A Sequential Monte Carlo Approach to Endogenous Time Varying Parameter Models

Evan Hart

March 8, 2024

Abstract

I propose a method to evaluate the likelihood of a nonlinear model with time-varying parameters and endogenous variables. Existing techniques to estimate time-varying parameter models with endogenous variables are restricted to conditionally linear models. The proposed approach modifies a Sequential Monte Carlo filter to evaluate the likelihood of a nonlinear process with an endogenous variable. The modified filter augments the typical measurement and state equations with an equation incorporating instrumental variables. I evaluate the performance of a Bayesian estimator based on the likelihood calculation using simulations and find that the approach generates accurate estimates of both parameters and the unobserved time-varying parameter.

1 Introduction

Complex models are becoming more prevalent and essential to understanding macroeconomic dynamics. Many of these models allow for nonlinearities and shifts in parameters. Cogley and Sargent (2005) use evidence about drift and stochastic volatility to infer that monetary policy rules have changed and that the persistence of inflation itself has drifted over time. Fernández-Villaverde et al. (2007) show how the pricing parameters' movements correlate with inflation, casting doubt on the empirical relevance of Calvo models.

Models with time-varying parameters (TVP) are very popular. Cogley and Sargent (2005), Primiceri (2005), and Koop and Korobilis (2013) focused on time-varying structural vector autoregressions. Authors have also introduced this concept into dynamic stochastic general equilibrium models, such as Young (2004) and Fernández-Villaverde et al. (2007). Researchers in the empirical macroeconomic literature have investigated conditionally linear TVP models with endogenous regressors; see Peersman and Pozzi (2004) and Kim (2006). However, these papers rely on a method that substitutes the fitted value from the instrumenting equation for the actual value, creating an issue of heteroscedasticity. Kim and Kim (2011) address this issue using a Kalman filter for linear models. This method has not been extended to nonlinear models.

A method to estimate a TVP model with nonlinearities, where the time-varying parameter is on an endogenous variable and instrumental variables are available, is not previously defined in the literature. I fill this gap in the literature by adapting the Sequential Monte Carlo filter to run a joint estimation procedure. The Sequential Monte Carlo filter, also called a particle filter, finds the likelihood of the data conditioned on a given set of parameters. It does this by proposing a set of states for the time-varying parameter. The probability of each state is calculated and used to sample from the initial proposal. I adjust the likelihood to a multivariate normal distribution in which the possible errors are paired with the instrumental error. We then iterate this procedure for each period. The likelihood of the data is the product across states of the average likelihood over the accepted particles for each state.

I evaluate the proposed approach using Monte Carlo simulations, where I fix static model parameters at their true values and use the proposed filter to estimate the time-varying parameter. I first demonstrate that the un-modified sequential Monte Carlo filter produces a biased estimate of the time-varying parameter when there is endogeneity. I then show that the modified sequential Monte Carlo filter produces estimates of the time-varying parameter that are approximately unbiased.

The rest of the paper has the following structure. First, I develop a TVP model with endogenous regressors. Then, I explain the traditional method that can be used to estimate such a model but fails to account for the endogeneity. I go on to explain a joint estimation modification to the existing particle filter. Section 5 presents Monte Carlo experiments, in which I compare the performance of the proposed joint method to the traditional method.

2 State Space Model

The state space model provides the foundation for time series data analysis. It is widely used for forecasting, structural, and unstructured modeling. First I lay out a standard system, with normal errors. Then I introduce the system of interest, in which a regressor is endogenous. A particle filter can be adapted to estimate such a system.

The particle filter estimates the likelihood of a data set given a state space model and a set of parameters to define that model. The particle filter does this with a number of assumptions. The time path is conditional on given parameters, the distribution assumptions of the stochastic terms, and the parameters to define those distributions. We must use a filter for estimation when an unobserved state variable follows a time series process. When the measurement or the transition equation is nonlinear, the particle filter can be used instead of the Kalman filter. Consider the following generic state space model:

$$y_t = \lambda_t x_t + \nu_t, \tag{1}$$

$$\lambda_t = \lambda_{t-1} + \omega_t, \qquad \omega \backsim N(0, \sigma_\omega^2). \tag{2}$$

The measurement equation (1) is defined by a linear or nonlinear function: h(), a set of parameters: θ, state variable λ_t, and a stochastic error term: ν_t. Exogenous variables are organized in matrix: X_t. The transition equation (2) gives structure to the process that governs the evolution of the time-varying parameter. The procedure for the traditional particle filter will be discussed in terms of normally distributed and independent identical draws of ν and ω for each period; the procedure can be extended to other distributions.

