A Central Limit Theorem for Endogenous Locations and Complex Spatial Interactions:*

Joris Pinkse†, Lihong Shen‡, and Margaret Slade§

October 2005

Abstract

We provide a new central limit theorem (CLT) for spatial processes under weak conditions which, unlike existing results, are plausible for most economic applications. In particular, our CLT is designed for problems that have some, but not necessarily all, of the following features: i) Agents choose the locations of observations to maximize profits, welfare, or some other objective. ii) The objects that are chosen (e.g., stores or brands) interact with one another. For example, they can be substitutes or complements. iii) Interaction can be complex. In particular, interaction between i and j need not depend only on the distance between the locations of i and j, but can also depend on distance to or location of other observations k, or possibly on the number of other such observations.

*We thank Tim Conley, Harry Kelejian, two anonymous referees, and seminar participants at cemmap, the universities of Tilburg and Warwick, and the Tinbergen Institute for their valuable comments. Margaret Slade would like to acknowledge support from the Leverhulme Foundation.

†Department of Economics, The Pennsylvania State University, 608 Kern Graduate Building, University Park PA 16802, joris@psu.edu

‡Department of Economics, The Pennsylvania State University, 608 Kern Graduate Building, University Park PA 16802, lshen@psu.edu

§Department of Economics, The University of Warwick, Coventry CV4 7AL, United Kingdom, m.slade@warwick.ac.uk

1 Introduction

In this paper we develop a new central limit theorem (CLT) for spatially dependent processes that allows applied researchers to work with a rich set of models and broad classes of data under assumptions that are more plausible in many economic applications than those that are made in the existing spatial literature.

It has become common to note that there are obvious differences between spatial and time–series data. The differences that are most often noted are that: i) time is one–dimensional whereas space is of higher dimension, ii) time is unidirectional whereas space has no natural direction, iii) time–series observations are usually evenly spaced whereas spatial observations are rarely located on a regular grid, and iv) time–series observations are draws from a continuous process whereas, with spatial data, it is common for the sample and the population to be the same (e.g., the set of all firms in a market).

Our paper is principally concerned with other differences that have received less attention in the literature. Indeed, the theoretical literature has thus far, implicitly or explicitly, treated spatial dependence as a simple multivariate extension of time–series dependence.[1] Observations are typically regarded as draws from a stationary underlying process (SUP), see e.g. Bolthausen (1982). In many interesting economic applications, however, spatial dependence is nonstationary, where by nonstationarity we mean that the joint distribution can depend on locations, not just on distance between locations, and not that a unit root is present. More problematic than nonstationarity is the fact that the characteristics of the spatial process can depend on the number of observations. Finally, both the location of economic observations and the total number of observations can be endogenous. Our paper deals with all of those eventualities.

To illustrate the advantages of our CLT, we begin with an enumeration of examples for which the assumptions that underlie previous theoretical work do not hold. All of our examples are spatial. However, we use the term 'space' in a broad sense to denote characteristic as well as geographic space, as has been common in economics since the time of Hotelling (1929).

Industrial Organization

Our first example, which is from industrial organization (IO), involves spatial price competition among retail outlets in a local market. Suppose that there are n firms in an imperfectly competitive market (e.g., hairdressers), and that those firms choose prices simultaneously. In that game, the price that salon i would like to charge depends on what j intends to charge. It is not true, however, that all i, j pairs compete equally, and it seems reasonable to assume that dependence decays with

[1] There is a vast literature in which the spatial dependence structure is assumed known up to a finite–dimensional parameter vector. Asymptotic results for such processes can be established under the assumption of independence after an appropriate transformation. See e.g. Anselin (1988) for an early treatise. The results presented here do not presume that the spatial dependence structure is known.

geographic distance. However, dependence between two salons is determined, not only by the distance between them, but also by the proximity of other rivals. Furthermore, when a new salon is introduced into the market, the nature of competition between the existing salons changes; the new salon is not simply a theretofore unobserved point of an SUP. Finally, the location of the new salon is apt to be endogenous. Indeed, firms attempt to choose locations where demand is high and costs are low.

