

# A Unified Approach to Nonlinearity, Structural Change and Outliers

Paolo Giordani\*

*School of Economics  
University of New South Wales*

Robert Kohn†

*School of Economics  
University of New South Wales*

Dick van Dijk‡

*Econometric Institute  
Erasmus University Rotterdam*

ECONOMETRIC INSTITUTE REPORT EI 2005-09

March 2005

## Abstract

This paper demonstrates that the class of conditionally linear and Gaussian state-space models offers a general and convenient framework for simultaneously handling nonlinearity, structural change and outliers in time series. Many popular nonlinear time series models, including threshold, smooth transition and Markov-Switching models, can be written in state-space form. It is then straightforward to add components that capture parameter instability and intervention effects. We advocate a Bayesian approach to estimation and inference, using an efficient implementation of Markov Chain Monte Carlo sampling schemes for such linear dynamic mixture models. The general modelling framework and the Bayesian methodology are illustrated by means of several examples. An application to quarterly industrial production growth rates for the G7 countries demonstrates the empirical usefulness of the approach.

**Keywords:** State-space models; Markov-switching models; Threshold models; Bayesian inference; Business cycle asymmetry

**JEL Classification Codes:** C11, C22, E32.

---

\*School of Economics, University of New South Wales, Sydney NSW 2052, ACT 2600 Australia, e-mail: [p.giordani@unsw.edu.au](mailto:p.giordani@unsw.edu.au)

†School of Economics, University of New South Wales, Sydney NSW 2052, ACT 2600 Australia, e-mail: [r.kohn@unsw.edu.au](mailto:r.kohn@unsw.edu.au)

‡Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, e-mail: [djvandijk@few.eur.nl](mailto:djvandijk@few.eur.nl) (corresponding author)

# 1 Introduction

Nonlinearity, structural change, and outliers are prominent features of many economic and financial time series. Consider, for example, the quarterly industrial production growth rates for the G7 countries over the period 1961Q2-2004Q1, shown in panel (a) of Figures 1-7: The extended periods of economic expansions with positive average growth rates are occasionally interrupted by short recessionary periods with negative growth, hinting at the presence of nonlinear regime-switching. For most countries growth also appears to have become more stable during the second half of the sample period, suggesting a permanent structural change in the variability of the series. Finally, the extremely large negative and positive growth rates in 1968Q2-3 for France, 1969Q4-1970Q1 for Italy, and 1974Q1-2 for the UK, among others, are indicative of outliers.

To date, nonlinearity, structural change, and outliers are considered mainly in isolation; that is, most empirical research focuses exclusively on one of these properties, using more or less *ad hoc* procedures for handling the other features. In empirical studies involving nonlinear models, for example, it is not uncommon to encounter statements such as: ‘The time series has been adjusted for outliers prior to estimation’. Such an approach is understandable, given that only few formal attempts have been made to incorporate different features simultaneously in time series models. The exceptions include Lundbergh, Teräsvirta and van Dijk (2003) and Anderson and Low (2005), who develop smooth transition models with structural change in the parameters; Chan and Cheung (1994) and Gabr (1998), who consider outlier-robust GM estimation of threshold and bilinear models, respectively; and Battaglia and Orfei (2005), who propose an iterative procedure for outlier detection and estimation in nonlinear time series models. However, the dangers of focusing on a specific time series feature while virtually ignoring others are evident, as the neglected properties may seriously distort inference concerning the feature of interest. For example, outliers may both hide and spuriously suggest the presence of nonlinearity, see van Dijk, Franses and Lucas (1999) and Koop and Potter (2000). Hence, there is a need for expanding the time series analyst’s toolkit to enable simultaneous treatment of different important features, in particular nonlinearity, structural change and outliers.

The aim of this paper is to show that the class of conditionally linear and Gaussian state-space models offers a general and convenient framework for this purpose. In particular, many popular nonlinear time series models, including threshold, smooth transition and Markov-Switching models, can be written in state-space form. Doing so, it is relatively straightforward to add components that capture structural change in parameters and intervention effects such as additive and innovation outliers. The state-space representations of such models in fact are linear dynamic mixture models, in the sense that they are linear and Gaussian, conditional on a vector of latent random variables that are Markov. Estimation and inference in such models has been greatly facilitated by recent advances in Bayesian statistics. In particular, Gerlach, Carter and Kohn (2000) propose a statistically efficient Markov Chain Monte Carlo sampling scheme and show how to implement this in a computationally efficient manner.

We should note at the outset that we are not the first to show that popular nonlinear time series models can be written in state-space form, see Hamilton (1994) and Kim and Nelson (1999a), among others. However, it has not been recognized previously that using the state-space format allows the inclusion of components for structural change and outliers. Similarly, the advantages of the Bayesian approach for estimation and inference in nonlinear models have been known for quite some time, see Kim and Nelson (1999a), Koop and Potter (1999a), and Bauwens, Lubrano and Richard (1999). Conventional sampling algorithms run into severe problems though when these models are extended to allow for structural change and outliers. In fact, they often even break down completely. The sampling algorithm of Gerlach *et al.* (2000) essentially solves this problem.

The plan of the paper is as follows. Section 2 describes the general state-space framework, and provides three worked examples that demonstrate how this framework can be used for accommodating outliers and various forms of parameter instability in non-linear time series models. Section 3 discusses the efficient Bayesian sampling algorithm for estimating the parameters in these models, with additional details provided in the Appendix. Section 4 illustrates the empirical usefulness of our methodology by estimating a Markov-Switching model for quarterly industrial production growth rates for the G7 countries, allowing for additive and innovation outliers and for instability in the mean and volatility

of growth. Section 5 describes a limited Monte Carlo simulation experiment, which is used to gauge the small sample properties of our approach. Section 6 concludes.

## 2 State-space framework

Consider the general state-space model

$$y_t = g_t + \mathbf{h}'_t \mathbf{x}_t + \gamma_t u_t, \quad (1)$$

$$\mathbf{x}_t = \mathbf{f}_t + \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{\Gamma}_t \mathbf{v}_t, \quad (2)$$

where  $\{y_t\}_{t=1}^n$  is the observed scalar time series variable of interest,  $\mathbf{x}_t$  is the state vector, and the errors  $u_t$  and  $\mathbf{v}_t$  are assumed to be independent and (multivariate) standard normal. The system matrices  $g_t$ ,  $\mathbf{h}_t$ ,  $\gamma_t$ ,  $\mathbf{f}_t$ ,  $\mathbf{F}_t$ , and  $\mathbf{\Gamma}_t$  are determined, up to a set of unknown parameters  $\boldsymbol{\theta}$ , by the value of  $\mathbf{K}_t$ , where  $\mathbf{K} = (\mathbf{K}_1, \dots, \mathbf{K}_n)'$  is an unobserved first-order Markov process such that  $p(\mathbf{K}_t | \mathbf{K}_1, \dots, \mathbf{K}_{t-1}) = p(\mathbf{K}_t | \mathbf{K}_{t-1})$ . Many popular nonlinear time series models can be expressed in the form of (1) and (2). The following examples demonstrate that, within this state-space framework, it is fairly straightforward to generalize these models to accommodate outliers and parameter instability.

*Example 1 - Markov-Switching model with structural change in mean:* Consider the first-order autoregressive model with a Markov-Switching mean, as popularized by Hamilton (1989),

$$y_t = \mu_t + \phi(y_{t-1} - \mu_{t-1}) + \sigma_e e_t, \quad (3)$$

$$\mu_t = \nu + \delta K_{\delta t}, \quad (4)$$

where  $e_t$  is standard normal, and  $K_{\delta t} \in \{0, 1\}$  with transition probabilities  $p_{ij,\delta} = p(K_{\delta t} = j | K_{\delta t-1} = i)$  for  $i, j = 0, 1$ . Hence, the mean of  $y_t$  conditional on  $K_{\delta t}$  (and on all information up to and including  $t-1$ ) is given by  $\mu_t$ , and is equal to  $\nu$  in case  $K_{\delta t} = 0$  and equal to  $\nu + \delta$  in case  $K_{\delta t} = 1$ .

The model (3)-(4) can be expressed in the state-space form (1)-(2) by setting  $\mathbf{x}_t = (y_t - \mu_t, \mu_t)'$ ,  $u_t = 0$ ,  $\mathbf{v}_t = e_t$ , and  $\mathbf{K}_t = K_{\delta t}$ , and defining the system matrices as  $g_t = 0$ ,  $\mathbf{h}_t = (1, 1)'$ ,  $\gamma_t = 0$ ,  $\mathbf{f}_t = (0, \nu + \delta K_{\delta t})'$ ,

$$\mathbf{F}_t = \begin{pmatrix} \phi & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{and} \quad \mathbf{\Gamma}_t = \begin{pmatrix} \sigma_e \\ 0 \end{pmatrix}.$$

Note that in this case only  $\mathbf{f}_t$  depends on the Markov process  $\mathbf{K}_t$ .

The Markov-Switching model has become very popular for modelling business cycle asymmetry in output growth  $y_t$ , where the different states of  $K_{\delta t}$  correspond with periods of high and low growth (and are intended to match expansions and recessions). In such applications the sample period usually covers a long time span, in order to include a reasonable number of business cycles or regime switches, which is necessary to obtain accurate estimates of the transition probabilities  $p_{ij,\delta}$ . However, in that case it is not uncommon to observe structural changes in the ‘base-line’ average growth rate  $\nu$  during the sample period, for example due to changes in productivity, see Dolmas, Raj and Slottje (1999) and Kim and Piger (2002), among others. In the state-space framework we can allow for such structural instability in  $\nu$  in a straightforward, yet flexible manner. In particular, we may replace  $\nu$  in (4) with  $\nu_t$  and model this as  $\nu_t = \nu_{t-1} + \sigma_o K_{o_t} o_t$ , where  $o_t$  is standard normal and  $K_{o_t} \in \{0, 1\}$  with transition probabilities  $p_{ij,o}$  (which may also be set such that  $K_{o_t}$  is independent of  $K_{o,t-1}$ ). Hence,  $\nu_t$  is allowed to change every time period, but is not forced to change at any point in time. Defining  $\mu_t = \nu_t + \delta_t$  with  $\delta_t = \delta K_{\delta t}$ , the extended Markov-Switching model is given by

$$y_t = \nu_t + \delta_t + \phi(y_{t-1} - (\nu_{t-1} + \delta_{t-1})) + \sigma_e e_t, \quad (5)$$

$$\nu_t = \nu_{t-1} + \sigma_o K_{o_t} o_t, \quad (6)$$

$$\delta_t = \delta K_{\delta t}. \quad (7)$$

It is straightforward to rewrite the model (5)-(7) in the state-space form of (1)-(2) by setting  $\mathbf{x}_t = (y_t - \nu_t - \delta_t, \nu_t)'$ ,  $u_t = 0$ ,  $\mathbf{v}_t = (e_t, o_t)$ , and defining the system matrices accordingly.

*Example 2 - Smooth transition models with time-varying parameters:* Consider the first-order smooth transition autoregressive (STAR) model

$$y_t = \phi_{1t} y_{t-1} + \phi_{2t} y_{t-1} \Lambda(s_t; \lambda, c) + \sigma_e e_t, \quad (8)$$

where  $\Lambda(s_t; \gamma, c)$  may be the logistic function

$$\Lambda(s_t; \lambda, c) = \frac{1}{1 + \exp(-\lambda(s_t - c))}, \quad \lambda > 0, \quad (9)$$

and the (observable) transition variable  $s_t$  often is taken to be a lagged dependent variable  $y_{t-d}$  with  $d > 0$ , but can also be an exogenous variable ( $s_t = z_t$ , say) or time ( $s_t = t$ ); see Teräsvirta (1994) and van Dijk, Franses and Teräsvirta (2002). In the standard STAR model the autoregressive parameters are assumed to be constant,  $\phi_{it} = \phi_i$ ,  $i = 1, 2$ , but recently proposed extensions allow for parameter instability. Lundbergh, Teräsvirta and van Dijk (2003) develop the time-varying STAR (TV-STAR) model, in which the AR parameters evolve according to a smooth transition mechanism as  $\phi_{it} = \phi_{i1} + \phi_{i2}\Lambda(t; \lambda, \tau)$ , where  $\Lambda(t; \lambda, \tau)$  is a logistic function as in (9). This effectively boils down to a single (potentially) gradual structural change in  $\phi_i$ ,  $i = 1, 2$ , centered at  $t = \tau$ . Anderson and Low (2005) consider a more flexible form of instability by assuming that the AR parameters behave as random walks, that is

$$\phi_{it} = \phi_{i,t-1} + \sigma_{o_i} o_{it}, \quad i = 1, 2,$$

where  $o_{it}$  is standard normal. Again, it is straightforward to show that this so-called random walk STAR (RW-STAR) model can be written in the state-space form of (1) and (2), this time by defining the state vector as  $\mathbf{x}_t = (\phi_{1t}, \phi_{2t})$ . The model may also be extended by using mixtures. For example, we could have  $\phi_{it} = \phi_{i,t-1} + \sigma_{o_i} K_{o_{it}} o_{it}$  with  $K_{o_{it}} \in \{0, 1\}$ ,  $i = 1, 2$ , such that the AR parameters are not forced to change continuously.

