

A two-country model of stochastic output with changes in regime

Kerk L. Phillips*

University of Rochester, Rochester, NY 14627, USA, and University of Michigan, Ann Arbor, MI 48109, USA

Received February 1990, revised version received September 1990

This paper evaluates the transmission of business cycles between countries and presents evidence that world-wide shocks dominate any transmission of business cycles. I examine industrial production and look at the relationships of the United States with Canada, Germany and the United Kingdom. In the process, I present a specialized Markov-switching time-series model and evaluate it relative to other more familiar stochastic models. I show that for the data sets considered, Markov-switching models have much better goodness of fit, but larger forecast variance relative to ARMA or VAR models.

1. Introduction

There are many ways to examine the international transmission of business cycles. Layton (1987) uses monthly observations of NBER 'growth cycle' time series to test if the U.S. growth cycle Granger-causes that of Australia. Boehm and Moore (1984) use a turning point approach.

Hamilton (1989) introduced a statistical model of output that can be slightly modified to examine business cycle transmission. In his model the log of output follows a random walk with drift. The drift parameter follows a two-state Markov process. When the drift parameter takes on the lower of the two values the economy is said to be in recession. When it takes on the higher value the economy is in a boom. Using this method with quarterly data for the United States, Hamilton estimates the high-growth value to be 1.16% and the low growth value to be -0.36% . The diagonal elements of the transition matrix for his two-state process are 0.905 and 0.755, implying an expected length of 10.5 quarters for booms and 4.1 quarters for recessions.

*I am indebted to Alan Stockman for insight and guidance. Working papers graciously provided by James D. Hamilton were vital to the completion of this paper and are acknowledged with gratitude. Thanks are due also to an anonymous referee for helpful comments, to Tom Cooley and to the participants of the international workshop at the University of Rochester. This research was supported in part by the Alfred P. Sloan Foundation's Workshop on International Markets at the University of Rochester. I, of course, am responsible for all errors.

His estimation technique also gives inferences about the state of the Markov process for each observation. These inferences match up quite well with the NBER's dating of U.S. business cycle turning points.

It is a logical next step to extend Hamilton's analysis to a two-country setup. In such a case there could be a high- and low-growth parameter for each country for a total of four possible states of the world. The nature of the transition matrix will have strong implications for the international transmission of business cycles.

It should be stressed that such a model cannot attempt to explain the economic forces at work in the transmission of business cycles; rather, it attempts to characterize the behavior of economies. Just as knowing which dynamic processes accurately describe the behavior of GNP helps to build better economic models of output, it is hoped that the results of this paper will be of some help in constructing two-country business cycle models.

The remainder of this paper is organized as follows. Section 2 presents some empirical evidence on the growth rates of industrial production for the countries to be examined. This evidence includes average growth rates, cross-country correlations and parameters from a version of Hamilton's single-country model. The paper's emphasis on growth rates as opposed to levels or deviations from trend is motivated by testing the data for unit roots. Section 3 discusses the stochastic process assumed for output in the two-country model. Section 4 briefly presents my two-country modifications to the filter and estimation procedure given in Hamilton (1989). Section 5 discusses the implications of various cases of business cycle transmission for the transition matrix of the Markov process. Tests for various types of transition matrices are presented. Section 6 evaluates the model in this paper relative to other statistical models using various criteria. Predictive accuracy and goodness of fit are examined relative to a random walk, various ARMA models, and VARs. Section 7 presents the results of estimation. The linkage of the United States with three major trading partners is evaluated. Section 8 gives some concluding remarks.

2. Empirical evidence on growth rates

The data used in this paper are quarterly observations of seasonally adjusted industrial production. For Canada, the United Kingdom and the United States the data are taken from the IMF's International Financial Statistics magnetic tape. Data for Germany are from Deutsche Bundesbank publications. All series run from 1961:1 to 1989:3.

The average quarterly growth rates for the sample period are: 0.97% for Canada, 0.78% for Germany, 0.51% for the United Kingdom and 0.91% for the United States. Table 1 shows the correlations of quarterly growth rates. The highest correlation is 0.62 for the United States and Canada.

Table 1
Correlations of quarterly growth rates of industrial production.

| | Canada | Germany | U.K. |
|---------|--------|---------|-------|
| Germany | 0.313 | – | – |
| U.K. | 0.205 | 0.215 | – |
| U.S. | 0.625 | 0.383 | 0.300 |

Table 2
Augmented Dickey–Fuller tests for a unit root.

| | Test 1 | Test 2 | Test 3 |
|-------------------------------------------------------------|-------------------|------------|--------------------|
| Canada | 2.11 ^a | –2.36 | –2.06 |
| Germany | 1.88 ^b | –2.29 | –1.89 |
| U.K. | 2.58 ^a | –1.30 | –2.23 |
| U.S. | 2.59 ^a | –1.85 | –3.49 ^c |
| ^a Reject unit root hypothesis at 98% confidence. | unit root | hypothesis | at 98% |
| ^b Reject unit root hypothesis at 95% confidence. | unit root | hypothesis | at 95% |
| ^c Reject unit root hypothesis at 98% confidence. | unit root | hypothesis | at 98% |

Hamilton’s model and my two-country version both assume a unit root for output. This assumption needs to be examined. Table 2 presents a series of augmented Dickey–Fuller tests for a unit root. The first model used to construct the test statistic includes 12 lags of the first-difference of log output on the right-hand side. Model two adds a constant term, while model three adds both a constant and a deterministic trend. The test statistic used is the *t*-statistic from OLS regressions. Generally, the null hypothesis of a unit root cannot be rejected except for models without a constant term.

A simplified version of Hamilton’s GNP process was estimated for all four countries for the period 1962:2 to 1983:3. Since first-differencing results in the loss of one observation, this subsample allows me later to compare the model directly with ARMA and VAR models using up to four lags of the growth rate. The results are reported in table 3. The model is identical to Hamilton (1989) except an AR(1) term is used, rather than an AR(4). If we let *y_t* be the growth of output, the stochastic process assumed is

$$y_t = \mu_1 s_t + \mu_2 (1 - s_t) + \varepsilon_t; \quad \varepsilon_t = \rho \varepsilon_{t-1} + u_t; \quad u_t \sim N(0, \sigma^2), \tag{1}$$

where *s_t* is the state of the economy; 1 in boom and 0 in recession. The state is driven by a Markov process with the following transition matrix:

Table 3
Results from single-country estimation.

| | Canada | Germany | U.K. | U.S. |
|------------|---------|---------|---------|---------|
| Likelihood | −201.57 | −225.15 | −231.94 | −192.74 |
| μ_1 | 1.34 | 1.42 | 0.87 | 1.29 |
| μ_2 | −2.71 | −1.47 | −2.88 | −2.91 |
| σ^2 | 1.73 | 2.31 | 2.98 | 1.31 |
| ρ | 0.24 | −0.10 | −0.14 | 0.52 |
| π_{11} | 0.97 | 0.90 | 0.96 | 0.95 |
| π_{22} | 0.71 | 0.66 | 0.59 | 0.56 |
| p_1 | 1.00 | 1.00 | 1.00 | 1.00 |

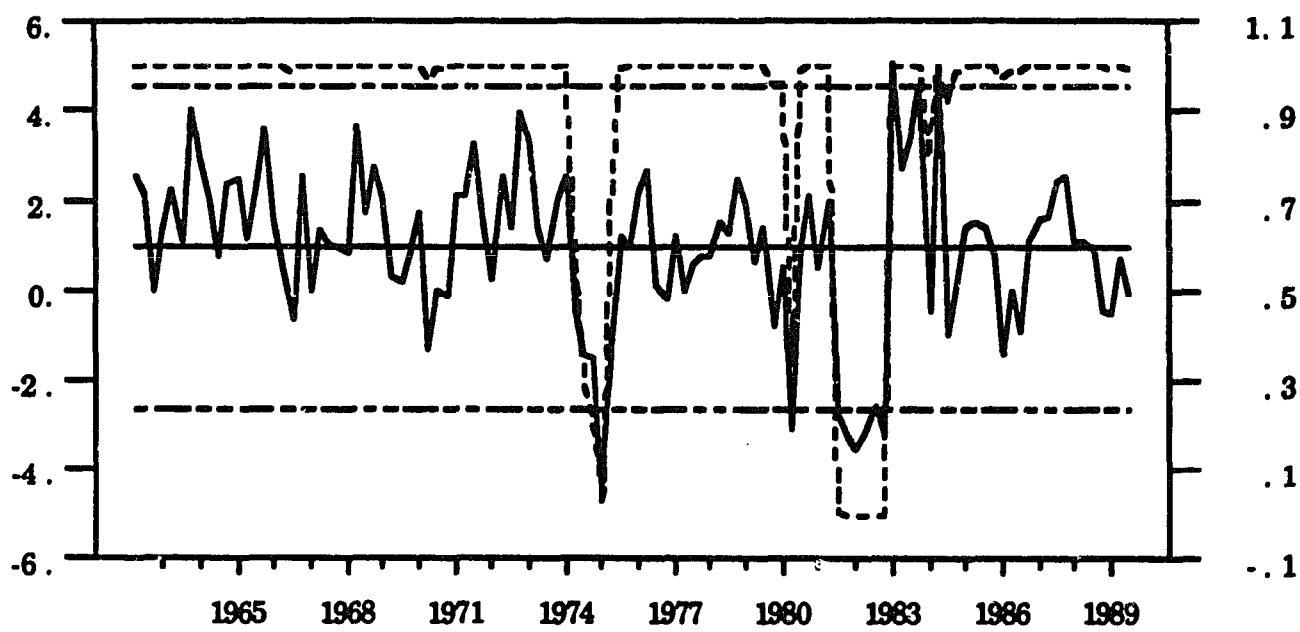


Fig. 1. Canada.

