a Hidden Markov Model (HMM). As in all HMMs, it is supposed that the observed time series is dependent on some unobservable series that is defined to have the Markov property. In this model, the hidden time series contains only the discrete values that identify the regime as either normal or bubble ( $s_t = 0$  or  $s_t = 1$ , respectively) and the volatility as either “low” or “high” ( $\hat{\sigma}_t = 0$  or  $\hat{\sigma}_t = 1$ ). For the volatility,  $\hat{\sigma}_t = 0$  indicates that  $\sigma_t$  in the process model should be  $\sigma_l$ , the low value of volatility, while  $\hat{\sigma}_t = 1$  indicates that  $\sigma_t = \sigma_l \sigma_m$  should be used in the process model, where  $\sigma_m > 1$  is the ratio of high-volatility to low-volatility. Thus, there are four possible values for the hidden state, shown in Table 1.

The state is probabilistically determined in the state-particle filter described in the next section of this chapter. At a given time step, the value of the hidden Markov chain determines which of the two regime-process models from the previous section should be used, as well as which of two possible values should be used as  $\sigma_t$  in those equations. A long-form equation for the whole model would be:

$$x_t = (\beta_1 x_{t-1} - (1 - \beta_1) \alpha_t) (s_t - 1) + \beta_2 x_{t-1} s_t + \sigma_l (\sigma_m)^{\hat{\sigma}_t} \epsilon_t \quad (7)$$

This notation gives the appropriate process model for each of the four possible hidden states. Although this is a satisfactory equation, the model is much easier to handle as two separate processes with two possible volatility values. In the program and throughout this paper, the model is treated in the easier-to-handle fashion, rather than using Equation 7.

It is important to remember that the main goal of the model in this thesis is to determine when the asset is in a bubble, and so the most important output of the entire thing is the probability that  $s_t = 1$ .

### Before We Continue

The algorithm used in this paper consists of two nested particle-filtering algorithms. It is presented “inside-out”; that is, the first particle filter, referred to as the state-particle filter, depends on the set of parameters,  $\theta$ , and is used to determine the probability that the time series is in any one of the four possible states presented in Table 1; the second particle filter, referred to as the  $\theta$ -particle

Normal Regime, High Volatility $(s_t, \hat{\sigma}_t) = (0, 1)$ $x_t = (1 - \beta_1)\alpha_t + \beta_1 x_{t-1} + \sigma_l \sigma_m \epsilon_t$	Bubble Regime, High Volatility $(s_t, \hat{\sigma}_t) = (1, 1)$ $x_t = \beta_2 x_{t-1} + \sigma_l \sigma_m \epsilon_t$
Normal Regime, Low Volatility $(s_t, \hat{\sigma}_t) = (0, 0)$ $x_t = (1 - \beta_1)\alpha_t + \beta_1 x_{t-1} + \sigma_l \epsilon_t$	Bubble Regime, Low Volatility $(s_t, \hat{\sigma}_t) = (1, 0)$ $x_t = \beta_2 x_{t-1} + \sigma_l \epsilon_t$

Table 1: Table of States

filter, uses auxiliary variables created by the state-particle filter as probabilistic observations and is used to draw a sample from the joint distribution of parameters. This is explained in detail in the following sections, but was worth explaining here as it can seem confusing at first.

## Parameters

The parameters, which are elements in the vector  $\theta$ , decide the behavior of the state-particle filter and thus directly affect the determined probability of each state. There are eight parameters necessary for the algorithm.

- $p_n$  - The length of a normal regime is determined by a geometric distribution; each time step is independent, which is reasonable if the time steps are sufficiently far apart in real-time, and at any given timestep the probability of changing from a normal regime to a bubble regime is given by the parameter  $p_{normal}$ .
- $p_b$  - Similarly, the length of a bubble regime is described by a geometric distribution; each time step is again assumed to be independent, and the probability of returning from a bubble regime to a normal regime at any given timestep is the parameter  $p_{bubble}$ .

The use of geometric distributions to time regime changes allows sufficient flexibility for unexpectedly fast switches and for unexpectedly long-lasting regimes. For a large sample size, meaning a time series with many regime switches, the geometric distributions will approximate the exponential distributions used by Fulop and Yu (2017). However, geometric distributions more closely align with the discrete nature of the time series and are, in the author’s opinion, the more intuitive

distribution. The probability of switching from a normal regime to a bubble regime has the Markov transition matrix:

$$P_s = \begin{bmatrix} 1 - p_n & p_n \\ p_b & 1 - p_b \end{bmatrix}$$

- $\sigma_l$  - As previously stated, the model assumes that the analyzed time series falls into either a low-volatility regime or a high-volatility regime. The value of  $\sigma_l$  is used as the standard deviation of the random term at the end of each process equation during low-volatility regimes.
- $\sigma_m$  - This parameter is the ratio of the high-volatility standard deviation to the low-volatility standard deviation. That is; if the standard deviation of the random term is  $\sigma_l$  in the low-volatility periods, then the standard deviation during the high-volatility periods is given by  $\sigma_l \sigma_m$ .
- $p_l$  - The length of a low-volatility regime is proposed to follow a geometric distribution. At any timestep, during a low-volatility regime, the probability of switching to a high-volatility regime (without changing from normal to bubble or vice verse) is  $p_{low}$ . This is an appropriate proposal for the same reasons given for  $p_{normal}$ .
- $p_h$  - For the same reasons and by the same arguments, the length of a high-volatility regime is said to follow a geometric distribution, and the probability of switching from high-volatility to low-volatility is  $p_{high}$ . This gives the volatility-regimes the transition matrix:

$$P_\sigma = \begin{bmatrix} 1 - p_l & p_l \\ p_h & 1 - p_h \end{bmatrix}$$

Switches from low-volatility to high-volatility regimes are independent of switches from normal regimes to bubbles regimes. This allows a switch from a high-volatility normal regime to a low-volatility normal regime, or from a low-volatility bubble regime to a low-volatility normal regime, to occur according to this transition matrix.

$$P_{s,\sigma} = \begin{bmatrix} (1 - p_n)(1 - p_l) & (1 - p_n)p_l & p_n(1 - p_l) & p_np_l \\ (1 - p_n)p_h & (1 - p_n)(1 - p_h) & p_np_h & p_n(1 - p_h) \\ p_b(1 - p_l) & p_bp_l & (1 - p_b)(1 - p_l) & (1 - p_b)p_l \\ p_bp_h & p_b(1 - p_h) & (1 - p_b)p_h & (1 - p_b)(1 - p_h) \end{bmatrix} \begin{bmatrix} (s_t, \sigma_t) = (0, 0) \\ (0, 1) \\ (1, 0) \\ (1, 1) \end{bmatrix}$$

- $\beta_1$  - The mean-reverting process corresponding to  $s_t = 0$  has an auto-regressive term and a moving-average term, making it an ARMA(1) process. The estimated coefficient for this auto-regressive term is  $\beta_1$ . If this coefficient is greater than one, the autoregressive term would overwhelm the mean-reverting term  $(1 - \beta_1)\alpha$ , and so it must be limited to  $\beta_1 \leq 1$ .
- $\beta_2$  - The explosive process corresponding to  $s_t = 1$  only has an auto-regressive term and a random term, making it an AR(1) process. To make it a properly explosive regime, the coefficient for this auto-regressive term must be limited to  $\beta_2 > 1$ .

In summary, the parameters necessary for the state-particle filter are the two auto-regressive coefficients  $\beta_1$  and  $\beta_2$ , the four switching-probabilities  $p_n$ ,  $p_b$ ,  $p_l$ , and  $p_h$ , the standard deviation of the random term,  $\sigma_l$ , and the ratio of low-volatility to high-volatility,  $\sigma_m$ . Thus, the vector  $\theta$  is:

$$\theta = [p_n, p_b, \sigma_l, \sigma_m, p_l, p_h, \beta_1, \beta_2]$$

## Auxiliary Variables

The parameter-learning algorithm, the  $\theta$ -particle filter, uses auxiliary variables to find the posterior probability of the parameter vector  $\theta$ . There are two auxiliary variables that are created by the state-particle filter:

- $h_t$  - This variable is the time since the last normal-to-bubble or bubble-to-normal switch. Its creation is described in detail in the next section.  $h_t$  can be used to find the probabilities  $P(p_l|h_t)$  and  $P(p_b|h_t)$ , depending on the regime at time  $t$ . How this auxiliary variable is used in the  $\theta$ -particle filter is described later.
- $u_t$  - This variable is similar to  $h_t$ ; it measures the number of timesteps since the last high-to-low volatility or low-to-high volatility switch. It can be used to find  $P(p_l|u_t)$  and  $P(p_h|u_t)$ , and its use by the  $\theta$ -particle filter is described in detail in a later section.

## State-Particle Filter

As explained earlier in this chapter, the time series being analyzed can be in one of four possible states. The small number of states makes this a perfect candidate for the Discrete Particle

Filter designed by Fearnhead (1998). This method is considered “optimal” because it explores every possible state rather than a random sample of states, with the unlikely states being filtered out during the resampling step. In the following explanation of the inner mathematics of the filter, dependence on the parameter vector  $\theta$  is suppressed to ease notation. Every step in the state-particle filter is dependent on  $\theta$ .

Suppose at time  $t$  there are  $N$  equally-weighted particles  $(s_t^i, \hat{\sigma}_t^i, h_t^i, u_t^i, w_t^i)$ ,  $i = 1, 2, \dots, N$ , in which  $w_t^i$  is the weight of the particle and all other elements are as previously defined. The goal of the particle filter is to track  $f(s_t, \hat{\sigma}_t | x_{1:t})$ , dependent on  $\theta$ , and the filter can be moved forward one timestep with the following recursion:

### Branching Out

At time  $t + 1$ , there are four potential states, shown in Table 1. The Discrete Particle Filter branches out into each possible state. Thus, four new particles must be created for each existing particle, one with each combination of  $s_{t+1}$  and  $\hat{\sigma}_{t+1}$ . These particles are

$$(s_{t+1}^{i,j} = k, \hat{\sigma}_{t+1}^{i,j} = l, h_{t+1}^{i,j}, u_{t+1}^{i,j}, w_t^i), j = \{1, 2, 3, 4\}, k = \{0, 1\}, l = \{0, 1\}$$

. The particle weight is not yet updated. The auxiliary variables are updated thusly:

- $h_{t+1}^{i,j}$  - For each new branch-particle, if  $s_{t+1}^{i,j} = s_t^i$ , then  $h_{t+1}^{i,j} = h_t^i + 1$ , while if  $s_{t+1}^{i,j} \neq s_t^i$ , then  $h_{t+1}^{i,j} = 0$ .
- $u_{t+1}^{i,j}$  - For each new branch-particle, if  $\hat{\sigma}_{t+1}^{i,j} = \hat{\sigma}_t^i$ , then  $u_{t+1}^{i,j} = u_t^i + 1$ . Otherwise, if  $\hat{\sigma}_{t+1}^{i,j} \neq \hat{\sigma}_t^i$ , then  $u_{t+1}^{i,j} = 0$ .

To update the weights, let the function  $f(s_{t+1}, \hat{\sigma}_{t+1} | s_t, \hat{\sigma}_t)$  give the probability of the new state  $(s_{t+1}^{i,j}, \hat{\sigma}_{t+1}^{i,j})$  given the previous state  $(s_t^i, \hat{\sigma}_t^i)$ . It is obvious that  $f(h_{t+1}, u_{t+1} | h_t, s_t, u_t, \hat{\sigma}_t)$  is equivalent to the state-probability function in the previous sentence. The values of this probability function follow the total state transition matrix.

Then, the new weights become:

$$w_{t+1|t}^{i,j} = f(s_{t+1}^{i,j}, \hat{\sigma}_{t+1}^{i,j} | s_t^i, \hat{\sigma}_t^i) w_t^i \quad (8)$$

This is the new prior distribution, and does not take the new observation  $x_{t+1}$  into account.