*Example 3 - Threshold model with additive and innovation outliers:* When the parameter  $\lambda$  in (9) is very large, the logistic function  $\Lambda(s_t; \lambda, c)$  effectively becomes an indicator function  $I_{\{s_t > c\}}$ , such that the STAR model (8) reduces to a threshold model, see Tong (1990) for an extensive discussion. The state-space framework is convenient for accommodating innovation and additive outliers in such a model. Consider the threshold model

$$y_t = z_t + \sigma K_{at} a_t, \tag{10}$$

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-1} I_{\{s_t > c\}} + \sigma K_{et} e_t, \tag{11}$$

where  $s_t$  is observable<sup>1</sup>,  $a_t$  and  $e_t$  are standard normal, and  $\sigma$  is a common scale factor for both shocks. Specifying  $K_{at}$  such that  $K_{at} = 0$  with probability  $1 - \pi_a$  and  $K_{at} = g_a > 0$  with probability  $\pi_a$ , the component  $\sigma K_{at} a_t$  in (10) captures additive outliers. Similarly,

---

<sup>1</sup>Note that this excludes “self-exciting” threshold models, where  $s_t = z_{t-d}$  for certain  $d > 0$ .

assuming that  $K_{et} = 1$  with probability  $1 - \pi_e$  and  $K_{et} = g_e > 1$  with probability  $\pi_e$ , innovation outliers are allowed for. This model can be written in the state-space form of (1)-(2) by setting  $\mathbf{x}_t = z_t$ ,  $u_t = a_t$ ,  $\mathbf{v}_t = e_t$ ,  $\mathbf{K}_t = (K_{at}, K_{et})'$ , and

$$\begin{aligned} g_t &= 0, & \mathbf{h}_t &= 1, & \gamma_t &= \sigma K_{at}, \\ \mathbf{f}_t &= 0, & \mathbf{F}_t &= \phi_1 + \phi_2 I_{\{s_t > c\}}, & \text{and } \mathbf{\Gamma}_t &= \sigma K_{et}. \end{aligned}$$

The following section discusses a Bayesian sampling approach that enables efficient inference on the latent variables and the parameters in nonlinear time series models, while allowing for outliers and parameter instability, as in the above examples.

### 3 Sampling methodology

Recent advances in Bayesian statistics have greatly simplified the task of estimating the parameters in conditionally linear and Gaussian state-space models of the form given in (1)-(2), where the system matrices depend on the latent Markov process  $\mathbf{K}$ . In particular, Gerlach *et al.* (2000) develop an efficient sampling algorithm for this class of models. Their contribution is in fact twofold. First, they argue that  $\mathbf{K}$  should be drawn without conditioning on the states  $\mathbf{x}_t$ , which is crucial for dealing with additive outliers and sudden structural breaks. Second, they demonstrate how the number of operations required to obtain draws of  $\mathbf{K}$  can be reduced dramatically, from  $O(n^2)$  to  $O(n)$  where  $n$  is the sample size. In addition, draws of the parameters  $\boldsymbol{\theta}$ , which can be obtained with standard Bayesian techniques in most cases, are less correlated and hence more efficient, because these are drawn conditional on  $\mathbf{K}$ . Therefore, if consecutive draws of  $\mathbf{K}$  are less correlated, draws of the other parameters are likely to be as well. The sampling scheme takes the general form:

1. Draw  $\mathbf{K}$  conditional on the parameter vector  $\boldsymbol{\theta}$  and on the data  $\mathbf{y} = (y_1, \dots, y_n)'$ .
2. Draw the states  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$  conditional on  $\mathbf{K}$ ,  $\mathbf{y}$ , and  $\boldsymbol{\theta}$ .
3. Draw  $\boldsymbol{\theta}$  conditional on  $\mathbf{x}$ ,  $\mathbf{y}$  and  $\mathbf{K}$ .

Steps 1 and 2 are identical for all models. Step 1 uses the algorithm of Gerlach *et al.* (2000), to which we refer for discussion. For completeness, a brief summary of the algorithm is given at the end of this section, with further details provided in the Appendix. Step 2 employs the Gibbs sampling procedure of Carter and Kohn (1994), with the algorithm of Durbin and Koopman (2001) being an interesting alternative. Step 3 is model-dependent. If we draw the parameter vector  $\boldsymbol{\theta}$  conditional on the states  $\mathbf{x}$ , we typically require only standard results in Bayesian inference, at least if conjugate priors are employed. Alternatively, drawing conditional on  $\mathbf{K}$  and  $\mathbf{y}$  with a Metropolis-Hastings step is also straightforward, as the Kalman filter can be used to evaluate the likelihood function  $\mathcal{L}(\mathbf{y}|\boldsymbol{\theta}, \mathbf{K})$ , in which case we do not need step 2 in the algorithm.

The above algorithm can be readily applied for estimating nonlinear time series models with intervention effects and parameter instability. We demonstrate this below by examining the sampling algorithm in detail for two of the examples from the previous section.

*Example 1 (ctd.):* Consider again Hamilton's (1989) Markov-Switching model with instability in mean as given in (5)-(7), but discarding the autoregressive component for simplicity:

$$y_t = \nu_t + \delta_t + \sigma_e e_t, \quad (12)$$

$$\nu_t = \nu_{t-1} + \sigma_o K_{ot} o_t, \quad (13)$$

$$\delta_t = \delta K_{\delta t}. \quad (14)$$

where  $e_t$  and  $o_t$  are standard normal,  $K_{\delta t} \in \{0, 1\}$ , with transition probabilities  $p_{ij,\delta} = p(K_{\delta t} = j | K_{\delta,t-1} = i)$  for  $i, j = 0, 1$ , and  $K_{ot} \in \{0, 1\}$ . It is convenient and parsimonious to assume that  $K_{ot}$  is independent of  $\mathbf{K}_{t-1} = (K_{o,t-1}, K_{\delta,t-1})'$  and of  $K_{\delta t}$ , so that  $p(\mathbf{K}_t | \mathbf{K}_{t-1}) = p(K_{ot}, K_{\delta t} | \mathbf{K}_{t-1}) = p(K_{ot})p(K_{\delta t} | K_{\delta,t-1})$ .

Let  $\pi_o \equiv p(K_{ot} = 1)$  and let  $\boldsymbol{\theta} = (\nu_0, \delta, \sigma_e, \sigma_o, \pi_o, p_{00,\delta}, p_{11,\delta})'$  be the vector of unknown model parameters. We assume a normal, and possibly dispersed prior on  $\nu_0$ . The prior on  $\delta$  is taken to be truncated normal, with support on the negative real line for identification purposes. The priors for  $\sigma_e^2$  and  $\sigma_o^2$  are inverse gamma with parameters  $(S_i, n_i)$ ,  $i = e, o$ . The prior for  $\pi_o$  is conveniently expressed through a beta distribution  $\pi_o \sim \text{Beta}(n_0, n_1)$ .



Since the beta distribution is conditionally conjugate for  $\pi_o$ ,  $n_0$  and  $n_1$  can be interpreted as pre-sample occurrences of  $K_{ot} = j$ ,  $j = 0, 1$ .<sup>2</sup> Hence,  $n_1/(n_0 + n_1)$  is the prior mean of  $\pi_o$ . We suggest to set  $n_0 + n_1$  equal to 100 or larger when trying to capture infrequent interventions. Finally, we use independent beta distributions as priors for  $p_{00,\delta}$  and  $p_{11,\delta}$ . The priors for these transition probabilities can be less informative than that for  $\pi_o$ , as in typical applications we may expect to observe a sizable number of regime-switches.

The complete sampling algorithm for this model is:

1. Draw  $\mathbf{K}$  conditional on  $\boldsymbol{\theta}$  and  $\mathbf{y}$ , as in Gerlach *et al.* (2000).
2. Given  $\mathbf{K}$ , draw  $\pi_o$ ,  $p_{00,\delta}$  and  $p_{11,\delta}$ . The posterior for  $\pi_o$  is beta, while  $p_{00,\delta}$  and  $p_{11,\delta}$  can be drawn efficiently with a Metropolis-Hastings step as explained in Geweke (2005).
3. Given  $\mathbf{K}$  and all parameters, draw  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)'$  as in Carter and Kohn (1994).
4. Let  $y_t^* = y_t - \nu_t$ . Then standard conjugate analysis applies to

$$y_t^* = \delta_t + \sigma_e e_t = \delta K_{\delta t} + \sigma_e e_t,$$

such that it is straightforward to draw  $\sigma_e^2$  and  $\delta$ . A rejection step can be used to enforce the restriction  $\delta < 0$ .

5. Finally, let  $S = \sum_{t=1}^n I(K_t = 1)(\nu_t - \nu_{t-1})^2$  and  $n = \sum_{t=1}^n I(K_t = 1)$ . Then the posterior for  $\sigma_o^2$  is inverse gamma with parameters  $(S + S_o, n + n_o)$ .

*Example 3 (ctd.):* Consider again the threshold model with additive and innovation outliers given in (10) and (11). Defining  $\mathbf{K}_t = (K_{at}, K_{et})'$ , the support of  $\mathbf{K}_t$  is  $\{(0, 1), (0, g_e), (g_a, 1)\}$ , representing a normal state, an innovation outlier and an additive outlier, respectively. Note that we do not allow additive and innovation outliers to occur simultaneously. For convenience, we assume that  $\mathbf{K}_t$  is an independent sequence; that is,  $p(\mathbf{K}_t | \mathbf{K}_{t-1}) = p(\mathbf{K}_t)$ , with  $p(0, 1) = \pi_1$ ,  $p(0, g_e) = \pi_2$ , and  $p(g_a, 1) = \pi_3$ .

---

<sup>2</sup>The interpretability of priors as additional (or pre-sample) observations is a general property of conjugate distributions (see, for example, Gelman et al., 1995, ch. 2)

We now give the priors for the parameter vector  $\boldsymbol{\theta} = (z_0, c, \phi_1, \phi_2, \sigma, g_a, g_e, \pi_1, \pi_2, \pi_3)'$ . The prior on  $z_0$  can be dispersed, such as  $z_0 \sim N(0, \alpha^2)$ , where  $\alpha$  is large relative to  $\sigma$ . We follow Koop and Potter (1999a,b) (and the frequentist literature) in our specification of the prior on the threshold  $c$ , by requiring at least  $w\%$  of observations in each regime and using a uniform distribution on the resulting set of ‘admissible’ values for  $c$ . It is convenient to choose conjugate priors for  $\phi_1$ ,  $\phi_2$ , and  $\sigma$ , such as a normal distribution for  $\phi_i|\sigma$ ,  $i = 1, 2$ , and an inverse-gamma distribution for  $\sigma^2$ . These can also be dispersed if desired. The priors for  $g_a$  and  $g_e$  could be truncated normals or inverse gamma, or any other distribution with positive support. It is important to make these priors quite informative when the sample at hand may have only a few, if any, outliers. In practice, fixing  $g_a$  and  $g_e$  works well when trying to capture large outliers, and we will do so in the empirical application in Section 4.<sup>3</sup> The prior on  $\pi_i$ ,  $i = 1, 2, 3$  is conveniently expressed through a Dirichlet distribution for  $p(\mathbf{K}_t)$ :

$$p(\mathbf{K}_t) \sim D(n_{01}, n_{02}, n_{03}).$$

We suggest to set  $n_0 = \sum_{i=1}^3 n_{0i}$  equal to 100 or larger when trying to capture infrequent interventions.

The complete sampling algorithm for this model is:

1. Given parameters  $\boldsymbol{\theta}$  and the data  $\mathbf{y}$ , draw  $\mathbf{K}$  as in Gerlach *et al.* (2000).
2. Given  $\mathbf{K}$  and  $\boldsymbol{\theta}$ , draw  $\mathbf{z}$  as in Carter and Kohn (1994).
3. Given  $\mathbf{K}$ , draw  $p(\mathbf{K}_t)$ . The posterior distribution is Dirichlet

$$p(\pi_1, \pi_2, \pi_3) \sim D(n_{01} + n_{S1}, n_{02} + n_{S2}, n_{03} + n_{S3}),$$

where  $n_{Sj}$  is the number of in-sample occurrences of state  $j$  of  $\mathbf{K}_t$ ,  $j = 1, 2, 3$ .

4. Given  $\mathbf{z}$ , draw  $c$  as explained below.

---

<sup>3</sup>On the other hand, fixing  $g_a$  is not advisable when trying to capture a steady stream of small measurement errors.

5. Given  $c$ ,  $\mathbf{z}$  is linear and Gaussian, and hence the distribution of  $(\sigma^2, \phi_1, \phi_2)$  conditional on  $c$  and  $\mathbf{z}$  has a standard conjugate form given our choice of priors; see Bauwens *et al.* (1999) and Koop and Potter (1999a,b).

The conditional distribution of  $c$  given  $\mathbf{z}$  in step (4) is non-standard, but  $c$  can be generated using a Metropolis-Hastings step. Alternatively, its conditional distribution can be quickly tabulated at each iteration, as described in Bauwens *et al.* (1999). As long as the sample size  $n$  is not too large, the latter alternative is computationally attractive, because a Metropolis-Hastings step for generating  $c$  requires a careful design as the conditional distribution of  $c$  is often multimodal.

### 3.1 Discussion

Bayesian estimation of nonlinear time series models is relatively straightforward in the absence of outliers and parameter shifts, see Koop and Potter (1999a,b), Bauwens *et al.* (1999), and Kim and Nelson (1999a) for threshold, smooth transition and Markov switching models, respectively. Kim and Nelson (1998) estimate a regime-switching model with non-observable and time-varying states. Their model is also expressed in the conditionally linear and Gaussian state-space form of (1) and (2), with all system matrices potentially depending on a latent Markov vector  $\mathbf{K}_t$ . The first difference between their work and ours is that we improve on the sampling of  $\mathbf{K}$  by using the results in Gerlach *et al.* (2000). The second difference is that our approach to modeling structural change is different and that we place considerable emphasis on outliers, which Kim and Nelson (1998) do not consider. These two differences are closely related, with the second stemming from the first. Our algorithm for drawing  $\{\mathbf{K}_t\}_{t=1}^n$  not only increases sampling efficiency, but allows us to estimate models that cannot be analyzed with Kim and Nelson's (1998) approach. More specifically, while their sampler is less efficient than ours for any regime-switching model with non-constant states, it just cannot handle additive outliers and certain forms of sudden parameter instability.

The sampling scheme of Kim and Nelson (1998) was a major contribution to the study of state-space models with Markov switching. However, it is now possible to estimate a much larger class of conditionally linear and Gaussian state-space models using the

methodology of Gerlach *et al.* (2000). Gerlach *et al.* (2000) draw  $\mathbf{K}$  without conditioning on the states, which is statistically more efficient than drawing  $\{\mathbf{K}_t\}_{t=1}^n$  conditionally on  $\{\mathbf{x}_t\}_{t=1}^n$ , as in Carter and Kohn (1994) (which Kim and Nelson (1998) follow). The difference in efficiency can be appreciable when  $\mathbf{K}_t$  and  $\mathbf{x}_t$  are dependent. In particular, if the dependence between  $\mathbf{K}_t$  and  $\mathbf{x}_t$  is sufficiently high, the sampler of Carter and Kohn (1994) breaks down completely. This situation is particularly relevant for additive outliers, since these are typically either zero or small (measurement error) but occasionally large, and for infrequent parameter shifts.