To incorporate endogeneity into the model, I assume that the joint distribution between V_t and the disturbance to the equation linking an endogenous variable $x_{1,t} \in X_t$ to the instrument z_t is known. Specifically, I make a joint normality assumption:

$$x_t = mz_t + \xi_t \tag{3}$$

$$\nu, \xi \backsim N(0, \Sigma), \qquad \Sigma = \begin{bmatrix} \sigma_\nu^2 & \rho * \sigma_\nu * \sigma_\xi \\ \rho * \sigma_\nu * \sigma_\xi & \sigma_\xi^2 \end{bmatrix} \tag{4}$$

Equation (**??**) links the endogenous variable to the instrumental variable. (4) identifies the endogeneity in the system. I assume joint normality. This is a crucial assumption. However, it is a common assumption made in the literature. The control function literature approach uses a similar assumption about the error terms. An example of the control function literature: Heckman (1979) makes a similar assumption for a linear model with an endogeneity issue.

We can gain some intuition from this simple model in the following way. Suppose the covariance factor ρ is positive. As ξ gets larger, so does the conditional mean of ν. Therefore, as ξ gets larger, positive nonzero values of ν should have a higher likelihood. This means that states with error of zero will be selected less during the second sample period. A univariate distribution fixes the mean at zero and fails to adjust to the larger expected value of ν, biasing the estimate of λ_t upward.

3 Review of the Particle Filter

The particle filter can evaluate the likelihood of data given a state space model, a set of parameters to define that model, and exogenous independent variables. The state space model is represented by equations: (1), (2), and (3). The particle filter finds the likelihood by estimating a set of states for each period, then taking the average likelihood of the sample for each period and taking the product between periods. The particle filter finds a sample for each period's time-varying parameter through a double-sampling process. The first sample is a naive sample based on the parameters of the stochastic process governing the time series process. The second sample is an informed sample that uses the relative probability of the states to resample from the naive sample states.

The first sample of states is found using the transition equation, this sample is denoted: $\lambda_0^{t|t-1,i}$. Under the random walk assumption, each state is the sum of the previous value of λ and a random variable ω. This fact is expressed in our notation: $t|t-1$, each period λ_t is conditioned on the previous value of λ_{t-1}. For each period, N values are simulated for ω_t from the model distribution. The first sample of $\omega_0^{t,i}$ is denoted by i from 1 to N, notice that the first sample is not conditioned on the previous value of λ. The first sample of states now has the following mathematical expression:

$$\lambda_0^{t|t-1,i} = \lambda^{t-1|t-2,i} + \omega_0^{t,i}.$$

At time $t = 1$, the previous value: λ^0 is not conditioned on a previous value, and we assume all N paths start at the same value of λ^0.

The second sample is an informed sample that uses the relative probability of the states. The probability for each state is the likelihood of the error created by that state divided by the total likelihood from all the sample errors. The likelihood of an error is the value of the density function at the error: $f(\nu_0^{t|t-1,i})$. We must know the distribution and have a suggestion for the parameters that define the distribution. The error is the difference between each prediction and the observation value. The prediction is found using the measurement equation. From these predictions, we find the initial sample of errors $\nu_0^{t|t-1,i}$:

$$y_t = h(X_t, \theta, \lambda_0^{t|t-1,i}) + \nu_0^{t|t-1,i}.$$

We then resample based on the relative probability of the suggestions:

$$p(\lambda_0^{t|t-1,i}) = \frac{f(\nu_0^{t|t-1,i})}{\sum_{i=1}^{N} f(\nu_0^{t|t-1,i})}$$

This final sample is saved, now denoted $\lambda^{t|t-1,i}$ with corresponding values of $\nu^{t|t-1,i}$, and the process is repeated for period t+1.