Using the probability function for the new observation,  $f(x_{t+1} | s_{t+1}, \hat{\sigma}_{t+1})$ , the posterior probability has the relationship

$$f(s_{t+1}, \hat{\sigma}_{t+1} | x_{1:t+1}) \propto f(x_{t+1} | s_{t+1}, \hat{\sigma}_{t+1})f(s_{t+1}, \hat{\sigma}_{t+1}, s_t, \hat{\sigma}_t | x_{1:t})$$

This relationship can be exploited by assigning each particle the unnormalized weight

$$\hat{w}_{t+1}^{i,j} = f(x_{t+1} | s_{t+1}^{i,j}, \hat{\sigma}_{t+1}^{i,j})w_{t+1|t}^{i,j} \quad (9)$$

At this point, the marginal likelihood of  $x_{t+1}$  can be approximated by

$$\hat{p}(x_{t+1} | x_{1:t}) = \frac{1}{4N} \sum_{i=1}^N \sum_{j=1}^4 \hat{w}_{t+1}^{i,j} \quad (10)$$

This likelihood must be saved at each time step, as it is used by the  $\theta$ -particle filter when the Particle Marginal Metropolis-Hastings update must be performed. In practice, the marginal likelihood becomes quite close to zero and so the loglikelihood is used.

## Resampling

The *Branching Out* step created a total of  $4N$  particles, and if it were repeated the number of particles would grow exponentially and quickly become computationally unmanageable. The particles will now be resampled down to the original sample size,  $N$ . Before resampling, the weights need to be normalized. The normalized weights are:

$$w_{t+1}^{i,j} = \frac{\hat{w}_{t+1}^{i,j}}{\sum_{i=1}^N \sum_{j=1}^4 \hat{w}_{t+1}^{i,j}} \quad (11)$$

Once  $N$  particles have been resampled according to the normalized weights, these resampled particles are reset to equal weights  $w_{t+1}^i = \frac{1}{N}$ . The particles at time  $t + 1$  consist of  $(s_{t+1}^i, \hat{\sigma}_{t+1}^i, h_{t+1}^i, u_{t+1}^i, w_{t+1}^i t^i)$ ,  $i = 1, 2, \dots, N$ , and the state-particle filter is ready to continue to the next timestep.

The probability of being in any state at time  $t$  is approximated by

$$P(s_t = k, \hat{\sigma}_t = l) \approx \hat{p}(s_t = k, \hat{\sigma}_t = l) = \frac{1}{N} \sum_{i=1}^N 1_{s_t=k, \hat{\sigma}_t=l} \quad (12)$$

In this,  $1_{s_t=k, \hat{\sigma}_t=l}$  is defined as a function that is 1 when  $s_t = k$  and  $\hat{\sigma}_t = l$ , and 0 elsewhere.

Pseudocode for the state-particle filter is in Algorithm 1 and Figure 1 shows the movement of state-particles between quadrants over time.

---

**Algorithm 1** State-Particle Filter

---

- 1: Initialize the Filter {Initialization can be treated separately}
  - 2: **for**  $t = 0$  **do**
  - 3:   Set the distribution  $P(s_0, \hat{\sigma}_0)$  arbitrarily
  - 4:   Sample  $N$  particles from  $P(s_0, \hat{\sigma}_0)$ . Call these particles  $\mathbb{P}_0^i$
  - 5:   Set  $h_0^i = 0$  and  $u_0^i = 0$  in each particle
  - 6:   Set all weights  $w_0^i = 1$
  - 7:   Set the likelihood  $\hat{p}(x_0) = 1$
  - 8: **end for**
  - 9: **for**  $t \geq 1$  **do**
  - 10:   Create all possible successor particles  $\mathbb{P}_t^{i,j}$ , including elements  $h_t^{i,j}$  and  $u_t^{i,j}$
  - 11:   Calculate new weights  $w_{t|t-1}^{i,j} = f(s_t^{i,j}, \hat{\sigma}_t^{i,j} | s_{t-1}^i, \hat{\sigma}_{t-1}^i) w_{t-1}^i$
  - 12:   Attach new information and calculate the unnormalized weights  $\hat{w}_t^{i,j} = f(x_t | s_t^{i,j}, \hat{\sigma}_t^{i,j}) w_{t|t-1}^{i,j}$
  - 13:   Store the likelihood  $\hat{p}(x_t | x_{1:t-1}) = \frac{1}{4N} \sum_{i=1}^N \sum_{j=1}^4 \hat{w}_t^{i,j}$
  - 14:   Normalize the weights  $w_t^{i,j} = \frac{\hat{w}_t^{i,j}}{\sum_{i=1}^N \sum_{j=1}^4 \hat{w}_t^{i,j}}$
  - 15:   Resample  $N$  particles according to the normalized weights
  - 16:   Store the probability of a bubble state  $\hat{p}(s_t = 1 | x_{1:t}) = \frac{1}{N} \sum_{i=1}^N 1_{s_t^i=1}$
  - 17: **end for**
-

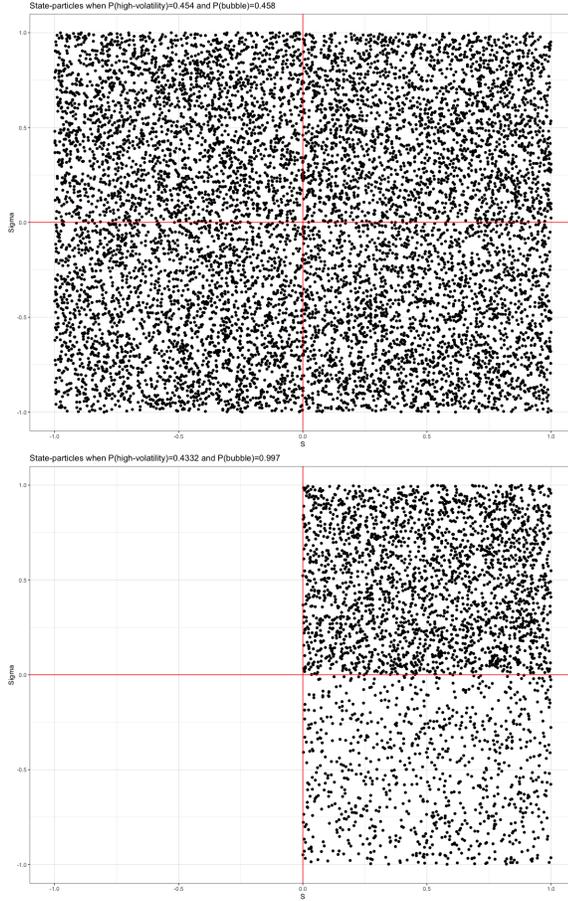


Figure 1: Position of State-Particles at time 1 and time 160

## $\theta$ -Particle Filter

Each step of the state-particle filter is dependent on the parameter-vector  $\theta$ , which in practice is not known. How could the probability of switching from a normal regime to a bubble regime be found if  $p_n$  is unknown? To conquer this problem, this model uses a secondary particle filter whose purpose is to learn  $\theta$ . Unlike other possible answers to the problem, such as using Maximum-Likelihood Estimates, this nested-particle filter approach allows the entire analysis to be done sequentially. In this model, the parameters need only be subjected to weak assumptions.

The Marginalized-Resample-Move algorithm simultaneously published by Fulop and Li (2013) and Chopin et al. (2013), also known as  $SMC^2$ , is a relatively efficient method of sequential parameter learning. Chopin et al. (2013) gives the more rigorous justification of the technique.

In it, the auxiliary variables  $h_t$  and  $u_t$ , which are created by the state-particle filter, are used to track the posterior probability of the fixed parameters,  $P(\theta|x_{1:t})$ . Because  $\theta$  is made of continuous variables that would be difficult if not impossible to discretize, tracking this posterior requires the use of Sampling-Importance Resampling (SIR). Let  $\pi$  refer to the instrumental probability density. Then, at each time  $t$ , there are  $M$   $\theta$ -particles  $(\pi_t^m, \theta_t^m)$ ,  $m = 1, 2, \dots, M$ , in which  $\pi_t^m$  represents the normalized importance weight. In the nested particle filter, each  $\theta$ -particle has  $N$  state-particles attached that form a sample drawn from  $f(s_t, \hat{\sigma}_t | x_{1:t}, \theta_t^m)$ . The values of interest gained by the sample of state-particles are the marginal likelihood  $\hat{p}(x_t | x_{1:t-1}, \theta^m)$  and the auxiliary variables  $h_t$  and  $u_t$ , the length of the present process and volatility regimes.

Let  $\delta_t^{i,m}$  be the auxiliary variables  $h_t$  and  $u_t$  in the  $i^{\text{th}}$  state-particle of the  $m^{\text{th}}$   $\theta$ -particle. The density of these random variables  $\delta_t^{i,m}$  is

$$\psi(\delta_{1:t}^{i,m} | x_{1:t}, \theta^m) = \prod_{l=1}^t \psi(\delta_l^{i,m} | \delta_{l-1}^{i,m}, x_l, \theta^m) \quad (13)$$

Within each  $\theta$ -particle, this density is approximated by  $\hat{\psi}(\delta_t | x_{1:t}) = \frac{1}{N} \sum_{i=1}^N 1_{\delta_t^{i,m} = \delta_t}$ .

The target distribution of this parameter-learning algorithm is

$$\gamma_t(\theta) = p(\theta | x_{1:t}) \propto p(x_{1:t} | \theta)p(\theta) = \prod_{l=1}^t p(x_l | x_{1:l-1}, \theta)p(\theta) \quad (14)$$

Define:

$$\hat{p}(x_t | x_{1:t-1}, \theta) \equiv \hat{p}(x_t | \delta_t, \delta_{t-1}, \theta) \quad (15)$$

From this prior, the following posterior can be found:

$$p(\theta, \delta_{1:t} | x_{1:t}) \propto p(\theta) \prod_{l=1}^t \hat{p}(x_l | \delta_l, \delta_{l-1}, \theta) \psi(\delta_l | \delta_{l-1}, x_l, \theta) \hat{p}(x_1 | \delta_1, \theta) \psi(\delta_1 | x_1, \theta) \quad (16)$$

In Equation 16, the components at time 1 are separated so that they can be treated according to a different starting distribution, or perhaps set to specific values as may be appropriate in some circumstances. An approximation for the target distribution in Equation 14, used in the model, is

$$\hat{\gamma}(\theta, \delta_{1:t}) \propto \prod_{l=1}^t \hat{p}(x_l | x_{1:t-1}, \delta_{1:l-1}, \theta) \hat{\psi}(\delta_l | x_{1:l}, \delta_{1:l-1}) p(\theta) \quad (17)$$

This uses the approximate likelihood from the state-particle filter, as well as the approximate density of the auxiliary variables. Both of these approximations are unbiased (Del Moral, 2004), and so the

target  $p(\theta | x_{1:t})$  is the marginal distribution of the density in Equation 16, and the two distributions have the same normalizing constant (Fulop & Li, 2013).

The  $\theta$ -particle filter uses a combination of SIR methods and the mathematics explained above to approximate a sample from the intractable distribution of  $\theta$ . First,  $M$   $\theta$ -particles must be sampled from the prior distribution  $\theta_0^m \sim p(\theta)$ , and each of them should have identical importance weight  $\pi_0^m = \frac{1}{M}$ . This creates a cloud of particles  $(\pi_0^m, \theta_0^m)$  that is distributed according to  $p(\theta)$ . To each of these  $M$   $\theta$ -particles, attach  $N$  state-particles created according to the previous section.

Now, assume the particle filter has moved forward to time  $t$ . There are now  $M$   $\theta$ -particles  $(\pi_t^m, \theta_t^m)$  that represent  $p(\theta | x_{1:t})$ . To include the next observation, the  $\theta$ -particles are moved forward such that  $\theta_{t+1}^m = \theta_t^m$ . The weights are then updated according to

$$\hat{\pi}_{t+1}^m = \pi_t^m \times \hat{p}(x_{t+1} | x_{1:t}, \delta_{1:t+1}^m, \theta_{t+1}^m) \quad (18)$$

The normalized weights for time  $t + 1$  are then  $\pi_{t+1}^m = \frac{\hat{\pi}_{t+1}^m}{\sum_{i=1}^M \hat{\pi}_{t+1}^i}$ , and the new  $\theta$ -particles are  $(\pi_{t+1}^m, \theta_{t+1}^m)$ .

This reweighting leaves those less-likely values of  $\theta$  with small probabilities and the more-likely values with higher probabilities. Thus, if the reweighting is repeated, the weight for unlikely values will approach zero, effectively removing those values from the sample. The Effective Sample Size (ESS) is tracked by

$$ESS_{\theta,t} = \frac{\left( \sum_{m=1}^M \pi_t^m \right)^2}{\sum_{m=1}^M (\pi_t^m)^2} \quad (19)$$

as per Chopin et al (2013). When the ESS drops below a chosen value  $B$ , which by rule of thumb is often set  $B = \frac{M}{2}$  (Fulop & Li, 2013), the sample must be rejuvenated by the *Resampling* and *Move* steps of the algorithm.

