Wirtschafts- und Sozialwissenschaftliche Fakultät
der Christian-Albrechts-Universität
Institut für Volkswirtschaftslehre

Seminar Paper Financial Economics
Estimation of Financial Markets Model from Kouwenberg & Zwinkels (2015) with MCMC methods

Prof. Dr. Thomas Lux
Summer Term 2024

Projensdorfer Straße 155
24106 Kiel
Handed in on 30.06.2024

Luca Preuße
Enrolment Number 1182372
Quantitative Economics, 3rd term

Table of Contents

1. Introduction 1
2. Model 2
3. Estimation Methods 3
4. MCMC 4
5. Data and Initialization 6
6. Baseline Results 7
7. Nonparametric Estimation of the Likelihood 10
8. Bandwidth Results 12
9. Summary 15
References 16

Appendix **17**

A.1 Model Details 17
A.2 MLE and Posterior Derivation 19
A.3 Data 21
A.4 Convergence 22
A.5 Bandwidth Selection 24

1. Introduction

Representative agent models that assume homogeneous rational agents and efficient markets are often incapable of explaining stylized facts in financial time series as well as the recent financial crisis (Lux & Zwinkels 2018, Kouwenberg & Zwinkels 2015). Therefore there has been recent interest in models that address agent-heterogeneity modeling the behavior and interactions of different types of agents and their effect on economic aggregates. These so called agent-based models (ABMs) can in principle create complex dynamics matching the complex dynamics we observe in real data.

As ABMs can be very expensive to simulate and estimate, the literature on estimation has been growing recently due to advances in computational power.

Estimation can be roughly dissected into Bayesian and frequentist methods as well as likelihood and moment-based methods. Estimation is often complicated by an analytically unavailable likelihood as well as unavailable theoretical moment conditions so that simple likelihood or moment approaches cannot be applied or have to be adapted. This issue can for example be overcome by writing the model in a nonlinear nongaussian state-space notation and estimating the likelihood with the help of a particle filter. Alternatively the model can be simulated and an estimation algorithm in an outer loop can take advantage of that simulated data. Both procedures however increase computation time drastically.

Most established in the literature are simulated minimum distance (SMD) methods which try to minimize a distance between simulated and observed data. However in the comparative study from Platt (2020) Bayesian methods - or as he argues generally likelihood based methods - outperform SMD methods. Numerical maximum likelihood can be used based on a particle filter estimation of the likelihood (Blevins 2016) or a nonparametric kernel density estimator (KDE) of the likelihood with simulated data (Kristensen & Shin 2012). Both the particle filter (Andrieu et al. 2010) and KDE (Grazzini et al. 2017, Zhang et al. 2023) can also be integrated into a Markov chain Monte Carlo (MCMC) or Sequential Monte Carlo (SMC) algorithm performing Bayesian inference. Grazzini et al. (2017) further suggest to use parametric estimation of the likelihood instead of nonparametric to decrease computation time and - at least in the models considered in their paper - without increasing the error due to misspecification too much. Another recent approach reconciling SMD methods and a Bayesian framework is Approximate Bayesian Computation which performs Bayesian inference by rejection sampling that is not likelihood based but based on a distance of moments between simulated and actual data. This approach is appealing as it makes more use

of moment information than SMD (Lux 2023) while keeping computation time comparatively small (Grazzini et al. 2017).

In this paper I will first present an ABM for the housing market that actually gives rise to a likelihood that is available analytically. I will estimate this model using Bayesian methods and a standard MCMC algorithm. Then I will alter this algorithm by replacing the analytical likelihood with a KDE estimate of the likelihood as introduced in Grazzini et al. (2017). I will analyze the role of the bandwidth of the KDE estimate for the efficiency of the whole MCMC sampling making use of the existence of an efficient sampler due to the analytically available likelihood. This may give important insights for applications of this algorithm for models which do not give rise to a computable likelihood.

2. Model

The model considered in this paper is the Multi-Agent housing market model from Kouwenberg & Zwinkels (2015) (from now on K&Z 2015) which will be presented here briefly to establish notation and intuition. For details I refer to the original paper or to A.1.

In the model the supply side is assumed to consist of constructors and the demand side to consist of consumers and investors. Aggregate constructor and consumer housing supply and demand depend linearly on the past period log-real house price (P_t). Investors are assumed to follow one of two rules to forecast prices. The fundamentalist rule is based on an error-correction mechanism that assumes a reversion of the prices towards a fundamental value F_t. The chartist rule forecasts prices by using the autocorrelation of housing returns considering the L most recent returns and extrapolating them. Investors demand is then a linear function of the expected future returns. The fraction of fundamentalists W_t is determined by a logit-switching rule that is based on past forecast performances of the two rules. The past forecast performance is calculated by absolute error loss using the most recent K observations.[1]

Under these assumptions one can solve for the equilibrium price which is given as a function of some deep parameters θ and past values of P_t and F_t. To get to the actual price some random noise ϵ_t is added so assuming one knows the lag orders K and L one can write:

[1] W_t is completely determined by the parameters and past observations and is therefore not a latent state variable.

$$P_t = f(X_t, \theta) + \epsilon_t \tag{1}$$
$$X_t = (P_{t-1}, F_{t-1}, ...P_{t-K-L-1})$$
$$\theta = (c, d, \alpha, \beta, \gamma)$$

The exact form of f is given in A.1. For θ, c and d summarize the parameters of the aggregate consumers demand and constructors supply. α<0 determines the speed of reversion of the actual house price towards its fundamental value used in the fundamentalist rule. β>0 is the extrapolation parameter in the chartist rule and γ>0 captures the sensitivity of the different types of investors to switch between rules due to better forecast performance. K&Z (2015) show that this model under certain parametrization - that is consistent with the data - leads to stable limit cycles and can therefore endogenously explain boom and bust cycles of the housing market.

3. Estimation Methods

Assuming $\epsilon_t \sim N(0,\sigma^2)$ according to equation (1) $P_t|X_t \sim N(f(X_t,\theta),\sigma^2)$.
Following K&Z (2015) F_t will be treated as exogenous and will be dropped in the conditioning set to simplify notation. Using the conditional-marginal factorization we can then given some known presample values written as the vector P_0 write the likelihood of the full sample $P_{1:T}$

$$p(P_{1:T}|\theta, h) = p(P_1|P_0, \theta, h)p(P_2|P_{0:1}, \theta, h)...p(P_T|P_{1:T-1}, \theta, h) \tag{2}$$
$$\propto h^{T/2} exp(-0.5h \sum_{t=1}^{T} (P_t - f(X_t, \theta))^2)$$

where $h=\sigma^{-2}$
Given this likelihood one estimation approach is to numerically maximize the (log-) likelihood with respect to θ and h. This is the estimation approach used in K&Z (2015).

In this paper we will however consider Bayesian estimation methods. The goal of Bayesian estimation is not to find the (assumed to be fixed in population) values θ and h that maximize the likelihood but to find the full posterior distribution $p(\theta,h|P_{1:T})$ so the distribution of θ and h after seeing the data. From this one can also derive a marginal posterior of θ and, if one

desires point estimation for a comparison to frequentist methods, one can compute the posterior mean, median or mode that all minimize a different expected loss function respectively (Platt 2020).

From Bayes rule it follows that the posterior is proportional to $p(P_{1:T}|\theta,h)p(\theta,h)$. The first term in the product is the likelihood while the second term is a prior distribution representing prior beliefs about the parameters.

Advantages of Bayesian estimation are that one can model prior beliefs explicitly and that one aims to find a whole posterior distribution instead of just a single point estimate. This might uncover interesting properties of the model like multiple local maxima which would be lost by just reporting a global maximum of the likelihood. Often in agent based models only flat priors are considered (for example Lux & Zwinkels 2018 or Platt 2020). Here we will do the same. Specifically I set

$$p(\theta, h) \propto h^{-1} * 1(h > 0) \tag{3}$$

This prior can be seen as a special case of an independent Normal-Gamma-prior with prior variances going to infinity so the improper prior becomes flat and uninformative (details in A.2).

