A Comparison of Alternative Asymptotic Frameworks to Analyze a Structural Change in a Linear Time Trend

Ai Deng[*] **Pierre Perron**[†]

Boston University **Boston University**

This version: August 26, 2005

Abstract

This paper considers various asymptotic approximations to the finite sample distribution of the estimate of the break date in a simple one-break model for a linear trend function that exhibits a change in slope, with or without a concurrent change in intercept. The noise component is either stationary or has an autoregressive unit root. Our main focus is on comparing the so-called "bounded-trend" and "unbounded-trend" asymptotic frameworks. Not surprisingly, the "bounded-trend" asymptotic framework is of little use when the noise component is integrated. When the noise component is stationary, we obtain the following results. If the intercept does not change and is not allowed to change in the estimation, both frameworks yield the same approximation. However, when the intercept is allowed to change, whether or not it actually changes in the data, the "bounded-trend" asymptotic framework completely misses important features of the finite sample distribution of the estimate of the break date, especially the pronounced bimodality that was uncovered by Perron and Zhu (2005) and shown to be well captured using the "unbounded-trend" asymptotic framework. Simulation experiments confirm our theoretical findings, which expose the drawbacks of using the "bounded-trend" asymptotic framework in the context of structural change models.

Keywords: change-point, confidence intervals, shrinking shifts, bounded trend, level shift.

[*]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (dengai@bu.edu).
[†]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

1 Introduction

Estimating and forming confidence intervals for break dates in structural change models is an important issue of practical interest (see, e.g., Perron, 2005, for an extensive review). To obtain confidence intervals, the most common approach is to use the asymptotic distribution of the estimated break dates, though other approaches are possible such as methods based on inverting a test statistic (see, e.g., Siegmund, 1988). In the context of models with stationary regressors, it is well known since the work of Hinkley (1970) that the limit distribution of the MLE (or other estimates) depends on the finite sample distribution of the errors. A solution to this problem is to consider an asymptotic framework whereby the magnitude of the change shrinks at a suitable rate as the sample size increases, in which case the limit distribution is invariant to the finite sample distribution of the errors (see, e.g., Yao, 1987, and Picard, 1985). For comprehensive results in linear regression models, see Bai (1997) for a single break, and Bai and Perron (1998) for the multiple break case. In general, the shrinking shift asymptotic framework provides a reliable guide unless the magnitude of the change is very small, in which case the confidence intervals are liberal, or very large, in which case they are conservative (see, e.g., Bai and Perron, 2005).

In the context of regression models with trending regressors, there is an additional layer of complexity related to the way in which the trend is specified. Consider the case where the regression function is a simple linear trend. One approach is simply to specify the regression function as, say, $\mu + \beta t$, as done in Perron and Zhu (2005), henceforth referred to as PZ, and Bai, Lumsdaine and Stock (1998). A more common approach is to specify it in the form $\mu + \beta(t/T)$, so that the trend is bounded and indeed restricted to be in the interval $[0, 1]$, see, e.g., Bai (1997). This is a special case of a "bounded-trend" asymptotic framework analyzed by Andrews and McDermott (1995). A common argument is that the "bounded-trend" asymptotic framework yields more tractable results, but little is known about which approach delivers a better approximation to the distribution of the estimates of the break dates and the parameters of the trend function.

The aim of this paper is to inquire about which asymptotic framework provides the most reliable approximation to the finite sample distribution of the estimate of the break date in a simple one-break model for a linear trend function of the form

$$y_t = \mu + \beta t + u_t$$

where y_t is a scalar observable variable and u_t is a noise component that can be stationary, $I(0)$, or have an autoregressive unit root, $I(1)$. While a simple model, it is nevertheless an important one in practice. For example, it is often of interest to assess whether the rate of growth of a series such as real GDP exhibits a change and if so, to form a confidence interval for the break date. Allowing the errors to be $I(0)$ or $I(1)$ permits the series to be trend- or difference-stationary, see e.g., Nelson and Plosser (1982) and Perron (1989).

The main findings are the following. First, if the noise function u_t is $I(1)$, the "bounded-trend" asymptotic framework is of little use. This is not surprising since with a "bounded-trend" specification, the noise then dominates the signal. Of more interest are the results pertaining to the case with a stationary noise function. If the intercept μ does not change concurrently with the slope β and the fitted intercept is not allowed to change, then all frameworks are basically equivalent. However, if the fitted intercept is allowed to change, whether or not μ changes, the "bounded-trend" asymptotic framework completely misses important features of the finite sample distribution of the estimate of the break date, especially the pronounced bimodality that was uncovered by PZ and shown to be well captured using the alternative "unbounded-trend" asymptotic framework.

The rest of the paper is organized as follows. Section 2 presents the various models considered and summarizes the relevant results from PZ pertaining to the "unbounded-trend" asymptotic framework. Section 3 provides results about limit distributions using the "bounded-trend" asymptotic framework, with fixed and shrinking shifts. Section 4 presents the results from simulation experiments aimed to assess the quality of the approximation to the finite sample distribution of the various asymptotic frameworks. Section 5 offers brief concluding remarks and an appendix some technical derivations.

2 The Models

Throughout, it is assumed that some variable of interest, y_t, is the sum of some systematic part d_t and a random component, u_t, i.e.

$$y_t = d_t + u_t.$$

The models analyzed differ according to the assumptions made about both components d_t and u_t. For the random component u_t, we specify $E(u_t) = 0$ and alternatively one of the following two assumptions:

- Assumption 1: $u_t \sim I(0)$. More specifically u_t is such that $T^{-1/2} \sum_{t=1}^{[Tr]} u_t \Rightarrow \sigma W(r)$ where $\sigma^2 = \lim_{T\to\infty} T^{-1} E(\sum_{t=1}^{T} u_t)^2$ exists and is strictly positive. Here "$\Rightarrow$" denotes weak convergence in distribution (under the sup metric) and $W(r)$ is the unit Wiener process.

- Assumption 2: $u_t \sim I(1)$. More specifically $u_t = \sum_{j=1}^{t} \varepsilon_j$ where the sequence ε_t is assumed to be $I(0)$ as defined in Assumption 1.

Remark 1 *There are many sets of sufficient conditions to guarantee that the weak convergence result stated in Assumption 1 holds. One that is fairly general is that used in Phillips and Perron (1988), namely* (a) $\sup_t E\left|u_t\right|^{\gamma+\eta} < \infty$ *for some* $\gamma > 2$ *and* $\eta > 0$ *and* (b) $\{u_t\}_1^\infty$ *is strong mixing with mixing numbers* α_m *that satisfy* $\sum_1^\infty \alpha_m^{1-2/\gamma} < \infty$*. Alternatively, we can assume that* u_t *is a linear process such that* $u_t = \sum_{i=0}^\infty c_i e_{t-i}$ *where* $\{e_t, \mathcal{F}_{t-1}\}$ *is a martingale difference sequence with* $\mathcal{F}_{t-1}$ *the filtration to which* e_t *is adapted. Also* $\sum_{i=0}^\infty i|c_i| < \infty$ *(see Phillips and Solo, 1992). Either sets of conditions include the popular stationary and invertible ARMA processes.*

For the systematic component d_t, we consider 3 cases. As a matter of notation, the time of break is denoted T_1 and the break fraction as $\lambda = T_1/T$.

• **Models I.a and I.b: Joint broken trend with $I(1)$ or $I(0)$ errors.** Here, d_t is a first-order linear trend with a one time change in slope such that the trend function is joined at the time of break, specified by:

$$d_t = \mu_1 + \beta_1 t + \beta_b B_t \tag{1}$$

where B_t is a dummy variable for the slope change defined by

$$B_t = \begin{cases} 0, & \text{if } t \leq T_1 \\ t - T_1, & \text{if } t > T_1 \end{cases}$$

Here, the slope coefficient changes from β_1 to $\beta_1 + \beta_b$ at the time point T_1. However, the trend function is continuous at T_1. For this reason, this specification is referred to as a "joint broken trend".

• **Models II.a and II.b: Local disjoint broken trend with $I(1)$ or $I(0)$ errors.** Here, d_t is a first-order linear trend with a one time change in intercept and slope which is such that in the absence of an intercept change, the trend function is joined at the time of break, i.e.,

$$d_t = \mu_1 + \beta_1 t + \mu_b C_t + \beta_b B_t \tag{2}$$

where C_t is a dummy variable for the level shift defined by

$$C_t = \begin{cases} 0, & \text{if } t \leq T_1 \\ 1, & \text{if } t > T_1 \end{cases}$$

Note that μ_b and β_b capture the change in the intercept and slope coefficients. At the break point T_1, the slope changes by β_b and the level shifts by μ_b, which is negligible compared to the level of the series $\mu_1 + \beta_1 T_1$, hence the label "local disjoint segmented trend".

• **Models III.a and III.b: Global Disjoint broken trend with $I(1)$ or $I(0)$ errors.** The third specification is similar except that the trend function is not restricted to be joined at the time of break (in the absence of a change in intercept). If one wants to model a permanent shift in the level of the series such that the trend function is discontinuous at the break date even asymptotically, we can specify the DGP as

$$d_t = \mu_1 + \beta_1 t + \mu_b C_t + \beta_b B_t^{dj} \tag{3}$$

where

$$B_t^{dj} = \begin{cases} 0, & \text{if } t \leq T_1 \\ t, & \text{if } t > T_1 \end{cases}$$

We label this model as a "global disjoint segmented trend" since, in contrast to the previous "local disjoint segmented trend", the implied (relative to the overall level of the trend function) level shift at the break date converges to $\beta_b/\beta_1 \neq 0$ as $T \to \infty$, since $d_{T_1+1} - d_{T_1} = \beta_1 + \mu_b + \beta_b T_1$. Note that Model III is the one used by Bai, Lumsdaine and Stock (1998) with shrinking shifts.

Hence, we have six different models labelled as follows: I.a) joint broken trend with $I(1)$ errors; I.b) joint broken trend with $I(0)$ errors; II.a) local disjoint broken trend with $I(1)$ errors; II.b) local disjoint broken trend with $I(0)$ errors; III.a) global disjoint broken trend with $I(1)$ errors; III.b) global disjoint broken trend with $I(0)$ errors. Note that, in empirical applications, using Model II or III yields exactly the same results for the estimates of the parameters T_1, μ_1, β_1 and β_b. Nevertheless, the two specifications yield drastically different asymptotic results, in particular pertaining to the rate of convergence and the asymptotic distribution of the estimated break date. Limiting results obtained from Model II (local disjoint trend) provide good approximations to the finite sample distributions when the shift in level is small while those from Model III will do so when the shift in level is large. Hence, both asymptotic frameworks are complementary.