$$\begin{bmatrix} \pi_{11} & 1 - \pi_{11} \\ 1 - \pi_{22} & \pi_{22} \end{bmatrix}. \tag{2}$$

The parameters estimated and reported are: μ_1 and μ_2 , the high and low growth rates; σ^2 , the variance of the error term, π_{11} and π_{22} , the transition probabilities; and ρ , the autocorrelation coefficient. In addition, the probability of the date zero state was treated as a free parameter and is reported as p_1 .

The growth rates for each country are plotted in figs. 1–4. The solid straight line traces the mean value and the irregularly dashed lines show the mean value plus and minus two standard deviations. The scale on the left-hand side of each panel is for these series. In addition, the full-sample

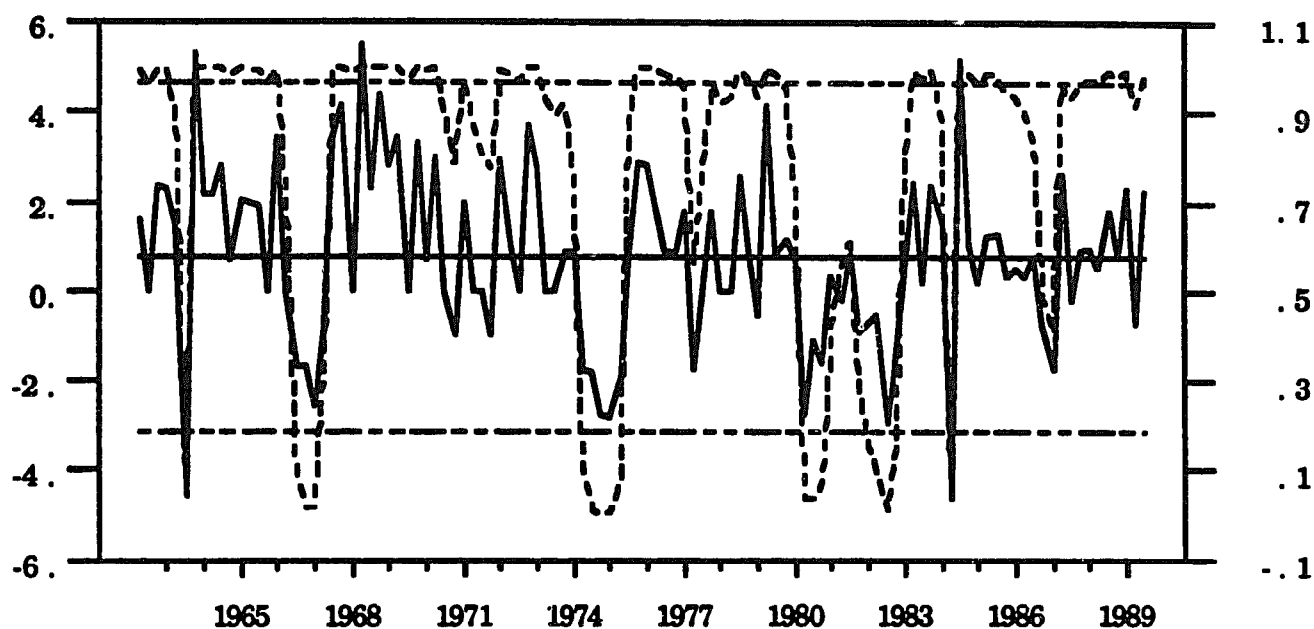


Fig. 2. Germany.

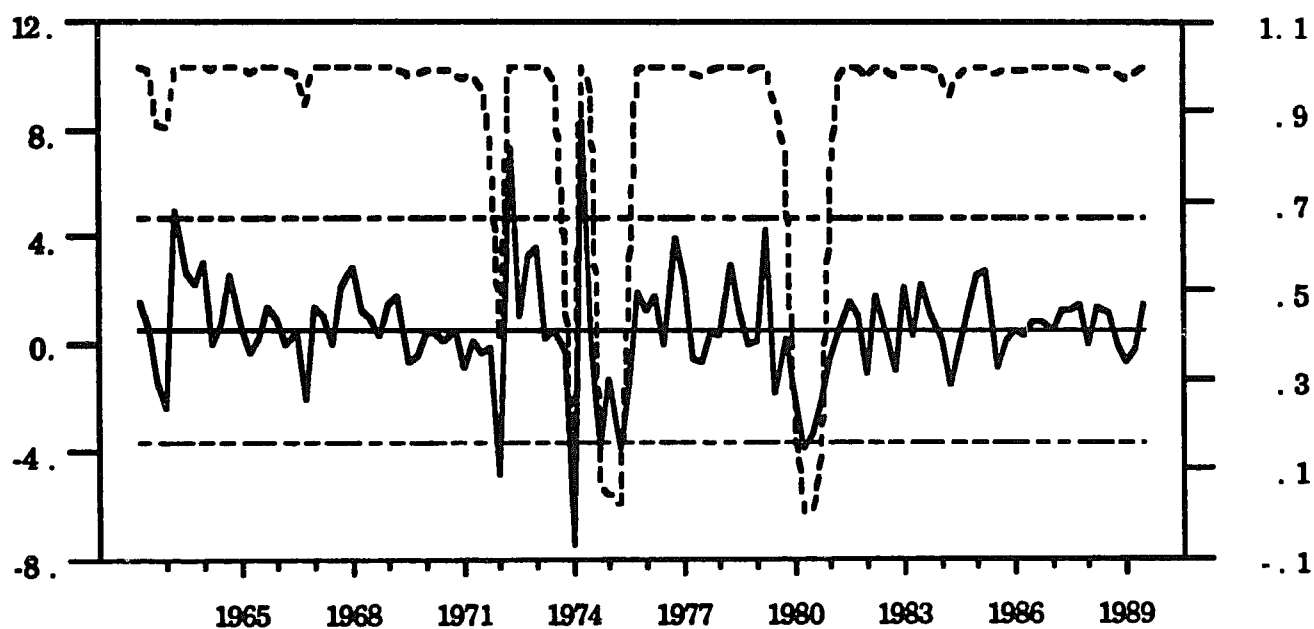


Fig. 3. United Kingdom.

smoothed probabilities of the boom state are plotted with a regular dashed line. The scale for these probabilities is shown on the right-hand side of the panel.