4. MCMC

Since $f(X_t,\theta)$ is a general function, the posterior distribution $p(\theta,h|P_{1:T})$ is of unknown form. To uncover its properties one can use MCMC algorithms. The idea is to generate draws $\theta_{(1)},\theta_{(2)}$... from a converging Markov chain with its stationary distribution being the posterior distribution. After convergence these draws can then be used to approximate posterior moments by sample averages. Two of the most important MCMC algorithms, which I will use to estimate the model, are Gibbs sampling and the Random Walk Metropolis Hastings algorithm (RWMH).

Gibbs sampling is based on dividing the parameters of interest into different subsets (here in θ and h) and using the conditional-marginal factorization

$$p(\theta|P_{1:T}) = \int p(\theta, h|P_{1:T})dh = \int p(\theta|P_{1:T}, h)p(h|P_{1:T})dh$$

and likewise

$$p(h|P_{1:T}) = \int p(h|P_{1:T}, \theta)p(\theta|P_{1:T})d\theta$$

which implies that taking a draw from a marginal posterior from one subset, conditioning on that draw and drawing from the other subset is the same as drawing from the joint posterior. The draw can then be used as a new draw from the marginal to condition on to take a draw from the conditional of the other subset. For some starting value this process can then be repeated as many times as desired to generate draws from the posterior. Under some generality conditions, draws from this algorithm converge towards draws from the posterior regardless of the starting value (see for example Roberts & Smith 1994). This algorithm is useful when the conditional posterior is easy to take draws from. This is the case here as it can be shown that $p(h|P_{1:T},\theta)$ follows a Gamma distribution (A.2).

$p(\theta|P_{1:T},h)$ however is still an unknown distribution and therefore I use RWMH to sample from it. This algorithm proposes draws θ^* from a multivariate normal $\theta^* \sim N(\theta_{(s-1)},\Sigma)$. Then one calculates the ratio

$$R = \frac{p(\theta^*|P_{1:T}, h)}{p(\theta_{(s-1)}|P_{1:T}, h)} = \frac{p(P_{1:T}|\theta^*, h)}{p(P_{1:T}|\theta_{(s-1)}, h)} \tag{4}$$

$\theta_{(s)}$ is set to θ^* with probability min(R,1) and else is set to $\theta_{(s-1)}$. This again generates a Markov chain with convergence towards the posterior regardless of the starting value (Roberts & Smith 1994). This algorithm requires a tuning parameter Σ which represents the step size of the proposal distribution. Is this step size too large, many draws will be from space with no posterior density and will therefore be rejected. Is this step size too small, many draws will be in close proximity to the old draw leading to an acceptance probability close to 1. Both cases are undesirable for the efficiency of the algorithm. This implies however that the overall acceptance rate is informative about the efficiency of the algorithm and can therefore be targeted. A rule of thumb suggests acceptance probabilities between 0.23-0.44, which is based on considering normal distributions with infinite and one dimension(s) as the target distribution (Gelman et al. 1996). This targeting can be either done by some automatic updating of Σ (Rosenthal 2011) or by manual tuning of Σ. I will resort to the latter as this worked well for this rather small model and the results were robust against different Σ as well. The full algorithm looks as follows:

RWMH (within Gibbs) algorithm

1. Set tuning parameter Σ and some starting value $\theta_{(0)}$. Further set the desired number of draws S_1 and a number of burn-in draws S_0 that will be discarded in the end to ensure convergence
2. Draw $h_{(s)}$ from $p(h|\theta_{(s-1)},P_{1:T})$ which follows a Gamma distribution
3. Draw θ^* from $N(\theta_{(s-1)},\Sigma)$
4. Calculate R according to equation (4) setting $h=h_{(s)}$ [2]
5. Accept θ^* with probability min(R,1). Set $\theta_{(s)}=\theta^*$ if θ^* is accepted, else set it to $\theta_{(s-1)}$
6. Repeat steps 2-5 $S=S_0+S_1$ times and use the last S_1 draws as draws from the posterior

5. Data and Initialization

Following K&Z (2015) the data to calculate P_t and F_t is based on the aggregate stock of owner-occupied housing in the United States developed by Davis et al. (2008). The data on prices and rents can be found at the American Enterprise Institute at https://www.aei.org/historical-land-price-indicators/ under "Aggregate U.S. Rent-Price Ratio". The quarterly data is available from 1960 Q1 until 2018 Q2. Since the data is available in nominal terms and the model assumes real prices, I transform them into real values using the CPI "Consumer Price Index for All Urban Consumers: All Items in U.S. City Average" available at FRED.

F_t is constructed as in K&Z (2014) assuming F_t is equal to the present value of all future expected rent payments and using real-time available historical observations as a rolling average to estimate expected rent growth and rent yield. Plots on P_t and F_t can be found in A.3.

To select K and L I follow K&Z (2015) and apply a grid search over all combinations of K and L between 1 and 12 and choose the one that maximizes the likelihood.[3] This procedure selects K=1 and L=3. A Monte-Carlo simulation calibrated around the maximum likelihood estimates (MLE) and the actual chosen lag number shows that this procedure selects the true Lag-Order 897 out of 1000 times.[4]

The starting value for the MLE has been selected similar as in K&Z (2015) using the estimates from a restricted model setting $\gamma=0$ and then using $\gamma=2$ as a starting value which is

[2] To avoid computational issues it is useful to calculate R=exp(log(R)) and apply the log to simplify R to avoid numerical roundings that lead to 0/0 or ∞/∞ expressions.
[3] The sample was cut such that for all values of K and L the same sample size could be used to calculate the Likelihood.
[4] Such simulations have also been used to make sure that the methods estimate the true parameters consistently to rule out coding errors.

close to the estimates in the paper. Then the function fmincon integrated in Matlab was used to find the maximum numerically. The starting value for the MCMC algorithm is then set to the MLE.

Σ has been tuned manually with a small chain length by starting with $\Sigma \propto I_5$ adjusting a constant multiplier such that the overall acceptance rate is between 0.01 and 0.99 and then using the resulting draws variance to build the diagonal variance matrix V and then finally setting $\Sigma=\lambda V$ adjusting λ such that the overall acceptance rate is within the rule of thumb range.

The results are based on mixing 4 chains running in parallel with a burn-in of S_0=10,000 and final draws for each chain S_1=1,000,000.[5]

6. Baseline Results

Figure 1 shows the histograms of the MCMC draws and Table 1 summarizes some statistics as well as the maximum likelihood estimates. Noticeable differences to K&Z (2015) is that I find a much lower (in absolute) speed of reversion towards the fundamental price (α=-0.63 in their paper) and a weaker tendency to switch between the forecasting rules (γ=2.18 in their paper). The value of α found in K&Z (2014), who estimate a special case of the model used here, is closer to our estimates with −0.1692.

Generally the MLE and posterior statistics are quite close to each other (within 1- posterior standard deviation range). One noteworthy exception is α for which the posterior mean is larger than the MLE. The reason is that the posterior of α has multiple local maxima which can be seen in Figure 1 and as pointed out by Platt (2020) Bayesian point estimation with uninformative priors is only equivalent to MLE for the posterior mode (which is indeed close to the MLE for α) and if mode and mean/median differ, they also differ from the MLE generally.

[5] Some discussion on convergence can be found in A.4 but generally convergence did not seem to be an issue. Also experiments with a much longer chain and a Σ 1000 times larger than the actual Σ, leading to a very low acceptance rate, gave similar results so the whole posterior space seems to be explored. Also different starting values, specifically the estimates of K&Z (2015), made no noticable difference with a larger burn-in.