*Resample:* When  $ESS_{\theta,t} < B$ , the  $\theta$ -particles are resampled with replacement according to their normalized weights  $\pi_t^m$ , and the weights are then set to  $\pi_t^m = \frac{1}{M}$ . For each resampled  $\theta$ , the marginal likelihoods and the state-particles must be attached. The resampled cloud of particles now has  $ESS_{\theta,t} = M$ , and is still distributed according to  $\hat{\gamma}(\theta, \delta_{1:t})$ .

*Move:* This is the most computationally expensive part of the entire algorithm. To move

the  $\theta$ -particles, a Particle Marginal Metropolis-Hastings (PMMH) update, from Andrieu et al (2010), is performed on each.  $M$  new proposed  $\theta^*$ 's are drawn from a proposal distribution, in this case chosen to be a Multivariate normal with  $(\mu, \Sigma)$  chosen to be the weighted average and weighted covariance of the resampled  $\theta$ -particles at time  $t$ . Then, the algorithm is repeated for times  $1 : t$ , after which there are  $M$  pairs of  $\theta$ -particles and  $\theta^*$ -particles to be compared.

- Let the acceptance probability be

$$\alpha = \min \left( 1, \frac{\hat{p}(x_{1:t}, \delta_{1:t}^* | \theta^*) h_t(\theta^m | \theta^*)}{\hat{p}(x_{1:t}, \delta_{1:t}^m | \theta^m) h_t(\theta^* | \theta^m)} \right)$$

- With probability  $\alpha$ , replace  $\theta^m$  and attached particles with  $\theta^*$ . With probability  $(1 - \alpha)$ , keep the original values.

The final probability of being in any of the four possible regimes from the state-particle filter depends on  $\theta$ , and is found by

$$E(1_{s_t=k, \hat{\sigma}_t=l} | x_{1:t}) = E[E(1_{s_t=k, \hat{\sigma}_t=l} | x_{1:t}, \theta) | x_{1:t}] \approx \sum_{m=1}^M \pi_t^m \frac{1}{N} \sum_{n=1}^N 1_{s_t=k, \hat{\sigma}_t=l} \quad (20)$$

That is: The probability of being in a regime, taking  $\theta$  into account, is the probability of being in that regime in each state-particle filter with given parameters times the weight of those parameters created by the  $\theta$ -particle filter. Take the probability of being in a bubble state from step 16 of the state-particle algorithm and multiply it by the weight of the attached  $\theta$ -particle, then sum that over all  $\theta^m$ .

As the time series gets longer, the move step becomes increasingly computationally expensive. However, as the  $\theta$ -particle filter gets closer to representing the true distribution  $\gamma(\theta)$  these resample-move updates are needed less often. More detail on this is given in the discussion of the model's success on simulated data in Chapter 4, Results.

The pseduocode for this algorithm is in Algorithm 2.

---

**Algorithm 2**  $\theta$ -Particle Filter

---

- 1: Initialize the Filter {Initialization can be treated separately}
  - 2: **for**  $t = 0$  **do**
  - 3:   Select starting elements for  $M$   $\theta$ -particles by some convenient distribution
  - 4:   Set  $\theta$ -particle weights  $\pi_0^m = 1$
  - 5:   For each  $\theta$ -particle, initialize an attached state-particle filter
  - 6: **end for**
  - 7: **for**  $t \geq 1$  **do**
  - 8:   Move the state-particle filter forward from  $t - 1$  to  $t$
  - 9:   Calculate the marginal likelihood  $\hat{p}(x_{1:t} | x_{1:t-1}, \delta_{1:t}^m, \theta^m)$
  - 10:   Reweight  $\theta$ -particles according to  $\hat{\pi}_t^m = \pi_{t-1}^m \times \hat{p}(x_{1:t} | x_{1:t-1}, \delta_{1:t}^m, \theta^m)$
  - 11:   Normalize the weights  $\hat{\pi}_t^m$  from the previous step, and set the normalized weights as  $\pi_t^m$
  - 12:   Find  $ESS_{\theta,t} = \frac{\left(\sum_{m=1}^M \pi_t^m\right)^2}{\sum_{m=1}^M (\pi_t^m)^2}$
  - 13:   **if**  $ESS_{\theta,t} < B$  **then**
  - 14:     Resample  $\theta$ -particles according to normalized weights  $\pi_t^m$
  - 15:     Calculate the weighted mean vector and covariance matrix  $(\mu, \Sigma)$
  - 16:     Draw  $M$  new  $\theta^*$ -particles from a multivariate normal distribution  $(\mu, \Sigma)$
  - 17:     Run the state-particle and  $\theta$ -particle filters from time  $t = 1$  to time  $t$
  - 18:     Find the acceptance probability  $\alpha = \min\left(1, \frac{\hat{p}(x_{1:t}, \delta_{1:t}^* | \theta^*) \times h_t(\theta^m | \theta^*)}{\hat{p}(x_{1:t}, \delta_{1:t}^m | \theta^m) \times h_t(\theta^* | \theta^m)}\right)$  {The multivariate normal distribution is symmetric, and so the  $h_t$  terms can be canceled out in this case, leaving only the marginal likelihoods}
  - 19:     Draw a random number,  $U$ , from a Uniform distribution  $[0, 1]$
  - 20:     **if**  $U < \alpha$  **then**
  - 21:       Replace  $\theta^m$  with  $\theta^*$
  - 22:     **else**
  - 23:       Keep  $\theta^m$
  - 24:     **end if**
  - 25:   **end if**
  - 26: **end for**
-

## Nested Particle Filter

Now that the two particle filters have been explained individually, this section will focus on how they fit together and will end with the algorithm in its entirety. It is important to realize that the state-particle filter occurs entirely within the  $\theta$ -particle filter; the parameters used in the state-particle filter come from the  $\theta$ -particle filter.

There are  $M$   $\theta$ -particles, and each one of those has  $N$  state-particles attached. Each total particle, a state-particle attached to a  $\theta$ -particle, is herein referred to as a  $\theta$ -state. There are a total of  $MN$   $\theta$ -states. It is recommended by Chopin et al (2013) that  $M$  be at least the minimum of the number of timesteps to be filtered or 1,000; however, both Fulop and Li (2013) and Svensson and Schon (2016) show that the algorithm can return an accurate posterior distribution even with 100 or fewer  $\theta$ -particles. Neither of the time series analyzed in this paper has more than 1,000 timesteps;  $M$  is set at 500, and the largest viable  $N$  was 40 for a total of 20,000  $\theta$ -states. More particles is better, and the small number of particles is one of the limitations described in Chapter 5.

After several successive resamples of the  $\theta$ -particles, the elements covariance matrix may approach zero, and so the covariance scaling factor  $c$  was introduced. In the final model, the multivariate distribution from which new  $\theta$ -particles are drawn is defined by  $(\mu, c\Sigma)$ . The PMMH acceptance rate is tracked as well: When too many  $\theta^*$  are accepted ( $50\% <$ ), the new and old particles are too close together and  $c$  is multiplied by 1.05; when too few are accepted ( $< 15\%$ ), the proposal distribution is too wide and  $c$  is reduced to  $0.95c$ , with a minimum value of 0.05. The covariance matrix is not allowed to be zero. The sample of  $\theta$ -particles can never fully degenerate.

Pseudocode for the complete nested particle filter is given by Algorithms 3-5 at the end of this chapter. A flowchart explaining how the information moves through the algorithm is presented in Figure 2.

## Bubble-Stamping Loss Function

The raw particle-filter-based probability of being a bubble regime may be a poor bubble-stamping indicator. The event of interest is not necessarily the bubble itself, but the switch between

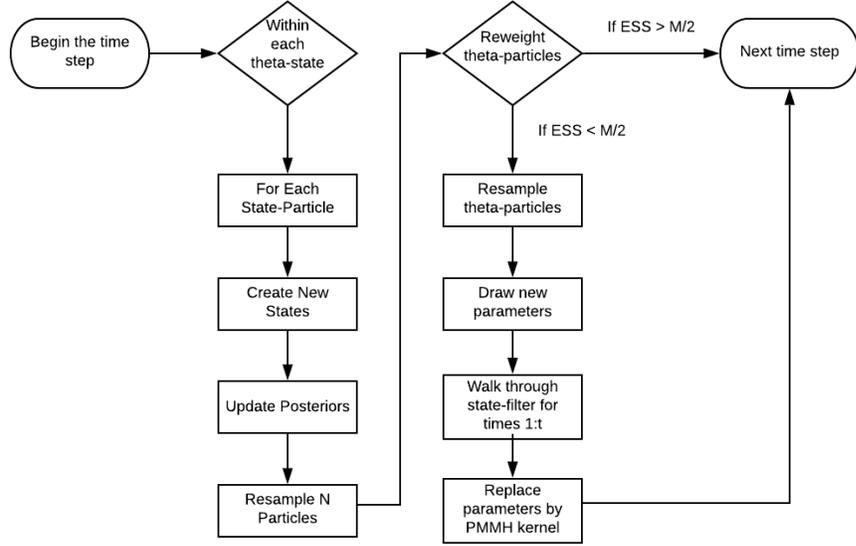


Figure 2: Nested Algorithm Flowchart

regimes. During the life of a bubble, the probability of being in a bubble still fluctuates. Thus, a loss function is introduced to improve the bubble-stamping procedure. This loss function is closely related to that of Fulop and Yu (2017).

Say that any misspecification of the regime results in some abstract loss. Let the loss from erroneously stamping a bubble regime be  $l_t^b$ , and the loss from erroneously stamping a normal regime is  $l_t^n$ . Let  $r_t$  be 1 when the timestep is stamped as a bubble and 0 otherwise. Thus, the total loss from misspecifying the regime at any timestep is

$$L_t(s_t, r_t) = l_t^b r_t \mathbf{1}_{s_t=0} + l_t^n (1 - r_t) \mathbf{1}_{s_t=1}$$

Note how the loss variables are multiplied by the number of particles in the opposite regime. For the two possible values of  $r_t$ , the expected loss is easy to find.

$$E[L_t(s_t, r_t = 1)] = l_t^b P(\mathbf{1}_{s_t=0} | x_{1:t})$$

$$E[L_t(s_t, r_t = 0)] = l_t^n P(\mathbf{1}_{s_t=1} | x_{1:t})$$

The expected loss is minimized by stamping the start of a bubble when  $l_t^b P(\mathbf{1}_{s_t=0} | x_{1:t}) < l_t^n P(\mathbf{1}_{s_t=1} | x_{1:t})$ , and the expected loss is minimized by stamping the start of a normal regime when

$l_t^n P(1_{s_t=1}|x_{1:t}) < l_t^b P(1_{s_t=0}|x_{1:t})$ . With some algebraic manipulation, it is optimal to stamp the start of a bubble when

$$\frac{l_t^b}{l_t^n} < \frac{P(1_{s_t=1}|x_{1:t})}{P(1_{s_t=0}|x_{1:t})}$$

and it is optimal to stamp the beginning of a normal regime when

$$\frac{l_t^b}{l_t^n} < \frac{P(1_{s_t=0}|x_{1:t})}{P(1_{s_t=1}|x_{1:t})}$$

Since the losses are subjective, the left-hand ratio can be set by the user;  $\zeta = \frac{l_t^b}{l_t^n}$ , with higher  $\zeta$  indicating stronger persistence of regimes.

At any given time, the time series is either in a bubble regime or a normal regime; there are no other options. Thus,  $P(1_{s_t=1}|x_{1:t}) + P(1_{s_t=0}|x_{1:t}) = 1$ . Using this, the following algebra is possible.

$$\begin{aligned} \zeta &< \frac{P(1_{s_t=1} | x_{1:t})}{1 - P(1_{s_t=1} | x_{1:t})} \\ (1 - P(1_{s_t=1} | x_{1:t}))\zeta &< P(1_{s_t=1} | x_{1:t}) \\ \zeta - \zeta P(1_{s_t=1} | x_{1:t}) &< P(1_{s_t=1} | x_{1:t}) \\ \zeta &< P(1_{s_t=1} | x_{1:t}) + \zeta P(1_{s_t=1} | x_{1:t}) \\ \zeta &< P(1_{s_t=1} | x_{1:t})(1 + \zeta) \\ \frac{\zeta}{1 + \zeta} &< P(1_{s_t=1} | x_{1:t}) \end{aligned}$$

The final effect of the loss function is: stamp the beginning of a bubble when the probability of being in a bubble crosses  $\frac{\zeta}{1+\zeta}$ , and stamp the end of that bubble when the probability cross back underneath  $\frac{1}{1+\zeta}$ .