3. A two-country stochastic process

Hamilton (1989) models growth as the sum of an autoregressive process and a Markov process in the trend. I will work in this paper with a similar formulation. Letting y_t denote growth of output in period t :

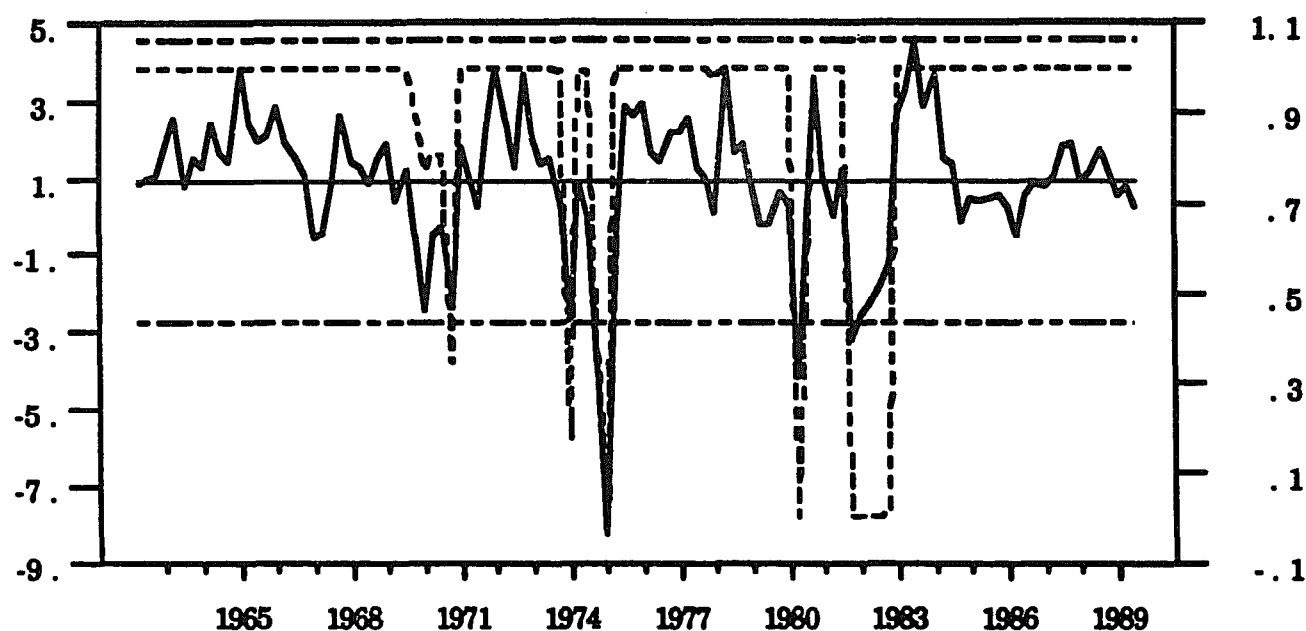


Fig. 4. United States.

$$y_t = n_t + \varepsilon_t, \quad (3)$$

$$n_t = \mu_1 s_{1t} + \mu_2 s_{2t} + \mu_3 s_{3t} + \mu_4 s_{4t},$$

where $s_{it} = 1$, if state at date t is state i ,

$$= 0 \quad \text{otherwise,} \quad (4)$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t; \quad u_t \sim N(0, \Sigma). \quad (5)$$

Eq. (5) allows the error terms to be vector-autocorrelated.

For a two-country model, y_t , n_t , ε_t , u_t and for all four μ 's are two-by-one vectors. I will use an 'h' superscript to denote the element of each vector corresponding to the home country and an 'f' for the foreign country. Σ is a two-by-two matrix with diagonal elements σ^h , σ^f and off-diagonal element σ^{hf} , ρ is a two-by-two matrix of autocorrelation coefficients like those computed by standard VARs.

There are two possible states for each country; the four different combinations of these will be the four different states in the Markov process. I will use the following convention to define these four states throughout this paper. State 1: both countries are in high growth. State 2: the home country is in low growth while the foreign country is in high growth. State 3: the home country has high growth and the foreign has low. State 4: both countries have low growth. This convention gives the following values to the four μ vectors:

$$\mu_1 = \begin{bmatrix} \mu_1^h \\ \mu_1^f \end{bmatrix}; \quad \mu_2 = \begin{bmatrix} \mu_2^h \\ \mu_1^f \end{bmatrix}; \quad \mu_3 = \begin{bmatrix} \mu_1^h \\ \mu_2^f \end{bmatrix}; \quad \mu_4 = \begin{bmatrix} \mu_2^h \\ \mu_2^f \end{bmatrix}, \quad (6)$$

where $\mu_1 > \mu_2$ for both countries.

There is also a transition matrix for the Markov process which is a four-by-four matrix of probabilities, π_{ij} , where $\pi_{ij} = \Pr(s_t = j | s_{t-1} = i)$. Of course these probabilities must sum to one over j for each i .

It is assumed that the y_t vector is observed, but the state, s_t , and the vector of shocks, ε_t , are unobserved. The objective is to construct an estimation procedure that will yield estimates of each of the four μ 's, the three σ 's in the Σ matrix and the twelve independent π 's.

4. Filter and estimation

The following section is an application of Hamilton's (1991) paper to the specific model outlined above.

The basic filter takes as inputs the data series and the parameters; it produces as outputs the log-likelihood and a set of smoothed probabilities. I will follow the convention of referring to the entire parameter set as θ . The filter goes through the data one observation at a time. It takes the probabilities of each possible prior two-period history conditional on observations of y up to time $t-1$ and updates them according to the following formulas. The notation $\Pr(\cdot)$ refers to a discrete probability set, while $P(\cdot)$ denotes a probability density function. First, it computes the joint probability density of s_t , s_{t-1} and y_t conditional on observations 1 to $t-1$. This is a set of 16 values:

$$\begin{aligned} & \Pr(s_t, s_{t-1}, y_t | y_{t-1}, \dots, y_1; \theta) \\ &= \sum_{s_{t-2}=1}^4 \Pr(s_{t-1}, s_{t-2} | y_{t-1}, \dots, y_1; \theta) \Pr(s_t | s_{t-1}; \theta) P(y_t | \Omega_t; \theta). \end{aligned} \quad (7)$$

Since u_t is distributed normally, the probability density of y_t conditioned on Ω_t ($\Omega_t = \{y_{t-1}, s_t, s_{t-1}\}$), which is used in (7), is given by

$$P(y_t | \Omega_t; \theta) = \frac{1}{2\pi |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} u_t' \Sigma^{-1} u_t \right\}. \quad (8)$$

where $u_t = y_t - \mu_{s_t} - \rho(y_{t-1} - \mu_{s_{t-1}})$. The value of $\Pr(s_t | s_{t-1})$ is just the appropriate π_{ij} element of the transition matrix.

Since there is a VAR term included in the model, I begin the filter on the

second observation. This ensures there is a value of y_{t-1} for eq. (8). I must also make provisions for the initial state probabilities, $\Pr(s_0)$. One approach is to constrain them to the values of an unconditional draw from the Markov process. Instead of this, I opted to treat these as free parameters and estimate them. I define these three free parameters as $p_i = \Pr(s_0 = i)$, for $i = 1, 2, 3$. Then, $\Pr(s_1 | s_0) \Pr(s_0)$ gives the probabilities in the right-hand side of eq. (7).

Next, the filter calculates the probability density of y_t conditional on observations 1 to $t-1$ by summing the left-hand side of (7) over s_t and s_{t-1} :

$$P(y_t | y_{t-1}, \dots, y_1; \theta) = \sum_{s_t=1}^4 \sum_{s_{t-1}=1}^4 P(s_t, s_{t-1}, y_t | y_1, \dots, y_1; \theta). \quad (9)$$

Finally, it calculates the probabilities of s_t and s_{t-1} conditional on observations 1 to t , by dividing the left-hand side of (7) by the left-hand side of (9):

$$\Pr(s_t, s_{t-1} | y_t, \dots, y_1; \theta) = \frac{P(s_t, s_{t-1}, y_t | y_{t-1}, \dots, y_1; \theta)}{P(y_t | y_{t-1}, \dots, y_1; \theta)}. \quad (10)$$

The EM algorithm used in this paper requires full-sample smoothed probabilities. These are probabilities of a particular two-period history at some date τ conditional on observations 1 to T . This requires that the filter also calculates and updates the following set of 256 probabilities:

$$\begin{aligned} & \Pr(s_t, s_{t-1}, s_{\tau}, s_{\tau-1} | y_t, \dots, y_1; \theta) \\ &= \frac{\sum_{s_{t-2}=1}^4 \Pr(s_{t-1}, s_{t-2}, s_{\tau}, s_{\tau-1} | y_{t-1}, \dots, y_1; \theta) \Pr(s_t, s_{t-1}) P(y_t | \Omega_t; \theta)}{P(y_t | y_{t-1}, \dots, y_1; \theta)} \end{aligned} \quad (11)$$

When $t = \tau$, $\Pr(s_t, s_{t-1}, s_{\tau}, s_{\tau-1} | y_t, \dots, y_1; \theta)$ is simply $\Pr(s_t, s_{t-1} | y_t, \dots, y_1; \theta)$, which is the set of 16 probabilities already calculated in (10). Thus, each pass through the filter it calculates a set of 16 probabilities and updates a set of 256 probabilities for each past observation.

Once (11) is calculated the filter moves to observation $t+1$ and repeats the process. The log-likelihood can be calculated by simply taking the log of the left-hand side of (9) and summing over all passes through the filter.

