Asymptotic distribution theory for break point estimators in models estimated via 2SLS

Alastair R. Hall

University of Manchester[1]

Sanggohn Han

Hyundai Research Institute

and

Otilia Boldea

Tilburg University

July 4, 2008

[1]Corresponding author. Economics, SoSS, University of Manchester, Manchester M13 9PL, UK. Email: alastair.hall@manchester.ac.uk

Abstract

In this paper, we present a limiting distribution theory for the break point estimator in a linear regression model estimated via Two Stage Least Squares under two different scenarios regarding the magnitude of the parameter change between regimes. First, we consider the case where the parameter change is of fixed magnitude; in this case the resulting distribution depends on distribution of the data and is not of much practical use for inference. Second, we consider the case where the magnitude of the parameter change shrinks with the sample size; in this case, the resulting distribution can be used to construct approximate large sample confidence intervals for the break point. The finite sample performance of these intervals are analyzed in a small simulation study and the intervals are illustrated via an application to the New Keynesian Phillips curve.

JEL classification: C12, C13

Keywords: Structural Change, Multiple Break Points, Instrumental Variables Estimation.

1 Introduction

Econometric time series models are based on the assumption that the economic relationships, or "structure", in question are stable over time. However, with samples covering extended periods, this assumption is always open to question and this has led to considerable interest in the development for statistical methods for detecting structural instability.[1]

In designing such methods, it is necessary to specify how the structure may change over time and a popular specification is one in which the parameters of the model are subject to discrete shifts at unknown points in the sample. This scenario can be motivated by the idea of policy regime changes.[2] Within this type of setting, the main concern is to estimate economic relationships in the different regimes and compare them. However, since not all policy changes may impact the economic relationship of interest, an important precursor to this analysis is the identification of the points in the sample, if any, at which the parameters change. This raises the issue of how to perform inference about the location of the so-called "break points", that is the points in the sample at which the parameters change, and motivates the interest to obtain a limiting distribution theory for break point estimators.[3] It is the latter which is the focus of this paper.

There is a literature in time series on the limiting distribution of break point estimators for estimation of changes in mean of process; see Hinckley (1970), Picard (1985), Bhattacharya (1987), Yao (1987), Bai (1994, 1997b). A limiting distribution theory has also been presented in the context of linear regression models estimated via Ordinary Least Squares (OLS). Bai (1997b) considers the case in which there is only one break. He presents two alternative limit theories for the break point estimator. One assumes the magnitude of change between the regimes is fixed; this turns out to depend on distribution of data. The other assumes the magnitude of the parameter change is shrinking with the sample size: this theory approach leads to practical methods for inference about the location of the break point. Bai and Perron (1998) consider

[1]See *inter alia* Andrews and Fair (1988), Ghysels and Hall (1990a,b), Andrews (1993), Andrews and Ploberger (1994), Sowell (1996), Hall and Sen (1999) as well as the other references below.

[2]For example, Bai (1997b) explores the impact of changes in monetary policy on the relationship between the interest rate and the discount factor in the US, and Zhang, Osborn, and Kim (2007) explore the impact of monetary policy changes on the Phillips curve.

[3]The term "change point" is also used in the literature to denote the points in the sample at which the parameter values change.

the case of multiple break points that are estimated simultaneously. They present a limiting distribution theory for the break point estimators based on the assumption that the parameter change is shrinking as the sample size increases; this can be used by practitioners to perform inference about the location of the break points.

One maintained assumption in Bai's (1997b) and Bai and Perron's (1998) analyses is that the regressors are uncorrelated with the errors so that OLS is an appropriate method of estimation. This is a leading case, of course, but there are also many cases in econometrics where the regressors are correlated with the errors and so OLS yields inconsistent estimators. It is desirable, therefore, to develop comparable methods for inference about the break fraction in this more general setting. Hall, Han, and Boldea (2007) consider the estimation of linear regression models with multiple breaks via 2SLS. Their framework allows for the regressors to be correlated with errors and for multiple breaks in the structural equation of interest. Their focus is on developing methods for inference about the parameters in the structural equation of interest. To establish these results, they prove the consistency of the break fraction estimators and also find their rate of convergence under the assumption that the magnitude of the parameter shift is fixed. However, they do not consider the distribution of the break point estimators.

In this paper, we derive the distribution of the break point estimator in the 2SLS model under the assumption that the parameter change is of fixed magnitude. This distribution is shown to be the natural extension of Bai's (1997b) result for OLS estimators and consequently shares the property that it depends on distribution of the data. Therefore, we also explore the distribution of the break point estimator when magnitude of the parameter change shrinks with the sample size. We establish the rate of convergence of the estimator and also a limiting distribution theory. Once again, this distribution is shown to be the natural generalization of the corresponding distribution for OLS estimation. As in the OLS case, this distribution can be used to perform inferences about the break point. We report results from a small simulation study that indicates this limiting distribution provides a good approximation to the finite sample behaviour of the estimated break fraction when the true break fraction is not too large. The use of these intervals is illustrated via an empirical application to the New Keynesian Phillips curve.

An outline of the paper is as follows. Section 2 lays out the model, estimation framework and certain preliminary results. Section 3 presents the large sample distribution of the break point estimator in the case where the change in the parameters across regimes is of fixed magnitude.

Section 4 presents a corresponding theory under the assumption that the magnitude of the parameter change is shrinking as the sample size increases. This section also reports the results from our simulation study. Section 5 contains the empirical application and Section 6 concludes. All proofs are relegated to a mathematical appendix.

2 The model, estimation framework and preliminary results

Consider the following linear structural equation model

$$y_t \;=\; x_t'\beta^0_{x,i} \;+\; z_{1,t}'\beta^0_{z_1,i} \;+\; u_t, \quad i=1,2 \qquad t=T^0_{i-1}+1,\ldots,T^0_i; \qquad T^0_0=0 \text{ and } T^0_2=T \quad (1)$$

in which the vector $x_t = (1, x_{t,2}, \cdots, x_{t,p_1})'$ is correlated with the error term u_t, and $z_{1,t}$ is a $p_2 \times 1$ vector of explanatory variables that are uncorrelated with u_t and includes the intercept. We define $p = p_1 + p_2$ and $\beta^0_i = ({\beta^0_{x,i}}', {\beta^0_{z_1,i}}')'$. The error term, u_t, is assumed to have a mean of zero.

Notice that this equation has a single break point at sample observation $t = T_1$. For the majority of what follows, it is assumed that the total number of break points is known to be one, but the location of the break point is not known. However, we consider the extension of certain results for the one break model to the multiple break model at the end of Section 4.

Suppose that a researcher is interested in estimating the coefficients $\beta^0 = (\beta^{0\prime}_1, \beta^{0\prime}_2)'$. It is well known that, in view of the correlation between x_t and u_t, OLS estimation of (1) would yield inconsistent estimators of the regression parameters. Instead, we consider estimation based on the Two-Stage Least Squares (2SLS) principle. To implement 2SLS, it is necessary to estimate the reduced form for x_t. In this paper, we assume this reduced form is as follows,

$$x_t' = z_t'\Delta_0 + v_t' \quad (2)$$

where $z_t = (z_{t,1}, z_{t,2}, \ldots, z_{t,q})' = (z_{1,t}', z_{2,t}')'$, $\Delta_0 = (\delta_{1,0}, \delta_{2,0}, \ldots, \delta_{p_1,0})$ with dimension $q \times p_1$ and each $\delta_{j,0}$ for $j = 1, \ldots, p_1$ has dimension $q \times 1$. The instrument vector, z_t, is assumed to be uncorrelated with both the error term in the reduced form, v_t, and the error term in the structural equation of interest, u_t.

To estimate β^0, the researcher must also estimate the break point. The estimation process proceeds as follows.

On the first stage of the 2SLS estimation, the reduced form in (2) is estimated via OLS to obtain,

$$\hat{x}_t' = z_t{}'\hat{\Delta}_T = z_t{}'(\sum_{t=1}^{T} z_t z_t{}')^{-1} \sum_{t=1}^{T} z_t x_t{}' \tag{3}$$

On the second stage of the 2SLS estimation, the regression coefficients are estimated for each partition of the sample ($T_0 = 0, T_1, T_2 = T$) such that $T_1 > q$ and $T - T_1 > q$ via

$$\hat{\beta}(T_1) \;=\; \arg\min_{\beta(T_1)} \sum_{i=1}^{2} \sum_{t=T_{i-1}+1}^{T_i} (y_t - \hat{x}_t'\beta_{x,i} - z_{1,t}'\beta_{z_1,i})^2 \Big|_{\beta=\beta(T_1)}$$

where $\beta = (\beta_1', \beta_2')'$. The estimator of the break point is then

$$\hat{T}_1 \;=\; \arg\min_{T_1} \sum_{i=1}^{2} \sum_{t=T_{i-1}+1}^{T_i} (y_t - \hat{x}_t'\beta_{x,i} - z_{1,t}'\beta_{z_1,i})^2 \Big|_{\beta=\beta(T_1)} \tag{4}$$

and the associated estimator of β^0 is

$$\hat{\beta} \;=\; \hat{\beta}(\hat{T}_1) \tag{5}$$

Hall, Han, and Boldea (2007) focus on inference about the parameters β^0 in a generalized version of the above model that allows for m breaks in the sample. Specifically, they derive the limiting distributions of both $\hat{\beta}$ and also various tests for parameter variation. However, to establish these results, they need to prove certain convergence results regarding the break point estimators. These results are also relevant to our analysis of the limiting distribution of the break point estimator, and so we summarize them below in a lemma. Rather than present Hall, Han, and Boldea's (2007) results for the m break model, we specialize them to the single break model being considered here. To present these results, we must state the assumptions under which they are derived. These assumptions are also imposed in our analysis of the limiting distribution of the break point estimator.

Let "$\Longrightarrow$" denote weak convergence in the space $D[0,1]$ under the skorohod metric, and [.] denote the integer part of the quantity in the brackets.

Assumption 1 *Let $b_t = (u_t, v_t')'$ and $\mathcal{F} = \sigma - field\{\ldots, z_{t-1}, z_t, \ldots, b_{t-2}, b_{t-1}\}$. Assume b_t is a martingale difference relative to $\{\mathcal{F}_t\}$ and $sup_t E[\|b_t\|^4] < \infty$.*

Assumption 2 *$rank\{\,\Upsilon_0\,\} = p$ where $\Upsilon_0 = (\Delta_0, \Pi)$, $\Pi' = (I_{p_2}, 0_{p_2\times(q-p_2)})$, I_a denotes the $a \times a$ identity matrix and $0_{a\times b}$ is the $a \times b$ null matrix.*

Assumption 3 *There exists an* $l_0 > 0$ *such that for all* $l > l_0$*, the minimum eigenvalues of* $A_{il} = (1/l)\sum_{t=T_i^0+1}^{T_i^0+l} z_t z_t{}'$ *and of* $A_{il}^* = (1/l)\sum_{t=T_i^0-l}^{T_i^0} z_t z_t{}'$ *are bounded away from zero for all* $i = 1, 2$.

Assumption 4 $T_1^0 = [T\lambda_1^0]$*, where* $0 < \lambda_1^0 < 1$.

Assumption 5 $T^{-1}\sum_{t=1}^{[Tr]} z_t z_t' \xrightarrow{p} Q_{ZZ}(r)$ *uniformly in* $r \in [0,1]$ *where* $Q_{ZZ}(r)$ *is positive definite for any* $r > 0$ *and strictly increasing in* r.

A few comments on these assumptions are in order. Assmption 1 includes the restriction that $E[z_t u_t] = 0$, and thus, that the instruments, z_t, are orthogonal to the structural equation error, u_t. Assumptions 2 and 5 imply the standard rank condition for identification in IV estimation in the linear model[4] because Assumptions 1, 2 and 5 together imply that

$$T^{-1}\sum_{t=1}^{[Tr]} z_t(x_t', z_{1,t}') \;\Rightarrow\; Q_{ZZ}(r)\Upsilon_0 \;=\; Q_{Z,[X,Z_1]}(r) \text{ uniformly in } r \in [0,1]$$

where $Q_{Z,[X,Z_1]}(r)$ has rank equal to p for any $r > 0$. Assumption 3 requires that there are enough observations near the true break points so that they can be identified. Assumption 4 is a standard requirement to allow the development of an asymptotic theory. It implies that each segment increases proportionately as the sample size increases. Assumption 5 is standard for multiple linear regressions. It rules out perfect linear dependencies among z_t.

Within this framework, the break point is indexed by the break fraction λ_1^0. Let $\hat{\lambda}_1 = \hat{T}_1/T$ be the estimator of λ_1. Hall, Han, and Boldea (2007)[Theorems 1 & 2] establish the following properties of this break fraction estimator.

Lemma 1 *Let* y_t *be generated by (1),* x_t *be generated by (2),* $\hat{x}_t$ *be generated by (3) and Assumptions 1-5 hold, then (i)* $\hat{\lambda}_1 \xrightarrow{p} \lambda_1^0$*; (ii) for every* $\eta > 0$*, there exists* C *such that for all large* T*,* $P(T|\hat{\lambda}_1 - \lambda_1^0| > C) < \eta$.

Therefore, the break fraction estimator deviates from the true break fractions by a term of order in probability T^{-1}.

[4] See *e.g.* Hall(2005)[p.35].

3 Fixed magnitude of shift in the regression parameters

In this section, we present a limiting distribution for the break point estimator within our single break model under the assumption that the shift in the regression parameters is of fixed magnitude. To simplify the notation, we now denote the true break point by k_0, that is $k_0 = T_1^0$, and denote the break point estimator by $\hat{k}$, that is $\hat{k} = \hat{T}_1$.

In the previous section, $\hat{k}$ is defined via the minimization of the residual sum of squares on the second stage of the 2SLS estimation. However, for the derivation of the limiting distribution theory, it is more convenient to redefine $\hat{k}$ via the maximization involving quasi-Wald statistics for testing parameter stability in this second stage regression; the prefix "quasi" refers here to that fact that the statistics in question are proportional to the Wald statistics as is described below. To present this alternative - but equivalent - definition, it is necessary to introduce a reparameterization of the regression model in the second stage of the 2SLS.

First consider the case where the true break point is known. In this case, the second stage regression model is:

$$y_t = x_t'\beta_{x,1}^0 + z_{1,t}'\beta_{z_1,1}^0 + \tilde{u}_t, \qquad t = 1, \cdots, k_0 \tag{6}$$

$$y_t = x_t'\beta_{x,2}^0 + z_{1,t}'\beta_{z_1,2}^0 + \tilde{u}_t, \qquad t = k_0 + 1, \cdots, T \tag{7}$$

where

$$\begin{aligned} \tilde{u}_t &= (x_t - \hat{x}_t)'\beta_{x,1}^0 + u_t, \qquad t = 1, \ldots k_0 \\ &= (x_t - \hat{x}_t)'\beta_{x,2}^0 + u_t, \qquad t = k_0 + 1, \ldots T \end{aligned}$$

Equation (7) can be reaparameterized as follows

$$y_t = \hat{x}_t'\beta_{x,1}^0 + z_{1,t}'\beta_{z_1,1}^0 + \hat{x}_t'(\beta_{x,2}^0 - \beta_{x,1}^0) + z_{1,t}'(\beta_{z_1,2}^0 - \beta_{z_1,1}^0) + \tilde{u}_t \tag{8}$$

Equations (6) and (8) can be then combined to yield

$$Y = W\beta_1^0 + W_0\theta^0 + \tilde{U} \tag{9}$$

where $Y = (y_1, \ldots, y_T)'$, $W = (w_1, w_2, \ldots, w_T)'$, $w_t = (\hat{x}_t', z_{1,t}')'$, $W_0 = (0, \ldots, 0, w_{k_0+1}, \ldots, w_T)'$, $\theta^0 = \beta_2^0 - \beta_1^0$, and $\tilde{U} = (\tilde{u}_1, \ldots, \tilde{u}_T)'$. A test of parameter stability, *i.e.* $\beta_1^0 = \beta_2^0$, can be performed by estimating (9) and then calculating the Wald test for $H_0 : \theta^0 = 0$.

Suppose now that k_0 is unknown. In this case, a strategy for testing parameter stability is as follows. For each possible break point, k, estimate the analagous version of (9)

$$Y \;=\; W\beta_1(k) \,+\, W_2\theta(k) \,+\, \text{“}error\text{”}$$

where $W_2 = (0, \ldots, 0, w_{k+1}, \ldots, w_T)'$, and calculate the Wald statistic for the null hypothesis that $\theta(k) = 0$. For our purposes, it turns out to be more convenient to consider inference based on the quasi-Wald statistic,[5]

$$\xi_W(k) = \frac{\hat{\theta}(k)'(W_2' M_W W_2)\hat{\theta}(k)}{\hat{\sigma}^2(k)}$$

where $M_W = I - W(W'W)^{-1}W'$, $\hat{\sigma}^2(k) = S_T(k)/(T-2p)$, $[\hat{\beta}_i(k), \hat{\theta}(k)]$ and $S_T(k)$ is the residual sum of squares from OLS regression of Y on W and W_2. Inference is then based on $\sup_{k\in([\pi T],[(1-\pi)T])} \xi_W(k)$ where $\pi \in (0, 0.5)$.[6]

In Section 2, the estimated break point is defined as

$$\hat{k} = \arg \min_{1\leq k\leq T} S_T(k) \tag{10}$$

The following proposition establishes two alternative characterizations of $\hat{k}$ based on $\{\xi_W(k)\}$.

Proposition 1 *Let y_t, x_t and $\hat{x}_t$ be generated respectively by (1), (2) and (3) then we have:*

(i) $\hat{k} \;=\; \arg\max_{1\leq k\leq T} \xi_W(k)$.

(ii) $\hat{k} \;=\; \arg\max_{1\leq k\leq T} V_T(k)$ *where* $V_T(k) = \hat{\theta}(k)'(W_2' M_W W_2)\hat{\theta}(k)$.

Part (i) of this proposition states that $\hat{k}$ is the break point associated with the $\sup_k \xi_W(k)$ statistic; part (ii) states that $\hat{k}$ is the choice of break point that maximizes the numerator of the Wald statistics. The latter is more useful for our subsequent analysis.

