NONLINEAR COINTEGRATING REGRESSION UNDER WEAK IDENTIFICATION

by

Xiaoxia Shi and Peter C.B. Phillips

COWLES FOUNDATION PAPER NO. 1355

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2012

http://cowles.econ.yale.edu/

Econometric Theory, **28**, 2012, 509–547.
doi:10.1017/S0266466611000648

NONLINEAR COINTEGRATING REGRESSION UNDER WEAK IDENTIFICATION

Xiaoxia Shi
University of Wisconsin

Peter C.B. Phillips
Yale University, University of Auckland
University of Southampton, and Singapore Management University

An asymptotic theory is developed for a weakly identified cointegrating regression model in which the regressor is a nonlinear transformation of an integrated process. Weak identification arises from the presence of a loading coefficient for the nonlinear function that may be close to zero. In that case, standard nonlinear cointegrating limit theory does not provide good approximations to the finite-sample distributions of nonlinear least squares estimators, resulting in potentially misleading inference. A new local limit theory is developed that approximates the finite-sample distributions of the estimators uniformly well irrespective of the strength of the identification. An important technical component of this theory involves new results showing the uniform weak convergence of sample covariances involving nonlinear functions to mixed normal and stochastic integral limits. Based on these asymptotics, we construct confidence intervals for the loading coefficient and the nonlinear transformation parameter and show that these confidence intervals have correct asymptotic size. As in other cases of nonlinear estimation with integrated processes and unlike stationary process asymptotics, the properties of the nonlinear transformations affect the asymptotics and, in particular, give rise to parameter dependent rates of convergence and differences between the limit results for integrable and asymptotically homogeneous functions.

1. INTRODUCTION

Nonlinear models provide an important means of extending the conventional linear cointegrating structures that are now commonly used in applied work. Nonlinearities provide a mechanism for controlling and modifying the random wandering characteristics of unit root time series, leading to a much wider range of possible

Our thanks to two referees and the co-editor, Pentti Saikkonen, for helpful comments on the original version. The paper originated in a 2008 Yale take-home examination. The first complete draft was circulated in December 2009. Xiaoxia Shi gratefully acknowledges support from the Cowles Foundation via a Carl Arvid Anderson Fellowship at Yale University. Peter Phillips thanks the NSF for support under grant SES 06-47086 and 09-56687. Address correspondence to Peter Phillips, Cowles Foundation, Yale University, P.O. Box 208281, New Haven, CT 06520, USA; e-mail: peter.phillips@yale.edu.

© Cambridge University Press 2011

response functions in regressions with such time series. For instance, integrable transformations of integrated time series attenuate outliers rather than proportionately transmit their effects as in linear cointegrating systems. Transformations of this type are valuable in modeling uneven output responses to economic fundamentals such as those that can occur in the presence of market interventions or regulatory regimes like exchange rate target zones.

Another useful property of nonlinear transformations is that they can modify the characteristics of nonstationary series, including their memory attributes. Modifications of this type are helpful in modeling time series like asset returns, which have near martingale difference characteristics, in terms of economic fundamentals that may behave much more like integrated time series. In such cases, the effects of the stochastic trend in the fundamentals are sufficiently attenuated to be negligible, except perhaps over long time periods where the drift in asset returns becomes perceptible. A useful mechanism for capturing such effects is to utilize loading coefficients on the nonlinear response functions that are allowed to be local to zero. The cointegrating effects then become "small," and they are only weakly identified. This approach gives flexibility in modeling the effects of fundamentals on returns and offers the potential for improvements over linear models in predicting asset returns using near integrated predictor processes, whose role has recently been emphasized in the work of Campbell and Yogo (2006) and others.

The goal of the present paper is to deal with such formulations and develop an asymptotic theory that retains its validity for small cointegrating effects. In particular, we study nonlinear cointegration models of the following form:

$$Y_t = \beta g(X_t, \pi) + u_t, \tag{1.1}$$

where X_t is an $I(1)$ process, Y_t is a dependent variable, not necessarily $I(1)$, u_t is an error term (to be specified more precisely later), $g(x, \pi)$ is a nonlinear transformation of x whose form is known up to a parameter π, and β is a loading coefficient that measures the importance of the nonlinear regression effect.

Models like (1.1) have the attractive feature that they can relate processes of different integration orders. As intimated earlier, this feature may be especially appealing in modeling and predicting stock market returns. Stock returns commonly behave as martingale differences, whereas the variables that are used in prediction are often $I(1)$, as discussed in Marmer (2008), leading to a potential imbalance in a regression formulation. Accordingly, any relationship between stock return levels and stochastic trend predictors is inevitably weak because of the efficiency of modern stock markets. In terms of the model (1.1), this consideration may be captured for a wide class of possible regression functions simply by permitting the true value of the loading coefficient to be close to zero. To develop an orderly asymptotic theory that accommodates this possibility, the model may be formulated to allow the true parameter, β_n, to drift to zero as the sample size $n \to \infty$. Then, if Y_t denotes stock returns and X_t denotes an $I(1)$ regressor embodying economic fundamentals, the behavior of Y_t will closely follow u_t. If u_t is a martingale difference, then Y_t may be regarded as local to a martingale difference sequence, where the locality is affected by the form of the function g, the nonstationary nature of x_t, and the magnitude of the localizing loading coefficient β_n. Such a relationship may be considered to be weakly identifying.

When a relationship such as (1.1) is weak, the nonlinear least squares (NLS) estimators $\left(\hat{\beta}_n, \hat{\pi}_n\right)$ of the true parameters (β_n, π_n) do not behave as standard asymptotic theory for nonstationary time series (Park and Phillips, 2001 [hereafter PP]) predicts even in large samples. In the extreme case, when $\beta_n = \beta_0 = 0$, π_0 is not identified and the estimator $\hat{\pi}_n$ cannot reasonably be expected to be anywhere near π_0, although standard asymptotic theory, which proceeds under the assumption that $\beta_0 > 0$, would imply that $\hat{\pi}_n$ is consistent and asymptotically normal. Similar discrepancies between standard asymptotic theory and the finite-sample distributions of NLS estimators exist when β_0 is close to zero.

The present paper explores these issues associated with potentially weak identification. The main contribution of the paper is to provide a local asymptotic theory that can approximate the finite-sample distributions uniformly well even when β_0 is close to zero. The new asymptotic theory is used to construct robust confidence intervals for the NLS estimators $\left(\hat{\beta}_n, \hat{\pi}_n\right)$ and may be further developed to use in the construction of forecasting intervals that take account of potentially small cointegrating effects. The critical values used to construct confidence intervals are nonstandard, as sometimes occurs in nonstationary regression, but these can be simulated. The robust confidence intervals are shown to have correct asymptotic size, indicating that they have good finite-sample coverage probabilities irrespective of identification strength.