All specifications discussed can be expressed in matrix notation as

$$Y = X_{T_1}\gamma + U$$

where $Y' = [y_1, ..., y_T]$, $U' = [u_1, ..., u_T]$, $X'_{T_1} = [x(T_1)_1, ..., x(T_1)_T]$, $\gamma' = (\mu_1, \beta_1, \mu_b, \beta_b)$ and where, for Models I, $x(T_1)'_t = [1, t, B_t]$, for Models II, $x(T_1)'_t = [1, t, C_t, B_t]$, and for Models

III, $x(T_1)'_t = [1, t, C_t, B_t^{dj}]$. Note that the matrix X_{T_1} depends on the postulated value of the break date T_1. Since the parameters are assumed to be obtained using a global least-squares criterion, we have the following estimate for the break date

$$\hat{T}_1 = \arg\min_{T_1} Y' \left(1 - P_{T_1}\right) Y$$

where P_{T_1} is the projection matrix constructed using X_{T_1}, i.e., $P_{T_1} = X_{T_1} \left(X'_{T_1} X_{T_1}\right)^{-1} X'_{T_1}$. Denoting by $X_{\hat{T}_1}$ the matrix X constructed using the least-squares estimate of the break date $\hat{T}_1$, the least-squares estimate of the coefficients γ is:

$$\hat{\gamma} = (X'_{\hat{T}_1} X_{\hat{T}_1})^{-1} X'_{\hat{T}_1} Y$$

and the resulting Sum of Squared Residuals is, for an estimated break fraction $\hat{\lambda} = \hat{T}_1/T$,

$$SSR(\hat{\lambda}) = \sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} (y_t - x(\hat{T}_1)'_t \hat{\gamma})^2 = Y'(I - P_{\hat{T}_1})Y$$

where $P_{\hat{T}_1}$ is the projection matrix associated with $X_{\hat{T}_1}$, i.e. $P_{\hat{T}_1} = X_{\hat{T}_1}(X'_{\hat{T}_1} X_{\hat{T}_1})^{-1} X'_{\hat{T}_1}$.

The true values of the unknown coefficients will be denoted with a 0 superscript, i.e. $\gamma^0 = (\mu_1^0, \beta_1^0, \mu_b^0, \beta_b^0)'$, T_1^0, $\lambda^0 = T_1^0/T$; $X_{T_1^0}$ is the matrix of regressors constructed using the true value T_1^0 for the break date, and $P_{T_1^0}$ is the associated projection matrix, i.e. $P_{T_1^0} = X_{T_1^0}(X'_{T_1^0} X_{T_1^0})^{-1} X'_{T_1^0}$. So the true data generating process is assumed to be

$$Y = X_{T_1^0} \gamma^0 + U.$$

Results pertaining to the limit distribution of $\hat{T}_1$ for each cases are reproduced below from PZ. We assume throughout that $\beta_b^0 \neq 0$ and $\lambda^0 \in (0, 1)$.

Theorem 1 *1. In Model I.a, $\sqrt{T}(\hat{\lambda} - \lambda) \to^d N(0, 2\sigma^2/(15(\beta_b^0)^2))$;*

2. In Model I.b, $T^{3/2}(\hat{\lambda} - \lambda) \to^d N(0, 4\sigma^2/[\lambda^0 \left(1 - \lambda^0\right) (\beta_b^0)^2])$;

3. For Model II.a, define $\xi_1 \equiv [\int_0^1 W(r)\,dr,\ \int_0^1 rW(r)\,dr,\ \int_{\lambda^0}^1 W(r)\,dr,\ \int_{\lambda^0}^1 \left(r - \lambda^0\right) W(r)\,dr]'$, $\xi_2 = [0, 0, W\left(\lambda^0\right), \int_{\lambda^0}^1 W(r)\,dr]'$, $\xi_3 \equiv \int_0^{\lambda^0} [(3r^2 - 2r\lambda^0)/\left(\lambda^0\right)^2] dW(r)$, $\xi_4 \equiv \int_{\lambda^0}^1 [(r-1)\left(3r - 2\lambda^0 - 1\right) / \left(1 - \lambda^0\right)^2] dW(r)$,

$$\Omega_1 \equiv \begin{bmatrix} \frac{4}{\lambda^0} & -\frac{6}{\left(\lambda^0\right)^2} & \frac{2}{\lambda^0} & \frac{6}{\left(\lambda^0\right)^2} \\ -\frac{6}{\left(\lambda^0\right)^2} & \frac{12}{\left(\lambda^0\right)^3} & -\frac{6}{\left(\lambda^0\right)^2} & -\frac{12}{\left(\lambda^0\right)^3} \\ \frac{2}{\lambda^0} & -\frac{6}{\left(\lambda^0\right)^2} & \frac{4}{\lambda^0\left(1-\lambda^0\right)} & 6\frac{1-2\lambda^0}{\left(\lambda^0\right)^2\left(1-\lambda^0\right)^2} \\ \frac{6}{\left(\lambda^0\right)^2} & -\frac{12}{\left(\lambda^0\right)^3} & 6\frac{1-2\lambda^0}{\left(\lambda^0\right)^2\left(1-\lambda^0\right)^2} & 12\frac{3\left(\lambda^0\right)^2-3\lambda^0+1}{\left(\lambda^0\right)^3\left(1-\lambda^0\right)^3} \end{bmatrix}$$

$$\Omega_2 \equiv \begin{bmatrix} -\frac{4}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & -\frac{2}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} \\ \frac{12}{(\lambda^0)^3} & -\frac{36}{(\lambda^0)^4} & \frac{12}{(\lambda^0)^3} & \frac{36}{(\lambda^0)^4} \\ -\frac{2}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & 4\frac{2\lambda^0-1}{(\lambda^0)^2(1-\lambda^0)^2} & \frac{12}{(\lambda^0)^3}\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0-1)^3} \\ -\frac{12}{(\lambda^0)^3} & \frac{36}{(\lambda^0)^4} & \frac{12}{(\lambda^0)^3}\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0-1)^3} & \frac{36}{(\lambda^0)^4}\frac{4(\lambda^0)^3-6(\lambda^0)^2+4\lambda^0-1}{(\lambda^0-1)^4} \end{bmatrix}$$

Also define $Z(m)$ as follows: $Z(0) = 0$, $Z(m) = Z_1(m)$ for $m < 0$ and $Z(m) = Z_2(m)$ for $m > 0$, with

$$\begin{aligned} Z_1(m) &= \left(\beta_b^0\right)^2 |m|^3/3 + m^2\sigma\beta_b^0\xi_4 + m\sigma^2\left[2\xi_2'\Omega_1\xi_1 - \xi_1'\Omega_2\xi_1\right], \ m < 0 \\ Z_2(m) &= \left(\beta_b^0\right)^2 |m|^3/3 + m^2\sigma\beta_b^0\xi_3 + m\sigma^2\left[2\xi_2'\Omega_1\xi_1 - \xi_1'\Omega_2\xi_1\right], \ m > 0 \end{aligned}$$

Then, $\sqrt{T}(\hat{\lambda} - \lambda) \to^d \arg\min_m Z(m)$;

4. *For Model II.b, define a stochastic process $S(m)$ on the set of integers as follows: $S(0) = 0$, $S(m) = S_1(m)$ for $m < 0$ and $S(m) = S_2(m)$ for $m > 0$, with*

$$\begin{aligned} S_1(m) &= \sum_{k=m+1}^{0}\left(\mu_b^0 + \beta_b^0 k\right)^2 - 2\sum_{k=m+1}^{0}\left(\mu_b^0 + \beta_b^0 k\right)u_k, \qquad m = -1, -2, ... \\ S_2(m) &= \sum_{k=1}^{m}\left(\mu_b^0 + \beta_b^0 k\right)^2 + 2\sum_{k=1}^{m}\left(\mu_b^0 + \beta_b^0 k\right)u_k, \qquad m = 1, 2, ... \end{aligned}$$

If $\{u_t\}$ is strictly stationary with a continuous distribution, S^* is a two-sided random walk with drift, and $T(\hat{\lambda} - \lambda) \to^d \arg\min_m S(m)$.

5. *In models III.a and III.b, $|\hat{\lambda} - \lambda^0| = o_p(T^{-3})$.*

For a detailed discussion of these results, see PZ. However, some important features need to be stressed to understand some comparisons to be made later. First, for Models I.a and I.b, the limiting distributions of the break date do no depend on the structure of the errors and remain the same irrespective of the nature of the serial correlation. This is in stark contrast to results obtained in a stationary context in which case the limiting distribution of the estimated break date, in this fixed shift case, not only depends on the properties of the residuals but in particular on their exact distribution (see, e.g., Bai, 1997). With shrinking shifts, the exact distribution of the errors is no longer present but the nature of the serial correlation still affects the limit distribution. For Model I with an "unbounded-trend" asymptotic framework, there is no need to resort to shrinking shifts asymptotic approximations. As expected, the rate of convergence is higher with $I(0)$ errors. The limit distributions depend on the exact nature of the errors only for Model II with $I(0)$ errors. Most importantly, comparing the results for Models I and II (with either I(1) or I(0) errors), the level shift plays an important role in the limiting distribution of the estimated break date. Suppose that the data generating process specifies no level shift, i.e. $\mu_b^0 = 0$. In Model I, no level shift is allowed in the regression while in Model II it is allowed via the regressor C_t. The results show that introducing such an irrelevant regressor changes the rate of convergence of the estimated break date and its asymptotic distribution, which shows strong bimodality as shown in PZ.

3 Bounded-trend Asymptotic Results.

To describe the bounded trend asymptotic framework, we follow the treatment of Andrews and McDermott (1995). Let T^* denote the actual size of the sample available. The idea is to embed the series of interest $(y_1, ..., y_{T^*})$ in a triangular array of random variables such that the shape of the trend function is mimicked in the limit but its magnitude remains bounded. With the triangular array denoted $\{y_{T,1}, ..., y_{T,T}\}$ with $y_{T,t} = d_{T,t} + u_t$, the number of rows is T and it is this value that is increased to infinity when doing the asymptotic. The embedding is achieved by specifying a trend functions of the form $d_{T,t} = d_{tT^*/T}$ so that $d_{T^*,t} = d_t$ for all $t \leq T$. In our cases, this leads to trend function of the following forms. For Models I,

$$d_t = \mu_1 + \beta_1 \frac{t}{T} T^* + \beta_b \frac{B_t}{T} T^*,$$

for Models II:

$$d_t = \mu_1 + \beta_1 \frac{t}{T} T^* + \mu_b C_t + \beta_b \frac{B_t}{T} T^*,$$

and for Models III:

$$d_t = \mu_1 + \beta_1 \frac{t}{T} T^* + \mu_b C_t + \beta_b \frac{B_t^{dj}}{T} T^*.$$

The following Theorem, proved in the appendix states the limit distributions of the estimate $\hat{T}_1 \equiv T\hat{\lambda}$.

Theorem 2 *1. For Model I.a,*

$$\hat{\lambda} \rightarrow^d \arg\max_{\lambda \in \Lambda} \begin{pmatrix} \int_0^1 W(r)\,dr & \int_0^1 rW(r)\,dr & \int_{\lambda^0}^1 \left(r - \lambda^0\right) W(r)\,dr \end{pmatrix}$$

$$\times \begin{pmatrix} 1 & \frac{1}{2} & \frac{(1-\lambda)^2}{2} \\ \frac{1}{2} & \frac{1}{3} & \frac{(1-\lambda)^2(\lambda+2)}{6} \\ \frac{(1-\lambda)^2}{2} & \frac{(1-\lambda)^2(\lambda+2)}{6} & \frac{(1-\lambda)^3}{3} \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 W(r)\,dr \\ \int_0^1 rW(r)\,dr \\ \int_{\lambda^0}^1 \left(r - \lambda^0\right) W(r)\,dr \end{pmatrix}$$

2. *For Model I.b,* $\sqrt{T}(\hat{\lambda}-\lambda^0) \to^d N(0, 4\sigma^2/[\lambda^0\left(1-\lambda^0\right)\left(T^*\beta_b^0\right)^2])$.