It follows trivially from Proposition 1(ii) that

$$\hat{k} \;=\; \arg\max_k [V_T(k) \,-\, V_T(k_0)], \tag{11}$$

[5]The prefix “quasi” refers to that fact that the denominator is calculated using the residuals from the regression equation on the second step of the 2SLS estimation rather than using the “residuals” form the structural equation evaluated at the 2SLS estimates.

[6]Using the Wald statistic, this approach to testing is proposed by Quandt (1960) in the context of linear models estimated via OLS. Andrews (1993) derives the limiting distribution of this statistic in the case of nonlinear dynamic models estimated via Generalized Method of Moments.

and it is convenient to work with this definition of $\hat{k}$. This leads us to consider $V_T(k) - V_T(k_0)$. It is shown in the appendix that

$$V_T(k) - V_T(k_0) \;=\; -|k_0 - k| G_T(k) + H_T(k), \qquad \text{for all } k \tag{12}$$

where $G_T(k)$ is defined as follows

$$G_T(k) \;=\; \frac{\theta^{0'}[W_0' M_W W_0 - W_0' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W W_0]\theta^0}{|k_0 - k|}, \qquad \text{for } k \neq k_0 \tag{13}$$

$$\;=\; \theta^{0'}\theta^0, \qquad \text{for } k = k_0 \tag{14}$$

and

$$\begin{aligned} H_T(k) \;=\;& 2\theta^{0'}(W_0' M_W W_2)(W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} \;- \\ & 2\theta^{0'}(W_0' M_W \tilde{U}) + \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} \\ & -\tilde{U}' M_W W_0 (W_0' M_W W_0)^{-1} W_0' M_W \tilde{U} \end{aligned} \tag{15}$$

Equation (12) is used to deduce the large sample behaviour of $V_T(k) - V_T(k_0)$ which in turn plays a key role in the derivation of the limiting distribution of break point estimator under the assumption of fixed magnitude of shift in regression parameters. In order to present these results, we introduce the following notation. Define $W_\triangle$ as follows

$$\begin{aligned} W_\triangle &= W_2 - W_0 = (0, \cdots, 0, w_{k+1}, \cdots, w_{k_0}, 0, \cdots, 0)' \quad \text{for } k < k_0 \\ W_\triangle &= -(W_2 - W_0) = (0, \cdots, 0, w_{k_0+1}, \cdots, w_k, 0, \cdots, 0)' \quad \text{for } k > k_0 \\ W_\triangle &= 0 \quad \text{for } k = k_0 \end{aligned}$$

and define Ξ as

$$\begin{aligned} \Xi &= 1 \quad \text{for } k_0 > k \\ \Xi &= -1 \quad \text{for } k_0 < k \end{aligned}$$

Notice that

$$W_2 = W_0 + W_\triangle \cdot \Xi \tag{16}$$

Proposition 2 *Under Assumptions 1 - 5, we have:*

$$V_T(k) - V_T(k_0) = -\theta^{0'} W_\triangle' W_\triangle \theta^0 + 2\theta^{0'} W_\triangle' \tilde{U} \cdot \Xi + o_p(1) \tag{17}$$

The limiting distribution of $\hat{k}$ is determined by the limiting behaviour of the terms on the right hand side of (17). However, without further restrictions, this limiting distribution is intractable. A similar problem is encountered by Bai (1997b) in his analysis of the break points in models estimated by OLS. However, he circumvented this problem by restricting attention to strictly stationary processes.[7] We impose the same restriction here.

Assumption 6 *The process* $\{z_t, u_t, v_t\}_{t=-\infty}^{\infty}$ *is strictly stationary.*

To facilitate the presentation of the limiting distribution of $\hat{k}$, we introduce a stochastic process $R^*(m)$ on the set of integers that is defined as follows:

$$R^*(m) = \begin{cases} R_1(m) & : \quad m < 0 \\ 0 & : \quad m = 0 \\ R_2(m) & : \quad m > 0 \end{cases}$$

with

$$\begin{aligned} R_1(m) &= -\theta^{0'}\Upsilon_0' \sum_{t=m+1}^{0} z_t z_t' \Upsilon_0 \theta^0 + 2\theta^{0'}\Upsilon_0' \left(\sum_{t=m+1}^{0} z_t u_t + \sum_{t=m+1}^{0} z_t v_t' \beta_{x,1}^0 \right) \\ & \quad \text{for } m = -1, -2, \cdots \\ R_2(m) &= -\theta^{0'}\Upsilon_0' \sum_{t=1}^{m} z_t z_t' \Upsilon_0 \theta^0 - 2\theta^{0'}\Upsilon_0' \left(\sum_{t=1}^{m} z_t u_t + \sum_{t=1}^{m} z_t v_t' \beta_{x,2}^0 \right) \\ & \quad \text{for } m = 1, 2, \cdots \end{aligned}$$

We note that if (z_t, u_t, v_t) is independent over t then the process $R^*(m)$ is a two-sided random walk with stochastic drifts.

It is necessary to impose a restriction on the random variables that drive $R^*(m)$.

Assumption 7 $-(z_t'\Upsilon_0\theta^0)^2 \pm 2\theta^{0'}\Upsilon_0'(z_t u_t + z_t v_t' \beta_{x,i}^0)$ *has a continuous distribution for* $i = 1, 2$.

We now present the main result of this section.

Theorem 1 *Under Assumptions 1-7, we have*

$$\hat{k} - k_0 \longrightarrow_d \arg\max_m \quad R^*(m)$$

[7] This approach is also pursued by Bhattacharya (1987), Picard (1985) and Yao (1987).

Remark 1: To derive the probability function of the limiting distribution, it is necessary to know both β^0 and the distribution of (z_t', u_t, v_t').

Remark 2: It is interesting to contrast our Proposition 2 with Bai's (1997b)[Proposition 2] in which the limiting distribution of $\hat{k}$ is presented for the case in which x_t and u_t are uncorrelated and (1) is estimated via OLS. In the latter case, Bai (1997b) shows that $\hat{k} - k_0 \longrightarrow_d \arg\max_m W^*(m)$ where $W^*(m)$ has the same structure as $R^*(m)$ but its behaviour is driven by

$$b(x_t, u_t) \; = \; \theta^{0\prime} x_t' x_t \theta^0 \pm 2 x_t u_t.$$

In contrast, the limiting distribution in Proposition 2 is driven by $b(z_t' \Upsilon_0, u_t + v_t' \beta_{x,i}^0)$. Therefore the limiting distribution in Proposition 2 is the same as would be obtained from Bai's (1997b)[Proposition 2] if y_t is regressed on $E[x_t|z_t]$ and $z_{1,t}$ using OLS.

In view of Remark 1, the limiting distribution in Proposition 2 is not useful for inference in general because of its dependence on unknowns. To circumvent this problem, we consider an alternative asymptotic approximation that is derived under the assumption that the magnitude of the parameter shift is shrinking with the sample size. This is the topic of the next section.

4 Shrinking magnitude of shift in the regression parameters

In this section, we derive the limiting distribution for $\hat{k}$ under the assumption that the magnitude of the parameter change shrinks as the sample size increases. Our analysis follows the same approach as Bai (1997b) used in his derivation of the analogous results for the case where x_t and u_t are uncorrelated and the model is estimated via OLS.

The data generation process is the same as in the previous section with the one exception that the parameter vector in the second regime is now dependent on T and so is denoted by $\beta_{2,T}^0$. We similarly index the magnitude of the change in parameters by T, and write $\theta_T^0 \equiv \beta_{2,T}^0 - \beta_1^0$. It is assumed that θ_T^0 behaves as follows.

Assumption 8 *As* $T \to \infty$, $\theta_T^0 \longrightarrow 0$ *and* $T^{1/2-\alpha}\theta_T^0 \longrightarrow \infty$ *for some* $\alpha \in (0, 1/2)$.

Note that under this assumption, θ_T^0 converges to zero at slower rate than $T^{-1/2}$.

Lemma 1 provides the convergence rate of $\hat{k}$ under the assumption that magnitude of the parameter change is fixed and so does not apply in our setting here. Therefore, we begin by deriving the companion result when the magnitude of change is shrinking.

It is once again convenient to work with the definition of $\hat{k}$ in (11). Note that (12) is still valid but with the following redefinitions of $G_T(k)$ and $H_T(k)$,

$$G_T(k) = \frac{\theta_T^{0'}[W_0'M_W W_0 - W_0'M_W W_2(W_2'M_W W_2)^{-1}W_2'M_W W_0]\theta_T^0}{|k_0 - k|} \qquad \text{for } k \neq k_0 \quad (18)$$

$$G_T(k) = \theta_T^{0'}\theta_T^0 \qquad \text{for } k = k_0 \quad (19)$$

and

$$\begin{aligned} H_T(k) &= 2\theta_T^{0'}(W_0'M_W W_2)(W_2'M_W W_2)^{-1}W_2'M_W\tilde{U} - 2\theta_T^{0'}(W_0'M_W\tilde{U}) \\ &+ \tilde{U}'M_W W_2(W_2'M_W W_2)^{-1}W_2'M_W\tilde{U} \\ &- \tilde{U}'M_W W_0(W_0'M_W W_0)^{-1}W_0'M_W\tilde{U} \end{aligned} \quad (20)$$

To establish the convergence rate, it suffices to consider break points within the set $K(C) = \{k : |k - k_0| > C\|\theta_T^0\|^{-2} \text{ and } T\eta \leq k \leq (1-\eta)T\}$ for a small number $\eta > 0$. The following proposition establishes a property of $G_T(k)$ that is useful in our subsequent analysis.

Proposition 3 *Under Assumptions 2, 3 and 5, there exists a $\gamma > 0$ such that for every $\epsilon > 0$ there exists $C < \infty$ such that*

$$\inf_{k \in K(C)} G_T(k) \geq \gamma ||\theta_T^0||^2$$

with probability at least $1 - \epsilon$.

We also impose the following assumption.

Assumption 9 *For some real number $r > 2$ and constant $A_r < \infty$*

$$E\left\|\sum_{t=i}^{j} w_t\tilde{u}_t\right\|^r \leq A_r(j-i)^{r/2} \qquad \textit{for all } 1 \leq i \leq j \leq T$$

Assumption 9 facilitates the derivation of a bound on $P\left(\sup_{k \geq m} k^{-1}\left\|\sum_{t=1}^{k} w_t\tilde{u}_t\right\| > \zeta\right)$, for every $\zeta > 0$, that plays a crucial role in establishing the following convergence result.

Theorem 2 *Under the Assumptions 1-5, 8 and 9, we have: $\hat{k} = k_0 + O_p(||\theta_T^0||^{-2})$*

Remark 3: Theorem 2 states that the break point estimator converges to the true break point at a rate equal to the inverse of the square of the rate at which the difference between the regimes disappears. Note that this is the same rate of convergence as is exhibited by the corresponding statistic in the case where x_t and u_t are uncorrelated and the model is estimated by OLS; see Bai (1997b)[Proposition 1].

We now turn to the issue of characterizing the limiting distribution of $\hat{k}$. This distribution is deduced in three steps. The first step is to identify the functions of the data that determine the large sample behaviour of $V_T(k) - V_T(k_0)$; see Proposition 4 below. The second step is to use these dominant terms to characterize the limit behaviour of $V_T(k) - V_T(k_0)$. The third step is apply the continuous mapping theorem for the argmax functional to the local weak convergence limit of $V_T(k) - V_T(k_0)$ in order to deduce the limiting distribution of $\hat{k}$. The last two steps are combined in Theorem 3 below.

Since Theorem 2 states that $P(|\hat{k}-k_0| > C\|\theta_T^0\|^{-2}) < \eta$ for every $\eta > 0$, it suffices to consider the behaviour of $V_T(k) - V_T(k_0)$ only over $D(C) \equiv \{k : |k - k_0| \le C\|\theta_T^0\|^{-2}\}$. The following result identifies the terms that determine the large sample behaviour of $V_T(k) - V_T(k_0)$ over $D(C)$.

Proposition 4 *Under Assumptions 1-5, and 8*

$$V_T(k) - V_T(k_0) \;=\; -\theta_T^{0'} W_\triangle' W_\triangle \theta_T^0 + 2\theta_T^{0'} W_\triangle' \tilde{U} \cdot \Xi + o_p(1)$$

for all $k \in D(C)$.

To derive the limit behaviour of $V_T(k) - V_T(k_0)$, we impose the following assumptions on the second moments of the data.

Assumption 10 *$\{z_t\}$ is second-order stationary within each regime such that $Ez_tz_t' = Q_1$ for $t \le k_0$ and $Ez_tz_t' = Q_2$ for $t > k_0$.*

Assumption 11 *For regime $i, i = 1, 2$, the errors $\{u_t, v_t\}$ satisfy*

$$Var\left[\begin{pmatrix} u_t \\ v_t \end{pmatrix} |z_t\right] = \Omega_i = \begin{pmatrix} \sigma_i^2 & \gamma_i' \\ \gamma_i & \Sigma_i \end{pmatrix}$$

where Ω_i is a constant, positive definite matrix. σ_i^2 is a scalar and Σ_i is $p_1 \times p_1$ matrix.

It is also useful to define $\Omega_i^{1/2}$ and $Q_i^{1/2}$ to be the symmetric matrices satisfying $\Omega_i = \Omega_i^{1/2}\Omega_i^{1/2}$ and $Q_i = Q_i^{1/2}Q_i^{1/2}$. Notice that $\Omega_i^{1/2}$ can be decomposed as

$$\Omega_i^{1/2} = \begin{pmatrix} N_1^{i'} \\ N_2^{i'} \end{pmatrix}$$

where $N_1^{i'}$ is a $1 \times (p_1 + 1)$ vector and $N_2^{i'}$ is $p_1 \times (p_1 + 1)$, and that, since $\Omega_i^{1/2}$ is symmetric, we have

$$\Omega_i = \begin{pmatrix} N_1^{i'} N_1^i & N_1^{i'} N_2^i \\ N_2^{i'} N_1^i & N_2^{i'} N_2^i \end{pmatrix} = \begin{pmatrix} \sigma_i^2 & \gamma_i' \\ \gamma_i & \Sigma_i \end{pmatrix}.$$

We also assume that a functional central limit theorem applies to relevant functions of the data.

Assumption 12 *$k_0^{-1/2} \sum_{t=1}^{[rk_0]} \{(u_t, v_t')'\} \otimes z_t \Longrightarrow \tilde{B}_1(r)$ and $(T-k_0)^{-1/2} \sum_{t=1}^{k_0+[r(T-k_0)]} \{(u_t, v_t')'\} \otimes z_t \Longrightarrow \tilde{B}_2(r)$ where $\tilde{B}_i(r)$ is a $q(p_1+1) \times 1$ and Gaussian process for $i = 1, 2$ and $E[\tilde{B}_i(r)\tilde{B}_i(s)'] = min(r, s)V_i$ where V_i is positive definite for $i = 1, 2$.*

If we define $V_i^{1/2}$ to be the symmetric matrix such that $V_i^{1/2}V_i^{1/2} = V_i$ then it follows from Assumptions 10 and 11 that $V_i = \Omega_i \otimes Q_i$ and $V_i^{1/2} = \Omega_i^{1/2} \otimes Q_i^{1/2}$ where $Q_i^{1/2}$ is the symmetric matrix such that $Q_i^{1/2}Q_i^{1/2} = Q_i$.

It is also convenient to reparameterize the shrinking magnitude of parameter change as follows.

Assumption 13 *$\theta_T^0 = \theta_0 \upsilon_T$, where υ_T is a positive number such that $\upsilon_T \to 0$ and $T^{(1/2)-\alpha}\upsilon_T \to \infty$ for some $\alpha \in (0, 1/2)$ and $\theta_0 \neq 0$.*

We now present the limiting distribution of the break point estimator.

Theorem 3 *Under Assumptions 1-5 and 9-12*

$$\frac{(\theta_T^{0'} \Upsilon_0' Q_1 \Upsilon_0 \theta_T^0)^2}{\theta_T^{0'} \Upsilon_0' \Phi_1 \Upsilon_0 \theta_T^0} (\hat{k} - k_0) \xrightarrow{d} \arg\max_s Z(s)$$

where

$$
\begin{aligned}
\xi &= \frac{\theta_0'\Upsilon_0'Q_2\Upsilon_0\theta_0}{\theta_0'\Upsilon_0'Q_1\Upsilon_0\theta_0} \\
\phi &= \frac{\theta_0'\Upsilon_0'\Phi_2\Upsilon_0\theta_0}{\theta_0'\Upsilon_0'\Phi_1\Upsilon_0\theta_0} \\
\Phi_i &= [(N_1^i + N_2^i\beta_{x,i}^0)' \otimes Q_i^{1/2}][(N_1^i + N_2^i\beta_{x,i}^0)' \otimes Q_i^{1/2}]' \text{ for } i = 1,2 \\
Z(s) &= \begin{cases} W_1(-s) - |s|/2 & : \quad s \leq 0 \\ \sqrt{\phi}W_2(s) - \xi s/2 & : \quad s > 0 \end{cases}
\end{aligned}
$$

and $W_i(s)$, $i = 1,2$ *be two independent Brownian motion processes defined on* $[0,\infty)$, *starting at the origin when* $s = 0$.

Remark 4: It is interesting to compare Theorem 3 with Bai's (1997b) Propostion 3 in which corresponding distribution for the case in which x_t and u_t are uncorrelated and the model is estimated by OLS. The two limiting distributions have the same generic structure but the definitions of ξ, ϕ, and Φ_i are different as is the scaling factor of $\hat{k} - k_0$. Inspection reveals that the result in Theorem 3 is equivalent to what would be obtained from applying Bai's (1997b) result to the case in which y_t is regressed on $E[x_t|z_t]$ and $z_{1,t}$ with error $u_t + v_t'\beta_{x,i}^0$.