This paper is most closely related to Cheng (2008); see also Cheng (2010). Cheng (2008) studies a weakly identified nonlinear regression model of the form (1.1) but in the cross-section context where both the regressor and the error are independent and identically distributed (i.i.d.). The present paper extends the limit theory to a nonstationary time series environment, in which the stochastic trend effect on Y_t is effectively small. As in Cheng (2008), we derive asymptotics of the NLS estimators under a drifting sequence of true values of β to characterize the behavior of NLS estimators when β_0 is close to zero. The limit theory reveals some important differences with the cross-section case. Unlike cross-section and stationary cases, it is shown that the effect of the drift rate in the loading coefficient β_n on the asymptotic theory depends on the shape characteristics of the function g and the parameter π_0. Correspondingly, there is interaction between the loading coefficient and nonlinear function effects when x_t is nonstationary. These dependencies reflect the nuances that arise in the impact of stochastic trends on outputs when the cointegrating association may be weak and nonlinear. These dependencies also affect inference, and their role will become clear in what follows.

The techniques used to derive the asymptotic distributions of nonlinear functions of integrated processes are mainly based on Park and Phillips (1999) and PP. PP provided building blocks for nonlinear cointegration asymptotics by establishing a limit theory for suitably standardized sample functions of quantities such as $g(X_t, \pi)$ and its derivatives, in addition to sample covariances of these quantities and u_t. For their results, PP require and prove only pointwise (in π) weak convergence of such sample covariances. In the present context, pointwise convergence is not enough because the covariance term contributes to the limit theory of the estimators when β_n drifts to zero. An important technical contribution of the present paper is to show that weak convergence of such sample covariances to certain mixed normal and stochastic integral limits holds uniformly over a compact space of π values. The new results are established by demonstrating stochastic equicontinuity of the sample covariance process. The uniform convergence results are of independent interest and useful in other extremum estimation problems involving nonlinear cointegration.

The paper is organized as follows. Section 2 lays out the model, the basic assumptions, and some embedding arguments used in the proofs. Section 3 introduces the NLS estimators of the loading coefficient and the nonlinear transformation coefficient. Section 4 develops the limit theory for the NLS estimators $\left(\hat{\beta}_n, \hat{\pi}_n\right)$ for integrable functions $g(\cdot, \pi)$ under various decay rates of the loading coefficient β_n. Section 5 develops analogous limit results for asymptotically homogeneous functions $g(\cdot, \pi)$. These results encompass the case where identification is strong enough to ensure that $\hat{\pi}_n$ is consistent but may still affect rates of convergence and the more extreme case where weak identification results in inconsistent estimation of π, leading to a random limit for $\hat{\pi}_n$ that reflects the weak identification. The latter outcome corresponds to results given in the partial identification literature (cf. Phillips, 1989; Stock and Wright, 2000). This section also proves a uniform weak convergence result to stochastic integrals. Section 6 discusses confidence interval construction. Section 7 concludes. The Appendixes provide proofs of the main results in the paper and some useful auxiliary lemmas.

2. THE MODEL AND BASIC ASSUMPTIONS

The model we consider is the following nonlinear regression model for a time series Y_t:

$$Y_t = \beta_0 g(X_t, \pi_0) + u_t, \tag{2.1}$$

where $g : R \times \Pi \to R$ is a known function, X_t and u_t are the regressors and regression errors, respectively, and $\theta_0 \equiv (\beta_0, \pi_0)'$ is the true parameter vector that lies in a parameter set $\Theta \equiv R \times \Pi \subset R^2$. We consider the case where X_t is an integrated process and u_t is a martingale difference sequence, specified more precisely later. Model (2.1) is a nonlinear cointegrating regression, but it differs from the nonlinear cointegrating regression considered in PP in an important way: the parameter π_0 is not identified in (2.1) if $\beta_0 = 0$ and only weakly identified if β_0 is close to zero.

The partial identification feature of model (2.1) invalidates standard NLS inference not only when $\beta_0 = 0$ but also when β_0 is close to zero. This point is discussed in Cheng (2008) in the context of cross-section nonlinear regression. We extend the limit theory to a nonstationary time series environment and construct suitable methods of inference. As in Cheng (2008), we derive asymptotics of the NLS estimators under a drifting sequence of true values (β_n, π_n) in an effort to characterize the behavior of NLS estimators when β_0 is close to zero. Unlike cross-section and stationary cases, however, the effect of the drift rate in β_n on the asymptotics depends on the shape characteristics of the function g and the parameter π_0. These dependencies affect inference, and their role will become clear in what follows.

We now complete the specification of model (2.1). We assume the generating mechanism of X_t is the unit root process

$$X_t = X_{t-1} + v_t, \qquad t = 1, 2, \ldots, n \tag{2.2}$$

and set $X_0 = 0$ for convenience, although $X_0 = o_{\text{a.s.}}\left(\sqrt{n}\right)$ will be sufficient for the results that follow and allows for moderately integrated initializations. Other possibilities for initialization might be considered (e.g., as in Phillips and Magdalinos, 2009) but, for brevity, are not pursued here. Similarly, the generating mechanism (2.2) for X_t may be replaced with a local to unity process without materially affecting results, which will be important in empirical applications such as those in Campbell and Yogo (2006). For the component time series u_t and v_t, we define the stochastic processes U_n and V_n on $[0, 1]$ by the standardized partial sums

$$U_n(r) = n^{-1/2} \sum_{t=1}^{[nr]} u_t \quad \text{and} \quad V_n(r) = n^{-1/2} \sum_{t=0}^{[nr]} v_{t+1}, \tag{2.3}$$

where $[r]$ denotes the largest integer not exceeding r.

The following high-level assumption is convenient and is closely related to similar assumptions in the literature, e.g., Assumption 2.1 in PP.

Assumption 2.1.

(a) $\sup_{r \in l0,1]} || (U_n(r), V_n(r)) - (U(r), V(r))|| \to_{\text{a.s.}} 0$ as $n \to \infty$, where (U, V) is a vector Brownian motion with

$$\operatorname{Var}\left(\begin{pmatrix} U(r) \\ V(r) \end{pmatrix}\right) = r \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix} \quad \text{for } r \in [0, 1],$$

where $\rho \in (-1, 1)$.

For each n, there exists a filtration $(\mathcal{F}_{n,t})$, $t = 0, \ldots, n$, such that

(b) $\left(u_t, \mathcal{F}_{n,t}\right)$ is a martingale difference sequence with $\mathrm{E}\left(u_t^2 | \mathcal{F}_{n,t-1}\right) = \sigma_u^2$ almost surely (a.s.) for all $t = 1, \ldots, n$ and $\sup_{1 \le t \le n} \mathrm{E}(|u_t|^q \, |\mathcal{F}_{n,t-1}) < \infty$ a.s. for some $q > 2$; and

(c) X_t is adapted to $\mathcal{F}_{n,t-1}$, $t = 1, \ldots, n$.

Remarks.