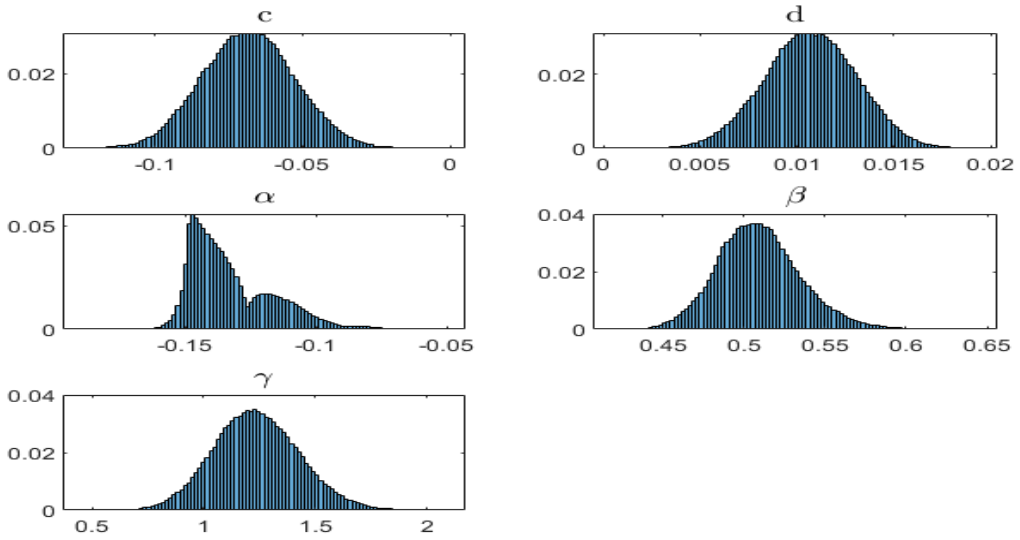

Figure 1. Histograms MCMC draws

Table 1. Posterior Statistics and MLE

Parameters	c	d	α	β	γ
MLE	-0.0749	0.0117	-0.1484	0.4986	1.2838
Mean	-0.0682	0.0107	-0.1327	0.5093	1.2362
Median	-0.0682	0.0107	-0.1372	0.5079	1.2305
Mode	-0.0678	0.0104	-0.1482	0.5069	1.2252
Standard Deviation	0.0156	0.0024	0.0160	0.0256	0.1904

MLE estimates of parameters and posterior statistics based on draws from RWMH algorithm. Acceptance rate is between 24.46%-24.52% across the 4 chains. The final number of observations after putting the required Lags for the model and the calculation of F_t in a pre-sample is 224. The mode was calculated using a histogram with 200 bins and choosing the midpoint of the bin with the largest number of observations..

One may further be interested in the estimated evolution of the share of fundamentalists/chartists, measured by W_t, over time. In this framework this would be the posterior

$$p(W_{1:T}|P_{1:T}) = \prod_{t=1}^{T} p(W_t|P_{1:T}, W_{1:t-1}) = \prod_{t=1}^{T} \int p(W_t|\theta, P_{1:T}, W_{1:t-1}) p(\theta|P_{1:T}, W_{1:t-1}) d\theta$$

Due to Bayes theorem

$$p(\theta|P_{1:T}, W_{1:t-1}) \propto p(W_{1:t-1}|P_{1:T}, \theta) p(\theta|P_{1:T})$$

which is only nonzero for θ that generated $W_{1:t-1}$ as W_t conditional on θ and the sample is known and not random (see A.1). For the same reason $p(W_t|\theta,P_{1:t},W_{1:t-1})=p(W_t|\theta,P_{1:t})$. This suggests to draw $W_{1:T}$ by drawing θ from the posterior and computing all $W_{1:T}$ given that draw. Samples from the posterior of $W_{1:T}$ can then be easily obtained as a byproduct of the posterior sampler of θ. The posterior median and 2.5% and 97.5% quantiles of W_t for each t are shown in Figure 2. As in K&Z (2014, 2015) we see a dominance of chartists preceding the financial crisis in 2008-2009, although less severe and for a shorter period of time compared to their papers, followed by an increase of fundamentalists during the crisis.

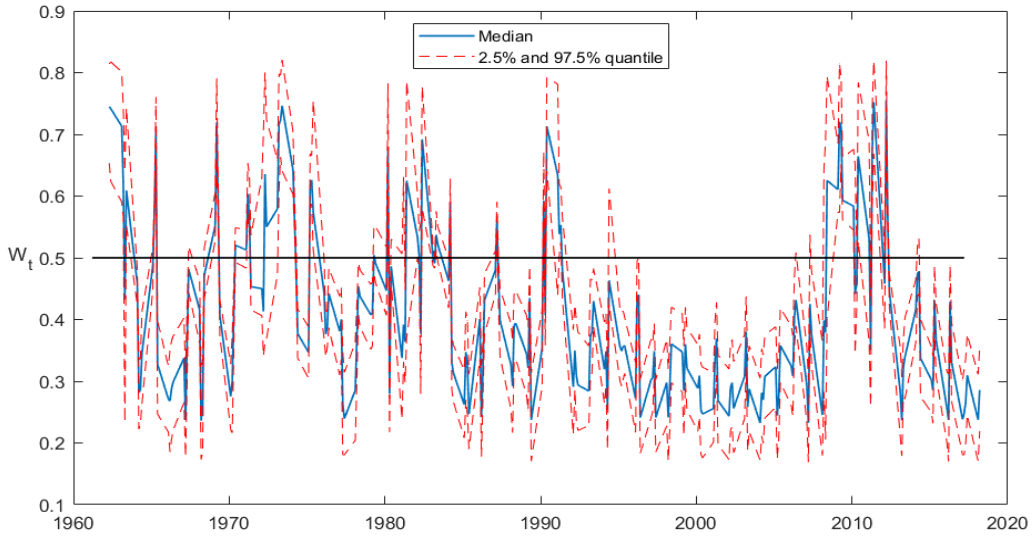

Figure 2 Posterior W_t (share fundamentalists) over time

The insight of K&Z (2015) is that the model endogenously produces stable boom-bust cycles for the housing market. They construct them by setting F_t to some constant roughly corresponding to P_t at 1985 and simulating the model without noise for the MLE estimates. Our Bayesian approach allows us to consider these cycles accounting for estimation uncertainty by drawing θ from the posterior and computing the limit cycles for each draw (without noise). Figure 3 visualizes 20 randomly drawn cycles and Table 2 shows some quantiles based on simulated cycles for 10,000 random draws of the posterior. Compared to K&Z (2015) I find for the representative limit cycle - given by the median - a larger period length and a smaller amplitude. Also the price fluctuates around a higher value (relative to the fundamental price). The limiting behavior however varies across different values of θ

considering P_t is measured in logs and the house price cycle may cover a range of up to 23.5% and not just 13.5% when taking estimation uncertainty into account.