When the filter has passed through each observation there will be a set of $256T$ conditional probabilities that have been updated. The full-sample smoothed probabilities that the filter outputs are $\Pr(s_{\tau}, s_{\tau-1} | y_T, \dots, y_1; \theta)$ and these are calculated by summing $\Pr(s_T, s_{T-1}, s_{\tau}, s_{\tau-1} | y_T, \dots, y_1; \theta)$ over s_T and

s_{T-1} from (11). Summing these probabilities over s_{t-1} will give a full-sample smoothed probability for each state at a given observation, $\Pr(s_t | y_T, \dots, y_1; \theta)$. Since the home country is in a boom in states 1 and 3, further summing these probabilities over $s_t=1$ and $s_t=3$ gives a full-sample smoothed probability that the home country is in its high-growth state at date t . Summing over $s_t=1$ and $s_t=2$ gives the corresponding probability for the foreign country.

An appendix giving a flow chart for the basic filter is available from the author upon request.

Hamilton (1991) shows how the EM algorithm of Dempster, Laird and Rubin (1977) can be used in conjunction with his basic filter to estimate the model's parameters. With an initial guess the log-likelihood and the full-sample inferences, $\Pr(s_t, s_{t-1} | y_T, \dots, y_1; \theta)$, can be computed. These inferences are used to update the five types of parameters; the μ 's, ρ 's, σ 's, π 's, and p 's. A second appendix showing the derivation of the updating equations for this model is available from the author. Once updated values of the parameters have been found they can be used to calculate a new value of the log-likelihood and full-sample inferences. Iterating in this method gives the maximum likelihood estimates of the parameters. The convergence criterion used in this paper is that each parameter estimate changes by less than 10^{-7} between iterations.

5. Implications of the transition matrix

The correlation of business cycles across countries will determine the nature of the transition matrix. For example, suppose the two countries each follow an independent regime-shifting process like that discussed in section 2. If the processes are truly independent, then the four-state transition matrix will look like the following:

$$\begin{bmatrix} \pi_{11}^h \pi_{11}^f & (1 - \pi_{11}^h) \pi_{11}^f & \pi_{11}^h (1 - \pi_{11}^f) & (1 - \pi_{11}^h) (1 - \pi_{11}^f) \\ (1 - \pi_{22}^h) \pi_{11}^f & \pi_{22}^h \pi_{11}^f & (1 - \pi_{22}^h) (1 - \pi_{11}^f) & \pi_{22}^h (1 - \pi_{11}^f) \\ \pi_{11}^h (1 - \pi_{22}^f) & (1 - \pi_{11}^h) (1 - \pi_{22}^f) & \pi_{11}^h \pi_{22}^f & (1 - \pi_{11}^h) \pi_{22}^f \\ (1 - \pi_{22}^h) (1 - \pi_{22}^f) & \pi_{22}^h (1 - \pi_{22}^f) & (1 - \pi_{22}^h) \pi_{22}^f & \pi_{22}^h \pi_{22}^f \end{bmatrix}. \quad (12)$$

If we are capturing shifts in regime that occur at business cycle frequencies we expect the value of the π 's to be in the range of 0.70–0.95. This means the diagonal elements will be large relative to the rest of the matrix. For example, if $\pi_{11}^h = \pi_{11}^f = 0.9$ and $\pi_{22}^h = \pi_{22}^f = 0.7$ we obtain the following matrix

$$\begin{bmatrix} 0.81 & 0.09 & 0.09 & 0.01 \\ 0.27 & 0.63 & 0.03 & 0.07 \\ 0.27 & 0.03 & 0.63 & 0.07 \\ 0.09 & 0.21 & 0.21 & 0.49 \end{bmatrix}. \quad (13)$$

If the two Markov processes are perfectly correlated, then they could be represented by a two-by-two transition matrix. However, the four-by-four matrix would look like the following:

$$\begin{bmatrix} \pi_{11} & 0 & 0 & (1-\pi_{11}) \\ - & - & - & - \\ - & - & - & - \\ (1-\pi_{22}) & 0 & 0 & \pi_{22} \end{bmatrix}. \quad (14)$$

The values in rows two and three are irrelevant since these states never occur.

The transition matrix can also allow for cases where one country leads the other. For example, suppose the foreign country is always in the same state the home country was in last period. This would be the case where the home country leads the foreign one into and out of recessions. In this case the transition matrix would look like the following:

$$\begin{bmatrix} \pi_{11}^h & (1-\pi_{11}^h) & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & (1-\pi_{22}^h) & \pi_{22}^h \end{bmatrix}. \quad (15)$$

A similar matrix can be constructed for the case where the foreign country leads. It is also possible to allow for expected leads of longer than one period. Since the Markov process is first-order, it is impossible to model the case where the lead is exactly two periods; this would require a second-order process. The matrix below illustrates a case where the expected length of the home country's lead into low growth is $1/(1-\alpha)$ and its expected lead into high growth is $1/(1-\beta)$:

$$\begin{bmatrix} \pi_{11}^h & (1-\pi_{11}^h) & 0 & 0 \\ 0 & \alpha & 0 & 1-\alpha \\ 1-\beta & 0 & \beta & 0 \\ 0 & 0 & (1-\pi_{22}^h) & \pi_{22}^h \end{bmatrix}. \quad (16)$$

It appears from the examples given above that this four-state Markov

process permits a great variety of cross-country business cycle transmissions to occur. One possibility that is rather difficult to model, however, is the case where two countries alternate their leads into and out of recessions. If the home country is a clear leader, as illustrated above, state 2 will always be followed by state 4 and state 3 will always be followed by state 1. However, if the two countries alternate leads, then state 2 could occur before state 4 if the home country leads into a recession and before state 1 if the foreign country leads out of a recession. A similar statement applies to state 3. This result is owing to the fact that the Markov process is first-order; only last period's state is allowed to influence the determination of this period's. For the case where leads alternate, this is clearly not the case; one would expect that a recovery by one country would not be followed by a recession immediately the next quarter. This implies that state 4 is less likely to follow state 2 or 3 if it also immediately preceded them. For cases where this occurs it will be necessary to look not only at the transition matrix, but also at the smoothed probabilities that generate the estimates. This will show how the states actually evolved over time.

Another problem that cannot be explained well would be high values for the parameters π_{23} and π_{32} . States 2 and 3 are when one country is in recession and the other in a boom. High values for these two transition probabilities imply that the states of the two countries will switch the next quarter. Such a scenario is not very appealing in that it does not seem to mimic business cycle behavior very well. The only case where high values might be consistent with business cycles is where the foreign country follows the home country out of a recession just as a new recession hits the home country. This would imply a recovery of only one or two quarters; but this is not inconsistent with the actual experience of the United States in 1980 and 1981 where the recovery was very short.

If the Markov processes driving states in each country are independent, then the four-state process estimated above can be represented by eq. (12). It is straightforward to conduct tests of independence using a likelihood ratio test. There are four transition probabilities in the restricted estimation. The normal equations for these are presented and solved for in the appendix at the end of this paper. With the value of the log-likelihood from this restricted estimation as well as the original unrestricted estimation above one can construct a likelihood ratio statistic. Since there were twelve independent transition probabilities in the original estimation and four when restricted, the statistic will be distributed $\chi^2(8)$.

Tests for transition matrices like those in eqs. (14) and (16) are also easy to construct. They require the model to be estimated with specific transition probabilities set to zero. Since one property of the filter and algorithm is that once a transition probability is set to zero it remains zero through all remaining iterations, one need simply set the initial guesses of the appropri-

ate transition probabilities to zero and the result will be the restricted parameter estimates. The test statistic for (14) is $\chi^2(12)$ and the statistic for (16) is $\chi^2(8)$. The parameters and log-likelihood values for the restricted and unrestricted estimations are given in table 4. The results of the likelihood ratio tests are given in table 5.

It is important to note here that likelihood ratio tests are based on the asymptotic properties of the estimators. The sample size is 110 observations, which might normally be a comfortable sample size for invoking asymptotic properties. However, some of the states of the Markov process occur only rarely. The transition probabilities are weighted sums of the full-sample smoothed probabilities. If one knew for certain when the economy was in each state, then these estimates would just be ratios of the number of times state j follows state i to the number of times state i occurs. If state j followed state i only once in the sample period, it is not reasonable to invoke asymptotic theory.

6. Evaluation of different models

I will consider two different criteria for evaluating the models in this paper relative to other, more common, ones. The first is accuracy of predictions and the second is goodness of fit. If the more complex models in this paper do not out-perform the others in at least one of these dimensions, then they are of little use.