(a) The stochastic processes (U_n, V_n) are defined on $D^2[0,1]$, where $D[0,1]$ is the space of càdlàg functions. As in PP, it is convenient to endow the space $D[0,1]$ with the uniform topology (see, e.g., Billingsley, 1968) and employ the Skorokhod representation.

(b) It is more common to have $\to_d$ instead of $\to_{\text{a.s.}}$ in Assumption 2.1(a). However, if $(U_n, V_n) \to_d (U, V)$, by the Skorokhod representation theorem, there exists a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$ supporting $\left(U_n^0, V_n^0\right)$ and $\left(U^0, V^0\right)$ such that

$$\left(U_n^0, V_n^0\right) =_d (U_n, V_n), \qquad \left(U^0, V^0\right) =_d (U, V), \quad \text{and}$$
$$\left(U_n^0, V_n^0\right) \to \left(U^0, V^0\right) \quad \text{a.s.} \tag{2.4}$$

For the purpose of deriving the consistency and the asymptotic distribution of the NLS estimator $(\hat{\beta}_n, \hat{\pi}_n)$, there is no loss of generality in assuming $(U_n, V_n) = \left(U_n^0, V_n^0\right)$ and $(U, V) = \left(U^0, V^0\right)$ and letting Assumption 2.1(a) hold. This assumption allows us to avoid repeated embedding arguments. When $(U_n, V_n) \to_d (U, V)$ holds instead of $(U_n, V_n) \to_{\text{a.s.}} (U, V)$, the results still hold with $\to_{\text{a.s}}$ and $\to_p$ replaced by $\to_d$ by virtue of the representation theory.

(c) The condition (c) in Assumption 2.1 that X_t is adapted to $\mathcal{F}_{n,t-1}$ is a simplifying assumption, and it is restrictive in linear cointegrating regression. But it is common in fully specified (cointegrating) regression models and allows for arguments based on martingale central limit theory, as in PP, for nonlinear cointegration. In the case of structural systems, where there is contemporaneous (and possibly serial cross) dependence between X_t and u_t, some modifications of the derivations and the results are required. The limit theory is especially complex in the case of models with integrable nonlinear functions, and it is not yet completely worked out in the literature even for the strongly identified case. In fact, when $g(\cdot, \pi)$ is an integrable function, substantially different proofs are needed, as shown by the limit theory in Jeganathan (2008) and Chang and Park (2011), the latter also for martingale difference u_t. Further, the limit theory involves only a partial invariance principle in the general case (Jeganathan, 2008). When $g(\cdot, \pi)$ is asymptotically homogeneous, the modifications that are required follow those in de Jong (2002) and Ibragimov and Phillips (2008, Thm. 3.1). Throughout the current paper, we will maintain condition (c), which is likely to be most relevant in prediction and in applied work on stock return regressions, to explore the effects of weak identification in nonlinear nonstationary models and to keep this paper to manageable length.

3. NONLINEAR LEAST SQUARES ESTIMATION

Let $\theta = (\beta, \pi)'$ and define the nonlinear least squares criterion function

$$Q_n(\theta) = n^{-1} \sum_{t=1}^{n} (Y_t - \beta g(X_t, \pi))^2 - n^{-1} \sum_{t=1}^{n} Y_t^2. \tag{3.1}$$

The NLS estimator $\hat{\theta}_n$ minimizes $Q_n(\theta)$ over Θ; i.e.,

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} Q_n(\theta). \tag{3.2}$$

Because the regression function is linear in β, it is convenient first to solve (3.2) for each fixed π, giving

$$\hat{\beta}_n(\pi) = \frac{\sum_{t=1}^{n} Y_t g(X_t, \pi)}{\sum_{t=1}^{n} g^2(X_t, \pi)}, \tag{3.3}$$

and then minimize the concentrated criterion function $Q_n(\pi) = Q_n(\hat{\beta}_n(\pi), \pi)$ for $\hat{\pi}_n$. The following condition is standard in extremum estimation.

Assumption 3.1. The parameter space Π of π is compact.

Following the framework of PP, in what follows we consider two possible families of g functions. These are the I-regular and the H-regular classes, and they will be discussed separately. We use the same definitions of these function classes as those in PP.

4. NLS FOR INTEGRABLE FUNCTIONS

This section considers integrable (more specially, I-regular as defined subsequently) classes of functions and examines the consistency, inconsistency, and asymptotic distributions of the NLS estimators $\hat{\beta}_n$ and $\hat{\pi}_n$ under drifting sequences of true parameters. Drifting sequences enable us to study cases where the parameters are weakly identified. We find that $\hat{\pi}_n$ and $\hat{\beta}_n$ are consistent and have an asymptotic distribution that is the same as in the strongly identified case considered in PP provided the true value of β drifts to zero at a rate slower than $n^{-1/4}$. When the true values β_n drift to zero at a faster rate, $\hat{\pi}_n$ is inconsistent, and the asymptotic distributions of $\hat{\pi}_n$ and $\hat{\beta}_n$ are nonstandard in comparison with the nonstationary limit theory of PP. Thus, weak identification is induced by a critical strip of $O\left(n^{-1/4}\right)$ around the origin in the loading coefficient β.

The following conditions are useful in the development of the limit theory. Assumption 4.1 is the same as Assumption 2.2(b) in PP except that the assumption on the characteristic function is stronger. The I-regularity conditions in Assumption 4.2 are adopted from Definition 3.3 of PP. Notice that the integrability of $|g(\cdot, \pi)|$ and T, together with the Lipschitz condition in Assumption 4.2(b), implies that $\int_{-\infty}^{\infty} T^2(x)\, dx, \int_{-\infty}^{\infty} g^2(x, \pi)\, dx < \infty$ for every $\pi \in \Pi$. Assumption 4.3

requires the function $g(\cdot,\pi)$ to be nondegenerate in the sense that $g^2(\cdot,\pi)$ has positive energy $\int_{-\infty}^{\infty} g^2(s,\pi)ds > 0$ for any $\pi \in \Pi$.

Assumption 4.1. In the generating mechanism of X_t, (2.2), $v_t = \varphi(L)\varepsilon_t = \sum_{k=1}^{\infty} \varphi_k \varepsilon_{t-k}$, with $\varphi(1) \neq 0$ and $\sum_{k=1}^{\infty} |\varphi_k| k < \infty$, and $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean zero and $\mathrm{E}|\varepsilon_t|^p < \infty$ for some $p > 4$, the distribution of which is absolutely continuous with respect to the Lebesgue measure and has characteristic function $c(\lambda)$ satisfying $\int_{-\infty}^{\infty} |c(\lambda)| d\lambda < \infty$.