Our second example from IO is similar to the first except that instead of having outlets that are located in geographic space, we consider brands of a differentiated product that are located in product–characteristic space (e.g., breakfast cereals). With obvious substitutions, the story is virtually the same — location and the number of brands matter in addition to distance in characteristic space, and both can be endogenous. There is also an important difference, however. It is less obvious what 'distance' means in product–characteristic space, and there is a richer set of possibilities to consider.[2]

Health Economics

Our next example is taken from the field of health economics, which has much in common with IO. We consider the formation of queues for hospital procedures, and our hypothetical data come from a complete set of hospitals in a region. It is clear that this is a spatial problem, since patients who face long waits in one hospital are apt to seek to have a procedure performed in a nearby hospital. In other words, there are spillover effects that decay with distance. Geographic distance therefore matters, but so do the hospitals' locations as well as the number of hospitals in the immediate area. Furthermore, hospital characteristics can also be important. For example, both the sorts of operations that are performed (e.g., heart transplants) and the expertise in performing those operations vary across hospitals. Hospitals that are 'close' in characteristic space should therefore also be closer substitutes. Finally, both the geographic locations of new hospitals and the characteristics of those hospitals will be endogenous, since they will be influenced by the length of queues.

International Economics

It is tempting to think that the problems that are the principal concern of the 'new economic geography' are spatial in nature. However, that is often not the case. Indeed, those models are more apt to deal with agglomeration than with location. Nevertheless, it is easy to come up with examples from international economics where our new CLT can be useful.

To illustrate, consider the problem of foreign direct investment. If there are fixed costs to setting up operations in a given country, such as learning a new language and writing contracts with suppliers, a firm will be more likely to enter a country where it already has factories. However,

[2] For example, Pinkse and Slade (2004) experiment with a number of measures of distance between brands of beer in product–characteristic space.

the advantage of a having a local factory should decay with its geographic distance from a firm's other operations. Furthermore, since there can be a critical number of operations after which economies of scale disappear, the number of factories that a firm already has in a region will also matter. In other words, both distance and location matter. Finally, both the locations and the number of factories that a firm establishes in a region are endogenous choices.

A second example in the international area is concerned with price convergence. In the presence of trade, competition for customers across countries has much in common with price competition within a country, and in both instances, competition should cause prices of identical products, as well as price levels, to tend towards equality. In other words, prices should converge. However, competition should be stronger when countries are in close geographic proximity as well as when they have similar characteristics (e.g., a common language). Nevertheless, in addition to distance, location matters. For example, the strength of competition and thus price convergence between Australia and New Zealand is apt to be stronger than between Uruguay and Peru, even though each pair shares a continent and a language and the distance between members of each pair is similar. One reason why convergence might be faster between Australia and New Zealand is that there are no countries in between. With this example, however, location is not endogenous, and, at least in the short run, neither is the number of countries.

Geology

Our final example, which comes from outside of economics, is concerned with resource appraisal. When a mineral deposit is discovered, mining companies would like to estimate the extent of reserves in that deposit, where reserves are defined as material that can be economically recovered at current prices using currently available technology. The standard way to assess reserves is to drill cores and take samples that can be processed in a lab to reveal mineral content. Drilling, however, is expensive. For that reason, companies use methods from spatial statistics to infer reserve values for an entire deposit from a small number of samples. This case seems to be closely analogous to a multivariate time–series problem. Indeed, each observation is a draw from a continuum of possible draws. Furthermore, the strength of dependence between two locations does not vary with the number of observations in close proximity. Nevertheless the number and locations of the drilling cites, which might be chosen to maximize the information content of the data for a given budget, are endogenous, implying that the analogy is not exact.