Remark 5: The density of $\arg\max_s Z(s)$ is characterized by Bai (1997b) and he notes that it is not symmetric if $\xi \neq 1$ or $\phi \neq 1$. It is possible to identify one special case in which $\xi = \phi = 1$, that is where $\Omega_1 = \Omega_2 = \Omega$, $Q_1 = Q_2 = Q$ and $\beta_{x,1}^0 = \beta_{x,2}^0$. Notice that this scenario includes the restriction that the parameters on the endogenous regressors do not change across regimes. The latter represents an important difference between the 2SLS case and the limiting distribution theory of the break fraction in the OLS model with exogenous regressors. For in the latter case, the restrictions for symmetry do not involve the constancy across regimes of any of the regression parameters.[8]

Remark 6: Although Theorem 3 is stated and proved for the one break model, it is easily extended to the multiple break model. Assumption 1 imposes a martingale difference structure on the errors, which is enough to ensure that the sample segments are asymptotically distinct, hence allowing for the analysis of the limiting distribution of the break-dates to be similar to the one break case. Specifically, with appropriate modification of the assumptions to fit the multiple break model, it can be shown that the distributional result in Theorem 3 holds for the n^{th} break

[8] See Bai (1997b)[pp. 554-555].

point, $\hat{k}$, only with quantities pertaining to the first and second regimes in the statement of theorem replaced by the analogous quantities in the n^{th} and $(n+1)^{th}$ regimes respectively.

The distributional result in Theorem 3 can be used to construct confidence intervals for k_0 as follows. To this end, we introduce the following definitions: $\hat{\theta} = \hat{\theta}(\hat{k})$, $\hat{\beta}_1 = \hat{\beta}_1(\hat{k})$, $\hat{\beta}_2 = \hat{\beta}_1 + \hat{\theta}$, $\hat{Q}_1 = \hat{k}^{-1}\sum_{t=1}^{\hat{k}} z_t z_t'$, $\hat{Q}_2 = (T-\hat{k})^{-1}\sum_{t=\hat{k}+1}^{T} z_t z_t'$, $\hat{\Omega}_1 = \hat{k}^{-1}\sum_{t=1}^{\hat{k}} \hat{b}_t \hat{b}_t'$, $\hat{\Omega}_2 = (T-\hat{k})^{-1}\sum_{t=\hat{k}+1}^{T} \hat{b}_t \hat{b}_t'$, $\hat{b}_t = [\hat{u}_t, \hat{v}_t']'$, $\hat{u}_t = y_t - w_t'\hat{\beta}_1$, for $t \leq \hat{k}$ and $\hat{u}_t = y_t - w_t'\hat{\beta}_2$, for $t > \hat{k}$, $\hat{v}_t = (x_t - \hat{\Delta}_T' z_t)$, $\hat{\Omega}_i^{1/2}$ is the symmetric matrix such that $\hat{\Omega}_i = \hat{\Omega}_i^{1/2}\hat{\Omega}_i^{1/2}$,

$$\hat{\Omega}_i^{1/2} = \begin{pmatrix} \hat{N}_1^{i'} \\ \hat{N}_2^{i'} \end{pmatrix}$$

where $\hat{\Omega}_i^{1/2}$ is partitioned conformably with $\Omega_i^{1/2}$,

$$\begin{aligned} \hat{\xi} &= \frac{\hat{\theta}'\hat{\Upsilon}_T'\hat{Q}_2\hat{\Upsilon}_T\hat{\theta}}{\hat{\theta}'\hat{\Upsilon}_T'\hat{Q}_1\hat{\Upsilon}_T\hat{\theta}}, \\ \hat{\phi} &= \frac{\hat{\theta}'\hat{\Upsilon}_T'\hat{\Phi}_2\hat{\Upsilon}_T\hat{\theta}}{\hat{\theta}'\hat{\Upsilon}_T'\hat{\Phi}_1\hat{\Upsilon}_T\hat{\theta}}, \\ \hat{\Phi}_1 &= [(\hat{N}_1^1 + \hat{N}_2^1\hat{\beta}_{x,1})' \otimes \hat{Q}_1^{1/2}][(N_1^1 + N_2^1\hat{\beta}_{x,1})' \otimes \hat{Q}_1^{1/2}]' \\ \hat{\Phi}_2 &= [(\hat{N}_1^2 + \hat{N}_2^2\hat{\beta}_{x,2})' \otimes \hat{Q}_2^{1/2}][(N_1^2 + N_2^2\hat{\beta}_{x,2})' \otimes \hat{Q}_2^{1/2}]', \end{aligned}$$

and $\hat{\Upsilon}_T = [\hat{\Delta}_T, \Pi]$.

It then follows that

$$\left(\hat{k} - \left[\frac{c_2}{\hat{H}}\right] - 1, \hat{k} - \left[\frac{c_1}{\hat{H}}\right] + 1\right) \tag{21}$$

is a $100(1-\alpha)$ percent confidence interval for k_0 where $[\cdot]$ denotes the integer part of the term in the brackets,

$$\hat{H} = \frac{(\hat{\theta}'\hat{\Upsilon}_T'\hat{Q}_1\hat{\Upsilon}_T\hat{\theta})^2}{\hat{\theta}'\hat{\Upsilon}_T'\hat{\Phi}_1\hat{\Upsilon}_T\hat{\theta}}$$

and c_1 and c_2 are respectively the $\alpha/2^{th}$ and $(1-\alpha/2)^{th}$ quantiles for $\arg\max_s Z(s)$ which can be calculated using equations (B.2) and (B.3) in Bai (1997b).

We conclude this section by reporting the results of a small simulation study that is designed to prove some insight into the accuracy of the limiting distribution in Theorem 3 as an approximation to the finite sample of the break fraction estimator. We consider designs with one break and two breaks.

In the single break case, the data generating process for the structural equation is:

$$\begin{aligned} y_t &= [1, x_t]'\beta_1^0 + u_t, \qquad \text{for } t = 1, \ldots, [T/2] \\ &= [1, x_t]'\beta_2^0 + u_t, \qquad \text{for } t = [T/2]+1, \ldots, T \end{aligned} \tag{22}$$

The reduced form equation for the scalar variable x_t is:

$$x_t = [1, z_t]'\delta + v_t, \qquad \text{for } t = 1, \ldots, T \tag{23}$$

where δ is $q \times 1$. The errors are generated as follows: $(u_t, v_t)' \sim IN(0_{2\times 1}, \Omega)$ where the diagonal elements of Ω are equal to one and the off-diagonal elements are equal to 0.5. The instrumental variables, z_t are generated via: $z_t \sim i.i.d\ N(0_{(q-1)\times 1}, I_{q-1})$. The specific parameter values are as follows: (i) $T = 60, 120, 240, 480$; (ii) $(\beta_1^0, \beta_2^0) = ([c, 0.1]', [-c, -0.1]')$, for $c = 0.3, 0.5, 1$; (iii) $q - 1 = 2, 4, 8$; (iv) δ is chosen to yield the population $R^2 = 0.5$ for the regression in (23).[9] For each configuration, 1000 simulations are performed.

Tables 1-3 report the empirical coverage of the 90%, 95% and 99% confidence intervals based on (21). It can be seen that the magnitude of c impacts on the quality of the approximation. If $c = 0.3$ then the confidence intervals are undersized at all samples sizes, although the empirical coverage is close to the nominal level at the largest sample for which $T = 480$; if $c = 0.5$ then the confidence intervals are undersized for $T = 60, 120$ but close to nominal level for $T = 240, 480$; if $c = 1$ then the empirical coverage exceeds the nominal level for the 90% and 95% nominal intervals for $T \geq 60$. For the $c = 1$ case, closer inspection of the empirical distribution of the break-point reveals that most of its probability mass is either at the true break-point or one observation off (only very rarely two or three data points off). Since, by construction, the break-point confidence intervals contains at least three points, if the break-point estimator is one data point off its true value, the confidence interval will necessarily contain the true value. Hence, over-coverage is unavoidable. Finally we note that the number of instruments has no discernable impact on the empirical coverage.

[9] For this model, $\delta_j = \sqrt{R^2/[(q-1)(1-\times R^2)]}$, with δ_j denoting the j^{th} element of δ, $j = 1, \ldots, q$; see Hahn and Inoue (2002).

For the two break case, the data generation process for the structural equation is:

$$\begin{aligned} y_t &= [1, x_t]'\beta_1^0 + u_t, && \text{for } t = 1, \ldots, [T/3] \\ &= [1, x_t]'\beta_2^0 + u_t, && \text{for } t = [T/3]+1, \ldots, [2T/3] \\ &= [1, x_t]'\beta_3^0 + u_t, && \text{for } t = [2T/3]+1, \ldots, T \end{aligned}$$

Three choices for β^0 are considered: $(\beta_1^0, \beta_2^0, \beta_3^0) = ([c, 0.1]', [-c, -0.1]', [c, 0.1]')$ where $c = 0.3, 0.5, 1$. All other aspects of the design are the same as the one break model.

The results are reported in Tables 4-6. It can be seen that the pattern of results is the same as in the single break case, although it is important to remember in making a comparison between the two models that in the two-break model the sub-samples are inevitably smaller. For $c = 0.3$, the empirical coverage tends to be too low - but tends towards the nominal level as the sample size increases, and is very close to the nominal levels at the largest sample size; for $c = 0.5$, the empirical coverage is approximately equal to the nominal level at $T \geq 240$; for $c = 1$, the empirical coverage exceeds the nominal level for the 90% and 95% intervals. For $c = 1$, closer inspection of the empirical distribution of the break point shows similar patterns as for the one break-case: a heavier mass than the nominal coverage at the true break-point or one observation off. Since the confidence intervals are again of at least three data points, over-coverage is not surprising.

Taken together, the two sets of simulation results suggest that the limiting distribution theory based on a shrinking amount of parameter change can provide a reasonable approximation for the designs for which the amount of change is smallest but not in the design with the largest amount of parameter variation. It would be interesting to develop a better understanding of the scenarios for which these intervals are appropriate but this is left to future research.

5 Empirical application

In this section, we assess the stability of the New Keynesian Phillips curve (NKPC), as formulated in Zhang, Osborn, and Kim (2007). The data is from the US, quarterly spanning 1981.1-2005.4. The definitions of the variables are the same as theirs: inf_t is the annualized quarterly growth rate of the GDP deflator, og_t is obtained from the estimates of potential GDP published by the Congressional Budget Office, and $inf^e_{t+1|t}$ is taken from the Michigan inflation expectations survey.[10] With this notation, the structural equation of interest is:

$$inf_t \;=\; c_0 \,+\, \alpha_f inf^e_{t+1|t} \,+\, \alpha_b inf_{t-1} \,+\, \alpha_{og} og_t \,+\, \sum_{i=1}^{3} \alpha_i \Delta inf_{t-i} \,+\, u_t \qquad (24)$$

where inf_t is inflation in (time) period t, $inf^e_{t+1|t}$ denotes expected inflation in period $t+1$ given information available in period t, og_t is the output gap in period t, u_t is an unobserved error term and $\theta = (c_0, \alpha_f, \alpha_b, \alpha_{og}, \alpha_1, \alpha_2, \alpha_3)'$ are unknown parameters. The variables $inf^e_{t+1|t}$ and og_t are anticipated to be correlated with the error u_t, and so (24) is commonly estimated via IV; *e.g.* see Zhang, Osborn, and Kim (2007) and the references therein.

Suitable instruments must be both uncorrelated with u_t and correlated with $inf^e_{t+1|t}$ and og_t. In this context, the instrument vector z_t commonly includes such variables as lagged values of expected inflation, the output gap, the short-term interest rate, unemployment, money growth rate and inflation. Hence, the reduced forms are:

$$inf^e_{t+1|t} \;=\; z_t'\delta_1 \,+\, v_{1,t} \qquad (25)$$

$$og_t \;=\; z_t'\delta_2 \,+\, v_{2,t} \qquad (26)$$

where:

$$z_t' = [1, inf_{t-1}, \Delta inf_{t-1}, \Delta inf_{t-2}, \Delta inf_{t-3}, inf^e_{t|t-1}, og_{t-1}, r_{t-1}, \mu_{t-1}, u_{t-1}]$$

with μ_t, r_t and u_t denoting respectively the M2 growth rate, the three-month Treasury Bill rate and the unemployment rate at time t.

Our sample comprises $T = 100$ observations.[11] Table 7 reports the instability tests for the structural equation, with a cut-off of $\epsilon = 0.15$[12]. The tests indicate evidence of a break at 2000:4, even though the BIC tends to favor the no break model. The parameter estimates for the structural equation are reported below, with standard errors in parentheses.

for 1981.1-2000.4:

$$inf_t \;=\; \underset{(0.89)}{-1.84} \,+\, \underset{(0.11)}{0.76}\, inf^e_{t+1|t} \,-\, \underset{(0.03)}{0.48}\, inf_{t-1} \,+\, \underset{(0.00)}{0.13}\, og_t \,-\, \underset{(0.02)}{0.42}\, \Delta inf_{t-1} \,-\, \underset{(0.01)}{0.36}\, \Delta inf_{t-2} \,-\, \underset{(0.01)}{0.36}\, \Delta inf_{t-3}$$

[10]While Zhang, Osborn, and Kim (2007) consider inflation expectations from different surveys as well, we focus for brevity on the Michigan survey only.

[11]We could have used more observations, but there is evidence of instability in the reduced forms before 1981.1 - see Zhang, Osborn, and Kim (2007).

[12]Different cut-offs yield similar results, indicating that the tests most likely do not suffer from end-of-sample problems.

for 2001.1-2005.4:

$$inf_t = \underset{(4.21)}{-0.69} + \underset{(0.46)}{1.08} inf^e_{t+1|t} - \underset{(0.29)}{0.68} inf_{t-1} + \underset{(0.04)}{0.16} og_t + \underset{(0.15)}{0.44} \Delta inf_{t-1} + \underset{(0.08)}{0.13} \Delta inf_{t-2} + \underset{(0.04)}{0.01} \Delta inf_{t-3}$$

The 99%, 95% and 90% confidence intervals for the break dates are: $[1998.1, 2001, 4]$,[1999.1, 2001.2] and $[1999.3, 2001.2]$, indicating a break around 1999-2001. These results are in line with Zhang, Osborn, and Kim's (2007) findings of a break around $2000-2001$. Using a more heuristical approach for estimating change-points, they find strong evidence for a break at 2000.4, location that coincides with ours. However, due to small sample issues, the parameter estimates in the second sub-sample should be interpreted with care.

6 Concluding remarks

In this paper, we present a limiting distribution theory for the break point estimator in a linear regression model estimated via Two Stage Least Squares under two different scenarios regarding the magnitude of the parameter change between regimes. First, we consider the case where the parameter change is of fixed magnitude; in this case the resulting distribution depends on distribution of the data and is not of much practical use for inference. Second, we consider the case where the magnitude of the parameter change shrinks with the sample size; in this case, the resulting distribution can be used to construct approximate large sample confidence intervals for the break point. These intervals are illustrated via an application to the New Keynesian Phillips curve.

Our results add to the literature on break point distributions. Previous contributions have concentrated on level shifts in univariate time series models or on parameter shits in linear regression models estimated via OLS in which the regressors are uncorrelated with the errors. Within our framework, the regressors of the linear regression model are allowed to be correlated with the error.

One limitation of our framework is that the parameters of the reduced form are assumed to remain constant throughout the sample. While this scenario is viable in block recursive systems of equations such as triangular systems of linear equations, it would be interesting to extend our results to the case in which the coefficients of the reduced form are also allowed to change.

Hall, Han, and Boldea (2007) consider inference in this more general setting. They propose a methodology in which the structural stability of the reduced form is assessed and its break points (if any) estimated first. The estimation of the reduced form then incorporates the information on the estimated break points. Their results assume the estimated break points in the reduced form exhibit similar convergence properties to those established here for the break points in the structural equation of interest. We therefore conjecture that the estimation of the break points in the reduced form would contaminate the limiting distribution of the break points in the structural equation. The verification or contradiction of this conjecture is beyond the scope of the current paper but is a topic of current research.

Mathematical Appendix

Proof of Proposition 1:

The argument is similar to Bai's (1997b) derivation of his equation (5). To show part (i), let $\bar{S}$ denote the sum of squared residuals by regressing Y on W alone. Then we obtain the identity[13]

$$\bar{S} - S_T(k) = \hat{\theta}(k)'(W_2' M_W W_2)\hat{\theta}(k) \tag{27}$$

The quasi-Wald statistic can be written as

$$\xi_W(k) = \left(\frac{T-2p}{p}\right)\left(\frac{\bar{S} - S_T(k)}{S_T(k)}\right)$$

Because $\bar{S}$ does not depend on k and the Wald statistic is a strictly decreasing transformation of $S_T(k)$, it follows immediately

$$\hat{k} = \arg\min_k S_T(k) = \arg\max_k \xi_W(k) \tag{28}$$

Part (ii) follows from (28) and (27). $\diamond$

Derivation of (12):

The LS estimator $\hat{\theta}(k)$ can be expressed as

$$\begin{aligned}
\hat{\theta}(k) &= (W_2' M_W W_2)^{-1} W_2' M_W Y \\
&= (W_2' M_W W_2)^{-1} W_2' M_W [W\beta_1^0 + W_0\theta^0 + \tilde{U}] \\
&= (W_2' M_W W_2)^{-1} W_2' M_W W\beta_1^0 + (W_2' M_W W_2)^{-1} W_2' M_W W_0\theta^0 \\
&\quad + (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} \\
&= (W_2' M_W W_2)^{-1} W_2' M_W W_0\theta^0 + (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U}
\end{aligned}$$

[13] See Amemiya (1985)[pp.31-33].

