A Quasi-Bayesian Analysis of Structural Breaks: China’s Output and Productivity Series

Xiao-Ming Li∗

Department of Commerce, Massey University—Albany, New Zealand

Abstract

A quasi-Bayesian model selection approach is employed to detect the number and dates of structural changes in China’s GDP and labour productivity data. It is shown that the predictive likelihood information criterion is valid only among models with well-behaved residuals.

Key words: structural change; predictive likelihood; GDP; labour productivity; China

JEL classification: C22; E30

1. Introduction

The literature on structural change in macroeconomic time series has been growing rapidly over the past decade. From a methodological point of view, such studies may be classified into two categories: one that adopts a hypothesis testing approach, and the other that takes a model selection approach (see Maddala and Kim, 1998, p. 406). Regarding hypothesis testing, there are two broad avenues to view. Tests of the unit root null hypothesis with or without structural change against the trend-stationarity alternative that allows for one or more breaks in the trend function fall into the first. Contributions in this area include Perron (1989), Inwood and Stengos (1991), Zivot and Andrews (1992), Raj and Slottje (1994), Perron (1997), Nunes et al. (1997), Lumsdaine and Papell (1997), Kanas (1998), Ben-David et al. (1999), and Newbold et al. (2001). The second line involves testing for alternative hypotheses against each other with respect to the number of breakpoints in the series, disregarding whether or not a unit root is present. Several testing techniques have recently been developed by Vogelsang (1997), Bai and Perron (1998), and Bai (1999) and applied in Ben-David and Papell (1998) and Ben-David and Papell (2000).

The problem of choosing between difference-stationary models and trend-stationary models that account for structural change and between models with

Received April 17, 2003, revised February 23, 2004, accepted March 8, 2004.

∗Correspondence to: Department of Commerce, Massey University—Albany, Private Bag 102904, North Shore Mail Centre, Auckland, New Zealand. Email: x.n.li@massey.ac.nz. The author thanks an anonymous Associate Editor and the Managing Editor for very useful comments and suggestions that helped improve the quality and presentation of this paper.
different numbers of breaks can also be tackled by a model selection approach. In fact, it should properly be so, as Maddala and Kim (1998) argue that ultimately determination of the number of breaks is best viewed as a model selection problem. However, the structural-change literature with model selection does not seem to be as vast as with hypothesis testing. Kim and Maddala (1991) propose a Bayesian approach which uses the BIC criterion for selecting the best model among the models with different numbers of breaks. A recent paper by Wang and Zivot (2000) also employs similar methods to select the most appropriate model from the data. While the above two studies may be categorised as following a Bayesian approach, Kashiwagi (1991) suggests what Maddala and Kim (1996a) term a “quasi-Bayesian” approach that uses the predictive likelihood in analysing the same problem of detecting structural changes. Bianchi (1995) extends this quasi-Bayesian approach to the case of autoregressive errors, and Bianchi and Kan (1996) apply the extended method in a study of the trend behaviour of the US GNP series. For differences between the Bayesian and quasi-Bayesian procedures, see Maddala and Kim (1996b).

More work needs to be done to compare the above two independently developed procedures for selecting the best model among alternatives with different numbers of breaks (Maddala and Kim, 1996b). This paper is part of the empirical work for that purpose. It reports the results of applying the quasi-Bayesian approach to detecting the structural breaks in China’s GDP and labour productivity data, which may be used later for comparisons with the results of applying the Bayesian approach. Section 2 briefly describes the approach employed, Section 3 presents empirical results, and Section 4 concludes.

2. Predictive Likelihood

Consider a time series model with multiple breaks as follows:

\[ y_t = a + bt + \sum_{i=1}^{n} a_i DU_{it} + \sum_{i=1}^{n} b_i DT_{it} + \epsilon_t. \]  

(1)

In this model, \( y_t \) is a time series under investigation, \( t \) represents time trend, \( n \) is the number of breakpoints, \( \epsilon_t \) denotes the error term, and \( a, b, a_i, \) and \( b_i \) are model parameters to be estimated. The break dummy variables, \( DU_{it} \) and \( DT_{it} \), take the following values:

\[ DU_{it} = \begin{cases} 1 & \text{if } t \geq TB_i \\ 0 & \text{otherwise} \end{cases} \]

\[ DT_{it} = \begin{cases} t - TB_i + 1 & \text{if } t \geq TB_i \\ 0 & \text{otherwise} \end{cases} \]

with the ranges of \( n \) break dates being:

\[ 1 < TB_1 < \ldots < TB_i < \ldots < TB_n < T. \]

Here \( T \) denotes the effective sample size and \( TB_1, \ldots, TB_n \) are \( n \) break dates. We as-
sume that the errors $e_t$ follow an AR($p$) process with the order $p$ to be determined.