## Simulated Data

When real data is used, the values of parameters cannot be truly known and any declaration as to the state of the time series is necessarily biased by subjective judgement. Thus, the proper way to evaluate the accuracy of the proposed model is with simulated data. The time series on which this model is tested was created to match the assumptions of the model:

- The time series follows either the mean-reverting or explosive process
- Volatility is either high or low
- The switches are independent and distributed according to geometric distributions.

First, the parameters are set:

$$\theta = \begin{bmatrix} p_n \\ p_b \\ \sigma_l \\ \sigma_m \\ p_l \\ p_h \\ \beta_1 \\ \beta_2 \end{bmatrix} \equiv \begin{bmatrix} 1/150 \\ 1/50 \\ 20 \\ 2 \\ 1/100 \\ 1/30 \\ 0.98 \\ 1.02 \end{bmatrix}$$

Then, with  $x_0 = 100$ , the time series followed a random walk  $x_t = x_{t-1} + \epsilon_t$ , where  $\epsilon_t \sim N(0, \sigma_l^2)$  for 120 time steps; this was necessary because the model uses a 120-period moving average for the mean-reverting regime. If a different moving-average window is discerned to be more effective, that number of periods will be required before the particle filter can be implemented.

Then, a total of 10 switches were drawn from the geometric distributions for regime and volatility; 6 from the regime distribution and 4 from the volatility distribution. The switches were drawn until the time series had at least two occurrences of each regime, and to make it sufficiently long for testing the computational complexity of the algorithm. The final length of the test data is 627 periods, with 507 usable for testing the particle filter.

Between switches, the time series was made to follow the proper regime process with the proper volatility. This is all shown in Table 2; wherein

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} (s_t, \hat{\sigma}_t) = (0, 0) \\ (0, 1) \\ (1, 0) \\ (1, 1) \end{bmatrix}$$

Section	Time	State	Process Function
1	[121, 138]	1	$x_t = (1 - \beta_1)\alpha_t + \beta_1x_{t-1} + \sigma_l\epsilon_t$
2	[139, 149]	3	$x_t = \beta_2x_{t-1} + \sigma_l\epsilon_t$
3	[150, 188]	4	$x_t = \beta_2x_{t-1} + \sigma_l\sigma_m\epsilon_t$
4	[189, 209]	3	$x_t = \beta_2x_{t-1} + \sigma_l\epsilon_t$
5	[210, 353]	1	$x_t = (1 - \beta_1)\alpha_t + \beta_1x_{t-1} + \sigma_l\epsilon_t$
6	[354, 405]	2	$x_t = (1 - \beta_1)\alpha_t + \beta_1x_{t-1} + \sigma_l\sigma_m\epsilon_t$
7	[406, 423]	1	$x_t = (1 - \beta_1)\alpha_t + \beta_1x_{t-1} + \sigma_l\epsilon_t$
8	[424, 456]	3	$x_t = \beta_2x_{t-1} + \sigma_l\epsilon_t$
9	[457, 486]	1	$x_t = (1 - \beta_1)\alpha_t + \beta_1x_{t-1} + \sigma_l\epsilon_t$
10	[487, 567]	3	$x_t = \beta_2x_{t-1} + \sigma_l\epsilon_t$
11	[568, 627]	1	$x_t = (1 - \beta_1)\alpha_t + \beta_1x_{t-1} + \sigma_l\epsilon_t$

Table 2: Simulated Data States

A plot showing the simulated time series is in Figure 3.

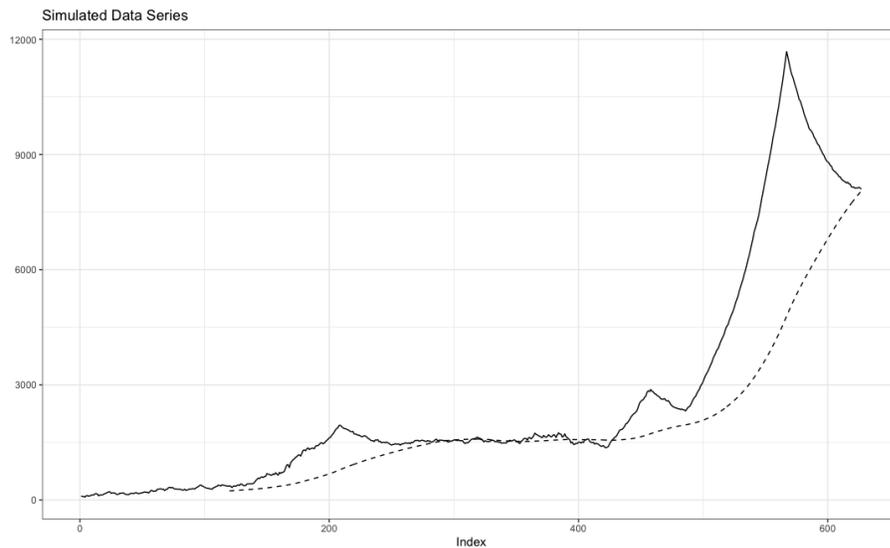


Figure 3: Simulated Data Series

---

**Algorithm 3** Nested Particle Filter: Part 1

---

- 1: Initialize both filters {Initialization can be treated separately}
  - 2: **for**  $t = 0$  **do**
  - 3:   Select starting elements for  $M$   $\theta$ -particles by any convenient distribution
  - 4:   Set  $\theta$ -particle weights  $\pi_0^m = 1$
  - 5:   **for** Each  $\theta$ -particle **do**
  - 6:     Set the distribution  $P(s_0, \hat{\sigma}_0)$  arbitrarily {Initialize an attached state-particle filter}
  - 7:     Sample  $N$  particles from  $P(s_0, \hat{\sigma}_0)$ . Call these particles  $\mathbb{P}_0^i$
  - 8:     Set  $h_0^i = 0$  and  $u_0^i = 0$  in each particle
  - 9:     Set all weights  $w_0^i = 1$
  - 10:     Set the likelihood  $\hat{p}(x_0) = 1$
  - 11:   **end for**
  - 12:   Call each combined  $\theta$ -particle and state-particle a  $\theta$ -state
  - 13: **end for**
-

---

**Algorithm 4** Nested Particle Filter: Part 2

---

- 1: **for**  $t \geq 1$  **do**
  - 2:   **for** Each  $\theta$ -state **do**
  - 3:     Create all possible successor particles  $\mathbb{P}_t^{i,j}$ , including elements  $h_t^{i,j}$  and  $u_t^{i,j}$
  - 4:     Calculate new weights  $w_{t|t-1}^{i,j} = f(s_t^{i,j}, \hat{\sigma}_t^{i,j} | s_{t-1}^i, \hat{\sigma}_{t-1}^i) w_{t-1}^i$
  - 5:     Attach new information and calculate the unnormalized weights  $\hat{w}_t^{i,j} = f(x_t | s_t^{i,j}, \hat{\sigma}_t^{i,j}) w_{t|t-1}^{i,j}$
  - 6:     Store the likelihood  $\hat{p}(x_t | x_{1:t-1}) = \frac{1}{4N} \sum_{i=1}^N \sum_{j=1}^4 w_t^{i,j}$
  - 7:     Normalize the weights  $w_t^{i,j} = \frac{\hat{w}_t^{i,j}}{\sum_{i=1}^N \sum_{j=1}^4 \hat{w}_t^{i,j}}$
  - 8:   **end for**
  - 9:   **for** Each  $\theta$ -particle, which includes  $N$  state-particles **do**
  - 10:     Resample  $N$  particles according to the normalized weights
  - 11:     Calculate the marginal likelihood  $\hat{p}(x_{1:t} | x_{1:t-1}, \delta_{1:t}^m, \theta^m)$
  - 12:     Store the probability of a bubble state  $\hat{p}(s_t = 1 | x_{1:t}) = \frac{1}{N} \sum_{i=1}^N 1_{s_t^i=1}$
  - 13:   **end for**
  - 14:   Reweight  $\theta$ -particles according to  $\hat{\pi}_t^m = \pi_{t-1}^m \times \hat{p}(x_{1:t} | x_{1:t-1}, \delta_{1:t}^m, \theta^m)$
  - 15:   Normalize the weights  $\hat{\pi}_t^m$  from the previous step, and set the normalized weights as  $\pi_t^m$
  - 16:   Find  $ESS_{\theta,t} = \frac{\left( \sum_{m=1}^M \pi_t^m \right)^2}{\sum_{m=1}^M (\pi_t^m)^2}$
  - 17: **end for**
-

---

**Algorithm 5** Nested Particle Filter: Part 3

---

1: **for**  $t \geq 1$ , continued **do**

2:   **if**  $ESS_{\theta,t} < B$  **then**

3:     Resample  $\theta$ -particles according to normalized weights  $\pi_t^m$

4:     Calculate the weighted mean vector and covariance matrix  $(\mu, \Sigma)$

5:     Draw  $M$  new  $\theta^*$ -particles from a multivariate normal distribution  $(\mu, \Sigma)$

6:     Run the state-particle and  $\theta$ -particle filters from time  $t = 1$  to time  $t$

7:     Find the acceptance probability  $\alpha = \min\left(1, \frac{\hat{p}(x_{1:t}, \delta_{1:t}^* | \theta^*) \times h_t(\theta^m | \theta^*)}{\hat{p}(x_{1:t}, \delta_{1:t}^m | \theta^m) \times h_t(\theta^* | \theta^m)}\right)$  {The multivariate normal distribution is symmetric, and so the  $h_t$  terms can be canceled out in this case, leaving only the marginal likelihoods}

8:     Draw a random number,  $U$ , from a Uniform distribution  $[0, 1]$

9:     **if**  $U < \alpha$  **then**

10:       Replace  $\theta^m$  with  $\theta^*$

11:     **else**

12:       Keep  $\theta^m$

13:     **end if**

14:   **end if**

15:   Calculate the final probability of a bubble state  $\hat{p}(s_t = 1) = \sum_{m=1}^M \pi_t^m \frac{1}{N} \sum_{n=1}^N 1_{s_t=1}$

16: **end for**

---

## Data

All of the data necessary for this analysis was available from the Federal Reserve Bank of St. Louis's economic database, FRED. The data is available free of charge and can be downloaded in .csv format. Monthly data is used in this analysis for three reasons: 1) most of the necessary data is measured monthly but not daily; 2) monthly data exhibits much less noise than higher-frequency data; 3) there are 546 months from June 1974 to December 2019, and the program becomes quite computationally expensive after around 600 timesteps.

Following the example of Wu (1995), the measure of money supply used in this thesis is M1. "M1 includes funds that are readily accessible for spending" (FRED, 2020a). This includes money that is outside of the U.S. Treasury or the vaults of depository institutions (i.e. consumer banks). It also includes demand deposits, traveler's checks, and other checkable deposits. Currency and the other included items are seasonally adjusted separately and then summed (FRED, 2020a). The M1 data on FRED is updated every month, and runs about one month behind. It is not available in real time, and investors could not use today's M1 to make investment decisions. Thus, this time series is lagged by two months; investors use January's measurement to make decisions about March, September's data to make decisions about October, and so on.

Again following Wu (1995), the total industrial production index (TPI) was used as a proxy for real income. This time series "measures the real output of all relevant establishments located in the United States, regardless of their ownership" (FRED, 2020b). It is not seasonally adjusted. Like M1, the data is updated monthly, and runs a little more than one month behind. Thus, the model in this thesis will assume that investor's use two-month old data in their analysis.

The "most broadly available and frequently used index to represent the price level of traded goods" (Berrendorf & Chen, 2013) is the Producer Price Index (PPI), and so that is used in this paper to represent the price index. It is better to use the price level for traded goods rather than non-traded goods, as Bettendorf and Chen (2013) showed that the price level of traded goods seems to better coincide with the behavior of the exchange rate. The PPI for Japan in the Federal Reserve Economic Database is a monthly time series, but is updated slightly more than two months late.