It would be easy to add applications taken from other fields to our list.[3] Indeed, our examples share some, but not necessarily all, of the following features: i) Agents choose the locations of observations to maximize profits, welfare, or some other objective. ii) The objects that are chosen (e.g.,

[3] For example, in the field of labor economics we could consider the problem of regional wage dispersion, in the field of political economy, we could consider local voting patterns, in the field of urban economics we could consider land development decisions, and in the field of epidemiology we could consider disease contagion.

stores or brands) interact with one another. For example, they can be substitutes or complements. iii) Interaction can be complex. In particular, interaction between i and j need not depend only on the distance between the locations of i and j, but can also depend on distance to or location of other observations k, or possibly on the number of other such observations. The assumptions that underlie our CLT, while less primitive than some, cover those eventualities.

Finally, the implication of our new CLT is not that estimators that have been proposed thus far do not work, but, to the contrary, to provide a rigorous justification for using them in a broader class of economic applications.

2 The new CLT

2.1 A Sketch

It was mentioned in the introduction that because of the wide range of situations covered, the conditions for the new CLT are more complex than is usually the case. We begin with a sequence of mean zero scalar–valued observations $\xi_{n1}, \ldots, \xi_{nn}$, whose distribution, dependence properties and values are for now unconstrained and can vary arbitrarily with the sample size n.[4] The objective here is to establish properties under which

$$S_n/\sigma_n \xrightarrow{\mathcal{D}} N(0,1), \tag{1}$$

where $S_n = \sum_{i=1}^n \xi_{ni}$ and $\sigma_n^2 = ES_n^2$. In the case of independent and identically distributed (i.i.d.) observations $x_1, \ldots, x_n$, whose properties are independent of the sample size, $S_n = n(\bar{x} - \mu_x)$ and $\sigma_n^2 = n\sigma_x^2$, with $\bar{x}$ and μ_x the sample and population mean of the x_i's, respectively, and σ_x^2 their population variance, in which case (1) reduces to the familiar

$$\sqrt{n}\frac{\bar{x} - \mu_x}{\sigma_x} \xrightarrow{\mathcal{D}} N(0,1).$$

Our central limit theorem will make use of an old idea (Bernstein, 1927) to allow for the stated desiderata. Bernstein's idea is to divide the data up into blocks in such a way that dependence between block averages decreases as the sample size increases. The situation here is more complicated than Bernstein's stationary time series design and hence so is our choice of blocks but the principle is the same.

We assume that the data can be partitioned up into J groups, where $J < \infty$ does not depend on the sample size. Such partitioning does not actually have to take place; there merely must be the possibility to do so. Each observation belongs to exactly one group and its membership can be different for each value of n. Our assumptions will require that groups $2, \ldots, J$ play no role

[4] Vector–valued ξ's are a simple extension thank to the Wold device.

asymptotically which in a stationary weak dependence setting would imply that the number of observations in these groups would be a negligble fraction of the total, asymptotically. The purpose of creating multiple groups is to create 'gaps' of increasing size between collections of observations ('subgroups') in the same group.

To illustrate this point, consider figure 1. In this example, space is one–dimensional and there are 2 groups (first index) consisting of 6 subgroups (second index) each. As the number of observations increases, so does the number of subgroups in each group and the number of observations in the subgroups (albeit not necessarily strictly or monotonically). So the 'distance' between subgroups in the same group increases and, because of our weak dependence condition stated below, the strength of dependence between subgroups hence decreases. In the limit observations in different subgroups of the same group are independent of one another.

If one treats the sum over all observations in each subgroup as a single 'observation' in the group, then we approximately have a triangular sequence of independent (albeit sample–size dependent) random variables, whose sum has a normal distribution under well–understood conditions. This means that the sums over all ξ_{ni}'s in each group converge to separate normal distributions. If the ratio of the number of observations in group 2 to the number in group 1 converges to zero, then the limiting distribution of S_n will be the same as that generated by group 1, i.e. the observations in group 2 are of no consequence for the asymptotic distribution. In multiple dimensions, the principle is identical albeit that the required number of groups J is larger; it is usually $J = 2^d$ if d is the number of dimensions. Nevertheless, all but one are of no consequence for the asymptotic distribution.

