CHOOSING INFORMATION VARIABLES FOR TRANSITION PROBABILITIES IN A TIME-VARYING TRANSITION PROBABILITY MARKOV SWITCHING MODEL

Andrew J. Filardo

DECEMBER 1998

RWP 98-09

Research Division
Federal Reserve Bank of Kansas City

Andrew J. Filardo is a senior economist at the Federal Reserve Bank of Kansas City. The views expressed in this paper are those of the author and do not necessarily represent those of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

Filardo e-mail: filardo@frbkc.org

Abstract

This paper discusses a practical estimation issue for time-varying transition probability (TVTP) Markov switching models. Time-varying transition probabilities allow researchers to capture important economic behavior that may be missed using constant (or fixed) transition probabilities. Despite its use, Hamilton's (1989) filtering method for estimating fixed transition probability Markov switching models may not apply to TVTP models. This paper provides a set of sufficient conditions to justify the use of Hamilton's method for TVTP models. In general, the information variables that govern time-variation in the transition probabilities must be conditionally uncorrelated with the state of the Markov process.

Keywords*:* Markov switching; time-varying transition probabilities; maximum likelihood estimation

JEL Classification*:* C22, C13

1. Introduction

Regime switching is an important economic and econometric issue. Economists often ask "what if" questions which involve economic implications of regime change. Good examples of this in the macroeconomics literature are changes in fiscal/monetary policies and in exchange rate regimes. Econometrically, analyzing regime switches is particularly difficult because econometricians rarely (directly) observe regime switches, but must infer them from the data. Inference about unobserved regime switching suffers from the curse of dimensionality. The dimensionality, for example, of the state space in a two regime model is 2^T, where T is the length of the data series. To overcome this curse, Goldfeld and Quandt (1973) and Cosslett and Lee (1985) pioneered methods to estimate regime switching models and to draw inferences about unobserved switching. Hamilton (1989) opened up these models to dynamic macroeconomic analysis by developing computational methods to deal with lagged dependent variables. Hamilton's fixed transition probability (FTP) Markov switching model has yielded important macroeconomic evidence of regime switching.

An extension of fixed transition probability Markov switching model to incorporate time-varying transition probabilities has offered another set of useful regime-switching models and has led to many interesting studies because of their intuitive appeal. In a time-varying transition probability (TVTP) Markov switching model, transition probabilities are allowed to vary with such information variables as the strength of the economy, deviations of fundamentals from actual values, and other leading indicators of change. Examples of these extensions show up in many fields. Researchers have used TVTP models to study business cycle fluctuations (Filardo, 1994), interest rate dynamics (Gray, 1996), bubbles and asset pricing (van Norden and Schaller 1997), and exchange rates (Diebold, et.al. 1994, Engel and Hakkio 1994).

A standard empirical approach to estimate the TVTP model is t o use both conditional MLE and filtering methods such as Hamilton's filter for FTP models (also see Kim, 1994 and Gray, 1996). Despite the apparent reasonableness of this approach, the conditions that justify this approach for TVTP models are generally not known. The complication in the TVTP model arises from the presence of additional data, $\boldsymbol{z}$, in the unconditional likelihood function. In general, the presence of the $\boldsymbol{z}$ data implies the need to jointly estimate the parameters of the $\boldsymbol{y}$ and $\boldsymbol{z}$ processes. Thus, knowing sufficient conditions to justify conditional MLE for the TVTP models is valuable because researchers can then ignore estimation of the parameters of the $\boldsymbol{z}$ process which are typically considered nuisance parameters.

This paper outlines sufficient conditions to justify conditional MLE and standard filtering for TVTP models. The choice of information variables, $\boldsymbol{z}$, in the transition probabilities is nontrivial. It requires judicious choices and keen understanding of the economic model, the econometric model, and the statistical properties of the variables in the transition probabilities. This paper shows that conditional exogeneity between the information variables, $\boldsymbol{z}_t$, and the unobserved state variable, $\boldsymbol{S}_t$, validates Hamilton's estimation approach for the TVTP model. This restriction provides some guidance on how to choose information variables for time-varying transition probabilities.