Therefore, investors will use January's data in their analysis for April and May's data for their analysis in July, etc.

The interest rate for the country of Japan,  $i^*(t)$  in the fundamental value equation, is the Immediate Rate from the Central Bank of Japan. This time series is monthly, and is updated just over one month late. Thus, like with M1 and TPI, the time series will be lagged two months; investors will use January's interest rate in their analysis for March, and so on. Up-to-date interest rates are available from many websites, and the Immediate Rate is not typically the interest rate used by traders, but it is used here because it is the only interest rate available for free on a monthly basis going back to 1974. Other interest rate series only go back as far as 1986.

The JPY/USD exchange rate is also gathered from FRED due to the trustability of the historical data. In that database, this series is updated monthly and runs more than one month behind, but investors will surely have access to up-to-date pricing, in real time, and so the exchange rate is not lagged.

Calculation of the interest-rate semi-elasticity is performed by linear regression. Using the normal, non-logarithmic values, a simple take on the relationship between interest rates and money demanded is

$$M = e^{ki}$$

or, after applying the natural logarithm,

$$\ln(M) = ki$$

where  $\ln(M)$  was referred to above by the lower-case  $m$ . Thus, the value  $k$  for the fundamental-price model can be found by a simple linear regression with  $i$  as the input variable and  $m$  as the response.

Once the data was all downloaded in .csv format, it was loaded into R Studio for analysis. The natural logarithm was taken for M1, Japan's PPI, USD/JPY, and U.S. TPI, but not for the Japanese interest rate. The coefficient  $k$  was calculated by the regression explained above, and  $a = \frac{k}{k+1}$ . The fundamental value  $z(t)$  is simply a linear combination of these variables, and this was subtracted from the natural log of the JPY/USD rate, thereby removing the fundamental value and leaving only the bubble component.

The exchange rate was returned to standard form by exponentiation. Then, the reciprocal of the exchange rate was taken, so that the entry and exit prices were listed in USD for more intuitive trading analysis.

## IV. Results

### Results from Simulated Data

The stamping of bubbles requires specification of  $\zeta$ , which indicates the (subjective) belief in the persistence of regimes. Thus, a desirable outcome of testing this algorithm on simulated data is the determination of the optimal value of  $\zeta$ ; the one that produces the most accurate stamps. See Figure 4. There is a tradeoff; up to a point, higher  $\zeta$  fills in true bubbles better, but it also lengthens erroneous bubbles. When  $\zeta$  is too high, no erroneous bubbles are stamped but the first bubble, which is very close to the beginning of the particle filter, is almost entirely missed. In a trading sense, higher  $\zeta$  is more risk averse and should result in fewer losing trades, but may have a negative effect on the end equity due to too few trades being undertaken. This will be explored in the real data.

### Effective Sample Size

The simulated data required PMMH updates 22 times, of which 19 occurred before the 50th timestep and 1 occurred after the 100th. The ESS reached 251.65 at filter-timestep 358 (478 for the whole data), but instead of dropping below 250 and triggering another update, it shot back up and no update was required. The Effective Sample Size is in Figure 6.

### Parameters

Histograms showing the distribution of parameters both before and after the particle filter are in Figure 5. On the simulated data, the parameters were sampled down to a single value.

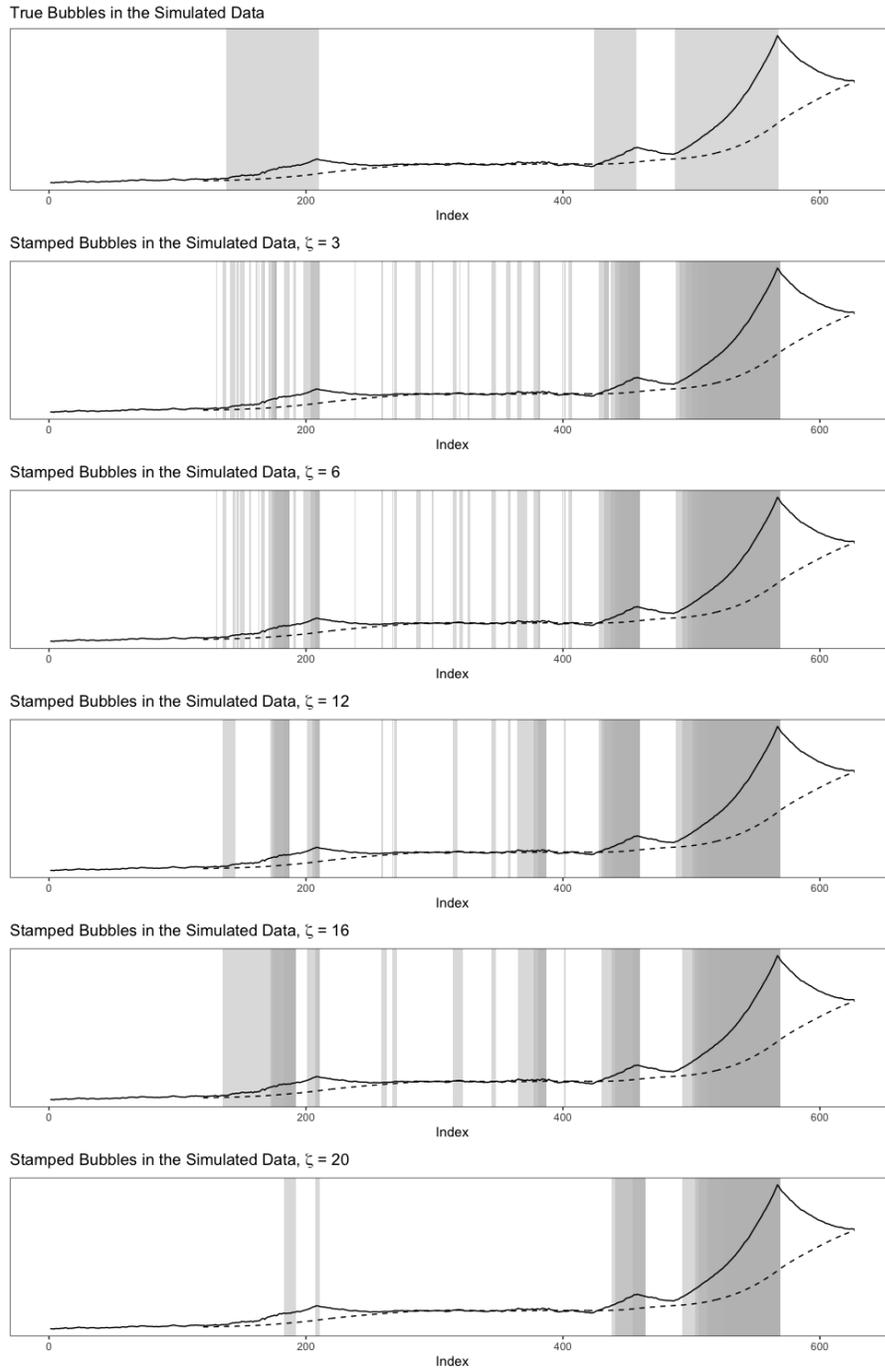


Figure 4: True and Stamped Bubbles

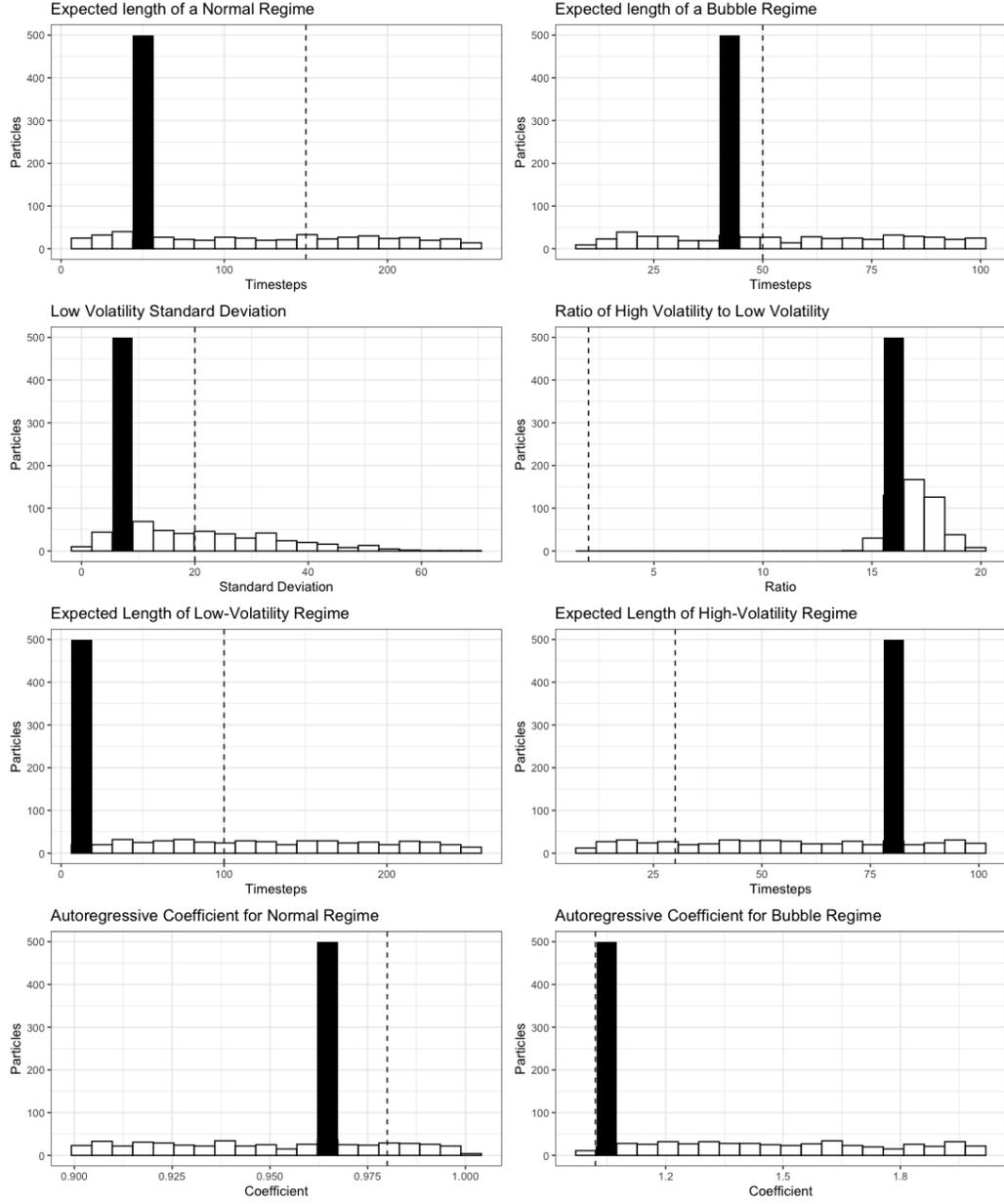


Figure 5: Histogram of prior (white) and posterior (black) distributions of parameters

The algorithm did not do especially well at estimating the expected length of regimes, possibly because there were few true switches. The expected length of regimes was generally underestimated. In Fulop and Yu (2017), the length of both process-regimes and volatility-regimes was overestimated. In this thesis, the learned parameter was quite close to the coefficient of the bubble process. The inaccuracy of the estimate of the ratio of high-to-low volatility regimes,  $\sigma_m$ ,

suffered from a poor choice of the prior distribution. This shows the importance of selecting a wide prior. It is also unexpected that the  $\theta$ -particles would be resampled down to a single value; the only way this could happen is if the weight of this particle was so relatively high that it was favored during repeated resampling steps, and then the new parameter-particles were all rejected during the last PMMH update. This is highly unlikely, but the results show that it is possible. This may also explain why the ESS behaved strangely; the single value of the parameters did not fit the area before timestep 358 very well, and so the ESS was falling, but they fit timestep 358 quite well and so the ESS suddenly sprang back up.