Assumption 4.2. The absolutely integrable function $g(\cdot,\pi)$ is I-regular on Π in the sense that

(a) for each $\pi_1, \pi_2 \in \Pi$, there exists a bounded, integrable function $T : R \to R_+$ such that $|g(x,\pi_1) - g(x,\pi_2)| \leq |\pi_1 - \pi_2| T(x)$; and
(b) for some constants $c > 0$ and $k > 6/(p-2)$ with $p > 4$ given in Assumption 4.1, the functions g and T satisfy $|g(x,\pi) - g(y,\pi)|$, $|T(x) - T(y)| \leq c|x-y|^k$ for all $\pi \in \Pi$, piecewise on each piece S_i of the common support $S = \cup_{i=1}^{m} S_i \subset R$.

Assumption 4.3. $\int_{-\infty}^{\infty} g^2(s,\pi)ds > 0$ for all $\pi \in \Pi$.

Lemma 4.1(iii) establishes the uniform convergence of the sample covariance between the regression function and the error term. The result is similar to the second part of Theorem 3.2 in PP. But our result is stronger because the convergence in distribution to a mixed normal limit holds uniformly over the parameter space Π. The stronger result is needed in this paper because the asymptotic distribution of the covariance term contributes to the asymptotic distribution of the NLS criterion function when we allow the true value of β to drift to zero with the sample size. In the lemma, we use the local time $L(1,0) = \lim_{\varepsilon\to 0} 1/2\varepsilon \int_0^1 1\{|V(r)| < \varepsilon\} dr$ of the Brownian motion process $V(r)$, and a secondary Gaussian process $Z(\pi)$ which is independent of $L(1,0)$. Parts (i) and (ii) of Lemma 4.1 are useful in the proof of Lemma 4.1(iii).

LEMMA 4.1. *Let Assumptions 2.1, 3.1, and 4.1–4.3 hold.*

(i) For all $\pi \in \Pi$, $\sup_n \mathrm{E}\left[n^{-1/2} \sum_{t=1}^{n} g^2(X_t,\pi)\right] < \infty$,
(ii) $\sup_n \mathrm{E}\left[n^{-1/2} \sum_{t=1}^{n} T^2(X_t)\right] < \infty$, *and*
(iii) the sequence of stochastic processes $\nu_n(\pi) : \pi \in \Pi$ *converges weakly to* $\nu(\pi) : \pi \in \Pi$, *where*

$$\nu_n(\pi) = n^{-1/4} \sum_{t=1}^{n} g(X_t,\pi) u_t,$$

$$\nu(\pi) = L(1,0)^{1/2} Z(\pi),$$

and $Z(\pi)$ *is a Gaussian process with covariance kernel*

$$k(\pi_a,\pi_b) = \sigma_u^2 \int_{-\infty}^{\infty} g(s,\pi_a) g(s,\pi_b) ds.$$

This uniform convergence result makes it possible to characterize the limiting form of the NLS criterion $Q_n(\pi)$ and hence find the asymptotic distribution of $\hat{\pi}_n$. We start with the following lemma, which establishes the asymptotic distribution of the centered NLS criterion function $D_n(\pi,\pi_n) := Q_n(\pi) - Q_n(\pi_n)$ (with appropriate scaling). In this lemma and the rest of the paper, $R_{[\pm\infty]}$ denotes the extended real line: $R \cup \{-\infty,+\infty\}$.

LEMMA 4.2. *Let Assumptions 2.1, 3.1, and 4.1–4.3 hold. Under drifting sequences of true parameters* $\{(\beta_n,\pi_n) \in \Theta\}$ *such that* $(n^{1/4}\beta_n,\pi_n) \to (c,\pi_0) \in R_{[\pm\infty]} \times \Pi$, *the following limits hold:*

(i) if $c = \pm\infty$, *then*

$$n^{1/2}\beta_n^{-2} D_n(\pi,\pi_n) \to_p D_I(\pi,\pi_0)$$

$$:= \left[\int_{-\infty}^{\infty} g^2(s,\pi_0)ds - \frac{\left(\int_{-\infty}^{\infty} g(s,\pi)g(s,\pi_0)ds\right)^2}{\int_{-\infty}^{\infty} g^2(s,\pi)ds} \right] L(0,1),$$

uniformly over $\pi \in \Pi$, *and*

(ii) if $c \in R$, *then* $\{nD_n(\pi,\pi_n) : \pi \in \Pi\}$ *converges weakly to* $D(c,\pi,\pi_0) : \pi \in \Pi$, *where*

$$D(c,\pi,\pi_0) := \left\{ cL(1,0)^{1/2} \left(\int_{-\infty}^{\infty} g^2(s,\pi_0)ds \right)^{1/2} + \frac{Z(\pi_0)}{\left(\int_{-\infty}^{\infty} g^2(s,\pi_0)ds\right)^{1/2}} \right\}^2$$

$$- \left\{ cL(1,0)^{1/2} \frac{\int_{-\infty}^{\infty} g(s,\pi_0)g(s,\pi)ds}{\left(\int_{-\infty}^{\infty} g^2(s,\pi)ds\right)^{1/2}} + \frac{Z(\pi)}{\left(\int_{-\infty}^{\infty} g^2(s,\pi)ds\right)^{1/2}} \right\}^2.$$

Assumption 4.4 rules out collinearity between g(s, π_1) and g(s, π_2) for $\pi_1 \neq \pi_2$ and ensures that $D(c,\cdot,\pi_0)$ has a unique minimum in Π with probability one.

Assumption 4.4. For every $a \neq 0$ and $\pi_1, \pi_2 \in \Pi$ with $\pi_1 \neq \pi_2$

$$\int_{-\infty}^{\infty} (g(s,\pi_1) - ag(s,\pi_2))^2 ds > 0.$$

LEMMA 4.3. *Suppose Assumptions 4.2–4.4 hold. For any* $c \in R$ *and* $\pi_0 \in \Pi$, $D(c,\cdot,\pi_0)$ *is continuous and has a unique minimizer in* Π *with probability one.*

We are now in a position to develop a limit distribution theory. Theorem 4.1 characterizes the limit behavior of $\hat{\pi}_n$ under different sequences of drifting true parameters. The outcomes depend critically on the limit behavior of β_n. If $n^{1/4}\beta_n$ is bounded as $n \to \infty$ then the data are insufficiently informative to deliver a consistent estimator, and $\hat{\pi}_n$ converges weakly to a random quantity, reflecting that lack of information. If $n^{1/4}\beta_n$ diverges, then there is sufficient information for consistent estimation. In that event, the rate of convergence of $\hat{\pi}_n$ is $n^{1/4}\beta_n$ and depends on the sequence β_n, as shown in Theorem 4.2.