We do not assume stationarity[5] and our 'blocks' do not have to be blocks; they are judiciously chosen subsets of observations. Absent stationarity, the groups that are asymptotically negligible are not necessarily the ones with the smallest number of observations, since the observations are allowed to have different variances.

The blocking method is not completely general. It is possible to create examples in which any observation i has zero dependence with all but an asymptotically negligible fraction of the observations but in which the blocking assumption is not satisfied.

2.2 Notation

The J groups are represented by $\mathcal{G}_{n1}, \ldots, \mathcal{G}_{nJ}$ and the m_{nj} subgroups of group j are $\mathcal{G}_{nj1}, \ldots, \mathcal{G}_{njm_{nj}}$. S_{nj} and S_{njt} denote the partial sums over observations in group j and subgroup t of group j,

[5] As mentioned informally in the introduction, we neither assume data to be on a regular grid nor do we assume that dependence is the same if the distance between two data points is the same.

respectively, i.e.

$$S_{nj} = \sum_{t=1}^{m_{nj}} S_{njt}, \quad S_{njt} = \sum_{s \in \mathcal{G}_{njt}} \xi_{ns}. \tag{2}$$

In analogy to $\sigma_n^2 = ES_n^2$, defined above, we let $0 < \sigma_{nj}^2 = ES_{nj}^2 < \infty$, and $0 < \sigma_{njt}^2 = ES_{njt}^2 < \infty$. Finally, $\varsigma_{nj}^2 = \sum_{t=1}^{m_{nj}} \sigma_{njt}^2$.

2.3 Weak Dependence

Strong mixing assumes that any two events defined on sets of observations become less dependent as the distance between the two sets becomes greater. The limitation of the strong mixing assumptions lies in the 'any event' portion of the definition. The strong–mixing assumption has the problem that reasonable sequences of random variables exist whose sample mean can easily be shown to have a limiting normal distribution but which do not satisfy the strong–mixing assumption, see e.g. Andrews (1984). An alternative but more complicated strategy to ours is to assume *near epoch dependence*, see e.g. Ibragimov (1962). We instead follow Doukhan and Louhichi (1999) and assume that for all functions f in some class of functions $\mathcal{F}$, defined below, the covariance between partial sums of observations in different sets declines as the distance between sets increases. Strong mixing would require such a decline of covariances of *all* functions of sets of variables instead of some scalar–argument functions of sums of variables. A more detailed discussion of the merits of the condition used here can be found in Doukhan and Louhichi (1999), albeit that Doukhan and Louhichi's assumptions are in the context of stationary data that do not change with the sample size.

We now state formally our weak dependence assumption.

Definition 1 *Let $\mathcal{F}$ be a collection of functions $\{f : \forall t \in \mathbb{R} : f(t) = t$ or $\exists u \in \mathbb{R} : \forall t : f(t) = e^{\iota ut}\}$, where ι is the imaginary number.*

Let V be the variance operator.

Assumption A *For any $j = 1, \ldots, J$, let $\mathcal{G}_n^*, \mathcal{G}_n^{**} \subset \mathcal{G}_{nj}$ be any sets for which*

$$\forall t = 1, \ldots, m_{nj} : \mathcal{G}_{njt} \cap \mathcal{G}_n^* \neq \emptyset \Rightarrow \mathcal{G}_{njt} \cap \mathcal{G}_n^{**} = \emptyset.$$

Then for any function $f \in \mathcal{F}$,

$$\left\| Cov\left(f\Big(\sum_{s \in \mathcal{G}_n^*} \xi_{ns}\Big), f\Big(\sum_{s \in \mathcal{G}_n^{**}} \xi_{ns}\Big)\right)\right\| \leq \sqrt{Vf\Big(\sum_{s \in \mathcal{G}_n^*} \xi_{ns}\Big)} \sqrt{Vf\Big(\sum_{s \in \mathcal{G}_n^{**}} \xi_{ns}\Big)} \alpha_{nj}, \tag{3}$$

for some 'mixing' numbers α_{nj} with

$$\lim_{n \to \infty} \sum_{j=1}^{J} m_{nj}^2 \alpha_{nj} = 0. \tag{4}$$

Note that all observations in $\mathcal{G}_n^*$ and $\mathcal{G}_n^{**}$ belong to the same group but that observations in the same subgroup cannot be both in $\mathcal{G}_n^*$ and $\mathcal{G}_n^{**}$. So assumption A requires a bound on a correlation of two quantities, each corresponding to different sets of subgroups of the same group.