Similarly, we have

$$\begin{aligned}\hat{\theta}(k_0) &= (W_0' M_W W_0)^{-1} W_0' M_W Y \\ &= (W_0' M_W W_0)^{-1} W_0' M_W [W \beta_1^0 + W_0 \theta^0 + \tilde{U}] \\ &= (W_0' M_W W_0)^{-1} W_0' M_W W_0 \theta^0 + (W_0' M_W W_0)^{-1} W_0' M_W \tilde{U} \\ &= \theta^0 + (W_0' M_W W_0)^{-1} W_0' M_W \tilde{U}.\end{aligned}$$

Thus, it follows that

$$\begin{aligned}V_T(k) &= \hat{\theta}(k)'(W_2' M_W W_2)\hat{\theta}(k) \\ &= [(W_2' M_W W_2)^{-1} W_2' M_W W_0 \theta^0 + (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U}]'(W_2' M_W W_2) \\ &\quad \times [(W_2' M_W W_2)^{-1} W_2' M_W W_0 \theta^0 + (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U}] \\ &= \theta^{0'}(W_0' M_W W_2)(W_2' M_W W_2)^{-1}(W_2' M_W W_2)(W_2' M_W W_2)^{-1} W_2' M_W W_0 \theta^0 \\ &\quad + \theta^{0'}(W_0' M_W W_2)(W_2' M_W W_2)^{-1}(W_2' M_W W_2)(W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} \\ &\quad + \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1}(W_2' M_W W_2)(W_2' M_W W_2)^{-1} W_2' M_W W_0 \theta^0 \\ &\quad + \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1}(W_2' M_W W_2)(W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} \\ &= \theta^{0'}(W_0' M_W W_2)(W_2' M_W W_2)^{-1}(W_2' M_W W_0)\theta^0 + 2\theta^{0'}(W_0' M_W W_2)(W_2' M_W W_2)^{-1} \\ &\quad \times (W_2' M_W \tilde{U}) + (\tilde{U}' M_W W_2)(W_2' M_W W_2)^{-1} W_2' M_W \tilde{U}\end{aligned}$$

and

$$\begin{aligned}V_T(k_0) &= \theta^{0'}(W_0' M_W W_0)(W_0' M_W W_0)^{-1}(W_0' M_W W_0)\theta^0 \\ &\quad + 2\theta^{0'}(W_0' M_W W_0)(W_0' M_W W_0)^{-1}(W_0' M_W \tilde{U}) + (\tilde{U}' M_W W_0)(W_0' M_W W_0)^{-1} \\ &\quad \times W_0' M_W \tilde{U} \\ &= \theta^{0'}(W_0' M_W W_0)\theta^0 + 2\theta^{0'}(W_0' M_W \tilde{U}) + (\tilde{U}' M_W W_0)(W_0' M_W W_0)^{-1} W_0' M_W \tilde{U}\end{aligned}$$

Therefore, we have

$$\begin{aligned}V_T(k) - V_T(k_0) &= \theta^{0'}[W_0' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W W_0 - W_0' M_W W_0]\theta^0 \\ &\quad + 2\theta^{0'}(W_0' M_W W_2)(W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - 2\theta^{0'}(W_0' M_W \tilde{U}) \\ &\quad + \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \tilde{U}' M_W W_0 (W_0' M_W W_0)^{-1} \\ &\quad \times W_0' M_W \tilde{U} \\ &= G_T(k) + H_T(k) \qquad \diamond.\end{aligned}$$

Proof of Proposition 2:

Before we investigate the convergence of $|k_0 - k|G_T(k)$, we let $\Xi = sgn(k_0 - k)$ for notational simplicity.

$$\begin{aligned}
|k_0 - k|G_T(k) &= \theta^{0'}[W_0'M_W W_0 - W_0'M_W W_2(W_2'M_W W_2)^{-1}W_2'M_W W_0]\theta^0 \\
&= \theta^{0'}[(W_2 - \Xi W_\triangle)'M_W(W_2 - \Xi W_\triangle) - (W_2 - \Xi W_\triangle)'M_W W_2 \\
&\quad \times \ (W_2'M_W W_2)^{-1}W_2'M_W(W_2 - \Xi W_\triangle)]\theta^0 \\
&= \theta^{0'}[W_2'M_W W_2 - W_2'M_W W_\triangle \Xi - \Xi W_\triangle' M_W W_2 + \Xi W_\triangle' M_W W_\triangle \Xi \\
&\quad - \ (W_2'M_W W_2 - \Xi W_\triangle' M_W W_2)(W_2'M_W W_2)^{-1}(W_2'M_W W_2 \\
&\quad - W_2'M_W W_\triangle \Xi)]\theta^0 \\
&= \theta^{0'}[W_2'M_W W_2 - W_2'M_W W_\triangle \Xi - \Xi W_\triangle' M_W W_2 + \Xi W_\triangle' M_W W_\triangle \Xi \\
&\quad - \ W_2'M_W W_2 + W_2'M_W W_\triangle \Xi + \Xi W_\triangle' M_W W_2 - \Xi W_\triangle' M_W W_2 \\
&\quad \times \ (W_2'M_W W_2)^{-1}W_2'M_W W_\triangle]\theta^0 \\
&= \theta^{0'}[W_\triangle' M_W W_\triangle - W_\triangle' M_W W_2(W_2'M_W W_2)^{-1}W_2'M_W W_\triangle]\theta^0 \qquad (29)
\end{aligned}$$

Now, we investigate (29) term by term. It is most convenient to begin with the second term on the right hand side of (29). Define $D(C) = \{k : |k - k_0| \leq C\}$. From Lemma 1(ii), it is sufficient to investigate the behaviour of $V_T(k) - V_T(k_0)$ over $D(C)$ for the establishment of the limiting distribution of the break point estimators. Over the set $D(C)$, $W_\triangle' M_W W_2$ consists of a sum of finite terms. Thus, it is clear that

$$||W_\triangle' M_W W_2|| = |k_0 - k|O_p(1) = O_p(1).$$

Since $||W_2'M_W W_2|| = O_p(T)$, the second term on the right hand side of (29) is bounded by $O_p(1) \cdot O_p(T^{-1}) \cdot O_p(1) = o_p(1)$.

Next, consider the first term on the right hand side of (29). We have

$$W_\triangle' M_W W_\triangle = W_\triangle' W_\triangle + W_\triangle' W(W'W)^{-1}W'W_\triangle.$$

Under our assumptions, we have

$$W_\triangle' W(W'W)^{-1}W'W_\triangle = |k_0 - k|O_p(1) \cdot O_p(T^{-1}) \cdot |k_0 - k|O_p(1) = O_p(1) \cdot O_p(T^{-1}) \cdot O_p(1) = o_p(1).$$

Thus, combining these results, it follows from (29) that

$$|k_0 - k|G_T(k) = \theta_0' W_\triangle' W_\triangle \theta_0 + o_p(1) \qquad (30)$$

Next, consider $H_T(k)$. We write

$$H_T(k) \;=\; H_T^{(1)}(k) \,+\, H_T^{(2)}(k) \tag{31}$$

where

$$\begin{aligned} H_T^{(1)}(k) &= 2\theta^{0'}(W_0'M_WW_2)(W_2'M_WW_2)^{-1}W_2'M_W\tilde{U} - 2\theta^{0'}(W_0'M_W\tilde{U}) \\ H_T^{(2)}(k) &= \tilde{U}'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_W\tilde{U} - \tilde{U}'M_WW_0(W_0'M_WW_0)^{-1}W_0'M_W\tilde{U} \end{aligned}$$

First consider $H_T^{(1)}(k)$. We have

$$\begin{aligned} H_T^{(1)}(k) &= 2\theta^{0'}(W_2 - W_\triangle\Xi)'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_W\tilde{U} - 2\theta^{0'}W_0'M_W\tilde{U} \\ &= 2\theta^{0'}(W_2'M_W\tilde{U} - \Xi W_\triangle'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_W\tilde{U}) - 2\theta^{0'}W_0'M_W\tilde{U} \\ &= 2\theta^{0'}W_2'M_W\tilde{U} - 2\Xi\theta^{0'}W_\triangle'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_W\tilde{U} - 2\theta^{0'}W_0'M_W\tilde{U} \\ &= 2\theta^{0'}(W_2 - W_0)'M_W\tilde{U} - 2\Xi\theta^{0'}W_\triangle'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_W\tilde{U} \\ &= \Xi[2\theta^{0'}W_\triangle'M_W\tilde{U} - 2\theta^{0}W_\triangle'M_WW_2(W_2'M_WW_2)^{-1}W_2'M_W\tilde{U}] \end{aligned} \tag{32}$$

We now investigate the convergence of each term on the right hand side of (32) in turn. Since $W_\triangle'W = O_p(1)$ over $D(C)$ and

$$(W'W)^{-1}W'\tilde{U} = (1/\sqrt{T})\cdot(W'W/T)^{-1}W'\tilde{U}/\sqrt{T} = T^{-1/2}O_p(1),$$

it follows that

$$\begin{aligned} W_\triangle'M_W\tilde{U} &= W_\triangle\tilde{U} - W_\triangle'W(W'W)^{-1}W'\tilde{U} \\ &= W_\triangle\tilde{U} - O_p(1)\cdot T^{-1/2}O_p(1) \\ &= W_\triangle\tilde{U} + o_p(1) \qquad \text{over } D(C) \end{aligned}$$

Similarly, we observe that

$$\begin{aligned} W_\triangle'M_WW_2 &= W_\triangle'W_2 - W_\triangle'W(W'W)^{-1}W'W_2 \\ &= O_p(1) - O_p(1)\cdot O_p(T^{-1})\cdot O_p(T) \\ &= O_p(1) \quad \text{over } D(C) \end{aligned}$$

and

$$(W_2'M_WW_2)^{-1}W_2'M_W\tilde{U} \;=\; O_p(T^{-1})\cdot O_p(T^{1/2}) = O_p(T^{-1/2})$$

Thus, the second term on the right hand side of (32) can be written as

$$W_{\triangle}' M_W W_2 \cdot (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} = O_p(1) \cdot O_p(T^{-1/2}) = o_p(1) \quad \text{over } D(C)$$

Thus, we have

$$\begin{aligned} H_T^{(1)}(k) &= \Xi[2\theta^{0'} W_{\triangle}' M_W \tilde{U} - 2\theta^0 W_{\triangle}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U}] \\ &= 2\Xi\theta^{0'} W_{\triangle}' \tilde{U} + o_p(1) \end{aligned}$$

Now consider $H_T^{(2)}(k)$. We have

$$\begin{aligned} H_T^{(2)}(k) &= \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \tilde{U}' M_W W_0 (W_0' M_W W_0)^{-1} W_0' M_W \tilde{U} \\ &= \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \tilde{U}' M_W (W_2 - W_{\triangle}\Xi) \cdot \\ &\quad \cdot [(W_2 - W_{\triangle}\Xi)' M_W (W_2 - W_{\triangle}\Xi)]^{-1} (W_2 - W_{\triangle}\Xi)' M_W \tilde{U} \\ &\equiv M - N \end{aligned} \tag{33}$$

where

$$\begin{aligned} M &\equiv \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} \\ N &\equiv \tilde{U}' M_W (W_2 - W_{\triangle}\Xi)[(W_2 - W_{\triangle}\Xi)' M_W (W_2 - W_{\triangle}\Xi)]^{-1} (W_2 - W_{\triangle}\Xi)' M_W \tilde{U} \end{aligned}$$

To investigate the limiting behavior of $H_T^{(2)}(k)$ over the set $D(C)$, it turns out to be most convenient to check the limit behavior of N first. Using the relationship $W_0 = W_2 - W_{\triangle}\Xi$, N can be written as

$$\begin{aligned} N &= \{\tilde{U}' M_W (W_2 - W_{\triangle}\Xi)/\sqrt{T}\}[(W_2 - W_{\triangle}\Xi)' M_W (W_2 - W_{\triangle}\Xi)/T]^{-1} \\ &\quad \times \{(W_2 - W_{\triangle}\Xi)' M_W \tilde{U}/\sqrt{T}\} \end{aligned} \tag{34}$$

We now consider the limit behaviour of the first term on the right hand side of (34). We have

$$\tilde{U}' M_W (W_2 - W_{\triangle}\Xi)/\sqrt{T} \;=\; \tilde{U}' M_W W_2/\sqrt{T} - \tilde{U}' M_W W_{\triangle}\Xi/\sqrt{T}$$

Since $\tilde{U}' M_W W_{\triangle}\Xi$ is the sum of $|k - k_0|$ terms and the total number of the added terms is bounded over $D(C) = \{k : |k - k_0| \le C\}$, it follows that

$$\begin{aligned}
\tilde{U}' M_W W_\triangle &= \tilde{U}'(I - W(W'W)^{-1}W')W_\triangle \\
&= \tilde{U}' W_\triangle - \tilde{U}' W (W'W)^{-1} W' W_\triangle \\
&= O_p(1)\sqrt{|k - k_0|} - O_p(T^{1/2}) O_p(T^{-1}) O_p(1)|k - k_0| \\
&= O_p(1) - O_p(T^{-1/2}) \\
&= O_p(1) \qquad\qquad \text{over } D(C)
\end{aligned}$$

and so

$$\tilde{U}' M_W W_\triangle \; = \; \tilde{U}' M_W W_2/\sqrt{T} + o_p(1).$$

Similarly, for the second term on the right hand side of (34) we have

$$\begin{aligned}
(W_2 - W_\triangle \Xi)' M_W (W_2 - W_\triangle \Xi)/T &= W_2' M_W W_2/T - (1/T)\{W_2' M_W W_\triangle \Xi \\
&\quad + \; \Xi W_\triangle' M_W W_2 - W_\triangle' M_W W_\triangle\} \\
&= W_2' M_W W_2/T + o_p(1)
\end{aligned}$$

Thus, it follows from (34) that

$$N \; = \; \left(\tilde{U}' M_W W_2/\sqrt{T} + o_p(1)\right) (W_2' M_W W_2/T + o_p(1))^{-1} \left(W_2' M_W \tilde{U}/\sqrt{T} + o_p(1)\right) \tag{35}$$

Therefore, it follows from (33) and (35) that

$$\begin{aligned}
H_T^{(2)}(k) &= M - N \\
&= \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \left(\tilde{U}' M_W W_2/\sqrt{T} + o_p(1)\right) \\
&\quad \times \; (W_2' M_W W_2/T + o_p(1))^{-1} \left(W_2' M_W \tilde{U}/\sqrt{T} + o_p(1)\right) \\
&= \left(\tilde{U}' M_W W_2/\sqrt{T}\right) (W_2' M_W W_2/T)^{-1} \left(W_2' M_W \tilde{U}/\sqrt{T}\right) \\
&\quad - \; \left(\tilde{U}' M_W W_2/\sqrt{T} + o_p(1)\right) (W_2' M_W W_2/T + o_p(1))^{-1} \left(W_2' M_W \tilde{U}/\sqrt{T} + o_p(1)\right) \\
&= o_p(1)
\end{aligned}$$

Thus, we have

$$\begin{aligned}
H_T(k) &= H_T^{(1)}(k) + H_T^{(2)}(k) \\
&= 2\Xi \theta^{0'} W_\triangle' \tilde{U} + o_p(1)
\end{aligned} \tag{36}$$

Finally, it follows from (12), (30) and (36) that

$$V_T(k) - V_T(k_0) \ = \ -\theta^{0'} W'_{\triangle} W_{\triangle}\theta^0 + 2\Xi\theta^{0'} W'_{\triangle}\tilde{U} + o_p(1) \qquad \diamond .$$

Proof of Theorem 1

First, we consider the case of $k < k_0$. Multiplying out, we obtain

$$\begin{aligned} -\theta^{0'} W'_{\triangle} W_{\triangle}\theta^0 + 2\theta^{0'} W'_{\triangle}\tilde{U}\cdot\Xi &= -\theta^{0'} W'_{\triangle} W_{\triangle}\theta^0 + 2\theta^{0'} W'_{\triangle}\tilde{U} \\ &= -\theta^{0'} \sum_{t=k+1}^{k_0} w_t w'_t + 2\theta^{0'} \sum_{t=k+1}^{k_0} w_t \tilde{u}_t \\ &= -\theta^{0'} \sum_{t=k+1}^{k_0} \hat{\Upsilon}'_T z_t z'_t \hat{\Upsilon}_T \theta^0 + 2\theta^{0'} \sum_{t=k+1}^{k_0} \hat{\Upsilon}'_T z_t \tilde{u}_t \end{aligned} \tag{37}$$

where $\hat{\Upsilon}_T = [\hat{\Delta}_T, \Pi]$. By substituting $\tilde{u}_t = u_t + v'_t\beta^0_{x,1} - z'_t[(Z'Z)^{-1}Z'V]\beta^0_{x,1}$ into (37), we obtain

$$\begin{aligned} -\theta^{0'} W'_{\triangle} W_{\triangle}\theta^0 + 2\theta^{0'} W'_{\triangle}\tilde{U}\cdot\Xi &= -\theta^{0'}\hat{\Upsilon}'_T \sum_{t=k+1}^{k_0} z_t z'_t \hat{\Upsilon}_T\theta^0 + 2\theta^{0'}\hat{\Upsilon}'_T (\sum_{t=k+1}^{k_0} z_t u_t \\ &\quad + \sum_{t=k+1}^{k_0} z_t v'_t \beta^0_{x,1}) + 2\theta^{0'}\hat{\Upsilon}'_T \sum_{t=k+1}^{k_0} z_t z'_t (Z'Z)^{-1} Z'V\beta^0_{x,1} \\ &= -\theta^{0'}\Upsilon'_0 \sum_{t=k+1}^{k_0} z_t z'_t \Upsilon_0\theta^0 + 2\theta^{0'}\Upsilon'_0 (\sum_{t=k+1}^{k_0} z_t u_t \\ &\quad + \sum_{t=k+1}^{k_0} z_t v'_t \beta^0_{x,1}) + o_p(1) \qquad \text{over } D(C) \end{aligned} \tag{38}$$

where the last equality comes from the following convergence result over $D(C)$

$$\sum_{t=k+1}^{k_0} z_t z'_t (Z'Z)^{-1} Z'V = |k_0 - k| O_p(1)\cdot O_p(T^{-1})\cdot O_p(T^{1/2}) = O_p(T^{-1/2}) = o_p(1).$$

From Assumption 6 it follows that $\{z_t, u_t, v_t\}_{t=k+1}^{k_0}$ and $\{z_t, u_t, v_t\}_{t=k-k_0+1}^{0}$ have the same joint distribution. Therefore, (38) has the same distribution as $R_1(k - k_0)$ over $D(C)$.

Similarly, for $k > k_0$

$$\begin{aligned} -\theta^{0'} W'_{\triangle} W_{\triangle}\theta^0 + 2\theta^{0'} W'_{\triangle}\tilde{U}\cdot\Xi &= -\theta^{0'} W'_{\triangle} W_{\triangle}\theta^0 - 2\theta^{0'} W'_{\triangle}\tilde{U} \\ &= -\theta^{0'}\Upsilon'_0 \sum_{t=k_0+1}^{k} z_t z'_t \Upsilon_0\theta^0 - 2\theta^{0'}\Upsilon'_0 (\sum_{t=k_0+1}^{k} z_t u_t \\ &\quad + \sum_{t=k_0+1}^{k} z_t v'_t \beta^0_{x,2}) + o_p(1) \end{aligned}$$

which has the same distribution as $R_2(k-k_0)$ over $D(C)$.