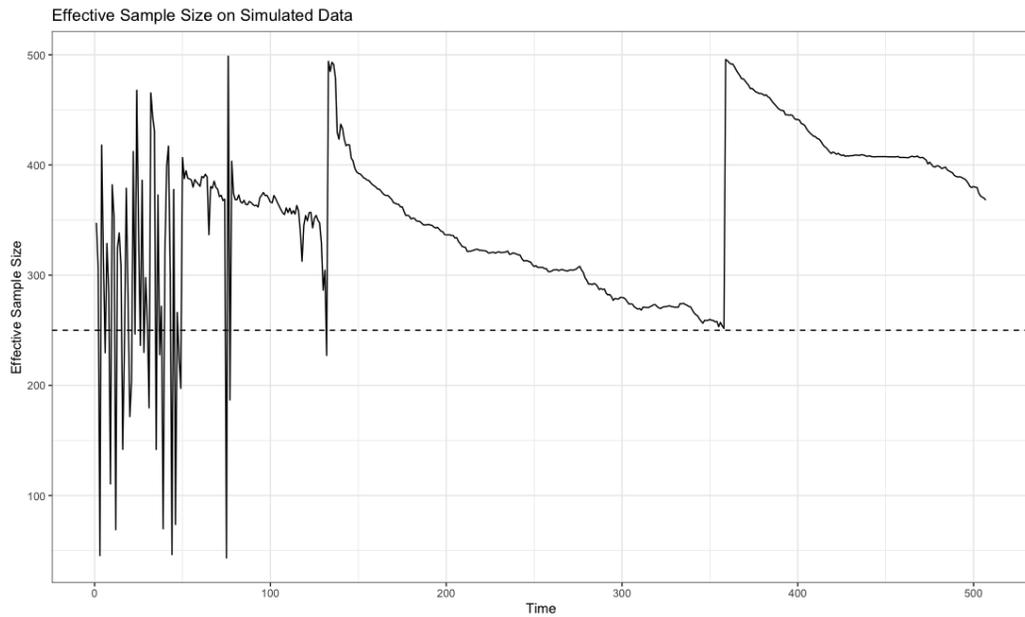


Figure 6: ESS on Simulated Data

## Empirical Results

A set of plots of the bubble component of the exchange rate, showing stamped bubbles for various values of  $\zeta$ , is in Figure 7. Two high and two low values are considered along with  $\zeta = 16$ , due to its performance on the simulated data. Regardless of  $\zeta$ , the algorithm finds bubbles in the late 1980s, from around June 1985 through May 1989, then several short bubbles from 1999 to 2003, followed by time of bubbles from March 2008 through October 2009, and finally between April

2015 and July 2016. Other than those areas in which  $\zeta$  agree that bubbles exist, there are many small differences and short bubbles stamped by one value but not another. Most interesting is the seeming disagreement about the early 1990s. Low  $\zeta$  finds strong bubbles there, while with high  $\zeta$ , those bubble disappear.

Those same bubbles are stamped on the real exchange rate in Figure 8. Analyzing the bubble component rather than the exchange rate itself made it possible to detect several bubbles even while the exchange rate was below the long-term average.

### Effective Sample Size

The effective sample size was not as stable in the real data as it was for the simulated data. At first, it varied widely (19 PMMH updates were required in the first 100 timesteps), but once a well-fit set of parameters was found, it declines slowly from timestep 94 to 288. After timestep 289, falling on June 1st, 1998, the ESS became unstable again and required 10 more PMMH updates. The effective sample size does not drop below 250 after timestep 340, September 1st, 2002. This plot is in Figure 9.

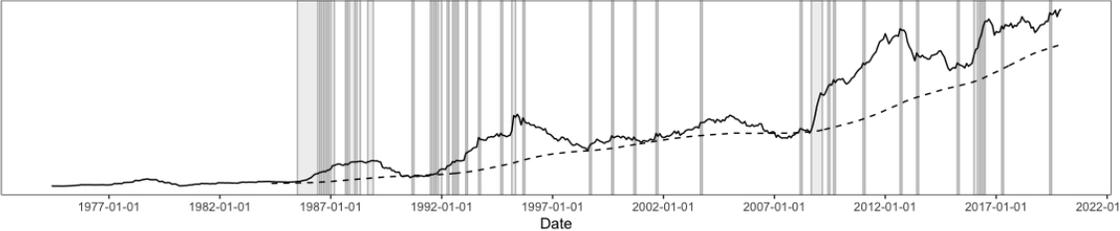
Table 3: Posterior Mean and Standard Deviation of Parameters

	$p_n$	$p_b$	$\sigma_l$	$\sigma_m$	$p_l$	$p_m$	$\beta_1$	$\beta_2$
Mean	134.61632	54.48070	8.30132	7.27419	174.21871	50.94670	0.96404	1.07997
St. Dev.	0.04667	0.03005	0.05769	0.08920	0.02802	0.06379	0.00046	0.01338

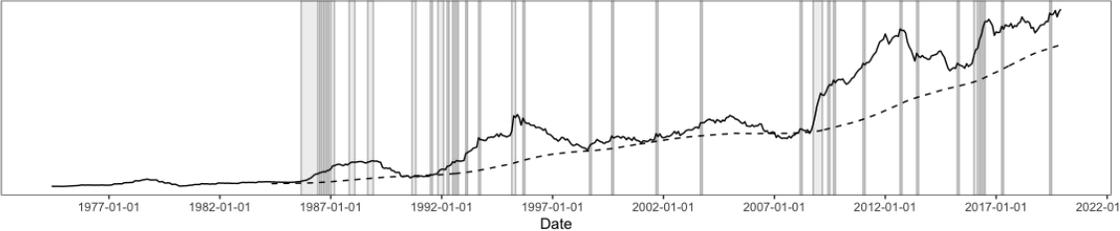
### Parameters

The starting distributions of parameters was the same as in the simulated data. Unlike in the simulated data, the  $\theta$ -particles did not degenerate to a single point. The mean and standard deviation of each parameter is shown in Table 3. The standard deviations are small, even though the covariance-scaling constant,  $c$ , is 4.11. In each PMMH update, the algorithm accepted a high proportion of proposed particles. This is a sign that the random-walk aspect of the Metropolis-Hastings sampler was not wide enough; the algorithm may not have been ergodic. Histograms

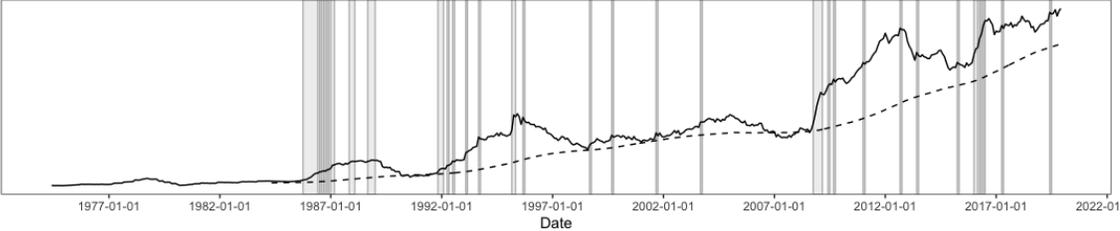
Stamped Bubbles in the Bubble Component,  $\zeta = 1$



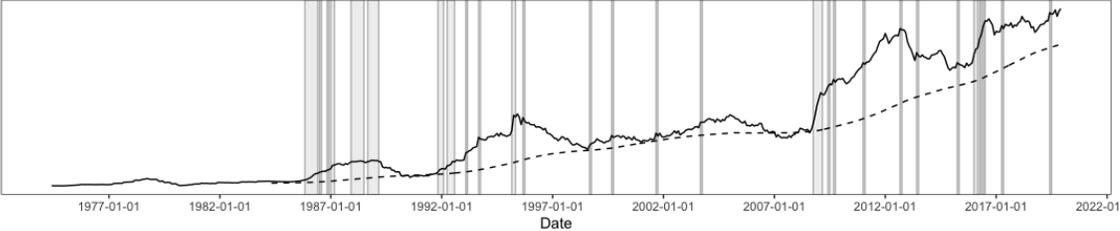
Stamped Bubbles in the Bubble Component,  $\zeta = 5$



Stamped Bubbles in the Bubble Component,  $\zeta = 16$



Stamped Bubbles in the Bubble Component,  $\zeta = 50$



Stamped Bubbles in the Bubble Component,  $\zeta = 99$

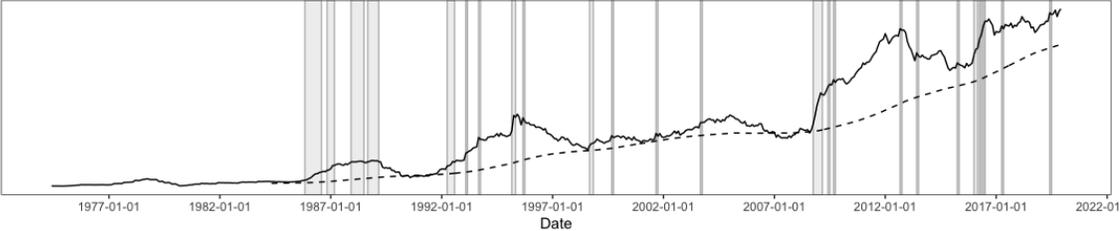


Figure 7: Stamped Bubbles in the Bubble Component of the Exchange Rate

showing the prior and posterior distributions of parameters are in Figure 10.

All of the parameters ended up inside the prior distribution; the width of the distributions tightened until all samples fall into a single bin. The most extreme parameters, compared to their

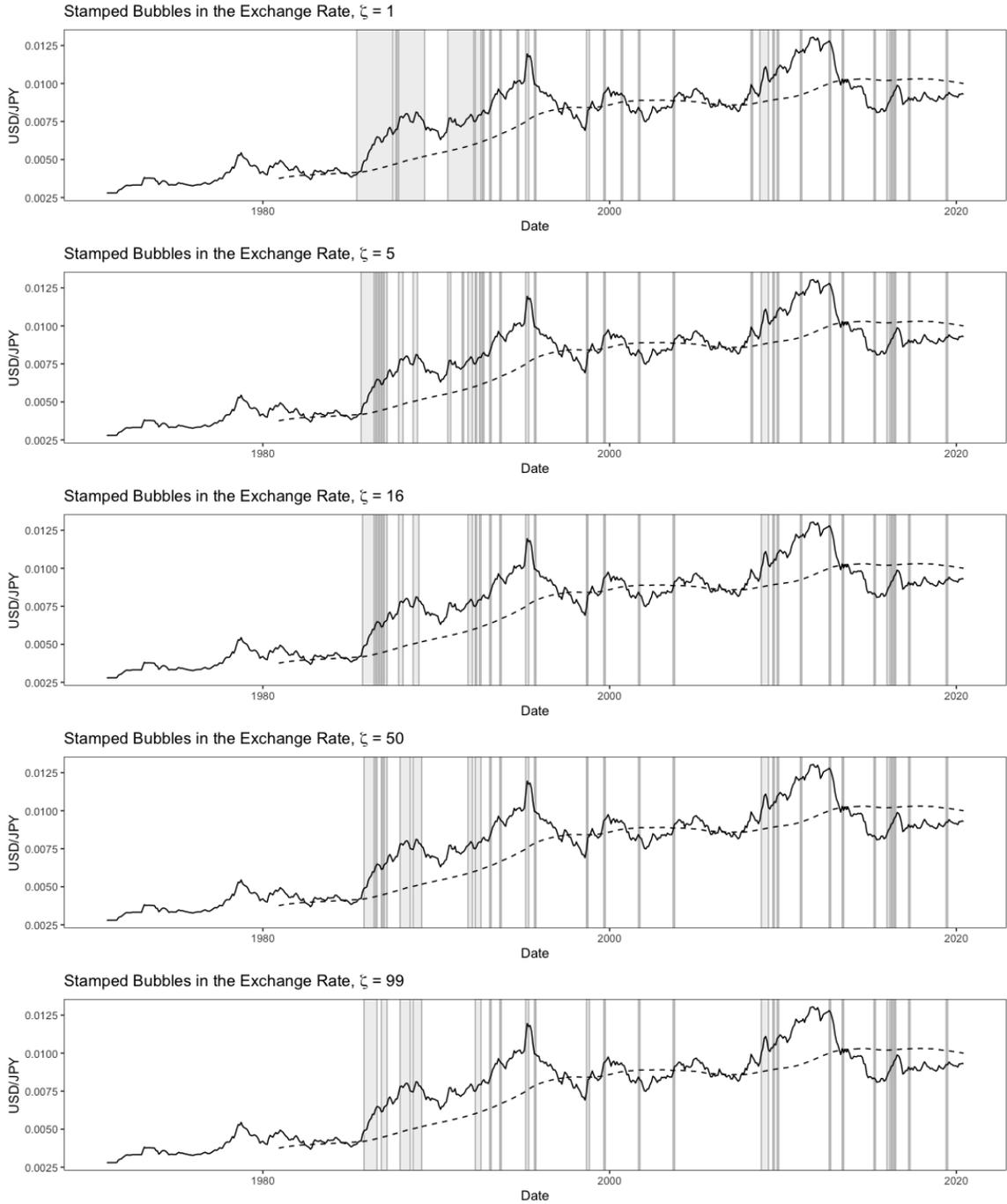


Figure 8: Stamped Bubbles in the Bubble Component of the Exchange Rate

priors, are  $\sigma_m$  and  $\beta_2$ :  $\sigma_m$  moved toward the upper curve of the prior, and  $\beta_2$  moved to the lower end of the prior.