3. *For Model II.a,*

$$\hat{\lambda} \;\to\; ^d \arg\max_{\lambda\in\Lambda} \begin{pmatrix} \int_0^1 W(r)\,dr & \int_0^1 rW(r)\,dr & \int_{\lambda^0}^1 W(r)\,dr & \int_{\lambda^0}^1 \left(r-\lambda^0\right) W(r)\,dr \end{pmatrix}$$
$$\times \begin{pmatrix} \frac{4}{\lambda^0} & -\frac{6}{\left(\lambda^0\right)^2} & \frac{2}{\lambda^0} & \frac{6}{\left(\lambda^0\right)^2} \\ -\frac{6}{\left(\lambda^0\right)^2} & \frac{12}{\left(\lambda^0\right)^3} & -\frac{6}{\left(\lambda^0\right)^2} & -\frac{12}{\left(\lambda^0\right)^3} \\ \frac{2}{\lambda^0} & -\frac{6}{\left(\lambda^0\right)^2} & \frac{4}{\lambda^0\left(1-\lambda^0\right)} & 6\frac{1-2\lambda^0}{\left(\lambda^0\right)^2\left(1-\lambda^0\right)^2} \\ \frac{6}{\left(\lambda^0\right)^2} & -\frac{12}{\left(\lambda^0\right)^3} & 6\frac{1-2\lambda^0}{\left(\lambda^0\right)^2\left(1-\lambda^0\right)^2} & 12\frac{3\left(\lambda^0\right)^2-3\lambda^0+1}{\left(\lambda^0\right)^3\left(1-\lambda^0\right)^3} \end{pmatrix} \begin{pmatrix} \int_0^1 W(r)\,dr \\ \int_0^1 rW(r)\,dr \\ \int_{\lambda^0}^1 W(r)\,dr \\ \int_{\lambda^0}^1 \left(r-\lambda^0\right) W(r)\,dr \end{pmatrix}$$

4. *For Model II.b, define a stochastic process* $S^*(m)$ *on the set of integers as follows:* $S^*(0)=0, S^*(m)=S_1^*(m)$ *for* $m<0$ *and* $S^*(m)=S_2^*(m)$ *for* $m>0$*, with*

$$S_1^*(m) = |m|\,\mu_b^0 - 2\sum_{k=m+1}^{0} u_k, \quad m=-1,-2,\ldots$$

$$S_2^*(m) = m\mu_b^0 + 2\sum_{k=1}^{m} u_k, \quad m=1,2\ldots$$

If $\{u_t\}$ *is strictly stationary with a continuous distribution,* S^* *is a two-sided random walk with drift, and* $T(\hat{\lambda}-\lambda) \to^d \arg\min_m S^*(m)$

5. *For Model III.a,* $\hat{\lambda} \Longrightarrow \arg\max\left\{3\left(\int_0^1 rW(r)\,dr\right)^2/\lambda^3 + 3\left(\int_\lambda^1 rW(r)\,dr\right)^2/[1-\lambda^3]\right\}$.

6. *For Model III.b, define a stochastic process* $Z^*(m)$ *on the set of integers as follows:* $Z^*(0)=0, Z^*(m)=Z_1(m)$ *for* $m<0$ *and* $Z^*(m)=Z_2(m)$ *for* $m>0$*, with*

$$Z_1(m) = |m|\left(\mu_b^0+\beta_b^0T^*\lambda^0\right) - 2\sum_{k=m+1}^{0} u_k, \quad m=-1,-2,\ldots$$

$$Z_2(m) = m\left(\mu_b^0+\beta_b^0T^*\lambda^0\right) + 2\sum_{k=1}^{m} u_k, \quad m=1,2\ldots$$

If $\{u_t\}$ *is strictly stationary with a continuous distribution,* Z^* *is a two-sided random walk with drift, and* $T(\hat{\lambda}-\lambda) \to^d \arg\min_m Z^*(m)$.

Consider first the cases with $I(1)$ errors (I.a, II.a and III.a). In all cases, the estimate of the break fraction is not consistent. Moreover, the limit distributions do not involve any parameters of the model, in particular the magnitude of the change in slope. Also, these limit distributions are the same that obtains assuming no change in the trend function. This is fairly intuitive since the trend function is bounded while the noise component, being $I(1)$, is stochastically unbounded. Hence, the noise component dominates any signal in the trend. As shown in PZ, the "unbounded-trend" asymptotic distribution provides a good approximation to the finite sample distribution in the case of Model I.a. For Model II.a, it fails to deliver a limit distribution that involves the magnitude of the intercept shift, which influences greatly the finite sample distribution. However, for this case, PZ provides a stochastic expansion which provides a very accurate approximation.

Consider now the cases with $I(0)$ errors. First, in the case of Model I.b., for which no intercept shift is present nor permitted, the results are the same as in the "unbounded-trend" asymptotics. The reason for that is that the mapping from the asymptotic result to the approximation of the finite sample distribution is done by setting $T^* = T$. It is then easy to see that the same approximation applies in both asymptotic frameworks. Things are, however, very different when the intercept is allowed to change, as in Models II and III.

Consider the limit distribution for Model II.b. Of special relevance is the fact that the limit distribution in the "bounded-trend" asymptotic is independent of the value of the slope change β_b^0, unlike for the "unbounded-trend" asymptotic distribution. More interestingly, the two limit distributions would be the same if both the slope and level shifts are zero (though strictly, the "unbounded-trend" limit distribution is valid only if the change in slope is non-zero). This insensitivity of the limit distribution to the value of the change in slope implies automatically a bad finite sample approximation unless the slope change is very small (since the finite sample distribution is highly sensitive to changes in β_b^0).

The limit distribution in Model III.b does depend on the value of the slope shift β_b^0. However, it is easy to see that it is the same as for Model II.b, but with a level shift $\mu_b^0 + \beta_b^0 T^* \lambda^0$ instead of μ_b^0. The reason for this is that we can write Model III in the form of Model II with the corresponding change in the level shift. What transpires from these results is that once allowance is made for a level shift, the importance of the slope shift is masked and no longer has a first-order effect on the limit distribution. This contrasts sharply with the "unbounded-trend" asymptotic framework, whereby both the slope and level shifts influence the shape of the limit distribution.

The most intriguing feature of the "bounded-trend" asymptotic result is the following. Consider, for simplicity, the case where the errors u_t are $i.i.d.$ and suppose the true process involves a trend with a change in slope but with both segments joined at the time of break. If one uses Model I, where no allowance is made for a concurrent level shift at the time of break, the asymptotic approximation is $T^{3/2}(\hat{\lambda} - \lambda^0) \approx N(0, 4\sigma^2/[\lambda^0 \left(1 - \lambda^0\right) \left(\beta_b^0\right)^2])$, the same as for the "unbounded-trend" asymptotic framework. Consider now simply allowing for the possibility of a level shift, i.e., using the regression pertaining to Model II. Then the asymptotic approximation is $T(\hat{\lambda} - \lambda) \approx \arg\min_m 2W^*(m)$ with $W^*(m)$ a two-sided random walk defined by $W^*(m) = -\sum_{k=m+1}^{0} u_k$ for $m < 0$ and by $W^*(m) = \sum_{k=1}^{m} u_k$ for $m > 0$. Hence, introducing an irrelevant level shift regressor not only reduces the rate of convergence (as in the "unbounded-trend" asymptotic framework) but also completely eliminate the influence of the magnitude of the change in slope β_b^0 (unlike what occurs in the "unbounded-trend" asymptotic framework).

3.1 Bounded-trend with Shrinking Shifts Asymptotic Framework

We next consider the bounded-trend asymptotic framework with shrinking shifts specified as follows:

- Assumption 3: Let $\delta_T = (\mu_b, \beta_b)$ and $\delta_0 = \left(\mu_b^0, \beta_b^0\right) \neq 0$, then $\delta_T = \delta_0 v_T$, where v_T is a positive number such that $v_T \to 0$, and $T^{1/2-\alpha} v_T \to \infty$ for some $\alpha \in (0, 1/2)$.

In what follows, for reasons discussed above, we only consider the case with $I(0)$ errors. The limit distributions of the estimate of the break date $\hat{T}_1$ are stated in the following Theorem, whose proof is omitted as the arguments closely parrallel those for the proof of Theorem 1.

Theorem 3 *Under Assumption 3, we have,*

1. For Model I.b, $\sqrt{T} v_T(\hat{\lambda} - \lambda^0) \to^d N(0, 4\sigma^2/[\lambda^0 \left(1 - \lambda^0\right) \left(T^* \beta_b^0\right)^2])$.

2. For Model II.b, define a stochastic process $H(m)$ as follows: $H(0) = 0$, $H(m) = H_1(m)$ for $m < 0$ and $H(m) = H_2(m)$ for $m > 0$, with

$$H_1(m) = |m| \mu_b^0 - 2\psi_1 W_1(-m), m \leq 0$$

$$H_2(m) = m\mu_b^0 + 2\psi_2 W_2(m), m > 0$$

where W_i, $i = 1, 2$, be two independent standard Wiener processes defined on $[0, \infty)$, $\psi_1 = \lim E(k_0^{-1/2} \sum_{t=1}^{k_0} u_t)^2$ and $\psi_2 = \lim E((T - k_0)^{-1/2} \sum_{t=k_0+1}^{T} u_t)^2$. If $\{u_t\}$ is strictly stationary with a continuous distribution, H is a two-sided random walk with drift, and $Tv_T^2(\hat{\lambda} - \lambda) \rightarrow^d \arg\min_m H(m)$

3. For Model III.b, we have from Bai (1997),

$$\frac{\delta_T' g(\lambda_0) g(\lambda_0) \delta_T}{\psi_1^2} T(\hat{\lambda} - \lambda_0) \rightarrow^d \arg\max_s J(s)$$

where

$$J(s) = \begin{cases} W_1(-s) - |s|/2, & \text{if } s \leq 0 \\ \sqrt{\psi_2/\psi_1} W_2(s) - |s|/2, & \text{if } s > 0 \end{cases}$$

with ψ_1 and ψ_2 defined as above. And, $g(\lambda_0 T^*) = (1, \lambda_0 T^*)'$.

For Model I.a, the results do not change qualitatively, the limit distribution is the same, only the rate of convergence is slower, as expected (the limit distribution was already independent of the finite sample distribution of the errors, so taking shrinking shifts is not expected to change that back). In Models II and III, using shriking shifts effectively implies limit distributions that are no longer dependent on the finite sample distribution of the errors. But now, the slope shift parameter β_b^0 no longer influences even the limit distribution for Model III. Since changes in this parameter have an important effect on the finite sample distribution, it can be concluded that using a shrinking shift asymptotic framework is of no help in this "bounded-trend" asymptotic framework.

4 Simulations

In this section, we use simulation to assess the adequacy of the approximations for the various asymptotic distributions. Our results pertains to Model II.b, which involves $I(0)$ errors and both fitted slope and intercept changes. Throughout, the number of replications is 5000 and all graphs for the probability density functions are obtained from the empirical distributions of the estimates using a kernel smoothing method as in PZ [1].

[1] That is, for a given set of statistics, say $\{X_i\}_{i=1,\ldots,N}$, the pdf at value x is estimated by $\widetilde{f}(x) = (N \cdot h_x)^{-1} \sum_{i=1}^{N} K((x - X_i)/h_x)$ where $K(\cdot)$ is the kernel function and h_x is the bandwidth. In our case, $N = 5000$ and we use the standard normal distribution as the kernel function. Since the estimates of the break date are discrete integers, the cross-validation method for choosing the optimal bandwidth does not work well in this case. As a rule of thumb, we simply let $h_x = 0.3\hat{\sigma}_x$ where $\hat{\sigma}_x$ is the estimated standard deviation of a given sample of statistics $\{X_i\}_{i=1,\ldots,N}$.