Thus, Proposition 2 implies that $V_T(k)-V_T(k_0)$ converges in distribution to $R^*(k-k_0)$ over the bounded set $D(C)$. Let $\hat{k}_C = \arg\max_{|k-k_0|\leq C} V_T(k)-V_T(k_0)$ and $m^*_C = \arg\max_{|m|\leq C} R^*(m)$. The uniform convergence of $V_T(k)-V_T(k_0)$ to $R^*(k-k_0)$ on any bounded set of integers (i.e. the difference $|k-k_0|$ is bounded) implies that $\hat{k}_C - k_0 \longrightarrow_d m^*_C$, and so,

$$|P(\hat{k}_C - k_0 = j) - P(m^*_C = j)| < \epsilon, \qquad \text{for all large } T \text{ and all } |j| \leq C. \tag{39}$$

To complete the proof, we must show that this convergence in distribution holds for the whole range and not just the bounded set $D(C)$. This is established as follows.

From Assumption 7, it follows that the limit distribution $R^*(m)$ is continuous. Thus, the process $R^*(m)$ has a unique maximum with probability one because $P(R^*(m)=R^*(m'))=0$ for $m \neq m'$. Now define $m^* = \arg\max_m R^*(m)$. Since $\theta^{0\prime}\Upsilon_0' \sum_{t=m+1}^{0} z_t z_t' \Upsilon_0 \theta^0 = O_p(m)$, it dominates $\theta^{0\prime}\Upsilon_0' \sum_{t=m+1}^{0} z_t u_t + \sum_{t=m+1}^{0} z_t v_t' \beta^0_{x,i} = O_p(m^{1/2})$. Similarly $\theta^{0\prime}\Upsilon_0' \sum_{t=1}^{m} z_t z_t' \Upsilon_0 \theta^0 = O_p(m)$ dominates $\theta^{0\prime}\Upsilon_0' \sum_{t=1}^{m} z_t u_t + \sum_{t=1}^{m} z_t v_t' \beta^0_{x,i} = O_p(m^{1/2})$. Therefore, we have $R^*(m) \longrightarrow -\infty$ with probability tending to 1 as $|m| \longrightarrow \infty$. Thus, m^* is $O_p(1)$. Therefore, we have that for every $\epsilon > 0$, there exists $C_1 < \infty$ such that

$$P(|m^*| > C_1) < \epsilon, \qquad \text{for all large } T. \tag{40}$$

From Lemma 1(ii), it follows that

$$P(|\hat{k} - k_0| > C_2) < \epsilon, \qquad \text{for all large } T. \tag{41}$$

Now, if $|\hat{k}-k_0| \leq C$ where $C = \max\{C_1, C_2\}$, then $\hat{k} = \hat{k}_C$ and if $|m^*| \leq C$ then $m^* = m^*_C$. For the next step in the argument, it is convenient to define three events, the union of which covers the whole sample space: these events are $\{|\hat{k}-k_0| \leq C \text{ and } |m^*| \leq C\}$, $\{|\hat{k}-k_0| > C\}$ and $\{|m^*| > C\}$. Notice that the first event $\{|\hat{k}-k_0| \leq C \text{ and } |m^*| \leq C\}$ is equivalent to the event $\{\hat{k} = \hat{k}_C \text{ and } m^* = m^*_C\}$ by the definition of $\hat{k}_C$ and m^*_C. Thus, when this event happens, we have the equality $P(\hat{k}-k_0 = j) - P(m^* = j) = P(\hat{k}_C - k_0 = j) - P(m^*_C = j)$. Since the union of other two events is the complement of the first event, it follows that

$$\begin{aligned} |P(\hat{k} - k_0 = j) - P(m^* = j)| &\leq |P(\hat{k}_C - k_0 = j) - P(m^*_C = j)| \\ &\quad + P(|\hat{k} - k_0| > C) + P(|m^*| > C) \\ &< 3\epsilon \end{aligned}$$

To complete the proof, note that ϵ can be arbitrarily small and C can be arbitrarily large. $\diamond$

To prove Proposition 3, we need the following lemma.

Lemma A.1 *The following two inequalities hold:*

$$\begin{aligned} &W_0' M_W W_0 - W_0' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W W_0 \\ &\geq W_\triangle' W_\triangle (W_2' W_2)^{-1} W_0' W_0 \quad \text{for } k < k_0 \end{aligned} \tag{42}$$

$$\begin{aligned} &W_0' M_W W_0 - W_0' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W W_0 \\ &\geq W_\triangle' W_\triangle (W'W - W_2' W_2)^{-1} (W'W - W_0' W_0) \quad \text{for } k \geq k_0 \end{aligned} \tag{43}$$

Proof of Lemma A.1:

Let $H = (W_2'W_2)^{-1} - (W'W)^{-1}$. First consider the case in which $k \leq k_0$. Since

$$\begin{aligned} W_0' M_W W_2 &= W_0'(I - W(W'W)^{-1}W')W_2 \\ &= W_0'W_2 - W_0'W(W'W)^{-1}W'W_2 = W_0'W_0 - W_0'W_0(W'W)^{-1}W_2'W_2 \\ &= W_0'W_0(W_2'W_2)^{-1}W_2'W_2 - W_0'W_0(W'W)^{-1}W_2'W_2 \\ &= W_0'W_0[(W_2'W_2)^{-1} - (W'W)^{-1}]W_2'W_2 = W_0'W_0 H W_2'W_2 \end{aligned}$$

and

$$\begin{aligned} W_2' M_W W_2 &= W_2'(I - W(W'W)^{-1}W')W_2 \\ &= W_2'W_2 - W_2'W(W'W)^{-1}W'W_2 = W_2'W_2 - W_2'W_2(W'W)^{-1}W_2'W_2 \\ &= W_2'W_2(W_2'W_2)^{-1}W_2'W_2 - W_2'W_2(W'W)^{-1}W_2'W_2 \\ &= W_2'W_2[(W_2'W_2)^{-1} - (W'W)^{-1}]W_2'W_2 = W_2'W_2 H W_2'W_2, \end{aligned}$$

it follows that

$$W_0' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W W_0 = W_0' W_0 H W_2' W_2 (W_2' W_2 H W_2' W_2)^{-1} \times W_2' W_2 H W_0' W_0 \quad (44)$$

Define $A = H^{1/2} W_2' W_2$. Since $I - A(A'A)^{-1}A'$ is a projection matrix, we have $I - A(A'A)^{-1}A' \geq 0$. Therefore, putting $C = (W_0'W_0)H^{1/2}(I - A(A'A)^{-1}A')H^{1/2}(W_0'W_0)$, we have

$$\begin{aligned} C &= W_0'W_0HW_0'W_0 - W_0'W_0H^{1/2}H^{1/2}W_2'W_2(W_2'W_2H^{1/2}H^{1/2}W_2'W_2)^{-1}W_2'W_2H^{1/2} \\ &\quad \times\ H^{1/2}W_0'W_0 \\ &= W_0'W_0HW_0'W_0 - W_0'W_0HW_2'W_2(W_2'W_2HW_2'W_2)^{-1}W_2'W_2HW_0'W_0 \quad (45) \\ &\geq 0 \quad (46) \end{aligned}$$

Since the second term in (45) is identical to (44), it suffices to show

$$W_0'M_WW_0 - W_0'W_0HW_0'W_0 \geq W_\triangle'W_\triangle(W_2'W_2)^{-1}W_0'W_0 \quad (47)$$

in order to establish (42). In fact, the equality holds in (47) because the left hand side of (47) is

$$\begin{aligned} W_0'M_WW_0 - W_0'W_0HW_0'W_0 &= W_0'(I - W(W'W)^{-1}W')W_0 - W_0'W_0HW_0'W_0 \\ &= W_0'W_0 - W_0'W(W'W)^{-1}W'W_0 - W_0'W_0HW_0'W_0 \\ &= W_0'W_0 - W_0'W_0(W'W)^{-1}W_0'W_0 - W_0'W_0HW_0'W_0 \\ &= W_0'W_0(W_0'W_0)^{-1}W_0'W_0 - W_0'W_0(W'W)^{-1}W_0'W_0 \\ &\quad - W_0'W_0HW_0'W_0 \\ &= W_0'W_0[(W_0'W_0)^{-1} - (W'W)^{-1} - H]W_0'W_0 \\ &= W_0'W_0[(W_0'W_0)^{-1} - (W_2'W_2)^{-1}]W_0'W_0, \quad (48) \end{aligned}$$

and so, since $W_2'W_2 = W_0'W_0 + W_\triangle'W_\triangle$, we have

$$\begin{aligned} W_0'M_WW_0 - W_0'W_0HW_0'W_0 &= (W_2'W_2 - W_\triangle'W_\triangle)[(W_0'W_0)^{-1} - (W_2'W_2)^{-1}]W_0'W_0 \\ &= [W_2'W_2(W_0'W_0)^{-1} - W_2'W_2(W_2'W_2)^{-1} - W_\triangle'W_\triangle(W_0'W_0)^{-1} \\ &\quad + W_\triangle'W_\triangle(W_2'W_2)^{-1}]W_0'W_0 \\ &= (W_2'W_2 - W_0'W_0 - W_\triangle'W_\triangle) + W_\triangle'W_\triangle(W_2'W_2)^{-1}W_0'W_0 \\ &= W_\triangle'W_\triangle(W_2'W_2)^{-1}W_0'W_0 \end{aligned}$$

Now consider the case with $k \geq k_0$. Define $W_2^* = (w_1, w_2, \cdots, w_k, 0, \cdots, 0)'$, $\quad W_0^* = (w_1, w_2, \cdots, w_{k_0}, 0, \cdots, 0)'$ and $N = (W_2^{*'} W_2^*)^{-1} - (W'W)^{-1}$. It then follows that

$$\begin{aligned}
W_0' M_W W_2 &= W_0' W_2 - W_0' W (W'W)^{-1} W' W_2 \\
&= W_2' W_2 - W_0' W_0 (W'W)^{-1} W_2' W_2 \\
&= W'W - W_2^{*'} W_2^* - (W'W - W_0^{*'} W_0^*)(W'W)^{-1}(W'W - W_2^{*'} W_2^*) \\
&= W'W - W_2^{*'} W_2^* - W'W + W_2^{*'} W_2^* + W_0^{*'} W_0^* - W_0^{*'} W_0^* (W'W)^{-1} W_2^{*'} W_2^* \\
&= W_0^{*'} W_0^* - W_0^{*'} W_0^* (W'W)^{-1} W_2^{*'} W_2^* \\
&= W_0^{*'} W_0^* (W_2^{*'} W_2^*)^{-1} W_2^{*'} W_2^* - W_0^{*'} W_0^* (W'W)^{-1} W_2^{*'} W_2^* \\
&= W_0^{*'} W_0^* [(W_2^{*'} W_2^*)^{-1} - (W'W)^{-1}] W_2^{*'} W_2^* \\
&= W_0^{*'} W_0^* N W_2^{*'} W_2^*
\end{aligned}$$

and

$$\begin{aligned}
W_2' M_W W_2 &= W_2' W_2 - W_2' W (W'W)^{-1} W' W_2 \\
&= W_2' W_2 - W_2' W_2 (W'W)^{-1} W_2' W_2 \\
&= (W'W - W_2^{*'} W_2^*) - (W'W - W_2^{*'} W_2^*)(W'W)^{-1}(W'W - W_2^{*'} W_2^*) \\
&= W'W - W_2^{*'} W_2^* - W'W + W_2^{*'} W_2^* + W_2^{*'} W_2^* - W_2^{*'} W_2^* (W'W)^{-1} W_2^{*'} W_2^* \\
&= W_2^{*'} W_2^* - W_2^{*'} W_2^* (W'W)^{-1} W_2^{*'} W_2^* \\
&= W_2^{*'} W_2^* [(W_2^{*'} W_2^*)^{-1} - (W'W)^{-1}] W_2^{*'} W_2^* \\
&= W_2^{*'} W_2^* N W_2^{*'} W_2^*.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
W_0' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W W_0 &= W_0^{*'} W_0^* N W_2^{*'} W_2^* (W_2^{*'} W_2^* N W_2^{*'} W_2^*)^{-1} \\
&\quad \times W_2^{*'} W_2^* N W_0^{*'} W_0^*
\end{aligned}$$

Let $B = N^{1/2} W_2^{*'} W_2^*$. Using the fact, $I - B(B'B)^{-1}B' \geq 0$, we have that $D = (W_0^{*'} W_0^*) N^{1/2} (I -$

$B(B'B)^{-1}B')N^{1/2}(W_0^{*'}W_0^*)$ satisfies

$$\begin{aligned} D &= W_0^{*'}W_0^*NW_0^{*'}W_0^* - W_0^{*'}W_0^*N^{1/2}N^{1/2}W_2^{*'}W_2^*(W_2^{*'}W_2^*N^{1/2}N^{1/2}W_2^{*'}W_2^*)^{-1} \\ &\quad \times W_2^{*'}W_2^*N^{1/2}N^{1/2}W_0^{*'}W_0^* \\ &= W_0^{*'}W_0^*NW_0^{*'}W_0^* - W_0^{*'}W_0^*NW_2^{*'}W_2^*(W_2^{*'}W_2^*NW_2^{*'}W_2^*)^{-1}W_2^{*'}W_2^*NW_0^{*'}W_0^* \\ &\geq 0 \end{aligned}$$

For the proof of (43), it suffices to show

$$\begin{aligned} W_0'M_WW_0 - W_0^{*'}W_0^*NW_0^{*'}W_0^* &\geq W_\triangle'W_\triangle(W'W - W_2'W_2)^{-1} \\ &\quad \times (W'W - W_0'W_0) \end{aligned} \tag{49}$$

In fact, the equality holds in (49) because the left hand side of (49) is

$$\begin{aligned} W_0'M_WW_0 - W_0^{*'}W_0^*NW_0^{*'}W_0^* &= W_0'(I - W(W'W)^{-1}W')W_0 - W_0^{*'}W_0^*NW_0^{*'}W_0^* \\ &= W_0'W_0 - W_0'W(W'W)^{-1}W'W_0 - W_0^{*'}W_0^*NW_0^{*'}W_0^* \\ &= W_0'W_0 - W_0'W_0(W'W)^{-1}W_0'W_0 - W_0^{*'}W_0^*NW_0^{*'}W_0^* \\ &= (W'W - W_0^{*'}W_0^*) - (W'W - W_0^{*'}W_0^*)(W'W)^{-1} \\ &\quad \times (W'W - W_0^{*'}W_0^*) - W_0^{*'}W_0^*NW_0^{*'}W_0^* \\ &= W'W - W_0^{*'}W_0^* - W'W + W_0^{*'}W_0^* + W_0^{*'}W_0^* \qquad (50) \\ &\quad - W_0^{*'}W_0^*(W'W)^{-1}W_0^{*'}W_0^* - W_0^{*'}W_0^*NW_0^{*'}W_0^* \\ &= W_0^{*'}W_0^* - W_0^{*'}W_0^*(W'W)^{-1}W_0^{*'}W_0^* - W_0^{*'}W_0^*NW_0^{*'}W_0^* \\ &= W_0^{*'}W_0^*[(W_0^{*'}W_0^*)^{-1} - (W'W)^{-1} - N]W_0^{*'}W_0^* \\ &= W_0^{*'}W_0^*[(W_0^{*'}W_0^*)^{-1} - (W_2^{*'}W_2^*)^{-1}]W_0^{*'}W_0^* \qquad (51) \end{aligned}$$

and so, since $W_2^{*'}W_2^* = W_0^{*'}W_0^* + W_\triangle' W_\triangle$, it follows that

$$\begin{aligned} W_0' M_W W_0 - W_0^{*'}W_0^* N W_0^{*'}W_0^* &= (W_2^{*'}W_2^* - W_\triangle' W_\triangle)[(W_0^{*'}W_0^*)^{-1} - (W_2^{*'}W_2^*)^{-1}]W_0^{*'}W_0^* \\ &= [W_2^{*'}W_2^*(W_0^{*'}W_0^*)^{-1} - W_2^{*'}W_2^*(W_2^{*'}W_2^*)^{-1} \\ &\quad - W_\triangle' W_\triangle (W_0^{*'}W_0^*)^{-1} + W_\triangle' W_\triangle (W_2^{*'}W_2^*)^{-1}]W_0^{*'}W_0^* \\ &= (W_2^{*'}W_2^* - W_0^{*'}W_0^* - W_\triangle' W_\triangle) + W_\triangle' W_\triangle (W_2^{*'}W_2^*)^{-1} \\ &\quad \times W_0^{*'}W_0^* \\ &= W_\triangle' W_\triangle (W_2^{*'}W_2^*)^{-1} W_0^{*'}W_0^* \\ &= W_\triangle' W_\triangle (W'W - W_2'W_2)^{-1}(W'W - W_0'W_0) \qquad \diamond . \end{aligned}$$

Proof of Proposition 3:

Suppose $k \leq k_0$. First notice that if $W_\triangle' W_\triangle$ is invertible then the matrix $A(k) = (k_0 - k)^{-1} W_\triangle' W_\triangle (W_2'W_2)^{-1} W_0'W_0$ is symmetric and positive definite because

$$\begin{aligned} A(k) &= (k_0 - k)^{-1} W_\triangle' W_\triangle (W_\triangle' W_\triangle + W_0'W_0)^{-1} W_0'W_0 \\ &= (k_0 - k)^{-1} W_\triangle' W_\triangle \left\{ (W_\triangle' W_\triangle)^{-1} [(W_\triangle' W_\triangle)^{-1} + (W_0'W_0)^{-1}]^{-1} (W_0'W_0)^{-1} \right\} W_0'W_0 \\ &= (k_0 - k)^{-1} [(W_\triangle' W_\triangle)^{-1} + (W_0'W_0)^{-1}]^{-1} \end{aligned}$$

From the symmetry of $A(k)$ and Lemma A.1, it follows that

$$G_T(k) \;\geq\; \theta_T^{0'} A(k) \theta_T^0 \;\geq\; \gamma_T(k) ||\theta_T||^2 \tag{52}$$

where $\gamma_T(k)$ is the minimum eigenvalue of $A(k)$. Therefore, the desired result can be established by showing that, with probability tending to one, $\gamma_T(k)$ is bounded away from zero as $k_0 - k$ increases.