As mentioned above, the variances of individual elements can be different. To ensure that each subgroup makes an asymptotically negligble contribution and that all groups other than group 1 are of no relevance for the asymptotic distribution, we make assumption B.

Assumption B

$$\lim_{n\to\infty} \max_{t\leq m_{nj}} \sigma_{njt}/\varsigma_{nj} = 0,\ j = 1,\ldots,J, \qquad \lim_{n\to\infty} \varsigma_{nj}/\varsigma_{n1} = 0,\ j = 2,\ldots,J. \tag{5}$$

Finally, we make a uniform integrability assumption which is a standard feature of CLT's, albeit in a different form.

Assumption C *For some sequence* $\{h_n\}$ *for which* $\lim_{n\to\infty} h_n = 0$ *and for all* $j = 1,\ldots,J$,

$$\lim_{n\to\infty} \max_{t\leq m_{nj}} E\left(\frac{S_{njt}^2}{\sigma_{njt}^2} I\left(\left|\frac{S_{njt}}{\varsigma_{nj}}\right| > h_n\right)\right) = 0. \tag{6}$$

A sufficient condition for assumption C is that for some $p > 1$,

$$E|S_{njt}|^{2p} = o\left(\sigma_{njt}^2 \varsigma_{nj}^{2p-2}\right), \quad j = 1,\ldots,J;\ t = 1,\ldots,m_{nj}.^6 \tag{7}$$

We can now state our theorems.

2.4 Our Main Results

Theorem 1 *If assumptions A–C hold, then*

$$\frac{S_n}{\sigma_n} \xrightarrow{\mathcal{D}} N(0,1). \tag{8}$$

The vector–version now follows immediately with the Wold device. Let $\vec{\xi}_{ni}$ be vector–valued and

$$\vec{S}_n = \sum_{i=1}^n \vec{\xi}_{ni}.$$

Let furthermore $\Sigma_n = V\vec{S}_n$.

Theorem 2 *If for any vector v with $||v|| = 1$, assumptions A–C are satisfied for $\xi_{ni} = v'\Sigma_n^{-1/2}\vec{\xi}_{ni}$, then*

$$\Sigma_n^{-1/2}\vec{S}_n \xrightarrow{\mathcal{D}} N(0,I). \tag{9}$$

All proofs are in an appendix.

[6] By the Hölder and Markov inequalities, $E\big(S_{njt}^2 I(|S_{njt}| > h_n\varsigma_{nj})\big) \leq (E|S_{njt}|^{2p})^{1/p}\big(P(|S_{njt}| > h_n\varsigma_{nj})\big)^{1-1/p} \leq E|S_{njt}|^{2p} h_n^{2-2p}\varsigma_{nj}^{2-2p}$.

References Cited

Andrews, D.W.K. (1984) "Non–Strong Mixing Autoregressive Processes," *Journal of Applied Probability* 21, 930–934.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models,* Kluwer (Dordrecht, The Netherlands).

Bernstein, S. (1927) "Sur l'Extension du Théorème du Calcul des Probabilités aux Sommes de Quantités Dependantes," *Mathematische Annalen* 97, 1–59.

Bolthausen, E. (1982) "On the Central Limit Theorem for Stationary Mixing Random Fields," *Annals of Probability* 10, 1047–1050.

Davidson, James (1994), *Stochastic Limit Theory,* Oxford University Press, Oxford.

Doukhan, P. and S. Louhichi (1999) "A New Weak Dependence Condition and Applications to Moment Inequalities," *Stochastic Processes and their Applications* 84, 312–342.