We now describe the sampler of Gerlach *et al.* (2000). For any variable  $z_t$ , let  $\mathbf{z}^{t,T} = (z_t, z_{t+1}, \dots, z_T)'$  and  $\mathbf{z} = (z_1, z_2, \dots, z_n)'$ , where  $n$  is the sample size. Gerlach *et al.* (2000) draw  $\mathbf{K}$  from

$$p(\mathbf{K}_t | \mathbf{y}, \mathbf{K}_{s \neq t}) \propto p(\mathbf{y} | \mathbf{K}) p(\mathbf{K}_t | \mathbf{K}_{s \neq t}) \propto p(\mathbf{y}^{t,n} | \mathbf{y}^{1,t-1}, \mathbf{K}) p(\mathbf{K}_t | \mathbf{K}_{s \neq t}).$$

For a given proposed value of  $\mathbf{K}_t$ ,  $p(\mathbf{K}_t | \mathbf{K}_{s \neq t})$  is easily evaluated from the transition probabilities, and  $p(\mathbf{y}^{t,n} | \mathbf{y}^{1,t-1}, \mathbf{K})$  can be computed with the Kalman filter in conditionally Gaussian models. Evaluating  $p(\mathbf{y}^{t,n} | \mathbf{y}^{1,t-1}, \mathbf{K})$  through the Kalman filter is straightforward, but requires  $O(n)$  operations, implying  $O(n^2)$  operations to draw  $\mathbf{K}$ . Gerlach *et al.* (2000)'s second contribution is to provide an algorithm to evaluate  $p(\mathbf{y}^{t,n} | \mathbf{y}^{1,t-1}, \mathbf{K})$  in one step and thus to draw  $\mathbf{K}$  in  $O(n)$  operations. Since  $\mathbf{K}_t$  can take a finite number of values, it can be drawn by computing  $p(\mathbf{K}_t | \mathbf{y}, \mathbf{K}_{s \neq t})$  for all possible values of  $\mathbf{K}_t$  and then normalizing. A more detailed description of the sampling algorithm is provided in the Appendix.

## 4 G7 industrial production growth

Output growth has been by far the most popular macro-economic application of regime-switching time series models. In particular, following Hamilton (1989) many attempts have been made to describe the apparent asymmetric behavior over the business cycle in US and international output by means of Markov-Switching models, see Clements and Krolzig (2002), Mills and Wang (2002) and Kim, Morley, and Piger (2005) for recent contributions,

and Hamilton and Raj (2002) for a survey.<sup>4</sup> Structural change in the properties of output also has been heavily investigated. While initially breaks in the mean of output growth were emphasized (see Perron, 1989, for example), more recently the focus has shifted to changes in output volatility (Kim and Nelson, 1999b; McConnell and Perez-Quiros, 2000; Blanchard and Simon, 2001; Stock and Watson, 2003,2005; among others). The occurrence of outliers in output has received considerably less attention, although Balke and Fomby (1994) and van Dijk *et al.* (1999) document the presence of aberrant observations in GDP and industrial production, respectively.

The state-space framework considered in this paper allows us to investigate all of the above-mentioned features simultaneously. Hence, in this section we apply a two-regime Markov-Switching model, extended to allow for structural changes in the mean and in the variance and for additive and innovation outliers, to quarterly growth rates of industrial production (IP) series for the G7 countries over the period 1961Q1-2004Q1.<sup>5</sup> In particular, we consider the model

$$y_t = z_t + \nu_t + \sigma_t K_{\delta t} \delta_t + \sigma_t K_{at} a_t, \quad (15)$$

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + \sigma_t K_{et} e_t, \quad (16)$$

$$\nu_t = \nu_{t-1} + \sigma_t K_{ot} o_t, \quad (17)$$

$$\sigma_t = \sigma_1 I_{\{t \leq \tau\}} + \sigma_2 I_{\{t > \tau\}}, \quad (18)$$

$$\delta_t = \delta_1 I_{\{t \leq \tau\}} + \delta_2 I_{\{t > \tau\}}, \quad (19)$$

where  $y_t$  is the quarterly IP growth rate in annualized percentage points, and  $a_t$ ,  $e_t$  and  $o_t$  are standard normal. Switching between the high- and low-growth states is determined by  $K_{\delta t} \in \{0, 1\}$ . For identification purposes we assume  $\delta_t < 0$  for all  $t$ , such that  $K_{\delta t} = 0$  (1) corresponds with the high-growth (low-growth) state. In (15) the parameter  $\delta_t$  is

---

<sup>4</sup>This approach is not without criticism. Hess and Iwata (1997) and Harding and Pagan (2002), for example, question the relevance of nonlinear models for describing business cycle properties (such as length and depth of recessions) of (mostly US) output, arguing that linear AR models capture many of these properties equally well. However, more recently Galvão (2002) and Morley and Piger (2005) demonstrate that nonlinear models do perform better in reproducing such features as the variability of growth rates in different business cycle phases and the strong negative correlation between the severity of a recession and the strength of the subsequent recovery phase.

<sup>5</sup>The time series are taken from the OECD Main Economic Indicators, and are seasonally adjusted.

multiplied by the standard deviation of the innovations  $\sigma_t$  to make it scale-free. The innovation variance changes once, at an unknown point  $\tau$ , while we allow  $\delta_t$  to change as well when  $\sigma_t^2$  does. Permanent shifts in the mean growth rate are captured by the time-varying process  $\nu_t$ . This variation is of the mixture type:  $K_{ot} \in \{0, 0.3\}$ , where the value 0.3 incorporates the idea that period-to-period changes in mean growth are unlikely to be very large. It also reflects the notion that some care is required in specifying this part of the model. In particular, if we allowed for large and frequent permanent changes in the mean growth rate, it would become more difficult to distinguish between such breaks in  $\nu_t$  and the Markov regime switches in  $\delta_t$ , particularly when regimes are very persistent. This also bears consequences for the prior distribution on  $p(K_{ot})$  discussed below. Additive outliers are allowed for through  $K_{at} \in \{0, 3, 5\}$ , while innovation outliers are captured by  $K_{et} \in \{1, 3\}$ , with  $K_{et} = 1$  representing a regular innovation. The regime-switching process  $K_{\delta t}$  is first-order Markov with transition probabilities  $p_{ij,\delta} = p(K_{\delta t} = j | K_{\delta,t-1} = i)$ . Outliers and structural breaks are assumed to be independent of the regime-switching process, such that  $p(K_{it} | K_{\delta t}, \sigma_t) = p(K_{it})$  for  $i = a, e$  and  $o$ . For example, the probability that an additive outlier occurs is the same in the high- and low-growth states, and is the same before and after the change in  $\sigma_t$  and  $\delta_t$ . Finally, we allow the lag length  $p$  to be a random variable. This is natural in a Bayesian context and avoids the problem that choosing lag length prior to estimation may lead to an incorrect choice if the sample contains additive outliers, for example, see Ronchetti (1997).

The model (15)-(19) can be written in the state-space form of (1)-(2) by setting  $\mathbf{x}_t = (\mu_t, z_t, \dots, z_{t-p+1})'$ ,  $u_t = a_t$ ,  $\mathbf{v}_t = (o_t, e_t)$ , and defining the system matrices as  $g_t = \sigma_t K_{\delta t} \delta_t$ ,  $\mathbf{h}_t' = (1, 1, 0, \dots, 0)'$ ,  $\gamma_t = \sigma_t K_{at}$ ,  $\mathbf{f}_t = 0$ ,

$$\mathbf{F}_t = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \rho_1 & \rho_2 & \rho_3 & \cdots & \rho_{p-1} & \rho_p \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad \text{and} \quad \mathbf{\Gamma}_t = \begin{pmatrix} \sigma_t K_{o,t} & 0 \\ 0 & \sigma_t K_{e,t} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}.$$

The parameters in the model are  $\boldsymbol{\theta} = (\phi_1, \dots, \phi_p, p, z_0, \dots, z_{-p}, \nu_0, \sigma_1, \sigma_2, \delta_1, \delta_2, \tau)'$ , together with the parameters governing the distribution of  $\mathbf{K}_t = (K_{at}, K_{et}, K_{ot}, K_{\delta t})'$ .

The prior for the lag length  $p$  assumes that no lags are skipped; that is,  $\phi_j \neq 0 \Rightarrow \phi_i \neq 0 \forall i < j$ . The support of  $p$  is given by  $\{0, 1, 2, 3, 4\}$ , with prior probabilities proportional to  $\{5, 4, 3, 2, 1\}$ . The prior for  $\phi_1, \dots, \phi_p$  given  $p$  and  $\sigma_i^2$  ( $i = 1, 2$ ) is

$$\phi_1, \dots, \phi_p | p, \sigma_i^2 \sim N(0, \frac{\sigma_i^2}{5} V_p^{-1}),$$

where  $V_p$  is the covariance matrix of  $(y_{t-1}, \dots, y_{t-p})'$ . This is a version of Zellner's (1986) g-prior. We assume that  $\sigma_1^2$  and  $\sigma_2^2$  are independent a priori, each with inverse-gamma prior

$$\sigma_i^2 \sim IG(5\hat{\sigma}^2, 5), \quad i = 1, 2,$$

where  $\hat{\sigma}^2$  is the residual variance from an AR(4) model for  $y_t$ . The priors for  $\delta_1$  and  $\delta_2$  are independent, centered around a value commonly found for quarterly output growth rates, and only mildly informative; that is  $\delta_i \sim N(-2.2, 1)$ , for  $i = 1, 2$ . The prior on  $\tau$  is uniform on the central 70% of the sample, such that we have at least 15% of the observations before and after the structural change in  $\sigma_t$  and  $\delta_t$ . Finally, the priors for  $(z_0, \dots, z_{-p})'$  and  $\nu_0$  are normal and independent, centered at zero and over-dispersed.

Let  $\mathbf{K}_t^* = (K_{at}, K_{et}, K_{ot})'$  such that  $\mathbf{K}_t = (\mathbf{K}_t^*, K_{\delta t})$ , and note that  $\mathbf{K}_t^*$  and  $K_{\delta t}$  are independent. The support of  $\mathbf{K}_t^*$  is given by the five triplets  $(0,1,0)$ ,  $(0,3,0)$ ,  $(3,1,0)$ ,  $(5,1,0)$ , and  $(0,1,0.3)$ , assuming that additive and innovation outliers cannot occur simultaneously and that outliers do not occur at times when the mean growth rate changes. The prior probabilities for the five possible states of  $\mathbf{K}_t^*$  are set equal to 0.95, 0.01, 0.03, 0.005, and 0.005, with  $n_{01} + \dots + n_{05} = 500$ . Finally, we assume beta priors for the transition probabilities  $p_{00,\delta}$  and  $p_{11,\delta}$  governing  $K_{\delta t}$ , centered around common estimates for quarterly output growth and only mildly informative; that is,  $p_{00,\delta} \sim \text{Beta}(0.9, 25)$  and  $p_{11,\delta} \sim \text{Beta}(0.85, 25)$ .

The sampling algorithm for the Markov-Switching model in Section 3 is now extended to include the autoregressive component in (16) and the structural change in  $\sigma_t$  and  $\delta_t$  in (18) and (19):

1. Draw  $\mathbf{K}$  conditional on  $\boldsymbol{\theta}$ ,  $p(\mathbf{K}_t)$  and  $\mathbf{y}$  as in Gerlach *et al.* (2000).
2. Let  $y_t^* = y_t - \sigma_t K_{at} a_t$ . Given  $\mathbf{K}$  draw  $\mathbf{y}^*$ ,  $\boldsymbol{\nu}$ , and  $\mathbf{z}$  as in Carter and Kohn (1994).

3. Given  $\mathbf{K}$  draw the probabilities for the five states of  $\mathbf{K}^*$ . The conditional distribution is Dirichlet

$$p(\mathbf{K}_t^*) \sim D(n_{01} + n_{S1}, n_{02} + n_{S2}, n_{03} + n_{S3}, n_{04} + n_{S4}, n_{05} + n_{S5}),$$

where  $n_{Sj}$  are the sample occurrences of state  $j$  for  $\mathbf{K}_t^*$ .

4. Given  $\mathbf{K}$ , draw  $p_{00,\delta}$  and  $p_{11,\delta}$  using a Metropolis-Hastings step as in Geweke (2005).  
 5. Draw  $\sigma_1$ ,  $\sigma_2$ , and  $\tau$  conditional on  $\mathbf{z}$  and  $\phi_1, \dots, \phi_p$  as in Bauwens *et al.* (1999).  
 6. Draw  $\delta_1$  and  $\delta_2$  conditional on  $\mathbf{z}$ ,  $\boldsymbol{\nu}$  and the parameters by applying standard conjugate analysis to

$$z_t^* = y_t^* - \nu_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p} = \sigma_t \delta K_{\delta t} + \sigma_t K_{et} e_t.$$

By conditioning on the variances  $\sigma_1^2$  and  $\sigma_2^2$ ,  $\delta_1$  and  $\delta_2$  are generated separately based on the sub-samples  $\{z_t^*\}_{t=1}^\tau$  and  $\{z_t^*\}_{t=\tau+1}^n$ , respectively. A rejection step enforces  $\delta_j < 0$ ,  $j = 1, 2$ .

7. Update the lag length  $p$  conditional on  $\mathbf{x}$ ,  $\tau$ ,  $\sigma_1^2$ , and  $\sigma_2^2$ . For this purpose, define  $\tilde{z}_t = z_t/\sigma_t$  such that

$$\tilde{\mathbf{z}}|p = \prod_{t=1}^n \frac{1}{\sqrt{2\pi v_t}} \exp\left(-\frac{f_t^2}{2v_t}\right), \quad (20)$$

where

$$f_t = \tilde{z}_t - \phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p}, \quad \text{and} \quad v_t = 1 + \tilde{\mathbf{z}}_t' \frac{1}{5} V_p^{-1} \tilde{\mathbf{z}}_t,$$

with  $\tilde{\mathbf{z}}_t = (\tilde{z}_{t-1}, \dots, \tilde{z}_{t-p})'$ . The result in (20) can be used to tabulate the distribution of  $p$  given  $\tilde{\mathbf{z}}$  and the prior.