To this end, note that Assumption 3 implies that, for large $k_0 - k$, $W_\triangle' W_\triangle = \sum_{t=k+1}^{k_0} w_t w_t' = \hat{\Upsilon}_T' \sum_{t=k+1}^{k_0} z_t z_t' \hat{\Upsilon}_T$ is positive definite with large probability and so $A(k)$ is invertible with large probability for large $k_0 - k$. Therefore, we can consider $A(k)^{-1}$. Since

$$A(k)^{-1} \;=\; (W_0'W_0)^{-1} W_2'W_2 [(k_0 - k)^{-1} W_\triangle' W_\triangle]^{-1},$$

it follows that

$$||A(k)^{-1}|| \;\leq\; ||(W_0'W_0)^{-1} W_2'W_2|| \cdot ||[(k_0 - k)^{-1} W_\triangle' W_\triangle]^{-1}||$$

Furthermore, we have

$$
\begin{aligned}
||(W_0'W_0)^{-1}W_2'W_2|| &\leq ||(W_0'W_0)^{-1}W'W|| \\
&\leq ||(W_0'W_0)^{-1}|| \cdot ||W'W|| \\
&= ||(\hat{\Upsilon}_T' Z_0' Z_0 \hat{\Upsilon}_T)^{-1}|| \cdot ||\hat{\Upsilon}_T' Z' Z \hat{\Upsilon}_T|| \qquad (53)
\end{aligned}
$$

where $Z_0 = (0, \cdots, 0, z_{k_0+1}, ..., z_T)'$ and $Z = (z_1, ..., z_T)'$ defined. It follows from (53) and Assumptions 2, 3 and 5 that $||(W_0'W_0)^{-1}W_2'W_2||$ is bounded. In addition, the minimum eigenvalue of $(k_0 - k)^{-1}W_\triangle' W_\triangle$ is bounded away from zero by Assumption 3 with large probability and so, $||[(k_0 - k)^{-1}W_\triangle' W_\triangle]^{-1}||$ is bounded with large probability for all large $k_0 - k$. Therefore, it follows that $||A(k)^{-1}||$ is bounded with large probability for all large $k_0 - k$ and hence that $\gamma_T(k)$ is bounded away from zero for all large $k_0 - k$ with large probability.

Suppose now that $k > k_0$. Let

$$
\begin{aligned}
B(k) &= (k_0 - k)^{-1} W_\triangle' W_\triangle (W'W - W_2'W_2)^{-1}(W'W - W_0'W_0) \\
&= (k_0 - k)^{-1} W_\triangle' W_\triangle (W_2^{*'} W_2^*)^{-1} W_0^{*'} W_0^*
\end{aligned}
$$

By similar arguments to $A(k)$, it follows that $B(k)$ is symmetric and is positive definite when $W_\triangle' W_\triangle$ is invertible.

Using Lemma A.1, it follows that

$$
G_T(k) \geq \theta_T^{0'} B(k) \theta_T^0 \geq \gamma_T^*(k) ||\theta_T||^2
$$

where $\gamma_T^*(k)$ is the minimum eigenvalue of $B(k)$. It remains to establish that, with probability tending to one, $\gamma_T^*(k)$ is bounded away from zero as $k_0 - k$ increases.

For large $k_0 - k$, $W_\triangle' W_\triangle = \sum_{t=k+1}^{k_0} w_t w_t' = \hat{\Upsilon}_T' \sum_{t=k+1}^{k_0} z_t z_t' \hat{\Upsilon}_T$ will be positive definite with large probability by Assumption 3. Also we have,

$$
\begin{aligned}
B(k)^{-1} &= (W_0^{*'} W_0^*)^{-1} W_2^{*'} W_2^* [(k_0 - k)^{-1} W_\triangle' W_\triangle]^{-1} \\
||B(k)^{-1}|| &\leq ||(W_0^{*'} W_0^*)^{-1} W_2^{*'} W_2^*|| \cdot ||[(k_0 - k)^{-1} W_\triangle' W_\triangle]^{-1}||
\end{aligned}
$$

and

$$
\begin{aligned}
||(W_0^{*'} W_0^*)^{-1} W_2^{*'} W_2^*|| &\leq ||(W_0^{*'} W_0^*)^{-1} W'W|| \\
&\leq ||(W_0^{*'} W_0^*)^{-1}|| \cdot ||W'W|| \\
&= ||(\hat{\Upsilon}_T' Z_0^{*'} Z_0^* \hat{\Upsilon}_T)^{-1}|| \cdot ||\hat{\Upsilon}_T' Z' Z \hat{\Upsilon}_T|| \qquad (54)
\end{aligned}
$$

where $Z_0^* = (z_1, \cdots, z_{k_0}, 0, \cdots, 0)'$. It follows from (54) and Assumptions 2 and 3 that $||(W_0^{*\prime} W_0^*)^{-1} W_2^{*\prime} W_2^*||$ is bounded. In addition, Assumption 3 implies the minimum eigenvalue of $(k_0 - k)^{-1} W_\triangle' W_\triangle$ is bounded away from zero with large probability. Therefore, $\gamma_T^*(k)$, is bounded away from zero as $k_0 - k$ increases with large probability. $\diamond$

To prove Theorem 2, we need the following lemma.

Lemma A.2 *Under Assumption 9, there exists a $B < \infty$ such that for every $\zeta > 0$ and $m > 0$*

$$P\left(\sup_{k \geq m} \frac{1}{k} \left\| \sum_{t=1}^{k} w_t \tilde{u}_t \right\| > \zeta\right) \leq \frac{B}{\zeta^4 m^2}$$

Proof of Lemma A.2: This follows from Serfling (1970)[Theorem 5.1]. This theorem states that under Assumption 8, for each $\zeta > 0$ there exists a constant $C_\zeta < \infty$ (depending on A_r and K_r) such that

$$P\left(\sup_{k \geq m} \left\| \frac{S_k}{k} \right\| > \zeta\right) \leq C_\zeta \cdot m^{-r/2} \qquad \text{for all } m \geq 1$$

where $C_\zeta = (A_r + K_r)(\frac{\zeta}{2})^{-r}(1 - 2^{-r/2})^{-1}$

Thus,

$$P\left(\sup_{k \geq m} \frac{1}{k} \left\| \sum_{t=1}^{k} w_t \tilde{u}_t \right\| > \zeta\right) \leq C_\zeta \cdot m^{-r/2}$$

By letting $r = 4$, $B = (A_r + K_r)(1/2)^{-r}(1 - 2^{-r/2})^{-1}$, we get the desired maximal inequality

$$P\left(\sup_{k \geq m} \frac{1}{k} \left\| \sum_{t=1}^{k} w_t \tilde{u}_t \right\| > \zeta\right) \leq \frac{B}{\zeta^4 m^2}$$

Proof of Theorem 2:

By definition, $\hat{k} = \arg\max_k V_T(k)$. Thus, $V_T(\hat{k}) \geq V_T(k_0)$. Therefore, it suffices to show that for each $\epsilon > 0$, there exists $C > 0$ such that

$$P(\sup_{k \in K(C)} V_T(k) \geq V_T(k_0)) < \epsilon \tag{55}$$

From (12), it follows that $V_T(k) \geq V_T(k_0)$ is equivalent to

$$(H_T(k)/|k_0 - k|) \geq G_T(k)$$

and so by Proposition 3, it suffices to prove that

$$P\left(\sup_{k \in K(C)} \left| \frac{H_T(k)}{k_0 - k} \right| \geq \gamma ||\theta_T^0||^2\right) < \epsilon. \tag{56}$$

From (20), $H_T(k)$ can be decomposed into two parts as follows

$$H_T(k) \;=\; H_T^{(1)}(k) \,+\, H_T^{(2)}(k)$$

where $H_T^{(1)}(k)$ and $H_T^{(2)}(k)$ are (re)defined as

$$H_T^{(1)}(k) \;=\; 2\theta_T^{0'} W_0' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - 2\theta_T^{0'} W_0' M_W \tilde{U}$$

and

$$H_T^{(2)}(k) \;=\; \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \tilde{U}' M_W W_0 (W_0' M_W W_0)^{-1} W_0' M_W \tilde{U}$$

First consider the expression $H_T^{(1)}(k)$. Recalling the definition $W_0 = W_2 - W_\triangle \Xi$ from (16), $H_T^{(1)}(k)$ can be transformed into

$$H_T^{(1)}(k) = \Xi[2\theta_T^{0'} W_\triangle' M_W \tilde{U} - 2\theta_T^{0} W_\triangle' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U}] \tag{57}$$

by similar logic to the derivation of (32) (except with θ^0 replaced by θ_T^0). The first term in the square brackets in (57) can be written as

$$W_\triangle' M_W \tilde{U} = W_\triangle' \tilde{U} - W_\triangle' W (W'W)^{-1} W' \tilde{U}$$

It is clear that $W_\triangle' W = |k_0 - k| O_p(1)$ and $(W'W)^{-1} W' \tilde{U} = (1/\sqrt{T})(W'W/T)^{-1} W' \tilde{U}/\sqrt{T} = T^{-1/2} O_p(1)$. Thus, the first term at (57) can be written as

$$\begin{aligned} \theta_T^{0'} W_\triangle' M_W \tilde{U} \;&=\; \theta_T^{0'} W_\triangle' \tilde{U} - ||\theta_T^0|| \cdot |k_0 - k| O_p(1) \cdot T^{-1/2} O_p(1) \\ &=\; \theta_T^{0'} W_\triangle' \tilde{U} - |k_0 - k| T^{-1/2} ||\theta_T^0|| O_p(1) \end{aligned} \tag{58}$$

Now, we consider each factor in the second term within the square bracket in (57). We have

$$\begin{aligned} W_\triangle' M_W W_2 \;&=\; W_\triangle' W_2 - W_\triangle' W (W'W)^{-1} W' W_2 \\ &=\; |k_0 - k| O_p(1) - |k_0 - k| O_p(1) \cdot O_p(T^{-1}) \cdot O_p(T) \\ &=\; |k_0 - k| O_p(1) \end{aligned}$$

and

$$(W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} \;=\; O_p(T^{-1}) \cdot O_p(T^{1/2}) = O_p(T^{-1/2})$$

Thus, the second term in (57) can be written satisfies

$$2\Xi\theta_T^{0'} W_\triangle' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} \;=\; 2||\theta_T^0|| \cdot |k_0 - k| O_p(1) \cdot O_p(T^{-1/2}) \Xi \tag{59}$$

Therefore, combining the results in (58) and (59) over the set $K(C)$

$$\begin{aligned} H_T^{(1)}(k) &= \Xi\left(2\theta_T^{0'} W_\triangle' \tilde{U} - 2|k_0 - k|T^{-1/2}||\theta_T^0||O_p(1) - 2|k_0 - k|T^{-1/2}||\theta_T^0||O_p(1)\right) \\ &= 2\theta_T^{0'} W_\triangle' \tilde{U}\Xi + |k_0 - k|T^{-1/2}||\theta_T^0||O_p(1)\Xi \end{aligned} \quad (60)$$

Now, consider the expression

$$\begin{aligned} H_T^{(2)}(k) &= \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \tilde{U}' M_W (W_2 - W_\triangle \Xi) \\ &\quad \times [(W_2 - W_\triangle \Xi)' M_W (W_2 - W_\triangle \Xi)]^{-1} (W_2 - W_\triangle \Xi)' M_W \tilde{U} \end{aligned}$$

The first term of $H_T^{(2)}(k)$ is easily seen to be bounded as

$$\begin{aligned} \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} &= O_p(T^{1/2}) \cdot O_p(T^{-1}) \cdot O_p(T^{1/2}) \\ &= O_p(1) \end{aligned} \quad (61)$$

The second term of $H_T^{(2)}(k)$ can be expanded further as follows

$$\begin{aligned} &\tilde{U}' M_W (W_2 - W_\triangle \Xi) \cdot [(W_2 - W_\triangle \Xi)' M_W (W_2 - W_\triangle \Xi)]^{-1} (W_2 - W_\triangle \Xi)' M_W \tilde{U} \\ &= \left(\tilde{U}' M_W W_2 - \tilde{U}' M_W W_\triangle \Xi\right) [W_2' M_W W_2 - W_2' M_W W_\triangle \Xi - \Xi W_\triangle' M_W W_2 \\ &\quad + W_\triangle' M_W W_\triangle]^{-1} \left(W_2' M_W \tilde{U} - W_\triangle' M_W \tilde{U} \Xi\right) \end{aligned} \quad (62)$$

Investigating the limit behavior of each term on the right hand side of (62) over the set $K(C)$, we have the following convergence results:

$$\begin{aligned} \tilde{U}' M_W W_2 &= O_p(T^{1/2}) \\ \tilde{U}' M_W W_\triangle &= \tilde{U}' W_\triangle - \tilde{U}' W (W'W)^{-1} W' W_\triangle \\ &= \sqrt{|k_0 - k|} \cdot O_p(1) - O_p(T^{1/2}) \cdot O_p(T^{-1}) \cdot |k_0 - k| O_p(1) \\ &= \sqrt{|k_0 - k|} O_p(1) - |k_0 - k| O_p(T^{-1/2}) \\ &= O_p(T^{1/2}) O_p(1) - O_p(T) O_p(T^{-1/2}) \\ &= O_p(T^{1/2}) - O_p(T^{1/2}) = O_p(T^{1/2}) \\ W_2' M_W W_\triangle &= O_p(1) \cdot |k_0 - k| = O_p(T) \\ W_2' M_W W_2 &= O_p(T) \end{aligned}$$

and

$$\begin{aligned} W'_\triangle M_W W_\triangle &= W'_\triangle W_\triangle - W'_\triangle W(W'W)^{-1}W'W_\triangle \\ &= |k_0-k|O_p(1) - |k_0-k|O_p(1)\cdot O_p(T^{-1})\cdot |k_0-k|O_p(1) \\ &= |k_0-k|O_p(1) - |k_0-k|^2 O_p(T^{-1}) \\ &= O_p(T)O_p(1) - O_p(T^2)O_p(T^{-1}) \\ &= O_p(T) - O_p(T) = O_p(T) \end{aligned}$$

Thus, the second term of $H_T^{(2)}(k)$ is

$$\begin{aligned} &\left(O_p(T^{1/2}) - O_p(T^{1/2})\right)\left[O_p(T) - O_p(T) - O_p(T) + O_p(T)\right]^{-1}\left(O_p(T^{1/2}) - O_p(T^{1/2})\right) \\ &= O_p(T^{1/2})O_p(T^{-1})O_p(T^{1/2}) \\ &= O_p(1) \end{aligned} \tag{63}$$

Finally, combining (31), (60), and (61), we obtain

$$\begin{aligned} H_T(k) &= \left(2\theta_T^{0'}W'_\triangle\tilde{U}\Xi + |k_0-k|T^{-1/2}||\theta_T^0||O_p(1)\Xi\right) + (O_p(1)+O_p(1)) \\ &= \left(2\theta_T^{0'}W'_\triangle\tilde{U}\Xi + |k_0-k|T^{-1/2}||\theta_T^0||O_p(1)\Xi\right) + O_p(1) \end{aligned}$$

Thus, over the set $K(C)$, we have

$$\frac{H_T(k)}{|k_0-k|} = \frac{2}{|k_0-k|}\theta_T^{0'}W'_\triangle\tilde{U}\Xi + T^{-1/2}||\theta_T^0||O_p(1) + \frac{O_p(1)}{|k_0-k|} \tag{64}$$

We now show that (56) follows from (64). This is proved by investigating the probability limit behavior of the supremum of each term in (64) over $K(C)$ and then using the Triangle inequality to establish (56).

By the symmetry of the argument, it suffices to consider only the case for $k < k_0$. Now consider each term on the right hand side of (64) in turn.

(i) Consider $\frac{2}{k_0-k}\theta_T^{0'}W'_\triangle\tilde{U}$. We have:

$$P\left(\sup_{k\in K(C)}\left|\frac{2}{k_0-k}\theta_T^{0'}W'_\triangle\tilde{U}\right| > \frac{\gamma||\theta_T^0||^2}{3}\right) \le P\left(\sup_{k\le k_0 - C||\theta_T^0||^{-2}}\left|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0} w_t\tilde{u}_t\right| > \frac{\gamma||\theta_T^0||}{6}\right)$$

Using Lemma A.2 with $\zeta = \gamma||\theta_T^0||/6$ and $m = C||\theta_T^0||^{-2}$, the right-hand side above is bounded by

$$B\frac{6^4}{\gamma^4||\theta_T^0||^4}\cdot\frac{1}{C^2||\theta_T^0||^{-4}} = B\frac{1296}{\gamma^4 C^2} < \frac{\epsilon}{3}$$

for large C.

(ii) Consider $T^{-1/2}||\theta_T^0||O_p(1)$. We have:

$$P\left(\left|T^{-1/2}||\theta_T^0||O_p(1)\right| > \frac{\gamma||\theta_T^0||^2}{3}\right) = P\left(\frac{|O_p(1)|}{T^{1/2}||\theta_T^0||} > \frac{\gamma}{3}\right) < \frac{\epsilon}{3}$$

because $(T^{1/2}||\theta_T^0||)^{-1} \longrightarrow 0$.

