

## TRANSITION MODELING AND ECONOMETRIC CONVERGENCE TESTS

BY PETER C. B. PHILLIPS AND DONGGYU SUL<sup>1</sup>

A new panel data model is proposed to represent the behavior of economies in transition, allowing for a wide range of possible time paths and individual heterogeneity. The model has both common and individual specific components, and is formulated as a nonlinear time varying factor model. When applied to a micro panel, the decomposition provides flexibility in idiosyncratic behavior over time and across section, while retaining some commonality across the panel by means of an unknown common growth component. This commonality means that when the heterogeneous time varying idiosyncratic components converge over time to a constant, a form of panel convergence holds, analogous to the concept of conditional sigma convergence. The paper provides a framework of asymptotic representations for the factor components that enables the development of econometric procedures of estimation and testing. In particular, a simple regression based convergence test is developed, whose asymptotic properties are analyzed under both null and local alternatives, and a new method of clustering panels into club convergence groups is constructed. These econometric methods are applied to analyze convergence in cost of living indices among 19 U.S. metropolitan cities.

KEYWORDS: Club convergence, relative convergence, common factor, convergence, log  $t$  regression test, panel data, transition.

### 1. INTRODUCTION

IN THE PAST DECADE, the econometric theory for dynamic panel regressions has developed rapidly alongside a growing number of empirical studies involving macro, international, regional, and micro economic data. This rapid development has been stimulated both by the availability of new data sets and by the recognition that panels help empirical researchers to address many new economic issues. For example, macro aggregated panels such as the Penn World Table (PWT) data have been used to investigate growth convergence and evaluate the many diverse determinants of economic growth. Durlauf and Quah (1999) and Durlauf, Johnson, and Temple (2005) provided excellent overviews of this vast literature and the econometric methodology on which it depends. Similarly, micro panel data sets such as the PSID have been extensively used to analyze individual behavior of economic agents across section and over time; see Ermisch (2004) and Hsaio (2003) for recent overviews of micro panel research. A pervasive finding in much of this empirical panel data research is the importance of individual heterogeneity. This finding has helped researchers to build more realistic models that account for heterogeneity, an example being the renewed respect in macroeconomic modeling for micro foundations that accommodate individual heterogeneity; see Browning, Hansen, and Heckman

<sup>1</sup>Our thanks go to two referees and a co-editor for helpful comments on an earlier version. Phillips acknowledges partial research support from a Kelly Fellowship and the NSF under Grant SES 04-142254 and SES 06-47086.

(1999), Krusell and Smith (1998), Givenen (2005), and Browning and Carro (2006).

Concerns about capturing heterogeneous agent behavior in economic theory and modeling this behavior in practical work have stimulated interest in the empirical modeling of heterogeneity in panels. One popular empirical model involves a common factor structure and idiosyncratic effects. Early econometric contributions of this type analyzed the asymptotic properties of common factors in asset pricing models (e.g., Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986, 1988)). Recent studies have extended these factor models in several directions and developed theory for the determination of the number of common factors and for inference in panel models with nonstationary common factors and idiosyncratic errors (e.g., Bai (2003, 2004), Bai and Ng (2002, 2006), Stock and Watson (1999), Moon and Perron (2004), Phillips and Sul (2006)). There is much ongoing work in the econometric development of the field to better match the econometric methods to theory and to the needs of empirical research.

To illustrate some of the issues, take the simple example of a single factor model

$$(1) \quad X_{it} = \delta_i \mu_t + \epsilon_{it},$$

where  $\delta_i$  measures the idiosyncratic distance between some common factor  $\mu_t$  and the systematic part of  $X_{it}$ . The econometric interpretation of  $\mu_t$  in applications may differ from the prototypical interpretation of a “common factor” or aggregate element of influence in micro or macro theory. The factor  $\mu_t$  may represent the aggregated common behavior of  $X_{it}$ , but it could also be any common variable of influence on individual behavior, such as an interest rate or exchange rate. The model then seeks to capture the evolution of the individual  $X_{it}$  in relation to  $\mu_t$  by means of its two idiosyncratic elements: the systematic element ( $\delta_i$ ) and the error ( $\epsilon_{it}$ ).

The present paper makes two contributions in this regard. First, we extend (1) in a simple manner by allowing the systematic idiosyncratic element to evolve over time, thereby accommodating heterogeneous agent behavior and evolution in that behavior by means of a time varying factor loading coefficient  $\delta_{it}$ . We further allow  $\delta_{it}$  to have a random component, which absorbs  $\epsilon_{it}$  in (1) and allows for possible convergence behavior in  $\delta_{it}$  over time in relation to the common factor  $\mu_t$ , which may represent some relevant aggregate variable or possible representative agent behavior. The new model has a time varying factor representation

$$(2) \quad X_{it} = \delta_{it} \mu_t,$$

where both components  $\delta_{it}$  and  $\mu_t$  are time varying and there may be some special behavior of interest in the idiosyncratic element  $\delta_{it}$  over time. As discussed

in Section 4, we model the time varying behavior of  $\delta_{it}$  in semiparametric form as

$$(3) \quad \delta_{it} = \delta_i + \sigma_i \xi_{it} L(t)^{-1} t^{-\alpha},$$

where  $\delta_i$  is fixed,  $\xi_{it}$  is iid(0, 1) across  $i$  but weakly dependent over  $t$ , and  $L(t)$  is a slowly varying function (like  $\log t$ ) for which  $L(t) \rightarrow \infty$  as  $t \rightarrow \infty$  (see equation (24)). This formulation ensures that  $\delta_{it}$  converges to  $\delta_i$  for all  $\alpha \geq 0$ , which therefore becomes a null hypothesis of interest. If this hypothesis holds and  $\delta_i = \delta_j$  for  $i \neq j$ , the model still allows for transitional periods in which  $\delta_{it} \neq \delta_{jt}$ , thereby incorporating the possibility of transitional heterogeneity or even transitional divergence across  $i$ . As shown later, further heterogeneity may be introduced by allowing the decay rate  $\alpha$  and slowly varying function  $L(t)$  to be individual specific.

Such formulations accommodate some recent models of heterogeneous agent behavior. For example, heterogeneous discount factor models typically assume that the heterogeneity is transient and that the discount factors become homogeneous in the steady state (e.g., Uzawa (1968), Lucas and Stokey (1984), Obstfeld (1990), Schmidt-Grohe and Uribe (2003), Choi, Mark, and Sul (2006)). In such cases,  $\delta_{it}$  contains information relating to these assumed characteristics. The parameter of interest is then  $\delta_{it}$ , and particular attention is focused on its temporal evolution and convergence behavior.

The second contribution of the paper addresses this latter issue and involves the development of an econometric test of convergence for the time varying idiosyncratic components. Specifically, we develop a simple regression based test of the hypothesis  $H_0: \delta_{it} \rightarrow \delta$  for some  $\delta$  as  $t \rightarrow \infty$ . The approach has several features that make it useful in practical work. First, the test does not rely on any particular assumptions concerning trend stationarity or stochastic nonstationarity in  $X_{it}$  or  $\mu_i$ . Second, the nonlinear form of the model (2) is sufficiently general to include a wide range of possibilities in terms of the time paths for  $\delta_{it}$  and their heterogeneity over  $i$ . By focusing on  $\delta_{it}$ , our approach delivers information about the transition path of  $\delta_{it}$  and allows for the important case in practice where individual behavior may be transitionally divergent.

The remainder of the paper is organized into eight sections. Section 2 motivates our approach in terms of some relevant economic examples of factor models in macroeconomic convergence, labor income evolution, and stock prices. A major theme in our work is the analysis of long run equilibrium and convergence by means of a transition parameter,  $h_{it}$ . This parameter is constructed directly from the data  $X_{it}$  and is a functional of  $\delta_{it}$  that provides a convenient relative measure of the temporal evolution of  $\delta_{it}$ . Under certain regularity conditions, we show in Section 3 that  $h_{it}$  has an asymptotic representation in a standardized form that can be usefully interpreted as a relative transition path for economy  $i$  in relation to other economies in the panel.

Section 4 introduces a new regression test of convergence and a procedure for clustering panel data into clubs with similar convergence characteristics.

We call the regression test of convergence the  $\log t$  test because it is based on a time series linear regression of a cross section variance ratio of the  $h_{it}$  on  $\log t$ . This test is very easy to apply in practice, involving only a simple linear regression and a one-sided regression coefficient test with standard normal critical values. The asymptotic properties of this test are obtained and a local asymptotic power analysis is provided. The regression on which this test is based also provides an empirical estimate of the speed of convergence. This section provides a step by step procedure for practical implementation of this test and its use as a clustering algorithm to find club convergence groups. An analysis of the statistical properties of the convergence test and club convergence clustering algorithm is given in the [Appendix](#).

Section 5 reports the results of some Monte Carlo experiments that evaluate the performance of the convergence test in finite samples. The experiments are set up to include some practically interesting and relevant data generating processes.

Section 6 contains an empirical application of our methods to test for convergence in the cost of living across 19 metropolitan U.S. cities using consumer price indices. The empirical results reveal no convergence in cost of living among U.S. cities. Apparently, the cost of living in major metropolitan cities in California is increasing faster than in the rest of the United States, while the cost of living in St. Louis and Houston is decreasing relative to the rest of the United States.

Section 7 concludes the paper. The [Appendices](#) contain technical material and proofs.

## 2. TIME VARYING FACTOR REPRESENTATION AND CONVERGENCE

Panel data  $X_{it}$  are often usefully decomposed as

$$(4) \quad X_{it} = g_{it} + a_{it},$$

where  $g_{it}$  embodies systematic components, including permanent common components that give rise to cross section dependence, and  $a_{it}$  represents transitory components. For example, the panel  $X_{it}$  could comprise log national income data such as the PWT, regional log income data such as the 48 contiguous U.S. state log income data, regional log consumer price index data, or personal survey income data among many others. We do not assume any particular parametric specification for  $g_{it}$  and  $a_{it}$  at this point, and the framework includes many linear, nonlinear, stationary, and nonstationary processes.

As it stands, the specification (4) may contain a mixture of both common and idiosyncratic components in the elements  $g_{it}$  and  $a_{it}$ . To separate common from idiosyncratic components in the panel, we may transform (4) to the form of (2), namely

$$(5) \quad X_{it} = \begin{pmatrix} g_{it} + a_{it} \\ \mu_t \end{pmatrix} \mu_t = \delta_{it} \mu_t \quad \text{for all } i \text{ and } t,$$

where  $\mu_t$  is a single common component and  $\delta_{it}$  is a time varying idiosyncratic element. For example, if  $\mu_t$  represents a common trend component in the panel, then  $\delta_{it}$  measures the relative share in  $\mu_t$  of individual  $i$  at time  $t$ . Thus,  $\delta_{it}$  is a form of individual economic distance between the common trend component  $\mu_t$  and  $X_{it}$ . The representation (5) is a time varying factor model of the form (2) in which  $\mu_t$  is assumed to have some deterministic or stochastically trending behavior that dominates the transitory component  $a_{it}$  as  $t \rightarrow \infty$ . Factoring out a common trend component  $\mu_t$  in (5) leads naturally to specifications of the form (3) for  $\delta_{it}$ . Additionally, under some regularity conditions it becomes possible to characterize a limiting relative transition path for  $X_{it}$ , as discussed in Section 3. When both  $g_{it}$  and  $a_{it}$  behave like  $I(0)$  variables over time, this limiting characterization is not as relevant and factoring such as (5) with transition properties for  $\delta_{it}$  like those of (3) are less natural.<sup>2</sup>

The following examples illustrate how the simple econometric representation (5) usefully fits in with some micro- and macroeconomic models that are commonly used in applied work.

*Economic Growth:* Following Parente and Prescott (1994), Howitt and Mayer-Foulkes (2005), and Phillips and Sul (2006), and allowing for heterogeneous technology progress in a standard neoclassical growth model, log per capita real income,  $\log y_{it}$ , can be written as

$$(6) \quad \log y_{it} = \log y_i^* + (\log y_{i0} - \log y_i^*)e^{-\beta_{it}} + \log A_{it} = a_{it} + \log A_{it},$$

where  $\log y_i^*$  is the steady state level of log per capita real effective income,  $\log y_{i0}$  is the initial log per real effective capita income,  $\beta_{it}$  is the time varying speed of convergence rate, and  $\log A_{it}$  is the log of technology accumulation for economy  $i$  at time  $t$ . The relationship is summarized in (6) in the terms  $a_{it}$  and  $\log A_{it}$ , where  $a_{it}$  captures transitional components and  $\log A_{it}$  includes permanent components. Within this framework, Phillips and Sul (2006) further decomposed  $\log A_{it}$  as

$$\log A_{it} = \log A_{i0} + \gamma_{it} \log A_t,$$

writing current technology for country  $i$  in terms of initial technology accumulation,  $\log A_{i0}$ , and a component,  $\gamma_{it} \log A_t$ , that captures the distance of country  $i$  technology from publicly available advanced technology,  $\log A_t$ , at time  $t$ . The coefficient  $\gamma_{it}$  that measures this distance may vary over time and across country. If advanced technology  $\log A_t$  is assumed to grow at a constant rate  $a$ , then

$$\log y_{it} = \left( \frac{a_{it} + \log A_{i0} + \gamma_{it} \log A_t}{at} \right) at = \delta_{it} \mu_t,$$

<sup>2</sup>Nonetheless, when factor representations like (5) do arise in the  $I(0)$  case, some related modeling possibilities for the transition curves are available and these will be explored in later work.

corresponding to (5) and  $\delta_{it}$  may be modeled according to (3). Phillips and Sul called  $\delta_{it}$  a *transition parameter* and  $\mu_t$  a *common growth component*. Both components are of interest in this example. In the analysis of possible growth convergence or divergence over time and in the study of heterogeneous transition paths across economies, the time varying component  $\delta_{it}$  is especially important.

*Labor Income:* In labor economics (e.g., Katz and Autor (1999), Moffitt and Gottschalk (2002)), personal log real income  $\log y_{it}$  within a particular age group is commonly decomposed into components of permanent income,  $g_{it}$ , and transitory income,  $a_{it}$ , so that

$$\log y_{it} = g_{it} + a_{it}.$$

Typically, transitory income is interpreted as an idiosyncratic component and permanent income is regarded as having some common (possibly stochastic) trend component. Again, this model may be rewritten as in (5) by factoring out the common stochastic trend component. The main parameter of interest then becomes the time profile of the personal factor loading coefficient  $\delta_{it}$ . The evolution of this parameter may then be modeled in terms of individual attributes and relevant variables, such as education, vocational training, or job experience.

For example, gender wage differences might be examined by modeling wages as  $\log y_{it} = \delta_{it}\mu_t$ , with  $\delta_{it}$  satisfying

$$\delta_{it} \rightarrow \begin{cases} \delta_M & \text{for } i \in M \text{ (male),} \\ \delta_F & \text{for } i \in F \text{ (female),} \end{cases}$$

where  $\mu_t$  represents a common overall wage growth component. Alternatively, wages might be modeled as  $\log y_{it} = \delta_{Mit}\mu_{Mt} + \delta_{Fit}\mu_{Ft}$  with possibly distinct male and female growth components  $\mu_{Mt}$  and  $\mu_{Ft}$  that both influence overall wage growth but with coefficients satisfying

$$(7) \quad \begin{aligned} \delta_{Mit} &\rightarrow \delta_M, & \delta_{Fit} &\rightarrow 0 & \text{for } i \in M, \\ \delta_{Mit} &\rightarrow 0, & \delta_{Fit} &\rightarrow 1 & \text{for } i \in F. \end{aligned}$$

Then  $\log y_{it} = (\delta_{Mit}(\mu_{Mt}/\mu_{Ft}) + \delta_{Fit})\mu_{Ft} = \delta_{it}\mu_t$ , in which case the transition coefficient  $\delta_{it}$  may diverge if trend wage growth is higher for males than females. In either case, we can model wages as a single common factor without loss of generality within each convergent subgroup, and overall convergence and divergence may be assessed in terms of the time evolution of  $\delta_{it}$ .

Importantly also, by analyzing subgroup-convergent behavior among the idiosyncratic transition coefficients  $\delta_{it}$ , one may locate the sources of divergence in a panel. Suppose, for instance, that wage inequality arises because of

gender differences as well as certain other factors. By identifying convergence clubs in the wage transition coefficients and analyzing the characteristics of these clubs, the sources of wage inequality may be identified empirically.

*Stock Price Factor Modeling:* Models with a time varying factor structure have been popular for some time in finance. For example, Fama and French (1993, 1996) modeled stock returns  $R_{it}$  as

$$(8) \quad R_{it} = \gamma_{1,it}\theta_{1t} + \gamma_{2,it}\theta_{2t} + \gamma_{3,it}\theta_{3t} + \epsilon_{it},$$

where the  $\theta_{st}$  are certain “common” determining factors for stock returns, while the  $\gamma_{s,it}$  are time varying factor loading coefficients that capture the individual effects of the factors. It has often been found convenient in applied research to assume that the time varying loading coefficients are constant over short time periods. Ludvigson and Ng (2007), for instance, recently estimated the number of common factors in a model of the form (8) based on time invariant factor loadings. On the other hand, Adrian and Franzoni (2005) relaxed the assumption and attempted to estimate time varying loadings by means of the Kalman filter under the assumption that the factor loadings follow an AR(1) specification.

Alternatively, as in Menzly, Santos, and Veronesi (2002), we may model stock prices,  $X_{it}$ , instead of stock returns in (8) with multiple common factors, writing

$$(9) \quad X_{it} = \sum_{j=1}^J \delta_{j,it}\mu_{jt} + e_{it} = \left( \sum_{j=1}^J \delta_{j,it} \frac{\mu_{jt}}{\mu_{1t}} + \frac{e_{it}}{\mu_{1t}} \right) \mu_{1t} + e_{it} = \delta_{it}\mu_t,$$

so that the time varying multiple common factor structure can be embedded in the framework (5) of a time varying single common factor structure. If the common trend elements in (9) are drifting  $I(1)$  variables of the form

$$\mu_{jt} = m_j t + \sum_{s=1}^t e_{js} \quad \text{for } j = 1, \dots, J, \quad \text{with } m_1 \neq 0,$$

then

$$\frac{\mu_{jt}}{\mu_{1t}} = \frac{m_j t + \sum_{s=1}^t \epsilon_{js}}{m_1 t + \sum_{s=1}^t \epsilon_{1s}} = \frac{m_j}{m_1} + o_p(1)$$

and we have

$$\delta_{it} = \sum_{j=1}^J \delta_{j,it} \frac{m_j}{m_1} \{1 + o_p(1)\}, \quad \mu_t = \mu_{1t}.$$

Convergence occurs if  $\delta_{j,it} \rightarrow \delta_j \forall j$  as  $t \rightarrow \infty$  and then  $\delta_{it} \rightarrow \sum_{j=1}^J \delta_j(m_j/m_1) = \delta$ . It is not necessary to assume that there is a dominant common factor for this representation to hold. Moreover, as in (7), we may have certain convergent subgroups  $\{G_a: a = 1, \dots, A\}$  of stocks for which  $\delta_{j,it} \rightarrow \delta_j^a$  for  $i \in G_a$  and then  $\delta_{it} \rightarrow \delta^a = \sum_{j=1}^J \delta_j^a(m_j/m_1)$  for  $i \in G_a$ . In such cases, the  $X_{it}$  diverge overall, but the panel may be decomposed into  $A$  convergent subgroups. We will discuss how to classify clusters of convergent subgroups later in Section 4.

### 2.1. Long Run Equilibrium and Convergence

An important feature of the time varying factor representation is that it provides a new way to think about and model long run equilibrium. Broadly speaking, time series macroeconomics presently involves two categories of analysis: long run equilibrium growth on the one hand and short run dynamics on the other. This convention has enabled extensive use of cointegration methods for long run analysis and stationary time series methods for short run dynamic behavior. In the time varying factor model, the use of common stochastic trends conveniently accommodates long run comovement in aggregate behavior without insisting on the existence of cointegration and it further allows for the modeling of transitional effects. In particular, idiosyncratic factor loadings provide a mechanism for heterogeneous behavior across individuals and the possibility of a period of transition in a path that is ultimately governed by some common long run stochastic trend.

If two macroeconomic variables  $X_{it}$  and  $X_{jt}$  have stochastic trends and are thought to be in long run equilibrium, then the time series are commonly hypothesized to be cointegrated and this hypothesis is tested empirically. Cointegration tests are typically semiparametric with respect to short run dynamics and rely on reasonably long time spans of data. However, in micro panels such long run behavior is often not empirically testable because of data limitations that result in much shorter panels. In the context of the nonlinear factor model (5), suppose that the loading coefficients  $\delta_{it}$  slowly converge to  $\delta$  over time, but the data available to the econometrician are limited. The difference between two time series in the panel is given by  $X_{it} - X_{jt} = (\delta_{it} - \delta_{jt})\mu_t$ . If  $\mu_t$  is unit root nonstationary and  $\delta_{it} \neq \delta_{jt}$ , then  $X_{it}$  is generally not cointegrated with  $X_{jt}$ . But since  $\delta_{it}$  and  $\delta_{jt}$  converge to some common  $\delta$  as  $t \rightarrow \infty$ , we may think of  $X_{it}$  and  $X_{jt}$  as being *asymptotically* cointegrated. However, even in this case, if the speed of divergence of  $\mu_t$  is faster than the speed of the convergence of  $\delta_{it}$ , the residual  $(\delta_{it} - \delta_{jt})\mu_t$  may retain nonstationary characteristics and standard cointegration tests will then typically have low power to detect the asymptotic comovement.

To fix ideas, suppose

$$(10) \quad \delta_{it} \rightarrow \begin{cases} \delta^a & \text{for } i \in G_a, \\ \delta^b & \text{for } i \in G_b, \end{cases}$$



so that there is convergence within each of the two subgroups  $G_a$  and  $G_b$ . Under model (3) for the transition coefficients, the following relation then holds between series  $X_{it}$  and  $X_{jt}$  for  $i \in G_a$  and  $j \in G_b$ :

$$X_{it} - \frac{\delta_a}{\delta_b} X_{jt} = \left( \delta_{it} - \frac{\delta_a}{\delta_b} \delta_{jt} \right) \mu_t = \left\{ \sigma_i \xi_{it} - \frac{\delta_a}{\delta_b} \sigma_j \xi_{jt} \right\} \frac{\mu_t}{L(t)t^\alpha}.$$

Hence,  $X_{it} - \delta_a/\delta_b X_{jt}$  is  $I(0)$  when  $\mu_t = O_p(L(t)t^\alpha)$ , and then any two series from subgroups  $G_a$  and  $G_b$  are cointegrated. For example, when  $\mu_t = \mu L(t)t^\alpha + \sum_{s=1}^t \zeta_s$  for some stationary sequence  $\zeta_s$ , then each individual series  $X_{it}$  follows a unit root process with nonlinear drift, and is cointegrated with other series in  $G_a$  with cointegrating vector  $(1, -1)$  and is cointegrated with series in  $G_b$  with cointegrating vector  $(1, -\delta_a/\delta_b)$ . However, if  $L(t)^{-1}t^{-\alpha}\mu_t$  diverges (e.g., when  $\alpha = 1/2$  and  $\mu_t = O_p(t)$ ), then the series  $X_{it}$  and  $X_{jt}$  are not cointegrated even though we have convergence (10) in subgroups  $G_a$  and  $G_b$  when  $\alpha > 0$  and

$$\begin{aligned} \delta_{it} - \frac{\delta_a}{\delta_b} \delta_{jt} &\rightarrow_p 0 \quad \text{for } i \in G_a, \quad j \in G_b, \\ \delta_{it} - \delta_{jt} &\rightarrow_p 0 \quad \text{for } i, j \in G_a. \end{aligned}$$

In effect, the speed of convergence is not fast enough to ensure cointegrated behavior.

These examples show that for economists to analyze comovement and convergence in the context of individual heterogeneity, and to analyze evolution in the heterogeneity over time and across groups, some rather different econometric methods are needed. In particular, under these conditions, conventional cointegration tests do not serve as adequate tests for convergence. Clearly, the two hypotheses of cointegration and convergence are related but have distinct features. As the above examples illustrate, even though there may be no empirical support for cointegration between two series  $X_{it}$  and  $X_{jt}$ , it does not mean there is an absence of comovement or convergence between  $X_{it}$  and  $X_{jt}$ .

Accordingly, a simple but intuitive way to define “relative” long run equilibrium or convergence between such series is in terms of their ratio rather than their difference or linear combinations. That is, relative long run equilibrium exists among the  $X_{it}$  if

$$(11) \quad \lim_{k \rightarrow \infty} \frac{X_{it+k}}{X_{jt+k}} = 1 \quad \text{for all } i \text{ and } j.$$

In the context of (5), this condition is equivalent to convergence of the factor loading coefficients

$$(12) \quad \lim_{k \rightarrow \infty} \delta_{it} = \delta.$$

On the other hand, if  $X_{it}$  and  $X_{jt}$  are cointegrated, then the ratio  $X_{it}/X_{jt}$  typically converges to a constant or a random variable, the former occurring when the series have a nonzero deterministic drift.

## 2.2. Relative Transition

In the general case of (5), the number of observations in the panel is less than the number of unknowns in the model. It is therefore impossible to estimate the loading coefficients  $\delta_{it}$  directly without imposing some structure on  $\delta_{it}$  and  $\mu_t$ . Both parametric and nonparametric structures are possible. For example, if  $\delta_{it}$  evolved according to an AR(1), while  $\mu_t$  followed a random walk with a drift, it would be possible to estimate both  $\delta_{it}$  and  $\mu_t$  by a filtering technique such as the Kalman filter. Alternatively, as we show below, under some regularity conditions, it is possible to use a nonparametric formulation in which the quantities of interest are a transition function (based on  $\delta_{it}$ ) and a growth curve (based on  $\mu_t$ ). Some further simplification for practical purposes is possible by using a relative version of  $\delta_{it}$  as we now explain.

Since  $\mu_t$  is a common factor in (5), it may be removed by scaling to give the relative loading or transition coefficient

$$(13) \quad h_{it} = \frac{X_{it}}{\frac{1}{N} \sum_{i=1}^N X_{it}} = \frac{\delta_{it}}{\frac{1}{N} \sum_{i=1}^N \delta_{it}},$$

which measures the loading coefficient  $\delta_{it}$  in relation to the panel average at time  $t$ . We assume that the panel average  $N^{-1} \sum_{i=1}^N \delta_{it}$  and its limit as  $N \rightarrow \infty$  differ from zero almost surely, so that  $h_{it}$  is well defined by the construction (13). In typical applications,  $X_{it}$ ,  $\mu_t$ , and  $\delta_{it}$  are all positive, so the construction of this relative coefficient presents no difficulty in practice. Like  $\delta_{it}$ ,  $h_{it}$  still traces out a transition path for economy  $i$ , but now does so in relation to the panel average. The concept is useful in the analysis of growth convergence and measurement of transition effects, as discussed in some companion empirical work (Phillips and Sul (2006)) where  $h_{it}$  is called the *relative transition parameter*.

Some properties of  $h_{it}$  are immediately apparent. First, the cross sectional mean of  $h_{it}$  is unity by definition. Second, if the factor loading coefficients  $\delta_{it}$  converge to  $\delta$ , then the relative transition parameters  $h_{it}$  converge to unity. In this case, in the long run, the cross sectional variance of  $h_{it}$  converges to zero, so that we have

$$(14) \quad \sigma_t^2 = \frac{1}{N} \sum_{i=1}^N (h_{it} - 1)^2 \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

Later in the paper, this property will be used to test the null hypothesis of convergence and to group economies into convergence clusters.

3. ASYMPTOTIC RELATIVE TRANSITION PATHS

In many empirical applications, the common growth component  $\mu_t$  will have both deterministic and stochastic elements, such as a unit root stochastic trend with drift. In that case,  $\mu_t$  is still dominated by a linear trend asymptotically. In general, we want to allow for formulations of the common growth path  $\mu_t$  that may differ from a linear trend asymptotically, and a general specification allows for the possibility that some individuals may diverge from the common growth path  $\mu_t$ , while others may converge to it. These extensions involve some technical complications that can be accommodated by allowing the functions to be regularly varying at infinity (that is, they behave asymptotically like power functions). We also allow for individual standardizations for  $X_{it}$ , so that expansion rates may differ, as well as imposing a common standardization for  $\mu_t$ . Appendix A provides some mathematical details of how these extensions and standardizations can be accomplished so that the modeling framework is more general. The present section briefly outlines the impact of these ideas and shows how to obtain a nonparametric formulation of the model (5) in which the quantities of interest are a nonparametric transition function  $\delta(\cdot)$  and a growth curve  $\mu(\cdot)$ .

In brief, we proceed as follows. Our purpose is to standardize  $X_{it}$  in (5) so that the standardized quantity approaches a limit function that embodies both the common component and the transition path. To do so, it is convenient to assume that there is a suitable overall normalization of  $X_{it}$  for which we may write equation (5) in the standardized form given by (15) below. Suppose the standardization factor for  $X_{it}$  is  $d_{iT} = T^{\gamma_i} W_i(T)$  for some  $\gamma_i > 0$  and some slowly varying function<sup>3</sup>  $W_i(T)$ , so that  $X_{it}$  grows for large  $t$  according to the power law  $t^{\gamma_i}$  up to the effect of  $W_i(t)$  and stochastic fluctuations. We may similarly suppose that the common trend component  $\mu_t$  grows according to  $t^\gamma Z(t)$  for some  $\gamma > 0$  and where  $Z$  is another slowly varying factor. Then we may write

$$(15) \quad \frac{1}{d_{iT}} X_{it} = \frac{1}{T^{\gamma_i} W_i(T)} \left( \frac{a_{it} + g_{it}}{\mu_t} \right) \mu_t = \delta_{iT} \left( \frac{t}{T} \right) \mu_T \left( \frac{t}{T} \right) + o(1),$$

where we may define the sample functions  $\mu_T$  and  $b_{iT}$  as

$$(16) \quad \mu_T \left( \frac{t}{T} \right) = \left( \frac{t}{T} \right)^\gamma \frac{Z(\frac{t}{T} T)}{Z(T)} \quad \text{and} \quad \delta_{iT} \left( \frac{t}{T} \right) = \left( \frac{t}{T} \right)^{\gamma_i - \gamma} \frac{W_i(\frac{t}{T} T) Z(T)}{W_i(T) Z(\frac{t}{T} T)},$$

as shown in Appendix A.

<sup>3</sup>That is,  $W_i(aT)/W_i(T) \rightarrow 1$  as  $T \rightarrow \infty$  for all  $a > 0$ . For example, the constant function,  $\log(T)$ , and  $1/\log(T)$  are all slowly varying functions.

Now suppose that  $t = [Tr]$ , the integer part of  $Tr$ , so that  $r$  is effectively the fraction of the sample  $T$  corresponding to observation  $t$ . Then, for such values of  $t$ , (15) leads to the asymptotic characterization

$$(17) \quad \frac{1}{d_{iT}} X_{it} \sim \delta_{iT} \left( \frac{[Tr]}{T} \right) \mu_T \left( \frac{[Tr]}{T} \right) \sim \delta_{iT}(r) \mu_T(r).$$

In (17),  $\mu_T(r)$  is the sample growth curve and  $\delta_{iT}(r)$  is the sample transition path (given  $T$  observations) for economy  $i$  at time  $T$ . It is further convenient to assume that these functions converge in some sense to certain limit functions as  $T \rightarrow \infty$ . For instance, the requirement that  $\delta_{iT}$  and  $\mu_T$  satisfy

$$(18) \quad \mu_T(r) \rightarrow_p \mu(r), \quad \delta_{iT}(r) \rightarrow_p \delta_i(r) \quad \text{uniformly in } r \in [0, 1],$$

where the limit functions  $\mu(r)$  and  $\delta_i(r)$  are continuous or, at least, piecewise continuous, seems fairly weak. By extending the probability space in which the functions  $\delta_{iT}$  and  $\mu_T$  are defined, (18) also includes cases where the functions may converge to limiting stochastic processes.<sup>4</sup> The limit functions  $\mu(r)$  and  $\delta_i(r)$  represent the common steady state growth curve and limiting transition curve for economy  $i$ , respectively. Further discussion, examples, and some general conditions under which the formulations (17) and (18) apply are given in Appendix A.