8. Update  $\phi_1, \dots, \phi_p$  conditional on  $\sigma_1$ ,  $\sigma_2$  and  $p$ , applying standard conjugate analysis to  $z_t$ .

Table 1 presents the estimation results, in terms of posterior means and standard deviations of the unknown parameters, complemented by the graphs in Figures 1-7. Several interesting conclusions emerge. First, for all countries the model identifies distinct high-growth and low-growth regimes, in the sense that the posterior mean of  $K_{\delta t}$ , as shown in



panel (b) of Figures 1-7, is generally close to either 0 or 1. In fact, the estimates of  $\nu_t$  and  $\delta_t$  are such that the regimes for  $K_{\delta t} = 0$  and 1 correspond with positive and negative mean growth rates, respectively. (We return to the properties of  $\nu_t$  and  $\delta_t$  in more detail below.) We examine how well these regimes of  $K_{\delta t}$  match business cycle expansions and recessions, by computing turning points in IP using the quarterly version of the Bry-Boschan algorithm (BBQ) developed by Harding and Pagan (2002a).<sup>6</sup> The resulting recessions (periods between peaks and troughs) are shown as shaded areas in the graphs. It is seen that these correspond quite closely with the periods when  $K_{\delta t} = 1$  for Canada, Germany, Italy, the UK and the US.<sup>7</sup> The match appears to be less good for France and Japan. The estimates of the transition probabilities imply reasonable phase lengths. For the US, the posterior means of  $p_{00,\delta}$  and  $p_{11,\delta}$  are such that on average the high-growth and low-growth regimes last approximately 16 and 5 quarters, respectively. The lengths of expansions are similar for the other countries (except Italy), while recessions generally are somewhat shorter.

Second, several additive outliers are identified, see panel (e). Most of these aberrant observations can be related to atypical economic events, such as the nationwide strikes in France in 1968Q2, in Italy in 1969Q4, and in the UK in 1972Q1 and 1979Q1, the strike in the metal industry in Germany in 1984Q2, and the coal miners strike in the UK in 1984-5. The outlier in the UK in 1974Q1 can be explained by the three-day working week that was introduced in an attempt to restrict energy use to counter the acute power crisis. Innovation outliers do not seem to be present in the IP series, as the posterior mean of  $K_{\epsilon t}$  is very close to 1 for all  $t$  and for all countries.

---

<sup>6</sup>In the BBQ algorithm turning points are defined as follows. A peak (trough) is said to occur at quarter  $t$  if the level of IP is above (below) the level at  $t - 2$ ,  $t - 1$ ,  $t + 1$  and  $t + 2$ . This is combined with censoring rules to ensure that recession phases (between peaks and troughs) and expansion phases (between troughs and peaks) of the cycle have a minimum duration of two quarters and that complete cycles have a minimum duration of 5 quarters. See Harding and Pagan (2002b) for an interesting discussion of the BBQ algorithm vis-à-vis Markov-Switching models.

<sup>7</sup>For the UK, the model does not “recognize” the short two-quarter recessions identified by the BBQ algorithm as such, except to some extent the ones occurring in 1984 and 1985. The same applies to Germany, although in that case there is a moderate increase in the posterior mean of  $K_{\delta t}$  during short recessions, see Figure 3(b). On the other hand, it should be noted that the BBQ algorithm appears to be sensitive to the presence of outliers. For example, when the algorithm is applied to the posterior mean of the additive outlier corrected series  $y_t - \sigma_t K_{\epsilon t} a_t$ , no recessions are identified in 1962-3 and 1982 for the UK. Similarly, for France the recession in 1967-8 disappears after removing the additive outliers.

Third, there is strong evidence for a reduction in volatility for all countries. The posterior means of  $\sigma_2$  are substantially smaller than the posterior means of  $\sigma_1$ , with declines ranging from just over 20% for Canada and Japan to more than 40% for the other countries, and even 50% for the US. The posterior distributions of the break date  $\tau$ , shown in panel (f) of Figures 1-7, indicate that the volatility break occurred during the first half of the 1980s for all countries except the US. For the US, most posterior probability mass is located in the second half of the 1970s, while the mode occurs in 1981Q1. Note that this is earlier than the volatility break identified by Kim and Nelson (1999b) and McConnell and Perez-Quiros (2000), among others, who date it at 1984Q1. This difference may obviously be due to the additive outliers we identify to have occurred in 1978Q2, 1980Q2 and 1980Q4.

Fourth, for Canada, France, Japan and the UK we find that  $\delta_2$  is smaller than  $\delta_1$  in absolute value. This strengthens the reduction in the depth of the business cycle, defined as the difference between average growth in expansions and recessions and measured by  $|\sigma_t \delta_t|$ , due to the decline in volatility. For Germany and the US,  $|\delta_2| > |\delta_1|$ , but the changes in  $\delta_t$  are not sufficiently large to compensate for the reduction in volatility, such that the business cycle still becomes less deep after the volatility change.

Fifth, for all countries except the US, the ‘baseline’ mean growth rate  $\nu_t$  changes considerably over time, although the patterns of change vary. For France and Germany, we observe a sharp decline during the 1970s. The same holds for Italy and Japan, although the change in mean growth in these countries started already in the 1960s and continued in the 1980s, respectively. For Canada, growth steadily declined over the entire forty year sample period, although an acceleration occurred during the first half of the 1970s. For the UK, the decline in mean growth only started after 1985. Finally, for the US at first sight Figure 7(c) suggests a continuous decline in  $\nu_t$  during the first half of the sample period followed by an increase during the 1990s. However, inspecting the scale of the vertical axis shows that mean growth was in fact almost constant.

Sixth, the posterior of the lag length  $p$  is very close to a degenerate distribution at  $p = 0$  for France, Germany, Italy and the UK and at  $p = 1$  for Japan and the US. For Canada,  $p = 0$  and  $p = 1$  each are drawn about half of the iterations. The posterior

means of  $\phi_1$  for Japan and the US appear to be reasonable, in the sense that they are in the range of values commonly found in AR models for quarterly output growth rates.

## 5 Simulation evidence

Coping with aberrant observations, such as outliers and level shifts, in the context of linear time series models has been relatively well studied. Various approaches for dealing with such interventions have been developed, including iterative detection-and-removal procedures, see Chen and Liu (1993), and robust estimation methods, see Lucas, Franses and van Dijk (2005). Extending these procedures to nonlinear time series models has proved difficult so far.<sup>8</sup> We would like to emphasize that the state-space approach coupled with efficient Bayesian estimation is a major step forward in this respect, in the sense that it can handle quite general forms of outliers and structural change. This section further illustrates this attractive feature of the state-space approach through a simulation study.

The data generating process (DGP) is a simplified version of the model for the quarterly IP growth rates given in (15)-(19) and broadly reflects US IP data characteristics. The sample size is  $n = 200$ . The autoregressive order  $p$  is 1, with  $\phi_1 = 0.5$ . The structural break in volatility occurs at  $\tau = 100$ , with the standard deviation declining from  $\sigma_1 = 0.1$  to  $\sigma_2 = 0.05$ . The difference between mean growth rates in the two regimes of  $K_{\delta t}$  does not change, such that  $\delta_1 = \delta_2 \equiv \delta$ . We set  $\delta$  equal to  $-3$ , with transition probabilities for  $K_{\delta t}$  equal to  $p_{00,\delta} = 0.95$  and  $p_{11,\delta} = 0.85$ . Similarly, we do not consider innovation outliers by setting  $K_{\epsilon t} = 1$  for all  $t = 1, \dots, n$ , and there are no changes in the average growth rate  $\nu_t$  ( $K_{\nu t} = 0$  for all  $t = 1, \dots, n$ ). We consider three cases in terms of contamination with additive outliers:

1. None:  $K_{at}a_t = 0$  for all  $t = 1, \dots, n$ .
2. Medium: Five additive outliers are included, with fixed and identical locations and magnitudes in all samples. We set  $K_{at}a_t = -5$  for  $t = 30$  and  $190$ ,  $K_{at}a_t = 5$  for

---

<sup>8</sup>Recently, Battaglia and Orfei (2005) develop an outlier detection scheme for nonlinear models, but the usefulness of this approach still needs to be established. Chan and Cheung (1994) and Gabr (1998) consider a GM approach for (outlier-)robust estimation of threshold and bilinear models, respectively. However, as shown by Giordani (2005), the applicability of such methods may be limited.

$t = 90, 110$  and  $130$ , and  $K_{at}a_t = 0$  for all other  $t$ .

3. Heavy: 20 additive outliers are included; 15 of these with  $|K_{at}a_t| = 5$  and five with  $|K_{at}a_t| = 7$ . Their locations and signs are randomized, but kept fixed across samples.

We use 200 replications in the simulation experiment. For each replication we estimate two models nested within (15)-(19). The only restrictions imposed in the first model are that innovation outliers are not allowed for ( $K_{et} = 1$  for all  $t = 1, \dots, n$ ) and that  $\delta_t$  does not change ( $\delta_1 = \delta_2 \equiv \delta$ ), as in the DGP. Structural change in  $\nu_t$  as in (17) is included in the model, and in this respect it differs from the DGP. For computational speed, the lag length  $p$  is not treated as unknown but fixed at the true value 1. The support of  $\mathbf{K}_t^* = (K_{at}, K_{ot})'$  is given by  $(0,0)$ ,  $(3,0)$ ,  $(5,0)$ , and  $(0,0.3)$ , with Dirichlet prior, that is  $p(\mathbf{K}_t^*) \sim D(0.95n_0, 0.03n_0, 0.015n_0, 0.005n_0)$  and  $n_0 = 500$ . Note that these prior probabilities over- and underestimate the true probabilities of an additive outlier in the medium and heavy contamination cases, respectively. Obviously, this probability is also overestimated in the no contamination case. The prior probabilities for  $p_{00,\delta}$  and  $p_{11,\delta}$  are taken to be the DGP probabilities, with a mildly informative beta prior,  $p_{00,\delta} \sim \text{Beta}(0.9, 25)$  and  $p_{11,\delta} \sim \text{Beta}(0.85, 25)$ . The prior for  $\delta$  is normal, centered at  $-3$  with a standard deviation equal to 1. The prior on  $\mu_0$  is sparse. The second model does incorporate regime-switches through  $K_{\delta t}$  and a one-time change in volatility  $\sigma_t$ , but does not allow for additive outliers and mean shifts; that is,  $K_{at} = 0$  and  $K_{ot} = 0$  for  $t = 1, \dots, n$ . Prior specifications are identical to those for the first model.

Each model is estimated using 1000 iterations after a short burn-in period of 100. This is more than sufficient to arrive at the invariant distributions, particularly since estimation is initialized from the true parameter values and the correct states.<sup>9</sup> Table 2 summarizes the simulation results, by showing averages of the posterior means of the unknown parameters in the two models across the 200 replications, together with the corresponding root mean square error (RMSE). In addition, we report a statistic that measures the ability of the models to identify the regimes of  $K_{\delta t}$  correctly. Let  $\mu_t =$

---

<sup>9</sup>This should not be a problem since it is applied to both models, and therefore it should not affect the comparison.

$\nu_t + \sigma_t K_{\delta t} \delta_t$ . We compute  $\mu_t^{\text{DGP}}$  using the true values of  $\nu_t$ ,  $\sigma_t$ ,  $\delta$  and  $K_{\delta t}$ , and  $\mu_t^{\text{M}}$  as the average value (across iterations) of  $\mu_t$  in the estimated models. We then obtain the RMSE of  $\mu_t$  as the square root of the average of  $(\mu_t^{\text{DGP}} - \mu_t^{\text{M}})^2$  across observations  $t = 1, \dots, n$  and across the 200 samples. The resulting number (which should be compared with the standard deviation of the innovations) indicates how accurately the model is estimating the local mean of the process, and is therefore informative on how well it captures the regime-switching nonlinearity. Finally, for the first model we report the average fraction of correctly detected outliers (CDO), computed as the average number of iterations for which the model identifies an additive outlier ( $K_{at} \neq 0$ ) at locations where these indeed occur. For the DGP with no contamination, we report the average posterior mean probability of occurrence of an additive outlier.

These results lead to the following three conclusions. First, allowing for additive outliers in the model is not harmful when in fact there is no outlier contamination in the DGP. The average posterior means of all parameters are close to their true values for both models, while the RMSEs are of comparable magnitude. The parameter  $\delta$  seems to be affected most, with the average posterior mean being somewhat below the true value, while the RMSE for this parameter in the model including outlier effects is higher than in the model without. Note that the average posterior mean probability of occurrence of an additive outlier, reported in the last row (CDO), is equal to 0.025, compared with a prior probability of 0.045. Second, in the presence of additive outliers, ignoring these in the model leads to quite severe biases in the parameter estimates. In particular, the autoregressive parameter  $\phi$  is biased towards zero, while the standard deviations of the shocks  $\sigma_1$  and  $\sigma_2$  are inflated. This corresponds with well-known results from robust estimation of linear autoregressive models, see Denby and Martin (1979) and Bustos and Yohai (1986). We also observe that  $\delta$  is biased towards zero, while the transition probabilities  $p_{00,\delta}$  and  $p_{11,\delta}$  are underestimated, suggesting too frequent regime switches. Third, correctly accounting for additive outliers in the model removes most of these problems. The average posterior means for the first model indicate some bias in the parameter estimates, but these are of a much smaller magnitude than in the model that does not account for the outliers. This also leads to more accurate estimates of the local mean of the series, as

shown by the substantial reduction in RMSE for  $\mu_t$ . Finally, we note that, on average, more than 80% of the outliers are correctly identified.

## 6 Conclusion

This paper argues that the class of conditionally linear and Gaussian state-space models is well-suited for treating nonlinearity, structural change and outliers in time series simultaneously. Popular nonlinear time series models, such as threshold, smooth transition and Markov-Switching models, can be written in state-space format. It is then relatively straightforward to augment these models with components that capture structural change and outliers. The examples in Section 2 illustrate the general modelling framework and demonstrate that common types of intervention effects, such as additive and innovation outliers, and parameter instability, such as instantaneous structural breaks and random walk type behavior, are easily accommodated.