(iii) By imposing the restriction $k < k_0$ to the set $K(C)$, we get $k \leq k_0 - C||\theta_T^0||^{-2}$ which implies

$$\left|\frac{1}{k_0 - k}\right| \leq \frac{1}{C}||\theta_T^0||^2$$

Thus, for $k < k_0$

$$\begin{aligned} P\left(\sup_{k\in K(C)}\left|\frac{O_p(1)}{k_0 - k}\right| > \frac{\gamma||\theta_T^0||^2}{3}\right) &< P\left(\sup_{k\in K(C)}\|\theta_T^0\|^2\left|\frac{O_p(1)}{C}\right| > \frac{\gamma||\theta_T^0||^2}{3}\right) \\ &= P\left(\left|\frac{O_p(1)}{C}\right| > \frac{\theta}{3}\right) \\ &< \frac{\epsilon}{3} \qquad \text{for large } C. \end{aligned}$$

Combining (i)-(iii), we obtain:

$$\begin{aligned} P\left(\sup_{k\in K(C)}\left|\frac{H_T(k)}{k_0 - k}\right| > \gamma||\theta_T^0||^2\right) &\leq P\left(\sup_{k\in K(C)}|\frac{2}{k_0 - k}\theta_T^{0'}W_\triangle'\tilde{U}| > \frac{\gamma||\theta_T^0||^2}{3}\right) \\ &+ P\left(\left|T^{-1/2}||\theta_T^0||O_p(1)\right| > \frac{\gamma||\theta_T^0||^2}{3}\right) + P\left(\sup_{k\in K(C)}\left|\frac{O_p(1)}{k_0 - k}\right| > \frac{\gamma||\theta_T^0||^2}{3}\right) \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \qquad \diamond. \end{aligned}$$

Proof of Proposition 4:

First consider $|k_0 - k|G_T(k)$. We have

$$\begin{aligned} |k_0 - k|G_T(k) &= \theta_T^{0'}[W_0'M_W W_0 - W_0'M_W W_2(W_2'M_W W_2)^{-1}W_2'M_W W_0]\theta_T^0 \\ &= \theta_T^{0'}[(W_2 - \Xi W_\triangle)'M_W(W_2 - \Xi W_\triangle) - (W_2 - \Xi W_\triangle)'M_W W_2 \\ &\quad \times (W_2'M_W W_2)^{-1}W_2'M_W(W_2 - \Xi W_\triangle)]\theta_T^0 \\ &= \theta_T^{0'}[W_\triangle'M_W W_\triangle - W_\triangle'M_W W_2(W_2'M_W W_2)^{-1}W_2'M_W W_\triangle]\theta_T^0 \qquad (65) \end{aligned}$$

First, consider the second term on the right hand side of (65). Since $\|W_\triangle'M_W W_2\| = O_p(1)\|\theta_T^0\|^{-2}$ and $\|(W_2'M_W W_2)^{-1}\| = O_p(T^{-1})$ over the set $D(C)$, we have under Assumption 8 that

$$\begin{aligned} \|\theta_T^{0'}W_\triangle'M_W W_2(W_2'M_W W_2)^{-1}W_2'M_W W_\triangle\theta_T^0\| &\leq \|\theta_T^0\|^2 \cdot O_p(1)\|\theta_T^0\|^{-2} \cdot O_p(T^{-1}) \cdot O_p(1)\|\theta_T^0\|^{-2} \\ &= o_p(1). \qquad (66) \end{aligned}$$

Next, consider the first term on the right hand side of in (65), $\theta_T^{0'} W_\triangle' M_W W_\triangle \theta_T^0$. Since

$$\theta_T^0 W_\triangle' W(W'W)^{-1}W'W_\triangle \theta_T^0 = \|\theta_T^0\|^2 \cdot O_p(1)\|\theta_T^0\|^{-2} \cdot O_p(T^{-1}) \cdot O_p(1)\|\theta_T^0\|^{-2} = o_p(1)$$

under Assumption 8, we have

$$\theta_T^{0'} W_\triangle' M_W W_\triangle \theta_T^0 \;=\; W_\triangle' W_\triangle \theta_T^0 \;+\; o_p(1) \tag{67}$$

From (66)-(67), it follows that

$$|k_0 - k| G_T(k) \;=\; \theta_T^{0'} W \triangle' W_\triangle \theta_T^0 + o_p(1) \tag{68}$$

Next, consider $H_T(k)$. We have

$$\begin{aligned} H_T(k) &= 2\theta_T^{0'}(W_0' M_W W_2)(W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - 2\theta_T^{0'}(W_0' M_W \tilde{U}) \\ &\quad + \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \tilde{U}' M_W W_0 (W_0' M_W W_0)^{-1} W_0' M_W \tilde{U} \\ &= H_T^{(1)}(k) \;+\; H_T^{(2)}(k) \end{aligned} \tag{69}$$

where $H_T^{(i)}(k)$ are (re)defined as

$$\begin{aligned} H_T^{(1)}(k) &= 2\theta_T^{0'}(W_0' M_W W_2)(W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - 2\theta_T^{0'}(W_0' M_W \tilde{U}) \\ H_T^{(2)}(k) &= \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \tilde{U}' M_W W_0 (W_0' M_W W_0)^{-1} W_0' M_W \tilde{U} \end{aligned}$$

In the following derivation, we show that $H_T^{(2)}(k) = o_p(1)$ and so the limit behaviour of $H_T(k)$ over $D(C)$ is dominated by the limit behavior of $H_T^{(1)}(k)$.

First, consider $H_T^{(1)}(k)$.

$$\begin{aligned} H_T^{(1)}(k) &= 2\theta_T^{0'}(W_2 - W_\triangle \Xi)' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - 2\theta_T^{0'} W_0' M_W \tilde{U} \\ &= 2\theta_T^{0'}(W_2' M_W \tilde{U} - \Xi W_\triangle' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U}) - 2\theta_T^{0'} W_0' M_W \tilde{U} \\ &= \Xi[2\theta_T^{0'} W_\triangle' M_W \tilde{U} - 2\theta_T^0 W_\triangle' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U}] \end{aligned} \tag{70}$$

We now investigate the convergence of each term on the right hand side of (70) in turn. Noticing that $W_\triangle' W = O_p(1)\|\theta_T^0\|^{-2}$ over $D(C)$ and $(W'W)^{-1}W'\tilde{U} = (1/\sqrt{T}) \cdot (W'W/T)^{-1} W'\tilde{U}/\sqrt{T} = T^{-1/2}O_p(1)$, the first term in (70) can be written as

$$\begin{aligned} \theta_T^0 W_\triangle' M_W \tilde{U} &= \theta_T^0 W_\triangle \tilde{U} - \theta_T^{0'} W_\triangle' W (W'W)^{-1} W' \tilde{U} \\ &= \theta_T^{0'} W_\triangle \tilde{U} - \|\theta_T\| \cdot O_p(1)\|\theta_T^0\|^{-2} \cdot T^{-1/2} O_p(1) \\ &= \theta_T^0 W_\triangle \tilde{U} - O_p(1)/(T^{1/2}\|\theta_T^0\|) \\ &= \theta_T^0 W_\triangle \tilde{U} + o_p(1) \qquad \text{over } D(C) \end{aligned} \tag{71}$$

Now consider the second term on the right hand side of (70). We observe that

$$\begin{aligned} W_{\triangle}' M_W W_2 &= W_{\triangle}' W_2 - W_{\triangle}' W (W'W)^{-1} W' W_2 \\ &= O_p(1) \|\theta_T^0\|^{-2} - O_p(1) \|\theta_T^0\|^{-2} \cdot O_p(T^{-1}) \cdot O_p(T) \\ &= O_p(1) \|\theta_T^0\|^{-2} \quad \text{over } D(C) \\ (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} &= O_p(T^{-1}) \cdot O_p(T^{1/2}) = O_p(T^{-1/2}) \end{aligned}$$

and so

$$\begin{aligned} \theta_T^{0'} W_{\triangle}' M_W W_2 \cdot (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} &= \|\theta_T^0\| \cdot O_p(1) \|\theta_T^0\|^{-2} \cdot O_p(T^{-1/2}) \qquad (72) \\ &= O_p(1) / (T^{1/2} \|\theta_T^0\|) \\ &= o_p(1) \quad \text{over } D(C) \qquad (73) \end{aligned}$$

Thus, combining (71)-(73), we have

$$\begin{aligned} H_T^{(1)}(k) &= \Xi [2\theta^{0'} W_{\triangle}' M_W \tilde{U} - 2\theta^0 W_{\triangle}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U}] \\ &= 2\Xi \theta^{0'} W_{\triangle}' \tilde{U} + o_p(1) \qquad (74) \end{aligned}$$

Next, we prove that $H_T^{(2)}(k) = o_p(1)$. We have

$$\begin{aligned} H_T^{(2)}(k) &= \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \tilde{U}' M_W W_0 (W_0' M_W W_0)^{-1} W_0' M_W \tilde{U} \\ &= \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} - \tilde{U}' M_W (W_2 - W_{\triangle} \Xi) \\ &\quad \times [(W_2 - W_{\triangle} \Xi)' M_W (W_2 - W_{\triangle} \Xi)]^{-1} (W_2 - W_{\triangle} \Xi)' M_W \tilde{U} \\ &\equiv M - N \end{aligned}$$

where M and N are (re)defined as

$$\begin{aligned} M &\equiv \tilde{U}' M_W W_2 (W_2' M_W W_2)^{-1} W_2' M_W \tilde{U} \\ N &\equiv \tilde{U}' M_W (W_2 - W_{\triangle} \Xi) [(W_2 - W_{\triangle} \Xi)' M_W (W_2 - W_{\triangle} \Xi)]^{-1} (W_2 - W_{\triangle} \Xi)' M_W \tilde{U} \end{aligned}$$

To investigate the limiting behavior of $H_T^{(2)}(k)$ over the set $D(C)$, it turns out to be most convenient to check the limit behavior of N first. Using the relationship $W_0 = W_2 - W_{\triangle} \Xi$, N can be written as

$$\begin{aligned} N &= \{\tilde{U}' M_W (W_2 - W_{\triangle} \Xi) / \sqrt{T}\} \cdot [(W_2 - W_{\triangle} \Xi)' M_W (W_2 - W_{\triangle} \Xi) / T]^{-1} \\ &\quad \times \{(W_2 - W_{\triangle} \Xi)' M_W \tilde{U} / \sqrt{T}\} \qquad (75) \end{aligned}$$

We now investigate the limiting behaviour of the terms on the right hand side of (75). Since

$$\begin{aligned}\tilde{U}'M_W W_\triangle \Xi/\sqrt{T} &= W_\triangle \tilde{U}\Xi/\sqrt{T} + o_p(1) \\ &= O_p(1)\sqrt{\|\theta_T\|^{-2}/T} + o_p(1) \\ &= o_p(1) + o_p(1) \\ &= o_p(1) \qquad \text{over } D(C) \equiv \{k : |k-k_0| \leq C\|\theta_T^0\|^{-2}\},\end{aligned}$$

it follows that

$$\begin{aligned}\tilde{U}'M_W(W_2 - W_\triangle \Xi)/\sqrt{T} &= \tilde{U}'M_W W_2/\sqrt{T} - \tilde{U}'M_W W_\triangle \Xi/\sqrt{T} \\ &= \tilde{U}'M_W W_2/\sqrt{T} + o_p(1) \qquad (76)\end{aligned}$$

Similarly, we have

$$\begin{aligned}(W_2 - W_\triangle \Xi)'M_W(W_2 - W_\triangle \Xi)/T &= W_2'M_W W_2/T - (1/T)\{W_2'M_W W_\triangle \Xi \\ &\quad + \Xi W_\triangle' M_W W_2 - W_\triangle' M_W W_\triangle\} \\ &= W_2'M_W W_2/T + o_p(1) \qquad (77)\end{aligned}$$

Thus, combining (75)-(77), we have

$$N = \left(\tilde{U}'M_W W_2/\sqrt{T} + o_p(1)\right)\left(W_2'M_W W_2/T + o_p(1)\right)^{-1}\left(W_2'M_W \tilde{U}/\sqrt{T} + o_p(1)\right)$$

Therefore, we have

$$\begin{aligned}H_T^{(2)}(k) &= M - N \qquad (78) \\ &= \tilde{U}'M_W W_2(W_2'M_W W_2)^{-1}W_2'M_W\tilde{U} - \left(\tilde{U}'M_W W_2/\sqrt{T} + o_p(1)\right) \\ &\quad \times \left(W_2'M_W W_2/T + o_p(1)\right)^{-1}\left(W_2'M_W\tilde{U}/\sqrt{T} + o_p(1)\right) \\ &= \left(\tilde{U}'M_W W_2/\sqrt{T}\right)\left(W_2'M_W W_2/T\right)^{-1}\left(W_2'M_W\tilde{U}/\sqrt{T}\right) \\ &\quad - \left(\tilde{U}'M_W W_2/\sqrt{T} + o_p(1)\right)\left(W_2'M_W W_2/T + o_p(1)\right)^{-1}\left(W_2'M_W\tilde{U}/\sqrt{T} + o_p(1)\right) \\ &= o_p(1) \qquad (79)\end{aligned}$$

It follows from (69), (74) and (79) that

$$H_T(k) = 2\Xi\theta_T^{0'}W_\triangle'\tilde{U} + o_p(1) \qquad (80)$$

Finally, it follows from (12), (68) and (80) that

$$
\begin{aligned}
V_T(k) - V_T(k_0) &= -|k_0 - k| G_T(k) + H_T(k) \\
&= \left(-\theta_T^{0'} W_\triangle' W_\triangle \theta_T^0 + o_p(1)\right) + \left(2\Xi \theta_T^{0'} W_\triangle' \tilde{U} + o_p(1)\right) \\
&= -\theta_T^{0'} W_\triangle' W_\triangle \theta_T^0 + 2\Xi \theta_T^{0'} W_\triangle' \tilde{U} + o_p(1)
\end{aligned}
$$

which is the desired result. $\diamond$.

Proof of Theorem 3:

Theorem 2 proved that $\hat{k} = k_0 + O_p(\|\theta_T^0\|^{-2})$. Since $\theta_T^0 = \theta_0 \upsilon_T$ under Assumption 13, we have $\hat{k} = k_0 + O_p(\upsilon_T^{-2})$. Therefore, it suffices to derive the limiting process of $V_T(k) - V_T(k_0)$ for $k = k_0 + [s\upsilon_T^{-2}]$ and $s \in [-C, C]$.

We first consider $s \leq 0$ (that is, $k \leq k_0$). From Proposition 4,

$$
\begin{aligned}
V_T(k) - V_T(k_0) &= -\theta_T^{0'} W_\triangle' W_\triangle \theta_T^0 + 2\theta_T^{0'} W_\triangle' \tilde{U} \cdot \Xi + o_p(1) \\
&= -\theta_0' (\upsilon_T^2 \sum_{t=k+1}^{k_0} w_t w_t') \theta_0 + 2\theta_0' (\upsilon_T \sum_{t=k+1}^{k_0} w_t \tilde{u}_t) + o_p(1) \\
&= -\theta_0' \Upsilon_0' (\upsilon_T^2 \sum_{t=k+1}^{k_0} z_t z_t') \Upsilon_0 \theta_0 + 2\theta_0' \Upsilon_0' (\upsilon_T \sum_{t=k+1}^{k_0} z_t \tilde{u}_t) + o_p(1) \qquad (81)
\end{aligned}
$$

For the first term in (81) we know that it involves $|[s\upsilon_T^{-2}]|$(the absolute value of $[s\upsilon_T^{-2}]$) observations of z_t. Thus, by Assumptions 2, 5 and 10,

$$
\upsilon_T^2 \sum_{t=k+1}^{k_0} z_t z_t' \Longrightarrow |s| Q_1 \qquad (82)
$$

The second term in (81) can be further expanded as follows

$$
\begin{aligned}
\upsilon_T \sum_{t=k+1}^{k_0} z_t \tilde{u}_t &= \upsilon_T \sum_{t=k+1}^{k_0} z_t u_t + \upsilon_T \sum_{t=k+1}^{k_0} z_t v_t' \beta_{x,1}^0 \\
&\quad - \upsilon_T \sum_{t=k+1}^{k_0} z_t z_t' [(Z'Z)^{-1} Z'V] \beta_{x,1}^0 \qquad (83)
\end{aligned}
$$

We investigate the limit behavior of each term on the right hand side of (83). To this end, it is useful to introduce the following definitions. Let $\tilde{B}_i(r) = V^{1/2} B_i(r)$ where $B_i(r) = [B_i^{(1)}(r)', B_i^{(2)}(r)']$ is $[q \times (p_1+1)] \times 1$ standard Brownian motion corresponding to the i^{th} regime, and $B_i^{(1)}(r)$ is $q \times 1$ and $B_i^{(2)}(r)$ is $qp_1 \times 1$. Also define $B_i^{mat}(r)$ is defined as $vec\{B_i^{mat}(r)\} = B_i^{(2)}(r)$.

We now consider $\upsilon_T \sum_{t=k+1}^{k_0} z_t u_t$. By Assumption 12,

$$(-s\upsilon_T^{-2})^{-1/2} \sum_{t=k+1}^{k_0} z_t u_t \Rightarrow (N_1^{1'} \otimes Q_1^{1/2}) B_i^{(1)}(1)$$

Denoting the rescaled version of this standard Brownian motion process by $W_1 \equiv \sqrt{t} B_i^{(1)}(1)$, we can rewrite the above convergence result as

$$\begin{aligned} \upsilon_T \sum_{t=k+1}^{k_0} z_t u_t &\Rightarrow \sqrt{-s}(N_1^{1'} \otimes Q_1^{1/2}) B_1(1) \\ &= (N_1^{1'} \otimes Q_1^{1/2}) W_1(-s) \end{aligned}$$

Similarly, the limit process of the second term of (83) can be shown to be

$$\upsilon_T \sum_{t=k+1}^{k_0} z_t v_t' \Rightarrow Q_1^{1/2} W_1^{mat}(-s) N_2$$

where $W_1^{mat}(-s) \equiv \sqrt{t} B_1^{mat}(1)$. Finally, the last term of (83) behaves as follows,

$$\begin{aligned} \upsilon_T \sum_{t=k+1}^{k_0} z_t z_t' [(Z'Z)^{-1} Z'V] \beta_{x,1}^0 &= T^{-1/2} \upsilon_T^{-1} s \cdot (s\upsilon_T^{-2})^{-1} \sum_{t=k+1}^{k_0} z_t z_t' \\ &\quad \times [(T^{-1} Z'Z)^{-1} T^{-1/2} Z'V] \beta_{x,1}^0 \\ &= T^{-1/2} \upsilon_T^{-1} O_p(1) = o_p(1) \end{aligned}$$

Thus, combining the results on (83), we obtain

$$\upsilon_T \sum_{t=k+1}^{k_0} z_t \tilde{u}_t \Rightarrow \quad (N_1^{1'} \otimes Q_1^{1/2}) W_1(-s) + Q_1^{1/2} W_1^{mat}(-s) N_2^1 \beta_{x,1}^0$$

Using $vec(A_1 A_2 A_3) = (A_3' \otimes A_1) vec(A_2)$, we have

$$\begin{aligned} Q_1^{1/2} W_1^{mat}(-s) N_2^1 \beta_{x,1}^0 &= vec(Q_1^{1/2} W_1^{mat}(-s) N_2^1 \beta_{x,1}^0) \\ &= (\beta_{x,1}^{0'} N_2^{1'} \otimes Q_1^{1/2}) vec(W_1^{mat}(-s)) \\ &= (\beta_{x,1}^{0'} N_2^{1'} \otimes Q_1^{1/2}) W_1(-s) \end{aligned}$$

Therefore, we have

$$\begin{aligned} \upsilon_T \sum_{t=k+1}^{k_0} z_t \tilde{u}_t &\Longrightarrow \left[(N_1^{1'} \otimes Q_1^{1/2}) + (\beta_{x,1}^{0'} N_2^{1'} \otimes Q_1^{1/2}) \right] W_1(-s) \\ &= \left[(N_1^{1'} + \beta_{x,1}^{0'} N_2^{1'}) \otimes Q_1^{1/2}) \right] W_1(-s) \\ &= \left[(N_1 + N_2^1 \beta_{x,1}^0)' \otimes Q_1^{1/2} \right] W_1(-s) \end{aligned} \tag{84}$$

Then $\theta_0' \Upsilon_0' (\upsilon_T \sum_{t=k+1}^{k_0} z_t \tilde{u}_t)$ has an asymptotic distribution of $(\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0)^{1/2} W_1(-s)$ where $W_1(\cdot)$ is a rescaled Brownian motion process defined on $[0, \infty)$.