The breakpoints (including their number and timing) are to be searched for with the aid of the predictive likelihood information criterion as briefly described below.

If $p = 0$, the errors are serially independent and so the predictive likelihood criterion suggested by Kashiwagi (1991) is relevant, and its value, $PL$, is computed as:

$$PL = -\frac{T}{2} - \ln(T) \left( \frac{T}{2} \right) \ln(2\pi\sigma^2) - \frac{T}{2} \ln(2\pi\Omega_p),$$

where $\sigma^2$ is the variance of $e_t$ and $k$ is defined as $k = 2(n+1) - m_1$. Here $m_1$ is the number of the periods that have only one observation.

If $p > 0$, the errors $e_t$ become an autoregressive process of order $p$:

$$e_t = \mu_1 + \mu_2 e_{t-1} + \ldots + \mu_p e_{t-p} + \epsilon_t \sim N(0, \sigma^2 p),$$

where $\mu_1, \ldots, \mu_p$ are parameters and $\epsilon_t$ represents a normally, independently and identically distributed (NIID) error term. In this case, a heuristic criterion proposed by Bianchi and Kan (1996) may be employed. It is expressed as:

$$PL = -\frac{T}{2} - \ln(T) \left( \frac{T}{2} \right) \ln(2\pi\sigma^2) - \frac{T}{2} \ln(2\pi\Omega_p),$$

with $k = 2(n+1) - m_1 + p$. The idea behind adding $p$ to the expression for $k$ is to penalise for the number of autoregressive terms in the errors.

Equation (1) is estimated with the ordinary least square (OLS) method when $p$ is assumed to be 0 and with the maximum likelihood method when $p$ is assumed to be greater than 0. In the former case, use (2) to calculate $PL$. In the latter case, simply deducting from the estimated value of the log-likelihood function the bias correction term $T(k+1)/(T-k-2)$ will give $PL$.

3. Empirical Results


Tables 1 to 4 display the models with the number of breaks associated with the highest $PL$ values under four different assumptions about the error terms. They also report diagnostic statistics about the behaviour of the error terms. In particular, both the Ljung-Box $Q$ and the Breusch-Godfrey ($B-G$) statistics are used to check for the presence of serial correlation. Since the former is valid only for a pure autoregressive process while the latter is also valid in the case where such variables as trends and dummies appear as regressors, we mainly rely on the latter for inference.
If only the predictive likelihood criterion is used, one would certainly pick out the models in Table 1 and conclude that GDP has 10 structural breaks and labour productivity has 9. However, although the $Q$ statistics suggest that the assumption of independent errors cannot be rejected at the 5% level, the $B$-$G$ statistics indicate a decisive rejection at a lower than 1% level for the two series. In addition, the productivity data suffer the problem of heteroscedasticity. Thus, the models that assume $e_t$ to be an NIID error are not valid, and the two numbers of breakpoints suggested by the highest $PL$ values may well be overestimated.

<table>
<thead>
<tr>
<th>Series</th>
<th>n</th>
<th>PL</th>
<th>$Q(10)$</th>
<th>$B-G(10)$</th>
<th>Normality</th>
<th>ARCH(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>10</td>
<td>91.860</td>
<td>16.358 [0.090]</td>
<td>28.668 [0.001]</td>
<td>0.233 [0.890]</td>
<td>5.817 [0.121]</td>
</tr>
<tr>
<td>Productivity</td>
<td>9</td>
<td>99.283</td>
<td>18.248 [0.051]</td>
<td>26.655 [0.003]</td>
<td>1.396 [0.497]</td>
<td>8.631 [0.035]</td>
</tr>
</tbody>
</table>


The models with AR(1) and AR(3) specifications for $e_t$ do not qualify for a clean bill of health either. Tables 2 and 4 show that both the $Q$ and $B$-$G$ statistics signal serial correlation in $e_t$, although normality and homoscedasticity are present. The only valid models seem to be those given in Table 3: incorporating the AR(2)
assumption for \( e_t \) now makes \( e_t \) a NIID process, as suggested by all the reported diagnostic test statistics. Accordingly, 90.681 and 91.963 are the only valid highest PL values for the GDP and productivity data respectively, and the associated numbers of breaks are 5 for the former and 4 for the latter. Table 5 presents more details about the PL values and break dates for the models with AR(2) error terms, and the two columns with bold numbers represent the two best models among them for the two series respectively.