We advocate the use of Bayesian techniques for estimation and inference in such linear dynamic mixture models. In particular, the contribution of Gerlach *et al.* (2000) enables dealing with additive outliers and sudden structural breaks in these models in the first place. Furthermore, it provides an efficient implementation of the required Markov Chain Monte Carlo sampling scheme that converges rapidly to the posterior distribution.

The application of our approach to quarterly IP growth rates for the G7 countries demonstrates its empirical usefulness. Using the state-space framework, we simultaneously identify nonlinear regime-switching, structural changes in mean and in volatility, and additive outliers in these series.

# Appendix A The Gerlach *et al.* (2000) sampling algorithm

This Appendix briefly describes the sampling algorithm developed by Gerlach *et al.* (2000), slightly modified as we impose independence between innovations in the transition and measurement equations.

Consider the conditionally Gaussian state-space model

$$y_t = g_t + \mathbf{h}'_t \mathbf{x}_t + \gamma_t u_t \quad (\text{A.1})$$

$$\mathbf{x}_t = \mathbf{f}_t + \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{\Gamma}_t \mathbf{v}_t \quad (\text{A.2})$$

where  $u_t$  and  $\mathbf{v}_t$  are independent and standard normal and the system matrices  $g_t$ ,  $\mathbf{h}_t$ ,  $\gamma_t$ ,  $\mathbf{f}_t$ ,  $\mathbf{F}_t$  and  $\mathbf{\Gamma}_t$  may all depend on the vector Markov process  $\mathbf{K}_t$ .

The sampling scheme of Gerlach *et al.* (2000) generates  $\mathbf{K}_t$  from the density  $p(\mathbf{K}_t | \mathbf{y}, \mathbf{K}_{s \neq t})$  for  $t = 1, \dots, n$  without conditioning on the states  $\mathbf{x}_t$ . The crucial thing to notice is that

$$\begin{aligned} p(\mathbf{K}_t | \mathbf{y}, \mathbf{K}_{s \neq t}) &\propto p(\mathbf{y} | \mathbf{K}) p(\mathbf{K}_t | \mathbf{K}_{s \neq t}) \\ &\propto p(y_t | \mathbf{y}^{1,t-1}, \mathbf{K}^{1,t}) p(\mathbf{y}^{t+1,n} | \mathbf{y}^{1,t}, \mathbf{K}) p(\mathbf{K}_t | \mathbf{K}_{s \neq t}), \end{aligned} \quad (\text{A.3})$$

where for any variable  $z_t$ ,  $\mathbf{z}^{t,T} = (z_t, z_{t+1}, \dots, z_T)'$  and  $\mathbf{z} = (z_1, z_2, \dots, z_n)'$ , where  $n$  is the sample size. For each value of  $\mathbf{K}_t$ , the right side of (A.3) is evaluated as follows. The term  $p(\mathbf{K}_t | \mathbf{K}_{s \neq t})$  is obtained from the prior. The term  $p(y_t | \mathbf{y}^{1,t-1}, \mathbf{K}^{1,t})$  is obtained from  $p(\mathbf{x}_{t-1} | \mathbf{y}^{1,t-1}, \mathbf{K}^{1,t-1})$  using one step of the Kalman filter. Traditional sampling algorithms use  $n - t + 1$  steps of the Kalman filter given the current values of  $\mathbf{K}^{t,n}$  to obtain the term  $p(\mathbf{y}^{t+1,n} | \mathbf{y}^{1,t}, \mathbf{K})$ . Therefore it requires  $O(n)$  operations to generate each  $\mathbf{K}_t$ , and hence  $O(n^2)$  operations to generate  $\mathbf{K}$ . The crucial innovation of Gerlach *et al.* (2000)'s algorithm is that the term  $p(\mathbf{y}^{t+1,n} | \mathbf{y}^{1,t}, \mathbf{K})$  is obtained in *one step* after an initial set of backward recursions. This reduces the number of operations required to generate the complete vector  $\mathbf{K}$  to  $O(n)$ . Before giving the recursion for generating  $\mathbf{K}$ , we state several crucial results that are used in the algorithm. Proofs of these lemmas can be found in Gerlach *et al.* (2000).

**Lemma 1** Let  $N_{t+1} = \text{var}(y_{t+1} | \mathbf{x}_t, \mathbf{K}^{1,t+1})$ . Then

$$N_{t+1} = \mathbf{h}'_{t+1} \mathbf{\Gamma}_{t+1} \mathbf{\Gamma}'_{t+1} \mathbf{h}_{t+1} + \gamma_{t+1}^2$$

and

$$\begin{aligned} E(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{y}_{t+1}, \mathbf{K}) &= \mathbf{a}_{t+1} + \mathbf{A}_{t+1} \mathbf{x}_t + \mathbf{B}_{t+1} \mathbf{y}_{t+1} \\ \text{var}(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{y}_{t+1}, \mathbf{K}) &= \mathbf{C}_{t+1} \mathbf{C}'_{t+1} \end{aligned}$$

where

$$\begin{aligned} \mathbf{a}_{t+1} &= (I - \mathbf{B}_{t+1} \mathbf{h}'_{t+1}) \mathbf{f}_{t+1} - \mathbf{B}_{t+1} g_{t+1} \\ \mathbf{A}_{t+1} &= (I - \mathbf{B}_{t+1} \mathbf{h}'_{t+1}) \mathbf{F}_{t+1} \\ \mathbf{B}_{t+1} &= \mathbf{\Gamma}_{t+1} \mathbf{\Gamma}'_{t+1} \mathbf{h}_{t+1} \mathbf{N}_{t+1}^{-1} \\ \mathbf{C}_{t+1} \mathbf{C}'_{t+1} &= \mathbf{\Gamma}_{t+1} (I - \mathbf{\Gamma}'_{t+1} \mathbf{h}_{t+1} \mathbf{N}_{t+1}^{-1} \mathbf{h}'_{t+1} \mathbf{\Gamma}_{t+1}) \mathbf{\Gamma}'_{t+1}. \end{aligned} \quad (\text{A.4})$$

The expression on the right hand side of (A.4) can be factored as  $\mathbf{C}_{t+1}\mathbf{C}'_{t+1}$  where the matrix  $\mathbf{C}_{t+1}$  is either null or has full column rank. Then

$$\mathbf{x}_{t+1} = \mathbf{a}_{t+1} + \mathbf{A}_{t+1}\mathbf{x}_t + \mathbf{B}_{t+1}\mathbf{y}_{t+1} + \mathbf{C}_{t+1}\boldsymbol{\zeta}_{t+1},$$

where  $\boldsymbol{\zeta}_t \sim N(0, I)$ , independent of  $\mathbf{x}_t$  and  $y_{t+1}$  (conditional on  $\mathbf{K}$ ).

**Lemma 2** For  $t = 1, \dots, n-1$ , the density  $p(\mathbf{y}^{t+1,n}|\mathbf{x}_t, \mathbf{K})$  is independent of  $\mathbf{K}^{1,t}$  and can be expressed as

$$p(\mathbf{y}^{t+1,n}|\mathbf{x}_t, \mathbf{K}) \propto \exp\left\{-\frac{1}{2}(\mathbf{x}_t\boldsymbol{\Omega}_t\mathbf{x}_t - 2\boldsymbol{\mu}'_t\mathbf{x}_t)\right\},$$

where  $\boldsymbol{\Omega}$  and  $\boldsymbol{\mu}$  are computed recursively starting from

$$\boldsymbol{\Omega}_n = 0, \quad \boldsymbol{\mu}_n = 0,$$

and moving backward

$$\begin{aligned} \boldsymbol{\Omega}_t &= \mathbf{A}'_{t+1}(\boldsymbol{\Omega}_{t+1} - \boldsymbol{\Omega}_{t+1}\mathbf{C}_{t+1}\mathbf{D}_{t+1}^{-1}\mathbf{C}'_{t+1}\boldsymbol{\Omega}_{t+1})\mathbf{A}_{t+1} + \mathbf{F}'_{t+1}\mathbf{h}_{t+1}\mathbf{N}_{t+1}^{-1}\mathbf{h}'_{t+1}\mathbf{F}_{t+1}, \\ \boldsymbol{\mu}_t &= \mathbf{A}'_{t+1}(I - \boldsymbol{\Omega}_{t+1}\mathbf{C}_{t+1}\mathbf{D}_{t+1}^{-1}\mathbf{C}'_{t+1})(\boldsymbol{\mu}_{t+1} - \boldsymbol{\Omega}_{t+1}(\mathbf{a}_{t+1} + \mathbf{B}_{t+1}y_{t+1})) \\ &\quad + \mathbf{F}'_{t+1}\mathbf{h}_{t+1}\mathbf{N}_{t+1}^{-1}(y_{t+1} - g_{t+1} - \mathbf{h}'_{t+1}\mathbf{F}_{t+1}) \end{aligned}$$

where

$$\mathbf{D}_{t+1} = \mathbf{C}'_{t+1}\boldsymbol{\Omega}_{t+1}\mathbf{C}_{t+1} + I.$$

**Lemma 3** Let  $\mathbf{m}_t = E(\mathbf{x}_t|\mathbf{y}^{1,t}, \mathbf{K})$ ,  $\mathbf{V}_t = \text{var}(\mathbf{x}_t|\mathbf{y}^{1,t}, \mathbf{K})$  and  $r_t = \text{var}(y_t|\mathbf{y}^{1,t-1}, \mathbf{K})$ . The Kalman filter for the state-space model (A.1) is given by

$$R_t = \mathbf{h}'_t\mathbf{F}_t\mathbf{V}_{t-1}\mathbf{F}'_t\mathbf{h}_t + \mathbf{h}'_t\boldsymbol{\Gamma}_t\boldsymbol{\Gamma}'_t\mathbf{h}_t + \gamma_t^2, \quad (\text{A.5})$$

$$\mathbf{m}_t = (I - \mathbf{J}_t\mathbf{h}'_t)(\mathbf{f}_t + \mathbf{F}_t\mathbf{m}_{t-1}) + \mathbf{J}_t(y_t - g_t)/r_t \quad (\text{A.6})$$

$$\mathbf{V}_t = \mathbf{F}_t\mathbf{V}_{t-1}\mathbf{F}'_t + \boldsymbol{\Gamma}_t\boldsymbol{\Gamma}'_t - \mathbf{J}_t\mathbf{J}'_t/r_t \quad (\text{A.7})$$

where

$$\mathbf{J}_t = \mathbf{F}_t\mathbf{V}_{t-1}\mathbf{F}'_t\mathbf{h}_t + \boldsymbol{\Gamma}_t\boldsymbol{\Gamma}'_t\mathbf{h}_t.$$

The conditional density  $p(y_t|\mathbf{y}^{1,t-1}, \mathbf{K}^{1,t})$  is given by

$$p(y_t|\mathbf{y}^{1,t-1}, \mathbf{K}^{1,t}) \propto r_t^{-1/2} \exp\left(-\frac{1}{2r_t}(y_t - g_t - \mathbf{h}'_t\mathbf{f}_t + \mathbf{h}'_t\mathbf{F}_t\mathbf{m}_{t-1})^2\right). \quad (\text{A.8})$$

We can write  $\mathbf{V}_t = \mathbf{T}_t\mathbf{T}'_t$ , where the matrix  $\mathbf{T}_t$  has full column rank (if  $\mathbf{V}_t \neq 0$ ) or is null (if  $\mathbf{V}_t = 0$ ). Then, conditional on  $\mathbf{K}$ ,

$$\mathbf{x}_t = \mathbf{m}_t + \mathbf{T}_t\boldsymbol{\xi}_t,$$

where  $\boldsymbol{\xi}_t \sim N(0, I)$  and independent of  $\mathbf{y}^{1,t}$ .

**Lemma 4** Using the results of Lemma 3, it follows that

$$\begin{aligned} p(y^{t+1,n}|\mathbf{y}^{1,t}, K) &= \int p(\mathbf{y}^{t+1,n}|\mathbf{x}_t, \mathbf{K}^{t+1,n})p(\boldsymbol{\xi}_t|\mathbf{K}^{1,t})d\boldsymbol{\xi}_t \\ &\propto |\mathbf{T}'_t\boldsymbol{\Omega}_t\mathbf{T}_t|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{m}'_t\boldsymbol{\Omega}_t\mathbf{m}_t - 2\boldsymbol{\mu}'_t\mathbf{m}_t - \boldsymbol{\phi}'_t(\mathbf{T}'_t\boldsymbol{\Omega}_t\mathbf{T}_t + I)^{-1}\boldsymbol{\phi}_t)\right\}, \end{aligned} \quad (\text{A.9})$$

where

$$\boldsymbol{\phi}_t = \mathbf{T}'_t(\boldsymbol{\mu}_t - \boldsymbol{\Omega}_t\mathbf{m}_t).$$



The recursion for generating  $\mathbf{K}$  in  $O(n)$  operations comprises two main steps:

1. Given the current value of  $\mathbf{K}$ , calculate  $\boldsymbol{\Omega}_t$  and  $\boldsymbol{\mu}_t$  for  $t = n - 1, \dots, 1$ , using the recursions in Lemma 2.
2. Given  $E(\mathbf{x}_0)$  and  $\text{var}(\mathbf{x}_0)$ , perform the following for  $t = 1, \dots, n$ :
  - (a) Obtain  $r_t$ ,  $\mathbf{m}_t$  and  $\mathbf{V}_t$  from  $\mathbf{m}_{t-1}$  and  $\mathbf{V}_{t-1}$  as in Lemma 3.
  - (b) Obtain  $p(y_t|\mathbf{y}^{1,t-1}, \mathbf{K}^{1,t})$  as in Lemma 3, and  $p(\mathbf{y}^{t+1,n}|\mathbf{y}^{1,t}, \mathbf{K})$  as in Lemma 4.
  - (c) Obtain  $p(\mathbf{K}_t|\mathbf{y}, \mathbf{K}_{s \neq t})$  for all values of  $\mathbf{K}_t$  by normalization of

$$p(\mathbf{K}_t|\mathbf{y}, \mathbf{K}_{s \neq t}) \propto p(y_t|\mathbf{y}^{1,t-1}, \mathbf{K}^{1,t})p(\mathbf{y}^{t+1,n}|\mathbf{y}^{1,t}, \mathbf{K})p(\mathbf{K}_t|\mathbf{K}_{s \neq t}).$$

Then draw  $\mathbf{K}_t$ .