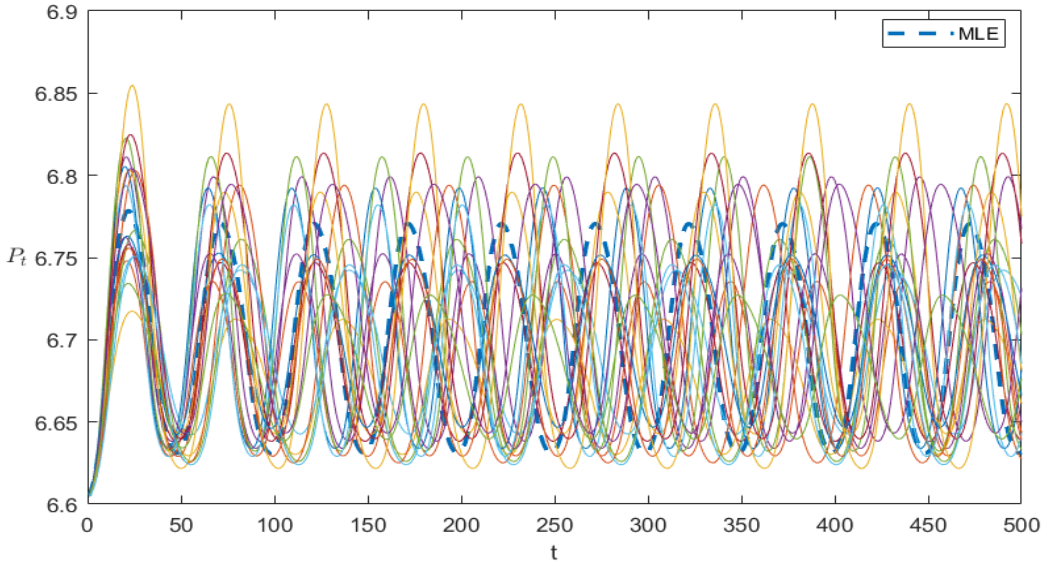

Figure 3. Stable Limit Cycles

Figure shows 20 different stable cycles computed with no noise, setting $F_t=P_{1985Q1}$ for all periods and θ randomly drawn from the posterior as well as the cycle under the MLE.

Table 2. Posterior Quantiles of stable limit cycles

Quantile	2.5%	50%	97.5%	MLE	K&Z
Period length	45	52	64	49	44
Minimum	0.0178	0.0310	0.0650	0.0282	0.007
Maximum	0.1054	0.1668	0.2681	0.1678	0.222
Amplitude	0.0657	0.1341	0.2346	0.1395	0.215

Quantiles for characteristics of stable limit cycle of P_t (log-real price) based on 10,000 random draws from the posterior and simulations without noise for every draw. Minimum and Maximum in differences to F_t. $F_t=P_{1985Q1}$ for all t. Values for K&Z (2015) based on Panel A Figure 2 in their paper. Periods in quarters.

7. Nonparametric Estimation of the Likelihood

The method considered for estimation here while appropriate for the model of choice is too specific for estimation of general ABMs as it requires a computable likelihood to evaluate the acceptance probability in the RWMH algorithm. Typically in ABMs the likelihood cannot be simply computed. Grazzini et al. (2017) suggest methods that work around this problem. Platt

(2020) shows that one of these approaches - the nonparametric KDE of the likelihood - fairs well against many frequentist alternative methods for a variety of models.

The suggested approach is to replace the likelihood in the MCMC algorithm with an estimate using a simulated sample and KDE.[6] As simplification I only apply this method to estimate the conditional posterior of p(θ|h,y) in step 4 of the MCMC algorithm and keep the rest of the algorithm as before.

KDE aims to estimate a density by fitting kernels at each observation and weighthing them based on the distance to the point at which the density is supposed to be evaluated. Formally for a density f(x) and a sample X_i of size N the KDE is

$$\hat{f}(x) = \frac{1}{N * bw} \sum_{i=1}^{N} K(\frac{X_i - x}{bw})$$

where bw is a bandwidth parameter, which plays a similar role as the binwidth of a histogram, and K(u) is a symmetric kernel density function integrating to 1 with E(u)=0. Here we will only consider the Gaussian kernel so K(u) is the standard normal pdf. While not applicable for all models we will here consider conditional estimation of the likelihood which should be more efficient and deals appropriately with the nonstationary data (Grazzini et al. 2017).[7] We then replace step 4 of the MCMC algorithm with

$$\hat{R} = \frac{\hat{p}(P_{1:T}|\theta^*, h_{(s)})}{\hat{p}(P_{1:T}|\theta_{(s-1)}, h_{(s)})} = \frac{\prod_{t=1}^{T} \hat{p}(P_t|P_{0:t-1}, \theta^*, h_{(s)})}{\prod_{t=1}^{T} \hat{p}(P_t|P_{0:t-1}, \theta_{(s-1)}, h_{(s)})} \tag{5}$$

$$\hat{p}(P_t|P_{0:t-1}, \theta, h_{(s)}) = \hat{f}(P_t) = \frac{1}{N * bw} \sum_{i=1}^{N} K(\frac{P_{t,i} - P_t}{bw})$$

where $P_{t,i}$ is simulated N times under parameter values θ and $h_{(s)}$ keeping all observations before t fixed.

[6] The simulation is done with the same seed for the proposed parameter and the old parameter which helps to smooth the Likelihood making the sampling more efficient (Zhang et al. 2023).

[7] Grazzini et al. (2017) mainly consider a less precise but more general approach for stationary data which is to run one long simulation to estimate the marginal density of the data and estimate the likelihood as a product of these densities. As pointed out by Platt (2020) this makes a (usually) wrong independence assumption which may lead to identification issues for some models. Gatti & Grazzini (2020) account for this by estimating the joint density of current and past observations or alternatively consider the limited-information likelihood discussed in Kristensen & Shin (2012) for more general estimation based on this idea.

The bandwidth is an important parameter to choose for the KDE. Higher bandwidths are associated with a larger bias but smaller variance of the KDE so choosing the bandwidth solves a bias-variance tradeoff.

Grazzini et al. (2017) choose the bandwidth by applying Silverman's rule of thumb which sets

$$bw = C * 1.059\hat{\sigma}N^{-0.2} \tag{6}$$

with C=1 where $\hat{\sigma}$ is the estimated standard deviation of the simulated sample. This bandwidth leads to the KDE being a consistent estimator of f(x) and minimizes the asymptotic mean integrated squared error (AMISE) assuming that f(x) is normally distributed (Silverman 1988 p. 45).

As far as I am aware, other bandwidths for KDE to estimate the likelihood have not been considered much in the literature. Platt (2020) does not make an explicit statement about the bandwidth and refers to Grazzini et al. (2017) for details. Zhang et al. (2023) use KDE of the likelihood integrated into an SMC algorithm and also make use of a Silverman bandwidth. Kristensen & Shin (2012) apply numerical MLE with a KDE of the likelihood and consider bandwidths in ranges of 15% around Silverman's rule of thumb and do not find big differences. MLE is however conceptually different from Bayesian inference and I will also consider bandwidths that are further away from the rule of thumb than just 15%.

Note that the actual density that is approximated in (5) is normal (equation (2)) and therefore the rule of thumb is indeed the AMISE minimizing bandwidth in this case.

8. Bandwidth Results

Figure 4 and Table 3 show the results based on S_0=10,000 and S_1=100,000 MCMC draws of this algorithm for different bandwidths.[8] KDE was applied using N=1000 simulated values.

Kullback-Leibler divergence (KLD) is a non-symmetric distance measure ≥0 between two densities is defined as

$$KLD = \int_{-\infty}^{\infty} log(\frac{p(\theta|P_{1:T})}{p(\theta_{bw}|P_{1:T})})p(\theta|P_{1:T})\, d\theta$$

We can use KLD to quantify the difference between the baseline results and the results from this KDE approach for the different bandwidths. The nominator is estimated using

[8] I reduced the MCMC draws compared to baseline because of generally higher computational demand. The time for one model run was roughly 6-7 hours for different bandwidths while the baseline RWMH with S_1=1,000,000 took 360 seconds (which included the sampling of W_t).

multivariate KDE with the baseline results and the denominator with the results for the respective bandwidth. The bandwidth matrix is set by using a multivariate generalization of Silverman's rule. Note that this density estimation is quite precise as we have 400,000 draws to compute the estimate of the posterior density (and 4,000,000 for the baseline). We then approximate the integral by a sample average over 1000 random draws from the baseline results.