Using the error terms from the final sample of the states, we can calculate the likelihood of the data. The per-period likelihood is the average likelihood of the final sample:

$$f(\nu^{t|t-1}|\theta) = \frac{1}{N}\sum_{i=0}^{N} f(\nu^{t|t-1,i}|\theta).$$

Independence of ν across periods allows us to take the product of the per period likelihood to find the likelihood of all the data:

$$L(Y^T|\theta) = \prod_{t=1}^{T} f(\nu^{t|t-1}|\theta).$$

The procedure defined above will provide a likelihood function that asymptotically peaks at the correct value of the model parameters provided that X_t is exogenous. However, if X_t is endogenous, the likelihood function will peak at the wrong values of model parameters. Estimation routines built on this likelihood calculation will then produce biased estimates of model parameters and the time-varying process λ_t. In the next section I present a modified version of the particle filter that corrects for this bias.

4 A Modified Particle Filter Incorporating Instrumental Variables

If x_t is endogenous, an estimate of our time varying parameters using an uncorrected particle filter will be inconsistent. A particle filter using a joint distribution between the instrumental and measurement error will estimate consistent parameters. Endogeneity here is represented by a correlation between the variable and the measurement error. Assuming we have instrumental variables, we can estimate an instrumental error, ξ_t, that contains the part of x_t that is correlated with v_t:

$$x_t = mz_t + \xi_t$$

The particular filter procedure remains the same. However, the likelihood used for the second sample is a multivariate distribution. After taking the initial naive sample and finding the error term ν, we can use a joint distribution between the measurement and instrumental error to find the likelihood of each state. The probability that a state is selected now considers the correlation between ξ_t and $\nu^{t|t-1,i}$:

$$p(\lambda^{t|t-1,i}) = \frac{g(\nu^{t|t-1,i}, \xi_t)}{\sum_{i=1}^{N} g(\nu^{t|t-1,i}, \xi_t)}$$

Using the error terms from the final sample of the states and the instrumental error, we can calculate the per period likelihood as the average likelihood between the instrumental error and the final sample state errors:

$$g(\nu_t, \xi_t|\theta, m) = \frac{1}{N} \sum_{i=0}^{N} g(\nu_{t,i}, \xi_t|\theta, m),$$

Independence still holds across periods, allowing us to take the product of the per-period likelihood to find a likelihood that will give unbiased estimates of model parameters:

$$L(Y^T|\theta, m) = \prod_{t=1}^{T} g(\nu_t, \xi_t|\theta, m).$$

5 Benefits of an Extended Filter

To analyze the benefits of the particle filter, I will compare the output of an extended particle filter to a traditional particle filter. The comparison is made across models and levels of covariance. It is beneficial for the accuracy of the time-varying parameter and the accuracy of the static parameters to include instruments when there is endogeniety. When covariance is zero, the extended particle filter makes slightly less accurate estimates.

Any maximum likelihood method can be used to search over possible parameters; I choose to implement a Bayesian Markov Chain Monte Carlo. In this estimation method, we collect samples of the parameters of interest to specify a posterior distribution of the parameters. The sampler needs specific prior beliefs about those parameters. I choose to implement flat priors for all my variables. Flat priors will provide a clean starting point to see this bias. We have a burn-in period for each iteration of the MH sampler, after which we collect 2000 samples of the parameters. Then, I take the median of these samples. I simulate 20 data sets of T=250 and collect a point estimate (median) at each iteration. In tables, I collect the mean of the estimates and standard deviation between the samples.

To evaluate the bias, I find the deviation between the true value and the filtered estimate of the TVP. Then, plot the average deviation at each period. Finally, I collected static estimates and the MSE of the TVP.