I will compare a variety of models. First, I will consider univariate specifications which include: a random walk, various ARMA processes, and Hamilton's regime-shifting model. Secondly, I examine two kinds of two-country specifications: VARs with various lag lengths and the two-country model in section 3.

To measure predictive ability, I consider the nature of the one-period-ahead forecast errors. There are two criteria to consider. First, a good model will have an unbiased forecast, i.e. the expected value of the forecast error should be zero. And second, the variance of the forecast error should be as small as possible. To compare the different models on equal footing I estimated each one over the same set of data. To allow for models with lagged right-hand-side variables, such as an AR(4), the sample was begun four quarters after the first available observation. I used a very simple set of assumptions to calculate the in-sample forecast. First, I assume the parameters of the process under consideration are constant over time. Secondly, I assume the true parameters are the ones obtained from estimation. This implies that the forecast error for the ARMA and VAR models is just the residual. By construction the mean of the AR and VAR models must be zero. Thus, examination of the variance will determine which has the best predictive power. The results are presented in table 6.

Table 4
Results of two-country estimation.

| | | | | | | | | | |
|--------------------------------|-------|-----------------|------|------|---------------------------|------|------|------|------------|
| Baseline model | | | | | U.S. and Canada | | | | |
| | | | | | Log-likelihood: -370.13 | | | | |
| | Home | Foreign | | | Transition matrix | | | | P vector |
| μ_1 | 1.21 | 1.27 | | | 0.96 | 0.01 | 0.01 | 0.02 | 1.00 |
| μ_2 | -3.11 | -3.16 | | | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| ρ matrix | | Σ matrix | | | | | | | |
| 0.46 | | 0.23 | 1.26 | 0.36 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 0.32 | | 0.12 | 0.36 | 1.54 | 0.42 | 0.00 | 0.00 | 0.58 | 0.00 |
| Perfect correlation [eq. (14)] | | | | | Log-likelihood: -373.66 | | | | |
| | Home | Foreign | | | Transition matrix | | | | P vector |
| μ_1 | 1.17 | 1.24 | | | 0.97 | 0.03 | | | 1.00 |
| μ_2 | -3.30 | -3.28 | | | 0.43 | 0.57 | | | 0.00 |
| ρ matrix | | Σ matrix | | | | | | | |
| 0.46 | | 0.21 | 1.40 | 0.30 | | | | | |
| 0.36 | | 0.08 | 0.30 | 1.64 | | | | | |
| Baseline model | | | | | U.S. and Germany | | | | |
| | | | | | Log-likelihood: -370.13 | | | | |
| | Home | Foreign | | | Transition matrix | | | | P vector |
| μ_1 | 1.27 | 1.40 | | | 0.90 | 0.01 | 0.07 | 0.03 | 1.00 |
| μ_2 | -3.02 | -1.59 | | | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 |
| ρ matrix | | Σ matrix | | | | | | | |
| 0.52 | | -0.04 | 1.32 | 0.40 | 0.58 | 0.00 | 0.34 | 0.08 | 0.00 |
| 0.11 | | -0.14 | 0.40 | 2.21 | 0.06 | 0.00 | 0.33 | 0.61 | 0.00 |
| Perfect correlation [eq. (14)] | | | | | Log-likelihood: -411.67 | | | | |
| | Home | Foreign | | | Transition matrix | | | | P vector |
| μ_1 | 1.28 | 1.00 | | | 0.96 | 0.04 | | | 1.00 |
| μ_2 | -3.05 | -1.56 | | | 0.44 | 0.56 | | | 0.00 |
| ρ matrix | | Σ matrix | | | | | | | |
| 0.52 | | -0.01 | 1.32 | 0.48 | | | | | |
| 0.09 | | -0.04 | 0.48 | 3.29 | | | | | |
| U.S. leads [eq. (16)] | | | | | Log-likelihood: -412.15 | | | | |
| | Home | Foreign | | | Transition matrix | | | | P vector |
| μ_1 | 1.32 | 1.08 | | | 0.97 | 0.03 | 0.00 | 0.00 | 0.00 |
| μ_2 | -2.64 | -1.73 | | | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| ρ matrix | | Σ matrix | | | | | | | |
| 0.49 | | -0.03 | 1.50 | 0.56 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 0.04 | | -0.06 | 0.56 | 3.14 | 0.00 | 0.00 | 0.34 | 0.66 | 0.00 |

Table 4 (continued)

| | | | | | | | | | |
|--------------------------------|-------|-----------------|-------|-------|-------------------------|------|---------|------|----------|
| Baseline model | | | | | U.S. and U.K. | | | | |
| | | | | | Log-likelihood: -414.02 | | | | |
| | Home | Foreign | | | Transition matrix | | | | P vector |
| μ_1 | 1.25 | 0.85 | | | 0.93 | 0.01 | 0.06 | 0.00 | 1.00 |
| μ_2 | -3.37 | -2.56 | | | 0.00 | 0.74 | 0.00 | 0.26 | 0.00 |
| ρ matrix | | Σ matrix | | 0.57 | 0.00 | 0.00 | 0.43 | 0.00 | |
| | 0.55 | 0.01 | 1.35 | -0.11 | 0.51 | 0.00 | 0.39 | 0.10 | 0.00 |
| | 0.33 | -0.18 | -0.11 | 2.82 | | | | | |
| Independent [eq. (12)] | | | | | Log-likelihood: -419.12 | | | | |
| | Home | Foreign | | | Transition matrices | | | | P vector |
| | | | | | Home | | Foreign | | |
| μ_1 | 1.25 | 0.65 | | | 0.95 | 0.05 | 0.98 | 0.02 | 1.00 |
| μ_2 | -2.81 | -7.54 | | | 0.50 | 0.50 | 1.00 | 0.00 | 0.00 |
| ρ matrix | | Σ matrix | | 0.50 | 0.50 | 1.00 | 0.00 | 0.00 | |
| | 0.19 | 0.51 | 1.29 | 0.29 | | | | | 0.00 |
| | 0.45 | 0.08 | 0.29 | 3.04 | | | | | |
| Perfect correlation [eq. (14)] | | | | | Log-likelihood: -419.10 | | | | |
| | Home | Foreign | | | Transition matrix | | | | P vector |
| μ_1 | 1.30 | 0.70 | | | 0.96 | 0.04 | | | 1.00 |
| μ_2 | -3.20 | -1.61 | | | 0.43 | 0.57 | | | 0.00 |
| ρ matrix | | Σ matrix | | | | | | | |
| | 0.49 | 0.01 | 1.32 | -0.04 | | | | | |
| | 0.44 | -0.05 | -0.04 | 3.52 | | | | | |

Table 5
Likelihood ratio tests.

| Test | U.S. and Canada | U.S. and Germany | U.S. and U.K. |
|----------------------|--------------------|--------------------|--------------------|
| Independent | 21.06 ^a | 18.17 ^a | 10.21 |
| Perfectly correlated | 7.07 | 10.77 | 10.17 |
| Home leads | 21.24 ^a | 11.73 | 18.91 ^b |
| Foreign leads | 26.75 ^a | 18.85 ^b | 23.27 ^b |

^aReject null at 99% confidence.
^bReject null at 95% confidence.

For the regime-shifting models the forecast error was calculated as follows. First, the current-information inferences from the filter were calculated. These give the probabilities for today's state based on all observations up to the present. These and the transition probabilities are used to weight the different μ 's to obtain a forecast for y_{t+1} . The forecast error is simply the difference between this and the value of y_{t+1} that actually occurred. A general equation for the forecast error is given below:

$$\varepsilon_{t+1}^F = y_{t+1} - \sum_{s_t} \Pr(s_t | y_t, \dots, y_1) \left\{ \sum_{s_{t+1}} \Pr(s_{t+1} | s_t) \mu_{s_{t+1}} \right\}$$

Table 6
Forecast variances for various statistical models.

| Model | U.S. | Canada | U.S. | Germany | U.S. | U.K. |
|-------------|------|--------|------|---------|------|------|
| Random walk | 3.34 | 3.20 | | 3.81 | | 4.34 |
| ARMA(1,0) | 2.41 | 2.67 | | 3.75 | | 4.34 |
| ARMA(2,0) | 2.38 | 2.57 | | 3.65 | | 4.34 |
| ARMA(4,0) | 2.33 | 2.39 | | 3.58 | | 4.20 |
| ARMA(0, 1) | 2.45 | 2.84 | | 3.76 | | 4.34 |
| ARMA(0,2) | 2.25 | 2.41 | | 3.51 | | 4.22 |
| ARMA(1, 1) | 2.36 | 2.61 | | 3.67 | | 3.85 |
| MS(1) | 2.45 | 2.65 | | 3.80 | | 4.76 |
| VAR(1) | 2.27 | 2.45 | 2.41 | 3.51 | 2.41 | 4.03 |
| VAR(2) | 2.17 | 2.42 | 2.37 | 3.46 | 2.38 | 4.00 |
| VAR(4) | 2.00 | 2.22 | 2.19 | 3.40 | 2.18 | 3.88 |
| VMS(1) | 3.35 | 4.09 | 3.06 | 3.57 | 3.51 | 5.14 |

$$-\rho \left\{ y_t - \sum_{s_t} \Pr(s_t | y_t, \dots, y_1) \mu_{s_t} \right\}. \quad (17)$$

Eq. (17) shows why the forecast variance from these models might be larger than from the linear models. Consider a two-state case where the only uncertainty is from the Markov process, i.e. the error terms are all zero. This means y must take on a value of either μ_1 or μ_2 . However, the uncertainty associated with the Markov process means that even if today's state is known with certainty, the forecast will not be μ_1 or μ_2 , but their transition probability weighted sum.