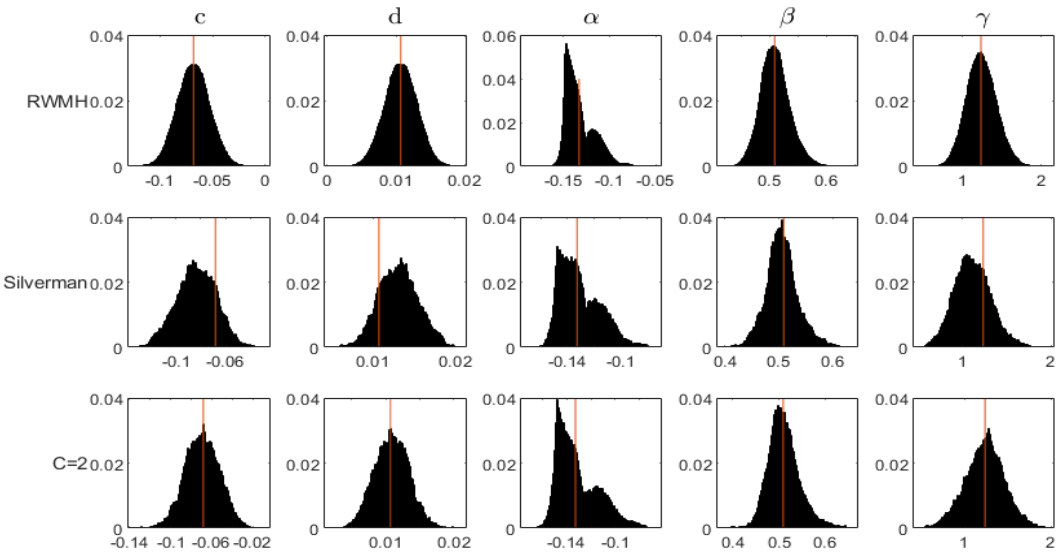

Figure 4. Posterior histograms for baseline and different bandwidths

Shows histograms of MCMC samples for baseline RWMH, using Silverman's rule of thumb bandwidth and using the bandwidth in (6) with C=2 which was the best performing in Table 3. The red line corresponds to the baseline posterior means.

Table 3. Posterior Moments and KLD to baseline for different bandwidths

	c	d	α	β	γ	KLD
Baseline	-0.0682 (0.0156)	0.0107 (0.0024)	-0.1327 (0.0160)	0.5093 (0.0256)	1.2362 (0.1904)	0
RWMH smaller chain	-0.0711 (0.0169)	0.0111 (0.0026)	-0.1352 (0.0151)	0.5075 (0.0250)	1.2387 (0.1897)	0.6081
C=1	-0.0836 (0.0161)	0.0131 (0.0024)	-0.1310 (0.0154)	0.5056 (0.0305)	1.1101 (0.2073)	2.7343
C=1.15	-0.0794 (0.0152)	0.0124 (0.0023)	-0.1314 (0.0156)	0.5051 (0.0293)	1.1446 (0.2102)	1.7640
C=1.25	-0.0776 (0.0157)	0.0122 (0.0024)	-0.1318 (0.0157)	0.5078 (0.0291)	1.1650 (0.2068)	1.2465
C=1.5	-0.0741 (0.0153)	0.0116 (0.0023)	**-0.1322** (0.0160)	0.5089 (0.0316)	1.1939 (0.2073)	0.7583
C=2	**-0.0694** (0.0171)	**0.0109** (0.0026)	-0.1314 (0.0172)	**0.5095** (0.0347)	**1.2271** (0.2283)	**0.5819**
C=3	-0.0665 (0.0211)	0.0104 (0.0032)	-0.1305 (0.0196)	0.5109 (0.0384)	1.2629 (0.2710)	0.7562
C=4	-0.0617 (0.0267)	0.0097 (0.0041)	-0.1262 (0.0233)	0.5088 (0.0524)	1.3057 (0.3530)	1.2470

Posterior means and standard deviation in brackets for baseline results (sampled with RWMH) and KDE approach for different bandwidth choices following (6). KDE algorithm with N=1000 simulations and S_0=10,000 and S_1=100,000 running 4 chains in parallel. RMWH smaller chain uses same chain length with the baseline algorithm. Last column shows estimated Kullback–Leibler divergence between RWMH and the KDE algorithm based on 1000 random draws from the RWMH algorithm and multivariate KDE. Closest means and KLD to baseline in bold.

We see improvements of the estimation by choosing bandwidths that are larger (although not too large) than the Silverman rule of thumb. This might be surprising as the actual density to approximate is normal for which the Silverman bandwidth is designed. The general approximation of the posterior is satisfying with the most efficient bandwidth even leading to a slightly lower estimated KLD to the baseline than the RWMH with the same chain length.

Some of the reasons for the failure of the Silverman rule of thumb are easy to understand. The KDE in the full algorithm is always evaluated at the same point - the actual data - so one does not necessarily care about the minimization of the AMISE which minimizes the error integrated over the whole space of the density and not just at the point of evaluation. Also what one is ultimately interested in within the MCMC algorithm is the precision of the estimation of min(R,1). Consider the extreme case for which $\theta^*=\theta_{(s-1)}$. Then obviously the

true R=1. Now consider a zero variance estimate of $p(P_t|P_{0:t-1},\theta,h)$ with arbitrary bias. This would lead to numerator and denominator in (5) to be the same so the estimate of R is equal to the true R with probability 1 so in this case ignoring bias and only caring about variance would lead to the best possible estimate of R even though the MSE of the KDE may be large.

A discussion on the bandwidth selection in practice can be found in A.5. Briefly I suggest to minimize a loss function between a pseudo-true and estimated R averaged over the evolution of the MCMC algorithm. For a more precise KDE estimation of the likelihood as the pseudo-true likelihood this works well in this case. An extension could be made by considering a particle filter estimation as the pseudo-true likelihood.

9. Summary

In this paper I estimated the agent based housing market model developed in Kouwenberg & Zwinkels (2015) with Bayesian methods based on an uninformative prior and an updated data set. Compared to their paper I find a weaker mean reversion towards the fundamental price and a weaker tendency to switch between forecasting rules (chartist and fundamentalist rules). I find a similar dominance of chartists in the decades preceding the financial crisis although the differences are less severe. My estimation results yield similar stable limit cycles as the original paper and our Bayesian approach allows us to account for estimation uncertainty when characterizing these cycles.

I further used the - for ABMs unusual - availability of an analytical likelihood function to evaluate the efficiency of a KDE estimate of the likelihood integrated in an MCMC sampling as introduced in Grazzini et al. (2017). I showed that the KDE Silverman rule of thumb bandwidth for such likelihood estimation is not optimal despite it being the AMISE minimizing bandwidth of the underlying density that needs to be estimated. Better bandwidth choices can increase the efficiency of the MCMC sampling and due to general computational demand of estimation of ABMs these efficiency gains may matter a lot in practice.

To what extent the latter results hold beyond the simplifications made here (conditional simulation, Gibbs-step with known conditional posterior, constant proposal variance) and applications for more challenging models is a task for further research. Further it would be interesting to compare the efficiency of this method with the most efficient bandwidth to alternative Bayesian methods like particle filter MCMC.

References

Andrieu, C., A. Doucet and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(3), 269–342.

Blevins, J. R. (2016). Sequential Monte Carlo Methods for Estimating Dynamic Microeconomic Models. Journal of Applied Econometrics, 31(5), 773-804.

Brock, W. A. and C. H. Hommes (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic Dynamics and Control, 22(8–9), 1235–1274.

Davis, M. A., A. Lehnert and R. F. Martin (2008). The Rent-Price Ratio for the Aggregate Stock of Owner-Occupied Housing. Review of Income and Wealth, 54(2), 279-284.