Combining (17) and (18), we have the following limiting behavior for the standardized version of  $X_{it}$ :

$$(19) \quad \frac{1}{d_{iT}} X_{it} \rightarrow_p X_i(r) = \delta_i(r) \mu(r).$$

With this limiting decomposition, we may think about  $\mu(r)$  as the limiting form of the common growth path and about  $\delta_i(r)$  as the limiting representation of the transition path of individual  $i$  as this individual moves toward the growth path  $\mu(r)$ . Representation (19) is sufficiently general to allow for cases where individuals approach the common growth path in a monotonic or cyclical fashion, either from below or above  $\mu(r)$ .

To illustrate (19), when  $\mu_t$  is a stochastic trend with positive drift, we have the simple standardization factor  $d_{iT} = T$  and then

$$T^{-1} \mu_{t=[Tr]} = m \frac{[Tr]}{T} + O_p(T^{-1/2}) \rightarrow_p mr$$

<sup>4</sup>For example, if  $\mu_t$  is a unit root process, then under quite general conditions we have the weak convergence  $T^{-1/2} \mu_{[Tr]} = \mu_T(r) \Rightarrow B(r)$  to a limit Brownian motion  $B$  (e.g., Phillips and Solo (1992)). After a suitable change in the probability space, we may write this convergence in probability, just as in (18).

for some constant  $m > 0$ . Similarly, the limit function  $\delta_i(r)$  may converge to  $\delta_i$  as  $T \rightarrow \infty$ . Combining the two factors gives the limiting path  $X_i(r) = \delta_i mr$  for individual  $i$ , so that the long run growth paths are linear across individuals. When there is convergence across individuals, we have limit transition curves  $\delta_i(r)$  each with the property that  $\delta_i(1) = \delta$ , for some constant  $\delta > 0$ , but which may differ for intermediate values (i.e.,  $\delta_i(r) \neq \delta_j(r)$  for some and possibly all  $r < 1$ ). In this case, each individual may transition in its own way toward a common limiting growth path given by the linear function  $X(r) = \delta mr$ . In this way, the framework permits a family of potential transitions to a common steady state.

Next we consider the asymptotic behavior of the relative transition parameter. Taking ratios to cross sectional averages in (15) removes the common trend  $\mu_t$  and leaves the standardized quantity

$$(20) \quad h_{iT} \left( \frac{t}{T} \right) = \frac{d_{iT}^{-1} X_{it}}{\frac{1}{n} \sum_{j=1}^n d_{jT}^{-1} X_{jt}} = \frac{\delta_{iT}(\frac{t}{T})}{\frac{1}{n} \sum_{j=1}^n \delta_{jT}(\frac{t}{T})},$$

which describes the relative transition of economy  $i$  against the benchmark of a full cross sectional average. Clearly,  $h_{iT}$  depends on  $n$  also, but we omit the subscript for simplicity because this quantity often remains fixed in the calculations. In view of (18), we have

$$(21) \quad h_{iT} \left( \frac{[Tr]}{T} \right) \rightarrow_p h_i(r) = \frac{\delta_i(r)}{\frac{1}{n} \sum_{j=1}^n \delta_j(r)} \quad \text{as } T \rightarrow \infty,$$

and the function  $h_i(r)$  then represents the limiting form of the relative transition curve for the individual  $i$ .

For practical purposes of implementation when the focus of interest is long run behavior in the context of macroeconomic data, it will often be preferable to remove business cycle components first. Extending (5) to incorporate a business cycle effect  $\kappa_{it}$ , we can write

$$X_{it} = \delta_{it} \mu_t + \kappa_{it}.$$

Smoothing methods offer a convenient mechanism for separating out the cycle  $\kappa_{it}$ , and we can employ filtering, smoothing, and regression methods to achieve this. In our empirical work with macroeconomic data, we have used two methods to extract the long run component  $\delta_{it} \mu_t$ . The first is the Whittaker–Hodrick–Prescott (WHP) smoothing filter.<sup>5</sup> The procedure is popular because

<sup>5</sup>Whittaker (1923) first suggested this penalized method of smoothing or “graduating” data and there has been a large subsequent literature on smoothing methods of this type (e.g., see Kitagawa and Gersch (1996)). The approach has been used regularly in empirical work in time series macroeconomics since the 1982 circulation of Hodrick and Prescott (1997).

of its flexibility, the fact that it requires only the input of a smoothing parameter, and does not require prior specification of the nature of the common trend  $\mu_t$  in  $X_{it}$ . The method is also suitable when the time series are short. In addition to the WHP filter, we employed a coordinate trend filtering method (Phillips (2005)). This is a series method of trend extraction that uses regression methods on orthonormal trend components to extract an unknown trend function. Again, the method does not rely on a specific form of  $\mu_t$  and is applicable whether the trend is stochastic or deterministic.

The empirical results reported in our applications below were little changed by the use of different smoothing techniques. The coordinate trend method has the advantage that it produces smooth function estimates and standard errors can be calculated for the fitted trend component. Kernel methods, rather than orthonormal series regressions, provide another general approach to smooth trend extraction and would also give standard error estimates. Kernel methods were not used in our practical work here because some of the time series we use are very short and comprise as few as 30 time series observations. Moreover, kernel method asymptotics for estimating stochastic processes are still largely unexplored and there is no general asymptotic theory to which we may appeal, although some specific results for Markov models have been obtained in work by Phillips and Park (1998), Guerre (2004), Karlsen and Tjøstheim (2001), and Wang and Phillips (2006).

Using the trend estimate  $\hat{\theta}_{it} = \widehat{b_{it}}\mu_t$  from the smoothing filter, the estimates

$$(22) \quad \hat{h}_{it} = \frac{\hat{\theta}_{it}}{\frac{1}{n} \sum_{i=1}^n \hat{\theta}_{it}}$$

of the transition coefficients  $h_{it} = \delta_{it}/(n^{-1} \sum_{i=1}^n \delta_{it})$  are obtained by taking ratios to cross sectional averages. Assuming a common standardization<sup>6</sup>  $d_{iT} = d_T$  for simplicity and setting  $t = [Tr]$ , we then have the estimate  $\hat{h}_i(r) = \hat{h}_{i[Tr]}$  of the limiting transition curve  $h_i(r)$  in (21). We can decompose the trend estimate  $\hat{\theta}_{it}$  as

$$(23) \quad \hat{\theta}_{it} = \theta_{it} + e_{it} = \left[ \delta_{it} + \frac{e_{it}}{\mu_t} \right] \mu_t,$$

where  $e_{it}$  is the error in the filter estimate of  $\theta_{it}$ . Since  $\mu_t$  is the common trend component, the condition  $e_{it}/\mu_t \rightarrow_p 0$  uniformly in  $i$  seems reasonable.<sup>7</sup>

<sup>6</sup>Alternatively, if the standardizations  $d_{iT}$  were known (or estimated) and were incorporated directly into the estimates  $\hat{\theta}_{it}$ , then  $\hat{h}_{it} = \hat{\theta}_{it}/(n^{-1} \sum_{i=1}^n \hat{\theta}_{it})$  would correspondingly build in the individual standardization factors. Accordingly,  $\hat{h}_{it}$  is an estimate of  $h_{it} = h_{iT}(\frac{t}{T})$  as given in (20).

<sup>7</sup>Primitive conditions under which  $e_{it}/\mu_t \rightarrow_p 0$  holds will depend on the properties of  $\mu_t$  and the selection of the bandwidth/smoothing parameter/regression number in the implementation

Then

$$\begin{aligned} \hat{h}_i(r) &= \frac{[\delta_{i\{Tr\}} + \frac{e_{i\{Tr\}}}{\mu_{\{Tr\}}}]}{\frac{1}{n} \sum_{i=1}^n [\delta_{j\{Tr\}} + \frac{e_{j\{Tr\}}}{\mu_{\{Tr\}}}] } = \frac{\delta_{iT}(\frac{t}{T})}{\frac{1}{n} \sum_{j=1}^n \delta_{jT}(\frac{t}{T})} + o_p(1) \\ &\rightarrow_p \frac{\delta_i(r)}{\frac{1}{n} \sum_{j=1}^n \delta_j(r)} = h_i(r), \end{aligned}$$

so that the relative transition curve is consistently estimated by  $\hat{h}_i(r)$ .

#### 4. MODELING AND TESTING CONVERGENCE

A general theory for the calculation of asymptotic standard errors of fitted curves of the type  $\hat{h}_i(r)$  that allow for deterministic and stochastic trend components of unknown form is presently not available in the literature and is beyond the scope of the present paper. Instead, we will confine ourselves to the important special case where the trend function involves a dominating stochastic trend (possibly with linear or polynomial drift) and the transition coefficient  $h_{it}$  is modeled semiparametrically. Our focus of attention is the development of a test for the null hypothesis of convergence and an empirical algorithm of convergence clustering.

As condition (12) states, under convergence, the cross sectional variation of  $\hat{h}_i(r)$  converges to zero as  $t \rightarrow \infty$ . We note, however, that decreasing cross sectional variation of  $\hat{h}_i(r)$  does not in itself imply overall convergence. For example, such decreasing cross sectional variation can occur when there is a local convergence within subgroups and overall divergence. Such a situation is plotted in Figure 1.

To design a statistical test for convergence, we need to take such possibilities of local subgroup convergence into account. As discussed earlier, the approach we use for this purpose is semiparametric and assumes the following general form for the loading coefficients  $\delta_{it}$ :

$$(24) \quad \delta_{it} = \delta_i + \sigma_{it}\xi_{it}, \quad \sigma_{it} = \frac{\sigma_i}{L(t)t^\alpha}, \quad t \geq 1, \quad \sigma_i > 0 \quad \text{for all } i,$$

where the components in this formulation satisfy the following conditions. Some generalization of (24) is possible, including allowance for individual specific decay rates  $\alpha_i$  and slowly varying functions  $L_i(t)$  that vary over  $i$ . These

---

of the filter. In the case of the WHP filter, this turns on the choice of the smoothing parameter ( $\lambda$ ) in the filter and its asymptotic behavior as the sample size increases. For instance, if  $\mu_t$  is dominated by a linear drift and  $\lambda \rightarrow \infty$  sufficiently quickly as  $T \rightarrow \infty$ , then the WHP filter will consistently estimate the trend effect. Phillips and Jin (2002) provided some asymptotic theory for the WHP filter under various assumptions about  $\lambda$  and the trend function.

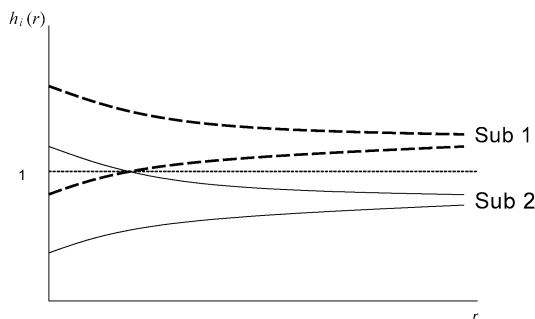


FIGURE 1.—Stylized club convergence with two subgroups.

extensions are discussed later in Remark 6. Our theory is now developed under (24) and the conditions below.

ASSUMPTION A1:  $\xi_{it}$  is iid(0, 1) with finite fourth moment  $\mu_{4\xi}$  over  $i$  for each  $t$ , and is weakly dependent and stationary over  $t$  with autocovariance sequence  $\gamma_i(h) = E(\xi_{it}\xi_{it+h})$  satisfying  $\sum_{h=1}^{\infty} h|\gamma_i(h)| < \infty$ . Partial sums of  $\xi_{it}$  and  $\xi_{it}^2 - 1$  over  $t$  satisfy the panel functional limit laws

$$(25) \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \xi_{it} \Rightarrow B_i(r) \text{ as } T \rightarrow \infty \text{ for all } i,$$

$$(26) \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} (\xi_{it}^2 - 1) \Rightarrow B_{2i}(r) \text{ as } T \rightarrow \infty \text{ for all } i,$$

where  $B_i$  and  $B_{2i}$  are independent and form independent sequences of Brownian motions with variances  $\omega_{ii}$  and  $\omega_{2ii}$ , respectively, over  $i$ .

ASSUMPTION A2: The limits

$$\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2 = v_\psi^2, \quad \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^4 = v_{4\psi},$$

$$\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2 \omega_{ii} = \omega_\xi^2, \quad \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^4 \omega_{2ii} = \omega_\eta^2,$$

$$\lim_{N \rightarrow \infty} N^{-2} \sum_{i=2}^N \sum_{j=1}^{i-1} \sigma_i^2 \sigma_j^2 \sum_{h=-\infty}^{\infty} \gamma_i(h) \gamma_j(h), \quad \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \delta_i = \delta$$

all exist and are finite and  $\delta \neq 0$ .



ASSUMPTION A3: Sums of  $\psi_{it} = \sigma_i \xi_{it}$  and  $\sigma_i^2(\xi_{it}^2 - 1)$  over  $i$  satisfy the limit laws

$$(27) \quad N^{-1/2} \sum_{i=1}^N \sigma_i \xi_{it} \Rightarrow N(0, v_\psi^2),$$

$$(28) \quad N^{-1/2} \sum_{i=1}^N \sigma_i^2(\xi_{it}^2 - 1) \Rightarrow N(0, v_{4\psi}(\mu_{4\xi} - 1))$$

as  $N \rightarrow \infty$  for all  $t$ , and the joint limit laws

$$(29) \quad T^{-1/2} N^{-1/2} \sum_{t=1}^T \sum_{i=1}^N \sigma_i \xi_{it} \Rightarrow N(0, \omega_\xi^2),$$

$$(30) \quad T^{-1/2} N^{-1/2} \sum_{t=1}^T \sum_{i=1}^N \sigma_i^2(\xi_{it}^2 - 1) \Rightarrow N(0, \omega_\eta^2),$$

$$(31) \quad T^{-1/2} \sum_{t=1}^T N^{-1} \sum_{i=2}^N \sum_{j=1}^{i-1} \sigma_i \sigma_j \xi_{it} \xi_{jt} \\ \Rightarrow N\left(0, \lim_{N \rightarrow \infty} N^{-2} \sum_{i=2}^N \sum_{j=1}^{i-1} \sigma_i^2 \sigma_j^2 \sum_{h=-\infty}^{\infty} \gamma_i(h) \gamma_j(h)\right)$$

hold as  $N, T \rightarrow \infty$ .

ASSUMPTION A4: The function  $L(t)$  in (24) is slowly varying (SV), increasing, and divergent at infinity. Possible choices for  $L(t)$  are  $\log(t + 1)$ ,  $\log^2(t + 1)$ , or  $\log \log(t + 1)$ .

Panel functional limit laws such as (25) and (26) in Assumption A1 are known to hold under a wide set of primitive conditions and were explored by Phillips and Moon (1999). These conditions allow for the variances  $\omega_{ii}$  to be random over  $i$ , in which case the limit in (25) is the mixture process  $B_i(r) = \omega_{ii}^{1/2} V_i(r)$ , where  $V_i$  is standard Brownian motion. The central limit results (27) and (28) hold under Assumptions A1 and A2, and also for cases where the components  $\xi_{it}$  are not identically distributed provided a uniform moment condition, such as  $\sup_i E(\xi_{it}^4) < \infty$ , holds. The joint limit laws (29)–(31) are high level conditions that hold under primitive assumptions of the type given in Phillips and Moon (1999).

In Assumption A4, the slowly varying function  $L(t) \rightarrow \infty$  as  $t \rightarrow \infty$ . In applications, it will generally be convenient to set  $L(t) = \log(t + 1)$  or a similar increasing slowly varying function. The presence of  $L(t)$  in (24) ensures that

$\delta_{it} \rightarrow_p \delta_i$  as  $t \rightarrow \infty$  even when  $\alpha = 0$ . Thus, when  $\delta_i = \delta$  for all  $i$ , the null hypothesis of convergence is the weak inequality constraint  $\alpha \geq 0$ , which is very convenient to test. In view of the fact that  $\delta_{it} \rightarrow_p \delta_i$  as  $t \rightarrow \infty$ , we also obtain a procedure for analyzing subgroup convergence. The presence of  $L(t)$  also assists in improving power properties of the test, as we discuss below.

The conditions for convergence in the model can be characterized as

$$\begin{aligned} \text{plim}_{k \rightarrow \infty} \delta_{it+k} = \delta & \quad \text{if and only if} \quad \delta_i = \delta \quad \text{and} \quad \alpha \geq 0, \\ \text{plim}_{k \rightarrow \infty} \delta_{it+k} \neq \delta & \quad \text{if and only if} \quad \delta_i \neq \delta \quad \text{or} \quad \alpha < 0. \end{aligned}$$

Note that there is no restriction on  $\alpha$  under divergence when  $\delta_i \neq \delta$ . However, we have a particular interest in the case of divergence with  $\delta_i \neq \delta$  and  $\alpha \geq 0$ , as this allows for the example considered in Figure 1 where there is the possibility of local convergence to multiple equilibria. This case is likely to be important in empirical applications where there is evidence of clustering behavior, for example, in individual consumption or income patterns over time. In such cases, we may be interested in testing whether elements in a panel converge within certain subgroups.

The remainder of this section develops an econometric methodology for testing convergence in the above context and provides a step by step procedure for practical implementation. In particular, we show how to test the null hypothesis of convergence, develop asymptotic properties of the test, including a local power analysis, and provide an intuitive discussion of how the test works. We also discuss a procedure for detecting panel clusters. Proofs and related technical material are given in the [Appendix](#).

#### 4.1. A Regression Test of Convergence

The following procedure is a regression  $t$  test of the null hypothesis of convergence

$$\mathcal{H}_0: \delta_i = \delta \quad \text{and} \quad \alpha \geq 0,$$

against the alternative  $\mathcal{H}_A: \delta_i \neq \delta$  for all  $i$  or  $\alpha < 0$ .

*Step 1:* Construct the cross sectional variance ratio  $H_1/H_t$ , where

$$(32) \quad H_t = \frac{1}{N} \sum_{i=1}^N (h_{it} - 1)^2, \quad h_{it} = \frac{X_{it}}{N^{-1} \sum_{i=1}^N X_{it}}.$$

*Step 2:* Run the following regression and compute a conventional robust  $t$  statistic  $t_{\hat{b}}$  for the coefficient  $\hat{b}$  using an estimate of the long run variance of

the regression residuals:

$$(33) \quad \log\left(\frac{H_1}{H_t}\right) - 2\log L(t) = \hat{a} + \hat{b} \log t + \hat{u}_t,$$

for  $t = [rT], [rT] + 1, \dots, T$  with  $r > 0$ .

In this regression we use the setting  $L(t) = \log(t + 1)$  and the fitted coefficient of  $\log t$  is  $\hat{b} = 2\hat{\alpha}$ , where  $\hat{\alpha}$  is the estimate of  $\alpha$  in  $H_0$ . Note that data for this regression start at  $t = [rT]$ , the integer part of  $rT$  for some fraction  $r > 0$ . As discussed below, we recommend  $r = 0.3$ .

*Step 3:* Apply an autocorrelation and heteroskedasticity robust one-sided  $t$  test of the inequality null hypothesis  $\alpha \geq 0$  using  $\hat{b}$  and a HAC standard error. At the 5% level, for example, the null hypothesis of convergence is rejected if  $t_{\hat{b}} < -1.65$ .

Under the convergence hypothesis,  $h_{it} \rightarrow 1$  and  $H_t \rightarrow 0$  as  $t \rightarrow \infty$  for given  $N$ . In Appendix B it is shown in (68) and (71) that  $H_t$  then has the logarithmic form

$$(34) \quad \log H_t = -2\log L(t) - 2\alpha \log t + 2\log \frac{v_{\psi N}}{\delta} + \epsilon_t,$$

with

$$(35) \quad \epsilon_t = \frac{1}{\sqrt{N}} \frac{\eta_{Nt}}{v_{\psi N}^2} - \frac{2}{\delta} \frac{1}{t^\alpha L(t)} \psi_t + \frac{1}{\delta^2} \frac{1}{t^{2\alpha} L(t)^2} \psi_t^2 + O_p\left(\frac{1}{N}\right),$$

where  $v_{\psi N}^2 = N^{-1}(1 - N^{-1}) \sum_{i=1}^N \sigma_i^2 \rightarrow v_\psi^2$  as  $N \rightarrow \infty$ ,  $\eta_{Nt} = N^{-1/2} \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1)$ , and  $\psi_t = N^{-1} \sum_{i=1}^N \sigma_{it} \xi_{it}$ . From (34) we deduce the simple regression equation

$$(36) \quad \log \frac{H_1}{H_t} - 2\log L(t) = a + b \log t + u_t,$$

where  $b = 2\alpha$ ,  $u_t = -\epsilon_t$ , and the intercept  $a = \log H_1 - 2\log(v_{\psi N}/\delta) = -2\log L(1) + u_1$  does not depend on  $\alpha$ .

Under convergence,  $\log(H_1/H_t)$  diverges to  $\infty$ , either as  $2\log L(t)$  when  $\alpha = 0$  or as  $2\alpha \log t$  when  $\alpha > 0$ . Thus, when the null hypothesis  $\mathcal{H}_0$  applies, the dependent variable diverges whether  $\alpha = 0$  or  $\alpha > 0$ . Divergence of  $\log(H_1/H_t)$  corresponds to  $H_t \rightarrow 0$  as  $t \rightarrow \infty$ . Thus,  $\mathcal{H}_0$  is conveniently tested in terms of the weak inequality null  $\alpha \geq 0$ . Since  $\alpha$  is a scalar, this null can be tested using a simple one-sided  $t$  test.

Under the divergence hypothesis  $\mathcal{H}_A$ , for instance, when  $\delta_i \neq \delta$  for all  $i$ ,  $H_t$  is shown in Appendix B to converge to a positive quantity as  $t \rightarrow \infty$ . Hence,

under  $\mathcal{H}_A$ , the dependent variable  $\log(H_1/H_t) - 2\log L(t)$  diverges to  $-\infty$  in contrast to the null  $\mathcal{H}_0$ , under which  $\log(H_1/H_t)$  diverges to  $\infty$ . The term  $-2\log L(t)$  in (36) therefore serves as a penalty that helps the test on the coefficient of the  $\log t$  regressor to discriminate the behavior of the dependent variable under the alternative from that under the null. In particular, when  $\alpha = 0$  and  $\delta_i \neq \delta$  for some  $i$ , the inclusion of  $\log L(t)$  produces a negative bias in the regression estimate of  $b$  since  $\log t$  and  $-2\log L(t)$  are negatively correlated. The  $t$  statistic for  $\hat{b}$  then diverges to negative infinity and the test is consistent even in this (boundary) case where  $\alpha = 0$ .

Discarding some small fraction  $r$  of the time series data helps to focus attention in the test on what happens as the sample size gets larger. The limit distribution and power properties of the test depend on the value of  $r$ . Our simulation experience indicates that  $r = 0.3$  is a satisfactory choice in terms of both size and power. Appendix B provides details of the construction of the regression equation under the null and alternative, and derives the asymptotic properties of the  $\log t$  test.

#### 4.2. Asymptotic Properties of the $\log t$ Convergence Test

The following result gives the limit theory for the least squares estimate  $\hat{b}$  of the slope coefficient  $b$  in the  $\log t$  regression equation (36) and the associated limit theory of the regression  $t$  test under the null  $\mathcal{H}_0$ .

**THEOREM 1**—Limit Theory under  $\mathcal{H}_0$ : *Let the panel  $X_{it}$  defined in (2) have common factor  $\mu_t$  and loading coefficients  $\delta_{it}$  that follow the generating process (24) and satisfy Assumptions A1–A4. Suppose that the convergence hypothesis  $\mathcal{H}_0$  holds and the regression equation (36) is estimated with time series data over  $t = [Tr], \dots, T$ , for some  $r > 0$ . Suppose further that if  $\alpha > 0$ ,  $T^{1/2}/(T^{2\alpha}L(T)^2N^{1/2}) \rightarrow 0$  and if  $\alpha = 0$ ,  $T^{1/2}/N \rightarrow 0$  as  $T, N \rightarrow \infty$ .*

(a) *The limit distribution of  $\hat{b}$  is*

$$(37) \quad \sqrt{NT}(\hat{b} - b) \Rightarrow N(0, \Omega^2),$$

where  $\Omega^2 = \omega_\eta^2/v_\psi^4\{(1-r) - (\frac{r}{1-r})\log^2 r\}^{-1}$ ,  $\omega_\eta^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_i^4 \omega_{2ii}$ , and  $v_\psi^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_i^2$ .

(b) *The limit distribution of the regression  $t$  statistic is*

$$t_{\hat{b}} = \frac{\hat{b} - b}{s_{\hat{b}}} \Rightarrow N(0, 1),$$

where

$$s_{\hat{b}}^2 = \widehat{\text{var}}_r(\hat{u}_t) \left[ \sum_{t=[Tr]}^T \left( \log t - \frac{1}{T - [Tr] + 1} \sum_{t=[Tr]}^T \log t \right)^2 \right]^{-1}$$

and  $\widehat{\text{Ivar}}_r(\hat{u}_t)$  is a conventional HAC estimate formed from the regression residuals  $\hat{u}_t = \log(H_1/H_t) - 2 \log L(t) - \hat{a} - \hat{b} \log t$  for  $t = [Tr], \dots, T$ . The estimate  $TN\hat{s}_b^2$  is consistent for  $\Omega^2$  as  $T, N \rightarrow \infty$ .

REMARK 1: As shown in Appendix B, to avoid asymptotic collinearity in the regressors (the intercept and  $\log t$ ), the regression (36) may be rewritten as

$$(38) \quad \log \frac{H_1}{H_t} - 2 \log L(t) = a^* + b \log \frac{t}{T} + u_t,$$

where  $a^* = a + b \log T$ . The estimate  $\hat{b}$  then has the form

$$\hat{b} - b = \left( \sum_{t=[Tr]}^T \tau_t u_t \right) \left( \sum_{t=[Tr]}^T \tau_t^2 \right)^{-1},$$

involving the demeaned regressor  $\tau_t = (\log \frac{t}{T} - \overline{\log \frac{t}{T}}) = \log t - \frac{1}{T-[Tr]+1} \times \sum_{t=[Tr]}^T \log t$ , where  $\overline{\log \frac{t}{T}} = \frac{1}{T-[Tr]+1} \sum_{t=[Tr]}^T \log \frac{t}{T}$ .

REMARK 2: The convergence rate for  $\hat{b}$  under the null of convergence is  $O(\sqrt{NT})$  and is the same for all  $\alpha \geq 0$ . Although the regression is based on only  $O(T)$  observations (specifically,  $T - [Tr]$  observations), the convergence rate is faster than  $\sqrt{T}$  because the dependent variable  $\log H_t$  involves a cross section average (32) over  $N$  observations and this averaging affects the order of the regression error  $u_t = -\epsilon_t$ , as is apparent in (35). In particular, the leading term of  $\epsilon_t$  is  $O_p(N^{-1/2})$  when the relative rate condition  $T^{1/2}/(T^{2\alpha}L(T)^2N^{1/2}) \rightarrow 0$  holds for  $\alpha > 0$  or when  $\frac{T^{1/2}}{N} \rightarrow 0$  holds if  $\alpha = 0$ . These rate conditions require that  $N$  does not pass to infinity too slowly relative to  $T$ . Otherwise the limit distribution (37) involves a bias term, as discussed in Appendix B.

REMARK 3: The quantity  $\Omega_u^2 = \omega_\eta^2/v_\psi^4$  in the asymptotic variance formula is the limit of a cross section weighted average of the long run variances  $\omega_{2ii}$  of  $\eta_{it} = \xi_{it}^2 - 1$ . Appendix B shows how this average long run variance can be estimated by a standard HAC estimate, such as the truncated kernel estimate  $\widehat{\text{Ivar}}_r(\hat{u}_t) = \sum_{l=-M}^M \frac{1}{T-[Tr]} \sum_{[Tr] \leq t, t+l \leq T} \hat{u}_t \hat{u}_{t+l}$  given in (93) and formed in the usual way from the residuals  $\hat{u}_t$  with bandwidth (truncation) parameter  $M$ . Of course, other kernels may be used and the same asymptotics apply for standard bandwidth expansion rates for  $M$  such as  $\frac{M}{\sqrt{T}} + \frac{1}{M} \rightarrow 0$ , as discussed in Appendix B.

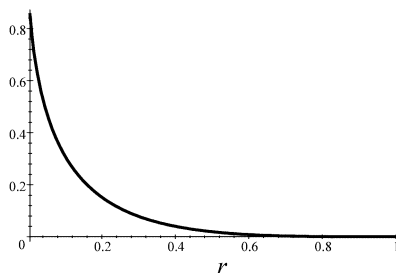


FIGURE 2.—The precision curve  $(1-r) - \left(\frac{r}{1-r}\right) \log^2 r$ .

REMARK 4: The precision of the estimate  $\hat{b}$  is measured by the reciprocal of  $\Omega^2$  and depends on the factor

$$(1-r) - \left(\frac{r}{1-r}\right) \log^2 r \rightarrow \begin{cases} 1, & r \rightarrow 0, \\ 0, & r \rightarrow 1. \end{cases}$$

So the asymptotic variance of  $\hat{b}$  diverges as  $r \rightarrow 1$ , corresponding to the fact that the fraction of the sample used in the regression goes to zero in this event. The precision curve is graphed in Figure 2.

REMARK 5: We call the one-sided regression  $t$  test based on  $t_b$  the log  $t$  test. To test the hypothesis  $b = 2\alpha \geq 0$ , we fit the regression model (36), or equivalently (38), over  $t = [Tr], \dots, T$  and compute the  $t$  statistic  $t_b = \hat{b}/s_b$ . As the following result shows, this test is consistent against alternatives where the idiosyncratic components  $\delta_{it}$  diverge (i.e., when  $b = 2\alpha < 0$ ) as well as alternatives where the  $\delta_{it}$  converge, but to values  $\delta_i$  that differ across  $i$ . Both cases seem important in practical applications and it is an advantage of the log  $t$  test that it is consistent against both.