5.1 Model I

The first test will use a nonlinear model with an additive error. This sample model was chosen because of its relation to a sticky information Phillips curve with drift in the rate of attentiveness. Endogeneity is known to be an issue when estimating the Phillips curve, and more sophisticated models will check for variation in the parameters. Each parameter will have a possible range shown in Table 5. The model has the following form:

measurement equation:

$$y_t = \frac{1-\lambda_t}{\lambda_t} x_{1,t} + \beta x_{2,t} + \nu_t,$$

transition equation:

$$\lambda_t = \lambda_{t-1} + \omega_t,$$

instrumental equation:

$$x_{1,t} = m z_t + \xi_t.$$

After gathering a deviation path for each data set, I find the average deviation at each period and plot this deviation in Figure 1. In the time series graph in which the correlation coefficient is zero, we can see that both the particle filter and the extended particle filter vary in deviation from -0.3 to 0.2, with the particle filter producing a slightly closer deviation to 0. The extended particel filter estimates a slightly negative covariance biasing the path downward, as seen in Tabel 1. However, when we introduce endogeneity, the traditional particle filter skews heavily downward with convergence to a deviation of -2. In contrast, the extended particle filter varies slightly around 0. The bias in the TVP creates a bias in the static parameters under a full estimation. The bias will occur because the peak likelihood value will no longer be at the correct parameters. This bias can also be seen in Tabel 2, the traditional particle filter has trouble identifying the static parameter β. Both filters do a good job estimating the initial value of the TVP, and both overestimate the variance of the stochastic part of the random walk that forms the structure of the TVP.

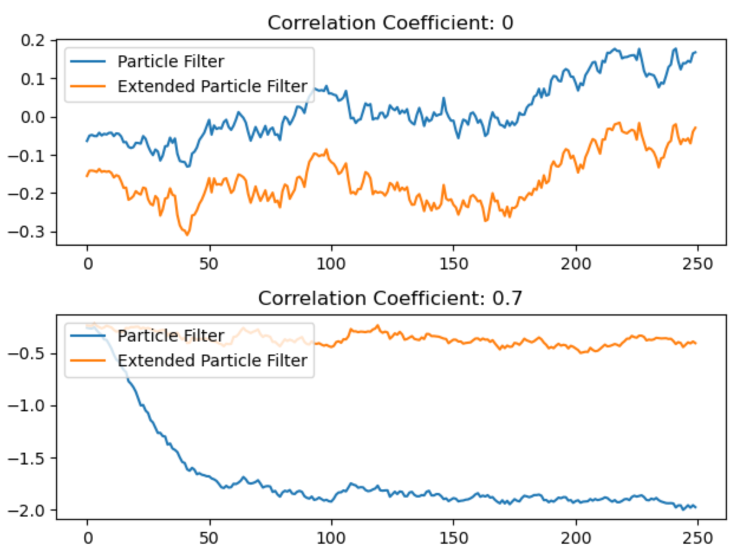

Figure 1: Absolute Deviation of λ_t, Model I.

	Extended Particle Filter	Particle Filter
β =1	0.959 (0.054)	0.976 (0.045)
$\sigma_\nu = 1$	0.957 (0.049)	0.963 (0.043)
λ_0 =4	3.845 (0.384)	3.934 (0.239)
σ_ω =0.1	0.167 (0.025)	0.151 (0.021)
σ_ξ =1	1.005 (0.049)	-
$\rho = 0$	-0.066 (0.074)	-
$m = 1$	0.981 (0.049)	-
Acceptance Rate	0.400	0.578

Table 1: Metropolis-Hastings Sampler Results for Model 1

	Extended Particle Filter	Particle Filter
$\beta = 1$	0.971 (0.041)	0.841 (0.036)
$\sigma_\nu = 1$	0.897 (0.054)	0.784 (0.036)
$\lambda_0 = 4$	3.735 (0.311)	3.736 (0.125)
$\sigma_\omega = 0.1$	0.181 (0.018)	0.167 (0.014)
$\sigma_\xi = 1$	1.000 (0.048)	-
$\rho = 0.7$	0.685 (0.041)	-
$m = 1$	1.055 (0.048)	-
Acceptance Rate	0.344	0.310

Table 2: Metropolis-Hastings Sampler Results for Model 1

5.2 Model II

The first test will use a linear model with a multiplicative error. This sample model was chosen to demonstrate the flexibility of the partilcke filter. The model has the following form: measurement equation:

$$y_t = \lambda_t x_{1,t} + x_{2,t} * \nu_t,$$

transition equation:

$$\lambda_t = \lambda_{t-1} + \omega_t,$$

instrumental equation:

$$x_{1,t} = mz_t + \xi_t.$$

6 The Role of Instruments

The joint normal assumption provides the structure needed to estimate the models and may allow us to model a relationship between the endogenous variable and the measurement error directly. A useful tool if a researcher doesn't have access to instruments. To test the effect of excluding instruments, I will simulate a model with instruments and compare a particle filter that uses those instruments and one that does not. To compare the effectiveness of the two filters, I will again provide the estimation results of the static parameters and the per period mean squared error of the TVP. I will again use an MH sampler and over 102 independent data sets; priors are in Table 6.