Gatti, D. D. and J. Grazzini (2020). Rising to the challenge: Bayesian estimation and forecasting techniques for macroeconomic Agent Based Models. Journal of Economic Behavior & Organization, 178, 875-902.

Gelman, A., G. O. Roberts and W. R. Gilks (1996). Efficient Metropolis Jumping Rules. Bayesian Statistics 5: Proceedings of the Fifth Valencia International Meeting (Oxford), 599-608.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting (Oxford), 169-194.

Grazzini, J., M. G. Richiardi and M. Tsionas (2017). Bayesian estimation of agent-based models. Journal of Economic Dynamics and Control, 77, 26-47.

Kouwenberg, R. and R. Zwinkels (2014). Forecasting the US housing market. International Journal of Forecasting, 30(3), 415-425.

Kouwenberg, R. and R. Zwinkels (2015). Endogenous Price Bubbles in a Multi-Agent System of the Housing Market. PLoS ONE 10(6): e0129070.

Kristensen, D. and Y. Shin (2012). Estimation of dynamic models with nonparametric simulated maximum likelihood. Journal of Econometrics, 167(1), 76-94.

Lux, T. and R. Zwinkels (2018). Empirical Validation of Agent-Based Models. Handbook of Computational Economics, 4, Ch. 8, 437-488.

Lux, T. (2023). Approximate Bayesian inference for agent-based models in economics: a case study. Studies in Nonlinear Dynamics & Econometrics, 27(4), 423-447.

Platt, D. (2020). A comparison of economic agent-based model calibration methods. Journal of Economic Dynamics and Control, 113.

Roberts, G. O. and A.F.M. Smith (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. Stochastic Processes and their Applications, 49(2), 207-216.

Rosenthal, J. S. (2011). Optimal proposal distributions and adaptive MCMC. Handbook of Markov chain Monte Carlo (Cambridge), 93-112.

Silverman, B. W. (1998). Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC, London.

Zhang, J., Q. Zhang, Y. Li and Q. Wang (2023). Sequential Bayesian inference for agent-based models with application to the Chinese business cycle. Economic Modelling, 126.

Appendix

A.1 Model Details

Here I write out the details for the model from Kouwenberg & Zwinkels (2015) that is the basis for this paper. Everything is identical to their paper except for slight differences in notation. All variables are in log-real terms.

Aggregate consumers housing demand is given by a linear function of the past period price (index) P_t

$$D_t^C = a_0 + b_0 P_{t-1} \tag{A1}$$

Similarly aggregate supply is given by

$$S_t^C = c_0 + d_0 P_{t-1} \tag{A2}$$

Investors demand for housing is a weighted average of the demand of chartists and fundamentalists.

$$D_t^I = W_t D_t^{fu} + (1 - W_t) D_t^{ch} \tag{A3}$$

Assuming that investors are mean-variance maximizers with the same risk aversion and belief about a constant variance of housing returns speculative demand is given by (shown by Brock & Hommes 1998):

$$D_t^j = k E_{t-1}^j (P_t - P_{t-1}) \tag{A4}$$

with k>0. Fundamentalists build their expectations on returns by assuming a mean reversion towards the fundamental price F_t

$$E_{t-1}^{fu}(P_t - P_{t-1}) = \alpha(P_{t-1} - F_{t-1}) \tag{A5}$$

with α<0 such that a reversion towards the fundamental price and negative returns are expected when the price is larger than the fundamental value.

Chartists build their expectations by extrapolating the past L observed housing returns

$$E_{t-1}^{ch}(P_t - P_{t-1}) = \beta \sum_{l=1}^{L}(P_{t-l} - P_{t-l-1}) \tag{A6}$$

The fraction of fundamentalists W_t is given by a logit-switching rule based on past observed forecast errors under the respective rule π_{t-1}^{fu} , π_{t-1}^{ch}

$$W_t = (1 + exp[\gamma(\frac{\pi_{t-1}^{fu} - \pi_{t-1}^{ch}}{\pi_{t-1}^{fu} + \pi_{t-1}^{ch}})])^{-1} \tag{A7}$$

γ>0 and represents the sensitivity towards forecasting performance. The forecast error is given by the sum of absolute error losses of the past K observed housing returns

$$\pi_{t-1}^{j} = \sum_{k=1}^{K} |E_{t-1-k}^{j}(P_{t-k} - P_{t-k-1}) - (P_{t-k} - P_{t-k-1})| \tag{A8}$$

Finally the change of prices is assumed to depend linearly on excess demand and unpredictable noise ϵ_t

$$P_t - P_{t-1} = f(D_t^C + D_t^I - S_t) + \epsilon_t = k * f(D_t^C/k + D_t^I/k - S_t/k) + \epsilon_t \tag{A9}$$

k and f are then set to 1 without loss of generality. Plugging in gives us

$$P_t = P_{t-1} + c + dP_{t-1} + W_t\alpha(P_{t-1} - F_{t-1}) + (1 - W_t)\beta \sum_{l=1}^{L}(P_{t-l} - P_{t-l-1}) + \epsilon_t \tag{A10}$$

with W_t as defined in (A7) and (A8), c=(a_0-c_0)/k, d=(b_0-d_0)/k.

This is the full form of equation (1). The highest Lag-Order that needs to be defined as a pre-sample is K+L+1 through plugging in the chartist rule in (A8).

To rule out coding errors when implementing this model I made sure to be able to replicate Figure 2 Panel A of Kouwenberg & Zwinkels (2015) by generating artificial data without noise with the estimates in their paper.

A.2 MLE and Posterior Derivation

Consider the log-likelihood from equation (2).

$$L(\theta,h) = log(p(P_{1:T}|\theta,h)) = constant + 0.5Tlog(h) - 0.5h\sum_{t=1}^{T}(P_t - f(X_t,\theta))^2$$

$f(X_t,\theta)$ is not continuously differentiable with respect to θ but note that since h>0 θ_{MLE} that maximizes the log-likelihood is the θ that maximizes

$$-0.5\sum_{t=1}^{T}(P_t - f(X_t,\theta))^2$$

so θ can be found numerically ignoring h.

The first order condition with respect to h is

$$\frac{dL}{dh} = 0.5\frac{T}{h} - 0.5\sum_{t=1}^{T}(P_t - f(X_t,\theta))^2 = 0$$

Using the MLE of θ this simplifies to

$$h_{MLE} = \frac{T}{\sum_{t=1}^{T}(P_t - f(X_t,\theta_{MLE}))^2}$$

Using the invariance property of the MLE we get $\sigma_{MLE}=h_{MLE}^{-0.5}$. or simply the root of the variance of the estimated residuals.

Consider the Gamma distribution with following notation

$$f(x|\nu,S) \propto x^{\frac{\nu-2}{2}}exp(-0.5x*\nu*S^2)*1(x>0)$$

with expectation S^{-2} and variance $2S^{-4}/\nu$.

A possible prior choice for θ and h is to assume they are independent and θ is multivariate normally distributed and h is Gamma distributed so

$$p(\theta,h) \propto exp(-0.5(\theta-\theta_0)'V^{-1}(\theta-\theta_0))*h^{\frac{\nu-2}{2}}exp(-0.5h*\nu*S^2)*1(h>0)$$

An uninformative choice sets their prior variance to infinity so ν,$V^{-1} \to 0$ which simplifies the improper prior to (3).