Table 6 shows that the linear models generally have a lower forecast variance than the regime-shifting models. Here, as in table 7, MS(i) stands for Hamilton's single-country Markov-switching model with an autoregressive process of order i . VMS(i) stands for the model in section 3 with a vector-autoregressive process of order i . For every country a VAR(4) has the smallest forecast variance of all models considered. The single-country Markov-switching models seem to do roughly as well as the ARMA models, the exception being the United Kingdom, where it has a forecast variance that is 10% higher than that from a random walk. The two-country Markov-switching models have considerably higher variances than the VARs and – except for Germany – they are higher than those from the single-country model.

To measure goodness of fit, I consider the value of the log-likelihood. In table 7 it is clear that regime-shifting models have a substantially better fit than standard linear models. The only case where the single-country model is outperformed by an ARMA model is for the United Kingdom, where an

Table 7
Values of log-likelihood for various statistical models.

| Models | U.S. | Canada | U.S. | Germany | U.S. | U.K. |
|-------------|---------|---------|---------|---------|---------|---------|
| Random walk | −222.25 | −220.09 | | −229.63 | | −236.81 |
| ARMA(1,0) | −204.46 | −210.08 | | −228.74 | | −236.81 |
| ARMA(2,0) | −203.73 | −208.01 | | −227.27 | | −236.79 |
| ARMA(4,0) | −202.64 | −203.90 | | −226.22 | | −235.03 |
| ARMA(0,1) | −205.27 | −213.44 | | −228.95 | | −236.81 |
| ARMA(0,2) | −200.59 | −204.47 | | −225.21 | | −235.23 |
| ARMA(1,1) | −203.39 | −208.88 | | −227.52 | | −230.20 |
| VMS(1) | −192.74 | −201.57 | | −225.15 | | −231.94 |
| VAR(1) | −392.69 | | −423.92 | | −434.46 | |
| VAR(2) | −388.78 | | −421.98 | | −432.97 | |
| VAR(4) | −382.50 | | −415.83 | | −425.41 | |
| MS(1) | −370.13 | | −406.28 | | −414.02 | |

all cases. It is also interesting that, regardless of the model chosen, the likelihood rises notably when two countries are estimated together. That is to say, the likelihood for the VARs is much higher than the sum of the two corresponding AR models. The same is true of the Markov-shifting models.

Using the ARMA models as a baseline, there are two ways to improve the goodness of fit. One is to allow for Markov-switching processes and the other is to allow cross-country effects by estimating a VAR. Table 7 shows that for the United States with Germany and the United Kingdom the former leads to a substantially larger increase in the likelihood than the latter. For the United States with Canada, the opposite is true.

7. Results of estimation

This section examines the results of estimation for each country pair. A baseline model and four models with restrictions on the transition matrix, as explained above, were estimated. To ensure that global and not local maxima of the likelihood function were reached, each model was first estimated with a weak convergence criterion for a variety of starting values. With only a few exceptions, the different starting values yielded final values that were roughly the same. The final values yielding the highest value of the likelihood were then used as starting values for estimation with a much smaller convergence criterion. In general, the estimation technique identifies recessions as periods of large negative growth and booms as periods of mild positive growth. Also, the recessions identified are shorter and the booms longer than the results given by NBER business cycle dating. Figs. 5–7 plot the full-sample smoothed probabilities of a boom for both the United States and the appropriate foreign country for the unrestricted baseline model reported in table 4.

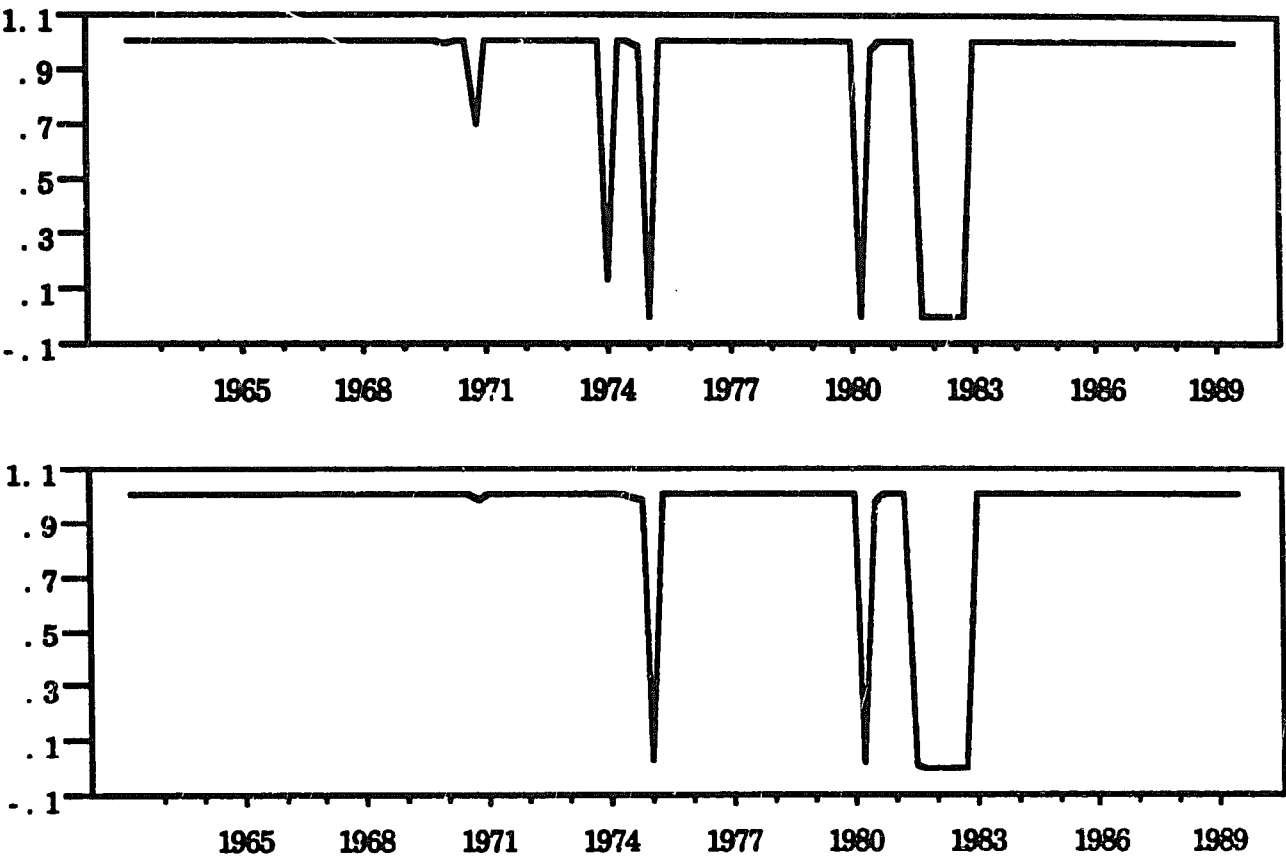


Fig. 5. United States/Canada.