THEOREM 4.1. *Suppose Assumptions 2.1, 3.1, and 4.1–4.4 hold. Under drifting sequences of true parameters* $\{(\beta_n,\pi_n)\in\Theta\}$ *such that* $\pi_n\to\pi_0$ *and* $n^{1/4}\beta_n\to c$ *for* $c\in R_{[\pm\infty]}$*, the following limits hold:*

(i) if $c=\pm\infty$*, then* $\hat{\pi}_n-\pi_n\to_p 0$*, and*

(ii) if $c\in R$*, then* $\hat{\pi}_n\to_d \tau_{I,\pi}(c,\pi_0)$*, where* $\tau_{I,\pi}(c,\pi_0)$ *is a random variable that minimizes* $D(c,\pi,\pi_0)$*.*

The following assumption imposes an I-regularity condition on the first and second derivatives of g with respect to π. To simplify notation, let $\dot{g}(x,\pi)=\partial g(x,\pi)/\partial\pi$ and $\ddot{g}(x,\pi)=\partial^2 g(x,\pi)/\partial\pi^2$. Assumption 4.5(b) implies that the matrix $\Sigma_{g\dot{g}}$ defined in (4.1) in Theorem 4.2 is positive definite.

Assumption 4.5.

(a) The functions $\dot{g}(\cdot,\pi)$ and $\ddot{g}(\cdot,\pi)$ are I-regular on Π; i.e., they satisfy Assumption 4.2, and

(b) for any $\pi\in\Pi$, there exists no $a\in R$ such that $\dot{g}(x,\pi)=ag(x,\pi)$ almost everywhere

Remark. Part (b) of the assumption is a rank condition. It typically holds if π and β are separately identifiable, which rules out formulations such as $\beta g(x;\pi)=\beta e^{\pi}f(x)$ in which the single parameter βe^{π} is strongly identified and conventional nonstationary limit theory for estimation of π applies (Park and Phillips, 1999).

Theorem 4.2 gives the asymptotic distribution of $\hat{\pi}_n$ when $n^{1/4}\beta_n\to c\in R_{[\pm\infty]}$.

THEOREM 4.2. *Suppose Assumptions 2.1, 3.1, and 4.1–4.5 hold. Under drifting sequences of true parameters* $\{(\beta_n,\pi_n)\in\Theta\}$ *such that* $\pi_n\to\pi_0$ *and* $n^{1/4}\beta_n\to c$*, the following limit behavior obtains:*

(i) if $c\in R$*, then* $n^{1/4}\hat{\beta}_n\to_d \tau_{I,\beta}(c,\pi_0)\equiv f_I(\tau_{I,\pi}(c,\pi_0))$*, where*

$$f_I(\pi):=\frac{\sigma_u\left(\int_{-\infty}^{\infty}g^2(s,\pi)ds\right)^{1/2}Z(\pi)+cL^{1/2}(1,0)\int_{-\infty}^{\infty}g(s,\pi)g(s,\pi_0)ds}{L^{1/2}(1,0)\int_{-\infty}^{\infty}g^2(s,\pi)ds},\ \textit{and}$$

(ii) if $c=\pm\infty$*,*

$$\begin{pmatrix} n^{1/4}(\hat{\beta}_n-\beta_n)\\ n^{1/4}\beta_n(\hat{\pi}_n-\pi_n)\end{pmatrix}\to_d\begin{pmatrix} T_{I,\beta}(\pi_0)\\ T_{I,\pi}(\pi_0)\end{pmatrix}:=\sigma_u\Sigma_{g\dot{g}}^{-1/2}L^{-1/2}(1,0)Z,$$

where $Z\sim N(0,I_2)$ *is independent of* $L(1,0)$*, and*

$$\Sigma_{g\dot{g}}:=\begin{pmatrix}\int_{-\infty}^{\infty}g^2(s,\pi_0)ds & \int_{-\infty}^{\infty}\dot{g}(s,\pi_0)g(s,\pi_0)ds\\ \int_{-\infty}^{\infty}\dot{g}(s,\pi_0)g(s,\pi_0)ds & \int_{-\infty}^{\infty}\dot{g}^2(s,\pi_0)ds\end{pmatrix}. \tag{4.1}$$

5. NLS FOR ASYMPTOTICALLY HOMOGENEOUS FUNCTIONS

This section considers asymptotically homogeneous (or H-regular) classes of functions and examines the consistency, inconsistency, and asymptotic distributions of the NLS estimators $\hat{\beta}_n$ and $\hat{\pi}_n$ under drifting sequences of true parameters. We find that $\hat{\pi}_n$ and $\hat{\beta}_n$ are consistent and have asymptotic distributions that are equivalent to those in PP when the true values of β drift to zero at a rate slower than $n^{1/2}$ times the asymptotic order of the nonlinear function g. When the true values β_n drift to zero faster, $\hat{\pi}_n$ is inconsistent, and the asymptotic distributions of $\hat{\pi}_n$ and $\hat{\beta}_n$ are again nonstandard in relation to PP. Weak identification in the present case occurs when the loading coefficient β lies in a critical strip around the origin whose order of magnitude depends on the asymptotic order of the function g.

To simplify notation, define the standardized quantity $X_{n,t} = n^{-1/2}X_t$. For a function $F(v,\pi)$, let $\int F(V,\pi)dU = \int_0^1 F(V(r),\pi)dU(r)$ and $\int F(V,\pi) = \int_0^1 F(V(r),\pi)dr$.

Assumption 5.1.

(a) $g(x,\pi)$ is H-regular on Π as defined in PP, with asymptotic order $\kappa(\lambda,\pi)$, limit homogeneous function $h(x,\pi)$, and residual $R(x,\lambda,\pi)$, where $\lambda \in R_+$. Let

$$h^*(x,\lambda,\pi) = \kappa^{-1}(\lambda,\pi)g(\lambda x,\pi) \equiv h(x,\pi) + \kappa^{-1}(\lambda,\pi)R(x,\lambda,\pi), \quad \textbf{(5.1)}$$

where $\kappa^{-1}(\lambda,\pi)R(x,\lambda,\pi) = o(1)$ for all $\pi \in \Pi$ as $\lambda \to \infty$.

(b) There exists a function $b: R \to R^+$ such that for all $x \in R$ and $\pi,\pi' \in \Pi$,

$$\sup_{\lambda \geq 1} \left| h^*(x,\lambda,\pi) - h^*(x,\lambda,\pi') \right| \leq b(x) \left| \pi - \pi' \right|.$$

(c) For all $\pi \in \Pi$ and $\delta > 0$, $\int_{|s|\leq\delta} h^2(s,\pi)ds > 0$.

(d) For $\pi \neq \pi'$ and $\delta > 0$, there is no $a \neq 0$ such that $\int_{|s|\leq\delta}(h(s,\pi) - ah(s,\pi'))^2 ds = 0$.

(e) $\lim_{\lambda\to\infty} \sup_{\pi\in\Pi} \kappa^{-1}(\lambda,\pi) = 0$.

Remarks.