Our first set of simulations is aimed to assess the extent to which the various approximations perform for a range of values for the slope and intercept changes. To that effect we simulate data from Model II with *i.i.d.* $N(0, \sigma^2)$ errors and the following combinations for the parameters of interest:

$$\begin{aligned} \mu_b^0 &= [-0.3, -0.1, 0, 0.1, 0.3] \\ \beta_b^0 &= [-0.1, -0.05, -0.03, 0.03, 0.05, 0.1]. \end{aligned}$$

The other parameters are set as in PZ, i.e., $\mu_1 = 1.72$, $\beta_1 = 0.03$ and $\sigma^2 = 0.1$. Since the errors are *i.i.d.* Normal, the fixed and shrinking shifts asymptotic distributions are the same. Hence, we make comparisons between the finite sample distribution, the "unbounded-trend" asymptotic approximation (labelled PZ) from Model II.b, and the "bounded-trend" asymptotic approximation (labelled BT), also from Model II.b.

The results are presented in Figures 1 to 5. They confirm the limited simulations reported in PZ to the effect that the "unbounded-trend" asymptotic distribution provides a very good approximation. In particular, it captures well the bimodality of the distribution, which the "bounded-trend" asymptotic distribution completely misses. The latter performs best when the intercept shift is large, in which case the bimodal nature of the finite sample distribution is the weakest, though the "unbounded-trend" limit distribution still performs better. When the level is small, the "bounded-trend" limit distribution completely misses most aspects of the finite sample distribution.

4.1 Simulations calibrated to empirical time series

While the previous simulation exercise is useful to assess the general properties of the approximations, it remains to show how the approximations fare in the context of data generating processes that are most likely to occur in practical applications. To that effect, we now consider a simulation design calibrated to historical (log) real per capita GDP series for a variety of countries. These are a subset of the series analyzed in PZ and they cover the period from 1870 to 1986 for 7 different countries for which the noise function was found to be $I(0)$ in Perron (1992): Australia, Canada, Denmark, France, Germany, the United Kingdom and the United States [2]. The comparison is made among 3 asymptotic approximations, namely, the "unbounded-trend", the "bounded-trend with fixed shifts" and the "bounded-trend with

[2] This data set is the same as used by Kormendi and Meguire (1990) and Perron (1992) and was obtained through the *Journal of Money, Credit and Banking* editorial office. All series are real GDP except for the United States for which real GNP is used. For the United States, the series is real GNP from the National Income and Products Accounts for the period 1929-1986, spliced to Romer's (1989) estimates for the period

shrinking shifts". Table I presents the relevant parameter estimates used to calibrate the simulations obtained by estimating the following model:

$$\begin{aligned} y_t &= d_t + u_t \\ d_t &= \mu_1 + \beta_1 t + \mu_b C_t + \beta_b B_t \\ u_t &= \sum_{j=1}^{p} \rho_j u_{t-j} + \varepsilon_t \end{aligned}$$

The noise component is modelled as an $AR(p)$ using BIC to select the order. From these estimates, we construct an estimate of the long-run variance (or spectral density function at frequency zero) of the errors using Andrews' (1991) data dependent method with AR(1) approximation. The errors are generated as $i.i.d.$ $N(0,1)$. The number of replications is 2000.

The results are presented in Figures 6-7. Although, there is a wide variety in the shapes of the finite sample distributions, overall the "unbounded-trend" asymptotic framework clearly performs best in providing a good approximation. For Australia, Canada, Denmark, France and Germany, it is indeed, very accurate. For Canada and Denmark, the "bounded-trend with shrinking-shift" approximation is also good; and for Germany, the "bounded-trend with fixed shift" is better than the "unbounted-trend" approximation in the center of the distribution while the reverse ranking holds in the tails, especially the right tail. For the United States, none of the approximations work very well due to the wide spread of the finite sample distribution caused by the fact that the shift in slope in very small, if at all present. For the United Kingdom, the model calibration is such that the break is identified with high precision in the sense that the finite sample distribution and all approximations are nearly degenerate at 0 (in fact only up to 3 values out of the 2000 replications are nonzero). Therefore, the kernel smoothed densities are not a good indication of the true nature of the distribution since it is greatly influenced by a few outlying values.

5 Conclusions

We have considered the adequacy of the various asymptotic approximations pertaining to the estimate of the break date in a simple linear trend model with a single change in slope

1870-1928. For the United Kingdom, the series is real GDP from Feinstein (1972) for the period 1870-1947 spliced to the International Financial Statistics (IFS) series of the IMF for the period 1948-1986. For the remaining countries, the series are indices of annual real GDP from Madison (1982) spliced to the postwar IFS data. The population series used are from the same sources. A logarithmic transformation is applied.

with a possible concurrent change in level. Our focus has been on comparing the limit distributions provided by the so-called "bounded-trend" and "unbounded-trend" frameworks. As expected, when the noise function u_t is $I(1)$, the "bounded-trend" asymptotic framework is of little use. Of more interest are the results pertaining to the case with a stationary noise function. If the intercept μ does not change concurrently with the slope β and the fitted intercept is not allowed to change, then all frameworks are basically equivalent. However, if the fitted intercept is allowed to change, whether or not μ changes, the "bounded-trend" asymptotic framework completely misses important features of the finite sample distribution of the estimate of the break date, especially the pronounced bimodality that was uncovered by PZ and shown to be well captured using the alternative "unbounded-trend" asymptotic framework. Simulation experiments confirm our theoretical findings, which expose the drawbacks of using the "bounded-trend" asymptotic framework in the context of structural change models.

References

[1] Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817-858.

[2] Andrews, D.W.K., McDermott, C.J., 1995. Nonlinear econometric models with deterministically trending variables. Review of Economic Studies 62, 343-360.

[3] Bai, J., 1997. Estimation of a change point in multiple regression models. The Review of Economics and Statistics 79, 551-563.

[4] Bai, J., Lumsdaine, R., Stock, J.H., 1998. Testing for and dating common breaks in multivariate time series. Review of Economic Studies 65, 395-432.

[5] Bai, J. Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47-78.

[6] Bai, J., Perron, P., 2005. Multiple structural change models: a simulation analysis. Forthcoming in Econometric Essays, D. Corbea, S. Durlauf and B.E. Hansen (eds.), Cambridge University Press.

[7] Feinstein, C.H., 1972. National income expenditure and output of the United Kingdom, vol. 6 in Studies in the National Income and Expenditure of the United Kingdom, Richard Stone (series editor), London: Cambridge University Press.

[8] Hinkley, D.V., 1970. Inference about the change-point in a sequence of random variables. Biometrika 57, 1-17.

[9] Kormendi, R.C., Meguire, P., 1990. A multicountry characterization of the nonstationarity of aggregate output. Journal of Money, Credit and Banking 22, 77-93.

[10] Maddison, A., 1982. Phases of Capitalist Development. London: Oxford University Press.

[11] Nelson, C.R., Plosser, C.I., 1982. Trends and random walks in macroeconomics time series: some evidence and implications. Journal of Monetary Economics 10, 139-162.

[12] Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 1361-1401.

[13] Perron, P., 1992. Trend, unit root and structural change: a multi-country study with historical data. Proceedings of the Business and Economic Statistics Section, American Statistical Association, 144-149.

[14] Perron, P., 2005. Dealing with structural breaks. Forthcoming in the Palgrave Handbook of Econometrics, Vol. 1, Econometric Theory.

[15] Perron, P., Zhu, X., 2005. Structural breaks with deterministic and stochastic trends. Journal of Econometrics, Forthcoming.

[16] Picard, D., 1985. Testing and estimating change-points in time series. Journal of Applied Probability 17, 841-867.

[17] Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335-346.

[18] Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Annals of Statistics 20, 971-1001.

[19] Romer, C., 1989. The prewar business cycle reconsidered: new estimates of gross national product, 1969-1928. Journal of Political Economy 97, 1-37.

[20] Siegmund, D., 1988. Confidence sets in change-point problems. International Statistical Review 56, 31-48.

[21] Yao, Y.-C., 1987. Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. Annals of Statistics 15, 1321-1328.

Appendix

From the properties of projections, we have for all T, $SSR(\hat{\lambda}) \leq SSR(\lambda^0)$, which implies the following inequality as shown in PZ,

$$\gamma^{0\prime}(X_{T_1^0} - X_{\hat{T}_1})'(I - P_{\hat{T}_1})(X_{T_1^0} - X_{\hat{T}_1})\gamma^0 + 2\gamma^{0\prime}(X_{T_1^0} - X_{\hat{T}_1})'(I - P_{\hat{T}_1})U + U'(P_{T_1^0} - P_{\hat{T}_1})U \leq 0 \tag{A.1}$$

Note also that

$$\begin{aligned} & \arg\min_{T_1} [SSR(T_1)] = \arg\min_{T_1} \left[SSR(T_1) - SSR(T_1^0)\right] \\ = \; & \arg\min_{T_1} \Big[\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I - P_{T_1})(X_{T_1^0} - X_{T_1})\gamma^0 + 2\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I - P_{T_1})U \\ & \quad + U'(P_{T_1^0} - P_{T_1})U\Big] \end{aligned}$$

This will be employed to derive the asymptotic distribution of the least-squares estimate of the break fraction, $\hat{\lambda} = \hat{T}_1/T$. Note that throughout we use the label $O(T^a)$ and $O_p(T^a)$ in its strict sense, i.e. meaning that the variables are not $o(T^a)$ and $o_p(T^a)$. We shall repeatedly use the following notation:

$$\begin{aligned} (XX) &\equiv \gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I - P_{T_1})(X_{T_1^0} - X_{T_1})\gamma^0 \\ (XU) &\equiv \gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I - P_{T_1})U \\ (UU) &\equiv U'(P_{T_1^0} - P_{T_1})U \end{aligned}$$

Proof of Theorem 1. We start by analyzing consistency and then the rate of convergence when the estimate is consistent. The results can be obtained from the following Lemma, which is similar to Lemma 1 in PZ.

Lemma A.1 *1. For Model I.a,*

$$(XX) = \left|T_1 - T_1^0\right|^2 O_p\left(T^{-1}\right), (XU) = \left|T_1 - T_1^0\right| O_p\left(T^{1/2}\right), (UU) = \left|T_1 - T_1^0\right| O_p(T)$$

2. For Model I.b,

$$(XX) = \left|T_1 - T_1^0\right|^2 O_p\left(T^{-1}\right), (XU) = \left|T_1 - T_1^0\right| O_p\left(T^{-1/2}\right), (UU) = \left|T_1 - T_1^0\right| O_p\left(T^{-1}\right)$$

3. For Model II.a,

$$(XX) = \left|T_1 - T_1^0\right| O_p(1), (XU) = \left|T_1 - T_1^0\right| O_p\left(T^{1/2}\right), (UU) = \left|T_1 - T_1^0\right| O_p(T)$$

4. For Model II.b,

$$(XX) = \left|T_1 - T_1^0\right| O_p(1), (XU) = \left|T_1 - T_1^0\right|^{1/2} O_p(1), (UU) = \left|T_1 - T_1^0\right|^{1/2} O_p\left(T^{-1/2}\right)$$

5. For Model III.a,

$$(XX) = \left|T_1 - T_1^0\right| O_p(1), (XU) = \left|T_1 - T_1^0\right| O_p\left(T^{1/2}\right), (UU) = \left|T_1 - T_1^0\right| O_p(T)$$

6. For Model III.b,

$$(XX) = \left|T_1 - T_1^0\right| O_p(1), (XU) = \left|T_1 - T_1^0\right|^{1/2} O_p(1), (UU) = \left|T_1 - T_1^0\right|^{1/2} O_p\left(T^{-1/2}\right)$$

The proof is similar to that of Lemma 1 in PZ and, hence, omitted. Using Lemma 1, it is relatively easy to establish the following results about consistency and rates of convergence.