**Table 5. The PL Values and Break Dates When the Residuals Are Assumed to Be AR(2)**

<table>
<thead>
<tr>
<th></th>
<th>( n = 0 )</th>
<th>( n = 1 )</th>
<th>( n = 2 )</th>
<th>( n = 3 )</th>
<th>( n = 4 )</th>
<th>( n = 5 )</th>
<th>( n = 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL</td>
<td>54.234</td>
<td>67.767</td>
<td>80.099</td>
<td>88.880</td>
<td>89.892</td>
<td><strong>90.681</strong></td>
<td>90.372</td>
</tr>
<tr>
<td>GDP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Productivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL</td>
<td>55.154</td>
<td>67.447</td>
<td>77.411</td>
<td>86.156</td>
<td><strong>91.963</strong></td>
<td>91.290</td>
<td>90.564</td>
</tr>
</tbody>
</table>

It is interesting to compare our results with those in Bianchi and Kan (1996). They show that the predictive likelihood approach is robust to different assumptions about the error terms \( e_t \), as the number of structural breaks turns out to be the same \( (n = 2) \), and the parameter estimates for different error specifications are not very different. However, this conclusion does not apply in our study here, since we find that the number and the dates of breaks are sensitive to different error processes assumed. To further verify whether this is indeed the case, we conducted a small Monte Carlo study. More specifically, we generated a random variable \( y_t \) using equation (1) but assuming 2 breaks (i.e., \( DU_{1t} \), \( DU_{2t} \), \( DT_{1t} \), and \( DT_{2t} \)) with their dates and the values of all coefficients being given. The error term \( e_t \) was generated following an AR(1) process with the autoregressive coefficient predetermined. When the (correct) AR(1) specification is used for the error structure, the computed highest PL value equals 74.375 which corresponds to the two true breakpoints. However, if the (incorrect) AR(0) process was assumed for the error term, we obtain a higher PL value of 78.260 associated with 4 breaks instead of 2; that is, we end up with an over-inflated number of breakpoints. This suggests that the validity of the PL model selection approach is conditional on the correct specification of the error structure, at least in this application. Merely accepting the highest PL value among all models with different structural changes and varying error processes could therefore result in overestimating or underestimating the number of breakpoints and possibly in misidentifying their locations.

The best models we have selected, i.e., the models that have the highest PL values and well-behaved residuals \( \varepsilon_t \), yield the following maximum likelihood estimates:
Based on equations (5) and (6), Figures 1 and 2 depict the long-run growth paths of China’s GDP and labour productivity. The growth rates of GDP during the six break periods can be calculated using the point estimates of the slopes given in (5). They are: 15.6% (1952–1956), 6.0% (1957–1960), 6.0% (1961–1965; note that the coefficient of $DT_{61}$ in (5) is insignificant), 9.0% (1966–1974), $-2.7\%$ (1975–1976), and 11.3% (1977–2000). By the same calculation method and using the relevant information provided in equation (6), we obtain the growth rates of labour productivity for the five break periods as follows: 8.6% (1952–1955), 1.1% (1956–1960), 4.2% (1961–1975), 6.5% (1976–1989), and 8.6% (1990–2000).

**Figure 1. The Long-Run Growth Path of Logarithmic Output**
These changes in long-run growth rates can all be related to particular historical events, but there are three notably surprising results to point out here. First, the Cultural Revolution (1966–1976) did not have a negative impact on labour productivity growth but had a positive growth effect on GDP, contrary to what people normally believe. Second, it is since the end of the Cultural Revolution in 1976, and not since the inauguration of economic reforms in 1979, that both GDP and labour productivity have experienced growth takeoff. That is, economic reforms did not impact the long-run growth paths of these two series, again a seemingly counter-intuitive result. Third, 1990 is the first year of implementing a drastic program of macroeconomic austerity and is also the immediately subsequent year of the 1989 political crisis (the “June 4th event”). However, how these unfavourable shocks should cause a rise in the long-run growth rate of labour productivity is a mystery. Clearly, the currently adopted approach is unable to solve these puzzles.

4. Conclusions

In this paper, a quasi-Bayesian model selection approach developed by Kashwagi (1991) and extended by Bianchi and Kan (1996) is adopted to detect the number and timing of structural changes in China’s GDP and labour productivity data. We have shown that one cannot mechanically or blindly use the highest value of the predictive likelihood information criterion to pick out the best model among models with and without well-behaved residuals: the criterion is valid only when the white-noise assumption about the error terms is satisfied. In our case, only within the class of models with AR(2) error terms can we compare the $PL$ values and use validly the highest one to determine the number and dates of breakpoints.
The empirical results suggest 5 breakpoints for GDP in 1957, 1961, 1966, 1975, and 1977 and 4 breakpoints for labour productivity in 1956, 1961, 1976, and 1990. These breakpoints divide the sample into different growth periods. While most of the estimated growth results are expected, some changed/unchanged long-run growth rates do not seem to agree with common sense. Accordingly, further research to adopt other model selection approaches such as the Bayesian one is needed to confirm these new findings or check their validity.

References