2. Proto-typical TVTP model

A proto-typical TVTP model with state-dependent means, predetermined right-side variables, and normally distributed errors is

$$\begin{aligned} y_t &= m^0 + \Phi(L)(y_{t-1} - m^{S_{t-1}}) + e_t \quad \text{if state } 0 \\ &= m^0 + \Phi(L)(y_{t-1} - m^{S_{t-1}}) + e_t \quad \text{if state } 1 \end{aligned} \qquad \textbf{(1)}$$

where $\Phi(L) = f_1 + f_2 L + \ldots + f_r L^{r-1}$ is the lag polynomial, $m^{S_t} = m_0 + m_1 S_t$ is the state-dependent mean, $e_t \sim N(0, s^2)$, and $S_t \in \{0, 1\}$. (This model can easily be extended to include state-dependent AR coefficients, Φ_{S_t}, state-dependent error processes, e_{s_t}, and other dependent variables, x_t.) The two-point stochastic process on S_t can be summarized by the transition matrix

$$P(S_t = s_t \mid S_{t-1} = s_{t-1}, \boldsymbol{z}_t) = \begin{bmatrix} q(\boldsymbol{z}_t) & 1 - p(\boldsymbol{z}_t) \\ 1 - q(\boldsymbol{z}_t) & p(\boldsymbol{z}_t) \end{bmatrix}, \qquad \textbf{(2)}$$

where the history of the economic indicator variables is $\boldsymbol{z}_t = \{z_t, z_{t-1}, \ldots,\}$.

With AR dynamics of order r, the conditional density f^* is

$$\begin{aligned} & f^*(y_t \mid y_{t-1}, \ldots, y_{-r}, \boldsymbol{z}_t) \\ & = \sum_{s_t=0}^{1} \cdots \sum_{s_{t-r}} f(y_t, S_t = s_t, S_{t-1} = s_{t-1}, \ldots, S_{t-r} = s_{t-r} \mid y_{t-1}, \ldots, y_{-r}, \boldsymbol{z}_t) \\ & = \sum_{s_t=0}^{1} \cdots \sum_{s_{t-r}}^{1} \hat{f}(y_t \mid S_t = s_t, \ldots, S_{t-r} = s_{t-r}, y_{t-1}, \ldots, y_{-r}) \qquad \textbf{(3)} \\ & \qquad \times P(S_t = s_t \mid S_{t-1} = s_{t-1}, \boldsymbol{z}_t) \\ & \qquad \times P(S_{t=1} = s_{t-1}, \ldots, S_{t-r} = s_{t-r} \mid y_{t-1}, \ldots, y_{-r}, \boldsymbol{z}_{t-1}) \end{aligned}$$

and the conditional log-density function is $\ln p(\boldsymbol{y} \mid \boldsymbol{z}, q) = \sum_{t=1}^{T} \ln[f^*(y_t \mid y_{t-1} \ldots y_{-r}, \boldsymbol{z}_t; q)]$.

As is evident in equation (3), the expected conditional density function is formed by conditioning on $\boldsymbol{z}$ and integrating the state of the economy, $\boldsymbol{S}_t$, out of the joint density function, $f(\boldsymbol{y}_T, \boldsymbol{z}_T, \boldsymbol{S}_T \mid q)$. The standard estimation approach in the literature is to recast the expected conditional density function in terms of θ and maximize the conditional likelihood

function $\pi(\theta \mid \boldsymbol{y}, \boldsymbol{z})$, given the observables $\boldsymbol{y}$ and $\boldsymbol{z}$. A key question arises about whether standard estimators from this conditional likelihood have the same statistical properties as those from the unconditional joint likelihood. The answer depends on the distributional relationships between the information variables, $\boldsymbol{z}_t$, the unobserved states, $\boldsymbol{S}_t$, and the observed dependent variable, $\boldsymbol{y}_t$. The next two sections investigate a set of distributional conditions to justify MLE estimation of the TVTP model based on the conditional likelihood function. The distributional conditions are derived from the ability to factor the unconditional likelihood into a concentrated likelihood function.