Hotelling, H., (1929) "Stability in Competition," *Economic Journal*, 39: 41–57.

Ibragimov, I.A. (1962) "Some Limit Theorems for Stationary Processes," *Theory of Probability and Its Applications* 7, 349–382.

Pinkse, J. and Slade, M.E. (2004), "Mergers, Brand Competition and the Price of a Pint," *European Economic Review* 48, 617–643.

Pinkse, J., Slade, M.E. and Shen, L. (2005) "Dynamic Spatial Discrete Choice Using One Step GMM: An Application to Mine Operating Decisions," Penn State mimeo.

Zhengyan, Lin and Lu Chuanrong (1996) *Limit Theory for Mixing Dependent Random Variables,* Kluwer, Dordrecht.

A Central limit theorem

PLEASE NOTE: Throughout the appendix we indicate when convenient the assumption, lemma, equation or well–known theorem an (in)equality is based on above that (in)equality; an assumption is indicated by its letter, a common theorem by its name, a lemma by the letter L followed by its number and an equation by its equation number in brackets.

A.1 Preliminaries

For this appendix, let $\tilde{S}_m = \sum_{j=1}^m \tilde{S}_{mj}$, where $\tilde{S}_{mj}$ is a real–valued random variable with mean zero and variance $0 < \tilde{\sigma}_{mj}^2 = E\tilde{S}_{mj}^2 < \infty$. Let further $\tilde{\sigma}_m^2 = E\tilde{S}_m^2$, and make the following assumptions.

AA1 For some sequence $\{\tilde{h}_m\}$ with $\lim_{m\to\infty} \tilde{h}_m = 0$,

$$\lim_{m\to\infty} \max_{j\le m} E\Big(\frac{\tilde{S}_{mj}^2}{\tilde{\sigma}_{mj}^2} I\big(\big|\tilde{S}_{mj}\big| > \tilde{h}_m\big)\Big) = 0, \tag{10}$$

AA2 For any function $f \in \mathcal{F}$ and any nonoverlapping groups $\tilde{\mathcal{G}}_1, \tilde{\mathcal{G}}_2$,

$$\left\|\mathrm{Cov}\left(f\Big(\sum_{t\in\tilde{\mathcal{G}}_1} \tilde{S}_{mt}\Big), f\Big(\sum_{t\in\tilde{\mathcal{G}}_2} \tilde{S}_{mt}\Big)\right)\right\| \le \sqrt{Vf\Big(\sum_{t\in\tilde{\mathcal{G}}_1} \tilde{S}_{mt}\Big)}\sqrt{Vf\Big(\sum_{t\in\tilde{\mathcal{G}}_2} \tilde{S}_{mt}\Big)}\tilde{\alpha}_m,$$

where $\tilde{\alpha}_m$ is such that $\lim_{m\to\infty} m^2\tilde{\alpha}_m = 0$.

AA3 $\lim_{m\to\infty} \max_{j\le m} \tilde{\sigma}_{mj} = 0$.

Lemma 1 *Let ι denote the imaginary number and let $\tilde{\phi}_{mj}(u) = Ee^{\iota u\tilde{S}_{mj}}$. For all u,*

$$\lim_{m\to\infty} \left| Ee^{\iota u\tilde{S}_m} - \prod_{j=1}^m \tilde{\phi}_{mj}(u) \right| = 0. \tag{11}$$

Proof: By lemma 6.1.2 of Zhengyan and Chuanrong (1996), the LHS in (11) is bounded by