REMARK 6—Models with Heterogeneous Decay Rates  $\alpha_i, L_i(t)$ : As indicated earlier, the framework based on (24) may be extended by allowing for individual specific decay rates in the loading coefficients. In such cases, both the rate parameter  $\alpha$  and the slowly varying function  $L(t)$  may vary over  $i$ . Relaxation of (24) in this way may be useful in some applications where there is greater heterogeneity across the population in terms of temporal responses to the common trend effect. For instance, some individuals may converge faster than others. Alternatively, in cases where there are subgroups in the population, there may be heterogeneity in the temporal responses among the different groups. To accommodate these extensions, we may replace (24) with the model

$$(39) \quad \delta_{it} = \delta_i + \sigma_{it} \xi_{it}, \quad \sigma_{it} = \frac{\sigma_i}{L_i(t)t^{\alpha_i}}, \quad t \geq 1, \quad \sigma_i > 0 \quad \text{for all } i,$$

where under the null hypothesis  $H_0$  the decay rates  $\alpha_i \geq 0 \forall i$  and where each of the slowly varying functions has the form  $L_i(t) = \log^{\beta_i} t$  for  $\beta_i > 0$  and  $t$  large, thereby satisfying Assumption A4. This specification should be sufficiently general to include most cases of practical interest, the most important extension being to allow for individual rate effects  $\alpha_i$ . To extend our asymptotic development to apply under (39), it is convenient to assume that the rates  $\alpha_i$ ,  $\beta_i$  and the standard error  $\sigma_i$  are drawn from independent distributions with smooth densities  $f_\alpha(\alpha)$ ,  $f_\beta(\beta)$ ,  $f_\sigma(\sigma)$  supported over  $\alpha \in [a, A]$  for some  $a \geq 0$ ,  $A > 0$ , and  $\beta \in [b, B]$  for some  $b, B > 0$  and  $\sigma \in [0, \infty)$ . In this more general framework, we still have  $h_{it} \rightarrow_p 1$  and  $H_t \rightarrow_p 0$  as  $t \rightarrow \infty$  under the convergence hypothesis  $H_0$ . Specifically, we have

$$h_{it} - 1 = \frac{\psi_{it} - \psi_t}{\delta + \psi_t}, \quad \psi_{it} = \frac{\sigma_i \xi_{it}}{L_i(t)t^{\alpha_i}},$$

where

$$\psi_t = N^{-1} \sum_{i=1}^N \frac{\sigma_i \xi_{it}}{L_i(t)t^{\alpha_i}}$$

has mean zero and, assuming  $\int_0^\infty \sigma^2 f_\sigma(\sigma) d\sigma < \infty$ ,

$$N^{-1} \sum_{i=1}^N \frac{\sigma_i^2}{(\log^{2\beta_i} t)t^{2\alpha_i}} \rightarrow \int_0^\infty \sigma^2 f_\sigma(\sigma) d\sigma \int_b^B \frac{f_\beta(\beta) d\beta}{\log^{2\beta} t} \int_a^A \frac{f_\alpha(\alpha) d\alpha}{t^{2\alpha}}$$

as  $N \rightarrow \infty$  by the strong law of large numbers. Using integration by parts, the following asymptotic expansions are obtained for large  $t$ :

$$\int_b^B \frac{f_\beta(\beta) d\beta}{\log^{2\beta} t} = \frac{f_\beta(b)}{(\log^{2b} t)(\log \log t)} \{1 + o_p(1)\}$$

and

$$\int_a^A \frac{f_\alpha(\alpha) d\alpha}{t^{2\alpha}} = \frac{f_\alpha(a)}{t^{2a} \log t} \{1 + o_p(1)\}.$$

Then  $\psi_t = O_p(N^{-1/2} t^{-2a} (\log^{2b} t)^{-1} (\log \log t)^{-1})$  and

$$\begin{aligned} H_t &= N^{-1} \sum_{i=1}^N \frac{(\psi_{it} - \psi_t)^2}{(\delta + \psi_t)^2} = \frac{N^{-1} \sum_{i=1}^N \psi_{it}^2}{\delta^2} \left\{ 1 + o_p\left(\frac{1}{N^{1/2} t^{2a} \log^{1+2b} t}\right) \right\} \\ &= \frac{\int_0^\infty \sigma^2 f_\sigma(\sigma) d\sigma f_\alpha(a) f_\beta(\beta)}{t^{2a} (\log^{2b} t) (\log \log t) \delta^2} \{1 + o_p(1)\}, \end{aligned}$$

since

$$\begin{aligned} N^{-1} \sum_{i=1}^N \psi_{it}^2 &= N^{-1} \sum_{i=1}^N \frac{\sigma_i^2 \xi_{it}^2}{L_i(t)^2 t^{2\alpha_i}} \\ &= N^{-1} \sum_{i=1}^N \frac{\sigma_i^2}{L_i(t)^2 t^{2\alpha_i}} + N^{-1} \sum_{i=1}^N \frac{\sigma_i^2 (\xi_{it}^2 - 1)}{L_i(t)^2 t^{2\alpha_i}} \\ &= \frac{\int_0^\infty \sigma^2 f_\sigma(\sigma) d\sigma f_\alpha(a) f_\beta(\beta)}{t^{2a} (\log^{2b} t) (\log \log t)} \{1 + o_p(1)\}. \end{aligned}$$

It follows that

$$\begin{aligned} (40) \quad \log H_t &= \log \left\{ \frac{\int_0^\infty \sigma^2 f_\sigma(\sigma) d\sigma f_\alpha(a) f_\beta(\beta)}{t^{2a} (\log^{2b} t) (\log \log t) \delta^2} \right\} \{1 + o_p(1)\} \\ &= \log \left\{ \delta^{-2} \int_0^\infty \sigma^2 f_\sigma(\sigma) d\sigma f_\alpha(a) f_\beta(\beta) \right\} - 2a \log t \\ &\quad - \log\{(\log^{2b} t) (\log \log t)\} + o_p(1) \\ &= 2 \log \frac{\nu}{\delta} - 2a \log t - 2 \log L(t) + o_p(1) \end{aligned}$$

with  $\nu = \int_0^\infty \sigma^2 f_\sigma(\sigma) d\sigma f_\alpha(a) f_\beta(\beta)$  and  $L(t)^2 = (\log^{2b} t) (\log \log t)$ . Thus, apart from the particular form of the slowly varying function  $L(t)$ ,  $\log H_t$  has the same specification for large  $t$  as (34) for the homogeneous decay rate case. Note that in (40) the term involving  $\log t$  has coefficient  $-2a$ , where  $a$  is the lower bound of the support of the decay rates  $\alpha_i$ . Corresponding to (40), we obtain a regression equation with the same leading systematic form as the model (36) given above. The regression equation is, in fact, identical to (36) when the slowly varying component in (39) is homogeneous across  $i$  and only the rate effects  $\alpha_i$  are heterogeneous. The approach to testing convergence using the  $\log t$  regression (40) can therefore be applied when the model is of the more general form (39), allowing for heterogeneity in the decay rates across the population. In such cases the coefficient of the  $\log t$  regressor is the lower bound of the decay rates across the population under the null. Under the alternative hypothesis of nonconvergence, when we allow for heterogeneity in the decay rates  $\alpha_i$ , we may have  $\alpha_i < 0$  for some  $i$  and  $\alpha_i > 0$  for other  $i$ . In such cases, there may be the possibility of subgroup convergence among those individuals with positive  $\alpha_i$  and this may be tested using the clustering algorithm described below.

**THEOREM 2—Test Consistency Under  $\mathcal{H}_A$ :** *Suppose the alternative hypothesis  $\mathcal{H}_A$  holds and the other conditions of Theorem 1 apply.*



(a) If  $\alpha \geq 0$  and  $\delta_i \sim \text{iid}(\delta, \sigma_\delta^2)$  with  $\sigma_\delta^2 > 0$ , then

$$\hat{b} \rightarrow_p 0, \quad t_{\hat{b}} = \frac{\hat{b}}{s_{\hat{b}}} \rightarrow -\infty,$$

as  $T, N \rightarrow \infty$ .

(b) If  $\gamma = -\alpha > 0$ ,  $\delta_i = \delta$  for all  $i$ , and  $T^{\gamma-1/2}/(\sqrt{N}L(T)) + \frac{1}{T} + \frac{1}{N} \rightarrow 0$ , then

$$\frac{\sqrt{NTL(T)}}{T^\gamma}(\hat{b} - b) \Rightarrow \frac{2}{\delta}N(0, Q_\xi^2),$$

where

$$Q_\xi^2 = \omega_\xi^2 \left[ \int_r^1 \left\{ \log s - \frac{1}{1-r} \int_r^1 \log p dp \right\}^2 s^{2\gamma} ds \right] \\ \times \left[ (1-r) - \left( \frac{r}{1-r} \right) \log^2 r \right]^{-1}$$

and  $t_{\hat{b}} = \hat{b}/s_{\hat{b}} \rightarrow -\infty$ .

(c) If  $\gamma = -\alpha > 0$ ,  $\delta_i = \delta$  for all  $i$ , and  $\sqrt{N}L(T)T^{-\gamma} + T^{-1} + N^{-1} \rightarrow 0$ , then

$$\hat{b} \rightarrow_p 0, \quad t_{\hat{b}} = \frac{\hat{b}}{s_{\hat{b}}} \rightarrow -\infty.$$

In all cases the test is consistent.

REMARK 7: The alternative hypothesis under (a) involves  $\delta_i \sim \text{iid}(\delta, \sigma_\delta^2)$  so that  $\delta_i \neq \delta$  for all  $i$ . As is clear from the proof of Theorem 2, it is sufficient for the result to hold that  $\delta_i \neq \delta$  for  $i \in G$ , some subgroup of the panel, and for  $N_G = \#\{i \in G\}$ , the number of elements in  $G$ , to be such that  $N_G/N \rightarrow \lambda > 0$  as  $N \rightarrow \infty$ . Test consistency therefore relies on the existence of enough economies with different  $\delta_i$ . The condition will be satisfied, for instance, in cases like that shown in Figure 1 where there are two convergence clubs with membership proportions  $\lambda$  and  $1 - \lambda$ . Obviously, the convergence null does not hold in this case but the cross sectional variation of the relative transition parameters measured by  $H_t$  may well decrease over time. Calculation reveals that

$$\lim_{\substack{N \rightarrow \infty \\ t \rightarrow \infty}} H_t = \frac{\lambda(1-\lambda)(\delta_A - \delta_B)^2}{(\lambda\delta_A + (1-\lambda)\delta_B)^2} := H_{AB},$$

where  $\delta_A = \lim_{i \in A, t \rightarrow \infty} \delta_{it}$  and  $\delta_B = \lim_{i \in B, t \rightarrow \infty} \delta_{it}$  for two subgroups  $G_A$  and  $G_B$  with membership shares  $\lambda = \lim_{N \rightarrow \infty} (N_{G_A}/N)$  and  $1 - \lambda = \lim_{N \rightarrow \infty} (N_{G_B}/N)$ . Clearly  $H_{AB}$  will be close to zero when the group means  $\delta_A$  and  $\delta_B$  are close.

REMARK 8: In part (a) of Theorem 2,  $\hat{b} \rightarrow_p 0$ . The heuristic explanation is that, when  $\alpha \geq 0$  and  $\delta_i \sim \text{iid}(\delta, \sigma_\delta^2)$ ,  $H_t$  tends to a positive constant, so that the dependent variable in (36) behaves like  $-2 \log L(t)$  for large  $t$ . Since  $\log L(t)$  is the log of a slowly varying function and grows more slowly than  $\log t$ , its regression coefficient on  $\log t$  is expected to be zero. More specifically, the regression of  $-2 \log L(t)$  on  $\log(\frac{t}{T})$  produces a slope coefficient that is negative and tends to zero like  $-\frac{2}{\log T}$ , as is shown in (102) in the Appendix. Since  $s_{\hat{b}} \rightarrow_p 0$  also and at a faster rate, the  $t$  ratio  $t_{\hat{b}}$  then diverges to negative infinity and the test is consistent.

REMARK 9: In part (b),  $\hat{b}$  is consistent to  $b = 2\alpha < 0$ , but at a reduced rate of convergence. The test is again consistent because  $t_{\hat{b}} = (\hat{b} - b)/s_{\hat{b}} + (b/s_{\hat{b}}) \rightarrow -\infty$  by virtue of the sign of  $b$ .

REMARK 10: In part (c), we again have  $\hat{b} \rightarrow_p 0$ . In this case, the  $\delta_{it}$  have divergent behavior and  $H_t = O_p(N)$ . Hence, in the time series regression (36), the dependent variable behaves like  $-2 \log L(t)$  for large  $t$ , and the slope coefficient is negative and tends to zero like  $-2/\log T$ , just as in part (a).

It is also interesting to analyze the local asymptotic properties of the  $\log t$  test. The following result analyzes the asymptotic consistency of the test for local departures from the null of the form

$$(41) \quad \mathcal{H}_{LA}: \delta_i \sim \text{iid}(\delta, c^2 T^{-2\omega}).$$

Such departures measure deviations from the null  $\mathcal{H}_0$  in terms of a distance  $|\delta_i - \delta|$  that is local to zero and of magnitude  $O_p(T^{-\omega})$  for some parameter  $\omega > 0$ . This local consistency result turns out to be useful in the clustering algorithm developed below.

**THEOREM 3—Local Asymptotic Consistency:** *Suppose the local alternative hypothesis  $\mathcal{H}_{LA}$  holds and the other conditions of Theorem 1 apply.*

(a) *Under (41) with  $\omega \leq \alpha$ ,  $\hat{b} \rightarrow_p 0$  and  $t_{\hat{b}} = \frac{\hat{b}}{s_{\hat{b}}} \rightarrow -\infty$  as  $T, N \rightarrow \infty$ . The test is consistent and the rate of divergence of  $t_{\hat{b}}$  is  $O((\log T)T^{1/2}/M^{1/2})$  for all choices of bandwidth  $M \leq T$ .*

(b) *Under the local alternative (41) with  $\omega > \alpha$  and when  $(T^{2(\omega-\alpha)})/(\sqrt{N} \times L(T)^2) \rightarrow 0$  as  $T, N \rightarrow \infty$ ,*

$$\hat{b} - b = -\frac{c^2}{v_{\psi N}^2} h(r) \frac{L(T)^2}{T^{2(\omega-\alpha)}} \{1 + o_p(1)\} \rightarrow_p 0,$$

$$t_{\hat{b}} \rightarrow \begin{cases} \infty, & \text{for } b = 2\alpha > 0, \\ -\infty, & \text{for } b = 2\alpha = 0, \end{cases}$$

where

$$h(r) = \left[ \int_r^1 \left\{ \log s - \frac{1}{1-r} \int_r^1 \log p \, dp \right\} s^{2\alpha} \, ds \right] \times \left[ (1-r) - \left( \frac{r}{1-r} \right) \log^2 r \right]^{-1}.$$

REMARK 11: In the proof of part (a), it is shown that

$$\hat{b} = -\frac{2}{\log T} + O_p\left(\frac{1}{\log^2 T}\right), \quad t_{\hat{b}} = -\frac{2}{\log T} \times O_p\left(\frac{(\log^2 T)T^{1/2}}{M^{1/2}}\right),$$

and so both  $\hat{b}$  and the  $t$  ratio  $t_{\hat{b}}$  have the same asymptotic behavior as in the case of fixed alternatives of the form  $\delta_i \sim \text{iid}(\delta, \sigma_\delta^2)$  considered in part (a) of Theorem 2. The reason for this equivalence is that under (41) the idiosyncratic effects have the form

$$\delta_{it} = \delta_i + \sigma_{it}\xi_{it} = \delta + \frac{\xi_i}{T^\omega} + \frac{\sigma_i \xi_{it}}{L(t)t^\alpha},$$

where the  $\xi_i$  are  $\text{iid}(0, c^2)$ . When  $\omega \leq \alpha$ , the final term is of smaller order than  $\delta_i = \delta + \xi_i/T^\omega$  and so the  $\log t$  regression has the same discriminatory power to detect the departure of  $\delta_i$  from  $\delta$  as it does in the case where the  $\delta_i$  are  $\text{iid}(\delta, \sigma_\delta^2)$ . We say that the test is *locally consistent* in the sense that it is consistent against local departures from the null of the form (41).

REMARK 12: When  $\omega > \alpha > 0$ , the test has negligible power to detect alternatives of the form (41). Since  $\omega > \alpha$ , this is explained by the fact that the alternatives are closer to the null than the convergence rate, so they elude detection. However, when  $\omega > \alpha = 0$ , the convergence rate of the idiosyncratic effects  $\delta_{it}$  is  $1/L(t)$  and is slower than any power rate. In this case, remarkably the test is consistent, although the divergence rate of the statistic is only  $O_p(T^{1/2}/M^{1/2})$ , which diverges when  $\frac{M}{T} \rightarrow 0$  (i.e., for standard bandwidth choices in HAC estimation). The consistency is explained by the fact that, even though the alternatives  $\delta_i \neq \delta$  are still very close to the null in (41), the rate of convergence of  $\delta_{it}$  is so slow that the test is able to detect the local departures from the null.

REMARK 13: Theorem 3 may be interpreted to include the case where there are additional individual effects in the formulation of the nonlinear factor model. For instance, suppose the panel  $X_{it}$  involves an additive effect so that

$$X_{it}^* = X_{it} + a_i = \left( \frac{a_i}{\mu_t} + \delta_{it} \right) \mu_t = \delta_{it}^* \mu_t$$

with

$$\delta_{it}^* = \delta_i + \frac{a_i}{\mu_i} + \frac{\sigma_i \xi_{it}}{L(t)t^\alpha} := \delta_i^* + \frac{\sigma_i \xi_{it}}{L(t)t^\alpha}.$$

Suppose the additive effect is common and  $a_i = a$  for all  $i$ . Then if  $\delta_i = \delta$  for all  $i$  and if the common trend  $\mu_t = O_p(t^\theta)$  with  $\theta > \alpha$ , we have

$$\delta_{it}^* = \delta + \frac{\sigma_i \xi_{it}}{L(t)t^\alpha} + o_p\left(\frac{1}{L(t)t^\alpha}\right),$$

in which case Theorem 1 holds when  $\alpha > 0$ . If  $\delta_i = \delta$  for all  $i$  and  $a_i \sim \text{iid}(0, c^2)$ , then the model is equivalent to that considered in Theorem 3, in which case the presence of the individual effects may be detectable, depending on the relative magnitudes of the decay parameters  $\theta$  and  $\alpha$ .

REMARK 14: Appendix B provides some discussion of the choice of the slowly varying function  $L(t)$  in terms of the induced asymptotic power properties. It is shown there that among choices such as  $L(t) = \log t$ ,  $\log_2 t = \log \log t$ , and  $\log_3 t = \log \log \log t$ , and where  $t$  is large, the choice  $L(t) = \log t$  is preferred in terms of asymptotic power. This choice was also found to work well in simulations and is recommended in practice.

#### 4.3. Club Convergence and Clustering

Rejection of the null of convergence does not imply there is no evidence of convergence in subgroups of the panel. Many possibilities exist as we move away from a strict null of full panel convergence. Examples include the possible existence of convergence clusters around separate points of equilibria or steady state growth paths, as well as cases where there may be both convergence clusters and divergent members in the full panel. If there are local equilibria or club convergence clusters, then it is of substantial interest to be able to identify these clusters, determine the number of clusters, and resolve individuals into respective groups. In the empirical growth literature, the great diversity in economic performance across countries has made searching for convergence clubs a central issue. For example, Canova (2004), Canova and Marcet (1995), Durlauf and Johnson (1995), and Quah (1996, 1997) all attempt to classify and identify convergence clubs.

Perhaps the simplest case for empirical analysis occurs when subgroups can be suitably categorized by identifying social or economic characteristics. For example, gender, education, age, region, or ethnicity could be identifying attribute variables. Under clustering by such covariates, convergence patterns within groups may be conducted along the lines outlined above using  $\log t$  regressions. For instance, if the convergence null for individual consumption behavior in a particular region (or age group) were rejected and it was suspected

that gender or ethnicity differences were a factor in the rejection  $\log t$  convergence tests could be rerun for different panels subgrouped according to gender and ethnicity to determine whether convergence was empirically supported within these subgroups.

Alternatively, if convergence subgroups can be determined by an empirical clustering algorithm, then it becomes possible to subsequently explore links between the empirical clusters and various social and economic characteristics. In this case, the club convergence grouping becomes a matter for direct empirical determination. A simple algorithm based on repeated  $\log t$  regressions is developed here to provide such an empirical approach to sorting individuals into subgroups.

To initiate the procedure, we start with the assumption that there is a “core subgroup”  $G_K$  with convergence behavior, that this subgroup contains at least  $K$  members, and that the subgroup is known. Below we discuss a method for detecting the initial core subgroup. Next we consider adding an additional individual ( $K + 1$ , say) to  $G_K$ . To assess whether the new individual belongs to  $G_K$ , we perform a  $\log t$  test. If  $K + 1$  belongs to  $G_K$ , the point estimate of  $b$  in the  $\log t$  test will not be significantly negative and the null hypothesis will be supported in view of Theorem 1. Otherwise, the point estimate of  $b$  will depend on the size of  $K$  and the extent of the deviation from the null. To see this, set  $\delta_i = \delta_A$  for  $i = 1, \dots, K$  and  $\delta_i = \delta_B$  for  $i = K + 1$ . The variation of  $\delta_i$  in the augmented subgroup is given by

$$(42) \quad \sigma^2 = \frac{1}{K + 1} \sum_{i=1}^{K+1} (\delta_i - \bar{\delta})^2 = \frac{K}{(K + 1)^2} (\delta_A - \delta_B)^2,$$

where

$$\bar{\delta} = \frac{1}{K + 1} \sum_{i=1}^{K+1} \delta_i = \frac{K\delta_A}{K + 1} + \frac{\delta_B}{K + 1}.$$

As  $K \rightarrow \infty$ ,  $\sigma^2 = O(K^{-1}) \rightarrow 0$  and  $\bar{\delta} \rightarrow \delta_A$ .

An asymptotic analysis of club convergence patterns in such cases can be based on local alternatives of the form

$$(43) \quad \delta_i \sim \text{iid } N(\delta, c^2 K^{-1}).$$

Appendix C provides such an analysis. It is shown there that when  $c^2 > 0$  and  $K/T^{2\alpha} \rightarrow 0$  as  $T \rightarrow \infty$ , the procedure is consistent in detecting departures of the form (43) for all bandwidth choices  $M \leq T$ . Given that  $\sigma^2 = O(K^{-1})$  in (42), this analysis also covers the case where  $\delta_i = \delta_A$  for  $i = 1, \dots, K$  and  $\delta_{K+1} = \delta_B \neq \delta_A$ . On the other hand, when  $\delta_i = \delta_A$  for  $i = 1, \dots, K + 1$ , the null hypothesis holds for  $N = K + 1$  and  $t_{\hat{b}} = (\hat{b} - b)/s_{\hat{b}} \Rightarrow N(0, 1)$ , as in Theorem 1.

When  $T^{2\alpha}/K \rightarrow 0$ , the alternatives (43) are very close to the null, relative to the convergence rate except when  $\alpha = 0$ . This case is analogous to case (b) of Theorem 3 and as that theorem shows, the test is inconsistent and unable to detect the departure from the null when  $\alpha > 0$ . However, when  $\alpha = 0$ , the convergence rate is slowly varying under the null, and Theorem 3 shows that the test is in fact consistent against local alternatives of the form (43). In effect, although the alternatives are very close (because  $K$  is large), the convergence rate is slow (slower than any power rate) and this suffices to ensure that the test is consistent as  $T \rightarrow \infty$ .

We now suggest the following method of finding a core subgroup  $G_K$ . When there is evidence of multiple club convergence as  $T \rightarrow \infty$ , this is usually most apparent in the final time series observations. We therefore propose that the panel be clustered initially according to the value of the final time series observation (or some average of the final observations). After ordering in this way, size  $k$  subgroups,  $G_k = \{1, \dots, k\}$  for  $\{k = 2, \dots, N\}$ , may be constructed based on panel members with the  $k$  highest final time period observations. Within each of these subgroups, we may conduct  $\log t$  regression tests for convergence, denoting by  $t_k$  the test statistic from this regression using data from  $G_k$ . Next, we choose  $k^*$  to maximize  $t_k$  over all values for which  $t_k > c$  for  $k = 2, \dots, N$  and where  $c$  is some critical value. A precise algorithm based on these ideas is contained in the following step by step procedure to determine the clustering pattern and to provide a stopping rule for the calculations.

*Step 1: Last Observation Ordering.* Order individuals in the panel according to the last observation in the panel. In cases where there is substantial time series volatility in  $X_{it}$ , the ordering may be done according to the time series average,  $(T - [Ta])^{-1} \sum_{i=[Ta+1]}^T X_{it}$ , over the last fraction ( $f = 1 - a$ ) of the sample (for example,  $f = 1/3$  or  $1/2$ ).

*Step 2: Core Group Formation.* Selecting the first  $k$  highest individuals in the panel to form the subgroup  $G_k$  for some  $N > k \geq 2$ , run the  $\log t$  regression and calculate the convergence test statistic  $t_k = t(G_k)$  for this subgroup. Choose the core group size  $k^*$  by maximizing  $t_k$  over  $k$  according to the criterion:

$$(44) \quad k^* = \arg \max_k \{t_k\} \quad \text{subject to} \quad \min\{t_k\} > -1.65.$$

The condition  $\min\{t_k\} > -1.65$  plays a key role in ensuring that the null hypothesis of convergence is supported for each  $k$ . However, for each  $k$  there is the probability of a type II error. Choosing the core group size so that  $k^* = \arg \max_k \{t_k\}$  then reduces the overall type II error probability and helps ensure that the core group  $G_{k^*}$  is a convergence subgroup with a very low false

inclusion rate.<sup>8</sup> Our goal is to find a core convergence group in this test and then proceed in Step 3 to evaluate additional individuals for membership of this group. If there is a single convergence club with all individuals included, then the size of the convergence club is  $N$ ; when there are two or more convergence clubs, each club necessarily has membership less than  $N$ . If the condition  $\min\{t_k\} > -1.65$  does not hold for  $k = 2$ , then the highest individual in  $G_k$  can be dropped from each subgroup and new subgroups  $G_{2j} = \{2, \dots, j\}$  can be formed for  $2 \leq j \leq N$ . The step can be repeated with test statistics  $t_j = t(G_{2j})$ . If the condition  $\min\{t_j\} > -1.65$  is not satisfied for the first  $j = 2$ , the step may be repeated again, dropping the highest individuals in  $G_j$  and proceeding as before. If the condition does not hold for all such sequential pairs, then we conclude that there are no convergence subgroups in the panel. Otherwise, we have found a core convergence subgroup, which we denote  $G_{k^*}$ .

*Step 3: Sieve Individuals for Club Membership.* Let  $G_{k^*}^c$  be the complementary set to the core group  $G_{k^*}$ . Adding one individual in  $G_{k^*}^c$  at a time to the  $k^*$  core members of  $G_{k^*}$ , run the  $\log t$  test. Denote the  $t$  statistic from this regression as  $\hat{t}$ . Include the individual in the convergence club if  $\hat{t} > c$ , where  $c$  is some chosen critical value. We will discuss the choice of the critical value below and in the Monte Carlo section. Repeat this procedure for the remaining individuals and form the first subconvergence group. Run the  $\log t$  test with this first subconvergence group and make sure  $t_b > -1.65$  for the whole group. If not, raise the critical value,  $c$ , to increase the discriminatory power of the  $\log t$  test and repeat this step until  $t_b > -1.65$  with the first subconvergence group.

*Step 4: Stopping Rule.* Form a subgroup of the individuals for which  $\hat{t} < c$  in Step 3. Run the  $\log t$  test for this subgroup to see if  $t_b > -1.65$  and this cluster converges. If so, we conclude that there are two convergent subgroups in the panel. If not, repeat Steps 1–3 on this subgroup to determine whether there is a smaller subgroup of convergent members of the panel. If there is no  $k$  in Step 2 for which  $t_k > -1.65$ , we conclude that the remaining individuals diverge.

The application in Section 6 provides practical details and an illustration of the implementation of this algorithm. Table IV, in particular, lays out the sequence of steps involved in a specific application where there are multiple clusters.

<sup>8</sup>We might consider controlling the critical value based on the distribution of the  $\max_k\{t_k\}$  statistic over the cross section. However, since this distribution changes according to the true size and composition of the actual convergence subgroup (which is unknown), this approach is not feasible. Instead, the  $\max_k t_k$  rule is designed to be conservative in its selection of the core subgroup so that the false inclusion rate is small. Note that the rule (44) is used to determine only the membership of this core group. Subsequently, we apply individual  $\log t$  regression tests to assess membership of additional individuals. The performance of this procedure is found to be very satisfactory in simulations that are reported in Section 5.

## 5. MONTE CARLO EXPERIMENTS

The simulation design is based on the data generating process (DGP)

$$(45) \quad X_{it} = \delta_{it}\mu_t, \quad \delta_{it} = \delta_i + \delta_{it}^0, \\ \delta_{it}^0 = \rho_i \delta_{it-1}^0 + \epsilon_{it}, \quad \text{Var}(\epsilon_{it}) = \sigma_i^2 L(t+1)^{-2} t^{-2\alpha}$$

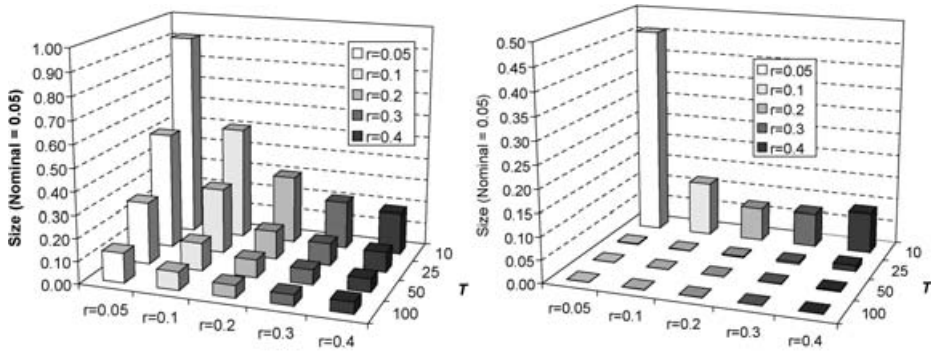
for  $t = 1, \dots, T$  and  $L(t+1) = \log(t+1)$ , so the slowly varying function  $L(t+1)^{-1}$  is well defined for all  $t \geq 1$ . We set  $\epsilon_{it} \sim \text{iid } N(0, \sigma_i^2 L(t+1)^{-2} t^{-2\alpha})$  and  $\rho_i \sim U[0, \rho]$  for  $\rho = 0.5, 0.9$ . To ensure that  $\delta_{it} \geq 0$  for all  $i$  and  $t$ , we control the range of  $\sigma_i$  by setting  $\sigma_i \sim U[0.02, 0.28]$  so that the 97.5% lower confidence limit for  $\delta_{it}$  at  $t = 1$  is greater than zero and then discard any trajectories that involve negative realizations. The simulation treats  $\delta_i$  and  $\rho_i$  as random variables drawn from the cross section population, so that for each iteration new values are generated. Since the common component  $\mu_t$  cancels out in the application of our procedure, there is no need to specify a parametric form for  $\mu_t$ .

The log  $t$  regression procedure relies on input choices of the initiating sample fraction  $r$  and the slowly varying function  $L(t)$ . In simulations, we explored a variety of possible choices; the full Monte Carlo results that consider these are available online.<sup>9</sup> Here we briefly report the results of varying these inputs and make some recommendations for applied work.

First we review the effects of varying the initiating sample fraction  $r$ . Various  $N$  and  $T$  combinations are considered under the DGP (45) and average rejection rates over these combinations were computed under the null hypothesis to assess the impact on test size. When  $\alpha = 0$ , test size is close to nominal size for sample sizes  $N, T \geq 50$  and  $r \geq 0.3$ . When the decay rate  $\alpha$  is small and nonzero, for example when  $\alpha \in (0, 0.4)$ , the rejection rate decreases as  $r$  increases for given  $N$  and  $T$ . The rejection rate also decreases rapidly as  $\alpha$  increases. Of course, when  $\alpha > 0$ , asymptotic theory shows that test size converges to zero since the  $t$  statistic diverges to positive infinity as  $N, T \rightarrow \infty$  in this case. Further, when  $r$  increases, test power declines because the effective sample size is smaller, which reduces discriminatory power. Thus, since  $\alpha$  is unknown, practical considerations suggest choosing a value of  $r$  for which size will be accurate when  $\alpha$  is close to zero, for which size is not too conservative when  $\alpha$  is larger, and for which power is not substantially reduced by the effective sample size reduction. The simulation results indicate that  $r \in [0.2, 0.3]$  achieves a satisfactory balance. When  $T$  is small or moderate ( $T \leq 50$ , say),  $r = 0.3$  seems to be a preferable choice to secure size accuracy in the test for small  $\alpha$ , and when  $T$  is large ( $T \geq 100$ , say), the choice  $r = 0.2$  seem satisfactory in terms of size and this choice helps to raise test power. Panels A and B of Figure 3 provide a visualization of the effects of different choices of  $r$  on actual size when  $\alpha = 0$  and  $\alpha = 0.1$  for various  $T$ .

<sup>9</sup>[http://homes.eco.auckland.ac.nz/dsul013/working/power\\_size.xls](http://homes.eco.auckland.ac.nz/dsul013/working/power_size.xls).





Panel A:  $\alpha = 0.0, N = 100,$

Panel B:  $\alpha = 0.1, N = 100$

FIGURE 3.—Effects of  $r$  on rejection rates under  $\mathcal{H}_0, \rho_i \in (0, 0.9),$  for nominal size 5%.