3. Markov switching and MLE estimation issues

Kiefer verified the desirable properties of MLE estimators for FTP Markov switching models with i.i.d. data. This section reviews his results and discusses how they relate to the estimators of the TVTP model.

Fixed transition probability model. Kiefer (1978) showed that MLE estimators of a i.i.d. FTP Markov switching model (without $\boldsymbol{z}$ variables) are consistent and asymptotically normally distributed. Kiefer's key lemma states: Let $p(\boldsymbol{y} \mid q)$ be the probability density function, $q = (q_1, \ldots, q_k)$ be the vector of unknown parameters, $y_1, \ldots, y_T$ be independent observations on $\boldsymbol{y}$, and the likelihood equations are given by $\frac{\partial \ln p(q)}{\partial q} = 0$, where $\ln p = \sum_{i=1}^{T} \ln f(y_i \mid q)$. Under quite general conditions on the derivatives of the likelihood (differentiability, boundedness, and positive definiteness), there exists a unique and consistent estimator $\hat{q}$ corresponding to a solution of the likelihood equations. Further, $\sqrt{T}(\hat{q} - q)$ is asymptotically normally distributed with mean zero and covariance $I(q)^{-1}$, where $I(q)$ is the Fisher information matrix.

Hamilton (1993) assumes that Kiefer's asymptotic distribution theory for an i.i.d. Markov switching model applies to the model with lagged dependent variables. The presence of the lagged dependent variables complicates the state space. Hamilton proposes a Kalman filter-like procedure to integrate the effect of the lagged states out of the marginal likelihood function. The integration simplifies the likelihood into one similar to the likelihood treated by Kiefer, thereby motivating the use of MLE estimators for Hamilton's FTP model.

TVTP model and factoring the likelihood function. With the introduction of $\boldsymbol{z}$ variables in the joint likelihood, it is not obvious that Hamilton's approach for the FTP model yields valid MLE estimators for the TVTP model. Generally, the likelihood equations from a conditional likelihood are not equivalent to the likelihood equations from an unconditional likelihood. If, however, the likelihood function can be concentrated, then they are equivalent.

A concentrated likelihood function simplifies joint estimation because it obviates the need to jointly estimate all the parameters in the likelihood function. Following Amemiya (1985), a concentrated (log) likelihood function is one that can be written as

$$\begin{aligned} \ln p_T(\hat{q}_1, \hat{q}_2) &= \max_{q \in \Theta} \ln p_{\mathrm{T}}(q) \\ &= \max_{q_1 \in \Theta_1} \ln p_{\mathrm{T}}(q_1) + \max_{q_2 \in \Theta_2} \ln p_{\mathrm{T}}(q_2) \end{aligned} \qquad \textbf{(4)}$$

The benefit of this type of likelihood function is clear. Given the factorization of the likelihood in equation (4), consistent estimates of q_1 can be calculated by solving for the roots of the concentrated likelihood equations, $\frac{\partial \ln p(q_1)'}{\partial q_1} = 0$, where $p(q_1)$ is the concentrated likelihood function of q_1. In other words, regardless of the value of q_2, q_1 is consistently estimated by

maximizing the concentrated likelihood function. The next section investigates sufficient conditions to factor the TVTP joint likelihood function into one piece that is the function of the parameters that determine the y process and another piece that is a function of the parameters that determine the z process. The key to understanding the conditions under which Hamilton's FTP approach can be extended to TVTP models comes from the conditions necessary to factor the joint likelihood into a concentrated likelihood.