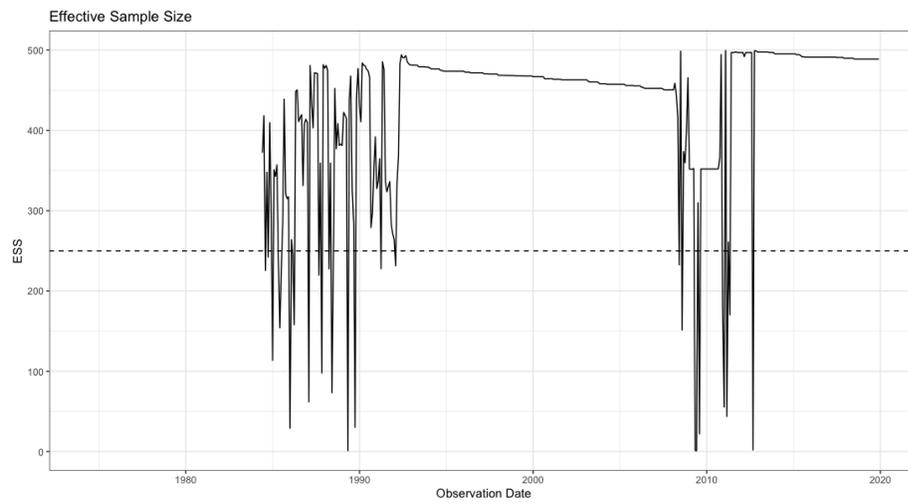


Figure 9: Effective Sample Size in analysis of the bubble component, over time

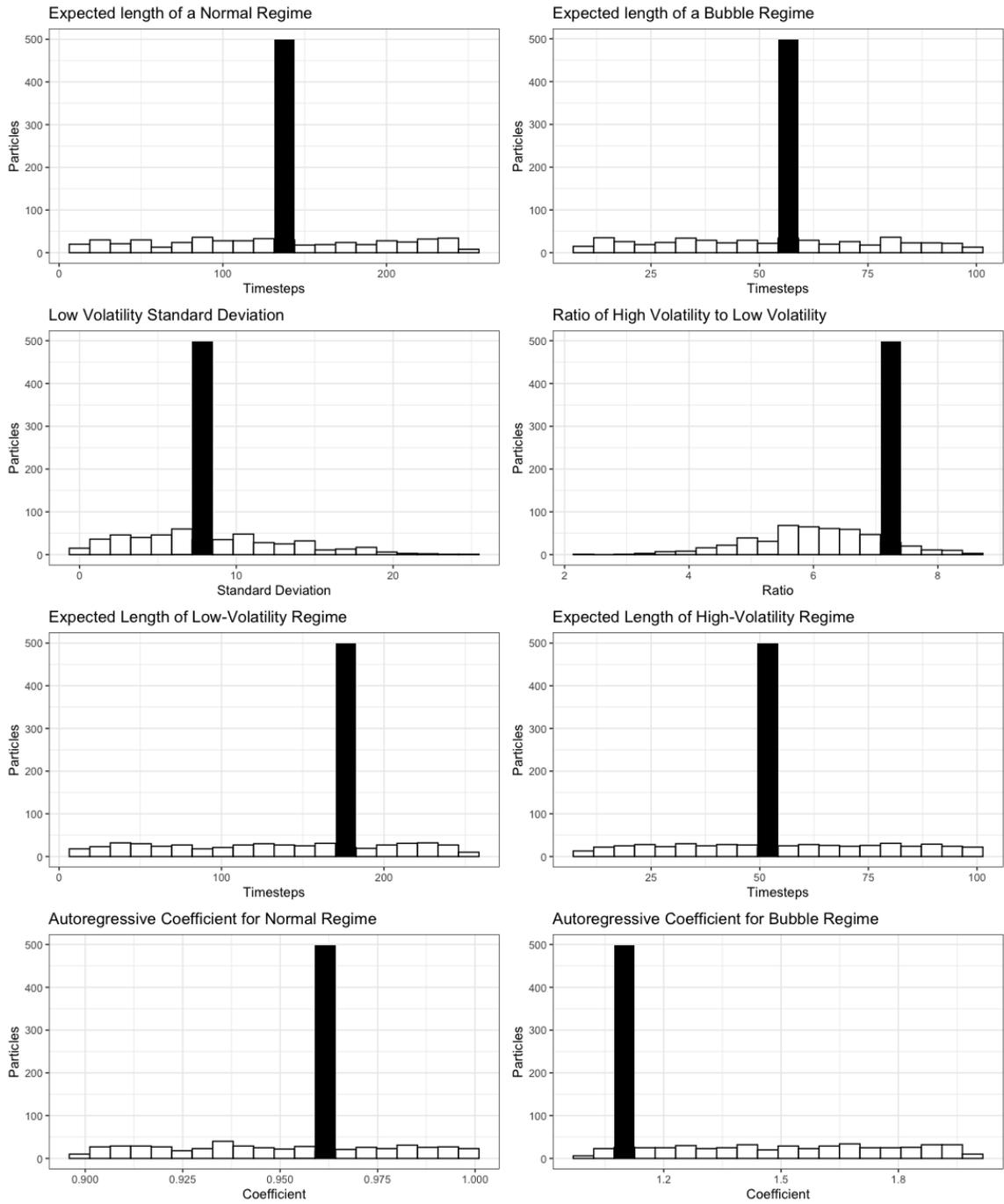


Figure 10: Histograms of the Distribution of Parameters in the Real Data

## V. Trading Strategy

Four trading strategies are considered. First, the buy-and-hold strategy; the investors buys Yen on June 1st, 1984, the first day the algorithm is active, and sells Yen on December 1st, 2019, the day the analysis ends. The second is a directional strategy; when the exchange rate is higher than in the previous month, the investor buys Yen, and the investor sells when it is lower. Third is a simple bubble-trading strategy wherein the investor buys when the start of a bubble is stamped and sells when the bubble is ended. The final strategy, the directional bubble strategy, combines strategies 2 and 3; the investor buys when the exchange rate goes up during a bubble, and sells either when the exchange rate falls or when the bubble ends.

For the bubble-based strategies, the choice of  $\zeta$  is important. All five values of  $\zeta$  compared in Figure 8 are tested. Each strategy starts with \$1 in June, 1984. All strategies buy as much as possible at each buy signal, then sell all of their holdings at each sell signal. Table 4 shows the end equity, Sharpe Ratio, and Sortino Ratio of each strategy.

There was a happy accident; by typo, the author tested  $\zeta = 999$  in addition to the others. The results for this are also in the Table 4. This very high value of  $\zeta$  returned low equity, but with the highest Sharpe ratio and by far the highest Sortino ratio of any tested strategy. Higher  $\zeta$  indicates more persistent regimes, a more conservative approach to bubble-stamping.

Table 4: Trade Statistics for Each Strategy

Strategy	End Equity	Sharpe Ratio	Sortino Ratio
Buy and Hold	2.1301		
Direction	5.5486	0.2221	1.0615
Simple Bubble, $\zeta = 1$	2.2789	0.2866	2.1923
Simple Bubble, $\zeta = 5$	2.0471	0.2862	2.145
Simple Bubble, $\zeta = 16$	1.7168	0.3216	1.8022
Simple Bubble, $\zeta = 60$	1.5537	0.3219	1.2843
Simple Bubble, $\zeta = 99$	1.7355	0.3229	1.7854
Simple Bubble, $\zeta = 999$	1.3494	0.4247	1.3713
Directional Bubble, $\zeta = 1$	1.7725	0.5643	2.9739
Directional Bubble, $\zeta = 5$	1.5955	0.5908	2.9214
Directional Bubble, $\zeta = 16$	1.5087	0.587	2.9557
Directional Bubble, $\zeta = 60$	1.4353	0.5039	2.5702
Directional Bubble, $\zeta = 99$	1.4522	0.5566	2.87
Directional Bubble, $\zeta = 999$	1.4477	0.7057	4.1868

The simple directional strategy has by far the highest end equity, but it has low Sharpe and Sortino ratios; this strategy faces more risk than the bubble strategies. The high ratios and low end equity of the high- $\zeta$  directional-bubbles strategy comes because it has only 2 losing trades out of 8; the win/loss ratio is 3:1. The average winning trade makes a 7.3% return, while the average loser is only 2.2%. By comparison, the simple directional strategy has a win/loss ratio of 24/31, 48 winners and 62 losers; the average winning trades has a 6.5% return, while the losing trades average a 1.8% loss.

The equity in the directional strategy is in Figure 11. The equity passes the buy-and-hold strategy in the mid 1980s, and doesn't face a big drawdown until the mid 1990s. The strategy faces

many drawdowns, for years at a time. Hence the low Sharpe ratio.



Figure 11: Equity Growth in the Simple Directional Strategy

The equity in the simple bubble strategies are in Figure 12; higher  $\zeta$  generally has a negative impact on the end equity, but also has comparatively shallow drawdowns. The only bubble strategy that has higher final wealth than the buy-and-hold strategy is the Simple Bubble Strategy with  $\zeta = 1$ .

The equity in the directional bubble strategies are in Figure 13; as in the simple bubble strategy, higher  $\zeta$  has a negative effect on the end equity. In the directional strategy, the highest  $\zeta$  eliminated all drawdowns except for those at the very end, which are flattened.

### **Ideas to Improve the Strategies**

The greatest flaw in the trading strategies analyzed here is that they only trade USD/JPY. The inclusion of more exchange rates in the strategy will allow the investor to diversify their portfolio and would reduce time spent outside the market. Including more exchange rates would require the particle filter to be run on several time series at once, making the algorithm much more computationally expensive. To run the filter on this one time series with 427 steps took several days; backtesting a multi-asset strategy based on this algorithm would be time intensive. Since the data is monthly, live-trading based on such a strategy should not be a problem.