- (d) Update  $\mathbf{m}_t$  and  $\mathbf{V}_t$  as in Lemma 3, using the generated value of  $\mathbf{K}_t$ .

## References

- Anderson, H.M. and C.N. Low (2005), Random walk smooth transition autoregressive models, in C. Milas, P. Rothman and D. van Dijk (eds.) *Nonlinear Time Series Analysis of Business Cycles*, North-Holland: Elsevier, to appear.
- Balke, N.S. and T.B. Fomby (1994), Large shocks, small shocks, and economic fluctuations: Outliers in macroeconomic time series, *Journal of Applied Econometrics* **9**, 181–200.
- Battaglia, F. and L. Orfei (2005), Outlier detection and estimation in nonlinear time series, *Journal of Time Series Analysis* **26**, 107–121.
- Blanchard, O.J. and J. Simon (2001), The long and large decline in US output volatility, *Brooking Papers on Economic Activity*, 135–174.
- Bauwens, L., M. Lubrano, and J.-F. Richard (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford: Oxford University Press.
- Bustos, O.H. and V.J. Yohai (1986), Robust estimates for ARMA models, *Journal of the American Statistical Association* **81**, 155–168.
- Carter, C. and R. Kohn (1994), On Gibbs sampling for state-space models, *Biometrika* **81**, 541–553.
- Chan, W.S. and S.H. Cheung (1994), On robust estimation of threshold autoregressions, *Journal of Forecasting* **13**, 37–49.
- Chen, C. and L-M. Liu (1993), Joint estimation of model parameters and outlier effects in time series, *Journal of the American Statistical Association* **88**, 284–297.
- Clements, M.P. and H.-M. Krolzig (2002), Can oil shocks explain asymmetries in the US business cycle?, *Empirical Economics* **27**, 185–204.
- Denby, L. and R.D. Martin (1979), Robust estimation of the first-order autoregressive parameter, *Journal of the American Statistical Association* **74**, 140–146.
- Dolmas, J., B. Raj and D. Slottje (1999), The US productivity slowdown: A peak through the structural break window, *Economic Inquiry* **37**, 226–241.
- Durbin, J. and S.J. Koopman (2001), *Time Series Analysis by State Space Methods*, Oxford: Oxford University Press.
- Galvão, A.B. (2002), Can non-linear time series models generate US business cycle asymmetric shape?, *Economics Letters* **77**, 187–194.
- Gelman, A., J.B. Carlin, H.S. Stern, D.B. Rubin (1995), *Bayesian Data Analysis*, London: Chapman & Hall.
- Gerlach, R., C. Carter and R. Kohn (2000), Efficient Bayesian inference for dynamic mixture models, *Journal of the American Statistical Association* **95**, 819–828.
- Geweke, J. (2005), *Contemporary Bayesian Econometrics and Statistics*, New York: John Wiley & Sons, to appear.
- Gabr, M.M. (1998), Robust estimation of bilinear time series models, *Communications in Statistics - Theory and Methods* **27**, 41–53.

- Giordani, P. (2005), A cautionary note on outlier robust estimation of threshold models, *Journal of Forecasting*, to appear.
- Hamilton, J.D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**, 357–384.
- Hamilton, J.D. (1994), State space models, in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics* Vol. **4**, pp. 3039–3080.
- Hamilton, J.D. and B. Raj (2002), New directions in business cycle research and financial analysis, *Empirical Economics* **27**, 149–162.
- Harding, D. and A. Pagan (2002a), Dissecting the cycle: A methodological investigation, *Journal of Monetary Economics* **49**, 365–381.
- Harding, D. and A. Pagan (2002b), A comparison of two business cycle dating methods, *Journal of Economic Dynamics and Control* **27**, 1681–1698 (with comment by J.D. Hamilton and rejoinder).
- Hess, G.D. and S.I. Iwata (1997), Measuring and comparing business-cycle features, *Journal of Business and Economic Statistics* **15**, 432–444.
- Kim, C.-J., and Nelson, C.R. (1998), Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime-switching, *Review of Economics and Economic Statistics* **80**, 188–201.
- Kim, C.-J., and Nelson, C.R. (1999a), *State-Space Models With Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications*, Cambridge: MIT Press.
- Kim, C.-J. and C.R. Nelson (1999b), Has the U.S. economy become more stable? A Bayesian approach based on a Markov-Switching model of the business cycle, *Review of Economics and Statistics* **81**, 608–616.
- Kim, C.-J. and J. Piger (2002), Common stochastic trends, common cycles, and asymmetry in economic fluctuations, *Journal of Monetary Economics* **49**, 1189–1211.
- Kim, C.-J., J. Morley and J. Piger (2005), Nonlinearity and the permanent effects of recessions, *Journal of Applied Econometrics*, to appear.
- Koop, G. and S.M. Potter (1999a), Bayes factors and nonlinearity: evidence from economic time series, *Journal of Econometrics* **88**, 251–281.
- Koop, G. and S.M. Potter (1999b), Dynamic asymmetries in US unemployment, *Journal of Business and Economic Statistics* **17**, 298–312.
- Koop, G. and S.M. Potter (2000), Nonlinearity, structural breaks or outliers in economic time series?, in W.A. Barnett, D.F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim and A. H. Würtz (eds.), *Nonlinear Econometric Modeling in Time Series Analysis*, Cambridge: Cambridge University Press, pp. 61–78.
- Lucas, A., P.H. Franses and D. van Dijk (2005), *Outlier Robust Analysis of Economic Time Series*, Oxford: Oxford University Press, to appear.
- Lundbergh, S., T. Teräsvirta and D. van Dijk (2003), Time-varying smooth transition autoregressive models, *Journal of Business and Economic Statistics* **21**, 104–121.
- McConnell, M.M. and G. Perez-Quiros (2000), Output fluctuations in the United States: What

- has changed since the early 1980s?, *American Economic Review* **90**, 1464–1476.
- Mills, T.C. and P. Wang (2002), Plucking models of business cycle fluctuations: Evidence from the G7 countries, *Empirical Economics* **27**, 255–276.
- Morley, J. and J. Piger (2005), The importance of nonlinearity in reproducing business cycle features, in C. Milas, P. Rothman and D. van Dijk (eds.) *Nonlinear Time Series Analysis of Business Cycles*, North-Holland: Elsevier, to appear.
- Perron, P. (1989), The great crash, the oil price shock, and the unit root hypothesis, *Econometrica* **57**, 1361–1401.
- Ronchetti, E. (1997), Robustness aspects of model choice, *Statistica Sinica* **7**, 327–338.
- Stock, J.H. and M.W. Watson (2003), Has the business cycle changed? Evidence and explanations, in *Monetary Policy and Uncertainty: Adapting to Changing Economy* symposium proceedings, Federal Reserve Bank of Kansas City, pp. 9–56.
- Stock, J.H. and M.W. Watson (2005), Understanding changes in international business cycle dynamics, *Journal of the European Economic Association*, to appear.
- Teräsvirta, T. (1994), Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.
- Tong, H. (1990), *Non-Linear Time Series: A Dynamical Systems Approach*, Oxford: Oxford University Press.
- van Dijk, D., P.H. Franses and A. Lucas (1999), Testing for smooth transition nonlinearity in the presence of outliers, *Journal of Business and Economic Statistics* **17**, 217–235.
- van Dijk, D., T. Teräsvirta and P.H. Franses (2002), Smooth transition autoregressive models - A survey of recent developments, *Econometric Reviews* **21**, 1–47.
- Zellner, A. (1986), On assessing prior distributions and Bayesian regression analysis with g-prior distributions, in P.K. Goel and A. Zellner (eds.) *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*, Amsterdam: North-Holland, pp. 233–243.

Table 1: Empirical results

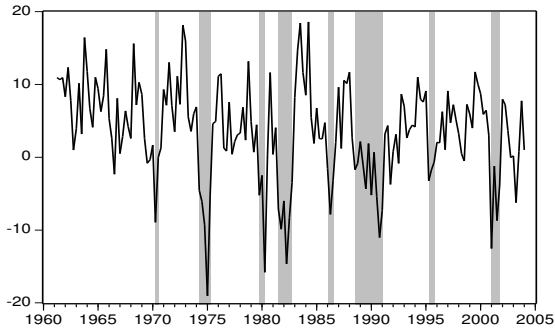
Parameter	Canada	France	Germany	Italy	Japan	UK	US
$\sigma_1$	5.22 (0.54)	6.22 (0.67)	6.66 (0.63)	8.34 (1.07)	6.28 (0.71)	5.43 (0.56)	5.03 (0.74)
$\sigma_2$	4.10 (0.49)	3.70 (0.36)	3.91 (0.47)	4.67 (0.49)	4.77 (0.50)	3.04 (0.33)	2.47 (0.21)
$\delta_1$	-2.28 (0.61)	-2.38 (1.11)	-1.44 (0.41)	-1.46 (0.54)	-3.47 (0.68)	-2.19 (0.55)	-2.58 (0.84)
$\delta_2$	-2.00 (0.43)	-1.30 (0.50)	-2.54 (0.50)	-1.56 (0.38)	-2.55 (0.69)	-2.08 (0.54)	-3.72 (0.48)
$p = \begin{cases} 0 \\ 1 \end{cases}$	0.40 0.58	0.96 0.04	0.98 0.02	0.94 0.06	0.00 0.97	0.95 0.05	0.00 0.99
$\phi_1$	0.21 (0.20)	0.00 (0.03)	0.00 (0.03)	-0.01 (0.05)	0.58 (0.10)	0.01 (0.05)	0.60 (0.08)
$p_{00,\delta}$	0.91 (0.05)	0.90 (0.09)	0.90 (0.04)	0.81 (0.06)	0.97 (0.02)	0.93 (0.03)	0.96 (0.02)
$p_{11,\delta}$	0.80 (0.06)	0.75 (0.09)	0.74 (0.06)	0.73 (0.06)	0.85 (0.05)	0.78 (0.06)	0.83 (0.05)

The table presents posterior means of the parameters in the model (15)-(19) estimated for quarterly IP growth rates for the G7 countries over the period 1961Q2-2004Q1. Posterior standard deviations are given in parentheses.

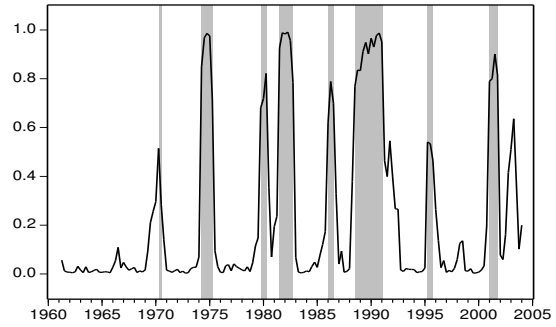
Table 2: Simulation results

Parameter	True	None		Medium		Heavy	
	Value	With	Without	With	Without	With	Without
$\delta$	-3.000	-3.298 (0.416)	-3.069 (0.288)	-3.314 (0.399)	-2.803 (0.385)	-3.585 (0.623)	-2.255 (0.814)
$\phi_1$	0.500	0.520 (0.074)	0.501 (0.071)	0.527 (0.078)	0.419 (0.114)	0.576 (0.108)	0.180 (0.322)
$\sigma_1$	0.100	0.100 (0.008)	0.105 (0.009)	0.099 (0.008)	0.121 (0.022)	0.094 (0.012)	0.206 (0.106)
$\sigma_2$	0.050	0.050 (0.004)	0.051 (0.004)	0.048 (0.004)	0.063 (0.013)	0.044 (0.006)	0.093 (0.043)
$\tau$	100	99.31 (3.618)	100.47 (3.444)	99.24 (4.097)	106.02 (7.272)	98.96 (3.829)	93.41 (9.076)
$p_{00,\delta}$	0.950	0.951 (0.015)	0.951 (0.014)	0.946 (0.016)	0.938 (0.019)	0.925 (0.029)	0.890 (0.069)
$p_{11,\delta}$	0.850	0.837 (0.034)	0.834 (0.034)	0.825 (0.042)	0.805 (0.050)	0.786 (0.070)	0.706 (0.151)
RMSE( $\mu_t$ )		0.054	0.052	0.052	0.063	0.061	0.087
CDO		0.025		0.831		0.822	

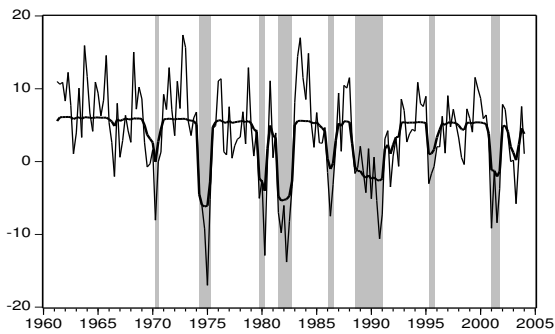
Note: The table presents results from the Monte Carlo simulation, where 200 samples of length  $n = 200$  are generated from the model nested in (15)-(19) as described in the text. Three levels of additive outlier contamination are considered: “None”, “Medium” (five AOs of absolute magnitude  $|K_{at}a_t| = 5$ ) and “Heavy” (20 AOs; 15 with  $|K_{at}a_t| = 5$  and five with  $|K_{at}a_t| = 7$ ). The cells contain averages of the posterior means of the parameters in the models with and without components that capture additive and innovation outliers (columns “With” and “Without”, respectively), with root mean square error (RMSE) in parentheses.  $\text{RMSE}(\mu_t)$  is defined as the square root of the average of  $(\mu_t^{\text{DGP}} - \mu_t^{\text{M}})^2$  across observations  $t = 1, \dots, n$  and across samples, where  $\mu_t = \nu_t + \sigma_t K_{\delta t} \delta_t$  and  $\mu_t^{\text{DGP}}$  and  $\mu_t^{\text{M}}$  are computed using the true and estimated parameters, respectively. CDO denotes the average fraction of correctly detected outliers (CDO), computed as the average number of iterations for which the model identifies an additive outlier ( $K_{at}a_t \neq 0$ ) at locations where these indeed occur. For the DGP with no contamination, CDO is the average posterior mean probability of occurrence of an additive outlier.