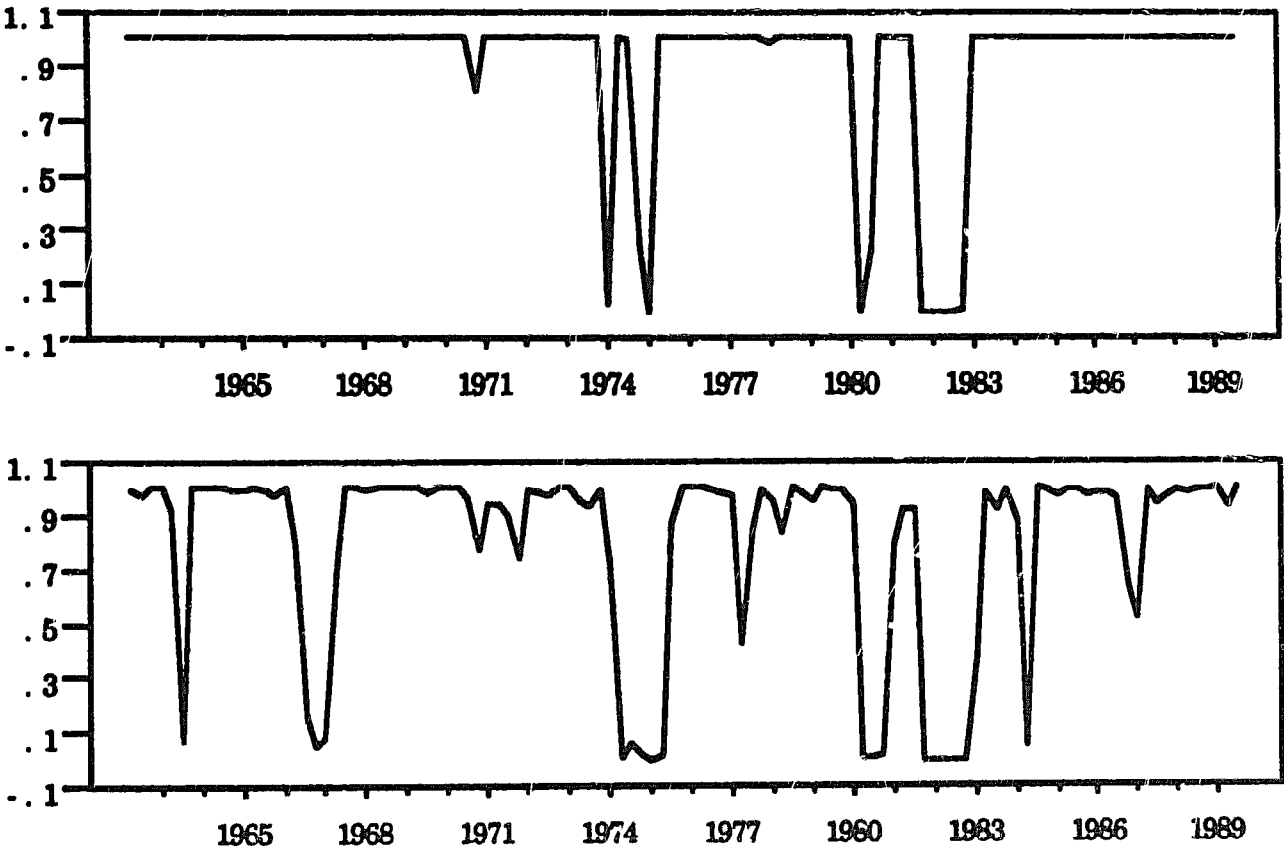


Fig. 6. United States/Germany.

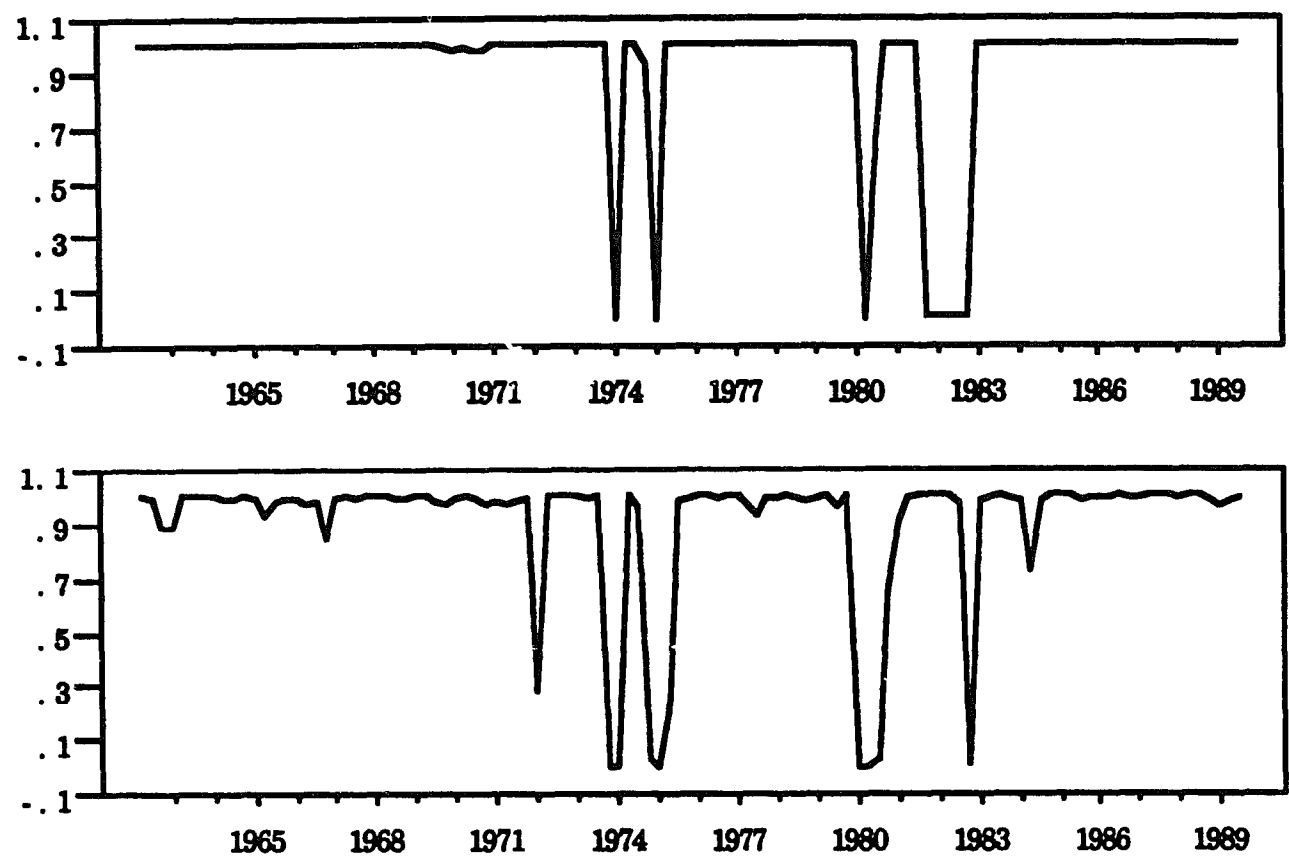


Fig. 7. United States/United Kingdom.

For the United States and Canada, table 5 shows that one can reject the hypotheses that the Markov processes driving the states are independent with 99% confidence. One can also reject the notion that either country leads the other. One cannot reject the hypothesis that the two processes are perfectly correlated, however. These means that both the United States and Canada will move into and out of recessions simultaneously. Booms and recessions have expected lengths of 25 quarters and 2.3 quarters, respectively. The values for high and low growth are close for both countries: around 1.2% and -3%. For comparison, the standard deviations of the error terms for the two countries are 1.1% and 1.2%. A glance at the ρ matrix shows that innovations in the U.S. growth rate series have an autocorrelation of 0.46. In addition, they have an autocorrelation with Canadian growth rates of around 0.34. This implies that while the two countries may move into and out of recessions together, a positive shock to U.S. growth seems to be transmitted to Canada in the following period.

For the United States and Germany, likelihood ratio tests reject the null of independent Markov processes as well as Germany leading the United States at 95% confidence. It is impossible to reject the nulls of either perfect correlation or the United States leading Germany. The high growth values for the United States are still approximately 1.3% and the low growth values are around -3%. High growth for Germany is between 1.0% and 1.4% per quarter, while low growth is between -1.5% and -1.7%. The standard

deviation of German innovations ranges between 1.5% and 1.8%. If the Markov processes are perfectly correlated the expected lengths of booms and recessions are 25 quarters and 2.3 quarters, very close to the estimates for the United States and Canada. If the United States leads, then Germany goes into and out of recessions the quarter immediately after the United States. Thus, booms for both countries have an expected length of 33 quarters and recessions a length of 2.9 quarters. The ρ matrix here shows that innovations for the United States are autocorrelated at around 0.5, while the effects of U.S. innovations on Germany and of Germany on either the United States or itself are small.