Next, we review the effects of varying the choice of  $L(t)$  in the formulation of the  $\log t$  regression equation. To do so, we standardize (45) so that  $\text{Var}(\epsilon_{it}) = \sigma_i^2 t^{-2\alpha}$ . Hence, the true model for  $\delta_{it}$  (and the DGP used in the simulation) does not involve a slowly varying function, whereas the fitted  $\log t$  regression involves some slowly varying function  $L(t)$ . The inclusion of  $L(t)$  in the regression then plays the role of a penalty function and we consider the effects of the presence of this variable in the form of the four functions  $L(t) = \log t, 2\log t, \log \log t,$  and  $2\log \log t,$  calculating the corresponding rejection rates for the tests in each case. Note that when  $\alpha = 0$  in the true DGP, convergence no longer holds and the test diverges to negative infinity, as in the asymptotics of Theorem 2. For  $\alpha > 0,$  test size should converge to zero as  $N, T \rightarrow \infty,$  just as in the correctly specified case. However, simulations show that for very small  $\alpha$  ( $\alpha = 0.01,$  say) there are substantial upward size distortions when  $L(t) = 2\log t,$  whereas the test is conservative for  $L(t) = \log t, 2\log \log t,$  and  $\log \log t.$  Test power is reduced when  $L(t) = 2\log \log t$  and  $\log \log t$  in comparison with  $L(t) = \log t.$  Hence, among these possibilities, the function  $L(t) = \log t$  produces the least size distortion and the best test power as  $N$  and  $T$  increase. A full set of simulation results is available online.<sup>10</sup>

Based on these experiments, we recommend setting  $r = 0.3$  and suggest  $L(t) = \log t$  for the slowly varying function in the  $\log t$  regressions. These settings are used in the remaining experiments.

For the remaining simulations we set  $T = 10, 20, 30, 40$  and  $N = 50, 100, 200.$  Since the size of the test is accurate to two decimal place when  $\alpha = 0$  and power is close to unity for moderately large  $T$  ( $T \geq 50),$  these results are not reported here. The number of replications was  $R = 2000.$  We consider the following four cases.

<sup>10</sup><http://homes.eco.auckland.ac.nz/dsul013/working/logtcomp.xls>.

*Case 1: Pure Convergence.* To check the size of the test, we set  $\delta_i = 1$  for all  $i$  and  $\alpha = 0.01, 0.05, 0.1, 0.2$ . When  $\alpha > 0.2$ , the test size is zero for all  $T$  and  $N$ , confirming that the limit theory is accurate for small  $T$  and  $N$  in this case. To measure the bias in the estimate of the speed of convergence,  $\hat{b}$ , we used  $\alpha = 0.05, 0.1$ , and  $0.5$ .

*Case 2: Divergence.* We set  $\delta_i \sim U[1, 2]$ .

*Case 3: Club Convergence.* We considered two equal sized convergence clubs in the panel with numbers  $S_1 = S_2 = 50$  and overall panel size  $N = 100$ . For the first panel, we set  $\delta_1 = 1$ , and for the second panel, we set  $\delta_2 = [1.1, 1.2, 1.5]$  to allow for different distances between the convergence clubs.

*Case 4: Sorting Procedure.* Two convergence clubs as in Case 3, but with  $\delta_1 = 1$  and  $\delta_2 = 1.2$ . We consider various convergence rates with  $\alpha = (0.01, 0.05, 0.1, 0.2)$  and with  $\rho = 0.5$ .

Table I gives the actual test size. The nominal size is fixed to be 5%. When the speed of convergence parameter  $\alpha$  is very close to zero, there is size distortion for small  $T$ . However, this distortion diminishes quickly when  $T$  increases or as  $\alpha$  increases. As  $\alpha$  increases, the test becomes conservative. Rejection rates in the test are expected to go to zero when  $\alpha > 0$  as  $T, N \rightarrow \infty$  since the limit theory in Theorem 1 shows that the regression  $t$  statistic (centered on the origin) diverges to positive infinity in this case.

TABLE I  
SIZE OF THE  $\log t$  TEST (5% NOMINAL SIZE)

$T$	$N$	$\rho \in [0, 0.5]$				$\rho \in [0, 0.9]$			
		$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$
10	50	0.30	0.21	0.13	0.04	0.25	0.18	0.10	0.03
10	100	0.40	0.26	0.12	0.02	0.32	0.22	0.10	0.01
10	200	0.56	0.32	0.12	0.01	0.41	0.24	0.08	0.00
20	50	0.15	0.09	0.03	0.00	0.13	0.07	0.03	0.00
20	100	0.18	0.08	0.01	0.00	0.14	0.05	0.01	0.00
20	200	0.23	0.07	0.01	0.00	0.17	0.04	0.00	0.00
30	50	0.09	0.05	0.01	0.00	0.10	0.04	0.01	0.00
30	100	0.14	0.03	0.00	0.00	0.10	0.02	0.00	0.00
30	200	0.14	0.02	0.00	0.00	0.10	0.01	0.00	0.00
40	50	0.09	0.03	0.01	0.00	0.07	0.03	0.00	0.00
40	100	0.09	0.02	0.00	0.00	0.08	0.02	0.00	0.00
40	200	0.11	0.01	0.00	0.00	0.08	0.01	0.00	0.00

TABLE II  
MEAN VALUES OF THE ESTIMATED SPEED OF CONVERGENCE

<i>T</i>	<i>N</i>	$\rho \in [0, 0.5]$			$\rho \in [0, 0.9]$		
		<i>b</i> = 0.1	<i>b</i> = 0.2	<i>b</i> = 1.0	<i>b</i> = 0.1	<i>b</i> = 0.2	<i>b</i> = 1.0
10	50	-0.09	0.00	0.84	-0.06	0.05	0.96
10	100	-0.11	0.00	0.84	-0.06	0.05	0.96
10	200	-0.10	0.00	0.84	-0.06	0.05	0.95
20	50	0.01	0.12	0.93	0.06	0.15	1.06
20	100	0.02	0.12	0.93	0.06	0.17	1.06
20	200	0.02	0.12	0.93	0.05	0.16	1.06
30	50	0.06	0.16	0.96	0.07	0.18	1.04
30	100	0.05	0.15	0.95	0.07	0.18	1.04
30	200	0.05	0.15	0.96	0.07	0.18	1.04
40	50	0.07	0.17	0.97	0.08	0.18	1.02
40	100	0.07	0.16	0.97	0.08	0.18	1.02
40	200	0.07	0.17	0.97	0.08	0.18	1.02

Table II shows the mean values of  $\hat{b}$ . When  $\alpha$  is small, there is a somewhat mild downward bias for small  $T$ , which arises from the correlation between the  $\log t$  regressor and the second order terms in  $u_t$ . The direction of the correlation is negative since in the expansion, the second order term of  $u_t$ ,  $L(t)^{-2}t^{-2\alpha}\psi_t^2/\delta^2$ , is negatively correlated with  $L(t)$ . The bias is dependent on the size of  $T$  and  $\alpha$  rather than  $N$ , just as the asymptotic theory predicts, when  $\alpha$  is small. This downward bias quickly disappears for larger  $T$  or as  $\alpha$  increases.

Table III shows the power of the test without size adjustment. For Case 2, the power becomes 1, irrespective of the values of  $\alpha$ ,  $T$ , and  $N$ . For Case 3, the  $\log t$  test distinguishes well whether there is club convergence or not, even with small  $T$  and  $\alpha$ , except when  $\delta_1$  is very close to  $\delta_2$ . For  $\delta_1 - \delta_2 = 0.1$ , the rejection rate is more than 50% with  $\alpha = 0.01$  for  $T = 10$ , and increases rapidly as  $T$  or  $N$  grows.

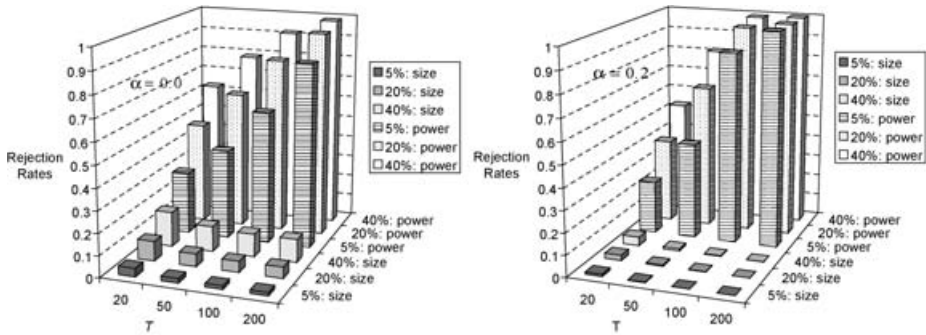
Figure 4 shows how the empirical clustering procedure suggested in the previous section works. Overall the results are encouraging. Panels A and B in Figure 4 display the size and power of the clustering test across various critical values with  $\alpha = 0$  and  $\alpha = 0.2$ , respectively. When  $\alpha > 0$ , the size of the clustering test—measuring the failure rate of including convergence members in the correct subconvergence club—goes to zero asymptotically since  $t_{\hat{b}}$  tends to positive infinity under the null of convergence as  $T \rightarrow \infty$ .

As asymptotic theory predicts, the size of the clustering test goes to zero as  $T$  increases in this case. Meanwhile, the power of the clustering test—the success rate in excluding nonconvergence members from the correct subconvergence club—goes to unity asymptotically regardless of the critical values used. However, in finite samples, test power is less than unity and, as larger critical values

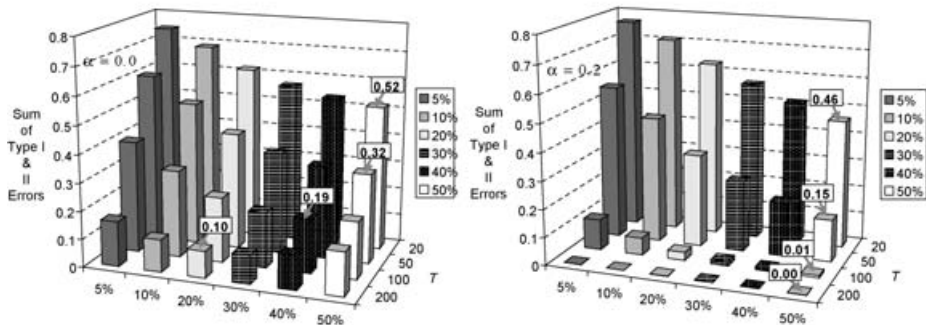
TABLE III  
THE POWER OF THE  $\log t$  TEST (5% TEST)

$T$	$N$	$\alpha$	$\delta_i - U[1, 2]$	$\delta_1 = 1$		
				$\delta_2 = 1.5$	$\delta_2 = 1.2$	$\delta_2 = 1.1$
10	50	0.01	1.00	1.00	0.93	0.57
10	100	0.01	1.00	1.00	0.99	0.77
10	200	0.01	1.00	1.00	1.00	0.92
20	50	0.01	1.00	1.00	0.99	0.61
20	100	0.01	1.00	1.00	1.00	0.82
20	200	0.01	1.00	1.00	1.00	0.97
30	50	0.01	1.00	1.00	1.00	0.69
30	100	0.01	1.00	1.00	1.00	0.89
30	200	0.01	1.00	1.00	1.00	0.99
40	50	0.01	1.00	1.00	1.00	0.78
40	100	0.01	1.00	1.00	1.00	0.95
40	200	0.01	1.00	1.00	1.00	1.00
10	50	0.05	1.00	1.00	0.93	0.59
10	100	0.05	1.00	1.00	1.00	0.77
10	200	0.05	1.00	1.00	1.00	0.91
20	50	0.05	1.00	1.00	0.99	0.64
20	100	0.05	1.00	1.00	1.00	0.81
20	200	0.05	1.00	1.00	1.00	0.96
30	50	0.05	1.00	1.00	1.00	0.73
30	100	0.05	1.00	1.00	1.00	0.91
30	200	0.05	1.00	1.00	1.00	0.99
40	50	0.05	1.00	1.00	1.00	0.84
40	100	0.05	1.00	1.00	1.00	0.96
40	200	0.05	1.00	1.00	1.00	1.00

are employed in the selection procedure, we do find higher power in the test. Panels C and D show the sum of the type I and II errors in this procedure against various significance levels when  $\alpha = 0$  and  $\alpha = 0.2$ , respectively. As  $T$  increases, the size and the type II error of the clustering test both go to zero. There is some trade-off between the type I and II errors, and in finite samples, the power gain by using higher significance level seems to exceed the size loss. Hence, for both cases  $\alpha = 0$  and  $\alpha = 0.2$ , the use of a sign test (that is, a test in which the critical value is zero at the 50% significance level) minimizes the sum of the type I and II errors for small  $T$  (that is,  $T = 20, 50$  in panel C). For larger values of  $T$  ( $T = 100, 200$  in panel C), a lower nominal significance level minimizes the sum of the two errors when  $\alpha = 0$  (in panel C, these nominal significance levels are 40% for  $T = 100$  and 20% for  $T = 200$ , and these cases are marked in the chart). When  $\alpha = 0.2$ , the sign test minimizes the sum of the type I and type II errors for all values of  $T$ , as is clear in panel D.



(Panels A & B: Size and Power are given for nominal 5%, 20%, 40% size settings)



(Panels C & D: Failure rate = Type I + Type II error rates)

FIGURE 4.—Effect of nominal critical value choice on test performance.

Figure 5 shows the finite sample performance of the core group selection procedure based on the  $\max_k t_k$  rule. Panel A in Figure 5 shows the false inclusion rates of nonconvergence members into a core group. As  $T$  increases or as  $\alpha$  increases, the  $\max t$  rule appears to sieve individuals very accurately. Even when  $\alpha = 0$ , more than 99% of the time the  $\max t$  rule does not include any nonconvergence member into a core group when  $T \geq 100$ . Panel B in Figure 5 shows the size of the core groups selected for various values of  $\alpha$  and  $T$ . As  $\alpha$  and  $T$  increase, the size of the core group increases steadily and approaches the true size of the convergence club (51) for some configurations.

6. EMPIRICAL APPLICATION TO CALCULATING THE COST OF LIVING

We provide an empirical application to illustrate the usefulness of the time varying nonlinear factor model and the operation of the  $\log t$  regression test for convergence and clustering. The example shows how to calculate a proxy

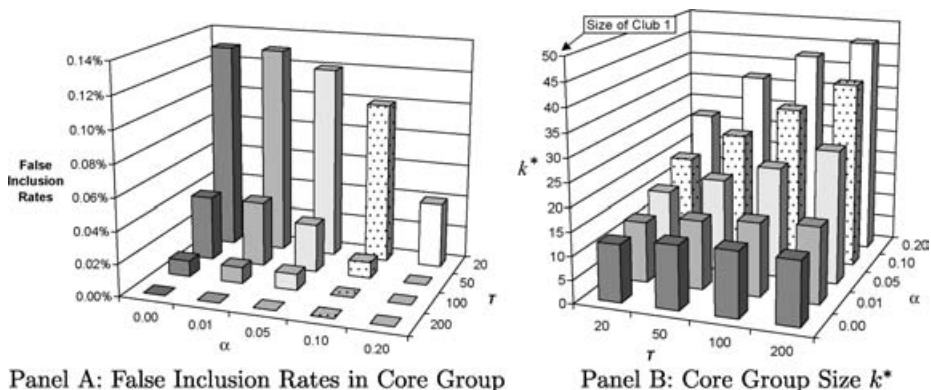


FIGURE 5.—Performance of the max  $t$  rule in core group selection.

for cost of living indices by using 19 consumer price indices (CPI's) for U.S. metropolitan areas. Measuring the cost of living by statistical indices has been a long-standing problem of econometrics that has many different contributions and much controversy.<sup>11</sup> A number of commercial web sites now provide various online cost of living indices. From a strict economic perspective, the most appropriate calculations for cost of living indices take account of the changing basket of commodities and services over time as well as nonconsumer price information such as local taxation, health and welfare systems, and economic infrastructure; while relevant, such matters are beyond the scope of many studies, including the present analysis. Here we constrain ourselves to working with cost of living indices obtained directly from commonly available consumer price information for 19 different metropolitan areas.

Our goal is to measure the relative cost of living across various metropolitan areas in the United States and to illustrate our empirical approach by examining evidence for convergence in the cost of living. We use the relative transition parameter mechanism to model individual variation, writing individual city CPI as

$$(46) \quad \log P_{it}^o = \delta_{it}^o \log P_t^o + e_{it},$$

where  $\log P_{it}^o$  is the log CPI for the  $i$ th city,  $\log P_t^o$  is the common CPI trend across cities, and  $e_{it}$  contains idiosyncratic business cycle components. The empirical application is to 19 major metropolitan U.S. cities from 1918 to 2001. Appendix D gives a detailed description of the data set.

It is well known that consumer price indices cannot be used to compare the cost of living across U.S. cities because of a base year problem. For example, if

<sup>11</sup>See the *Journal of Economic Perspectives*, 12, issue 1, for a recent special issue dealing with cost of living indices.

the base year were taken to be the last time period of observation, then the CPI indices would seem to converge because the last observations are identical. To avoid such artificial forms of convergence, we take the first observation as the base year and rewrite the data as  $\log P_{it} = \log(P_{it}^o/P_{i1}^o) = \log P_{it}^o - \log P_{i1}^o$ , from which we obtain  $\log P_{it} = [\delta_{it}^o - \delta_{i1}^o (\log P_{i1}^o / \log P_{i1}^o) + (e_{it} - e_{i1}) / \log P_{i1}^o] \log P_{i1}^o = \delta_{it} \log P_{i1}^o$ . The common price index  $P_{it}^o$  usually has a trend component, so that we have  $\log P_{it}^o = O_p(t^\alpha)$  for some  $\alpha > 0$ . For instance, if  $\log P_{it}^o$  follows a random walk with drift, we have  $\log P_{it}^o = a + \log P_{i,t-1}^o + \epsilon_t = at + \sum_{s=1}^t \epsilon_s = O_p(t)$ . Then  $\log P_{i1}^o / \log P_{i1}^o = o_p(1)$  and  $(e_{it} - e_{i1}) / \log P_{i1}^o = o_p(1)$  for large  $t$ , so the impact of the initial condition on  $\delta_{it}$  disappears as  $t \rightarrow \infty$ , and more rapidly the stronger the trend (or larger  $\alpha$ ). Cyclical effects are also of smaller magnitude asymptotically. Of course, these effects may be smoothed out using other techniques such as various filtering devices.

Figure 6 shows the cross sectional maximum, minimum, and median of the period-by-period log consumer price indices across the 19 U.S. cities. Due to the base year initialization, the CPI's in 1918 are identical, but the initial effects seem to have dissipated in terms of the observed dispersion in Figure 6 within two decades. To avoid the base year effect in our own calculations, we discard the first 42 annual observations. The relative transition parameters for 8 major metropolitan cities over the subsequent period 1960–2000 are plotted in Figure 7 after smoothing the CPI's using the WHP filter. The transition parameter curves provide relative cost of living indices across these metropolitan areas.

As is apparent in Figure 7, San Francisco shows the highest cost of living; Seattle is in second place. Chicago has the median cost of living among the 19 cities at the end of the sample and Atlanta has the lowest cost of living,

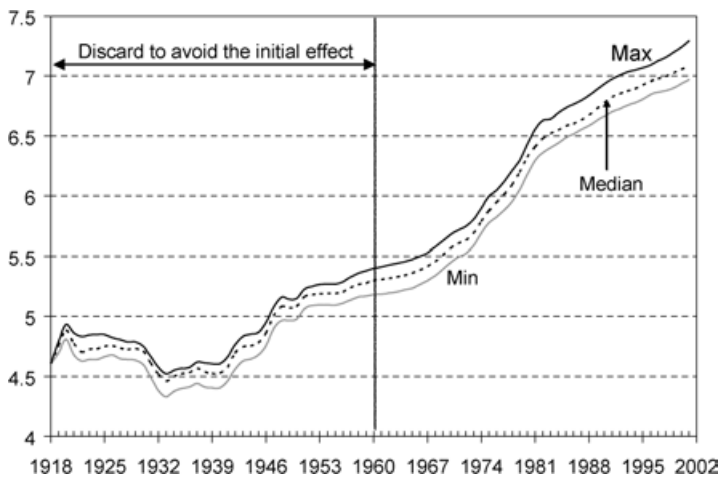


FIGURE 6.—Min, max, and median of consumer price indices.

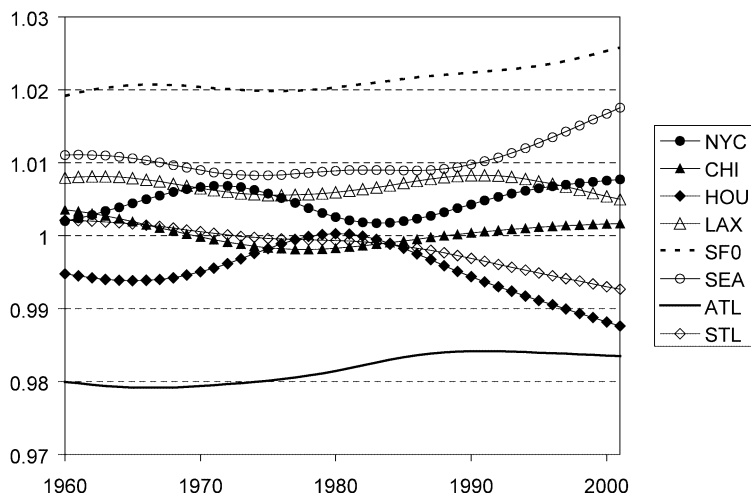


FIGURE 7.—Relative transition curves (relative cost of living).

again with little transition. Also apparent is that the cost of living indices in Houston and St. Louis have declined relatively since 1984, while those in New York, Seattle, and San Francisco have increased. The estimated equation for the overall log  $t$  regression with  $r = 1/3$  is

$$\log \frac{H_1}{H_t} - 2 \log \log t = 0.904 - 0.98 \log t, \quad (14.3) \quad (-51.4)$$

which implies that the null hypothesis of convergence in the relative cost of living is clearly rejected at the 5% level.

Next, we investigate the possibility of club convergence in cost of living indices among cities. Following the steps suggested in the previous section, we order the CPI's based on the last time series observation (Step 1) and display them in the first column of Table IV. Note that for further convenience (based on the convergence results we obtain below), we changed the order between New York (NYC) and Cleveland (CLE) metro. Based on this ordering, we choose San Francisco as the base city in the ordering, run the log  $t$  regression by adding further cities one by one, and calculate the  $t$  statistics until the  $t$  statistic is less than  $-1.65$  (Step 2). Proceeding in this way, we find that  $t_k = 6.1, -0.7, 1.4,$  and  $-7.8$  for  $k = \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\},$  and  $\{1, 2, 3, 4, 5\}$ , respectively. When we add Minnesota, the  $t_k$  statistic becomes  $t_k = -7.8$  and we stop adding cities. The  $t_k$  statistics are maximized for the group  $k = \{1, 2\}$  and so the core group is taken to be San Francisco and Seattle. Next, working from this core group, we add one city at a time and print out its  $t$  statistic in the third column of Table IV. We use the 50% critical value (or sign test), based on our findings in the Monte Carlo experiments. Only when



TABLE IV  
 CLUB CONVERGENCE OF COST OF LIVING INDICES AMONG 19 U.S. METROPOLITAN CITIES

Last <i>T</i> Order	Name	<i>t</i> value		Club	log <i>t</i> Test	Name	Step 1	Step 2	Club	log <i>t</i> Test	
		Step 1	Step 2								
1	SFO	Base	<b>Core</b>	S <sub>1</sub>	t <sub>S<sub>1</sub></sub> = 0.71					t <sub>S<sub>2</sub></sub> = 8.18 t <sub>S<sub>2</sub><sup>c</sup></sub> = -0.68	
2	SEA	<b>6.1</b>	<b>Core</b>								
4	NYC	1.4	0.7								
3	CLE	-0.7	-0.7	S <sub>1</sub> <sup>c</sup>	t <sub>S<sub>1</sub><sup>c</sup></sub> = -54.6	CLE	Base	<b>Core</b>	S <sub>2</sub>		
5	MIN	-7.8	-51.0			MIN	<b>1.0</b>	<b>Core</b>	S <sub>2</sub>		
6	LAX		-12.2			LAX	-1.7	-1.7	S <sub>2</sub> <sup>c</sup>		
7	POR		-2.4			POR		5.3	S <sub>2</sub>		
8	BOS		-3.7			BOS		13.9	S <sub>2</sub>		
9	CHI		-14.9			CHI		6.1	S <sub>2</sub>		
10	BAL		-28.8			BAL		-19.9	S <sub>2</sub> <sup>c</sup>		
11	PHI		-12.0			PHI		7.6	S <sub>2</sub>		
12	PIT		-35.6			PIT		-1.6	S <sub>2</sub> <sup>c</sup>		
13	CIN		-46.9			CIN		-18.1	S <sub>2</sub> <sup>c</sup>		
14	STL		-50.3			STL		-34.6	S <sub>2</sub> <sup>c</sup>		
15	DET		-124.4			DET		-4.9	S <sub>2</sub> <sup>c</sup>		
16	WDC		-16.7			WDC		-12.3	S <sub>2</sub> <sup>c</sup>		
17	HOU		-134.6			HOU		-28.0	S <sub>2</sub> <sup>c</sup>		
18	KCM		-116.5			KCM		-14.1	S <sub>2</sub> <sup>c</sup>		
19	ATL		-20.7			ATL		-67.2	S <sub>2</sub> <sup>c</sup>		

New York (NYC) is added to the core group, is the *t* statistic still positive. The log *t* regression with these three cities gives a *t* statistic of 0.71, and the null hypothesis of convergence cannot be rejected. Hence the first convergence club, S<sub>1</sub>, includes SFO, SEA, and NYC.

For the remaining 16 cities (S<sub>1</sub><sup>c</sup>), the log *t* test rejects the null of convergence even at the 1% level (t<sub>S<sub>1</sub><sup>c</sup></sub> = -54.6). Hence, we further analyze the data for evidence of club convergence among these 16 cities. Repeating the same procedure again, we find the next core group as Cleveland (CLE) and Minneapolis/St. Paul (MIN), and select 4 other cities (POR, BOS, CHI, and PHI) for the second subgroup, S<sub>2</sub>. The log *t* test with these 6 cities does not reject the null of convergence (t<sub>S<sub>2</sub></sub> = 8.2). Further, the log *t* test with the remaining 10 cities does not reject the null either (t<sub>S<sub>2</sub><sup>c</sup></sub> = -0.68) at the 5% level. Hence with the last group, there is rather weak evidence for convergence.

Figure 8 shows the relative transition parameters with the cross sectional means of the three convergence clubs. The transition curves indicate that the three clubs show some mild evidence of convergence until around 1982, but that after this there is strong evidence of divergence. In sum, the evidence is that the relative cost of living across 19 major U.S. metropolitan areas does not appear to be converging over time. However, there is some evidence of recent convergence clustering among three different metropolitan subgroups:

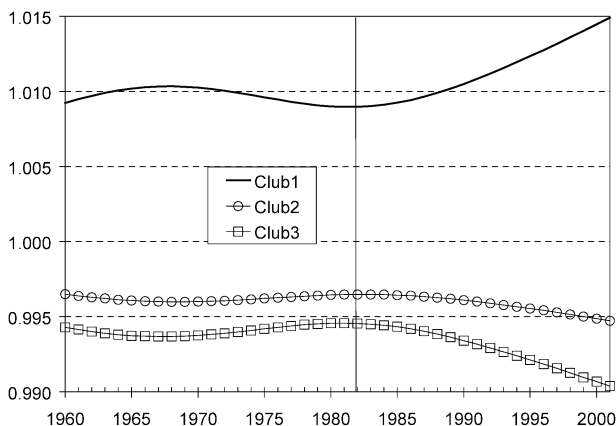


FIGURE 8.—Relative transition curves across clubs.

one with a very high cost of living, one with a moderate cost of living, and one that is relatively less expensive than the other two groups.

## 7. CONCLUSION

This paper has proposed a new mechanism for modeling and analyzing economic transition behavior in the presence of common growth characteristics. The model is a nonlinear factor model with a growth component and a time varying idiosyncratic component that allows for quite general heterogeneity across individuals and over time. The formulation is particularly useful in measuring transition toward a long run growth path or individual transitions over time relative to some common trend, representative, or aggregate variable. The formulation also gives rise to a simple and convenient time series regression test for convergence. This  $\log t$  convergence test further provides the basis for a stepwise clustering algorithm that is proposed for finding convergence clusters in panel data and analyzing transition behavior between clusters. The tests have good asymptotic properties, including local discriminatory power, and are particularly easy to apply in practice. Simulations show that the proposed  $\log t$  test and the clustering algorithm both work very well for values of  $T$  and  $N$  that are common in applied work. The empirical application reveals some of the potential of these new procedures for practical work.

Some extensions of the procedures seem worthwhile to pursue in later work. In particular, the procedures are developed here for panels of a scalar variable and will need to be extended when there are many variables. For example, to analyze issues of convergence and clustering in the context of potential relationships between two panel variables such as personal expenditure and income, the concepts and methods in the paper must be modified, possibly by

working with panel regression residuals or through panel vector autoregression and error correction formulations.

In addition, our procedures have been developed primarily for cases where there is a single common growth factor. The approach has the advantage that it is applicable regardless of the form of the generating mechanism for this common factor. In practical work, of course, there may be several relevant factors. General factor-analytic techniques are designed to address situations where there may be multiple factors and where the number of factors is unknown. These techniques have received attention in past econometric work (e.g., Sargent and Sims (1997)) and in recent large multidimensional panel modeling (e.g., Bai and Ng (2002)). In such work, investigators may use data based methods to search for the number of commonalities. In the present approach, there is a presumption that one growth factor dominates the commonalities for large  $t$  under the null hypothesis of convergence. But, as seen in some of the examples of Section 2 and the empirical application to city cost of living, sometimes the alternative hypothesis will be more relevant in practice. The methods developed here continue to apply in such situations, allowing for different subgroup behavior in which multiple factors may indeed be present, manifesting themselves in the form of the individual or subgroup transition effects.

*Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520-8281, U.S.A.; University of Auckland, New Zealand; and University of York, York, U.K.; peter.phillips@yale.edu*

*and*

*Dept. of Economics, University of Auckland, Auckland, New Zealand.*

*Manuscript received October, 2006; final revision received April, 2007.*

## APPENDIX A: STANDARDIZING GROWTH COMPONENTS

This appendix analyses how the growth components in the decomposition

$$X_{it} = a_{it} + g_{it} = \left( \frac{a_{it} + g_{it}}{\mu_t} \right) \mu_t = \delta_{it} \mu_t$$

may be standardized to yield the transition and growth curves discussed in Section 3. We let  $t \rightarrow \infty$  and characterize the limiting behavior of the components  $\delta_{it}$  and  $\mu_t$ .

We first proceed as if the growth components were nonstochastic. Suppose  $g_{it} = f_i(t)$  is regularly varying at infinity with power exponent  $\gamma_i$  (e.g., see Seneta (1976) for a discussion of regularly varying functions) so that

$$(47) \quad f_i(t) = t^{\gamma_i} W_i(t),$$

where  $W_i(t)$  is slowly varying at infinity, namely  $W_i(\lambda t)/W_i(t) \rightarrow 1$  as  $t \rightarrow \infty$  for all  $\lambda > 0$ . For example, we might have  $W_i(t) = \log t, \log^2 t$ , or  $\log \log t$ . Similarly, let  $\mu_t$  be regularly varying at infinity with power exponent  $\gamma > 0$  so that

$$(48) \quad \mu_t = t^\gamma Z(t)$$

for some slowly varying function  $Z(t)$ . The regular variation requirement means that  $f_i(t)$  and  $\mu_t$  both behave asymptotically very much like power functions for large  $t$ . In the simplest case where the common growth component is a linear drift (i.e.,  $\mu_t = t$ ) and  $g_{it}/t \rightarrow m$  for all  $i$  as  $t \rightarrow \infty$ , there is growth convergence and we have  $\gamma_i = \gamma = 1$  and  $W_i(t) = Z(t) = 1$ . Conditions (47) and (48) allow for a much wider variety of asymptotic behavior, including the possibility that individual  $i$  economy's growth may deviate from the common path (when  $\gamma_i \neq \gamma$ ) and that there may be a slowly varying component in the growth path. For example, if  $\gamma = 0$  and  $Z(t) = \log t$ , then  $\mu_t$  evolves logarithmically with  $t$  and growth is therefore slower than any polynomial rate.