4. Application to the proto-typical TVTP model

To simplify the notation, the T-period problem can be written as a two-period problem without loss of generality. For an economy that lasts two periods, the vector $(y_2, S_2, y_1, S_1, z_1)$ completely describes the economy. The random variables (y_2, S_2) and (y_1, S_1, z_1) are assumed to have a joint distribution which is summarized by the probability density function $p(y_2, y_1, z_1, S_1, S_2 \mid q \in \Theta)$. It is assumed that the (y_1, y_2, z_1) vector is observed by the econometrician, and importantly, S_1 and S_2 are not observed.

Given the marginal distributions of S_1 and S_2, the influence of S_1 and S_2 could be integrated out of the density function to yield an expected density function, $p(y_2, y_1, z_1 \mid q) = \sum_{S_2} \sum_{S_1} p(y_2, y_1, z_1, S_1, S_2 \mid q)$. In general, the distribution is intractable because of the curse of dimensionality in unobserved discrete state models. Another way to resolve this problem is to use the conditional marginal distributions of S_1 and S_2 that are of much lower dimensionality. The filters advocated by various researchers have offered different procedures to accomplish this. The issue of this article is to show auxiliary conditions for these filters to generally deliver estimators with desirable MLE properties. To this end, the joint

likelihood is rewritten in a form to show the conditions sufficient to guarantee the favorable properties. To do this, various distributional assumptions and identities need to be first outlined.

Assumptions implied by the proto-typical TVTP model. The proto-typical TVTP model implies the following four restrictions:

(A1) $p(y_1 \mid S_1, z_1)$ is independent of z_1, or $p(y_1 \mid S_1, z_1) = p(y_1 \mid S_1)$

(A2) $p(S_2 \mid S_1, z_1, y_1)$ is independent of y_1, $p(y_1 \mid S_1, z_1)$

(A3) $p(y_2 \mid S_2, S_1, z_1, y_1)$ is independent of (y_1, z_1, S_1), or

$$p(y_2 \mid S_2, S_1, z_1, y_1) = p(y_2 \mid S_2)$$

(A1) says that y_1 is independent of z_1 conditional on the state being known. This is directly verified by looking at equation (1). **(A2)** is implied by the Markovian evolution of the states. **(A3)** tells us that the past states S_1 and the information in *z* and *y* do not help to determine the value of *y* at time *2*. This essentially reduces equation (1) to an i.i.d. model. This assumption simplifies the notation but does not materially affect the results of this section.

Some useful distributional identities. Bayes' rule and integrating the probability density function with the condition distribution on the states yields some useful substitutions:

(S1)
$$\begin{aligned} p(y_2 \mid y_1, z_1) &= \sum\nolimits_{S_2} p(y_2, S_2 \mid y_1, z_1) p(S_2 \mid y_1, z_1) \\ p(S_2 \mid y_1, z_1) &= \sum\nolimits_{S_1} p(S_2 \mid S_1, y_1, z_1)\, p(S_2 \mid y_1, z_1) \end{aligned}$$

(S2)
$$= \sum\nolimits_{S_2} p(y_2, S_2 \mid y_1, z_1) p(S_2 \mid y_1, z_1) \quad \text{by } (\mathbf{A2})$$

(S3)
$$p(S_1 \mid y_1, z_1) = \frac{p(y_1, z_1 \mid S_1) p(S_1)}{p(y_1, z_1)} \qquad \text{by Bayes rule}$$

Deriving the concentrated likelihood function. Under these conditions, the likelihood function can be rewritten to isolate the influence of the z process. The joint likelihood function can be written as

$$\begin{aligned} p(y_2, y_1, z_1) &= p(y_2 \mid y_1, z_1) p(y_1, z_1) \\ &= \sum_{S_2} \sum_{S_1} p(y_2 \mid S_2) p(S_2 \mid S_1, z_1) \frac{p(y_1, z_1 \mid S_1) p(S_1)}{p(y_1, z_1)} p(y_1 \mid z_1) \\ &= \sum_{S_2} \sum_{S_1} [p(y_2 \mid S_2) p(y_1 \mid S_1)][p(S_2 \mid S_1, Z_1) p(S_1)] \frac{p(z_1 \mid y_1, S_1) p(y_1 \mid z_1)}{p(y_1, z_1)} \end{aligned}$$