Lemma A.2 *For Models, I.a, II.a and III.a (with $I(1)$ errors), $\hat{\lambda} - \lambda^0 = O_p(1)$ and the estimate of the break fraction is not consistent for the true break fraction. For Models I.b, II.b. and III.b (with $I(0)$ errors), $\hat{\lambda} \to_p \lambda^0$.*

Lemma A.3 *In the case of $I(0)$ errors, for Model I.b, $\hat{\lambda} - \lambda^0 = O_p(T^{-1/2})$, and for Models II.b and III.b, $\hat{\lambda} - \lambda^0 = O_p(T)$*

We are now in a position to derive the limit distributions. We start with the case where the errors are $I(1)$.

Proof for Models I.a, II.a and III.a: In all cases, as can be verified from Lemma A.1, the term (UU) dominates the other two. Since this term does not involve any parameter and, in particular, the magnitude of the changes, this immediately implies that the limit distribution is that corresponding to the no break case. We have for all three cases:

$$\begin{aligned}
\hat{\lambda} &= \arg\min_{\lambda\in[0,1]} \frac{U'(P_{T_1^0} - P_{T_1})U}{T^2} + o_p(1) = \arg\max_{\lambda\in[0,1]} \frac{U'P_{T_1}U}{T^2} + o_p(1) \\
&= \arg\max_{\lambda\in[0,1]} \frac{U'X_{T_1}\left(X_{T_1}'X_{T_1}\right)^{-1}X_{T_1}'U}{T^2}
\end{aligned}$$

For Model I.a, we have

$$T^{-1/2}X_{T_1}'U = \begin{pmatrix} T^{-1/2}\sum_{t=1}^{T} u_t \\ (T^*/T^{-3/2})\sum_{t=1}^{T} tu_t \\ (T^*/T^{-3/2})\sum_{t=T_1+1}^{T}(t-T_1)u_t \end{pmatrix} \Rightarrow \begin{pmatrix} \int_0^1 W(r)\,dr \\ T^*\int_0^1 rW(r)\,dr \\ T^*\int_{\lambda^0}^1\left(r-\lambda^0\right)W(r)\,dr \end{pmatrix} \quad \text{(A.2)}$$

and

$$T^{-1}X'_{T_1}X_{T_1} \to \begin{pmatrix} 1 & \frac{1}{2}T^* & \frac{(1-\lambda)^2}{2}T^* \\ \frac{1}{2}T^* & \frac{1}{3}\left(T^*\right)^2 & \frac{(1-\lambda)^2(\lambda+2)}{6}\left(T^*\right)^2 \\ \frac{(1-\lambda)^2}{2}T^* & \frac{(1-\lambda)^2(\lambda+2)}{6}\left(T^*\right)^2 & \frac{(1-\lambda)^3}{3}\left(T^*\right)^2 \end{pmatrix} \tag{A.3}$$

Hence,

$$\begin{aligned} \hat{\lambda} &= \arg\max_{\lambda\in\Lambda} \begin{pmatrix} \int_0^1 W(r)\,dr & T^*\int_0^1 rW(r)\,dr & T^*\int_{\lambda^0}^1 \left(r-\lambda^0\right)W(r)\,dr \end{pmatrix} \\ &\quad \times \begin{pmatrix} 1 & \frac{1}{2}T^* & \frac{(1-\lambda)^2}{2}T^* \\ \frac{1}{2}T^* & \frac{1}{3}\left(T^*\right)^2 & \frac{(1-\lambda)^2(\lambda+2)}{6}\left(T^*\right)^2 \\ \frac{(1-\lambda)^2}{2}T^* & \frac{(1-\lambda)^2(\lambda+2)}{6}\left(T^*\right)^2 & \frac{(1-\lambda)^3}{3}\left(T^*\right)^2 \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 W(r)\,dr \\ T^*\int_0^1 rW(r)\,dr \\ T^*\int_{\lambda^0}^1 \left(r-\lambda^0\right)W(r)\,dr \end{pmatrix} \end{aligned}$$

which reduces to the expression in Theorem 1. The limit distribution stated for Models II.a and III.a can be obtained analogously. We now consider the limit distributions for the cases with $I(0)$ errors.

Proof for Model I.b: Given the rate of convergence stated in Lemma A.3, we can restrict the search to the set of break dates that satisfy $|T_1 - T_1^0| = O_p(\sqrt{T})$. Then, the terms (XX) and (XU) are both $O_p(1)$, and $(UU) = o_p(1)$. Hence, we have $\hat{\lambda} = \arg\max_\lambda [(XX) + 2(XU) + o_p(1)]$. Without loss of generality, we set $\mu_1 = \beta_1 = 0$, since the estimates are invariant to these values and, hence, $\gamma^{0\prime} = (0, 0, \beta_b^0)$. Defining $m_T = |T_1 - T_1^0|/\sqrt{T}$, we have,

$$\begin{aligned} &T^{-1/2}X'_{T_1}(X_{T_1^0} - X_{T_1})\gamma^0 \\ &= \beta_b^0 \begin{pmatrix} \frac{1}{\sqrt{T}}\frac{T^*}{T}\left(\sum_{t=1}^{T_1-T_1^0} t + (T-T_1)(T_1-T_1^0)\right) \\ \frac{1}{\sqrt{T}}\left(\frac{T^*}{T}\right)^2\left(\sum_{t=1}^{T_1-T_1^0} t\,(T_1^0+t) + (T_1-T_1^0)\sum_{t=T_1+1}^{T} t\right) \\ \frac{1}{\sqrt{T}}\left(\frac{T^*}{T}\right)^2 (T_1-T_1^0)\sum_{t=1}^{T-T_1} t \end{pmatrix} \\ &= \beta_b^0 \begin{bmatrix} \left(1-\lambda^0\right)T^* & \frac{1-\left(\lambda^0\right)^2}{2}\left(T^*\right)^2 & \frac{\left(1-\lambda^0\right)^2}{2}\left(T^*\right)^2 \end{bmatrix} m_T + o_p(1) \end{aligned}$$

Hence, using (A.3),

$$\begin{aligned} &\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'X_{T_1}\left(X'_{T_1}X_{T_1}\right)^{-1}X'_{T_1}(X_{T_1^0} - X_{T_1})\gamma^0 \\ &= T^{-1/2}\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'X_{T_1}\left(T^{-1}X'_{T_1}X_{T_1}\right)^{-1}T^{-1/2}X'_{T_1}(X_{T_1^0} - X_{T_1})\gamma^0 \end{aligned}$$

$$
\begin{aligned}
&= (\beta_b^0)^2 \begin{pmatrix} \left(1-\lambda^0\right) T^* \\ \frac{1-\left(\lambda^0\right)^2}{2} \left(T^*\right)^2 \\ \frac{\left(1-\lambda^0\right)^2}{2} \left(T^*\right)^2 \end{pmatrix}' \begin{pmatrix} 1 & \frac{1}{2}T^* & \frac{(1-\lambda)^2}{2}T^* \\ \frac{1}{2}T^* & \frac{1}{3}\left(T^*\right)^2 & \frac{(1-\lambda)^2(\lambda+2)}{6}\left(T^*\right)^2 \\ \frac{(1-\lambda)^2}{2}T^* & \frac{(1-\lambda)^2(\lambda+2)}{6}\left(T^*\right)^2 & \frac{(1-\lambda)^3}{3}\left(T^*\right)^2 \end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix} \left(1-\lambda^0\right) T^* \\ \frac{1-\left(\lambda^0\right)^2}{2} \left(T^*\right)^2 \\ \frac{\left(1-\lambda^0\right)^2}{2} \left(T^*\right)^2 \end{pmatrix} m_T^2 + o_p(1) \\
&= (\beta_b^0)^2 \left[\frac{\left(1-\lambda^0\right)\left(4-\lambda^0\right)}{4} \left(T^*\right)^2 \right] m_T + o_p(1)
\end{aligned}
$$

Furthermore,

$$
\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(X_{T_1^0} - X_{T_1})\gamma^0 = (\beta_b^0)^2 \left(T^*\right)^2 \left(1-\lambda^0\right) m_T^2 + o_p(1)
$$

and

$$
(XX) = (\beta_b^0)^2 \left(T^*\right)^2 \left(1-\lambda^0\right) \lambda^0 m_T^2/4 + o_p(1).
$$

Next,

$$
\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'U = \beta_b^0 T^* \left(\sum_{t=T_1^0+1}^{T_1} \frac{t-T_1^0}{T} u_t + \sum_{t=T_1+1}^{T} \frac{T_1 - T_1^0}{T} u_t \right) \Rightarrow \beta_b^0 T^* m_T \int_{\lambda^0}^1 dW\left(r\right)
$$

using (A.2). Then,

$$
\begin{aligned}
&(XU) \\
&= \beta_b^0 m_T \sigma \left[T^* \int_{\lambda^0}^1 dW\left(r\right) - \left[\begin{array}{ccc} -T^* \frac{1-\lambda^0}{2} & \frac{3\left(1-\lambda^0\right)}{2\lambda^0} & \frac{3\left(2\lambda^0-1\right)}{2\lambda^0\left(1-\lambda^0\right)} \end{array}\right] \left[\begin{array}{c} \int_0^1 dW\left(r\right) \\ T^* \int_0^1 r dW\left(r\right) \\ T^* \int_{\lambda^0}^1 \left(r-\lambda^0\right) dW\left(r\right) \end{array}\right] \right] + o_p(1) \\
&= \beta_b^0 m_T \sigma T^* \zeta + o_p(1)
\end{aligned}
$$

where ζ is defined in PZ, eq. A-8. Collecting results, we have

$$
m_T^* = \arg\max_{m_T} \left[\left(T^* \beta_b^0\right)^2 \frac{2\left(1-\lambda^0\right)\lambda^0}{4} m_T^2 + \beta_b^0 m_T \sigma T^* \zeta + o_p(1) \right]
$$

and the result follows.