Set  $t = [Tr]$  for some  $r > 0$  representing the fraction of the overall sample  $T$  corresponding to observation  $t$ . Then under (47),

$$(49) \quad T^{-\gamma_i} g_{it} = T^{-\gamma_i} [Tr]^{\gamma_i} \frac{W_i(Tr)}{W_i(T)} W_i(T) \sim r^{\gamma_i} W_i(T)$$

and

$$T^{-\gamma} \mu_t = T^{-\gamma} [Tr]^\gamma \frac{Z(Tr)}{Z(T)} Z(T) \sim r^\gamma Z(T).$$

We deduce from this asymptotic behavior and (5) that

$$T^{-\gamma_i} X_{it} = \frac{a_{it} + g_{it}}{T^{\gamma_i}} = \frac{a_{it}}{T^{\gamma_i}} + \frac{g_{it}}{T^{\gamma_i}} \sim r^{\gamma_i} W_i(T),$$

$$T^{-\gamma} \mu_t \sim r^\gamma Z(T) = \mu(r) Z(T),$$

where  $\mu(r) = r^\gamma$ . Writing, as in (5),

$$\left( \frac{a_{it} + g_{it}}{\mu_t} \right) \mu_t = \delta_{it} \mu_t,$$

we then have

$$\begin{aligned} \frac{1}{T^{\gamma_i}} \left( \frac{a_{it} + g_{it}}{\mu_t} \right) \mu_t &= \frac{a_{it}}{T^{\gamma_i}} + \frac{g_{it}}{T^{\gamma_i}} \frac{T^\gamma}{\mu_t} \left( \frac{\mu_t}{T^\gamma} \right) \\ &= o(1) + \frac{g_{it}}{T^{\gamma_i}} \frac{T^\gamma}{\mu_t} \left( \frac{\mu_t}{T^\gamma} \right) \\ &\sim \{r^{\gamma_i - \gamma} J_i(T)\} \{r^\gamma Z(T)\} \\ &= \delta_{iT}^J(r) \mu_T^Z(r), \end{aligned}$$

where the ratio  $J_i(T) = W_i(T)/Z(T)$  is also slowly varying at infinity. Thus, the functions  $\delta_{iT}^J(r)$  and  $\mu_T^Z(r)$  are regularly varying and behave asymptotically like the power functions  $r^{\gamma_i-\gamma}$  and  $r^\gamma$ , at least up to slowly varying factors.

Next set  $d_{iT} = T^{\gamma_i} J_i(T) Z(T) = T^{\gamma_i} W_i(T)$ , so that the slowly varying components are factored into the standardization. Then, for  $t = [Tr]$ , we have

$$\begin{aligned} \frac{1}{d_{iT}} X_{it} &= \frac{1}{T^{\gamma_i} J_i(T) Z(T)} \left( \frac{a_{it} + g_{it}}{\mu_t} \right) \mu_t \\ &= \frac{a_{it}}{T^{\gamma_i} W_i(T)} + \frac{g_{it}}{T^{\gamma_i} W_i(T)} \left( \frac{T^\gamma Z(T)}{\mu_t} \right) \left( \frac{\mu_t}{T^\gamma Z(T)} \right) \\ &= o(1) + \frac{g_{it}}{T^{\gamma_i} W_i(T)} \left( \frac{T^\gamma Z(T)}{\mu_t} \right) \left( \frac{\mu_t}{T^\gamma Z(T)} \right) \\ (50) \quad &= o(1) + \delta_{iT} \left( \frac{t}{T} \right) \mu_T \left( \frac{t}{T} \right) \end{aligned}$$

$$(51) \quad \sim \delta_{iT}(r) \mu_T(r).$$

In (50), we define

$$(52) \quad \mu_T \left( \frac{t}{T} \right) = \frac{\mu_t}{T^\gamma Z(T)} = \frac{\mu_t}{t^\gamma Z(t)} \left( \frac{t^\gamma Z(t)}{T^\gamma Z(T)} \right) = \left( \frac{t}{T} \right)^\gamma \frac{Z(\frac{t}{T})}{Z(T)},$$

and in a similar manner,

$$(53) \quad \delta_{iT} \left( \frac{t}{T} \right) = \left( \frac{t}{T} \right)^{\gamma_i-\gamma} \frac{W_i(\frac{t}{T}) Z(T)}{W_i(T) Z(\frac{t}{T})}.$$

Then, for  $t = [Tr]$ , we have

$$(54) \quad \delta_{iT}(r) \rightarrow \delta_i(r) = r^{\gamma_i-\gamma}$$

and

$$(55) \quad \mu_T(r) \rightarrow \mu(r) = r^\gamma.$$

Relations (51)–(55) lead to a nonstochastic version of the stated result (18). For a stochastic version, we may continue to assume that the standardized representation (51) applies with an  $o_p(1)$  error uniformly in  $t \leq T$  and require that

$$\begin{aligned} \delta_{iT}(r) &\rightarrow_p \delta_i(r) = r^{\gamma_i-\gamma}, \\ \mu_T(r) &\rightarrow_p \mu(r) = r^\gamma \end{aligned}$$

uniformly in  $r \in [0, 1]$ , so that the limit transition function  $\delta_i(r)$  and growth curve  $\mu(r)$  are nonrandom functions.

More generally, the limit functions  $\delta_i(r)$  and  $\mu(r)$  may themselves be stochastic processes. For example, if the common growth component  $\mu_t$  in  $\log y_{it}$  is a unit root stochastic trend, then by standard functional limit theory (e.g., Phillips and Solo (1992)) on a suitably defined probability space

$$(56) \quad T^{-1/2} \mu_{[Tr]} = \mu_T(r) \rightarrow_p B(r)$$

for some Brownian motion  $B(r)$ . In place of (47), suppose that  $f_i(t) = g_{it}/\mu_t$  is stochastically regularly varying at infinity in the sense that  $f_i(t)$  continues to follow (47) for some power exponent  $\gamma_i$ , but with  $W_i(t)$  stochastically slowly varying at infinity, that is,  $W_i(\lambda t)/W_i(t) \rightarrow_p 1$  as  $t \rightarrow \infty$  for all  $\lambda > 0$ . Then, in place of (49) we have

$$\frac{1}{T^{\gamma_i}} \frac{g_{it}}{\mu_t} = \frac{[Tr]^{\gamma_i}}{T^{\gamma_i}} \frac{W_i(Tr)}{W_i(T)} W_i(T) \sim r^{\gamma_i} W_i(T).$$

Setting  $d_{iT} = T^{\gamma_i+1/2} W_i(T)$  and  $t = [Tr]$ , and working in the same probability space where (56) holds, we have

$$\begin{aligned} d_{iT}^{-1} X_{it} &= \frac{a_{it}}{T^{\gamma_i+1/2} W_i(T)} + \frac{1}{T^{\gamma_i} W_i(T)} \left( \frac{g_{it}}{\mu_t} \right) \left( \frac{\mu_t}{\sqrt{T}} \right) \\ &= o_p(1) + \delta_{iT}(r) \mu_T(r) \rightarrow_p \delta_i(r) B(r) \end{aligned}$$

with  $\delta_i(r) = r^{\gamma_i}$ . In this case the limiting common trend function is the stochastic process  $\mu(r) = B(r)$  and the transition function is the nonrandom function  $\delta_i(r) = r^{\gamma_i}$ .

## APPENDIX B: ASYMPTOTIC PROPERTIES OF THE $\log t$ CONVERGENCE TEST

### B.1. Derivation of the $\log t$ Regression Equation

We proceed with the factor model (2) and the semiparametric representation (24), written here as

$$(57) \quad \delta_{it} = \delta_i + \sigma_{it} \xi_{it} = \delta_i + \frac{\sigma_i \xi_{it}}{L(t)t^\alpha} := \delta_i + \frac{\psi_{it}}{L(t)t^\alpha}$$

for some  $\sigma_i > 0$ ,  $t \geq 1$  and where the various components satisfy Assumptions A1–A4. From (27) we have

$$(58) \quad \psi_{Nt} := \sqrt{N} \psi_t = \frac{1}{\sqrt{N}} \sum_{i=1}^N \psi_{it} \Rightarrow N(0, v_\psi^2) = \xi_{\psi t}, \quad \text{say,}$$

where  $v_\psi^2 = p \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \psi_{it}^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2$ . So  $\psi_t = O_p(N^{-1/2})$  and

$$\begin{aligned}
 (59) \quad \psi_t^2 &= N^{-1} \psi_{Nt}^2 = N^{-2} \sum_{i=1}^N \psi_{it}^2 + N^{-2} \sum_{i \neq j} \psi_{it} \psi_{jt} \\
 &= N^{-2} \sum_{i=1}^N \sigma_i^2 + N^{-2} \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1) + 2N^{-2} \sum_{i=2}^N \sum_{j=1}^{i-1} \psi_{it} \psi_{jt} \\
 &= O_p(N^{-1}).
 \end{aligned}$$

From (2) and the definition of  $h_{it}$ , we have

$$(60) \quad h_{it} - 1 = \frac{\delta_{it} - \frac{1}{N} \sum_i \delta_{it}}{\frac{1}{N} \sum_i \delta_{it}} = \frac{\delta_i - \bar{\delta} + (\psi_{it} - \psi_t)/(L(t)t^\alpha)}{\bar{\delta} + \psi_t/(L(t)t^\alpha)},$$

where  $\bar{\delta} = N^{-1} \sum_i \delta_i$ . Under the null  $\mathcal{H}_0$  of a homogeneous common trend effect, we have  $\delta_i = \delta$  for all  $i$  and  $\delta \neq 0$  in view of Assumption A2. Then

$$\begin{aligned}
 h_{it} - 1 &= \frac{1}{L(t)t^\alpha} \frac{\psi_{it} - \psi_t}{\delta + \psi_t/(L(t)t^\alpha)}, \\
 (h_{it} - 1)^2 &= \frac{(\psi_{it} - \psi_t)^2}{\psi_t^2 + L(t)^2 t^{2\alpha} \delta^2 + 2\delta L(t)t^\alpha \psi_t},
 \end{aligned}$$

and

$$(61) \quad H_t = \frac{1}{N} \sum_{i=1}^N (h_{it} - 1)^2 = \frac{\frac{1}{N} \sum_{i=1}^N (\psi_{it} - \psi_t)^2}{\psi_t^2 + L(t)^2 t^{2\alpha} \delta^2 + 2\delta L(t)t^\alpha \psi_t}.$$

Let  $\sigma_{\psi_t}^2 = N^{-1} \sum_{i=1}^N (\psi_{it} - \psi_t)^2 = N^{-1} \sum_{i=1}^N \psi_{it}^2 - \psi_t^2$ , so that by Assumptions A2 and A3 and (59) we have

$$\begin{aligned}
 (62) \quad \sigma_{\psi_t}^2 &= N^{-1} \sum_{i=1}^N \sigma_i^2 \xi_{it}^2 - N^{-2} \sum_{i=1}^N \sigma_i^2 - N^{-2} \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1) \\
 &\quad - 2N^{-2} \sum_{i=2}^N \sum_{j=1}^{i-1} \psi_{it} \psi_{jt} \\
 &= \frac{N-1}{N^2} \sum_{i=1}^N \sigma_i^2 + \frac{N-1}{N^2} \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1) - \frac{2}{N^2} \sum_{i=2}^N \sum_{j=1}^{i-1} \psi_{it} \psi_{jt} \\
 &= v_{\psi_N}^2 + N^{-1/2} \eta_{Nt} - N^{-1} \eta_{2Nt},
 \end{aligned}$$

where  $v_{\psi N}^2 = N^{-1}(1 - N^{-1}) \sum_{i=1}^N \sigma_i^2 \rightarrow v_{\psi}^2$  as  $N \rightarrow \infty$ ,  $\eta_{Nt} = N^{-1/2}(1 - N^{-1}) \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1)$ , and  $\eta_{2Nt} = 2N^{-1} \sum_{i=2}^N \sum_{j=1}^{i-1} \psi_{it} \psi_{jt}$ . In view of (28), we have

$$\eta_{Nt} \Rightarrow N(0, v_{4\psi}(\mu_{4\xi} - 1)) := \xi_{2\psi t}, \quad \text{say, as } N \rightarrow \infty,$$

so that  $\eta_{Nt} = O_p(1)$  as  $N \rightarrow \infty$ . Further, since the limit variate  $\xi_{2\psi t}$  depends on  $\{(\xi_{it}^2 - 1)\}_{i=1}^{\infty}$ , it retains the same dependence structure over  $t$  as  $(\xi_{it}^2 - 1)$ . Indeed, expanding the probability space in a suitable way, we may write  $\eta_{Nt} = \xi_{2\psi t} + o_p(1)$ , and partial sums over  $t$  satisfy a functional law

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_{2\psi t} \Rightarrow V_2(r),$$

where  $V_2$  is Brownian motion with variance  $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^4 \omega_{2ii}$ , which is a sequential functional convergence version of (30). Assumption A3 requires that the following central limit law hold jointly as both  $N, T \rightarrow \infty$ :

$$(63) \quad T^{-1/2} N^{-1/2} \sum_{t=1}^T \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1) \Rightarrow N \left( 0, \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^4 \omega_{2ii} \right).$$

Primitive conditions for this result may be developed along the lines of Phillips and Moon (1999). Note also that, in view of (30),  $N^{-1} T^{-1} \sum_{t=1}^T \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1)$  has an asymptotic mean squared error of order  $O(N^{-1} T^{-1})$ , so that

$$(64) \quad T^{-1} \sum_{t=1}^T \sigma_{\psi t}^2 = v_{\psi N}^2 + O_p(N^{-1/2} T^{-1/2}).$$

Finally, we observe that in view of the independence of the  $\xi_{it}$  across  $i$ , it follows by standard weak convergence arguments that

$$(65) \quad \eta_{2Nt} = 2N^{-1} \sum_{i=2}^N \sum_{j=1}^{i-1} \psi_{it} \psi_{jt} \Rightarrow 2 \int_0^1 U_t(r) dU_t(r),$$

where  $U_t(r)$  is a Brownian motion with variance  $v_{\psi}^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2$ . Thus,  $\eta_{2Nt} = O_p(1)$  as  $N \rightarrow \infty$ . Further, in view of (31), we have the joint convergence

$$\begin{aligned} & T^{-1/2} \sum_{t=1}^T N^{-1} \sum_{i=2}^N \sum_{j=1}^{i-1} \psi_{it} \psi_{jt} \\ & \Rightarrow N \left( 0, \lim_{N \rightarrow \infty} N^{-2} \sum_{i=2}^N \sum_{j=1}^{i-1} \sigma_i^2 \sigma_j^2 \sum_{h=-\infty}^{\infty} \gamma_i(h) \gamma_j(h) \right) \end{aligned}$$



as  $N, T \rightarrow \infty$ .

We now proceed with the derivation of the regression equation for  $H_t$ . Under  $\mathcal{H}_0$ , we can write

$$(66) \quad H_t = \frac{\sigma_{\psi_t}^2}{\psi_t^2 + L(t)^2 t^{2\alpha} \delta^2 + 2\delta L(t) t^\alpha \psi_t} \\ = \left( \frac{1}{L(t)^2 t^{2\alpha}} \right) \frac{\sigma_{\psi_t}^2 / \delta^2}{1 + L(t)^{-2} t^{-2\alpha} \psi_t^2 / \delta^2 + 2L(t)^{-1} t^{-\alpha} \psi_t / \delta}$$

and

$$H_1 = \frac{\sigma_{\psi_1}^2}{\psi_1^2 + L(1)^2 \delta^2 + 2\delta L(1) \psi_1},$$

which is independent of  $\alpha$ . Let  $\log H_1 = h_1$ . Taking logs yields

$$(67) \quad \log \frac{H_1}{H_t} = \log H_1 - \log H_t = h_1 - \log H_t$$

and, using (62), we have

$$(68) \quad \log H_t = \log \left[ \frac{v_{\psi_N}^2 + N^{-1/2} \eta_{Nt} - N^{-1} \eta_{2Nt}}{\delta^2} \right] - 2 \log L(t) - 2\alpha \log t \\ - \log \left\{ 1 + \frac{L(t)^{-2} t^{-2\alpha} \psi_t^2}{\delta^2} + \frac{2L(t)^{-1} t^{-\alpha} \psi_t}{\delta} \right\} \\ = -2 \log L(t) - 2\alpha \log t + \log \left\{ \frac{v_{\psi_N}^2}{\delta^2} \right\} + \epsilon_t,$$

where

$$(69) \quad \epsilon_t = \log \left[ 1 + \frac{N^{-1/2} \eta_{Nt}}{v_{\psi_N}^2} - \frac{\eta_{2Nt}}{N v_{\psi_N}^2} \right] \\ - \log \left\{ 1 + \frac{L(t)^{-2} t^{-2\alpha} \psi_t^2}{\delta^2} + \frac{2L(t)^{-1} t^{-\alpha} \psi_t}{\delta} \right\}.$$

Even if  $\alpha = 0$ , we can still expand the logarithm in the second term of the above expression for  $\epsilon_t$  since the slowly varying factor  $L(t)^{-1} \rightarrow 0$  for large  $t$ . Define  $\lambda_t = L(t)^{-1} t^{-\alpha} \psi_t / \delta$  and  $\zeta_t = \lambda_t^2 + 2\lambda_t$ , and using the expansion  $\log(1 + \zeta_t) = \zeta_t - \frac{1}{2} \zeta_t^2 + o(\zeta_t^3)$  and (59), we get

$$\log(1 + \zeta_t) = \lambda_t^2 + 2\lambda_t - \frac{1}{2} (\lambda_t^2 + 2\lambda_t)^2 + o_p(L(t)^{-3} t^{-3\alpha} \psi_t^3)$$

$$\begin{aligned}
 &= 2\lambda_t - \frac{1}{2}\lambda_t^4 - 2\lambda_t^3 - \lambda_t^2 + o_p(L(t)^{-3}t^{-3\alpha}\psi_t^3) \\
 &= \frac{2L(t)^{-1}t^{-\alpha}\psi_t}{\delta} - \frac{L(t)^{-2}t^{-2\alpha}\psi_t^2}{\delta^2} + O_p\left(\frac{1}{L(t)^3t^{3\alpha}N^{3/2}}\right),
 \end{aligned}$$

so that since  $N \rightarrow \infty$ ,

$$\begin{aligned}
 (70) \quad \epsilon_t &= \log\left\{1 + \frac{N^{-1/2}\eta_{Nt}}{v_{\psi N}^2} - \frac{\eta_{2Nt}}{Nv_{\psi N}^2}\right\} - \frac{2L(t)^{-1}t^{-\alpha}\psi_t}{\delta} + \frac{L(t)^{-2}t^{-2\alpha}\psi_t^2}{\delta^2} \\
 &\quad + O_p\left(\frac{1}{L(t)^3t^{3\alpha}N^{3/2}}\right)
 \end{aligned}$$

$$\begin{aligned}
 (71) \quad &= \left\{N^{-1/2}\frac{\eta_{Nt}}{v_{\psi N}^2} + \epsilon_{Nt}\right\} - \frac{2L(t)^{-1}t^{-\alpha}\psi_t}{\delta} + \frac{L(t)^{-2}t^{-2\alpha}\psi_t^2}{\delta^2} \\
 &\quad + O_p\left(\frac{1}{L(t)^3t^{3\alpha}N^{3/2}}\right),
 \end{aligned}$$

where

$$\begin{aligned}
 (72) \quad \epsilon_{Nt} &= -\frac{\eta_{2Nt}}{Nv_{\psi N}^2} - \frac{1}{2N}\frac{\eta_{Nt}^2}{v_{\psi N}^4} + O_p(N^{-3/2}) \\
 &= -\frac{1}{2N}\frac{E(\xi_{2\psi t}^2)}{v_{\psi N}^4} - \frac{\eta_{2Nt}}{Nv_{\psi N}^2} - \frac{1}{2N}\frac{\xi_{2\psi t}^2 - E(\xi_{2\psi t}^2)}{v_{\psi N}^4} + O_p(N^{-3/2}).
 \end{aligned}$$

Expressions (67) and (68) lead to the empirical regression equation

$$(73) \quad \log \frac{H_1}{H_t} - 2 \log L(t) = a + b \log t + u_t,$$

where

$$(74) \quad a = h_1 - 2 \log \frac{v_{\psi N}}{\delta}, \quad b = 2\alpha, \quad u_t = -\epsilon_t.$$

For  $t \geq [Tr]$  and  $r > 0$ , we may write

$$L(t)^{-1}t^{-\alpha} = \frac{1}{T^\alpha L(T)} \frac{L(T)}{L(t)} \frac{1}{(\frac{t}{T})^\alpha} = \frac{1}{T^\alpha L(T)} \frac{1}{(\frac{t}{T})^\alpha} \{1 + o(1)\}$$

and then

$$(75) \quad u_t = -\frac{1}{\sqrt{N}} \frac{1}{v_{\psi N}^2} \eta_{Nt} - \epsilon_{Nt} + \frac{2}{\delta} \frac{1}{t^\alpha L(t)} \psi_t - \frac{1}{\delta^2} \frac{1}{t^{2\alpha} L(t)^2} \psi_t^2$$

$$\begin{aligned}
 &+ O_p\left(\frac{1}{L(t)^3 t^{3\alpha} N^{3/2}}\right) \\
 &= -\frac{1}{\sqrt{N}} \frac{1}{v_{\psi_N}^2} \eta_{Nt} - \varepsilon_{Nt} + \frac{2}{\delta} \left[ \frac{\{1 + o(1)\}}{T^\alpha L(T)} \right] \frac{1}{\left(\frac{t}{T}\right)^\alpha} \psi_t \\
 &\quad - \frac{1}{\delta^2} \left[ \frac{\{1 + o(1)\}}{T^{2\alpha} L(T)^2} \right] \frac{1}{\left(\frac{t}{T}\right)^{2\alpha}} \psi_t^2 + O_p\left(\frac{1}{L(t)^3 t^{3\alpha} N^{3/2}}\right).
 \end{aligned}$$

Since  $\psi_t = O_p(N^{-1/2})$ ,  $\varepsilon_{Nt} = O_p(N^{-1})$ , and  $L(T) \rightarrow \infty$ , the first term of (75) dominates the behavior of the regression error  $u_t$  when  $\alpha \geq 0$ .

B.2. Proof of Theorem 1

In developing the limit theory, it is convenient to modify the regression equation (73) to avoid the singularity in the sample moment matrix that arises from the presence of an intercept and  $\log t$  in (73). Phillips (2007) provided a discussion and treatment of such issues in quite general regressions with slowly varying regressors that includes cases such as (73). It is simplest to transform to the equation

$$(76) \quad \log \frac{H_1}{H_t} - 2 \log L(t) = a^* + b \log \frac{t}{T} + u_t,$$

where  $a^* = a + b \log T$ . This transformation clearly does not affect the estimator of  $b$ .

Define the demeaned regressor

$$\tau_t = \left( \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right),$$

where  $\overline{\log \frac{t}{T}} = \frac{1}{T - [Tr] + 1} \sum_{t=[Tr]}^T \log \frac{t}{T}$ . Then, empirical regression of (76) over  $t = [Tr], [Tr] + 1, \dots, T$  for some  $r > 0$  yields

$$\hat{b} - b = \frac{\sum_{t=[Tr]}^T \tau_t u_t}{\sum_{t=[Tr]}^T \tau_t^2}.$$

Note that

$$\begin{aligned}
 (77) \quad \sum_{t=[Tr]}^T \tau_t^2 &= \sum_{t=[Tr]}^T \left( \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right)^2 \\
 &= T \left\{ \int_r^1 \left( \log s - \frac{1}{1-r} \int_r^1 \log p \, dp \right)^2 ds + o(1) \right\}
 \end{aligned}$$

$$\begin{aligned}
 &= T \left\{ \int_r^1 \log^2 s \, ds - \frac{1}{1-r} \left( \int_r^1 \log p \, dp \right)^2 + o(1) \right\} \\
 &= T \left\{ (1-r) - \left( \frac{r}{1-r} \right) \log^2 r + o(1) \right\}
 \end{aligned}$$

by Euler summation and direct evaluation of the integral

$$\int_r^1 \log^2 s \, ds - \frac{1}{1-r} \left( \int_r^1 \log p \, dp \right)^2 = (1-r) - \left\{ \frac{r}{1-r} \right\} \log^2 r.$$

Using (75) we have

$$\begin{aligned}
 (78) \quad \hat{b} - b &= -\frac{1}{v_{\psi N}^2} \frac{\frac{1}{\sqrt{NT}} \sum_{t=[Tr]}^T \tau_t \eta_{Nt} + \frac{1}{T} \sum_{t=[Tr]}^T \tau_t \varepsilon_{Nt}}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} \\
 &+ \frac{2}{\delta} \frac{\frac{1}{T} \sum_{t=[Tr]}^T \tau_t \psi_t / (t^\alpha L(t))}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} - \frac{1}{\delta^2} \frac{\frac{1}{T} \sum_{t=[Tr]}^T \tau_t \psi_t^2 / (t^{2\alpha} L(t)^2)}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} \\
 &+ O_p \left( \frac{1}{L(T)^3 T^{3\alpha} N^{3/2}} \right).
 \end{aligned}$$

Next observe that

$$\begin{aligned}
 (79) \quad \frac{1}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \eta_{Nt} &= \frac{1}{\sqrt{T}} \frac{1}{\sqrt{N}} \sum_{t=[Tr]}^T \tau_t \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1) \\
 &= \frac{1}{\sqrt{T} \sqrt{N}} \sum_{i=1}^N \sum_{t=[Tr]}^T \left( \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right) \sigma_i^2 (\xi_{it}^2 - 1) \\
 &\Rightarrow N \left( 0, \omega_\eta^2 \int_r^1 \left( \log s - \frac{1}{1-r} \int_r^1 \log p \, dp \right)^2 ds \right) \\
 &= N \left( 0, \omega_\eta^2 \left\{ (1-r) - \left( \frac{r}{1-r} \right) \log^2 r \right\} \right)
 \end{aligned}$$

in view of (29), where  $\omega_\eta^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^4 \omega_{2ii}$  and  $\omega_{2ii}$  is the long run variance of  $\xi_{it}^2$ . Also noting the fact that  $E(\xi_{2\psi t}^2)$  is constant over  $t$  and  $\sum_{t=[Tr]}^T \tau_t = 0$ , we find that

$$(80) \quad \frac{1}{T} \sum_{t=[Tr]}^T \tau_t \varepsilon_{Nt} = -\frac{1}{v_{\psi N}^2} \frac{1}{NT} \sum_{t=[Tr]}^T \tau_t \eta_{2Nt}$$

$$\begin{aligned}
 & -\frac{1}{2v_{\psi N}^4} \frac{1}{NT} \sum_{t=[Tr]}^T \tau_t (\xi_{2\psi t}^2 - E(\xi_{2\psi t}^2)) + O_p(N^{-3/2}) \\
 & = O_p\left(\frac{1}{\sqrt{TN}} + \frac{1}{N^{3/2}}\right),
 \end{aligned}$$

since both  $T^{-1/2} \sum_{t=[Tr]}^T \tau_t \eta_{2Nt}$  and  $T^{-1/2} \sum_{t=[Tr]}^T \tau_t (\xi_{2\psi t}^2 - E(\xi_{2\psi t}^2))$  are  $O_p(1)$ .  
 Further,

$$\begin{aligned}
 (81) \quad & \frac{\sqrt{TN}}{T} \sum_{t=[Tr]}^T \tau_t \frac{1}{t^\alpha L(t)} \psi_t \\
 & = \frac{1}{\sqrt{TN}} \sum_{i=1}^N \sum_{t=[Tr]}^T \tau_t \frac{1}{t^\alpha L(t)} \psi_{it} \\
 & = \frac{1}{\sqrt{TN}} \sum_{i=1}^N \sum_{t=[Tr]}^T \tau_t \left[ \frac{1}{T^\alpha L(T)} \frac{L(T)}{L(t)} \frac{1}{(\frac{t}{T})^\alpha} \psi_{it} \right] \\
 & = \left\{ \frac{1}{T^\alpha L(T)} \{1 + o_p(1)\} \right\} \\
 & \quad \times \frac{1}{\sqrt{TN}} \sum_{i=1}^N \sum_{t=[Tr]}^T \left( \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right) \left( \frac{t}{T} \right)^{-\alpha} \psi_{it} \\
 & \sim \frac{1}{T^\alpha L(T)} N \left( 0, \omega_\psi^2 \int_r^1 \left\{ \left( \log s - \frac{1}{1-r} \int_r^1 \log p \, dp \right)^2 s^{-2\alpha} \, ds \right\} \right) \\
 & = O_p\left(\frac{1}{T^\alpha L(T)}\right)
 \end{aligned}$$

and, when  $\alpha > 0$ , we have

$$\begin{aligned}
 & \frac{1}{\delta^2} \frac{\sqrt{TN}}{T} \sum_{t=[Tr]}^T \tau_t \frac{1}{t^{2\alpha} L(t)^2} \psi_t^2 \\
 & = \frac{1}{\delta^2} \frac{\sqrt{TN}}{T} \sum_{t=[Tr]}^T \tau_t \left[ \frac{1}{t^{2\alpha} L(t)^2} \right] \\
 & \quad \times \left\{ N^{-2} \sum_{i=1}^N \sigma_i^2 + N^{-2} \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1) + N^{-2} \sum_{\substack{i,j=1 \\ i \neq j}}^N \psi_{it} \psi_{jt} \right\}
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{\sqrt{TN}}{\delta^2} \frac{1}{T} \sum_{t=[Tr]}^T \tau_t \left[ \frac{1}{T^{2\alpha} L(T)^2} \left( \frac{L(T)}{L(t)} \right)^2 \frac{1}{\left(\frac{t}{T}\right)^{2\alpha}} \right] \frac{v_{\psi N}^2}{N\{1 - N^{-1}\}} \\
 &\quad + \frac{1}{\delta^2 N} \frac{1}{\sqrt{T}} \\
 &\quad \times \sum_{t=[Tr]}^T \tau_t \left[ \frac{1}{T^{2\alpha} L(T)^2} \left( \frac{L(T)}{L(t)} \right)^2 \frac{1}{\left(\frac{t}{T}\right)^{2\alpha}} \right] \frac{1}{\sqrt{N}} \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1), \\
 &\frac{1}{\delta^2 N^{1/2}} \frac{1}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \left[ \frac{1}{T^{2\alpha} L(T)^2} \left( \frac{L(T)}{L(t)} \right)^2 \frac{1}{\left(\frac{t}{T}\right)^{2\alpha}} \right] \frac{1}{N} \sum_{\substack{i,j=1 \\ i \neq j}}^N \psi_{it} \psi_{jt} \\
 (82) \quad &= \frac{v_{\psi N}^2 \sqrt{T} \{1 + o(1)\}}{\delta^2 \sqrt{N} T^{2\alpha} L(T)^2} \frac{1}{T} \sum_{t=[Tr]}^T \tau_t \left( \frac{t}{T} \right)^{-2\alpha} \\
 &\quad + \frac{\{1 + o(1)\}}{\delta^2 N T^{2\alpha} L(T)^2} \frac{1}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \left( \frac{t}{T} \right)^{-2\alpha} \frac{1}{\sqrt{N}} \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1)
 \end{aligned}$$

$$\begin{aligned}
 (83) \quad &+ \frac{\{1 + o(1)\}}{\delta^2 N^{1/2} T^{2\alpha} L(T)^2} \frac{1}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \left( \frac{t}{T} \right)^{-2\alpha} \left( \frac{1}{N} \sum_{\substack{i,j=1 \\ i \neq j}}^N \psi_{it} \psi_{jt} \right)
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{v_{\psi}^2}{\delta^2} \frac{T^{1/2}}{T^{2\alpha} L(T)^2 N^{1/2}} \\
 &\quad \times \int_r^1 \left\{ \left( \log s - \frac{1}{1-r} \int_r^1 \log p \, dp \right) s^{-2\alpha} \, ds \right\} \{1 + o_p(1)\} \\
 (84) \quad &+ O_p \left( \frac{1}{T^{2\alpha} L(T)^2 \sqrt{N}} \right),
 \end{aligned}$$

since

$$\begin{aligned}
 &\frac{1}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \left( \frac{t}{T} \right)^{-2\alpha} \frac{1}{\sqrt{N}} \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1) = O_p(1), \\
 &\frac{1}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \left( \frac{t}{T} \right)^{-2\alpha} \left( \frac{1}{N} \sum_{\substack{i,j=1 \\ i \neq j}}^N \psi_{it} \psi_{jt} \right) = O_p(1)
 \end{aligned}$$

in view of (30) and (31). When  $\alpha = 0$ , it is apparent that

$$\sum_{t=[Tr]}^T \tau_t \left(\frac{t}{T}\right)^{-2\alpha} v_{\psi}^2 = \sum_{t=[Tr]}^T \tau_t v_{\psi}^2 = 0$$

in line (82) of the earlier argument, in which case the first term of (84) is zero and the second term dominates, giving

$$(85) \quad \frac{1}{\delta^2} \frac{\sqrt{TN}}{T} \sum_{t=[Tr]}^T \tau_t \frac{1}{L(t)^2} \psi_t^2 = O_p\left(\frac{1}{L(T)^2 \sqrt{N}}\right).$$

From (78) we have

$$(86) \quad \begin{aligned} &\sqrt{NT}(\hat{b} - b) \\ &= -\frac{1}{v_{\psi N}^2} \frac{\frac{1}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \eta_{Nt} + \frac{\sqrt{N}}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \varepsilon_{Nt}}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} \\ &\quad + \frac{2}{\delta} \frac{\frac{\sqrt{N}}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \psi_t / (t^\alpha L(t))}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} - \frac{1}{\delta^2} \frac{\frac{\sqrt{N}}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \psi_t^2 / (t^{2\alpha} L(t)^2)}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} \\ &\quad + O_p\left(\frac{\sqrt{T}}{L(T)^3 T^{3\alpha} N}\right), \end{aligned}$$

and it then follows from (80)–(84) that when  $\alpha > 0$ ,

$$(87) \quad \begin{aligned} \sqrt{NT}(\hat{b} - b) &= -\frac{1}{v_{\psi N}^2} \frac{\frac{1}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \eta_{Nt}}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} + O_p\left(\frac{1}{N^{1/2}} + \frac{\sqrt{T}}{N}\right) \\ &\quad + O_p\left(\frac{1}{T^\alpha L(T)}\right) + O_p\left(\frac{T^{1/2}}{T^{2\alpha} L(T)^2 N^{1/2}}\right) \\ &\quad + O_p\left(\frac{\sqrt{T}}{L(T)^3 T^{3\alpha} N}\right) \\ &\Rightarrow \frac{1}{v_{\psi}^2} N \left(0, \omega_{\eta}^2 \left\{ (1-r) - \left(\frac{r}{1-r}\right) \log^2 r \right\}^{-1}\right) \end{aligned}$$

as  $T, N \rightarrow \infty$ , provided  $T^{1/2}/(T^{2\alpha} L(T)^2 N^{1/2}) \rightarrow 0$ . When  $\alpha = 0$ , we have, using (85),

$$\sqrt{NT}(\hat{b} - b) = -\frac{1}{v_{\psi N}^2} \frac{\frac{1}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \eta_{Nt}}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} + O_p\left(\frac{1}{N^{1/2}} + \frac{\sqrt{T}}{N}\right)$$

$$+ O_p\left(\frac{1}{L(T)}\right) + O_p\left(\frac{1}{L(T)^2\sqrt{N}}\right) + O_p\left(\frac{\sqrt{T}}{L(T)^3N}\right)$$

and precisely the same limit theory as (87) applies provided  $T^{1/2}/N \rightarrow 0$ .