The test will use a linear model with an additive error. This sample model was chosen because of its simplicity. The model has the following form:

measurement equation:

$$y_t = \lambda_t * x_{1,t} + \beta * x_{2,t} + \nu_t,$$

transition equation:

$$x_1 = mz_t + \xi_t,$$

instrumental equation:

$$\lambda_t = \lambda_{t-1} + \omega_t.$$

The exclusion of instruments has had a minimal effect on the parametrization I have chosen here. In Table 3, we can see that the estimation of static parameters is very close. However, the per period mean squared error of the TVP is slightly higher for the model that doesn't implement the instruments, as shown in Figure 2.

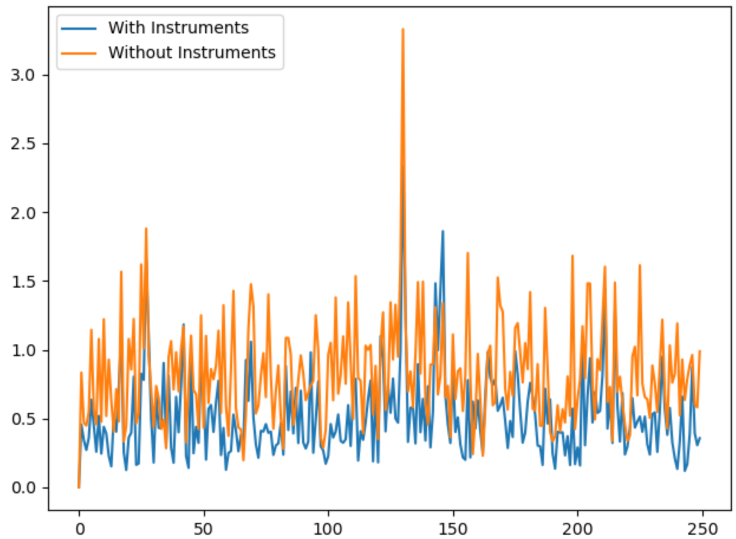

Figure 2: Per-Period Mean Squared Error of λ_t

	With Instruments	Without Instruments
$\sigma_\nu = 1$	0.911 (0.038)	0.981 (0.023)
β =1	0.982 (0.025)	0.993 (0.023)
λ_0 =1	1.009 (0.024)	0.997 (0.023)
σ_ω =1	1.012 (0.026)	1.005 (0.023)
σ_ξ =1	1.005 (0.023)	-
σ_x =1	-	1.041 (0.028)
$\rho_{\xi,\nu} = 0.7$	0.836 (0.052)	-
$\rho_{x,\nu}$	-	0.724 (0.019)
m=1	1.011 (0.023)	-
Acceptance Rate	0.394	0.312

Table 3: Metropolis-Hastings Sampler Results

7 Robustness to Misspecification

The method relies heavily on the joint normal assumption between the variables. To test the robustness of the extended particle filter to misspecification, I will compare particle filters when the true error has a joint t-distribution. One particle filter will use the normal assumption, while the other correctly specifies a t-distribution. I will compare the particle filters using an MH sampler and the following priors using 102 independent data sets. The model again has the following form:

measurement equation:

$$y_t = \lambda_t * x_{1,t} + \beta * x_{2,t} + \nu_t,$$

transition equation:

$$x_1 = mz_t + \xi_t,$$

instrumental equation:

$$\lambda_t = \lambda_{t-1} + \omega_t.$$

The change in the data-generating process of the ν and ξ to a t-distribution with three degrees of freedom doesn't seem to affect the particle filter that uses a normal distribution to analyze the system. The per-period MSE is slightly higher, and the estimates are exactly the same.