In the case of the United States with the United Kingdom, one can reject the nulls that either country leads the other, but cannot reject the nulls of independence or perfect correlation. High- and low-growth values for the United States are roughly the same as the two cases above. High growth for the United Kingdom ranges from 0.65% to 0.85% per quarter. Low growth varies widely depending on the restrictions. If the two countries are independent, then low growth for the United Kingdom is -7.5% per quarter and a 'recession' lasts only for a quarter. If the two countries are perfectly correlated, then low growth is -1.6% and we get booms and recessions with expected lengths similar to the two other cases. The standard deviation of U.K. innovations ranges between 1.7% and 1.9%. Innovations in the United States are autocorrelated with a coefficient around 0.5 (except under the assumption of independent processes where it is only 0.19). The value of ρ^{hf} is between 0.33 and 0.45, indicating that innovations in the United States will have a large influence on U.K. innovations next period. U.K. innovations have only a small autocorrelation, and small impact on U.S. innovations (again, except under the assumption of independent processes).

The evidence above on the transmission of business cycles is mixed. First, it appears from figs. 5–7 that there are three world-wide recessions: one in late 1974 associated with the first oil shock, one in early 1980, and one in late 1981 or early 1982. There is little evidence that the United States leads or follows into and out of these recessions; rather, all countries experienced these recessions and recoveries at the same time. Large swings in growth rates do not appear to begin in any single country and then gradually get transmitted to others. However, as can be seen by the large values of ρ^{hf} , at least in the case of Canada and the United Kingdom, shocks to U.S. growth this quarter are positively correlated with shocks to foreign countries next quarter. Since for all countries the variance of the error terms is small relative to the gap in boom and recession growth rates, this effect is only second-order. Thus, it appears that major movements in growth rates are caused by world-wide shocks and any transmission of business cycles by other mechanisms is small in comparison.

8. Conclusions

This paper evaluates different types of Markov-shifting processes relative to other each other and to other more common stochastic processes. The more standard linear ARMA and VAR models have better forecast properties, but the Markov-switching models have substantially better fit. In both cases it appears that adding information on a second country improves both the forecast and the fit of the model.

This paper has also addressed the issue of international transmission of business cycles. It appears that while recessions and booms do seem to occur simultaneously across countries, they are easily attributable to world-wide shocks and not the result of business cycle transmission from one country to another. Any such business cycles transmission that does occur is relatively minor. I should stress that the results of this paper apply only to the four countries considered. Any transmission that may occur between large industrial nations and less-developed countries or between large-small country pairs [see Phillips (1990)] is beyond the scope of this paper.

Appendix: Derivation of the likelihood ratio test

The purpose of this appendix is to show how the normal equations must be altered if certain types of restrictions are imposed. If the only parameters that have changed in the restricted models are the π_{ij} and they are now assumed to be functions of other underlying parameters which I will denote γ , and if these γ parameters are unrestricted, then the following analysis applies. This allows one to use the same basic filter as the unrestricted case and only change the calculations that update the γ parameters for each iteration of the EM algorithm.

First, consider the unrestricted estimation. One maximizes the following Lagrangian:

$$\mathcal{L} = L + \sum_{i=1}^4 \lambda_i \left\{ 1 - \sum_{j=1}^4 \pi_{ij} \right\}. \quad (18)$$

The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial \pi_{ij}} = \frac{\partial L}{\partial \pi_{ij}} - \lambda_i = 0. \quad (19)$$

Hamilton (1988) shows

$$\frac{\partial L}{\partial \pi_{ij}} = \frac{1}{\pi_{ij}} N_{ij}; \quad N_{ij} = \sum_{t=2}^T \Pr(s_t = j, s_{t-1} = i | y_T, \dots, y_1), \quad (20)$$

$$\lambda_i = D_i = \sum_{t=2}^T \Pr(s_{t-1} = i | y_T, \dots, y_1). \quad (21)$$

Now, consider the restricted case where the vector π is some function of the vector γ . The first-order condition for maximizing the Lagrangian with respect to the k th element of γ is

$$\frac{\partial \mathcal{L}}{\partial \gamma_k} = \sum_{\pi_{ij}} \frac{\partial \mathcal{L}}{\partial \pi_{ij}} \frac{\partial \pi_{ij}}{\partial \gamma_k} = \sum_{\pi_{ij}} \left\{ \frac{\partial L}{\partial \pi_{ij}} - \lambda_i \right\} \frac{\partial \pi_{ij}}{\partial \gamma_k} = 0. \quad (22)$$

Substituting in (20) and (21) gives:

$$\sum_{\pi_{ij}} \frac{1}{\pi_{ij}} N_{ij} \frac{\partial \pi_{ij}}{\partial \gamma_k} = \sum_{\pi_{ij}} D_i \frac{\partial \pi_{ij}}{\partial \gamma_k}. \quad (23)$$

One must now evaluate (23) on a case-by-case basis.

For the matrix in eq. (12), I will derive the estimate of π_{11}^h ; the other estimates are derived similarly.

Eq. (23) can be rewritten as

$$\begin{aligned} & \frac{\pi_{11}^f}{\pi_{11}^h \pi_{11}^f} N_{11} + \frac{-\pi_{11}^f}{(1-\pi_{11}^h) \pi_{11}^f} N_{12} + \frac{1-\pi_{11}^f}{\pi_{11}^h (1-\pi_{11}^f)} N_{13} + \frac{-(1-\pi_{11}^f)}{(1-\pi_{11}^h)(1-\pi_{11}^f)} N_{14} \\ & + \frac{1-\pi_{22}^f}{\pi_{11}^h (1-\pi_{22}^f)} N_{31} + \frac{-(1-\pi_{22}^f)}{(1-\pi_{11}^h)(1-\pi_{22}^f)} N_{32} + \frac{\pi_{22}^f}{\pi_{11}^h \pi_{22}^f} N_{33} + \frac{-\pi_{22}^f}{(1-\pi_{11}^h) \pi_{22}^f} N_{34} \\ & = \{ \pi_{11}^f - \pi_{11}^f + (1-\pi_{11}^f) - (1-\pi_{11}^f) \} D_1 + \{ (1-\pi_{22}^f) - (1-\pi_{22}^f) + \pi_{22}^f - \pi_{22}^f \} D_3. \end{aligned} \quad (24)$$

The right-hand side of (24) is zero and the equation reduces to

$$\frac{S_1}{\pi_{11}^h} - \frac{S_2}{(1-\pi_{11}^h)} = 0, \quad (25)$$

where $S_1 = (N_{11} + N_{13} + N_{31} + N_{33})$ and $S_2 = (N_{12} + N_{14} + N_{32} + N_{34})$. Solving for π_{11}^h and recalling (20) and (21) gives:

$$\pi_{11}^h = \frac{S_1}{S_1 + S_2}, \quad (26)$$

$$S_1 = \sum_{i=1,3} \sum_{j=1,3} \sum_{t=2}^T \Pr(s_t=j, s_{t-1}=i | y_T, \dots, y_1), \quad (27)$$

$$S_2 = \sum_{i=1,3} \sum_{j=2,4} \sum_{t=2}^T \Pr(s_t=j, s_{t-1}=i | y_T, \dots, y_1). \quad (28)$$

References

- Boehm, E.A. and G.H. Moore, 1984, New economic indicators for Australia, 1949–84, *Australian Economic Review* 68, 34–56.
- Cochrane, John H., 1988, How big is the random walk in GNP?, *Journal of Political Economy* 96, 893–920.
- Dempster, A.P., N.M. Laird and D.B. Rubin, 1977, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B* 39, 1–38.
- Engel, Charles, and James D. Hamilton, 1990, Long swings in the exchange rate: Are they in the data and do markets know it?, *American Economic Review* 80, 689–713.
- Hamilton, James D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357–384.
- Hamilton, James D., 1991, Analysis of time series subject to changes in regime, *Journal of Econometrics* 45, 39–70.
- King, Robert J., Charles I. Plosser, James Stock and Mark Watson, 1987, Stochastic trends and economic fluctuations, Rochester Center for Economic Research, Working Paper 79.
- Layton Allan P., 1987, Australian and U.S. growth cycle linkages, 1967–1983, *Journal of Macroeconomics* 9, 31–44.
- Nelson, Charles R., and Charles I. Plosser, 1982, Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics* 10, 139–162.
- Phillips, Kerk L., 1989, Full-sample versus long-lag inferences in Markov-switching time series models, University of Rochester, manuscript.
- Phillips, Kerk L., 1990, Business cycle transmission in large–small country pairs, University of Rochester, manuscript.