Since $p(y_1, z_1) = p(y_1 \mid z_1) p(z_1)$ by Bayes rule,

$$p(y_2, y_1, z_1) = \sum_{S_2} \sum_{S_1} [p(y_2 \mid S_2) p(y_1 \mid S_1)][p(S_2 \mid S_1, Z_1) p(S_1)] \frac{p(z_1 \mid y_1, S_1)}{p(z_1)}. \qquad \textbf{(5)}$$

Specifying the dependence on the parameter vector q will help to show how to concentrate the likelihood function. In particular, partition q into two subvectors (q_1, q_2), where q_1 includes the parameters that characterizes the y and S process, and q_2 includes the parameters that characterizes the z process. For example, in the univariate case, $y_t = g_1(q_1)$ and $z_t = g_2(q_2)$. Equation (5) can be rewritten to show the dependence on (q_1, q_2) as

$$p(y_2, y_1, z_1) = \sum_{S_2} \sum_{S_1} [p_{q_1}(y_2 \mid S_2) p_{q_1}(y_1 \mid S_1)][p_{q_1}(S_2 \mid S_1, z_1) p_{q_1}(S_1)] \frac{p_{q_2}(z_1 \mid y_1, S_1)}{p_{q_2}(z_1)}. \qquad \textbf{(6)}$$

This equation shows that estimation of the q_1 parameters generally requires simultaneous estimation with the parameters of the z process, q_2. To see this, notice that the last term in equation (6) is the only term that involves the parameters of the z process. Unless this term can be moved outside the double summation, then the likelihood cannot be concentrated and joint estimation is necessary.

More important, this equation reveals sufficient conditions to concentrate the likelihood function. If the last term in equation (6) is does not vary with S_1 and S_2, then it can be factored out of the summation. Different assumptions about the relationship between z_1 and S_1 will determine whether $\frac{p_{q_2}(z_1 \mid y_1, S_1)}{p_{q_2}(z_1)}$ is equal to 1, a constant k, or will vary with the state.

In general, the last term can be factored out of the integration if the following relationship between z_1 and S_1 holds:

(A4) Given y_1, z_1 is conditionally uncorrelated with S_1; $p_{q_2}(z_1 \mid y_1, S_1) = p_{q_2}(z_1 \mid y_1)$.

Under this assumption, the joint likelihood can be rewritten as

$$p_q(y_2, y_1, z_1) = \frac{p_{q_2}(z_1 \mid y_1)}{p_{q_2}(z_1)} \sum_{S_2} \sum_{S_1} \left[p_{q_1}(y_2 \mid S_2) p_{q_1}(y_1 \mid S_1) \right] \left[p_{q_1}(S_2 \mid S_1, z_1) p_{q_1}(S_1) \right] \qquad \textbf{(7)}$$

with $\frac{p_{q_2}(z_1 \mid y_1)}{p_{q_2}(z_1)}$ being factored out of the summation because it does not depend on either S_1 or S_2.

Assumption (**A4**) clarifies a class of econometric models of z_1 that generate valid maximum likelihood estimators. In models where z_1 is conditionally uncorrelated with (or independent) of S_1, the parameters of the z process of the concentrated out of the joint likelihood and therefore yield MLE estimators with their desirable statistical properties.

5. Extensions and practical applications

Assumption (**A4**) guarantees the factorization of the likelihood function into the concentrated likelihood function of the parameter q_1 and the concentrated likelihood function of the parameter q_2, so that the local maximum likelihood estimator q_1 is consistent and asymptotically normal. Assuming that q_1 is in the interior of Θ_1, p_{q_1} will be uniquely maximized regardless of the value of q_2, given that the roots of its concentrated likelihood equations are zero.