It follows that in both cases we have  $\sqrt{NT}(\hat{b} - b) \Rightarrow N(0, \Omega^2)$ , where

$$(88) \quad \Omega^2 = \frac{\omega_\eta^2}{v_\psi^4} \left\{ (1 - r) - \left(\frac{r}{1 - r}\right) \log^2 r \right\}^{-1},$$

$\omega_\eta^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^4 \omega_{2ii}$ , and  $v_\psi^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2$ , and where  $\omega_{2ii}$  is the long run variance of  $\xi_{it}^2$ . This gives the required result (a).

When the relative rate condition  $T^{1/2}/(T^{2\alpha}L(T)^2N^{1/2}) \rightarrow 0$  does not hold, the third term of (86) enters into the limit theory as a bias term. In particular, we have

$$\begin{aligned} & \sqrt{NT} \left\{ (\hat{b} - b) + \frac{1}{\delta^2} \frac{\frac{1}{T} \sum_{t=[Tr]}^T \tau_t \psi_t^2 / (t^{2\alpha}L(t)^2)}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} \right\} \\ &= -\frac{1}{v_{\psi N}^2} \frac{\frac{1}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \eta_{Nt} + \frac{\sqrt{N}}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \epsilon_{Nt}}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} + O_p\left(\frac{\sqrt{T}}{L(T)^3 T^{3\alpha} N}\right) \\ &= -\frac{1}{v_{\psi N}^2} \frac{\frac{1}{\sqrt{T}} \sum_{t=[Tr]}^T \tau_t \eta_{Nt}}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} + O_p\left(\frac{1}{N^{1/2}} + \frac{\sqrt{T}}{N}\right) + O_p\left(\frac{\sqrt{T}}{L(T)^3 T^{3\alpha} N}\right) \end{aligned}$$

and using (87) we have

$$\sqrt{NT} \left\{ (\hat{b} - b) + \frac{1}{\delta^2} \frac{\frac{1}{T} \sum_{t=[Tr]}^T \tau_t \psi_t^2 / (t^{2\alpha}L(t)^2)}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} \right\} \Rightarrow N(0, \Omega^2),$$

provided  $\frac{\sqrt{T}}{N} \rightarrow 0$ . In this case, there is an asymptotic bias of the form

$$-\frac{v_\psi^2}{\delta^2} \frac{1}{T^{2\alpha}L(T)^2N} \int_r^1 \left\{ \left( \log s - \frac{1}{1 - r} \int_r^1 \log p dp \right) s^{-2\alpha} ds \right\}$$

in the estimation of  $b$ . This bias is of  $O(T^{-2\alpha}L(T)^{-2}N^{-1})$  and will generally be quite small when  $\alpha > 0$ . The bias is zero when  $\alpha = 0$  because  $\int_r^1 (\log s - \frac{1}{1-r} \int_r^1 \log p dp) ds = 0$ , explaining the milder rate condition in this case.

### B.3. Asymptotic Variance Formula

Since the regressor in (76) is deterministic, we may consistently estimate the asymptotic variance  $\Omega^2$  in a simple way by estimating the long run variance of



$u_t$  using the least squares residuals  $\hat{u}_t$ . In particular, we may use the variance estimate

$$V(\hat{b}) = \widehat{\text{lvar}}_r(\hat{u}_t) \left[ \sum_{t=[Tr]}^T \tau_t^2 \right]^{-1},$$

where  $\widehat{\text{lvar}}_r(\hat{u}_t)$  is a consistent estimate of  $N^{-1}(\delta^2/v_\psi^2)\omega_\eta^2$ . To construct  $\widehat{\text{lvar}}_r(\hat{u}_t)$  we may use a conventional HAC estimate, as we now show.

We start by working directly with  $u_t$  and its autocovariance sequence. From (75) we have

$$\begin{aligned} (89) \quad \sqrt{N}u_t &= -\frac{\eta_{Nt}}{v_{\psi N}^2} + 2\frac{\{1+o(1)\}}{T^\alpha L(T)} \frac{1}{(\frac{1}{T})^\alpha} \sqrt{N}\psi_t + o_p(1) \\ &= -\frac{1}{v_{\psi N}^2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1) + o_p(1) \\ (90) \quad &= -\frac{1}{v_{\psi N}^2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \sigma_i^2 \eta_{it} + o_p(1) := -\frac{w_t}{v_{\psi N}^2} + o_p(1), \end{aligned}$$

where  $\eta_{it} = \xi_{it}^2 - 1$ , whose long run variance is  $\omega_{2ii}$ . The serial autocovariances of the leading term  $w_t = \frac{1}{\sqrt{N}} \sum_{i=1}^N \sigma_i^2 \eta_{it}$  are  $E(w_t w_{t+l}) = N^{-1} \sum_{i=1}^N \sigma_i^4 E(\eta_{it} \times \eta_{it+l})$ , and as  $M \rightarrow \infty$ , it follows that

$$\begin{aligned} \sum_{l=-M}^M E(w_t w_{t+l}) &= \frac{1}{N} \sum_{i=1}^N \sigma_i^4 \sum_{l=-M}^M E(\eta_{it} \eta_{it+l}) \\ &= \frac{1}{N} \sum_{i=1}^N \sigma_i^4 \left\{ \sum_{l=-\infty}^{\infty} E(\eta_{it} \eta_{it+l}) + o(1) \right\} \\ &\rightarrow \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^4 \omega_{2ii}, \end{aligned}$$

where  $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^4 \omega_{2ii}$ . Contributions to the long run variance of  $\sqrt{N}u_t$  from the second and higher order terms of (89) are of  $O(L(T)^{-2}T^{-2\alpha})$  for  $t \geq [Tr]$  and  $r > 0$ . Hence, the long run variance of  $\sqrt{N}u_t$  is given by

$$\text{lvar}(\sqrt{N}u_t) = \lim_{N \rightarrow \infty} \frac{1}{v_{\psi N}^4} \frac{1}{N} \sum_{i=1}^N \sigma_i^4 \omega_{2ii} = \frac{\omega_\eta^2}{v_\psi^4} := \Omega_u^2 \quad \text{say.}$$

Thus, the asymptotic variance formula (88) is

$$\Omega^2 = \frac{\Omega_u^2}{\{(1-r) - (\frac{r}{1-r}) \log^2 r\}^2}.$$

The denominator can be directly calculated or estimated in the usual manner with the moment sum of squares

$$\sum_{t=[Tr]}^T \tau_t^2 \sim T \left\{ (1-r) - \left( \frac{r}{1-r} \right) \log^2 r \right\},$$

as shown in (77) above. The numerator,

$$\Omega_u^2 = \lim_{N \rightarrow \infty} \frac{\delta^4}{v_{\psi N}^4} \frac{1}{N} \sum_{i=1}^N \sigma_i^4 \omega_{2ii},$$

is the limit of a weighted average of long run variances. As we next investigate, it may be estimated using a long run variance estimate with the residuals of the regression, namely  $\hat{\Omega}_u^2 = \widehat{\text{Ivar}}_r(\hat{u}_t)$ .

#### B.4. Estimation of the Weighted Average Long Run Variance

The sample serial covariances of the leading term  $w_t$  of the regression error  $u_t$  (using the available observations in the regression from  $t = [Tr], \dots, T$ ) have the form

$$\begin{aligned} (91) \quad & \frac{1}{T} \sum_{[Tr] \leq t, t+l \leq T} w_t w_{t+l} \\ &= \frac{1}{N} \sum_{i,j=1}^N \sigma_i^2 \sigma_j^2 \frac{1}{T} \sum_{[Tr] \leq t, t+l \leq T} \eta_{it} \eta_{jt+l} \\ &= \frac{1}{N} \sum_{i=1}^N \sigma_i^4 \frac{1}{T} \sum_{[Tr] \leq t, t+l \leq T} \eta_{it} \eta_{it+l} \\ &\quad + \frac{1}{\sqrt{T}} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^N \sigma_i^2 \left\{ \frac{1}{\sqrt{NT}} \sum_{j \neq i}^N \sigma_j^2 \sum_{[Tr] \leq t, t+l \leq T} \eta_{it} \eta_{jt+l} \right\} \right] \\ &= \frac{1}{N} \sum_{i=1}^N \sigma_i^4 \left\{ \frac{1}{T} \sum_{[Tr] \leq t, t+l \leq T} \eta_{it} \eta_{it+l} \right\} + O_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned}$$

By virtue of the usual process of HAC estimation for  $M \rightarrow \infty$  as  $T \rightarrow \infty$ , we have

$$\sum_{l=-M}^M \left\{ \frac{1}{T} \sum_{[Tr] \leq t, t+l \leq T} \eta_{it} \eta_{it+l} \right\} \rightarrow_p (1-r)\omega_{2ii}^2,$$

where the factor  $(1-r)$  reflects the fact that only the fraction  $1-r$  of the time series data is used in the regression. For  $M$  satisfying  $\frac{M}{\sqrt{T}} + \frac{1}{M} \rightarrow 0$  as  $T \rightarrow \infty$ , we find from (91) and standard HAC limit theory that

$$\begin{aligned} (92) \quad & \sum_{l=-M}^M \frac{1}{T} \sum_{[Tr] \leq t, t+l \leq T} w_t w_{t+l} \\ &= \frac{1}{N} \sum_{i=1}^N \sigma_i^4 \left\{ \sum_{l=-M}^M \frac{1}{T} \sum_{[Tr] \leq t, t+l \leq T} \eta_{it} \eta_{it+l} \right\} + o_p(1) \\ &= \frac{1}{N} \sum_{i=1}^N \sigma_i^4 \{ (1-r)\omega_{2ii}^2 + o_p(1) \} + o_p(1) \\ &\rightarrow_p (1-r) \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^4 \omega_{2ii}. \end{aligned}$$

Since the scaled regression residuals  $\sqrt{N}\hat{u}_t$  consistently estimate the quantities  $-w_t/v_{\psi N}^2$  in (90), we correspondingly have

$$\widehat{\text{ivar}}(\sqrt{N}\hat{u}_t) \rightarrow_p (1-r) \frac{1}{v_{\psi N}^4} \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^4 \omega_{2ii} = (1-r)\Omega_u^2.$$

If we use a standardization of  $1/(T - [Tr] + 1)$  rather than  $1/T$  in the sample serial covariances in (92), we have

$$\sum_{l=-M}^M \frac{1}{T - [Tr] + 1} \sum_{[Tr] \leq t, t+l \leq T} w_t w_{t+l} \rightarrow_p \Omega_u^2,$$

and the corresponding estimate (where the subscript  $r$  signifies the use of the scaling factor  $1/(T - [Tr] + 1)$  in the sample covariance formulae)

$$(93) \quad \widehat{\text{ivar}}_r(\sqrt{N}\hat{u}_t) = \sum_{l=-M}^M \frac{N}{T - [Tr] + 1} \sum_{[Tr] \leq t, t+l \leq T} \hat{u}_t \hat{u}_{t+l} \rightarrow_p \Omega_u^2.$$

The same behavior is observed for other HAC estimates constructed with different lag kernels.

Then the asymptotic variance estimate of  $\hat{b}$  is

$$s_b^2 = \widehat{\text{lvar}}_r(\hat{u}_t) \left[ \sum_{t=\lfloor Tr \rfloor}^T \tau_t^2 \right]^{-1} = \frac{1}{N} \widehat{\text{lvar}}_r(\sqrt{N}\hat{u}_t) \left[ \sum_{t=\lfloor Tr \rfloor}^T \tau_t^2 \right]^{-1} \\ \sim \frac{1}{NT} \Omega_u^2 \left\{ (1-r) - \left( \frac{r}{1-r} \right) \log^2 r \right\}^{-1} = \frac{1}{NT} \Omega^2,$$

and  $NTs_b^2 \rightarrow_p \Omega^2$  as  $N, T \rightarrow \infty$ . Accordingly, the  $t$  ratio  $t_b = (\hat{b} - b)/s_b \Rightarrow N(0, 1)$  and result (b) follows.

B.5. Proof of Theorem 2

(a)  $\alpha \geq 0$ . We assume  $\delta_i \sim \text{iid}(\delta, \sigma_\delta^2)$  and let  $\bar{\delta} = N^{-1} \sum_{i=1}^N \delta_i$ . Under this alternative, we have, from (60),

$$h_{it} - 1 = \frac{\delta_i - \bar{\delta} + (\psi_{it} - \psi_t)/(L(t)t^\alpha)}{\bar{\delta} + \psi_t/(L(t)t^\alpha)},$$

so that

$$(94) \quad H_t = \frac{\sigma^2}{\bar{\delta}^2} \frac{1 + (\sigma_{\psi_t}^2/\sigma^2)L(t)^{-2}t^{-2\alpha} + 2(\sigma_{\delta\psi_t}/\sigma^2)L(t)^{-1}t^{-\alpha}}{1 + L(t)^{-2}t^{-2\alpha}\psi_t^2/\bar{\delta}^2 + 2L(t)^{-1}t^{-\alpha}\psi_t/\bar{\delta}},$$

where  $\sigma^2 = N^{-1} \sum (\delta_i - \bar{\delta})^2 \rightarrow_p \sigma_\delta^2$  as  $N \rightarrow \infty$ ,

$$(95) \quad \sigma_{\delta\psi_t} = N^{-1} \sum_{i=1}^N (\delta_i - \bar{\delta})(\psi_{it} - \psi_t) = N^{-1/2} s_{Nt} + O_p(N^{-1}),$$

where  $s_{Nt} = N^{-1/2} \sum_{i=1}^N (\delta_i - \bar{\delta})\sigma_i \xi_{it} = O_p(1)$ , and  $\sigma_{\psi_t}^2 = v_{\psi_N}^2 + N^{-1/2} \eta_{Nt} + O_p(N^{-1})$  from (62) above. Note that  $H_t \rightarrow \sigma^2/\bar{\delta}^2 > 0$  as  $T, N \rightarrow \infty$ .

Taking logs in (94) and assuming  $\sigma^2 > 0$  (otherwise the null hypothesis holds), we have

$$\log H_t = 2 \log \frac{\sigma}{\bar{\delta}} + \epsilon_t,$$

where

$$(96) \quad \epsilon_t = \log\{1 + (\sigma_{\psi_t}^2/\sigma^2)L(t)^{-2}t^{-2\alpha} + 2(\sigma_{\delta\psi_t}/\sigma^2)L(t)^{-1}t^{-\alpha}\} \\ - \log\{1 + L(t)^{-2}t^{-2\alpha}\psi_t^2/\bar{\delta}^2 + 2L(t)^{-1}t^{-\alpha}\psi_t/\bar{\delta}\}.$$

The generating process for  $\log(H_1/H_t)$  therefore has the form, under the alternative,

$$(97) \quad \log \frac{H_1}{H_t} - 2 \log L(t) = \log H_1 - 2 \log \frac{\sigma}{\delta} - 2 \log L(t) - \epsilon_t,$$

while the fitted regression is

$$(98) \quad \log \frac{H_1}{H_t} - 2 \log L(t) = \hat{a}^* + \hat{b} \log \left( \frac{t}{T} \right) + \text{residual},$$

so that

$$(99) \quad \hat{b} = - \frac{\sum_{t=[Tr]}^T \tau_t \{2 \log L(t) + \epsilon_t\}}{\sum_{t=[Tr]}^T \tau_t^2}.$$

Note that  $t = Ta$  for some  $a > 0$  and for  $L(t) = \log t$ , we have

$$(100) \quad \begin{aligned} \log L(t) &= \log L(Ta) = \log \{ \log T + \log a \} = \log \left[ \log T \left\{ 1 + \frac{\log a}{\log T} \right\} \right] \\ &= \log \log T + \frac{\log \frac{t}{T}}{\log T} - \frac{1}{2} \frac{\log^2 \frac{t}{T}}{\log^2 T} + O \left( \frac{\log^3 \frac{t}{T}}{\log^3 T} \right), \end{aligned}$$

giving

$$(101) \quad \log L(t) - \overline{\log L(t)} = \frac{\log \frac{t}{T} - \overline{\log \frac{t}{T}}}{\log T} - \frac{1}{2} \frac{\log^2 \frac{t}{T} - \overline{\log^2 \frac{t}{T}}}{\log^2 T} + O \left( \frac{1}{\log^3 T} \right)$$

and

$$(102) \quad \begin{aligned} &\frac{\sum_{t=[Tr]}^T \tau_t \log L(t)}{\sum_{t=[Tr]}^T \tau_t^2} \\ &= \frac{\sum_{t=[Tr]}^T \tau_t \{ \log L(t) - \overline{\log L(t)} \}}{\sum_{t=[Tr]}^T \tau_t^2} \\ &= \frac{1}{\log T} \frac{T^{-1} \sum_{t=[Tr]}^T \tau_t (\log \frac{t}{T} - \overline{\log \frac{t}{T}})}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} \\ &\quad - \frac{1}{2 \log^2 T} \frac{T^{-1} \sum_{t=[Tr]}^T \tau_t (\log^2 \frac{t}{T} - \overline{\log^2 \frac{t}{T}})}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} + O \left( \frac{1}{\log^3 T} \right) \\ &= \frac{1}{\log T} \end{aligned}$$

$$\begin{aligned}
 & - \frac{1}{2 \log^2 T} \\
 & \times \frac{\int_r^1 (\log s - \frac{1}{1-r} \int_r^1 \log p dp) (\log^2 s - \frac{1}{1-r} \int_r^1 \log^2 p dp) ds}{\int_r^1 (\log s - \frac{1}{1-r} \int_r^1 \log p dp)^2 ds} \\
 & + O\left(\frac{1}{\log^3 T}\right) \\
 & = \frac{1}{\log T} - \frac{g(r)}{2 \log^2 T} + O\left(\frac{1}{\log^3 T}\right),
 \end{aligned}$$

where

$$g(r) = \frac{2(\log r)r^2 - (\log^3 r)r + 2(\log^2 r)r - 2(\log r)r - 4(1-r)^2}{(1-r)^2 - r \log^2 r}.$$

Hence, under the alternative, we have

$$\begin{aligned}
 (103) \quad \hat{b} & = - \frac{\sum_{t=[Tr]}^T \tau_t \epsilon_t}{\sum_{t=[Tr]}^T \tau_t^2} - \frac{2 \sum_{t=[Tr]}^T \tau_t \{\log L(t) - \overline{\log L(t)}\}}{\sum_{t=[Tr]}^T \tau_t^2} \\
 & = - \frac{\sum_{t=[Tr]}^T \tau_t \epsilon_t}{\sum_{t=[Tr]}^T \tau_t^2} - \frac{2}{\log T} + \frac{g(r)}{\log^2 T} + O_p\left(\frac{1}{\log^3 T}\right).
 \end{aligned}$$

Next consider  $\sum_{t=[Tr]}^T \tau_t \epsilon_t$ . Note that  $\sigma_{\delta \psi_t} = N^{-1/2} s_{Nt} + O_p(N^{-1}) = O_p(N^{-1/2})$  from (95),  $\psi_t = O_p(N^{-1/2})$  from (58), and  $\sigma_{\psi_t}^2 = v_{\psi N}^2 + N^{-1/2} \eta_{Nt} + O_p(N^{-1})$  from (62), where  $v_{\psi N}^2 = N^{-1} \sum_{i=1}^N \sigma_i^2 \rightarrow v_{\psi}^2$  as  $N \rightarrow \infty$  and  $\eta_{Nt} = N^{-1/2} \times \sum_{i=1}^N \sigma_i^2 (\xi_{it}^2 - 1) = O_p(1)$ . Hence, expanding (96) for  $t \geq [Tr]$  with  $r > 0$ , we get

$$\begin{aligned}
 (104) \quad \epsilon_t & = \frac{\sigma_{\psi_t}^2}{\sigma^2} \frac{1}{L(t)^2 t^{2\alpha}} + 2 \frac{\sigma_{\delta \psi_t}}{\sigma^2} \frac{1}{L(t) t^\alpha} - \frac{2}{\delta} \frac{1}{L(t) t^\alpha} \psi_t + O_p\left(\frac{1}{NL(T)^2 T^{2\alpha}}\right) \\
 & = \frac{v_{\psi N}^2}{\sigma^2} \frac{1}{L(t)^2 t^{2\alpha}} + \frac{1}{\sigma^2} \frac{N^{-1/2} \eta_{Nt}}{L(t)^2 t^{2\alpha}} + 2 \frac{\sigma_{\delta \psi_t}}{\sigma^2} \frac{1}{L(t) t^\alpha} - \frac{2}{\delta} \frac{1}{L(t) t^\alpha} \psi_t \\
 & \quad + O_p\left(\frac{1}{NL(T)^2 T^{2\alpha}}\right) \\
 & = \frac{v_{\psi N}^2}{\sigma^2} \frac{1}{L(t)^2 t^{2\alpha}} + O_p\left(\frac{1}{\sqrt{NL(T)} T^\alpha}\right).
 \end{aligned}$$

It follows that

$$\begin{aligned}
 (105) \quad & \frac{1}{T} \sum_{t=[Tr]}^T \tau_t \epsilon_t \\
 &= \frac{v_{\psi N}^2}{\sigma^2} \frac{1}{L(T)^2 T^{1+2\alpha}} \sum_{t=[Tr]}^T \tau_t \frac{L(T)^2}{L(t)^2} \left(\frac{t}{T}\right)^{-2\alpha} + O_p\left(\frac{1}{\sqrt{N}L(T)T^\alpha}\right) \\
 &= \frac{v_{\psi N}^2}{\sigma^2} \frac{1}{L(T)^2 T^{1+2\alpha}} \sum_{t=[Tr]}^T \tau_t \left(\frac{t}{T}\right)^{-2\alpha} \{1 + o(1)\} + O_p\left(\frac{1}{\sqrt{N}L(T)T^\alpha}\right) \\
 &= \frac{v_{\psi N}^2}{\sigma^2} \frac{1}{L(T)^2 T^{2\alpha}} \int_r^1 \left\{ \log s - \frac{1}{1-r} \int_r^1 \log p \, dp \right\} s^{-2\alpha} \, ds \{1 + o(1)\} \\
 &\quad + O_p\left(\frac{1}{\sqrt{N}L(T)T^\alpha}\right) \\
 &= \frac{v_{\psi N}^2}{\sigma^2} \frac{r^*(\alpha)}{L(T)^2 T^{2\alpha}} + O_p\left(\frac{1}{\sqrt{N}L(T)T^\alpha}\right),
 \end{aligned}$$

where

$$\begin{aligned}
 (106) \quad & r^*(\alpha) = \int_r^1 \left\{ \log s - \frac{1}{1-r} \int_r^1 \log p \, dp \right\} s^{-2\alpha} \, ds \\
 &= \begin{cases} (2\alpha + 2r\alpha \log r - r \log r - 2\alpha r + 2r^{2-2\alpha}\alpha + r^{1-2\alpha} \log r \\ \quad - 2r^{1-2\alpha}\alpha \log r - 2r^{1-2\alpha}\alpha) \\ \quad / ((2\alpha r - r + 1 - 2\alpha)(2\alpha - 1)) & \text{with } \alpha \neq \frac{1}{2}, \\ \frac{(\ln r)r - 2r + 2 + \ln r}{r} & \text{with } \alpha = \frac{1}{2}. \end{cases}
 \end{aligned}$$

Hence, from (103) and (77) we obtain

$$\begin{aligned}
 (107) \quad & \hat{b} = -\frac{\frac{1}{T} \sum_{t=[Tr]}^T \tau_t \epsilon_t}{\frac{1}{T} \sum_{t=[Tr]}^T \tau_t^2} - \frac{2}{\log T} + \frac{g(r)}{\log^2 T} + O_p\left(\frac{1}{\log^3 T}\right) \\
 &= -\left(\frac{v_{\psi N}^2}{\sigma^2}\right) \left(\frac{r^*(\alpha)}{\{(1-r) - (\frac{r}{1-r}) \log^2 r\}}\right) \left(\frac{1}{L(T)^2 T^{2\alpha}}\right) - \frac{2}{\log T} \\
 &\quad + \frac{g(r)}{\log^2 T} + O_p\left(\frac{1}{\log^3 T}\right)
 \end{aligned}$$

$$= -\left(\frac{v_{\psi N}^2}{\sigma^2}\right)\left(\frac{r(\alpha)}{L(T)^2 T^{2\alpha}}\right) - \frac{2}{\log T} + \frac{g(r)}{\log^2 T} + O_p\left(\frac{1}{\log^3 T}\right),$$

where

$$(108) \quad r(\alpha) = \frac{r^*(\alpha)}{(1-r) - \left(\frac{r}{1-r}\right) \log^2 r}.$$

When  $\alpha \geq 0$  and  $L(T) = \log T$ , the second term in (107) dominates and we have

$$(109) \quad \hat{b} = -\frac{2}{\log T} + \frac{g(r)}{\log^2 T} - \left(\frac{v_{\psi N}^2}{\sigma^2}\right)\left(\frac{r(\alpha)}{L(T)^2 T^{2\alpha}}\right) + O_p\left(\frac{1}{\log^3 T}\right).$$

Thus,  $\hat{b} \rightarrow_p 0$  in this case. Heuristically, this outcome is explained by the fact that  $H_t$  tends to a positive constant, so that the dependent variable in (98) behaves like  $-2\log L(t)$  for large  $t$ . Since  $\log L(t)$  is the log of a slowly varying function at infinity, its regression coefficient on  $\log t$  is expected to be zero. More particularly, the regression of  $-2\log L(t)$  on a constant and  $\log(\frac{t}{T})$  produces a slope coefficient that is negative and tends to zero like  $-\frac{2}{\log T}$ , as shown in (102).

Next consider the standard error of  $\hat{b}$  under the alternative. Writing the residual in (98) as  $\hat{u}_t$ , the long run variance estimate has the typical form

$$(110) \quad \widehat{\text{Ivar}}_r(\sqrt{N}\hat{u}_t) = \sum_{l=-M}^M \frac{N}{T - [Tr] + 1} \sum_{[Tr] \leq t, t+l \leq T} \hat{u}_t \hat{u}_{t+l}.$$

In view of (97) and (98), and with  $L(t) = \log t$ , we deduce from (101) and (102) that

$$\begin{aligned} (111) \quad \hat{u}_t &= -(\epsilon_t - \bar{\epsilon}) - 2\{\log L(t) - \overline{\log L(t)}\} - \hat{b} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\} \\ &= -(\epsilon_t - \bar{\epsilon}) - 2 \frac{\log \frac{t}{T} - \overline{\log \frac{t}{T}}}{\log T} - \hat{b} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\} + O_p\left(\frac{1}{\log^3 T}\right) \\ &= -\frac{v_{\psi N}^2}{\sigma^2} \left\{ \frac{1}{L(t)^2 t^{2\alpha}} - \frac{1}{L(T)^2 T^{2\alpha}} \frac{1-r^{1-2\alpha}}{(1-r)(1-2\alpha)} \right\} \\ &\quad + \left\{ \frac{v_{\psi N}^2}{\sigma^2} \frac{r(\alpha)}{L(T)^2 T^{2\alpha}} - \frac{g(r)}{\log^2 T} \right\} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\} + O_p\left(\frac{1}{\log^3 T}\right), \end{aligned}$$



using (107) and because, from (104),

$$\begin{aligned}
 (112) \quad & \frac{1}{T - [Tr]} \sum_{t=[Tr]}^T \epsilon_t \\
 &= \frac{v_{\psi N}^2}{\sigma^2} \frac{1}{T - [Tr]} \sum_{t=[Tr]}^T \frac{1}{L(t)^2 t^{2\alpha}} + O_p\left(\frac{1}{\sqrt{N}L(T)T^\alpha}\right) \\
 &= \frac{v_{\psi N}^2}{\sigma^2} \frac{1}{L(T)^2 T^{1+2\alpha}} \frac{1}{1-r} \sum_{t=[Tr]}^T \left(\frac{t}{T}\right)^{-2\alpha} + O_p\left(\frac{1}{\sqrt{N}L(T)T^\alpha}\right) \\
 &= \frac{v_{\psi N}^2}{\sigma^2} \frac{1}{L(T)^2 T^{2\alpha}} \frac{1}{1-r} \int_r^1 s^{-2\alpha} ds + O_p\left(\frac{1}{\sqrt{N}L(T)T^\alpha}\right) \\
 &= \frac{v_{\psi N}^2}{\sigma^2} \frac{1}{L(T)^2 T^{2\alpha}} \frac{1-r^{1-2\alpha}}{(1-r)(1-2\alpha)} + O_p\left(\frac{1}{\sqrt{N}L(T)T^\alpha}\right).
 \end{aligned}$$

In view of (111), we have, for  $|l| \leq M$ ,  $\frac{M}{T} \rightarrow 0$ , and  $t \geq [Tr]$  with  $r > 0$ ,

$$\begin{aligned}
 \hat{u}_{t+l} &= -(\epsilon_{t+l} - \bar{\epsilon}) - \hat{b} \left\{ \log \frac{t+l}{T} - \overline{\log \frac{t}{T}} \right\} \\
 &\quad - 2\{\log L(t+l) - \overline{\log L(t)}\} \\
 &= -\frac{v_{\psi N}^2}{\sigma^2} \left\{ \frac{1}{L(t)^2 t^{2\alpha}} [1 + o(1)] - \frac{1}{L(T)^2 T^{2\alpha}} \frac{1-r^{1-2\alpha}}{(1-r)(1-2\alpha)} \right\} \\
 &\quad + \left\{ \frac{v_{\psi N}^2}{\sigma^2} \frac{r(\alpha)}{L(T)^2 T^{2\alpha}} - \frac{g(r)}{\log^2 T} \right\} \left\{ \log \frac{t}{T} [1 + o(1)] - \overline{\log \frac{t}{T}} \right\} \\
 &\quad + O_p\left(\frac{\log L(T)}{\sqrt{N}L(T)T^\alpha}\right) + O_p\left(\frac{1}{\log^3 T}\right) \\
 &= \hat{u}_t \{1 + o(1)\}.
 \end{aligned}$$

Then

$$\begin{aligned}
 (113) \quad & \frac{1}{T - [Tr]} \sum_{[Tr] \leq t, t+l \leq T} \hat{u}_t \hat{u}_{t+l} \\
 &= \frac{1}{T - [Tr]} \sum_{[Tr] \leq t \leq T} \hat{u}_t^2 \{1 + o(1)\}
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{T - [Tr]} \frac{v_{\psi N}^4}{\sigma^4} \sum_{[Tr] \leq t \leq T} \left\{ \frac{1}{L(t)^2 t^{2\alpha}} - \frac{1}{L(T)^2 T^{2\alpha}} \frac{1 - r^{1-2\alpha}}{(1-r)(1-2\alpha)} \right\}^2 \\
 &\quad + \frac{1}{T - [Tr]} \left\{ \frac{v_{\psi N}^2}{\sigma^2} \frac{r(\alpha)}{L(T)^2 T^{2\alpha}} - \frac{g(r)}{\log^2 T} \right\}^2 \sum_{[Tr] \leq t \leq T} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\}^2 \\
 &\quad + \frac{2}{T - [Tr]} \frac{v_{\psi N}^2}{\sigma^2} \left\{ \frac{v_{\psi N}^2}{\sigma^2} \frac{r(\alpha)}{L(T)^2 T^{2\alpha}} - \frac{g(r)}{\log^2 T} \right\} \sum_{[Tr] \leq t \leq T} \frac{\log \frac{t}{T} - \overline{\log \frac{t}{T}}}{L(t)^2 t^{2\alpha}} \\
 &\quad + O_p \left( \frac{1}{\log^4 T} \right) \\
 &= O_p \left( \frac{1}{\log^4 T} \right),
 \end{aligned}$$

uniformly in  $l$  when  $L(T) = \log T$  and  $\alpha \geq 0$ . Hence, (110) becomes

$$\begin{aligned}
 \widehat{\text{Ivar}}_r(\sqrt{N}\hat{u}_t) &= \sum_{l=-M}^M \frac{N}{T - [Tr] + 1} \sum_{[Tr] \leq t, t+l \leq T} \hat{u}_t^2 \{1 + o(1)\} \\
 &= O_p \left( \frac{NM}{\log^4 T} \right)
 \end{aligned}$$

and so

$$\begin{aligned}
 (114) \quad s_b^2 &= \widehat{\text{Ivar}}_r(\hat{u}_t) \left[ \sum_{t=[Tr]}^T \tau_t^2 \right]^{-1} = O_p \left( \frac{M}{(\log^4 T)T} \right) \left[ \frac{1}{T} \sum_{t=[Tr]}^T \tau_t^2 \right]^{-1} \\
 &= O_p \left( \frac{M}{(\log^4 T)T} \right).
 \end{aligned}$$

Using (109) and (114), we see that the  $t$  ratio  $t_b$  has the asymptotic behavior, under the alternative,

$$\begin{aligned}
 (115) \quad t_b &= \frac{\hat{b}}{s_b} = -\frac{2}{\log T} \div O_p \left( \frac{M^{1/2}}{(\log^2 T)T^{1/2}} \right) \\
 &= -\frac{2}{\log T} \times O_p \left( \frac{(\log^2 T)T^{1/2}}{M^{1/2}} \right) \rightarrow -\infty
 \end{aligned}$$

for all  $\alpha \geq 0$  and all bandwidth choices  $M \leq T$ . It follows that the test is consistent. The divergence rate is  $O((\log T)T^{1/2}/M^{1/2})$ .