Proof for Model II.b: In this case, from Lemma A.3, $T(\hat{\lambda}-\lambda)=O_p(1)$ and we can restrict the analysis to break dates in a set such that $|T_1-T_1^0|=O_p(1)$. Following PZ, we define the following quantities. For $T_1^0\geq T_1$

$$g_1\left(T_1-T_1^0\right)=\sum_{t=T_1+1}^{T_1^0}\left[\mu_b^0+\beta_b^0\frac{t-T_1^0}{T}T^*\right],\quad h_1\left(T_1-T_1^0\right)=\sum_{t=T_1+1}^{T_1^0}\left[\mu_b^0+\beta_b^0\frac{t-T_1^0}{T}T^*\right]^2$$

and, for $T_1\geq T_1^0$,

$$g_2\left(T_1-T_1^0\right)=\sum_{t=T_1^0+1}^{T_1}\left[\mu_b^0+\beta_b^0\frac{t-T_1^0}{T}T^*\right],\quad h_2\left(T_1-T_1^0\right)=\sum_{t=T_1^0+1}^{T_1}\left[\mu_b^0+\beta_b^0\frac{t-T_1^0}{T}T^*\right]^2$$

Let $n=T_1-T_1^0$, and $k=t-T_1^0$, then

$$\begin{aligned}\text{For } n &< 0,\ g_1(n)=\sum_{k=n+1}^{0}\left[\mu_b^0+\beta_b^0\frac{k}{T}T^*\right],\quad h_1(n)=\sum_{k=n+1}^{0}\left[\mu_b^0+\beta_b^0\frac{k}{T}T^*\right]^2\\ \text{For } n &> 0,\ g_2(n)=\sum_{k=1}^{n}\left[\mu_b^0+\beta_b^0\frac{k}{T}T^*\right],\quad h_2(n)=\sum_{k=1}^{n}\left[\mu_b^0+\beta_b^0\frac{k}{T}T^*\right]^2\end{aligned}$$

Now consider driving T to infinity while keeping T^* fixed,

$$\begin{aligned}\text{For } n &< 0,\ g_1(n)=-n\mu_b^0+o_p(1),\quad h_1(n)=-n(\mu_b^0)^2+o_p(1)\\ \text{For } n &> 0,\ g_2(n)=n\mu_b^0+o_p(1),\quad h_2(n)=n(\mu_b^0)^2+o_p(1)\end{aligned}$$

since $n=T_1-T_1^0=O_p(1)$. Now, for $T_1>T_1^0$,

$$\begin{aligned}&\gamma^{0\prime}(X_{T_1^0}-X_{T_1})'(I-P_{T_1})(X_{T_1^0}-X_{T_1})\gamma^0\\ =&\sum_{t=T_1^0+1}^{T_1}\left[\mu_b^0+\beta_b^0\frac{t-T_1^0}{T}T^*\right]^2\\ &-\frac{1}{\sqrt{T}}\sum_{t=T_1^0+1}^{T_1}\left[\mu_b^0+\beta_b^0\frac{t-T_1^0}{T}T^*\right]x(T_1)_t'\left(\frac{X_{T_1}'X_{T_1}}{T}\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=T_1^0+1}^{T_1}x(T_1)_t\left[\mu_b^0+\beta_b^0\frac{t-T_1^0}{T}T^*\right]\end{aligned}$$

where $x(T_1)_t=(1,(t/T)T^*,0,0)$ for $T_1^0+1\leq t\leq T_1$. Note that $\sum_{t=T_1^0+1}^{T_1}[\mu_b^0+\beta_b^0\frac{t-T_1^0}{T}T^*]^2=h_2$ and

$$\frac{1}{\sqrt{T}}\sum_{t=T_1^0+1}^{T_1}\left[\mu_b^0+\beta_b^0\frac{t-T_1^0}{T}T^*\right]x(T_1)_t' \tag{A.4}$$

$$=\frac{1}{\sqrt{T}}g_2\begin{pmatrix}1 & \frac{T_1^0}{T}T^* & 0 & 0\end{pmatrix}+\frac{1}{\sqrt{T}}\sum_{t=T_1^0+1}^{T_1}\left[\mu_b^0+\beta_b^0\frac{t-T_1^0}{T}T^*\right]\begin{pmatrix}0 & \frac{t-T_1^0}{T}T^* & 0 & 0\end{pmatrix}$$

$$\leq \left\{ \frac{1}{\sqrt{T}} \left| g_2 \right| + \frac{1}{\sqrt{T}} \left| g_2 \right| \frac{T_1 - T_1^0}{T} T^* \right\} \begin{pmatrix} 1 & \frac{T_1^0}{T} T^* & 0 & 0 \end{pmatrix} = O_p \left(\left| g_2 \right| T^{-1/2} \right)$$

since $(T_1 - T_1^0)/T \to 0$. Also, $(T^{-1} X_{T_1}' X_{T_1})^{-1} = O_p(1)$, hence $\gamma^{0\prime}(X_{T_1^0} - X_{T_1})' P_{T_1} (X_{T_1^0} - X_{T_1}) \gamma^0 = O_p\left(\left| g_2 \right|^2 T^{-1} \right) = o_p(1)$. Therefore, we have $(XX) = h_2 + o_p(1)$ for $T_1 > T_1^0$, and $(XX) = h_1 + o_p(1)$ for $T_1^0 > T_1$. For the term (XU), we first have, for $T_1 > T_1^0$,

$$\begin{aligned}
& \gamma^{0\prime}(X_{T_1^0} - X_{T_1})' (I - P_{T_1}) U \\
= & \sum_{t=T_1^0+1}^{T_1} \left[\mu_b^0 + \beta_b^0 \frac{t - T_1^0}{T} T^* \right] u_t \\
& - \left[\frac{1}{\sqrt{T}} \sum_{t=T_1^0+1}^{T_1} \left[\mu_b^0 + \beta_b^0 \frac{t - T_1^0}{T} T^* \right] x(T_1)_t' \frac{1}{\sqrt{T}} \right] \left(\frac{X_{T_1}' X_{T_1}}{T} \right)^{-1} \frac{1}{\sqrt{T}} X_{T_1}' U \\
= & \sum_{t=T_1^0+1}^{T_1} \left[\mu_b^0 + \beta_b^0 \frac{t - T_1^0}{T} T^* \right] u_t + o_p(1)
\end{aligned}$$

where the last equality follows from (A.4). Hence,

$$(XU) = \begin{cases} \sum_{t=T_1^0+1}^{T_1} \left[\mu_b^0 \right] u_t + o_p(1) & \text{if } T_1 > T_1^0 \\ 0 & \text{if } T_1 = T_1^0 \\ -\sum_{t=T_1+1}^{T_1^0} \left[\mu_b^0 \right] u_t + o_p(1) & \text{if } T_1 < T_1^0 \end{cases}$$

and the result follows combining the limits of (XX) and (XU), since the term (UU) is dominated.

Proof for Model III.b: The result follows from Bai (1997) with simple modifications.

Table I: Parameter Estimates; Model II.b, Real per Capita GDP Series

	AUS	CAN	DEN	FRA	GER	UK	US
$\hat{\mu}_1$	1.0328	0.3469	2.1548	2.1110	0.9342	-2.2746	0.7019
$\hat{\beta}_1$	0.0021	0.0163	0.0162	0.0107	0.0138	0.0101	0.0155
$\hat{\mu}_b$	-0.1385	-0.2844	-0.2145	-0.4596	-0.2797	-0.2498	0.1800
$\hat{\beta}_b$	0.0199	0.0133	0.0142	0.0307	0.0300	0.0082	0.0001
$\hat{\psi}^2$	0.0189 (3)	0.0504 (3)	0.0105 (1)	0.0269 (2)	0.0629 (2)	0.0073 (2)	0.0399 (2)
T_1	1929	1930	1939	1943	1945	1919	1940

Note: $\hat{\psi}^2$ is the estimate of the spectral density function of u_t at frequency zero and the number in parenthesis is the lag order of the autoregression selected by the BIC.

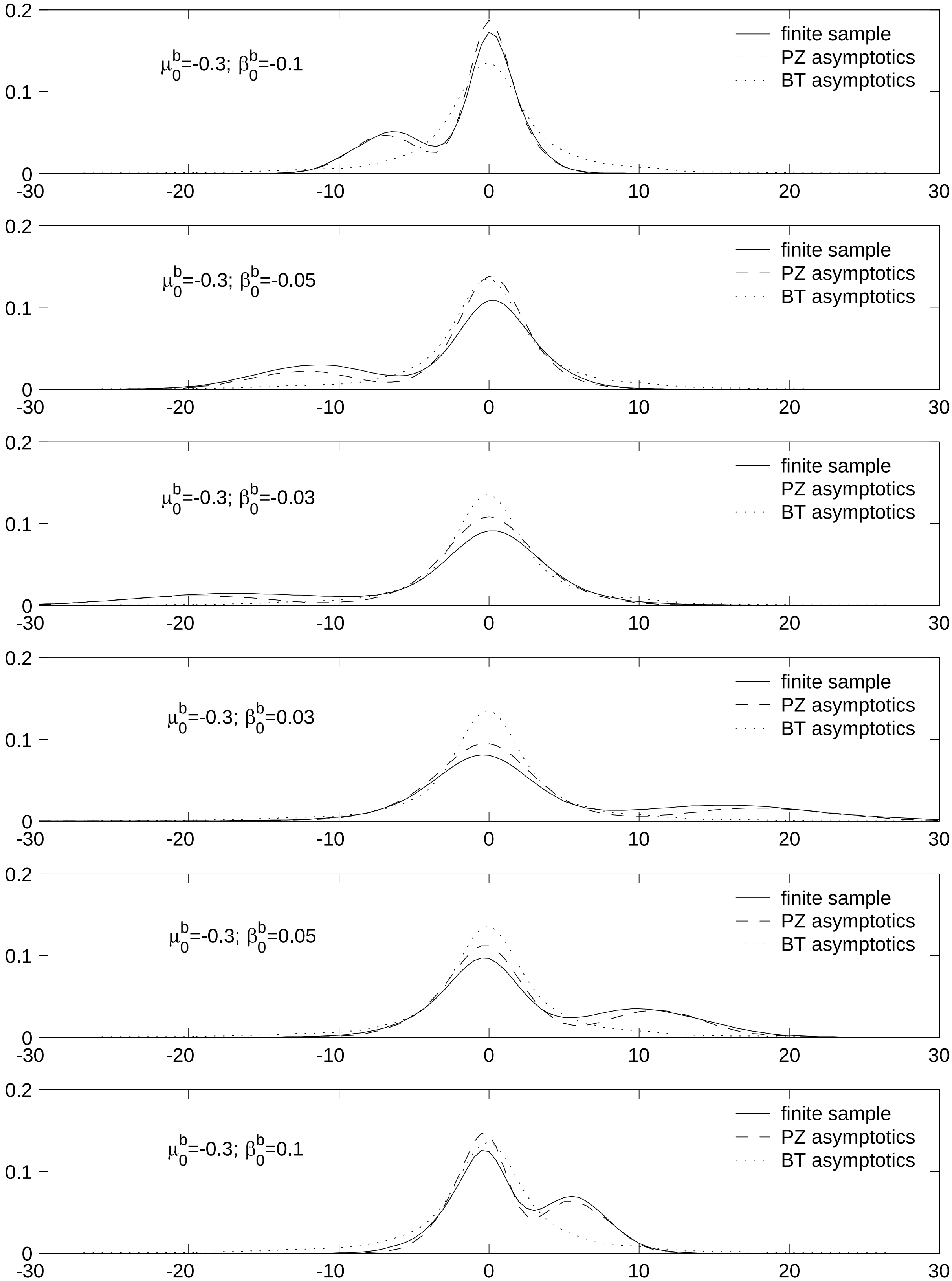

Figure 1: Finite sample versus asymptotic approximations: $\mu_b^0 = -0.3; \beta_b^0 = -0.1, -0.05, -0.03, 0.03, 0.05, 0.1$