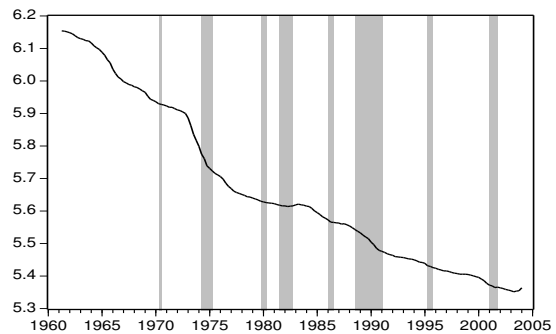
(a) Quarterly growth rate  $y_t$



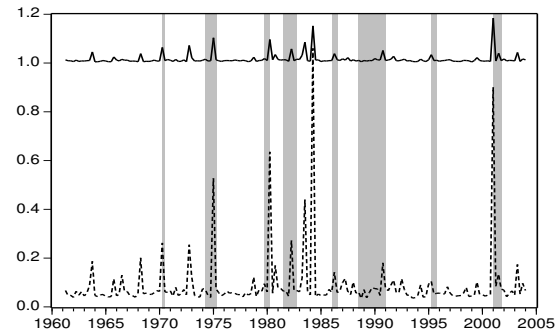
(b) Markov regime  $K_{\delta t}$



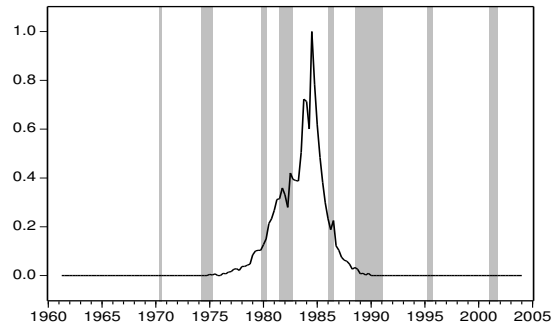
(c) Conditional mean  $\mu_t = \nu_t + \sigma_t \delta_t K_{\delta t}$



(d) Base-line growth rate  $\nu_t$

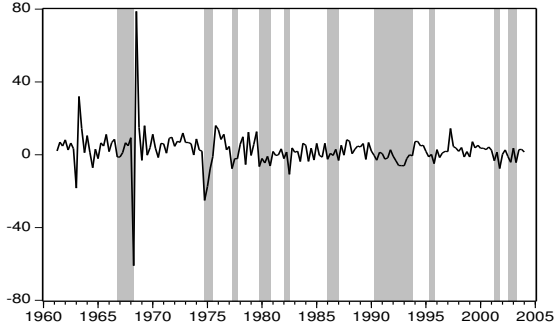


(e) Outliers  $K_{at}$  and  $K_{et}$

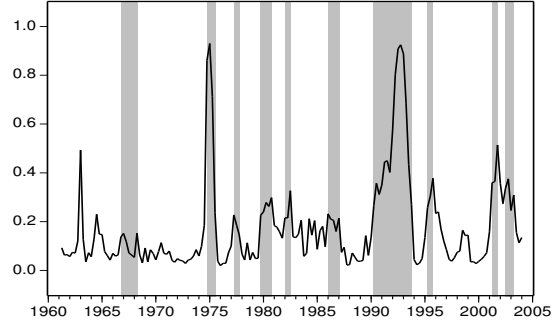


(f) Variance break date  $\tau$

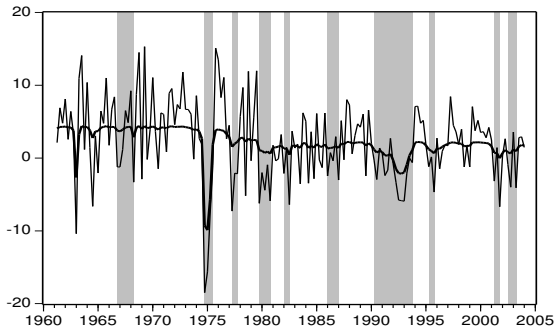
Figure 1: The graphs summarize the results of estimating the Markov-Switching model with outliers and structural change in mean and variance as given in (15)-(19) for quarterly Canadian IP growth rates over the period 1961Q2-2004Q1. The different panels contain: (a) the quarterly growth rate series  $y_t$ ; (b) the posterior mean of the regime-switching process  $K_{\delta t}$ ; (c) the posterior mean of the AO-corrected series  $y_t - \sigma_t K_{at} a_t$  (thin solid line) together with the posterior mean of  $\mu_t = \nu_t + \sigma_t K_{\delta t} \delta_t$  (thick solid line); (d) the posterior mean of  $\nu_t$ ; (e) the posterior means of  $K_{at}$  (dashed line) and  $K_{et}$  (solid line); and (f) the posterior distribution of  $\tau$ . Shaded areas indicate recessions as determined by the BBQ algorithm of Harding and Pagan (2002).



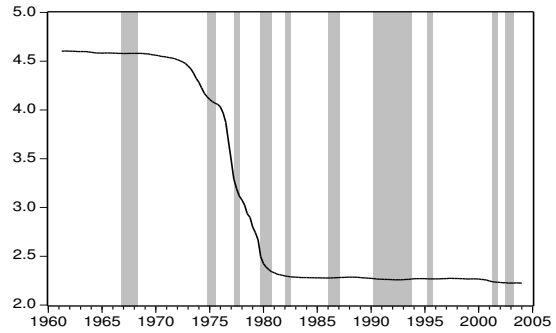
(a) Quarterly growth rate  $y_t$



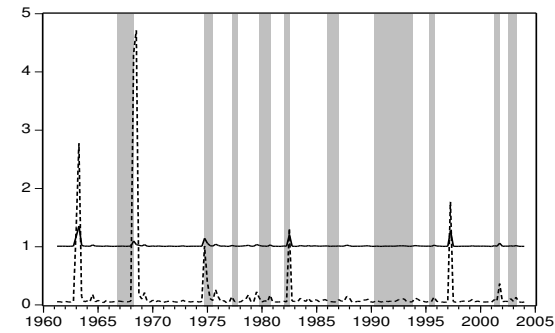
(b) Markov regime  $K_{\delta t}$



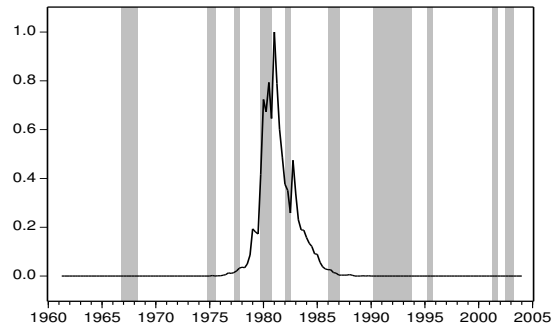
(c) Conditional mean  $\mu_t = \nu_t + \sigma_t \delta_t K_{\delta t}$



(d) Base-line growth rate  $\nu_t$



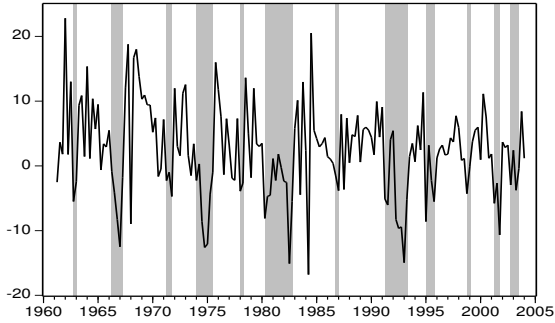
(e) Outliers  $K_{at}$  and  $K_{et}$



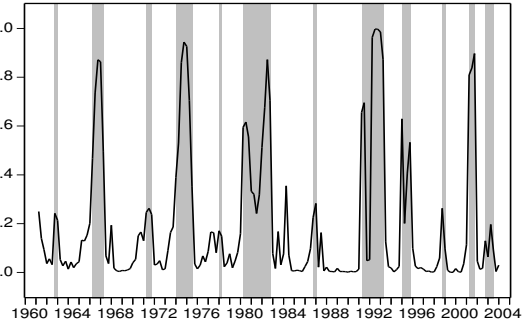
(f) Variance break date  $\tau$

Figure 2: The graphs summarize the results of estimating the Markov-Switching model with outliers and structural change in mean and variance as given in (15)-(19) for quarterly French IP growth rates over the period 1961Q2-2004Q1. See Figure 1 for details.

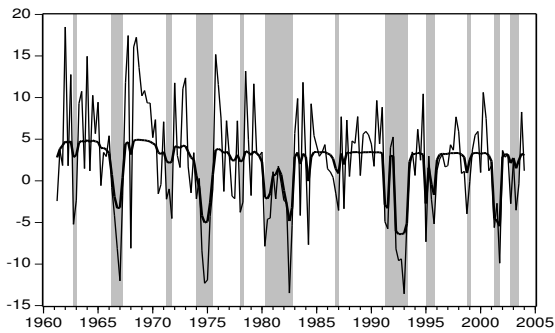




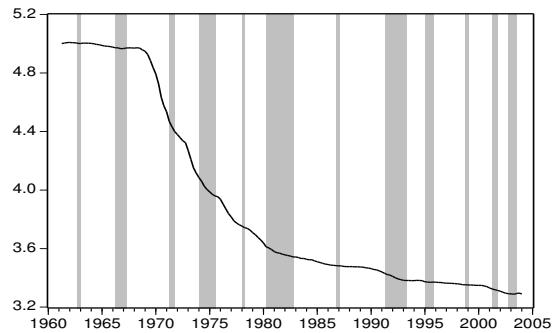
(a) Quarterly growth rate  $y_t$



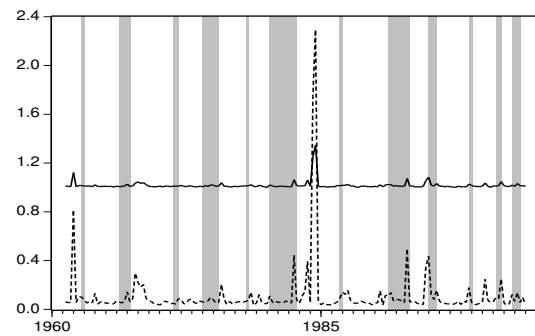
(b) Markov regime  $K_{\delta t}$



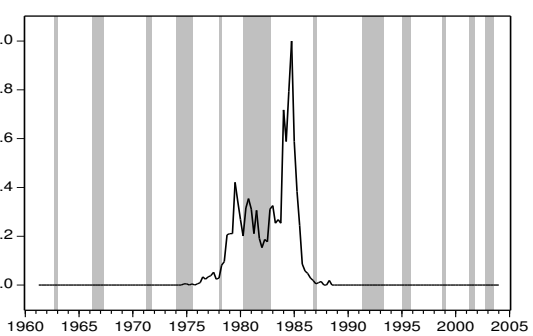
(c) Conditional mean  $\mu_t = \nu_t + \sigma_t \delta_t K_{\delta t}$



(d) Base-line growth rate  $\nu_t$

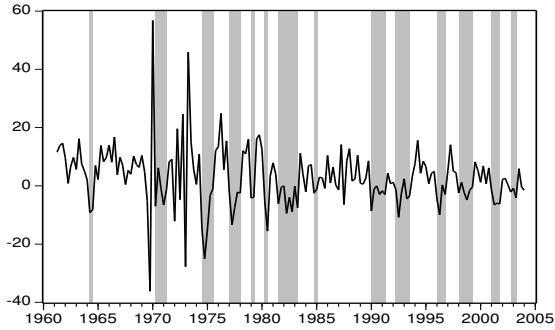


(e) Outliers  $K_{at}$  and  $K_{et}$

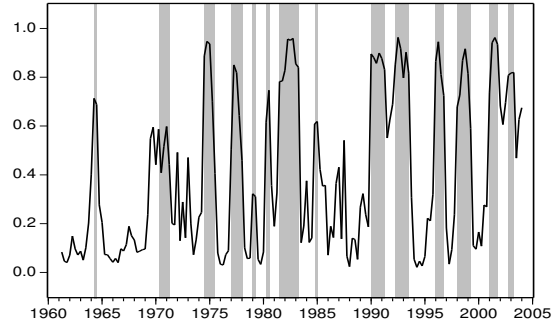


(f) Variance break date  $\tau$

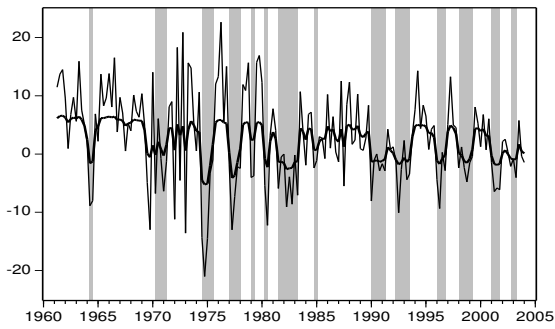
Figure 3: The graphs summarize the results of estimating the Markov-Switching model with outliers and structural change in mean and variance as given in (15)-(19) for quarterly German IP growth rates over the period 1961Q2-2004Q1. See Figure 1 for details.