(b)  $\alpha < 0$ . We consider the case where  $\alpha < 0$  and  $\delta_i = \delta$  for all  $i$ . The case  $\alpha < 0$  and  $\delta_i \neq \delta$  for all  $i$  may be treated in the same way as case (a) and is therefore omitted. Set  $\gamma = -\alpha > 0$ .

When  $\delta_i = \delta$  for all  $i$ , (73) and (74) continue to hold but with  $\alpha < 0$  and

$$(116) \quad \epsilon_t = \log \left[ 1 + N^{-1/2} \frac{\delta^2}{v_{\psi N}^2} \eta_{Nt} \right] - \log \{ 1 + L(t)^{-2} t^{-2\alpha} \psi_t^2 / \delta^2 + 2L(t)^{-1} t^{-\alpha} \psi_t / \delta \} + O_p(N^{-1}).$$

If  $T^\gamma / \sqrt{N} \rightarrow 0$ , then both logarithmic terms of  $\epsilon_t$  may still be expanded as  $T, N \rightarrow \infty$ , but now the second term dominates rather than the first. Thus, in place of (75), we get

$$(117) \quad u_t = -\epsilon_t = \frac{2}{\delta} \frac{t^\gamma}{L(t)} \psi_t + O_p \left( \frac{T^{2\gamma}}{L(t)^2 N} + \frac{1}{\sqrt{N}} \right),$$

where

$$\psi_t = \frac{1}{N} \sum_{i=1}^N \psi_{it} = \frac{1}{N} \sum_{i=1}^N \sigma_i \xi_{it} = O_p(N^{-1/2}).$$

The limit theory proceeds as in the proof of Theorem 1, but we now have

$$\begin{aligned} & \frac{\sqrt{NT}L(T)}{T^\gamma} (\hat{b} - b) \\ &= \frac{2}{\delta} \frac{(\sqrt{NT}L(T)) / (\sqrt{T}T^\gamma) \sum_{t=[Tr]}^T \tau_t(t^\gamma/L(t)) \psi_t}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} + o_p(1) \\ &= \frac{2}{\delta} \frac{\frac{1}{\sqrt{NT}} \sum_{t=[Tr]}^T \sum_{i=1}^N \tau_t(\frac{t}{T})^\gamma \sigma_i \xi_{it}}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} + o_p(1) \\ &\Rightarrow \frac{2}{\delta} N \left( 0, \frac{\omega_\xi^2 \int_r^1 \{ \log s - \frac{1}{1-r} \int_r^1 \log p dp \}^2 s^{2\gamma} ds}{(1-r) - (\frac{r}{1-r}) \log^2 r} \right) \end{aligned}$$

using Assumptions A2 and A3, where  $\omega_\xi^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2 \omega_{ii}$  and where

$$\frac{1}{T} \sum_{t=[Tr]}^T \tau_t \left( \frac{t}{T} \right)^{2\gamma} \rightarrow \int_r^1 \left\{ \log s - \frac{1}{1-r} \int_r^1 \log p dp \right\}^2 s^{2\gamma} ds.$$

Thus, if  $T^{\gamma-1/2} / (\sqrt{NT}L(T)) \rightarrow 0$ ,  $\hat{b}$  is still consistent, but at a reduced rate in comparison with the null and provided  $\gamma = -\alpha$  is not too large.

The behavior of the estimated standard error can be obtained in a similar manner to the derivation under the null, given above. In particular, in view of (117),

$$\begin{aligned} E(u_t u_{t+l}) &= \frac{4}{N^2 \delta^2} \sum_{i=1}^N \sigma_i^2 \frac{t^\gamma (t+l)^\gamma E(\xi_{it} \xi_{it+l})}{L(t)L(t+l)} \\ &= \frac{4}{N^2 \delta^2} \sum_{i=1}^N \sigma_i^2 \frac{t^{2\gamma} E(\xi_{it} \xi_{it+l})}{L(t)^2 t^{2\alpha}} \left(1 + O\left(\frac{M}{T}\right)\right), \end{aligned}$$

where  $|l| < M$  and  $\frac{M}{T} \rightarrow 0$  as  $T \rightarrow \infty$ . Since  $\omega_i^2 = \sum_{l=-\infty}^{\infty} E(\xi_{it} \xi_{it+l})$ , we have

$$\begin{aligned} \sum_{l=-M}^M E(u_t u_{t+l}) &= \frac{4}{N^2 \delta^2} \sum_{i=1}^N \sigma_i^2 \sum_{l=-M}^M \frac{t^{2\gamma} E(\xi_{it} \xi_{it+l})}{L(t)^2 (1 + O(\frac{M}{T}))} \\ &\sim \frac{4}{N^2 \delta^2} \sum_{i=1}^N \sigma_i^2 \frac{t^{2\gamma} \omega_{ii}^2}{L(t)^2} (1 + o(1)) \\ &= \frac{4}{\delta^2} \frac{t^{2\gamma}}{NL(t)^2} \omega_\xi^2 (1 + o(1)). \end{aligned}$$

The sample quantity is

$$\begin{aligned} &\frac{1}{T - [Tr]} \sum_{[Tr] \leq t, t+l \leq T} u_t u_{t+l} \\ &= \frac{4}{N^2 \delta^2} \sum_{i=1}^N \sigma_i^2 \frac{1}{T - [Tr]} \sum_{[Tr] \leq t, t+l \leq T} \frac{t^{2\gamma} \xi_{it} \xi_{it+l}}{L(t)^2} (1 + o(1)) \\ &= \frac{4}{N^2 \delta^2} \sum_{i=1}^N \sigma_i^2 \left\{ \frac{1}{T - [Tr]} \sum_{[Tr] \leq t, t+l \leq T} \frac{t^{2\gamma} E(\xi_{it} \xi_{it+l})}{L(t)^2} (1 + o(1)) \right. \\ &\quad \left. + \frac{1}{T - [Tr]} \sum_{[Tr] \leq t, t+l \leq T} \frac{t^{2\gamma} (\xi_{it} \xi_{it+l} - E \xi_{it} \xi_{it+l})}{L(t)^2} (1 + o(1)) \right\} \\ &= \frac{4}{N^2 \delta^2} \sum_{i=1}^N \sigma_i^2 E(\xi_{it} \xi_{it+l}) \left( \frac{1}{T - [Tr]} \sum_{Tr}^T \frac{1}{L(t)^2 t^{2\alpha}} \right) \{1 + o_p(1)\} \\ &= \frac{4}{N^2 \delta^2} \sum_{i=1}^N \sigma_i^2 E(\xi_{it} \xi_{it+l}) \frac{1}{T - [Tr]} \frac{T^{2\gamma}}{L(T)^2} \sum_{Tr}^T \frac{(\frac{t}{T})^{2\gamma}}{(L(t)^2)/(L(T)^2)} \\ &\quad \times \{1 + o_p(1)\} \end{aligned}$$

$$\begin{aligned}
 &= \frac{4}{N^2 \delta^2} \frac{T^{2\gamma}}{L(T)^2} \frac{\int_r^1 s^{2\gamma} ds}{1-r} \sum_{i=1}^N \sigma_i^2 E(\xi_{it} \xi_{it+l}) \{1 + o_p(1)\} \\
 &= \frac{4}{N^2 \delta^2} \frac{T^{2\gamma}}{L(T)^2} \frac{1-r^{1+2\gamma}}{(1+2\gamma)(1-r)} \sum_{i=1}^N \sigma_i^2 E(\xi_{it} \xi_{it+l}) \{1 + o_p(1)\},
 \end{aligned}$$

which gives

$$\begin{aligned}
 &\sum_{l=-M}^M \frac{1}{T - [Tr]} \sum_{[Tr] \leq t, t+l \leq T} u_t u_{t+l} \\
 &= \frac{4}{N^2 \delta^2} \frac{T^{2\gamma}}{L(T)^2} \frac{1-r^{1+2\gamma}}{(1+2\gamma)(1-r)} \sum_{i=1}^N \sigma_i^2 \sum_{l=-M}^M E(\xi_{it} \xi_{it+l}) \{1 + o_p(1)\} \\
 &= \frac{4}{N^2 \delta^2} \frac{T^{2\gamma}}{L(T)^2} \frac{1-r^{1+2\gamma}}{(1+2\gamma)(1-r)} \sum_{i=1}^N \sigma_i^2 \omega_{ii}^2 \{1 + o_p(1)\} \\
 &= \frac{4}{N \delta^2} \frac{T^{2\gamma}}{L(T)^2} \frac{1-r^{1+2\gamma}}{(1+2\gamma)(1-r)} \omega_\xi^2 (1 + o_p(1)).
 \end{aligned}$$

Similarly, we find that

$$\begin{aligned}
 \widehat{\text{lvar}}_r(\sqrt{N} \hat{u}_t) &= \sum_{l=-M}^M \frac{N}{T - [Tr] + 1} \sum_{[Tr] \leq t, t+l \leq T} \hat{u}_t \hat{u}_{t+l} \\
 &= \frac{4}{\delta^2} \frac{T^{2\gamma}}{L(T)^2} \frac{1-r^{1+2\gamma}}{(1+2\gamma)(1-r)} \omega_\xi^2 (1 + o_p(1)),
 \end{aligned}$$

and then

$$\begin{aligned}
 s_b^2 &= \widehat{\text{lvar}}_r(\hat{u}_t) \left[ \sum_{t=[Tr]}^T \tau_t^2 \right]^{-1} \\
 &= \frac{1}{NT} \widehat{\text{lvar}}_r(\sqrt{N} \hat{u}_t) \left[ \frac{1}{T} \sum_{t=[Tr]}^T \tau_t^2 \right]^{-1} \\
 &= \frac{4}{NT \delta^2} \frac{T^{2\gamma}}{L(T)^2} \frac{1-r^{1+2\gamma}}{(1+2\gamma)(1-r)} \omega_\xi^2 \left\{ (1-r) - \left( \frac{r}{1-r} \right) \log^2 r \right\}^{-1} \\
 &\quad \times (1 + o_p(1)).
 \end{aligned}$$

It follows that

$$t_{\hat{b}} = \frac{\hat{b}}{s_{\hat{b}}} = \frac{(\hat{b} - b)}{s_{\hat{b}}} + \frac{b}{s_{\hat{b}}} = \frac{b}{s_{\hat{b}}} + O_p(1)$$

and so, under the alternative with  $b = 2\alpha < 0$ , we have

$$\begin{aligned} \frac{b}{s_{\hat{b}}} &= \left( 2\alpha \left\{ (1-r) - \left( \frac{r}{1-r} \right) \log^2 r \right\}^{1/2} \right) \\ &\quad / \left\{ \frac{4}{N^2 T \delta^2} \frac{T^{2\gamma}}{L(T)^2} \frac{1-r^{1+2\gamma}}{(1+2\gamma)(1-r)} \omega_{\xi}^2 \right\}^{1/2} \rightarrow -\infty, \end{aligned}$$

confirming consistency of the test in this case. The divergence rate is  $O(L(T) \times T^{1/2-\gamma N})$ .

(c) Finally, consider the case where  $\gamma = -\alpha$  is such that  $T^\gamma / (\sqrt{N}L(T)) \rightarrow \infty$ . Again, (73) holds so that

$$(118) \quad \log \frac{H_1}{H_t} - 2 \log L(t) = a + b \log t + u_t$$

and the second term of (116) dominates, but now, for  $t \geq [Tr]$ , we have

$$\begin{aligned} u_t = -\epsilon_t &= \log \left\{ 1 + \frac{L(t)^{-2} t^{2\gamma} \psi_t^2}{\delta^2} + \frac{2L(t)^{-1} t^\gamma \psi_t}{\delta} \right\} + O_p(N^{-1/2}) \\ &= \log \left\{ \frac{t^{2\gamma}}{L(t)^2 N} \frac{(\sqrt{N} \psi_t)^2}{\delta^2} \right\} + O_p \left( \frac{\sqrt{N} L(T)}{T^\gamma} + \frac{1}{\sqrt{N}} \right) \\ &= -2 \log L(t) + 2\gamma \log t - \log N - \log \delta^2 + \log (\sqrt{N} \psi_t)^2 + o_p(1) \\ &= -2 \log L(t) + 2\gamma \log t - \log N - \log \delta^2 + \log \{ \xi_{\psi t} + o_p(N^{-1/2}) \}^2 \\ &\quad + o_p(1) \\ &= -2 \log L(t) + 2\gamma \log t + A_N + \xi_{wt} + o_p(1), \end{aligned}$$

where  $\xi_{wt} = \log \xi_{\psi t}^2 - E\{\log \xi_{\psi t}^2\}$  and  $A_N = E\{\xi_{wt}\} - \log N - \log \delta^2$ . Hence, (118) is equivalent to

$$(119) \quad \log \frac{H_1}{H_t} - 2 \log L(t) = a_N + w_t, \quad w_t = -2 \log L(t) + \xi_{wt} + o_p(1),$$

where  $a_N = a + A_N$  and the term in  $\log t$  drops out because  $2\gamma = -2\alpha = -b$ . The error ( $w_t$ ) in equation (119) therefore diverges to  $-\infty$  as  $T \rightarrow \infty$  and  $a_N, A_N = O(-\log N) \rightarrow -\infty$  as  $N \rightarrow \infty$ . This behavior is consistent with the fact (easily deduced from (61)) that  $H_t = O_p(N)$  in this case.

In view of (119), the fitted regression (98) behaves like a regression of  $-2\log L(t)$  on  $\log t$ , so that just as in case (a) and (109) above, we have

$$\begin{aligned} \hat{b} &= \frac{\sum_{t=[Tr]}^T \tau_t \{-2\log L(t) + \xi_{wt} + o_p(1)\}}{\sum_{t=[Tr]}^T \tau_t^2} \\ &= -\frac{2}{\log T} + O\left(\frac{1}{\log^2 T}\right). \end{aligned}$$

So  $\hat{b} \rightarrow_p 0$ , as in case (a).

Next consider the standard error. When  $L(t) = \log t$ ,

$$\log L(t) - \overline{\log L(t)} = \frac{\log \frac{t}{T} - \overline{\log \frac{t}{T}}}{\log T} + O\left(\frac{1}{\log^2 T}\right)$$

from (101), so that

$$\begin{aligned} \hat{u}_t &= -(\xi_{wt} - \bar{\xi}_w) - 2\{\log L(t) - \overline{\log L(t)}\} - \hat{b} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\} \\ &= -(\xi_{wt} - \bar{\xi}_w) - 2\frac{\log \frac{t}{T} - \overline{\log \frac{t}{T}}}{\log T} + \frac{2}{\log T} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\} \\ &\quad + O\left(\frac{1}{\log^2 T}\right) \\ &= -(\xi_{wt} - \bar{\xi}_w) + O\left(\frac{1}{\log^2 T}\right). \end{aligned}$$

Assuming the long run variance of  $\xi_{wt}$  exists and writing  $\omega_{\xi_w}^2 = \sum_{k=-\infty}^{\infty} E(\xi_{wt} \times \xi_{wt+k})$ , we have

$$\widehat{\text{Ivar}}_r(\hat{u}_t) = \sum_{l=-M}^M \frac{1}{T - [Tr] + 1} \sum_{[Tr] \leq t, t+l \leq T} \hat{u}_t \hat{u}_{t+l} = \omega_{\xi_w}^2 \{1 + o_p(1)\}$$

and then

$$\begin{aligned} s_{\hat{b}}^2 &= \widehat{\text{Ivar}}_r(\hat{u}_t) \left[ \sum_{t=[Tr]}^T \tau_t^2 \right]^{-1} = \frac{\omega_{\xi_w}^2}{T} \left[ \frac{1}{T} \sum_{t=[Tr]}^T \tau_t^2 \right]^{-1} \{1 + o_p(1)\} \\ &= \frac{\omega_{\xi_w}^2}{T} \left\{ (1-r) - \left( \frac{r}{1-r} \right) \log^2 r \right\}^{-1}. \end{aligned}$$

Hence

$$\begin{aligned}
 t_{\hat{b}} &= \frac{\hat{b}}{s_{\hat{b}}} \\
 &= \left\{ -\frac{2}{\log T} + O\left(\frac{1}{\log^2 T}\right) \right\} \\
 &\quad \div \left\{ \frac{\omega_{\xi_\omega}^2}{T} \left\{ (1-r) - \left(\frac{r}{1-r}\right) \log^2 r \right\}^{-1} \right\}^{1/2} \\
 &= -\frac{2\sqrt{T} \left\{ (1-r) - \left(\frac{r}{1-r}\right) \log^2 r \right\}^{1/2}}{\log T \omega_{\xi_\omega}} \{1 + o(1)\} \rightarrow -\infty
 \end{aligned}$$

and again the test is consistent. The divergence rate is  $O(T^{1/2}/\log T)$ .

B.6. Proof of Theorem 3

(a) Under the local alternative (41), we have

$$(120) \quad \delta_i \sim \text{iid}(\delta, c^2 T^{-2\omega}) \quad \text{for } \alpha \geq \omega > 0.$$

Under this alternative the DGP for  $\log(H_1/H_t)$  has the same form as in case (a) in the proof of Theorem 2, but with  $\bar{\delta} = \delta$  and

$$(121) \quad \sigma^2 = N^{-1} \sum (\delta_i - \delta)^2 = c^2 T^{-2\omega} \{1 + O_p(N^{-1/2})\} = O_p\left(\frac{1}{T^{2\omega}}\right),$$

$$\begin{aligned}
 (122) \quad \sigma_{\delta\psi_t} &= N^{-1} \sum_{i=1}^N (\delta_i - \delta)(\psi_{it} - \psi_t) = N^{-1/2} T^{-\omega} s_{TNt} \{1 + o_p(1)\} \\
 &= O_p(N^{-1/2} T^{-\omega})
 \end{aligned}$$

as  $N, T \rightarrow \infty$ . Thus, as in (94), we have

$$\begin{aligned}
 (123) \quad H_t &= \frac{1}{\delta^2} \frac{\sigma^2 + \sigma_{\psi_t}^2 L(t)^{-2} t^{-2\alpha} + 2(\sigma_{\delta\psi_t})L(t)^{-1} t^{-\alpha}}{1 + L(t)^{-2} t^{-2\alpha} \psi_t^2 / \delta^2 + 2L(t)^{-1} t^{-\alpha} \psi_t / \delta} \\
 &= \frac{1}{\delta^2} \frac{\sigma^2 + \sigma_{\psi_t}^2 L(t)^{-2} t^{-2\alpha} + 2(\sigma_{\delta\psi_t})L(t)^{-1} t^{-\alpha}}{1 + L(t)^{-2} t^{-2\alpha} \psi_t^2 / \delta^2 + 2L(t)^{-1} t^{-\alpha} \psi_t / \delta} \\
 &= O_p(T^{-2\omega}),
 \end{aligned}$$

since  $\omega \leq \alpha$ . It follows that  $H_t \rightarrow_p 0$  as  $T \rightarrow \infty$  for  $t \geq [Tr]$  and  $r > 0$ , as the model leads to behavior in  $H_t$  local to that under the null hypothesis. More



explicitly, we have, using (97), (96), and (123),

$$(124) \quad \log \frac{H_1}{H_t} - 2 \log L(t) = \log H_1 - 2 \log \frac{\sigma}{\delta} - 2 \log L(t) - \epsilon_t$$

$$= \log H_1 - 2 \log \frac{c}{\delta} + 2\omega \log T - 2 \log L(t) - \epsilon_t,$$

where

$$(125) \quad \epsilon_t = \log \left\{ 1 + \left( \frac{\sigma_{\psi_t}^2}{\sigma^2} \right) L(t)^{-2} t^{-2\alpha} + 2 \left( \frac{\sigma_{\delta\psi_t}}{\sigma^2} \right) L(t)^{-1} t^{-\alpha} \right\}$$

$$- \log \left\{ 1 + \frac{L(t)^{-2} t^{-2\alpha} \psi_t^2}{\bar{\delta}^2} + \frac{2L(t)^{-1} t^{-\alpha} \psi_t}{\bar{\delta}} \right\}$$

$$= \log \left\{ 1 + \left( \frac{\sigma_{\psi_t}^2}{c^2} \right) T^{2\omega} L(t)^{-2} t^{-2\alpha} + 2 \left( \frac{\sigma_{\delta\psi_t}}{c^2} \right) T^{2\omega} L(t)^{-1} t^{-\alpha} \right\}$$

$$- \log \left\{ 1 + \frac{L(t)^{-2} t^{-2\alpha} \psi_t^2}{\delta^2} + \frac{2L(t)^{-1} t^{-\alpha} \psi_t}{\delta} \right\}$$

$$= \left( \frac{\sigma_{\psi_t}^2}{c^2} \right) T^{2\omega} L(t)^{-2} t^{-2\alpha} + O_p \left( \frac{1}{L(T)^4 T^{4(\alpha-\omega)}} + \frac{1}{\sqrt{NL(T)} T^{2(\alpha-\omega)}} \right).$$

The fitted regression is again

$$(126) \quad \log \frac{H_1}{H_t} - 2 \log L(t) = \hat{a}^\dagger + \hat{b} \log \left( \frac{t}{T} \right) + \text{residual},$$

where now  $\hat{a}^\dagger = \log H_1 - 2 \log \frac{c}{\delta} + 2\omega \log T$  and, as in (103),

$$\hat{b} = - \frac{\sum_{t=[Tr]}^T \tau_t \{2 \log L(t) + \epsilon_t\}}{\sum_{t=[Tr]}^T \tau_t^2}$$

$$= - \frac{\sum_{t=[Tr]}^T \tau_t \epsilon_t}{\sum_{t=[Tr]}^T \tau_t^2} - \frac{2}{\log T} + \frac{g(r)}{\log^2 T} + O_p \left( \frac{1}{\log^3 T} \right).$$

Next,

$$(127) \quad \frac{1}{T} \sum_{t=[Tr]}^T \tau_t \epsilon_t = \frac{v_{\psi N}^2}{c^2} \frac{1}{L(T)^2 T^{1+2(\alpha-\omega)}} \sum_{t=[Tr]}^T \tau_t \frac{L(T)^2}{L(t)^2} \left( \frac{t}{T} \right)^{-2\alpha} \{1 + o_p(1)\}$$

$$= \frac{v_{\psi N}^2}{c^2} \frac{1}{L(T)^2 T^{2(\alpha-\omega)}}$$

$$\begin{aligned} & \times \int_r^1 \left\{ \log s - \frac{1}{1-r} \int_r^1 \log p \, dp \right\} s^{-2\alpha} \, ds \{1 + o_p(1)\} \\ & = \frac{v_{\psi N}^2}{c^2} \frac{r^*(\alpha)}{L(T)^2 T^{2(\alpha-\omega)}} \{1 + o_p(1)\}, \end{aligned}$$

where  $r^*(\alpha) = \int_r^1 \{ \log s - \frac{1}{1-r} \int_r^1 \log p \, dp \} s^{-2\alpha} \, ds$  is given in (106) above. We deduce that

$$(128) \quad \hat{b} = -\frac{2}{\log T} + \frac{g(r)}{\log^2 T} - \frac{v_{\psi N}^2}{c^2} \frac{r(\alpha)}{L(T)^2 T^{2(\alpha-\omega)}} + O_p\left(\frac{1}{\log^3 T}\right),$$

where  $r(\alpha)$  is given in (108), so that  $\hat{b} \rightarrow_p 0$  as  $T, N \rightarrow \infty$ . The result is therefore comparable to that under case (a) of Theorem 2.

Next consider the standard error of  $\hat{b}$  under the local alternative. Writing the residual in (126) as  $\hat{u}_t$ , the long run variance estimate has typical form

$$\widehat{\text{lvar}}_r(\sqrt{N}\hat{u}_t) = \sum_{l=-M}^M \frac{N}{T - [Tr] + 1} \sum_{[Tr] \leq t, t+l \leq T} \hat{u}_t \hat{u}_{t+l}.$$

In view of (125), (127), and (128), we have

$$\begin{aligned} \hat{u}_t &= -(\epsilon_t - \bar{\epsilon}) - 2\{\log L(t) - \overline{\log L(t)}\} - \hat{b} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\} \\ &= -(\epsilon_t - \bar{\epsilon}) - 2 \frac{\log \frac{t}{T} - \overline{\log \frac{t}{T}}}{\log T} - \hat{b} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\} + O_p\left(\frac{1}{\log^2 T}\right) \\ &= -\frac{v_{\psi N}^2}{c^2 T^{-2\omega}} \left\{ \frac{1}{L(t)^2 t^{2\alpha}} - \frac{1}{L(T)^2 T^{2\alpha}} \frac{1 - r^{1-2\alpha}}{(1-r)(1-2\alpha)} \right\} \\ &\quad + \left\{ \frac{v_{\psi N}^2}{c^2} \frac{r(\alpha)}{L(T)^2 T^{2(\alpha-\omega)}} - \frac{g(r)}{\log^2 T} \right\} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\} \\ &\quad + O_p\left(\frac{1}{\log^2 T}\right), \end{aligned}$$

using (112). Then, for  $|l| \leq M$ ,  $\frac{M}{T} \rightarrow 0$ , and  $t \geq [Tr]$  with  $r > 0$ , we have, as in case (a) of the proof of Theorem 2,

$$\begin{aligned} \hat{u}_{t+l} &= -(\epsilon_{t+l} - \bar{\epsilon}) - \hat{b} \left\{ \log \frac{t+l}{T} - \overline{\log \frac{t+l}{T}} \right\} - 2\{\log L(t) - \overline{\log L(t)}\} \\ &= -\frac{v_{\psi N}^2}{c^2 T^{-2\omega}} \left\{ \frac{1}{L(t)^2 t^{2\alpha}} [1 + o(1)] - \frac{1}{L(T)^2 T^{2\alpha}} \frac{1 - r^{1-2\alpha}}{(1-r)(1-2\alpha)} \right\} \end{aligned}$$

$$\begin{aligned}
 & + \left\{ \frac{v_{\psi N}^2}{c^2} \frac{r(\alpha)}{L(T)^2 T^{2\alpha}} - \frac{g(r)}{\log^2 T} \right\} \left\{ \log \frac{t}{T} [1 + o(1)] - \overline{\log \frac{t}{T}} \right\} \\
 & + O_p\left(\frac{1}{\sqrt{N}L(T)T^\alpha}\right) + O_p\left(\frac{1}{\log^2 T}\right) \\
 & = \hat{u}_t\{1 + o(1)\},
 \end{aligned}$$

so that, just as in (113), we find

$$\begin{aligned}
 & \frac{1}{T - [Tr]} \sum_{[Tr] \leq t, t+l \leq T} \hat{u}_t \hat{u}_{t+l} \\
 & = \frac{1}{T - [Tr]} \sum_{[Tr] \leq t \leq T} \hat{u}_t^2 \{1 + o(1)\} \\
 & = \frac{1}{T - [Tr]} \frac{v_{\psi N}^4}{c^4 T^{-4\omega}} \sum_{[Tr] \leq t \leq T} \left\{ \frac{1}{L(t)^2 t^{2\alpha}} - \frac{1}{L(T)^2 T^{2\alpha}} \frac{1 - r^{1-2\alpha}}{(1-r)(1-2\alpha)} \right\}^2 \\
 & + \frac{1}{T - [Tr]} \left\{ \frac{v_{\psi N}^2}{c^2} \frac{r(\alpha)}{L(T)^2 T^{2(\alpha-\omega)}} - \frac{g(r)}{\log^2 T} \right\}^2 \\
 & \times \sum_{[Tr] \leq t \leq T} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\}^2 \\
 & - \frac{2}{T - [Tr]} \frac{v_{\psi N}^2}{c^2 T^{-2\omega}} \left\{ \frac{v_{\psi N}^2}{c^2} \frac{r(\alpha)}{L(T)^2 T^{2(\alpha-\omega)}} - \frac{g(r)}{\log^2 T} \right\} \\
 & \times \sum_{[Tr] \leq t \leq T} \frac{\log \frac{t}{T} - \overline{\log \frac{t}{T}}}{L(t)^2 t^{2\alpha}} + O_p\left(\frac{1}{\log^4 T}\right) \\
 & = O_p\left(\frac{1}{\log^4 T}\right),
 \end{aligned}$$

uniformly in  $l$ , and when  $L(T) = \log T$  and  $\alpha \geq \omega > 0$ . The remainder of the proof follows that of case (a) in the proof of Theorem 2. In particular, we have

$$\begin{aligned}
 \widehat{\text{Ivar}}_r(\sqrt{N}\hat{u}_t) & = O_p\left(\frac{NM}{\log^4 T}\right), \\
 s_b^2 & = \widehat{\text{Ivar}}_r(\hat{u}_t) \left[ \sum_{t=[Tr]}^T \tau_t^2 \right]^{-1} = O_p\left(\frac{M}{(\log^4 T)T}\right),
 \end{aligned}$$

and

$$\begin{aligned} t_{\hat{b}} &= \frac{\hat{b}}{s_{\hat{b}}} = -\frac{2}{\log T} \div O_p\left(\frac{M^{1/2}}{(\log^2 T)T^{1/2}}\right) \\ &= -\frac{2}{\log T} \times O_p\left(\frac{(\log^2 T)T^{1/2}}{M^{1/2}}\right) \rightarrow -\infty \end{aligned}$$

for all  $\alpha \geq \omega > 0$  and all bandwidth choices  $M \leq T$ . Again, the divergence rate is  $O_p((\log T)T^{1/2}/M^{1/2})$ . Thus, the test is consistent against all local alternatives of the form (120) with  $\omega \leq \alpha$ .