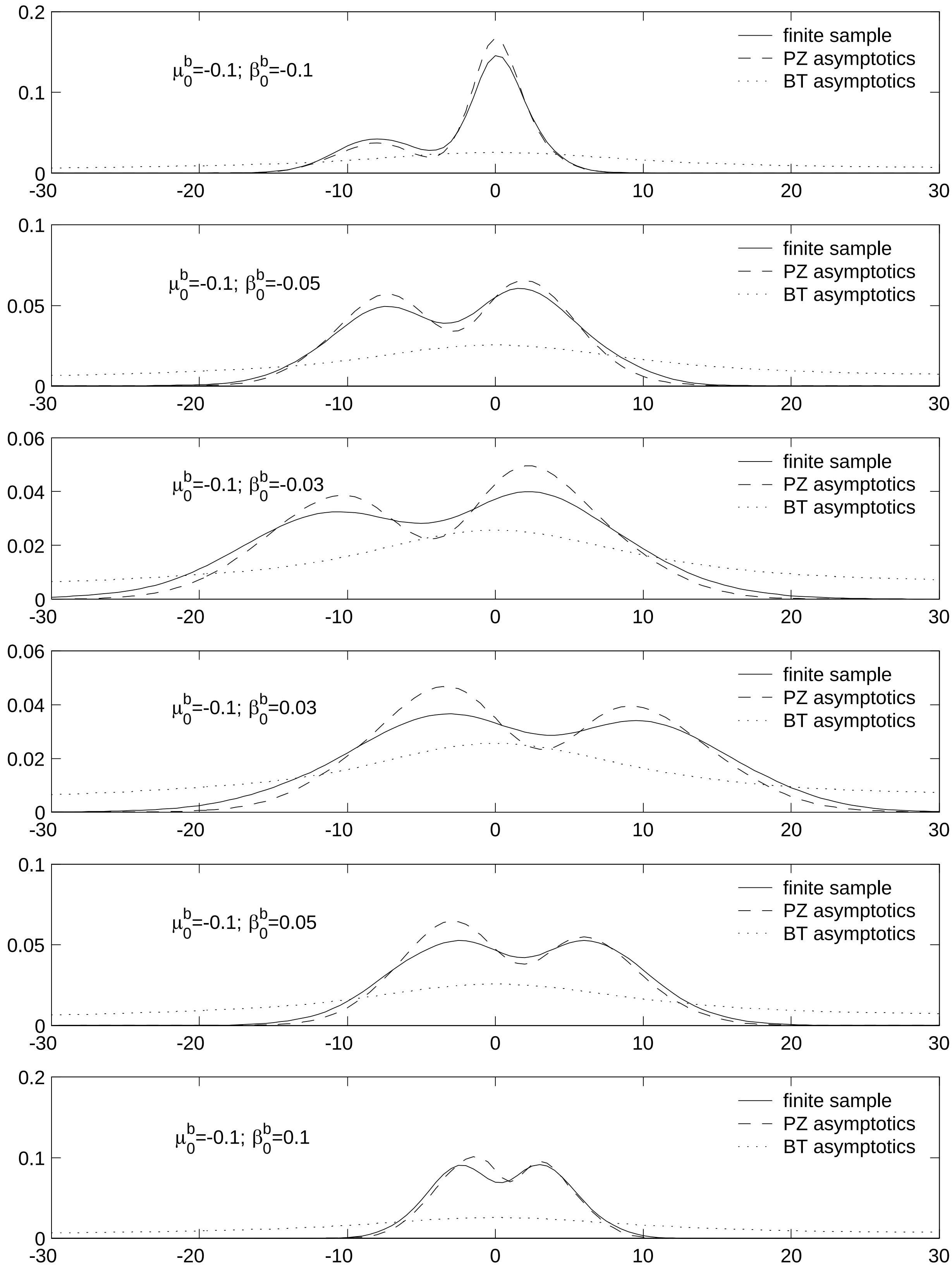

Figure 2: Finite sample versus asymptotic approximations: $\mu_b^0 = -0.1; \beta_b^0 = -0.1, -0.05, -0.03, 0.03, 0.05, 0.1$

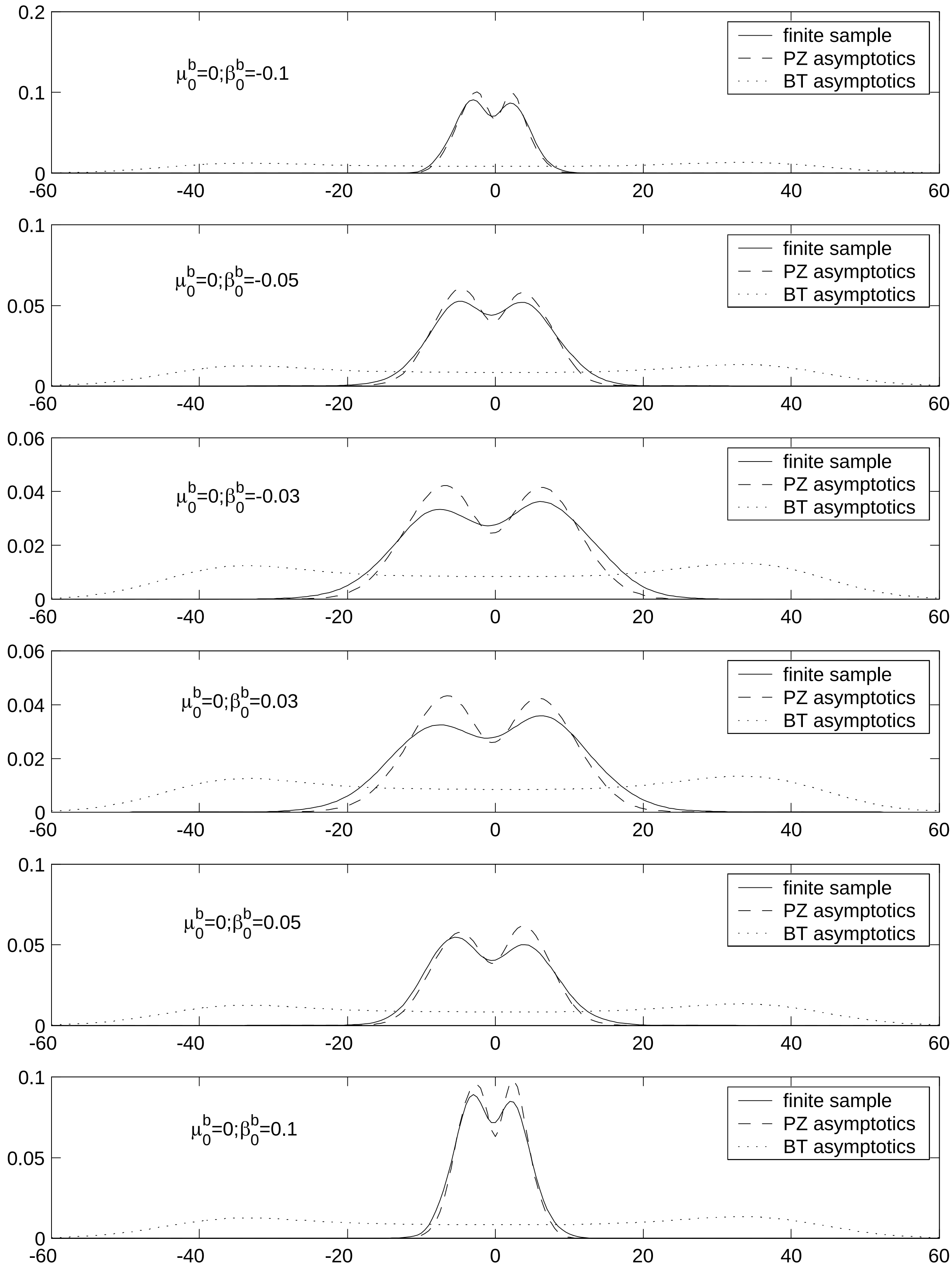

Figure 3: Finite sample versus asymptotic approximations: $\mu_b^0 = 0; \beta_b^0 = -0.1, -0.05, -0.03, 0.03, 0.05, 0.1$

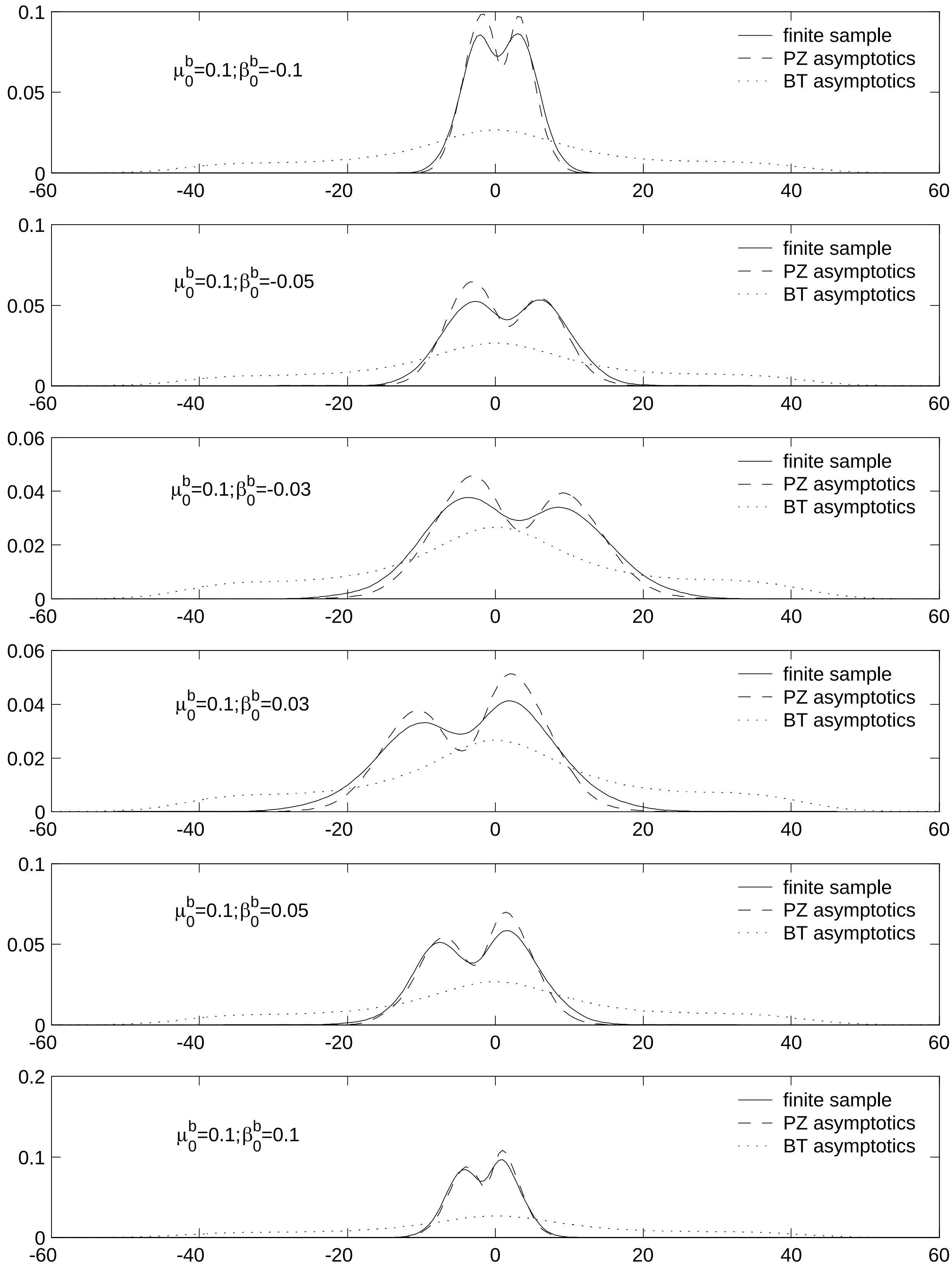

Figure 4: Finite sample versus asymptotic approximations: $\mu_b^0 = 0.1; \beta_b^0 = -0.1, -0.05, -0.03, 0.03, 0.05, 0.1$