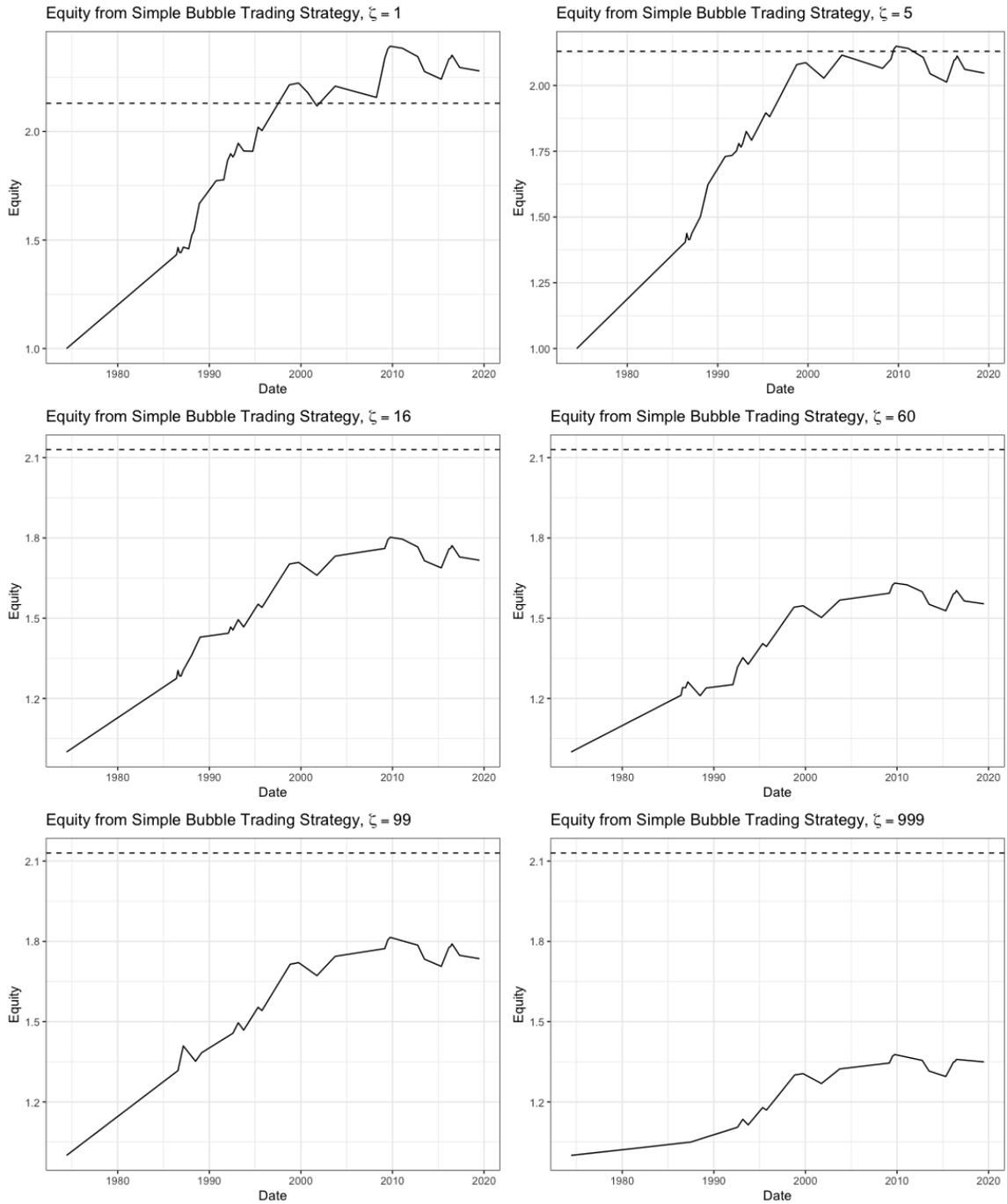


Figure 12: Equity plots for each of the Simple Bubble Strategies

The second flaw in the trading strategies is that they have relatively few trades; the highest Sharpe-ratio strategy has only 8 trades over 45 years! To improve this, it might be reasonable to apply the algorithm to higher-frequency data. Weekly data should be computationally feasible, and is

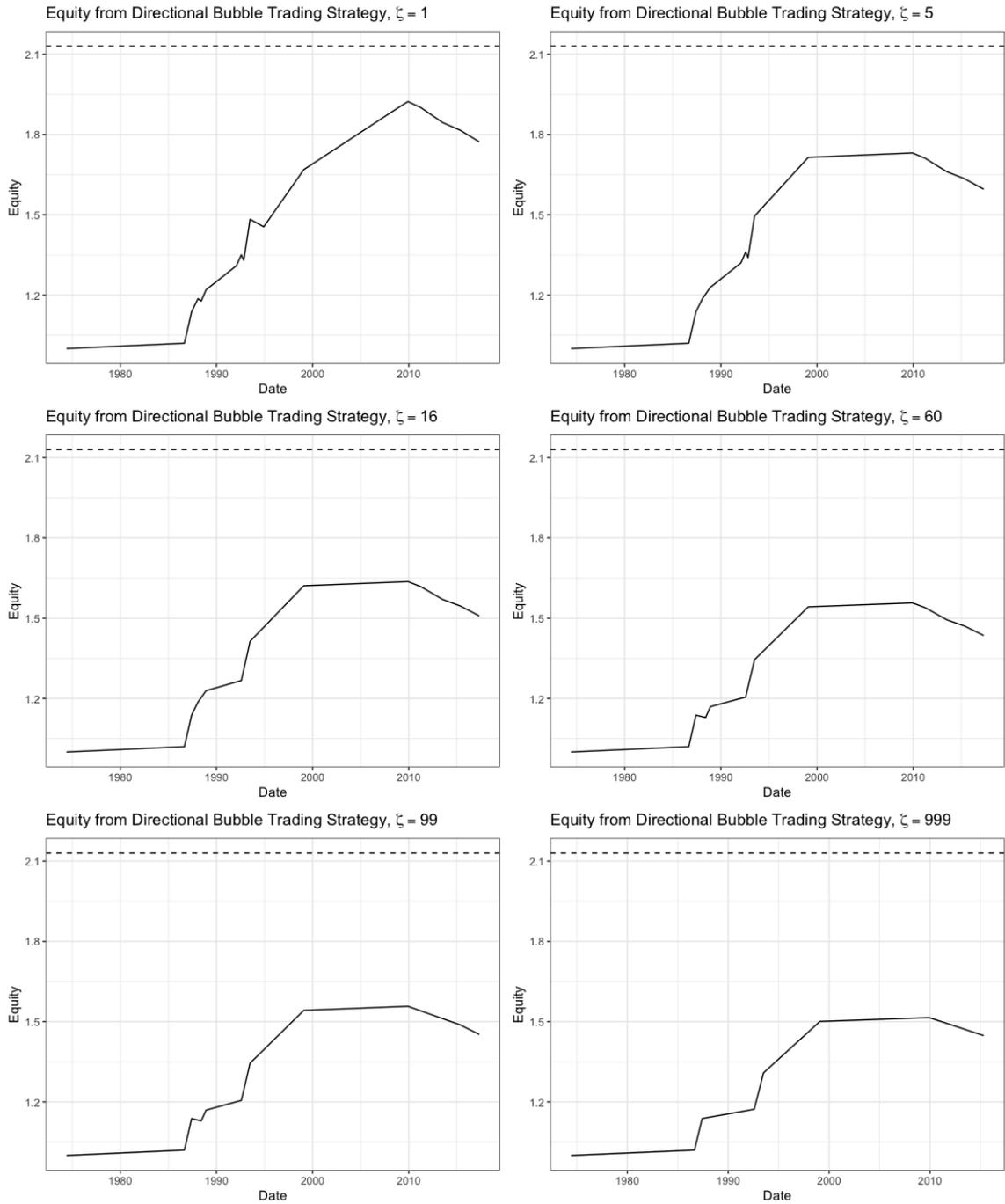


Figure 13: Equity plots for each of the Directional Bubble Strategies

the frequency used in Milunovich et al (2019) for a similar bubble-trading strategy. Daily data may also be feasible when there is no PMMH update. A PMMH update at  $t > 300$  took more than one day on this data.

A combination of the two improvements, weekly trading of several currencies, should make this strategy more worthwhile to a retail trader, although it would be quick time-consuming to backtest.

The best thing about the strategies tested is that the asset traded is among the most liquid and the trading indicators are long-lived. The bubble-trading strategies should have high capacity; increasing portfolio size should have little effect on the performance of the strategies.

## VI. References

- Al-Anaswah, N., & Wilfling, B. (2011). Identification of speculative bubbles using state-space models with Markov-switching. *Journal of Banking and Finance*, 35(2011), 1073-1086. <https://doi.org/10.1016/j.jbankfin.2010.09.021>
- Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov-chain Monte Carlo methods. *Journal of the Royal Statistical Society* 72(3) 269-342. <https://www.jstor.org/stable/40802151>
- Andrieu, C., Doucet, A., Punskeya, R. (2001). Sequential Monte Carlo methods for optimal filtering. In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential Monte Carlo methods in practice* (pp. 79-95). Springer. <https://doi.org/10.1007/978-1-4757-3437-9>
- Berquini, C., & Gilks, W. (2001). RESAMPLE-MOVE filtering with cross-model jumps. In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential Monte Carlo methods in practice* (pp. 117-138) Springer. <https://doi.org/10.1007/978-1-4757-3437-9>
- Bettendorf, T., & Chen, W. (2013). Are there bubbles in the Sterling-dollar exchange rate? New Evidence from sequential ADF tests. *Economics Letter* 120(2013), 350-353. <http://dx.doi.org/10.1016/j.econlet.2013.04.039>
- Cappe, O., Moulines, E., & Ryden, T. (2005). *Inference in Hidden Markov Models*. Springer.
- Charemza, W. W. (1996). Detecting stochastic bubbles on an East European foreign exchange market: An estimation/simulation approach. *Structural CHange and Economic Dynamics* 7(1996) 35-53.
- Chen, G., & Yan, C. (2011). A financial engineering approach to identify stock market bubbles. *Systems Engineering Procedia* 2(2011) 153-162. <https://doi.org/10.1016/j.sepro.2011.10.018>

- Chopin, N., Jacob, P., & Papaspiliopoulos, O. (2013). *SMC<sup>2</sup>*: An efficient algorithm for sequential analysis of state space models. *Royal Statistical Society*, 75(3), 397-426. <https://jsotr.org/stable/24772731>
- Crisan, D. (2001). Particle filters: A theoretical perspective. In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential Monte Carlo methods in practice* (pp. 17-41) Springer. <https://doi.org/10.1007/978-1-4757-3437-9>
- Doucet, A., de Freitas, N., & Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential Monte Carlo methods in practice*. Statistics for Engineering and Information Science. Springer.
- Fantazzini, D. (2016). The oil price crash in 2014/15: Was there a (negative) financial bubble? *Energy Policy* 96(2016) 383-396. <https://dx.doi.org/10.1016/j.enpol.2016.06.020>
- Fearnhead, P. (1998). *Sequential Monte Carlo methods in filter theory* [Unpublished doctoral dissertation]. Merton College, University of Oxford.
- Flood, R. P., & Garber, P. M. (1980). Market fundamentals versus price-level bubbles: The first tests. *Journal of Political Economy* 88(4), 745-770. <https://www.jstor.org/stable/1837311>
- Flood, R. P., Garber, P. M., & Scott, L. O. (1984). Multi-country tests for price-level bubbles. *Journal of Economic Dynamics and Control* 8(1984), 329-340. [https://doi.org/10.1016/0165-1889\(84\)90011-3](https://doi.org/10.1016/0165-1889(84)90011-3)
- Fulop, A., & Li, J. (2013). Efficient learning via simulation: A marginalized resample-move approach. *Journal of Economics* 176(2013) 146-161. <http://dx.doi.org/10.1016/j.jeconom.2013.05.002>
- Fulop, A., Li, J., & Yu, J. (2015). Self-exciting jumps, learning, and asset pricing implications. *The Society for Financial Studies* 28(3) 876-912. <https://www.jstor.org/stable/24465729>
- Fulop, A., & Yu, J. (2017). Bayesian analysis of bubbles in asset prices. *Econometrics* 4(47). <https://doi.org/10.3390/econometrics5040047>
- Hrabovska, Y. (2017). *A Markov-switching model for bubble detection in the stock market* [Unpublished master's thesis]. Universita Ca'Foscari Venezia.
- Kirman, A., Ricciotti, R. F., & Topol, R. L. (2007). Bubbles in foreign exchange rates. *Macroeconomic*

*Dynamics* 11(1) 102-123. <https://doi.org/10.1017/S1365100507060257>

- Macerinkiene, I., & Balciunas, A. (2013). Fundamental exchange rate forecasting models. Advantages and drawbacks. *KSI Transactions on Knowledge Society* 6(3) 9-17. <http://dx.doi.org/10.2139/ssrn.2633138>
- Milunovich, G., Shi, S., & Tan, D. (2018). Bubble detection and sector trading in real time. *Quantitative Finance* 19(2) 247-263. <https://doi.org/10.1080/14697688.2018.1459811>
- Okina, K. (1984). Rational expectations, bubbles, and foreign exchange markets. *BOJ Monetary and Economic Studies* 2(1) 81-117. <https://www.imes.boj.or.jp/research/papers/english/me2-1-3.pdf>
- Svensson, A., & Schon, T. B. (2016). *Comparing two recent particle filter implementations of Bayesian system identification*. Division of Systems and Control, Department of Information Technology, Uppsala University. <https://www.it.uu.se/research/publications/reports/2016-008/>
- Woo, W. T. (1987). Some evidence of speculative bubbles in the foreign exchange markets. *Journal of Money, Credit, and Banking* 19(4) 499-514. <https://www.jstor.org/stable/1992617>
- Wu, Y. (1995). Are there rational bubbles in foreign exchange markets? Evidence from an alternative test. *Journal of International Money and Finance* 14(1) 27-46.