(b) When  $\omega > \alpha$ , the alternative involves

$$(129) \quad \delta_i \sim \text{iid}(\delta, c^2 T^{-2\omega}) \quad \text{for } \omega > \alpha \geq 0,$$

so that  $\delta_i = \delta + O_p(T^{-\omega}) = \delta + o_p(T^{-\alpha})$  and the alternatives are closer to the null than in case (a). Now we have

$$\begin{aligned} \sigma^2 &= N^{-1} \sum_{i=1}^N (\delta_i - \delta)^2 = c^2 T^{-2\omega} \{1 + O_p(N^{-1/2})\} = o_p\left(\frac{1}{T^{2\alpha}}\right), \\ \sigma_{\delta\psi_t} &= N^{-1} \sum_{i=1}^N (\delta_i - \delta)(\psi_{it} - \psi_t) = N^{-1/2} T^{-\omega} s_{TNt} \{1 + o_p(1)\} \\ &= o_p(N^{-1/2} T^{-\alpha}), \end{aligned}$$

and (123) becomes

$$\begin{aligned} H_t &= \frac{1}{\delta^2} \frac{\sigma^2 + \sigma_{\psi_t}^2 L(t)^{-2} t^{-2\alpha} + 2(\sigma_{\delta\psi_t})L(t)^{-1} t^{-\alpha}}{1 + L(t)^{-2} t^{-2\alpha} \psi_t^2 / \delta^2 + 2L(t)^{-1} t^{-\alpha} \psi_t / \delta} \\ &= \frac{1}{\delta^2} \frac{\sigma_{\psi_t}^2 L(t)^{-2} t^{-2\alpha} + 2(\sigma_{\delta\psi_t})L(t)^{-1} t^{-\alpha} + c^2 T^{-2\omega} \{1 + O_p(N^{-1/2})\}}{1 + L(t)^{-2} t^{-2\alpha} \psi_t^2 / \delta^2 + 2L(t)^{-1} t^{-\alpha} \psi_t / \delta} \\ &= \frac{1}{L(t)^2 t^{2\alpha}} \frac{1}{\delta^2} \\ &\quad \times \frac{\sigma_{\psi_t}^2 + c^2 T^{-2\omega} t^{2\alpha} L(t)^2 \{1 + O_p(N^{-1/2})\} + 2(\sigma_{\delta\psi_t})L(t)t^\alpha}{1 + L(t)^{-2} t^{-2\alpha} \psi_t^2 / \delta^2 + 2L(t)^{-1} t^{-\alpha} \psi_t / \delta}, \end{aligned}$$

so that the behavior of  $H_t$  is asymptotically the same as under the null (cf. (66)). Taking logarithms, we have

$$\log H_t = -2 \log L(t) - 2\alpha \log t$$

$$\begin{aligned}
 & + \log \left\{ \frac{1}{\delta^2} \left[ v_{\psi_N}^2 + N^{-1/2} \eta_{Nt} + O_p(N^{-1}) \right. \right. \\
 & \left. \left. + c^2 T^{-2\omega} t^{2\alpha} L(t)^2 \{1 + O_p(N^{-1/2})\} + O_p\left(\frac{L(T)}{N^{1/2} T^{(\omega-\alpha)}}\right) \right] \right\} \\
 & - \log \left\{ 1 + \frac{L(t)^{-2} t^{-2\alpha} \psi_t^2}{\delta^2} + \frac{2L(t)^{-1} t^{-\alpha} \psi_t}{\delta} \right\} \\
 & = -2 \log L(t) - 2\alpha \log t + 2 \log \frac{v_{\psi_N}}{\delta} + \epsilon_t,
 \end{aligned}$$

where

$$\begin{aligned}
 \epsilon_t = \log & \left[ 1 + N^{-1/2} \frac{\eta_{Nt}}{v_{\psi_N}^2} + \frac{c^2}{v_{\psi_N}^2} T^{-2\omega} t^{2\alpha} L(t)^2 \{1 + O_p(N^{-1/2})\} \right. \\
 & \left. + O_p\left(\frac{L(T)}{N^{1/2} T^{(\omega-\alpha)}}\right) \right] \\
 & - \log \left\{ 1 + \frac{L(t)^{-2} t^{-2\alpha} \psi_t^2}{\delta^2} + \frac{2L(t)^{-1} t^{-\alpha} \psi_t}{\delta} \right\} + O_p(N^{-1}).
 \end{aligned}$$

These formulae lead to the empirical regression equation

$$\log \frac{H_1}{H_t} - 2 \log L(t) = a + b \log t + u_t,$$

where  $a = h_1 - 2 \log(v_{\psi_N}/\delta)$ ,  $b = 2\alpha$ , and  $u_t = -\epsilon_t$ , just as in (73) and (74), but now

$$\begin{aligned}
 (130) \quad u_t = & -\frac{1}{\sqrt{N}} \frac{\eta_{Nt}}{v_{\psi_N}^2} - \frac{c^2}{v_{\psi_N}^2} T^{-2\omega} t^{2\alpha} L(t)^2 \{1 + O_p(N^{-1/2})\} \\
 & + \frac{2}{\delta} \frac{1}{t^\alpha L(t)} \psi_t + O_p\left(\frac{1}{L(T)^2 T^{2\alpha} N} + \frac{L(T)}{N^{1/2} T^{(\omega-\alpha)}} + \frac{1}{N}\right) \\
 = & -\frac{c^2}{v_{\psi_N}^2} \frac{t^{2\alpha}}{T^{2\omega}} L(t)^2 \{1 + O_p(N^{-1/2})\} - \frac{1}{\sqrt{N}} \frac{\eta_{Nt}}{v_{\psi_N}^2} \\
 & + O_p\left(\frac{1}{N^{1/2} L(T) T^\alpha} + \frac{L(T)}{N^{1/2} T^{(\omega-\alpha)}}\right).
 \end{aligned}$$

The first term of (130) dominates the behavior of the regression error  $u_t$  when  $\omega > \alpha \geq 0$  and  $T^{2(\omega-\alpha)}/(\sqrt{N}L(T)^2) \rightarrow 0$  as  $T, N \rightarrow \infty$ . It follows that

$$\hat{b} - b = \frac{\sum_{t=[Tr]}^T \tau_t u_t}{\sum_{t=[Tr]}^T \tau_t^2}$$

$$\begin{aligned}
 &= -\frac{c^2}{v_{\psi N}^2} \frac{1}{T^{2\omega}} \frac{\frac{1}{\sqrt{NT}} \sum_{t=[Tr]}^T \tau_t t^{2\alpha} L(t)^2}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} - \frac{1}{v_{\psi N}^2} \frac{\frac{1}{\sqrt{NT}} \sum_{t=[Tr]}^T \tau_t \eta_{Nt}}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} \\
 &\quad + \frac{2}{\delta} \frac{\frac{1}{T} \sum_{t=[Tr]}^T \tau_t \psi_t / (t^\alpha L(t))}{T^{-1} \sum_{t=[Tr]}^T \tau_t^2} + O_p\left(\frac{1}{L(t)^2 T^{2\alpha} N}\right),
 \end{aligned}$$

and since

$$\begin{aligned}
 &\frac{1}{T} \sum_{t=[Tr]}^T \tau_t t^{2\alpha} L(t)^2 \\
 &= \frac{L(T)^2}{T^{1-2\alpha}} \sum_{t=[Tr]}^T \tau_t \left(\frac{t}{T}\right)^{2\alpha} \{1 + o(1)\} \\
 &= T^{2\alpha} L(T)^2 \int_r^1 \left\{ \log s - \frac{1}{1-r} \int_r^1 \log p \, dp \right\} s^{2\alpha} \, ds \{1 + o(1)\},
 \end{aligned}$$

we have

$$(131) \quad \hat{b} - b = -\frac{c^2}{v_{\psi N}^2} h(r) \frac{L(T)^2}{T^{2(\omega-\alpha)}} \{1 + o(1)\},$$

where

$$h(r) = \frac{\int_r^1 \left\{ \log s - \frac{1}{1-r} \int_r^1 \log p \, dp \right\} s^{2\alpha} \, ds}{(1-r) - \left(\frac{r}{1-r}\right) \log^2 r}.$$

Since  $\omega > \alpha$ ,  $\hat{b} \rightarrow_p b$  and  $\hat{b}$  is consistent.

Next, the regression residual has the form

$$\begin{aligned}
 \hat{u}_t &= -(\epsilon_t - \bar{\epsilon}) - (\hat{b} - b) \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\} \\
 &= -\frac{c^2}{v_{\psi N}^2} L(T)^2 \left\{ \frac{t^{2\alpha}}{T^{2\omega}} - \frac{1}{T^{2(\omega-\alpha)}} \frac{1-r^{1+2\alpha}}{2\alpha+1} \right\} \{1 + O_p(N^{-1/2})\} \\
 &\quad + \left\{ \frac{c^2}{v_{\psi N}^2} \frac{h(r)L(T)^2}{T^{2(\omega-\alpha)}} \right\} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\},
 \end{aligned}$$

since

$$\frac{1}{T} \sum_{t=[Tr]}^T t^{2\alpha} = \frac{1}{T^{1-2\alpha}} \sum_{t=[Tr]}^T \left(\frac{t}{T}\right)^{2\alpha} = T^{2\alpha} \int_r^1 s^{2\alpha} \, ds \{1 + o(1)\}$$

$$= T^{2\alpha} \frac{1 - r^{1+2\alpha}}{2\alpha + 1}.$$

As earlier, we find  $\hat{u}_{t+l} = \hat{u}_t\{1 + o(1)\}$  for  $|l| \leq M$ , and then

$$\begin{aligned} & \frac{1}{T - [Tr]} \sum_{[Tr] \leq t, t+l \leq T} \hat{u}_t \hat{u}_{t+l} \\ &= \frac{1}{T - [Tr]} \sum_{[Tr] \leq t \leq T} \hat{u}_t^2 \{1 + o(1)\} \\ &= \frac{1}{T - [Tr]} \frac{c^4}{v_{\psi N}^4} L(T)^4 \sum_{[Tr] \leq t \leq T} \left\{ \frac{t^{2\alpha}}{T^{2\omega}} - \frac{1}{T^{2(\omega-\alpha)}} \frac{1 - r^{1+2\alpha}}{2\alpha + 1} \right\}^2 \\ & \quad \times \{1 + O_p(N^{-1/2})\} \\ & \quad + \frac{L(T)^4}{T - [Tr]} \left\{ \frac{c^2}{v_{\psi N}^2} \frac{h(r)}{T^{2(\omega-\alpha)}} \right\}^2 \sum_{[Tr] \leq t \leq T} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\}^2 \\ & \quad - \frac{2}{T - [Tr]} \frac{c^2}{v_{\psi N}^2} \frac{c^2}{v_{\psi N}^2} \frac{h(r)L(T)^4}{T^{4(\omega-\alpha)}} \sum_{[Tr] \leq t \leq T} \left\{ \log \frac{t}{T} - \overline{\log \frac{t}{T}} \right\} \left( \frac{t}{T} \right)^{2\alpha} \\ & \quad \times \{1 + O_p(N^{-1/2})\} \\ &= O_p \left( \frac{L(T)^4}{T^{4(\omega-\alpha)}} \right) \end{aligned}$$

uniformly in  $|l| \leq M$ . Hence,

$$\begin{aligned} \widehat{\text{ivar}}_r(\hat{u}_t) &= \sum_{l=-M}^M \frac{1}{T - [Tr] + 1} \sum_{[Tr] \leq t, t+l \leq T} \hat{u}_t^2 \{1 + o(1)\} \\ &= O_p \left( \frac{ML(T)^4}{T^{4(\omega-\alpha)}} \right) \end{aligned}$$

and so

$$\begin{aligned} s_{\hat{b}}^2 &= \widehat{\text{ivar}}_r(\hat{u}_t) \left[ \sum_{t=[Tr]}^T \tau_t^2 \right]^{-1} = O_p \left( \frac{ML(T)^4}{T^{4(\omega-\alpha)} T} \right) \left[ \frac{1}{T} \sum_{t=[Tr]}^T \tau_t^2 \right]^{-1} \\ &= O_p \left( \frac{ML(T)^4}{T^{1+4(\omega-\alpha)}} \right). \end{aligned}$$

It follows that

$$t_{\hat{b}} = \frac{\hat{b}}{s_{\hat{b}}} = \left[ b - \frac{c^2}{v_{\psi N}^2} h(r) \frac{L(T)^2}{T^{2(\omega-\alpha)}} \{1 + o(1)\} \right] \times O_p \left( \frac{T^{1/2+2(\omega-\alpha)}}{M^{1/2} L(T)^2} \right)$$

$$\rightarrow \begin{cases} \infty, & \text{for } b = 2\alpha > 0, \\ -\infty, & \text{for } b = 2\alpha = 0. \end{cases}$$

Thus, when  $\alpha > 0$ , the test has no power to detect alternatives of the form (129), whereas when  $\omega > \alpha = 0$ , the test is consistent. In both cases, the alternatives  $\delta_i \neq \delta$  are close to the null because  $\omega > \alpha$ , but when  $\alpha = 0$ , the rate of convergence of  $\delta_{it}$  is slow (at a slowly varying rate) and the test is therefore able to detect the local departures from the null.

### B.7. Power and the Choice of the $L(t)$ Function

This section provides a short discussion on the choice of the  $L(t)$  function. Since the class of possible  $L(t)$  functions is vast, it is convenient to consider the restricted class of logarithmic and higher order logarithmic functions

(132)  $L(t) = \log_k t$  for integer  $k \geq 1$ ,

where  $\log_1 t = \log t$ ,  $\log_2 t = \log(\log t)$ , and so on. Since our concern is with situations where  $t$  is large in the regression asymptotics,  $L(t)$  and  $L(t)^{-1}$  are both well defined. Note that  $k$  can be any positive integer, but we confine attention below to the primary cases of interest where  $k = 1, 2$ . Higher order cases can be deduced by recursion.

From (132), we can rewrite (100) and (102) as

$$\begin{aligned} \log L(t) &= \log L\left(T \frac{t}{T}\right) = \log_k \left\{ \log T + \log \frac{t}{T} \right\} \\ &= \log_k \left[ \log T \left\{ 1 + \frac{\log \frac{t}{T}}{\log T} \right\} \right] \\ &= \begin{cases} \log_2 T + \frac{\log \frac{t}{T}}{\log T} - \frac{1}{2} \frac{\log^2 \frac{t}{T}}{\log^2 T} + O\left(\frac{\log^3 \frac{t}{T}}{\log^3 T}\right), & k = 1, \\ \log \left[ \log_2 T + \frac{\log \frac{t}{T}}{\log T} - \frac{1}{2} \frac{\log^2 \frac{t}{T}}{\log^2 T} + O\left(\frac{\log^3 \frac{t}{T}}{\log^3 T}\right) \right], & k = 2, \end{cases} \\ &= \begin{cases} \log_2 T + \frac{\log \frac{t}{T}}{\log T} - \frac{1}{2} \frac{\log^2 \frac{t}{T}}{\log^2 T} + O\left(\frac{\log^3 \frac{t}{T}}{\log^3 T}\right), & k = 1, \\ \log_3 T + \frac{\log \frac{t}{T}}{\log T \log_2 T} + O\left(\frac{\log^2 \frac{t}{T}}{\log^2 T \log_2 T}\right), & k = 2, \end{cases} \end{aligned}$$



extending (101). Correspondingly, under the alternative  $\alpha \geq 0$  and  $\delta_i \sim \text{iid}(\delta, \sigma_\delta^2)$  considered in case (a) of Theorem 2, the choice of  $L(t)$  affects the bias formula in (109). Since

$$\frac{\sum_{t=[Tr]}^T \tau_t \log L(t)}{\sum_{t=[Tr]}^T \tau_t^2} = \begin{cases} \frac{1}{\log T} + O\left(\frac{1}{\log^2 T}\right), & k = 1, \\ \frac{1}{\log T \log_2 T} + O\left(\frac{1}{\log^2 T \log_2 T}\right), & k = 2, \end{cases}$$

we find that

$$\hat{b} = \begin{cases} -\frac{2}{\log T} + O_p\left(\frac{1}{\log^2 T}\right), & k = 1, \\ -\frac{2}{\log T \log_2 T} + O\left(\frac{1}{\log^2 T \log_2 T}\right), & k = 2. \end{cases}$$

Proceeding as in the proof of Theorem 2 (case (a)) we find that

$$s_b^2 = \begin{cases} O_p\left(\frac{M}{(\log^4 T)T}\right), & k = 1, \\ O_p\left(\frac{M}{(\log_2^4 T)T}\right), & k = 2, \end{cases}$$

and then

$$t_b = \frac{\hat{b}}{s_b} = \begin{cases} -\frac{2}{\log T} \times O_p\left(\frac{(\log^2 T)T^{1/2}}{M^{1/2}}\right), & k = 1, \\ -\frac{2}{\log T} \times O_p\left(\frac{(\log_2^2 T)T^{1/2}}{M^{1/2}}\right), & k = 2. \end{cases}$$

So, the divergence rate and discriminatory power of the  $\log t$  test reduce as we change  $L(t)$  from  $\log t$  to  $\log_2 t = \log \log t$ . The test is still consistent for  $k = 2$ , provided

$$\frac{M \log T}{T \log_2^2 T} \rightarrow 0.$$

APPENDIX C: ASYMPTOTIC PROPERTIES OF THE CLUSTERING PROCEDURE

Section 4 develops a clustering procedure based on augmenting a core panel with  $K$  individuals where  $\delta_i = \delta_A$  for  $i = 1, \dots, K$  with additional individuals one at a time for which  $\delta_{K+1} = \delta_B$ , say. This appendix provides an asymptotic

analysis of that procedure. We assume that the size of the core group  $K \rightarrow \infty$  as  $N \rightarrow \infty$ . The variation of the  $\delta_i$  is then

$$(133) \quad \sigma^2 = \frac{1}{K+1} \sum_{i=1}^{K+1} (\delta_i - \bar{\delta})^2 = \frac{K}{(K+1)^2} (\delta_A - \delta_B)^2 = O(K^{-1}),$$

where

$$\bar{\delta} = \frac{1}{K+1} \sum_{i=1}^{K+1} \delta_i = \frac{K\delta_A}{K+1} + \frac{\delta_B}{K+1} = \delta_A + O(K^{-1})$$

as in (42) and  $\sigma^2$  depends on  $K$ . More generally, we can consider a panel with idiosyncratic coefficients

$$(134) \quad \delta_i \sim \text{iid}(\delta, c^2 K^{-1}), \quad \text{where } \frac{K}{T^{2\alpha}} \rightarrow 0 \quad \text{and } \alpha > 0,$$

so that  $K$  is small relative to  $T$ . In this case,  $\sigma^2 = c^2 K^{-1}$ , analogous to (133). In the same way as in (124) and (125), under this alternative the DGP for  $\log(H_1/H_t)$  has the form

$$\log \frac{H_1}{H_t} - 2 \log L(t) = \log H_1 - 2 \log \frac{c}{\delta} + \log K - 2 \log L(t) - \epsilon_t,$$

where

$$\epsilon_t = \left( \frac{\sigma_{\psi_t}^2}{c^2} \right) K L(t)^{-2} t^{-2\alpha} + O_p \left( \frac{K^2}{L(T)^4 T^{4\alpha}} + \frac{K}{\sqrt{N} L(T) T^{2\alpha}} \right).$$

The fitted regression can now be written as

$$\log \frac{H_1}{H_t} - 2 \log L(t) = \hat{a}^\dagger + \hat{b} \log \left( \frac{t}{T} \right) + \hat{u}_t,$$

where  $\hat{a}^\dagger = \log H_1 - 2 \log \frac{c}{\delta} + \log K$  and, as in (128) but with  $K/T^{2\alpha} \rightarrow 0$ , we find that

$$\hat{b} = -\frac{2}{\log T} + \frac{g(r)}{\log^2 T} - \frac{v_{\psi N}^2}{c^2} \frac{r(\alpha)K}{L(T)^2 T^{2\alpha}} + O_p \left( \frac{1}{\log^3 T} \right).$$

Proceeding as in the proof of Theorem 3, we find that

$$s_b^2 = \widehat{\text{lvar}}_r(\hat{u}_t) \left[ \sum_{t=[Tr]}^T \tau_t^2 \right]^{-1} = O_p \left( \frac{M}{(\log^4 T) T} \right)$$

and

$$\begin{aligned}
 t_{\hat{b}} &= \frac{\hat{b}}{s_{\hat{b}}} = -\frac{2}{\log T} \div O_p\left(\frac{M^{1/2}}{(\log^2 T)T^{1/2}}\right) \\
 &= -\frac{2}{\log T} \times O_p\left(\frac{(\log^2 T)T^{1/2}}{M^{1/2}}\right) \rightarrow -\infty
 \end{aligned}$$

for all  $\alpha > 0$ , for  $K$  satisfying  $K/T^{2\alpha} \rightarrow 0$ , and all bandwidth choices  $M \leq T$ . The test is therefore consistent against local alternatives of the form (134). In view of (133), this includes the case where  $\delta_i = \delta_A$  for  $i = 1, \dots, K$  with  $\delta_{K+1} = \delta_B \neq \delta_A$ . On the other hand, when  $\delta_i = \delta_A$  for  $i = 1, \dots, K$ , the null hypothesis holds for  $N = K$  and  $t_{\hat{b}} = (\hat{b} - b)/s_{\hat{b}} \Rightarrow N(0, 1)$  as in Theorem 1.

When  $T^{2\alpha}/K \rightarrow 0$ , the alternatives (134) are very close to the null, relative to the convergence rate except when  $\alpha = 0$ . This case may be treated as in the proof of case (b) of Theorem 3. Accordingly, the test is inconsistent and unable to detect the departure from the null when  $\alpha > 0$ . However, when  $\alpha = 0$  and the convergence rate is slowly varying under the null, the test is consistent against local alternatives of the form (134) just as in case (b) of Theorem 3. In effect, although the alternatives are very close (because  $K$  is large), the convergence rate is so slow (slower than any power rate) and this suffices to ensure the test is consistent as  $T \rightarrow \infty$ .

APPENDIX D: DATA FOR THE COST OF LIVING INDEX EXAMPLE

*Data Source:* Bureau of Labor Statistics

*Data:* 19 U.S. Cities CPI

*Time Period:* 1918–2001 (84 annual observations)

*List of Cities:* New York (NYC), Philadelphia (PHI), Boston (BOS), Cleveland (CLE), Chicago (CHI), Detroit (DET), Washington, DC (WDC). Baltimore (BAL), Houston (HOU), Los Angeles (LAX), San Francisco (SFO), Seattle (SEA), Portland (POR), Cincinnati (CIN), Atlanta (ATL), St. Louis (STL), Minneapolis/St. Paul (MIN), Milwaukee (KCM)

REFERENCES

ADRIAN, T., AND F. FRANZONI (2005): “Learning about Beta: Time-Varying Factor Loadings, Expected Returns, and the Conditional CAPM,” Mimeo, Federal Reserve Bank of New York. [1777]

BAI, J. (2003): “Inferential Theory for Factor Models of Large Dimensions,” *Econometrica*, 71, 135–172. [1772]

——— (2004): “Estimation Cross-Section Common Stochastic Trends in Nonstationary Panel Data,” *Journal of Econometrics*, 122, 137–183. [1772]

BAI, J., AND S. NG (2002): “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191–221. [1772,1813]

- (2006): “Evaluating Latent and Observed Factors in Macroeconomics and Finance,” *Journal of Econometrics*, 131, 507–537. [1772]
- BROWNING, M., AND J. CARRO (2006): “Heterogeneity and Microeconometrics Modelling,” Mimeo, University of Copenhagen. [1772]
- BROWNING, M., L. P. HANSEN, AND J. J. HECKMAN (1999): “Micro Data and General Equilibrium Models,” in *Handbook of Macroeconomics*, Vol. 1A. New York: Elsevier, Chap. 8. [1772]
- CANOVA, F. (2004): “Testing for Convergence Clubs: A Predictive Density Approach,” *International Economic Review*, 45, 49–77. [1798]
- CANOVA, F., AND A. MARCET (1995): “The Poor Stay Poor: Non-Convergence Across Countries and Regions,” Discussion Paper 1265, CEPR. [1798]
- CHAMBERLAIN, G., AND M. ROTHSCILD (1983): “Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets,” *Econometrica*, 51, 1305–1324. [1772]
- CHOI, C., N. MARK, AND D. SUL (2006): “Endogenous Discounting, the World Savings Gluts and U.S. Current Account,” Mimeo, University of Auckland. [1773]
- CONNOR, G., AND R. A. KORAJCSYK (1986): “Performance Measurement with Arbitrage Pricing Theory: A New Framework for Analysis,” *Journal of Financial Economics*, 15, 373–394. [1772]
- (1988): “Risk and Return in Equilibrium APT: Application of a New Test Methodology,” *Journal of Financial Economics*, 21, 255–290. [1772]
- DURLAUF, S. N., AND P. A. JOHNSON (1995): “Multiple Regimes and Cross-Country Growth Behaviour,” *Journal of Applied Econometrics*, 10, 365–384. [1798]
- DURLAUF, S. N., AND D. T. QUAH (1999): “The New Empirics of Economic Growth,” in *Handbook of Macroeconomics*, Vol. 1A. New York: Elsevier, Chap. 4. [1771]
- DURLAUF, S. N., P. A. JOHNSON, AND J. TEMPLE (2005): “Growth Econometrics,” Mimeo, University of Wisconsin. [1771]
- ERMISCH, J. (2004): “Using Panel Data to Analyze Household and Family Dynamics,” in *Researching Social and Economic Change: The Uses of Household Panel Studies*, ed. by D. Rose. London and New York: Routledge, 230–249. [1771]
- FAMA, E. F., AND K. R. FRENCH (1993): “Common Risk Factors in the Returns on Stocks and Bonds,” *Journal of Financial Economics*, 33, 3–56. [1777]
- (1996): “Multifactor Explanations of Asset Pricing Anomalies,” *Journal of Finance*, 51, 55–84. [1777]
- GIVENEN, F. (2005): “An Empirical Investigation of Labor Income Processes,” Mimeo, University of Rochester. [1772]
- GUERRE, E. (2004): “Design-Adaptive Pointwise Nonparametric Regression Estimation for Recurrent Markov Time Series,” Mimeo, Université Pierre et Marie Curie. [1784]
- HODRICK, R., AND E. PRESCOTT (1997): “Post War Business Cycles: An Empirical Investigation,” *Journal of Money Credit and Banking*, 29, 1–16. [1783]
- HOWITT, P., AND D. MAYER-FOULKES (2005): “R&D, Implementation and Stagnation: A Schumpeterian Theory of Convergence Clubs,” *Journal of Money, Credit and Banking*, 37, 147–177. [1775]
- HSIAO, C. (2003): *Analysis of Panel Data* (Second Ed.). Cambridge, U.K.: Cambridge University Press. [1771]
- KARLSEN, H. A., AND D. TJØSTHEIM (2001): “Nonparametric Estimation in Null Recurrent Time Series,” *The Annals of Statistics*, 29, 372–416. [1784]
- KATZ, L. F., AND D. H. AUTOR (1999): “Changes in the Wage Structure and Earnings Inequality,” in *Handbook of Labor Economics*, Vol. 3A. New York: Elsevier, Chap. 26. [1776]
- KITAGAWA, G., AND W. GERSCH (1996): *Smoothness Priors Analysis of Time Series*. New York: Springer Verlag. [1783]
- KRUSELL, P., AND A. A. SMITH (1998): “Income and Wealth Heterogeneity in the Macroeconomy,” *Journal of Political Economy*, 106, 867–896. [1772]
- LUCAS, R. E. JR., AND N. L. STOKEY (1984): “Optimal Growth with Many Consumers,” *Journal of Economic Theory*, 32, 139–171. [1773]

- LUDVIGSON, S. C., AND S. NG (2007): "The Empirical Risk-Return Relation: A Factor Analysis Approach," *Journal of Financial Economics*, 83, 171–222. [1777]
- MENZLY, L., T. SANTOS, AND P. VERONESI (2002): "The Time Series of the Cross Section of Asset Prices," Working Paper 9217, NBER. [1777]
- MOFFITT, R. A., AND P. GOTTSCHALK (2002): "Trends in the Transitory Variance on Earnings in the United States," *Economic Journal*, 112, C68–C73. [1776]
- MOON, H. R., AND B. PERRON (2004): "Testing for a Unit Root in Panels with Dynamic Factors," *Journal of Econometrics*, 122, 81–126. [1772]
- OBSTFELD, M. (1990): "Intertemporal Dependence, Impatience, and Dynamics," *Journal of Monetary Economics*, 26, 45–75. [1773]
- PARENTE, S. L., AND E. C. PRESCOTT (1994): "Barriers to Technology Adoption and Development," *Journal of Political Economy*, 102, 298–321. [1775]
- PHILLIPS, P. C. B. (2005): "Challenges of Trending Time Series Econometrics," *Mathematics and Computers in Simulation*, 68, 401–416. [1784]
- (2007): "Regression with Slowly Varying Regressors and Nonlinear Trends," *Econometric Theory*, 23, 557–614. [1821]
- PHILLIPS, P. C. B., AND S. JIN (2002): "Limit Behavior of the Whittaker (HP) Filter," Mimeo, Yale University. [1785]
- PHILLIPS, P. C. B., AND H. R. MOON (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111. [1787,1818]
- PHILLIPS, P. C. B., AND J. PARK (1998): "Nonstationary Density Estimation and Kernel Autoregression," Cowles Foundation Discussion Paper 1181. [1784]
- PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *The Annals of Statistics*, 20, 971–1001. [1782,1816]
- PHILLIPS, P. C. B., AND D. SUL (2003): "The Elusive Empirical Shadow of Growth Convergence," Discussion Paper 1398, Cowles Foundation, Yale University.
- (2006): "Economic Transition and Growth," Mimeo, University of Auckland. [1772, 1775,1780]
- QUAH, D. (1996): "Empirics for Economic Growth and Convergence," *European Economic Review*, 40, 1353–1375. [1798]
- (1997): "Empirics for Growth and Distribution: Polarization, Stratification and Convergence Clubs," *Journal of Economic Growth*, 2, 27–59. [1798]
- SARGENT, T. J., AND C. A. SIMS (1977): "Business Cycle Modeling Without Pretending to Have Too Much a priori Economic Theory," in *New Methods in Business Cycle Research*, ed. by C. A. Sims. Minneapolis: Federal Reserve Bank of Minneapolis, 45–109. [1813]
- SCHIMDT-GROHÉ, S., AND M. URIBE (2003): "Closing Small Open Economy Models," *Journal of International Economics*, 61, 163–185. [1773]
- SENETA, E. (1976): *Regularly Varying Functions*, Lecture Notes in Mathematics, Vol. 408. Berlin: Springer. [1813]
- STOCK, J. H., AND M. W. WATSON (1999): "Diffusion Indexes," Working Paper 6702, NBER. [1772]
- WANG, Q., AND P. C. B. PHILLIPS (2006): "Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression," Working Paper, Yale University. [1784]
- WHITTAKER, E. T. (1923): "On a New Method of Graduation," *Proceedings of the Edinburgh Mathematical Association*, 78, 81–89. [1783]
- UZAWA, H. (1968): "Time Preference and the Penrose Effect in a Two-Class Model of Economic Growth," *Journal of Political Economy*, 77, 628–652. [1773]