(a) The H-regularity concept in Assumption 5.1(a) was introduced in Park and Phillips (1999) and is illustrated in what follows. The definition includes a wide class of homogeneous, asymptotically homogeneous, and regularly varying functions and is discussed in PP. Assumption 5.1(b) is a Lipschitz continuity condition on $h^*(x,\lambda,\pi)$. The $\sup_{\lambda\geq 1}$ operation does not make the assumption more restrictive because $h^*(x,\lambda,\pi)$ converges to $h(x,\pi)$ as λ goes to infinity. For the same reason, Assumption 5.1(b) implies that $\left|h(x,\pi) - h(x,\pi')\right| \leq b(x)\left|\pi - \pi'\right|$ for all $x \in R$ and $\pi,\pi' \in \Pi$. Assumptions 5.1(c) and (d) guarantee the identification of β_0 and that of π_0 when

β_0 is not too close to zero. These assumptions along with Assumption 5.4 are the full-rank conditions.

(b) Assumption 5.1 is not very restrictive if g is smooth in π. An example is given next that satisfies parts (a)–(e). This assumption, along with Assumption 5.3, does rule out nonsmooth g functions, which is an interesting topic that is not covered by this paper.

The following example involves a typical asymptotically homogeneous function and demonstrates that Assumption 5.1 is not very restrictive.

Example
Let $g(x,\pi) = \left(1+x^2\right)^{\pi}$ and $\Pi = [\pi_a, \pi_b]$ with $0 < \pi_a < \pi_b < \infty$. Then,

$$g(\lambda x,\pi) = \lambda^{2\pi}\left(\lambda^{-2}+x^2\right)^{\pi} := \kappa(\lambda,\pi)\, h^*(x,\lambda,\pi), \quad \text{with } \kappa(\lambda,\pi) = \lambda^{2\pi}. \tag{5.2}$$

Clearly, $\inf_{\pi\in\Pi}\kappa(\lambda,\pi) = \lambda^{2\pi_a} \to \infty$ as $\lambda\to\infty$, the family $\{g(\cdot,\pi)\}$ is equicontinuous on Π, and $h(x,\pi) = x^{2\pi}$, which is homogeneous of order $\lambda^{2\pi}$ with $\int_{|s|\le\delta} s^{4\pi}ds > 0$ and $\int_{|s|\le\delta}(s^{2\pi} - s^{2\pi'})^2 ds > 0$ for all $\delta > 0$. The following equation implies that $g(x,\pi)$ satisfies Assumption 5.1(a):

$$\lim_{\lambda\to\infty}\sup_{|x|<C,\pi\in\Pi}\left|\left(\lambda^{-2}+x^2\right)^{\pi} - x^{2\pi}\right| = 0 \quad\text{and}\quad \sup_{|x|<C,\pi\in\Pi}\left|x^{2\pi}\right| < C^{2\pi_b}\vee 1 < \infty. \tag{5.3}$$

Assumption 5.1(b) holds because

$$\begin{aligned}\sup_{\lambda\ge 1}\left|\left(\lambda^{-2}+x^2\right)^{\pi} - \left(\lambda^{-2}+x^2\right)^{\pi'}\right| &= \sup_{\lambda\ge 1}\left|\left(\lambda^{-2}+x^2\right)^{\tilde\pi}\log\left(\lambda^{-2}+x^2\right)(\pi-\pi')\right| \\ &\le \left[\left(1+x^2\right)^{\pi_b}\left\{\log\left(1+x^2\right)\right.\right. \\ &\quad \left.\left.+\log\left(1+x^{-2}\right)\right\}\right]\left|\pi-\pi'\right|,\end{aligned} \tag{5.4}$$

where the equality holds for $\tilde\pi$ between π and π' by the mean value expansion and the inequality holds because

$$\sup_{\lambda\ge 1}\left(\lambda^{-2}+x^2\right)^{\tilde\pi} \le \left(1+x^2\right)^{\pi_b}$$

and

$$\begin{aligned}\sup_{\lambda\ge 1}\left|\log\left(\lambda^{-2}+x^2\right)\right| &\le \left|\log\left(1+x^2\right)\right| 1\{|x|\ge 1\} + \left|\log x^2\right| 1\{|x|<1\} \\ &\le \left|\log\left(1+x^2\right)\right| + \left|\log\left(1+x^{-2}\right)\right|.\end{aligned}$$

Assumptions 5.1(c) and (d) hold straightforwardly. Finally, we verify the validity of two additional conditions needed in later arguments. First, observe that

$$\frac{\kappa\left(n^{1/2},\pi_n\right)}{\kappa\left(n^{1/2},\pi_n'\right)} \sim n^{\pi_n-\pi_n'} \to 1, \quad \text{for } \pi_n-\pi_n' = o\left(\frac{1}{\log n}\right),$$

confirming a condition needed in Theorem 5.2. Next, the derivative function $\dot{g}(x,\pi) = \left(1+x^2\right)^{\pi}\log\left(1+x^2\right)$, whose asymptotic order is $\kappa_1(\lambda,\pi) = \lambda^{2\pi}\ln\lambda$, so that

$$\limsup_{\lambda\to\infty}\left(\frac{\kappa(\lambda,\pi)}{\kappa_1(\lambda,\pi)}\ln\lambda\right) = 1,$$

confirming the validity of a condition used in Assumption 5.4(b).

Assumption 5.2 places a uniform boundedness condition on the second moments of the limit homogeneous function h and the Lipschitz function b of Assumption 5.1.

Assumption 5.2.

(a) For all $\pi \in \Pi$, $\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} \mathrm{E}h^2(X_{n.t},\pi) < \infty$,
(b) $\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} \mathrm{E}b^2(X_{n,t}) < \infty$, and
(c) $\sup_{r\in[[0,1]}\mathrm{E}b^2(V(r)) < \infty$.

Remark. Parts (a) and (b) are helpful in establishing the stochastic equicontinuity of $n^{-1/2}\kappa^{-1}(n^{1/2},\pi)\sum_{t=1}^{n} g(X_t,\pi)u_t$. Part (c) is used to guarantee the existence of a random process $Y(\pi):\pi\in\Pi$ whose sample paths are continuous with probability one and satisfies $Y(\pi) = \int h(V,\pi)dU$ a.s. for every $\pi\in\Pi$. Lemma 5.1 formalizes the existence argument. But before presenting Lemma 5.1, we give a set of sufficient conditions for Assumption 5.2 because Assumption 5.2 itself, though convenient for proofs, may be hard to verify.[1] The lemma that follows is named SC5.2 to make it clear that it gives the "sufficient conditions for Assumption 5.2."

LEMMA SC5.2. *Suppose that*

(i) $b(x)$ *is symmetric and* $b(|x|)$ *is increasing in* $|x|$,
(ii) $\mathrm{E}b^2(|V(1)|+c) < \infty$, *for some* $c > 0$,
(iii) either b *is bounded or* $|X_{n,t}|$ *is first-order stochastically dominated by* $|V(1)|+c$ *for large enough* n, *and*
(iv) there exists a $\pi\in\Pi$, *such that*

$$\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n}\mathrm{E}h^2(X_{n.t},\pi) < \infty.$$

Then Assumption 5.2 holds.