Relaxing (**A3**) to allow lagged dependent variables would not materially change the results about the conditions to concentrate the likelihood function. Assumption (**A4**) is still sufficient condition to concentrate the likelihood function and establish the desirable MLE properties of the estimators. Despite the presence of lagged dependent variables, it is usually assumed that Kiefer's asymptotic distributional results apply to this case (even though as Hamilton (1994) points out there has not been a formal demonstration of this in the literature).

One practical issue in choosing valid information variables for a TVTP model is whether the transition probability information variables are uncorrelated with contemporaneous state. Many studies have used lagged information and assume that condition (**A4**) holds. This is fairly reasonable for many problems as long as z_{t-1} is considered to be predetermined with respect to S_t.

Using lagged information such as z_{t-1}, will not always be valid. This practice will be invalid in cases when the y_t responds to the underlying state of the economy with a lag. In such cases, the use of lagged z will violate condition (**A4**). In general, practitioners must be cognizant of the possibility that the state of the economy may be determined prior to realizations of y and z.

It is of special note that assumption (**A1**) in the proto-typical TVTP model was not used to derive (**A4**). Therefore, Hamilton's MLE approach is valid in the cases where z_1 is not independent of y_1. In other words, z is not restricted to the transition probability function but can be used as right-hand-side regressors in equation (1). In this case, the nonlinear relationship between the z in the output equation and z in the transition probability equation is sufficient to identify the parameters. Hamilton's methods are valid as long as z satisfies condition (**A4**).

Finally, in some applications contemporaneous z_t can be used in the time-varying transition probabilities and use the extended Hamilton filter to estimate the parameters. The main difficulty is verifying that condition (**A4**) holds. In most cases, direct verification is not possible because S is unobserved.

6. Conclusions

The Hamilton's Markov switching model has been a valuable tool in time-series econometrics to examine changes in regimes. This paper investigates the conditions under which Hamilton's filtering methods can be extended to time-varying transition probability Markov switching models. In general, the z_t variables that enter the transition probability functions must be contemporaneously conditionally uncorrelated with the unobserved state, S_t. If this condition is not met in a particular empirical application, other methods need to be employed to deliver

MLE estimators with the typical desirable properties. And, of course, joint estimation of the $\boldsymbol{y}$ and $\boldsymbol{z}$ is always an option.

Acknowledgements

The views expressed herein are those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or of the Federal Reserve System.

References

Amemiya, T. 1985. Advanced Econometrics. Cambridge: Harvard University Press.

Cosslett, S.R., and L.-F. Lee. 1985. Serial correlation in latent discrete variable models. Journal of Econometrics 27, 79-97.

Diebold, F.X., J.-H. Lee, and G.C. Weinbach. 1994. Regime switching with time-varying transition probabilities. In Nonstationary Time Series Analysis and Cointegration, C.P. Hargreaves, ed. Oxford and New York : Oxford University Press, 283-302.

Engel, C., and C.S. Hakkio. 1994. The distribution of exchange rates in the EMS. NBER working paper 4834.

Filardo, A.J. 1994. Business cycle phases and their transitions. Journal of Business and Economic Statistics 12, 299-308.

Filardo, A.J., and S.F. Gordon. 1998. Business cycle durations. Journal of Econometrics 85, 99-123.

Gray, S.F. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27-62.

Hamilton, J.D. 1989. A new approach to the economics analysis of nonstationary time series and the business cycle. Econometrica 57, 357-384.

Hamilton, J.D. Specification testing in Markov-switching time series models. Journal of Econometrics 70, 127-57.

Hamilton, J.D. 1994. State-space models. In R.F. Engle and D.L. McFadden, eds., Handbook of Econometrics, Vol. 4. Amsterdam: Elsevier Science.

Kiefer, N.M. 1978. Discrete parameter variation: Efficient estimation of a switching regression model. Econometrica 46, 427-34.

Kim, C.J. 1994. Dynamic linear models with Markov-switching. Journal of Econometrics 60, 1-22.

Schaller, H., and S. van Norden. 1997. Regime switching in stock market returns. Applied Financial Economics 7, 177-91.