$$\begin{aligned}
&\lim_{m\to\infty} \sum_{i=1}^{m-1}\sum_{t=i+1}^{m} \left| E\Big(\big\{e^{\iota u\tilde{S}_{mi}} - \tilde{\phi}_{mi}(u)\big\}(e^{\iota u\tilde{S}_{mt}} - 1)\prod_{s=t+1}^m e^{\iota u\tilde{S}_{ms}}\Big)\right| \\
&\overset{\text{AA2}}{\le} \lim_{m\to\infty} \sum_{i=1}^{m-1}\sum_{t=i+1}^{m} \tilde{\alpha}_m \sqrt{E\Big|e^{\iota u\tilde{S}_{mi}} - \tilde{\phi}_{mi}(u)\Big|^2}\sqrt{E\Big|(e^{\iota u\tilde{S}_{mt}} - 1)\prod_{s=t+1}^m e^{\iota u\tilde{S}_{ms}}\Big|^2} \\
&\le 2\lim_{m\to\infty} \sum_{i=1}^{m-1}\sum_{t=i+1}^{m} \tilde{\alpha}_m \le 2\lim_{m\to\infty} m^2\tilde{\alpha}_m = 0,
\end{aligned}$$

since $|e^{\iota t}| \le 1$. ✓

Lemma 2 *For any u and a sequence $\{\psi_m\}$ with $\lim_{m\to\infty}\psi_m = 0$,*

$$\left|\tilde\phi_{mj}(u) - (1 - u^2\tilde\sigma^2_{mj}/2)\right| \le \tilde\sigma^2_{mj}\psi_m, \tag{12}$$

Proof: By theorem 11.6 of Davidson (1994), the LHS in (12) is for any sequence $\{\tilde h_m\}$ bounded by

$$\begin{aligned} E\min\bigl(u^2\tilde S^2_{mj}, 6^{-1}|u\tilde S_{mj}|^3\bigr) &\le u^2 E\bigl(\tilde S^2_{mj} I(|\tilde S_{mj}| > \tilde h_m)\bigr) + 6^{-1}|u|^3 E\bigl(|\tilde S_{mj}|^3 I(|\tilde S_{mj}| \le \tilde h_m)\bigr) \\ &\le \tilde\sigma^2_{mj} u^2 E\left(\frac{\tilde S^2_{mj}}{\tilde\sigma^2_{mj}} I\bigl(|\tilde S_{mj}| > \tilde h_m\bigr)\right) + \tilde\sigma^2_{mj}\tilde h_m 6^{-1}|u|^3. \end{aligned}$$

Now pick $\psi_m = u^2 E\left(\frac{\tilde S^2_{mj}}{\tilde\sigma^2_{mj}} I\bigl(|\tilde S_{mj}| > \tilde h_m\bigr)\right) + \tilde h_m 6^{-1}|u|^3 = o(1)$ by assumption AA1. ✓

Lemma 3 *If $\lim_{m\to\infty}\sum_{j=1}^m \tilde\sigma^2_{mj} < \infty$, then for any u,*

$$\lim_{m\to\infty}\sum_{j=1}^m \left|\log\tilde\phi_{mj}(u) + \tilde\sigma^2_{mj}u^2/2\right| = 0. \tag{13}$$

Proof: For any a, b let $[a; b]$ denote some number between a and b. Then by the mean value theorem $\log\tilde\phi_{mj} = (\tilde\phi_{mj} - 1)/[\tilde\phi_{mj}; 1]$. Thus,

$$\begin{aligned} \sum_{j=1}^m \bigl|\log\tilde\phi_{mj} + \tilde\sigma^2_{mj}u^2/2\bigr| &= \sum_{j=1}^m \Bigl|\frac{\tilde\phi_{mj} - 1}{[\tilde\phi_{mj}; 1]} + \tilde\sigma^2_{mj}u^2/2\Bigr| \\ &= \sum_{j=1}^m \Bigl|\frac{\tilde\phi_{mj} - 1 + \tilde\sigma^2_{mj}u^2/2}{[\tilde\phi_{mj}; 1]} - \frac{\tilde\sigma^2_{mj}u^2}{2}\Bigl(\frac{1}{[\tilde\phi_{mj}; 1]} - 1\Bigr)\Bigr| \\ &\overset{\text{triangle}}{\le} \max_{j\le m}\frac{1}{[|\tilde\phi_{mj}|; 1]}\sum_{j=1}^m |\tilde\phi_{mj} - 1 + \tilde\sigma^2_{mj}u^2/2| + \max_{j\le m}\Bigl|\frac{1}{[|\tilde\phi_{mj}|; 1]} - 1\Bigr| u^2\sum_{j=1}^m \tilde\sigma^2_{mj}/2. \end{aligned} \tag{14}$$