Proof. First, it is immediate that (i) and (ii) imply Assumption 5.2(c) (by observing that $V(r) =_d \sqrt{r}V(1)$. Assumption 5.2(b) is implied by (i)–(iii) because $n^{-1}\sum_{t=1}^{n} b^2(X_{n,t}) \le \int_0^1 \left[b^2(|V(r)|+c)\right] dr$ for some version of $V(r)$ for large enough n and the right-hand side of the inequality has finite expectation. Assumption 5.2(a) is immediately implied by (iv), Assumption 5.1(b), and Assumption 5.2(b). ∎

Remark. Verifying the sufficient conditions in the preceding lemma is relatively straightforward because (i) b can always be chosen to be symmetric and monotonic on R_+,[2] (ii) $V(1) \sim N\left(0, \sigma_v^2\right)$, (iii) the boundedness of b is easy to check, and (iv) in many applications, $h^2(X_{n.t}, \pi)$ is a constant when $\pi = 0$. The only condition that can be hard to check is the stochastic domination of $|X_{n,t}|$. But we argue that it is often possible to guarantee this by imposing restrictions on the distribution of v_t and the correlation between them. We can also verify this condition empirically by inspecting the distribution of $X_{n,t}$ directly.

LEMMA 5.1. *Let Assumptions 3.1(a) and (b) and 5.2(c) hold. Then, there exists a random process $Y(\pi) : \pi \in \Pi$ that*

(i) has continuous sample paths with probability one and

(ii) satisfies $Y(\pi) = \int h(V,\pi)dU$ a.s. for every $\pi \in \Pi$.

Remark. Random processes indexed by π that satisfy (ii) in the preceding lemma are not necessarily unique (not even in an almost sure sense). That is, there may exist $Y(\pi)$, $Y'(\pi) : \pi \in \Pi$ that both satisfy (ii), but $Y(\pi) \neq Y'(\pi)\ \forall \pi \in \Pi$ a.s. However, under the given assumptions, the random process $Y(\pi)$ that satisfies both (i) and (ii) is unique in an almost sure sense.[3] To keep the notation intuitive, we let $\int h(V,\pi)dU : \pi \in \Pi$ denote the unique continuous process $Y(\pi)$ in Lemma 5.1. This should cause no confusion because previously the stochastic integral $\int h(V,\pi)dU$ was defined only for each $\pi \in \Pi$ and not as a random process indexed by π.

Lemma 5.2 establishes the uniform convergence of the sample covariance between the regression function and the error term. As in the case of integrable functions, the result is similar to the second part of Theorem 3.3 in PP but is stronger because the convergence holds uniformly over the parameter space. As before, the stronger result is needed here because the probability limit of the covariance term contributes to the asymptotic form of the NLS criterion function when we allow the true value of β to drift to zero as the sample size $n \to \infty$. The resulting uniform convergence to a parameterized stochastic integral is new and seems likely to be useful in other asymptotics involving nonstationary time series.

LEMMA 5.2. *Let Assumptions 2.1, 3.1, 5.1, and 5.2 hold. Then, uniformly in $\pi \in \Pi$,*

$$n^{-1/2}\kappa^{-1}(n^{1/2},\pi)\sum_{t=1}^{n} g(X_t,\pi)u_t \to_p \int h(V,\pi)dU.$$

As discussed earlier, we consider drifting sequences of true parameters $\{(\beta_n,\pi_n)\in\Theta\}$ such that $\kappa(n^{1/2},\pi_n)n^{1/2}\beta_n\to c$ for $c\in R_{[\pm\infty]}$. The rate $\kappa(n^{1/2},\pi_n)n^{1/2}$ is set so that, under the sequence $\{(\beta_n,\pi_n)\in\Theta\}$, the centered criterion function $D_n(\pi,\pi_n):=Q_n(\pi)-Q(\pi_n)$, when scaled properly, converges in probability to one function when $c=\pm\infty$ and to another function when $c\in R$. Lemma 5.3 establishes the respective probability limits.

LEMMA 5.3. *Let Assumptions 2.1, 3.1, 5.1, and 5.2 hold. Then under drifting sequences of true parameters $\{(\beta_n,\pi_n)\in\Theta\}$ such that $\pi_n\to\pi_0\in\Pi$ and $\kappa(n^{1/2},\pi_n)n^{1/2}\beta_n\to c\in R_{[\pm\infty]}$, the following limits hold:*

(i) if $c=\pm\infty$, $\kappa^{-2}(n^{1/2},\pi_n)\beta_n^{-2}D_n(\pi,\pi_n)\to D_H(\pi,\pi_0)$ a.s. uniformly over $\pi\in\Pi$ where

$$D_H(\pi,\pi_0):=\int h^2(V,\pi_0)-\frac{\left[\int h(V,\pi)h(V,\pi_0)\right]^2}{\int h^2(V,\pi)}.$$

(ii) If $c\in R$, then uniformly over $\pi\in\Pi$,

$$nD_n(\pi,\pi_n)\to_p \frac{\left[c\int h^2(V,\pi_0)+\int h(V,\pi_0)dU\right]^2}{\int h^2(V,\pi_0)} -\frac{\left[c\int h(V,\pi)h(V,\pi_0)+\int h(V,\pi)dU\right]^2}{\int h^2(V,\pi)}.$$

Lemma 5.4 shows that the probability limit of $nD_n(\pi,\pi_n)$ has a unique minimum with probability one, which guarantees that $\hat\pi_n$ has a well-defined limiting distribution.

LEMMA 5.4. *Let Assumptions 5.1 and 5.2 hold. For any $\pi_0\in\Pi$ and $c\in R$, the limit function*

$$\frac{\left[c\int h(V,\pi)h(V,\pi_0)+\int h(V,\pi)dU\right]^2}{\int h^2(V,\pi)} \tag{5.5}$$

is continuous in π and achieves a unique maximum in Π with probability one.

The theorem that follows establishes the consistency of $\hat\pi_n$ under drifting sequences of true parameters $\{(\beta_n,\pi_n)\in\Theta\}$ with $\kappa(n^{1/2},\pi_n)n^{1/2}\beta_n\to\pm\infty$ and gives the distributional limit of $\hat\pi_n$ under drifting sequences with $\kappa(n^{1/2},\pi_n)n^{1/2}$ $\beta_n\to c\in R$. In the latter case, there is insufficient information in the limit to ensure consistency, and $\hat\pi_n$ converges to a random quantity reflecting that lack of information.