Thus, it follows that for $s \leq 0$,

$$\begin{aligned} & V_T(k) - V_T(k_0) = V_T(k_0 + [s\upsilon_T^{-2}]) - V_T(k_0) \\ \Rightarrow \quad & -|s| \theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0 + 2\theta_0' \Upsilon_0' \left[(N_1^1 + N_2^1 \beta_{x,1}^0)' \otimes Q_1^{1/2} \right] W_1(-s) \\ = \quad & -|s| \theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0 + 2(\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0)^{1/2} W_1(-s) \end{aligned}$$

Similarly, for $s > 0$

$$\begin{aligned} & V_T(k_0 + [s\upsilon_T^{-2}]) - V_T(k_0) \\ \Rightarrow \quad & -|s| \theta_0' \Upsilon_0' Q_2 \Upsilon_0 \theta_0 + 2\theta_0' \Upsilon_0' \left[(N_1^2 + N_2^2 \beta_{x,2}^0)' \otimes Q_2^{1/2} \right] W_2(-s) \\ = \quad & -|s| \theta_0' \Upsilon_0' Q_2 \Upsilon_0 \theta_0 + 2(\theta_0' \Upsilon_0' \Phi_2 \Upsilon_0 \theta_0)^{1/2} W_2(-s) \end{aligned}$$

where $W_2(\cdot)$ is another Brownian motion process on $[0, \infty)$. The two processes W_1 and W_2 are independent because they are the limiting processes corresponding to the asymptotically independent regimes.

In summary,

$$\begin{aligned} & V_T(k_0 + [s\upsilon_T^{-2}]) - V_T(k_0) \Longrightarrow G(s) \\ \equiv \quad & \begin{cases} -|s| \theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0 + 2(\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0)^{1/2} W_1(-s) & : \quad s \leq 0 \\ -|s| \theta_0' \Upsilon_0' Q_2 \Upsilon_0 \theta_0 + 2(\theta_0' \Upsilon_0' \Phi_2 \Upsilon_0 \theta_0)^{1/2} W_2(-s) & : \quad s > 0 \end{cases} \end{aligned}$$

Now, we can invoke the continuous mapping theorem to conclude

$$\upsilon_T^2 (\hat{k} - k_0) \longrightarrow_d \arg\max_s G(s) \tag{85}$$

We now show that (85) implies the desired result. By a change of variable $s = bv$ with

$$b = \frac{\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0}{(\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0)^2}$$

it can be shown that

$$\arg\max_s G(s) \;=\; b \cdot \arg\max_v Z(v) \tag{86}$$

where $Z(v)$ is defined in equation (64). We now establish (86).

For $s \leq 0$

$$\begin{aligned}
G(s) &= -|s|\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0 + 2(\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0)^{1/2} W_1(-s) \\
&= -|bv| \cdot \theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0 + 2(\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0)^{1/2} W_1(-bv) \\
&= -|v| b \cdot \theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0 + 2(\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0)^{1/2} \sqrt{b} \cdot W_1(-v) \\
&= -|v| \frac{\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0}{(\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0)^2} \cdot \theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0 + 2(\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0)^{1/2} \frac{(\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0)^{1/2}}{\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0} W_1(-v) \\
&= -|v| \frac{\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0}{\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0} + 2 \frac{\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0}{\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0} W_1(-v)
\end{aligned}$$

Thus,

$$\begin{aligned}
\arg\max_s G(s) &= \arg\max_v \left\{ -|v| \frac{\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0}{\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0} + 2 \frac{\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0}{\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0} W_1(-v) \right\} \\
&= \arg\max_v \left\{ -\frac{|v|}{2} + W_1(-v) \right\} \frac{\theta_0' \Upsilon_0 \Phi_1 \Upsilon_0 \theta_0}{\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0} \\
&= \arg\max_v \left\{ -\frac{|v|}{2} + W_1(-v) \right\}
\end{aligned}$$

Similarly, for $s > 0$

$$\begin{aligned}
G(s) &= -s \cdot \theta_0' \Upsilon_0' Q_2 \Upsilon_0 \theta_0 + 2(\theta_0' \Upsilon_0' \Omega_2 \Upsilon_0 \theta_0)^{1/2} W_2(s) \\
&= -bv \cdot \theta_0' \Upsilon_0' Q_2 \Upsilon_0 \theta_0 + 2(\theta_0' \Upsilon_0' \Omega_2 \Upsilon_0 \theta_0)^{1/2} \sqrt{b} W_2(v) \\
&= -v \frac{\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0}{(\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0)^2} \theta_0' \Upsilon_0' Q_2 \Upsilon_0 \theta_0 + 2(\theta_0' \Upsilon_0' \Omega_2 \Upsilon_0 \theta_0)^{1/2} \frac{(\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0)^{1/2}}{\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0} W_2(v) \\
&= \frac{\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0}{\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0} \left[-\frac{\theta_0' \Upsilon_0' Q_2 \Upsilon_0 \theta_0}{\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0} v + 2 \left(\frac{\theta_0' \Upsilon_0' \Omega_2 \Upsilon_0 \theta_0}{\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0} \right)^{1/2} W_2(v) \right] \\
&= \frac{\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0}{\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0} \left[-\xi v + 2\sqrt{\phi} W_2(v) \right]
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\arg\max_s G(s) &= \arg\max_v \left\{ -\frac{\xi v}{2} + \sqrt{\phi} W_2(v) \right\} \frac{\theta_0' \Upsilon_0' \Phi_1 \Upsilon_0 \theta_0}{\theta_0' \Upsilon_0' Q_1 \Upsilon_0 \theta_0} \\
&= \arg\max_v \left\{ -\frac{\xi v}{2} + \sqrt{\phi} W_2(v) \right\}
\end{aligned}$$

Finally, the statement in Theorem 3 can be established in the following way. Since $V_T(k_0 + [s\upsilon_T^{-2}]) - V_T(k_0) \Rightarrow G(s)$ and $\arg\max_s G(s) = b \cdot \arg\max_\upsilon Z(\upsilon)$, we have

$$b^{-1}\upsilon_T^2(\hat{k} - k_0) \longrightarrow_d \arg\max_{\upsilon} Z(\upsilon).$$

. Using Assumption 13, we have

$$\begin{aligned} b^{-1}\upsilon_T^2 &= \frac{(\theta_0'\Upsilon_0'Q_1\Upsilon_0\theta_0)^2}{\theta_0'\Upsilon_0'\Phi_1\Upsilon_0\theta_0}\upsilon_T^2 \\ &= \frac{(\upsilon_T^{-1}\theta_T^{0\prime}\Upsilon_0'Q_1\Upsilon_0\upsilon_T^{-1}\theta_T^0)^2}{\upsilon_T^{-1}\theta_T^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\upsilon_T^{-1}\theta_T^0}\upsilon_T^2 \\ &= \frac{(\theta_T^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta_T^0)^2}{\theta_T^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta_T^0} \end{aligned}$$

and thus, it follows that

$$b^{-1}\upsilon_T^2(\hat{k} - k_0) = \frac{(\theta_T^{0\prime}\Upsilon_0'Q_1\Upsilon_0\theta_T^0)^2}{\theta_T^{0\prime}\Upsilon_0'\Phi_1\Upsilon_0\theta_T^0}(\hat{k} - k_0) \longrightarrow_d \arg\max_s Z(s), \qquad \diamond.$$

Table 1: Empirical coverage of break point confidence intervals

one break model with $(\beta_1^0;\beta_2^0)$*=(0.3,0.1,-0.3,-0.1)*

$q-1$	T	Confidence Interval 99 %	95 %	90 %
2	60	.90	.82	.75
	120	.95	.89	.85
	240	.97	.92	.87
	480	.99	.94	.89
4	60	.90	.80	.74
	120	.93	.86	.80
	240	.96	.92	.87
	480	.98	.94	.90
8	60	.91	.80	.74
	120	.94	.86	.81
	240	.97	.90	.86
	480	.98	.93	.89

Notes: Here $q-1$ is the number of instruments (excluding the intercept), and the column headed $100a\%$ gives the percentage of times (in 1000 simulations) the $100a\%$ confidence intervals for the break points contain the corresponding true values.

Table 2: Empirical coverage of break point confidence intervals

one break model with $(\beta_1^0;\beta_2^0)$=*(0.5,0.1,-0.5,-0.1)*

$q-1$	T	Confidence Interval		
		99 %	95 %	90 %
2	60	.95	.90	.86
	120	.97	.93	.89
	240	.98	.95	.92
	480	.99	.97	.92
4	60	.94	.88	.83
	120	.97	.93	.87
	240	.99	.93	.90
	480	.99	.95	.91
8	60	.94	.89	.85
	120	.97	.93	.88
	240	.98	.95	.91
	480	.99	.96	.92

Notes: For definitions see Table 1.

Table 3: Empirical coverage of break point confidence intervals

one break model with $(\beta_1^0;\beta_2^0)$*=(1,0.1;-1,-0.1)*

$q-1$	T	Confidence Interval		
		99 %	95 %	90 %
2	60	.99	.97	.96
	120	.99	.97	.96
	240	1.00	.98	.97
	480	1.00	.99	.98
4	60	.99	.98	.96
	120	1.00	.98	.97
	240	1.00	.98	.98
	480	1.00	.99	.98
8	60	.99	.97	.96
	120	.99	.98	.96
	240	.99	.98	.96
	480	.99	.98	.96

Notes: For definitions see Table 1.

Table 4: Empirical coverage of break point confidence intervals

two break model with $(\beta_1^0;\beta_2^0,\beta_3^0)$*=(0.3,0.1;-0.3,-0.1;0.3;0.1)*

k	T	Confidence Interval					
		1^{st} break			2^{nd} break		
		99 %	95 %	90 %	99 %	95 %	90 %
2	60	.91	.75	.66	.93	.81	.71
	120	.94	.82	.76	.95	.86	.78
	240	.97	.88	.81	.97	.92	.86
	480	.98	.94	.88	.98	.93	.88
4	60	.92	.76	.68	.90	.78	.70
	120	.94	.84	.76	.94	.86	.78
	240	.95	.87	.82	.97	.88	.82
	480	.98	.93	.88	.98	.93	.88
8	60	.92	.78	.70	.90	.79	.70
	120	.95	.83	.75	.94	.84	.76
	240	.96	.88	.81	.97	.88	.83
	480	.97	.92	.86	.98	.92	.88

Notes: For definitions see Table 1.

Table 5: Empirical coverage of break point confidence intervals

two break model with $(\beta_1^0; \beta_2^0, \beta_3^0)$*=(0.5,0.1;-0.5,-0.1;0.5;0.1)*

$q-1$	T	Confidence Interval					
		1^{st} break			2^{nd} break		
		99 %	95 %	90 %	99 %	95 %	90 %
2	60	.94	.86	.79	.94	.87	.84
	120	.96	.91	.89	.97	.92	.88
	240	.98	.95	.91	.98	.94	.90
	480	.99	.95	.92	.99	.96	.92
4	60	.94	.85	.78	.94	.87	.82
	120	.97	.91	.86	.98	.92	.87
	240	.98	.94	.90	.99	.94	.89
	480	.99	.96	.92	.99	.95	.91
8	60	.95	.85	.78	.95	.88	.82
	120	.97	.90	.86	.97	.91	.86
	240	.98	.93	.89	.98	.94	.89
	480	.99	.95	.92	.99	.97	.94

Notes: For definitions see Table 1.

Table 6: Empirical coverage of break point confidence intervals

two break model with $(\beta_1^0; \beta_2^0, \beta_3^0)$=*(1,0.1;-1,-0.1;1;0.1)*

k	T	Confidence Interval					
		1^{st} break			2^{nd} break		
		99 %	95 %	90 %	99 %	95 %	90 %
2	60	.98	.95	.94	.98	.96	.94
	120	.99	.98	.96	.99	.98	.97
	240	1.00	.98	.97	1.00	.99	.98
	480	1.00	.98	.97	.99	.98	.97
4	60	.99	.96	.94	.99	.96	.94
	120	.99	.97	.96	.99	.97	.96
	240	.99	.97	.96	1.00	.99	.98
	480	1.00	.98	.96	.99	.97	.96
8	60	.99	.96	.95	.99	.96	.93
	120	1.00	.98	.96	.98	.97	.96
	240	1.00	.98	.96	1.00	.98	.96
	480	1.00	.98	.98	.99	.98	.97

Notes: For definitions see Table 1.

Table 7: NKPC - stability statistics for structural equation

q-1	$q\times$ sup-F	F(k+1:k)	BIC
0	-	-	-0.020
1	13.76	10.29	-0.043
2	11.92	55.45	0.137
3	15.00	42.70	0.252
4	23.38	5.98	0.433
5	18.07	-	0.716

Notes: $q\times$ sup-F denotes the statistic for testing $H_0 : m = 0$ vs. $H_1 : m = k$, multiplied by q; F(k+1:k) is the statistic for testing $H_0 : m = k$ vs. $H_1 : m = k + 1$; BIC is the BIC criterion; see Hall, Han, and Boldea (2007) for further details. The percentiles for the statistics are for $k = 1, 2, \ldots$ respectively: (i) $q\times$ sup-F: (10%, 1%) significance level = (19.70, 26.71), (17.67, 21.87), (16.04, 19.42), (14.55, 17.44), (12.59,15.02); (ii) F(k+1:k): (10%, 1%) significance level =(21.79, 28.36), (22.87, 29.30), (24.06,29.86), (24.68, 30.52).

References

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press, Cambridge, MA, U.S.A.

Andrews, D. W. K. (1993). 'Tests for parameter instability and structural change with unknown change point', *Econometrica*, 61: 821–856.

Andrews, D. W. K., and Fair, R. (1988). 'Inference in econometric models with structural change', *Review of Economic Studies*, 55: 615–640.

Andrews, D. W. K., and Ploberger, W. (1994). 'Optimal tests when a nuisance parameter is present only under the alternative', *Econometrica*, 62: 1383–1414.

Bai, J. (1994). 'Least squares estimation of a shift in linear processes', *Journal of Time Series Analysis*, 15: 453–472.

——— (1997a). 'Estimating multiple breaks one at a time', *Econometric Theory*, 13: 315–352.

——— (1997b). 'Estimation of a change point in multiple regression models', *Review of Economics and Statistics*, 79: 551–563.

Bai, J., and Perron, P. (1998). 'Estimating and testing linear models with multiple structural changes', *Econometrica*, 66: 47–78.

Bhattacharya, P. K. (1987). 'Maximum Likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case', *Journal of Multivariate Analysis*, 23: 183–208.

Ghysels, E., and Hall, A. R. (1990a). 'Are consumption based intertemporal asset pricing models structural?', *Journal of Econometrics*, 45: 121–139.

——— (1990b). 'A test for structural stability of Euler condition parameters estimated via the Generalized Method of Moments', *International Economic Review*, 31: 355–364.

Hahn, J., and Inoue, A. (2002). 'A Monte Carlo comparison of various asymptotic approximations to the distribution of instrumental variables estimators', *Econometric Reviews*, 21: 309–336.

Hall, A. R., Han, S., and Boldea, O. (2007). 'Inference regarding multiple structural changes in linear models estimated via Two Stage Least Squares', Discussion paper, Economics, School of Social Studies, University of Manchester, Manchester, UK.

Hall, A. R., and Sen, A. (1999). 'Structural stability testing in models estimated by Generalized Method of Moments', *Journal of Business and Economic Statistics*, 17: 335–348.

Hinckley, D. (1970). 'Inference about the change points in a sequence of random variables', *Biometrika*, 57: 1–17.

Picard, D. (1985). 'Testing and estimating change points in time series', *Journal of Applied Probability*, 20: 411–415.

Quandt, R. (1960). 'Tests of the hypothesis that a linear regression obeys two separate regimes', *Journal of American Statistical Association*, 55: 324–330.

Serfling, R. J. (1970). 'Convergence properties of S_n under moment restrictions', *The Annals of Mathematical Statistics*, 41: 1235–1248.

Sowell, F. (1996). 'Optimal tests of parameter variation in the Generalized Method of Moments framework', *Econometrica*, 64: 1085–1108.

Yao, Y.-C. (1987). 'Approximating the distribution of the ML estimate of the change point in a sequence of independent r.v.'s', *Annals of Statistics*, 4: 1321–1328.

Zhang, C., Osborn, D., and Kim, D. (2007). 'The new Keynesian Phillips curve: from sticky inflation to sticky prices', *Journal of Money, Credit and Banking*, forthcoming.