The conditional posterior of h can be shown to follow a Gamma distribution

$$p(h|P_{1:T}, \theta) \propto p(h, \theta|P_{1:T}) \propto p(P_{1:T}|\theta, h)p(\theta, h)$$

$$\propto h^{\frac{T}{2}} exp(-0.5h \sum_{t=1}^{T} (P_t - f(X_t, \theta))^2) h^{-1} = h^{\frac{T-2}{2}} exp(-0.5h \sum_{t=1}^{T} (P_t - f(X_t, \theta))^2)$$

which is the same kernel as a Gamma distribution with $\nu = T$ and

$$\nu * S^2 = \sum_{t=1}^{T} (P_t - f(X_t, \theta))^2$$

A.3 Data

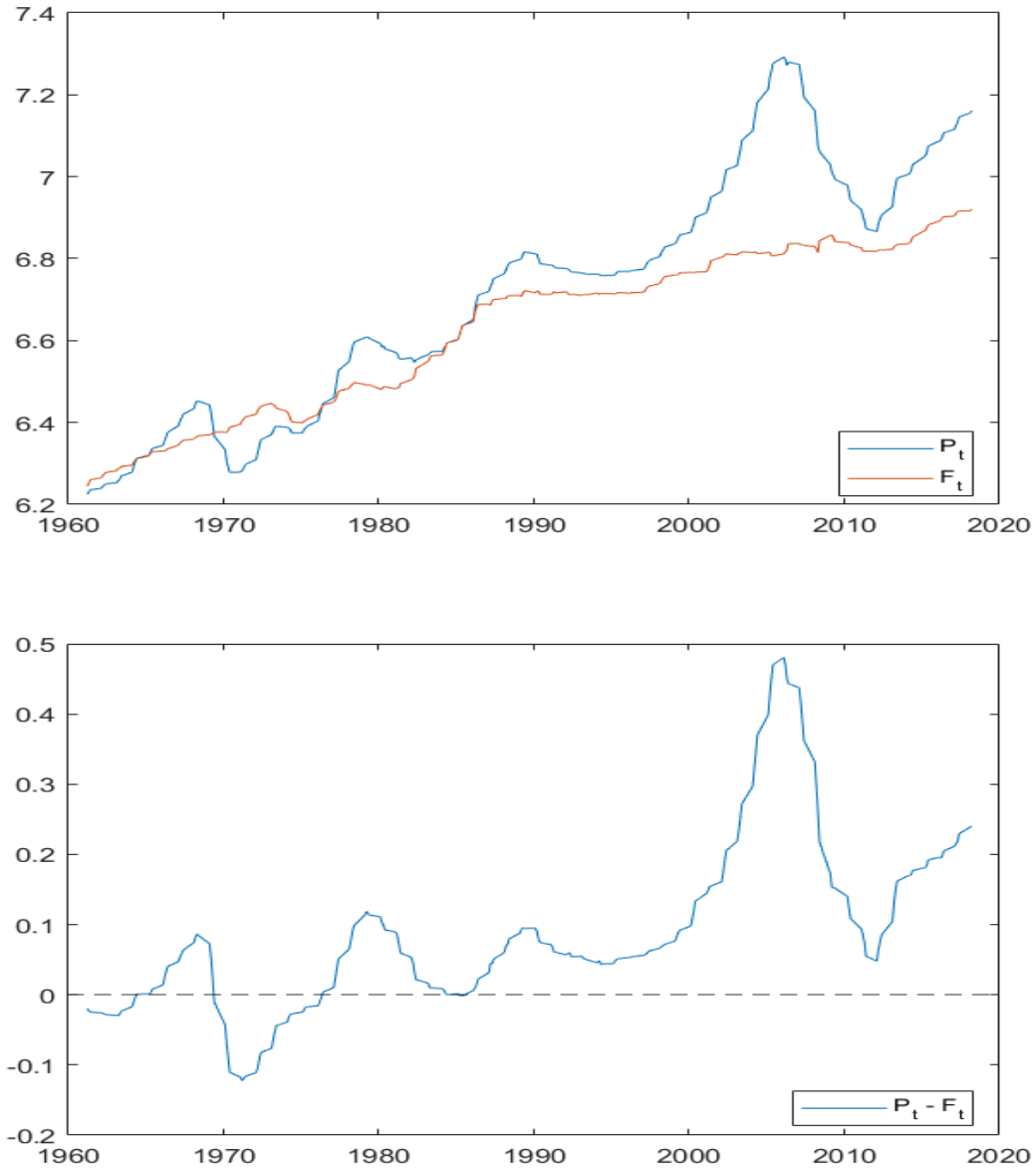

Figure A.1. Log Real Historical Prices and Fundamental Value and their Difference

A.4 Convergence

There exist several ways to test for convergence. Often it helps to just visually examine the evolution of the Markov chains and check whether they appear to follow the same stationary distribution which seems to be the case here (Figure A.2).

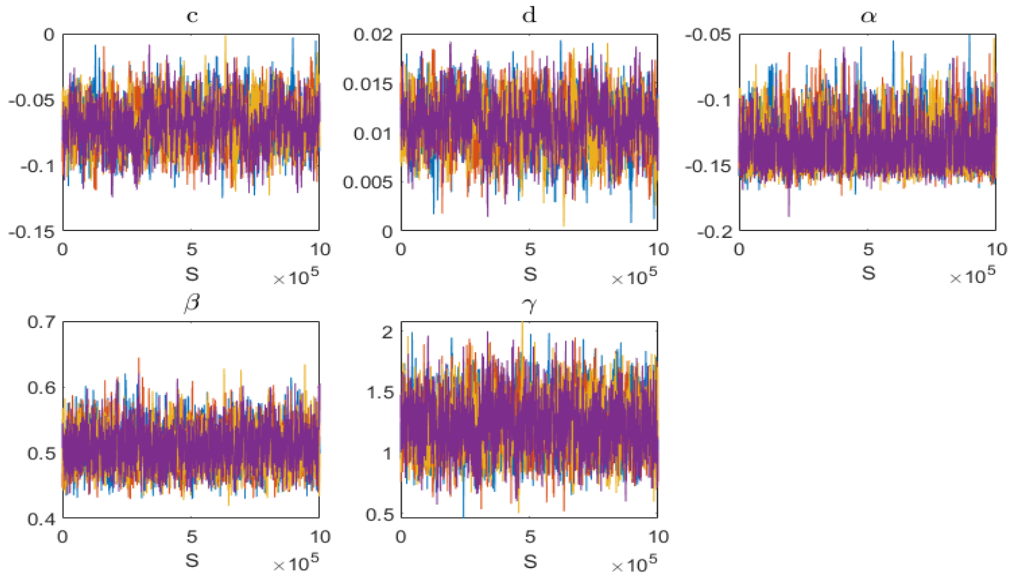

Figure A.2 Trace plots MCMC draws

Overlapping plots of draws after burn-in for 4 chains for baseline results.

One way to test for convergence is to test whether moments at different parts of the chain differ significantly (Geweke 1992). Consider the asymptotic distribution of a moment estimator from a Markov chain that has converged

$$T_A = A^{-1} \sum_{s=1}^{A} g(\theta_{(s),j}) \sim N(g(\theta_j), \frac{\lambda^2}{A})$$

$$\lambda^2 = Var(g(\theta_j)) + 2 * \sum_{k=1}^{A-1} \frac{A-k}{A} * Cov(g(\theta_{(s),j}), g(\theta_{(s-k),j}))$$

In practice the covariances used to estimate the long-run variance have to be truncated at some lag length (Newey West estimator). From this it follows that two independent statistics from the same stationary distribution follow for large A and B

$$T_A - T_B \sim N(0, \frac{\hat{\lambda_A}^2}{A} + \frac{\hat{\lambda_B}^2}{B})$$

which can be used to construct a standard normally distributed test statistics (CD). We do this here with the $g(\theta)=\theta$ for every parameter for every chain and following Geweke (1992) the subsample B is the last 50% (500,000) of the draws and A the first 10% draws so that the samples A and B are approximately independent because there are 400,000 draws in between for the autocorrelation to diminish. The truncation for the Newey West estimator is chosen at a fairly high value of 20,000 which roughly corresponds to the lag number it takes for the autocorrelation of the draws to disappear due to the high persistence of the RWMH algorithm. The results are summarized in Table A.1.