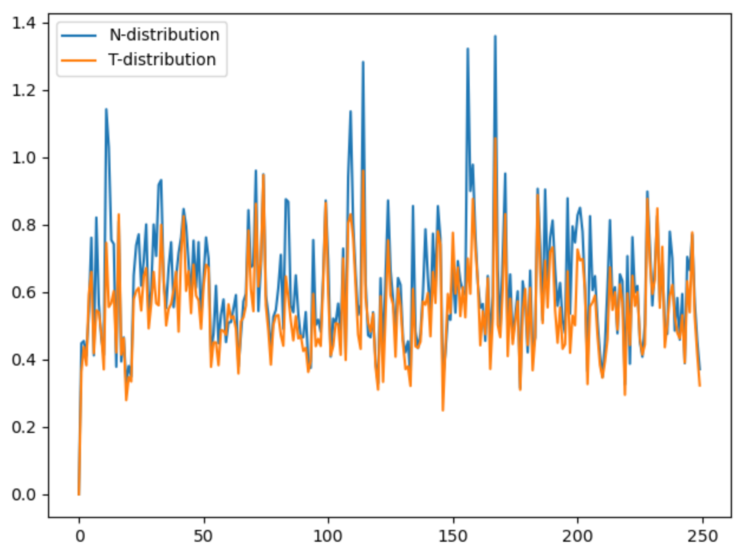

Figure 3: Per-Period Mean Squared Error of λ_t

	N-distribution	T-distribution
$\sigma_\nu = 1$	1.000 (0.022)	1.000 (0.022)
β =1	1.000 (0.022)	1.000 (0.022)
λ_0 =1	1.000 (0.022)	1.000 (0.022)
σ_ω =1	1.000 (0.022)	1.000 (0.022)
σ_ξ =1	1.001 (0.022)	1.000 (0.022)
$\rho = 0.7$	0.701 (0.016)	0.701 (0.016)
m=1	1.000 (0.022)	1.000 (0.022)
Acceptance Rate	0.184	0.379

Table 4: Metropolis-Hastings Sampler Results

8 Conclusion

A method to estimate a TVP model with nonlinearities, where the time-varying parameter is on an endogenous variable and instrumental variables are available, is not previously defined in the literature. I fill this gap in the literature by adapting the particle filter to run a joint estimation procedure. The IV estimation was shown to correct for bias in the TVP and static parameters when there is endogeneity. In addition, I show results demonstrating the flexibility of this method. We can run an estimation procedure without instruments. Finally, I show the method's robustness to misspecification.

References

Cogley, T. and Sargent, T. J. (2005). Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S. *Review of Economic Dynamics*, 8(2):262–302.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Cogley, T., and Schorfheide, F. (2007). How structural are structural parameters? [with comments and discussion]. *NBER Macroeconomics Annual*, 22:83–167.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–61.

Kim, C.-J. (2006). Time-varying parameter models with endogenous regressors. *Economics Letters*, 91(1):21–26.

Kim, Y. and Kim, C.-J. (2011). Dealing with endogeneity in a time-varying parameter model: joint estimation and two-step estimation procedures. *The Econometrics Journal*, 14(3):487–497.

Koop, G. and Korobilis, D. (2013). Large time-varying parameter vars. *Journal of Econometrics*, 177(2):185–198.

Peersman, G. and Pozzi, L. (2004). Determinants of consumption smoothing. Working papers of faculty of economics and business administration, ghent university, belgium, Ghent University, Faculty of Economics and Business Administration.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies*, 72(3):821–852.

Young, A. (2004). Labor's Share Fluctuations, Biased Technical Change, and the Business Cycle. *Review of Economic Dynamics*, 7(4):916–931.

9 Appendix

Hyperparameters

	beta	σ_ν	λ_0	σ_ω	M	σ_ξ	ρ
start	0	0	3	0	0	0	-1
end	2	2	5	0.2	2	2	1

Table 5: Metropolis-Hastings Priors Model 1

	Hyperparameters						
	beta	σ_ν	λ_0	σ_ω	M	σ_ξ	ρ
start	0	0	0	0	0	0	-1
end	2	2	2	2	2	2	1

Table 6: Metropolis-Hastings Priors, Role of Instruments