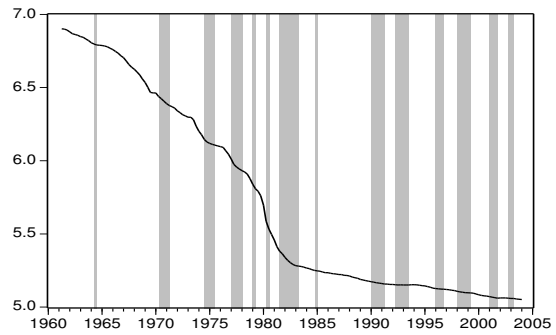
(a) Quarterly growth rate  $y_t$



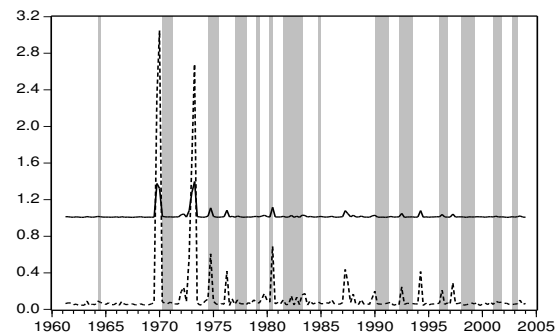
(b) Markov regime  $K_{\delta t}$



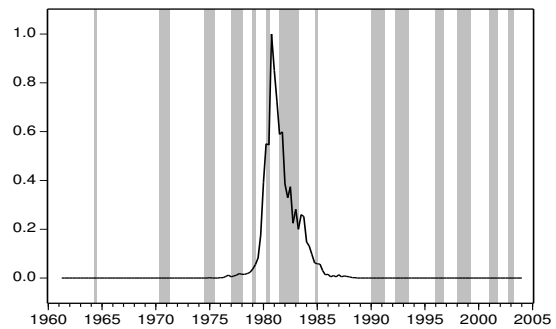
(c) Conditional mean  $\mu_t = \nu_t + \sigma_t \delta_t K_{\delta t}$



(d) Base-line growth rate  $\nu_t$

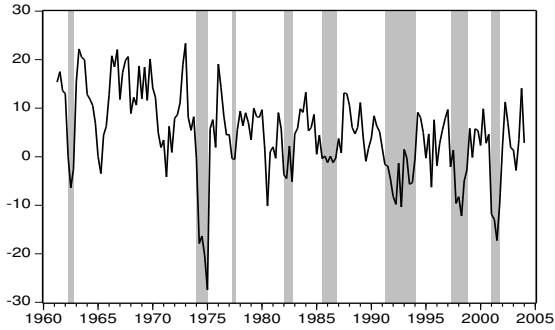


(e) Outliers  $K_{at}$  and  $K_{et}$

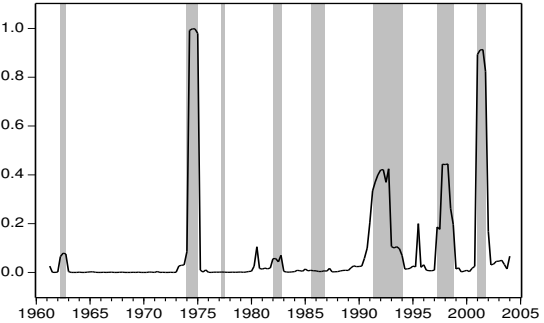


(f) Variance break date  $\tau$

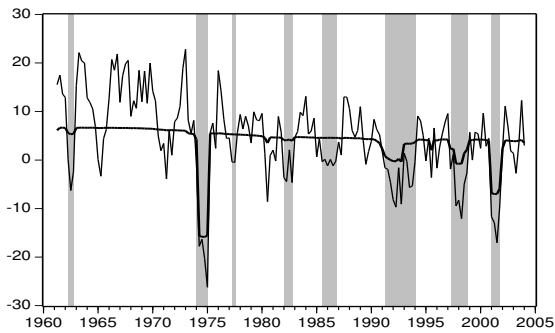
Figure 4: The graphs summarize the results of estimating the Markov-Switching model with outliers and structural change in mean and variance as given in (15)-(19) for quarterly Italian IP growth rates over the period 1961Q2-2004Q1. See Figure 1 for details.



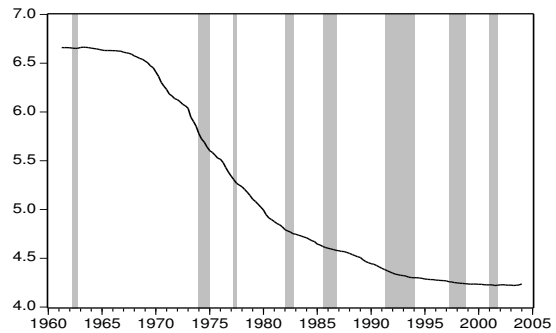
(a) Quarterly growth rate  $y_t$



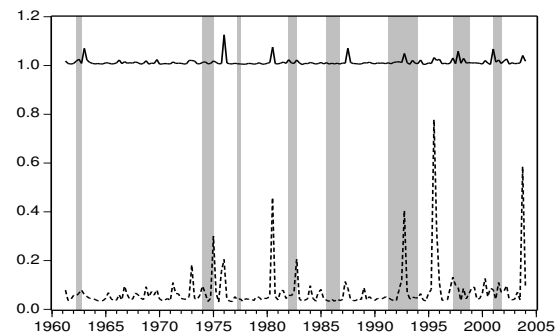
(b) Markov regime  $K_{\delta t}$



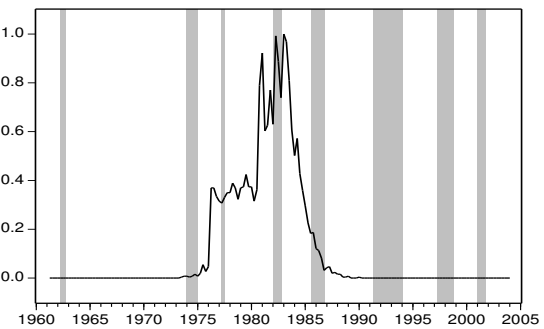
(c) Conditional mean  $\mu_t = \nu_t + \sigma_t \delta_t K_{\delta t}$



(d) Base-line growth rate  $\nu_t$

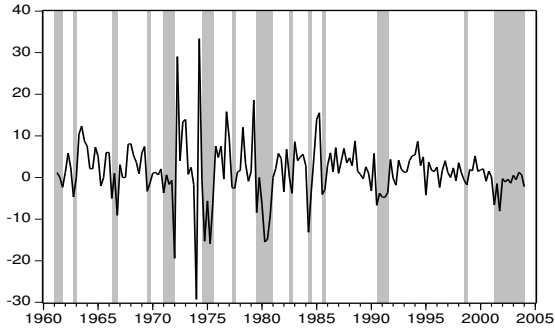


(e) Outliers  $K_{at}$  and  $K_{et}$

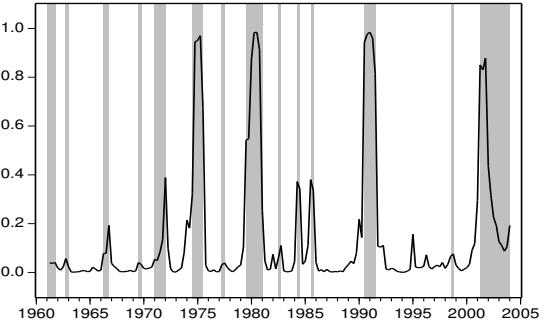


(f) Variance break date  $\tau$

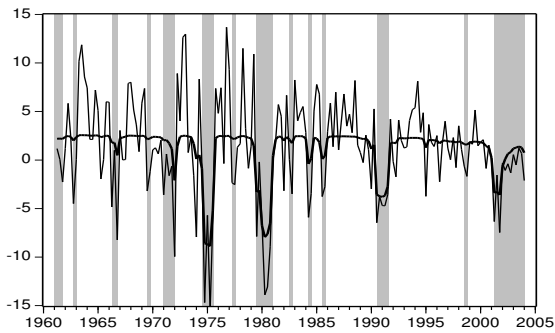
Figure 5: The graphs summarize the results of estimating the Markov-Switching model with outliers and structural change in mean and variance as given in (15)-(19) for quarterly Japanese IP growth rates over the period 1961Q2-2004Q1. See Figure 1 for details.



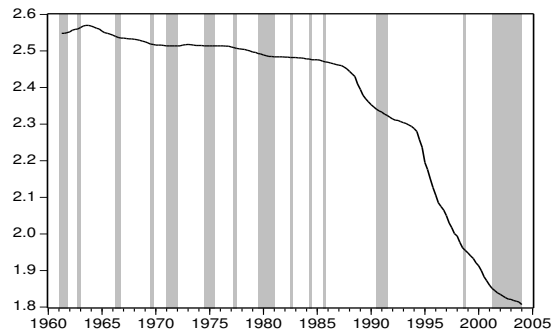
(a) Quarterly growth rate  $y_t$



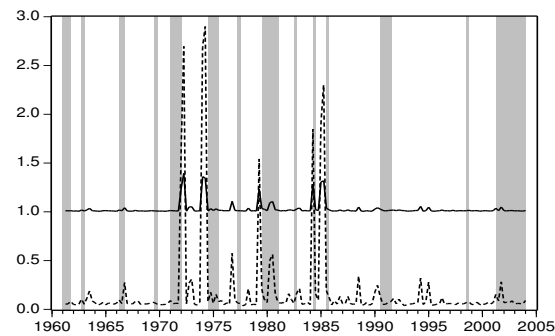
(b) Markov regime  $K_{\delta t}$



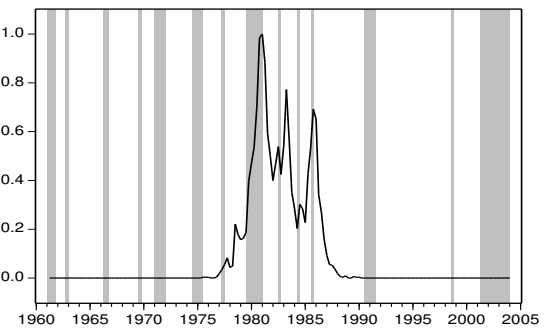
(c) Conditional mean  $\mu_t = \nu_t + \sigma_t \delta_t K_{\delta t}$



(d) Base-line growth rate  $\nu_t$

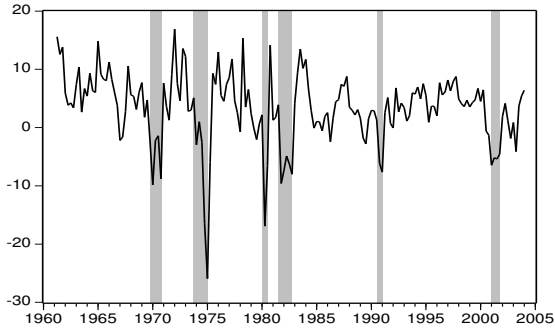


(e) Outliers  $K_{at}$  and  $K_{et}$

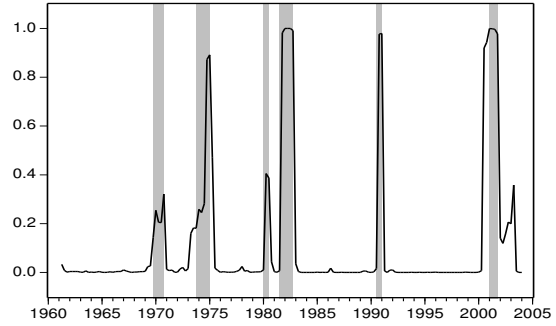


(f) Variance break date  $\tau$

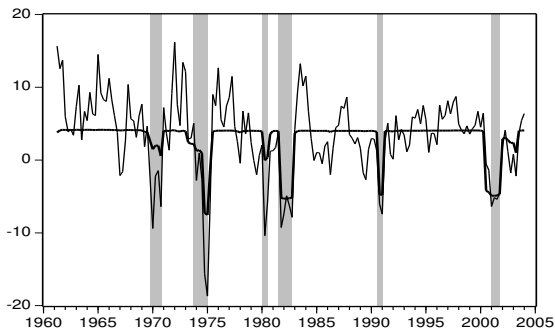
Figure 6: The graphs summarize the results of estimating the Markov-Switching model with outliers and structural change in mean and variance as given in (15)-(19) for quarterly UK IP growth rates over the period 1961Q2-2004Q1. See Figure 1 for details.



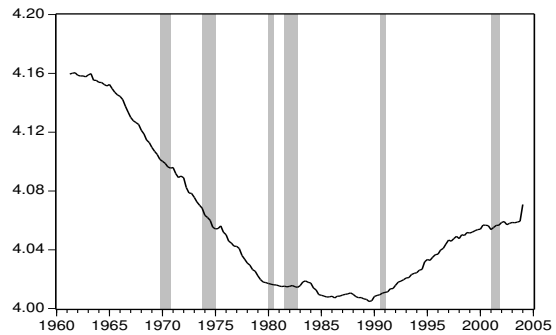
(a) Quarterly growth rate  $y_t$



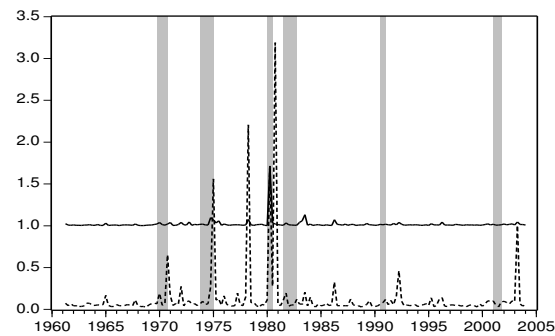
(b) Markov regime  $K_{\delta t}$



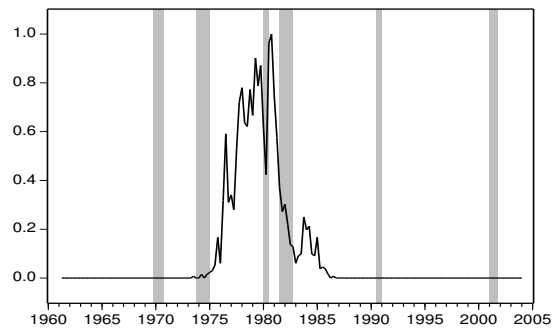
(c) Conditional mean  $\mu_t = \nu_t + \sigma_t \delta_t K_{\delta t}$



(d) Base-line growth rate  $\nu_t$



(e) Outliers  $K_{at}$  and  $K_{et}$



(f) Variance break date  $\tau$

Figure 7: The graphs summarize the results of estimating the Markov-Switching model with outliers and structural change in mean and variance as given in (15)-(19) for quarterly US IP growth rates over the period 1961Q2-2004Q1. See Figure 1 for details.