None of the statistics exceeds (in absolute) the 5% significance critical value of 1.96. From this and the graphical inspection I would not consider convergence to be a big problem despite the large autocorrelation. Also note that the results in Table 3 for a smaller chain are quite similar to the longer chain.

Table A.1 CD statistics

	c	d	α	β	γ
Chain 1	-1.4651	1.4676	-1.0111	0.3102	-1.0887
Chain 2	-0.6810	0.6944	-1.0492	0.7102	-0.6429
Chain 3	-0.3826	0.3764	0.1123	0.4182	-0.3986
Chain 4	0.0284	-0.0424	-0.2704	-1.6647	1.3356

CD statistics of difference in means for first 100,000 draws and last 500,000 draws (after burn-in). Newey West variance of mean calculated with 20,000 autocovariances. Under the null of convergence statistics are standard normally distributed.

A.5 Bandwidth Selection

To select a bandwidth for the KDE within the RWMH algorithm I propose to minimize

$$\int \left(\int E_{P_{t,i}}(L(min(R,1), min(\hat{R},1)))q(\theta^*|\theta)d\theta^* \right) p(\theta|P_{1:T})d\theta$$

with respect to the bandwidth choice. q(θ*|θ) is the normal distribution to generate proposal draws for the RWMH algorithm and L(x) is some loss function.

The intuition is to minimize an expected loss function between the actual and estimated acceptance rate averaged over the proposed draws averaged over the posterior space so that the average acceptance-rejection decisions of the algorithm, once it has converged, are as close as possible to the ones with the actual likelihood.

With a sample from the posterior and known likelihood we can search over a grid of bandwidths and for each bandwidth take M draws from the posterior, propose draws θ* and calculate the loss between the actual and KDE based estimated acceptance rate. If M is much smaller than the desired chain length of the full algorithm, the computation time for this selection relative to the computation time of running the whole chain is negligible. As min(R,1) has bounded range, absolute error loss seems to be a reasonable loss function.

In practice one will - at least if this algorithm is actually necessary - not know the likelihood. Approximate sampling from the posterior can still be done using for example the Silverman rule of thumb. Instead of the true R one can then use a pseudo-true R based on some estimate of the likelihood. This estimate can be much more precise than the actual estimate used in the full algorithm as it only needs to be computed M times.[9] One may for example use a KDE estimate of the likelihood with Silverman's rule with a much larger simulation size which I test here. Another alternative, which could be explored in further research, could be to use an unbiased estimator of the likelihood for example with the help of a particle filter (Andrieu et al. 2010).

[9] One can use the same seed for the samples from the posterior for this estimate as for every bandwidth that is searched over a grid so this estimate only has to be computed M times and not M times for each bandwidth that is tested.

Compactly the criterion is

$$\min_C \frac{1}{M}\sum_{s=1}^{M} |min(\frac{\prod_{t=1}^{T}\hat{p}^*(P_t|P_{0:t-1},\theta^*,h_{(s)})}{\prod_{t=1}^{T}\hat{p}^*(P_t|P_{0:t-1},\theta_{(s-1)},h_{(s)})},1) - min(\frac{\prod_{t=1}^{T}\hat{p}(P_t|P_{0:t-1},\theta^*,h_{(s)})}{\prod_{t=1}^{T}\hat{p}(P_t|P_{0:t-1},\theta_{(s-1)},h_{(s)})},1)|$$

$$\hat{p}^*(P_t|P_{0:t-1},\theta^*,h_{(s)}) = \frac{1}{bw_{SM} * N^*}\sum_{i=1}^{N^*} K(\frac{P_{t,i}-P_t}{bw_{SM}})$$

$$\hat{p}(P_t|P_{0:t-1},\theta^*,h_{(s)}) = \frac{1}{bw * N}\sum_{i=1}^{N} K(\frac{P_{t,i}-P_t}{bw})$$

$$bw_{SM} = 1.059\hat{\sigma}N^{*-0.2}$$

$$bw = C * 1.059\hat{\sigma}N^{-0.2}$$

with N* much larger than N, $\theta^* \sim N(\theta_{(s-1)}, \Sigma)$ and $\theta_{(s)}$ a random draw from an approximate posterior sample (for example generated with Silverman's rule).

Figure A.3 shows this approach based on samples from a Silverman rule bandwidth using the true likelihood, and pseudo-true likelihoods from KDEs with N*=25,000, 50,000 and 500,000 simulations and Silverman's rule. M is set to just 100. Bandwidths were tested over a grid of C in steps of 0.25. We see that this algorithm successfully selects the efficient bandwidth in Table 3 corresponding to C=2 for a sufficient simulation size to compute the pseudo-true likelihood. But even for the smallest simulation size considered here it picks a fairly efficient bandwidth. This was robust for different seeds and larger choices of M. Also considering mean square loss didn't change anything. Given the sample from the posterior, the bandwidth selection for N*=50,000 with a grid of 20 different bandwidth values, although most computation time is taken by the estimation of the pseudo-true likelihood, was roughly 15 minutes which is of minor order relative to the 6-7 hours it took to run a whole chain of 100,000 draws.

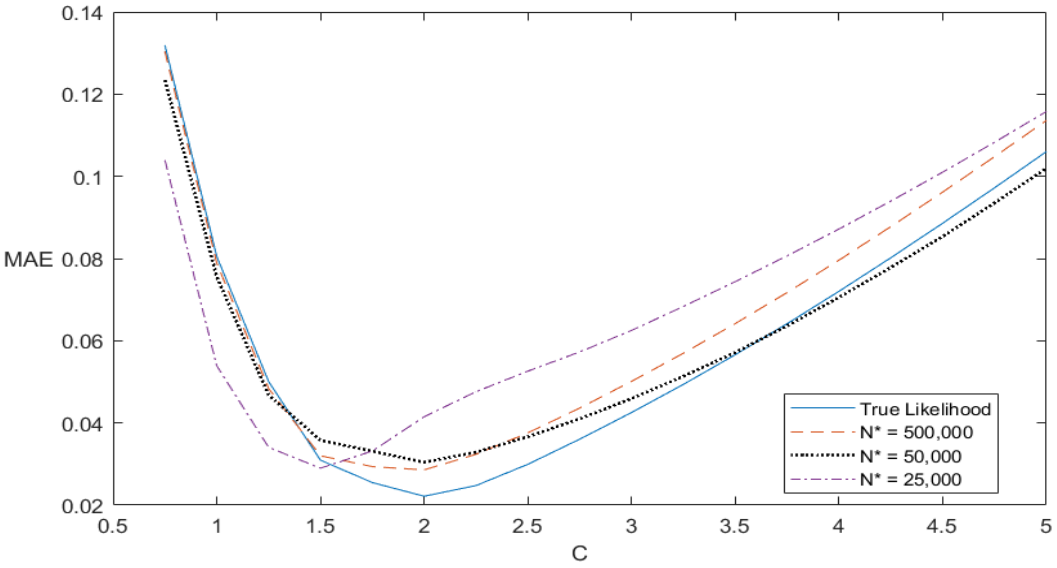

Figure A.3 Bandwidth Selection

Figure shows estimated mean absolute error (MAE) between estimated (with N=1000) and actual/pseudo-true acceptance rate (with true likelihood and Silverman bandwidth KDEs with N*=25,000, 50,000 and 500,000 respectively) averaged over M=100 proposal draws centered around random draws from posterior draws that are generated with Silverman bandwidth and N=1000. Bandwidths are considered according to equation (6) over a grid of C in steps of 0.25.