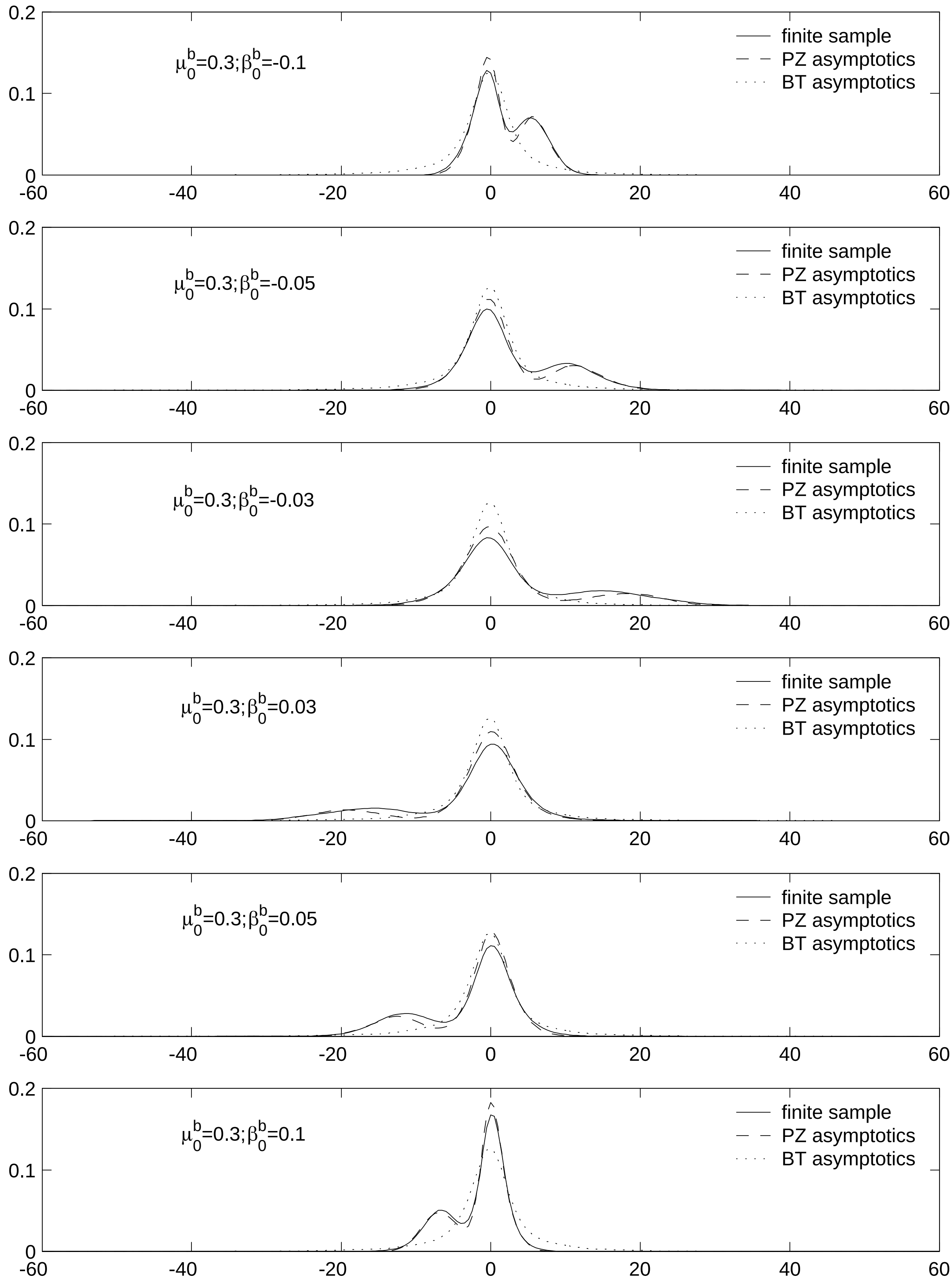

Figure 5: Finite sample versus asymptotic approximations: $\mu_b^0 = 0.3; \beta_b^0 = -0.1, -0.05, -0.03, 0.03, 0.05, 0.1$

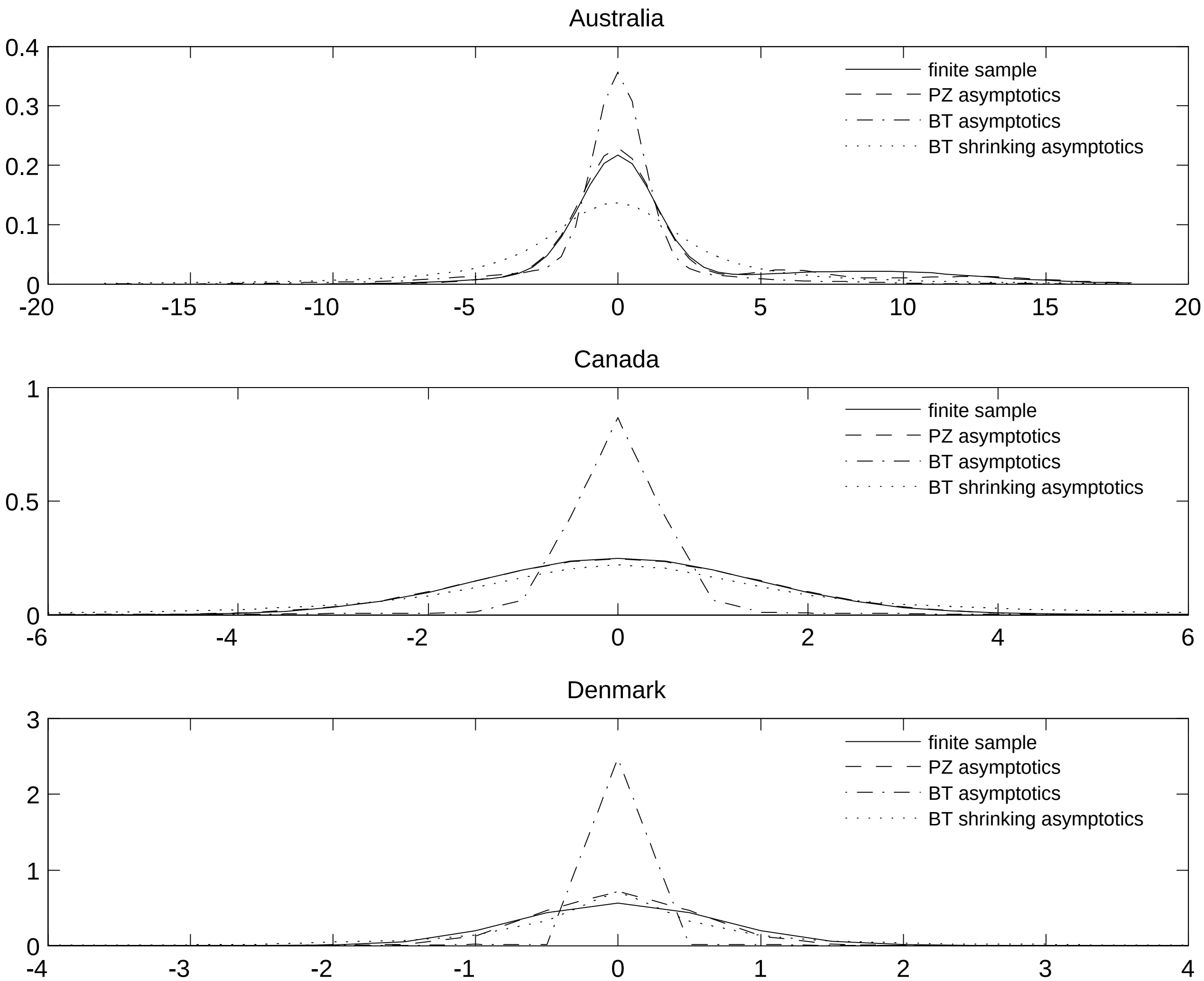

Figure 6: Finite sample versus asymptotic approximations: Empirically calibrated comparisons

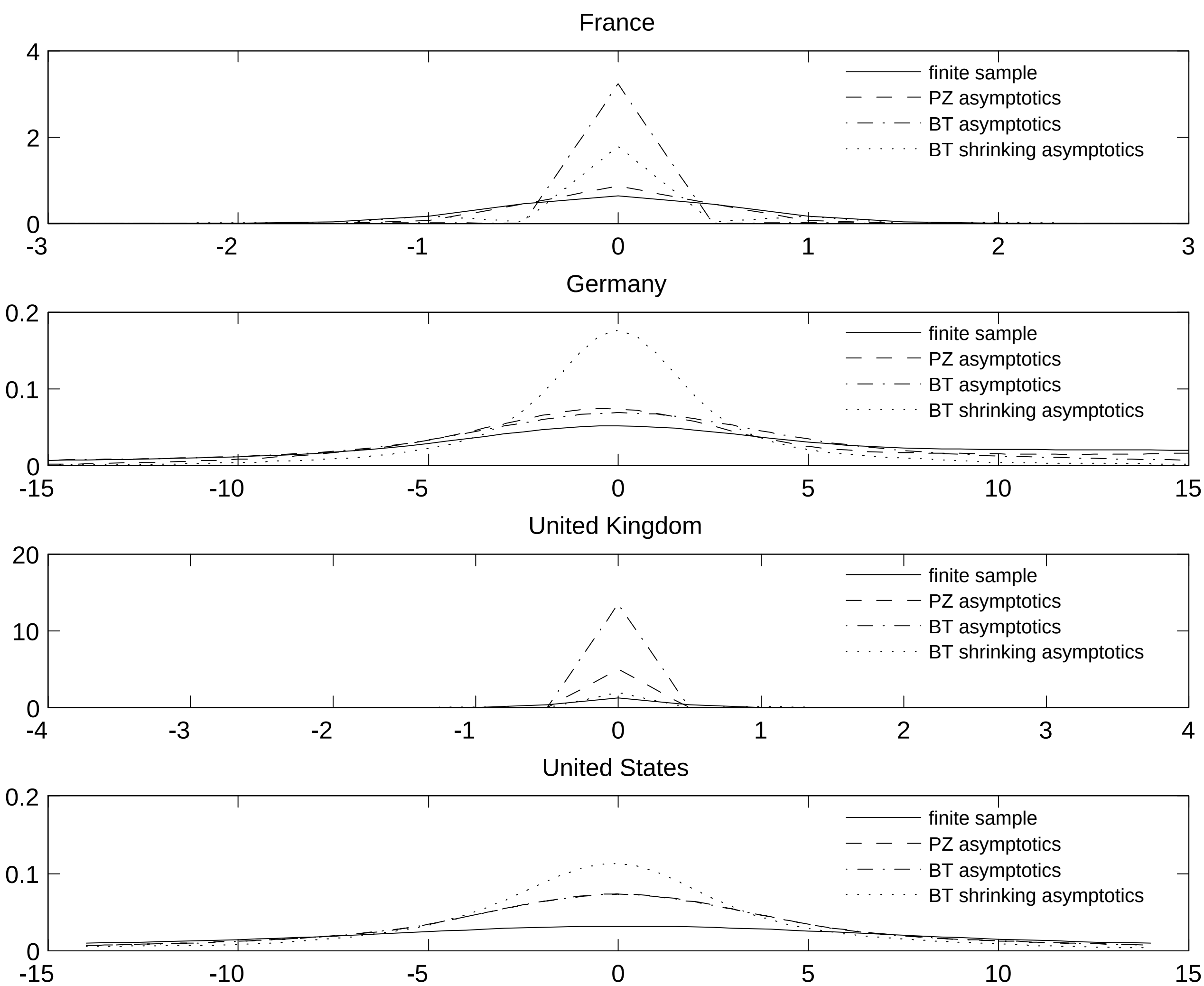

Figure 7: Finite sample versus asymptotic approximations: Empirically calibrated comparisons