By lemma 2 and the stated condition, (14) is $o(1)$ if $\max_{j\le m}\bigl(1 - [|\tilde\phi_{mj}|; 1]\bigr) = o(1)$. Now, again by theorem 11.6 of Davidson (1994), $|\tilde\phi_{mj} - 1| \le \tilde\sigma^2_{mj}u^2/2$. Hence,

$$\max_{j\le m}\bigl(1 - [|\tilde\phi_{mj}|; 1]\bigr) \le \max_{j\le m}|1 - \tilde\phi_{mj}| \le u^2\max_{j\le m}\tilde\sigma^2_{mj}/2 = o(1),$$

by assumption AA3. ✓

Lemma 4 *If $\lim_{m\to\infty}\sum_{j=1}^m \tilde\sigma^2_{mj} = 1$, then*

$$\tilde S_m \xrightarrow{\mathcal{D}} N(0, 1). \tag{15}$$

Proof: It suffices to show that for all u,

$$\lim_{m\to\infty}\log E e^{\iota u\tilde S_m} = -u^2/2.$$

Now,

$$\lim_{m\to\infty}\log Ee^{\iota u\tilde{S}_m}+u^2/2=\lim_{m\to\infty}\Big(\log Ee^{\iota u\tilde{S}_m}-\sum_{j=1}^{m}\log\tilde{\phi}_{mj}(u)\Big) \tag{16}$$

$$+\lim_{m\to\infty}\sum_{j=1}^{m}\Big(\log\tilde{\phi}_{mj}(u)+\tilde{\sigma}_{mj}^2u^2/2\Big) \tag{17}$$

$$+\Big(1-\lim_{m\to\infty}\sum_{j=1}^{m}\tilde{\sigma}_{mj}^2\Big)u^2/2. \tag{18}$$

Expressions (16)–(18) are zero by lemma 1, lemma 3 and the assumption made in the present lemma, respectively. ✓

A.2 Main Result

Lemma 5

$$\frac{S_{nj}}{\varsigma_{nj}}\xrightarrow{\mathcal{D}}N(0,1),\quad j=1,\ldots,J. \tag{19}$$

Proof: Let $m=m_{nj}$, choose $\tilde{S}_{mt}=S_{njt}/\varsigma_{nj}$, $t=1,\ldots,m_{nj}$, $\tilde{\sigma}_{mt}^2=E\tilde{S}_{mt}^2=\sigma_{njt}^2/\varsigma_{nj}^2$ and let $\tilde{h}_m=h_n$. Then assumptions AA1–AA3 are satisfied. Apply lemma 4. ✓

Proof of theorem 1: Note that

$$\frac{S_n}{\varsigma_{n1}}=\frac{S_{n1}}{\varsigma_{n1}}+\sum_{j=2}^{J}\frac{\varsigma_{nj}}{\varsigma_{n1}}\frac{S_{nj}}{\varsigma_{nj}}.$$

The first RHS term has a limiting $N(0,1)$ distribution by lemma 5 and the second RHS term is $o(1)$ by assumption B and lemma 5. ✓

Proof of theorem 2: Follows immediately from theorem 1 using the Wold device. ✓

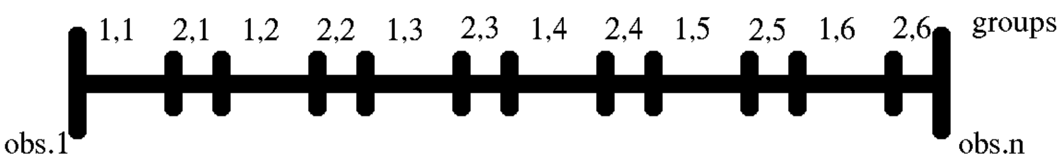

Figure 1: Bernstein Blocks