THEOREM 5.1. *Let Assumptions 2.1, 3.1, 5.1, and 5.2 hold. Under drifting sequences of true parameters$\{(\beta_n,\pi_n)\in\Theta\}$ such that $\pi_n\to\pi_0\in\Pi$ and*

$\kappa(n^{1/2}, \pi_n)n^{1/2}\beta_n \to c \in R_{[\pm\infty]}$, *the following limits hold:*

(i) if $c = \pm\infty$, *then* $\hat{\pi}_n - \pi_n \to_p 0$, *and*

(ii) if $c \in R$, *then* $\hat{\pi}_n \to_d \tau_{H,\pi}(c, \pi_0)$, *where* $\tau_{H,\pi}(c, \pi_0)$ *is a random variable that maximizes (5.5).*

Assumption 5.3 requires both the derivative functions $\dot{g}(x, \pi)$ and $\ddot{g}(x, \pi)$ to satisfy H-regularity conditions. These assumptions are needed to obtain the asymptotic distributions of the NLS estimators, and their asymptotic forms affect convergence rates.

Assumption 5.3.

(a) $\dot{g}(x, \pi), \pi \in \Pi$ is H-regular with asymptotic order $\kappa_1(\lambda, \pi)$, limit homogeneous function $h_1(x, \pi)$, and residual $R_1(x, \lambda, \pi)$,

(b) $\ddot{g}(x, \pi), \pi \in \Pi$ is H-regular with asymptotic order $\kappa_2(\lambda, \pi)$, limit homogeneous function $h_2(x, \pi)$, and residual $R_2(x, \lambda, \pi)$, and

(c) for $h_1^*(x, \lambda, \pi) = \kappa_1^{-1}(\lambda, \pi)\dot{g}(\lambda x, \pi)$ and $h_2^*(x, \lambda, \pi) = \kappa_2^{-1}(\lambda, \pi)\ddot{g}$ $(\lambda \mathrm{x} \pi)$, Assumptions 5.1(b) and 5.2 hold with h replaced by h_1 or h_2 and b replaced by b_1 or b_2.

Assumption 5.4(a) is part of the full-rank condition. Assumption 5.4(b) requires the asymptotic order of $\dot{g}$ to be larger than that of g by a certain factor. Part (b) is satisfied by most asymptotically homogeneous functions.

Assumption 5.4.

(a) For any $\pi \in \Pi$ and $\delta > 0$, there is no $a \neq 0$ such that $\int_{|s|\leq\delta}(h(s, \pi) - ah_1(s, \pi))^2 ds = 0$, and

(b) for any $\pi \in \Pi$, $\limsup_{\lambda\to\infty} |\kappa(\lambda, \pi)\kappa_1^{-1}(\lambda, \pi)| \log\lambda < \infty$.

Theorem 5.2 establishes the asymptotic distributions of the estimators under drifting sequences of true parameters. As the theorem shows, the estimators have the same asymptotic distributions as in Theorem 5.2 of PP when identification is strong—i.e., when $\kappa(n^{1/2}, \pi_n)n^{1/2}|\beta_n| \to \infty$. When identification is weak, the estimators have asymptotic distributions different from those given in PP.

For notational simplicity, let $\kappa_{n,\pi} = \kappa(n^{1/2}, \pi)$, $\kappa_{1,n,\pi} = \kappa_1(n^{1/2}, \pi)$, and $\kappa_{2,n,\pi} = \kappa_2(n^{1/2}, \pi)$.

THEOREM 5.2. *Suppose Assumptions 2.1, 3.1, and 5.1–5.4 hold. Under drifting sequences of true parameters* $\{(\beta_n, \pi_n) \in \Theta\}$ *such that* $\pi_n \to \pi_0 \in \Pi$ *and* $n^{1/2}\kappa_{n,\pi_n}\beta_n \to c \in R_{[\pm\infty]}$, *the following limits hold:*

(i) if $c \in R$, *then* $n^{1/2}\kappa_{n,\hat{\pi}_n}\hat{\beta}_n \to_p \tau_{H,\beta}(c, \pi_0) := f_H(\tau_{H,\pi}(c, \pi_0))$, *where*

$$f_H(\pi) := \frac{\int h(V, \pi)dU + c\int h(V, \pi)h(V, \pi_0)}{\int h^2(V, \pi)}. \tag{5.6}$$

(ii) if $c = \pm\infty$*, then* $n^{1/2}\beta_n\kappa_{1,n,\pi_n}(\hat{\pi}_n - \pi_n) \to_p T_{H,\pi}(\pi_0)$ *where*

$$T_{H,\pi} := \frac{\int h(V,\pi_0)h_1(V,\pi_0)\int h(V,\pi_0)dU - \int h^2(V,\pi_0)\int h_1(V,\pi_0)dU}{\int h_1^2(V,\pi_0)\int h^2(V,\pi_0) - \left[\int h(V,\pi_0)h_1(V,\pi_0)\right]^2}.$$

(iii) if $c = \pm\infty$ *and in addition,* $\kappa_{n,\pi_n}/\kappa_{n,\pi_n'} \to 1$ *whenever* $\pi_n - \pi_n' = o(1/\log n)$*. then* $n^{1/2}\kappa_{n,\pi_n}(\hat{\beta}_n - \beta_n) \to_p T_{H,\beta}(\pi_0)$*, where*

$$T_{H,\beta}(\pi_0) := \frac{\int h(V,\pi_0)dU}{\int h^2(V,\pi_0)} - \frac{\int h(V,\pi_0)h_1(V,\pi_0)}{\int h^2(V,\pi_0)} \times T_{H,\pi}(\pi_0).$$

These results, like those for integrable functions, reveal that the limit theory is affected by weak identification. In the present case, there is the additional complication that the convergence rates depend on the unknown parameters. A robust approach to inference needs to take account of these possibilities, which we now investigate.

6. CONFIDENCE INTERVALS

This section shows how to construct confidence intervals for the loading coefficient β and the nonlinear transformation parameter π. These intervals are robust in the sense that they allow for the possibility that identification may be weak. The approach is based on Theorems 4.2 and 5.2. The I-regular and the H-regular classes are treated separately. Special issues arise for the H-regular class because the drifting rate of the true values of β depends on the true values of the unknown parameter π.

We proceed in a general way and let γ be a generic notation for the relevant parameter and j denote a generic type of nonlinear transformation. In our model, γ may be either β or π, and j may be either I, standing for integrable type, or H, standing for asymptotically homogeneous type. Let $CI_{j,\gamma,n}(\alpha)$ denote the $1-\alpha$ percent confidence interval for parameter γ when the nonlinear transformation is of type j. For $\theta = (\beta, \pi)'$, let $\Pr_\theta$ be the probability function when the true parameter value is θ. At sample size n, the coverage probability of the confidence interval $CI_{j,\gamma,n}(1-\alpha)$ when the true parameter is θ is

$$CP_{j,\gamma,n}(\theta,\alpha) = \Pr_\theta(\gamma \in CI_{j,\gamma,n}(\alpha)). \quad \textbf{(6.1)}$$

This section constructs confidence intervals whose finite-sample coverage probabilities are uniformly controlled by the asymptotic size. The asymptotic size of $CI_{j,\gamma,n}$ is defined as

$$AsySZ_{j,\gamma}(\alpha) = \liminf_{n\to\infty} \inf_{\theta\in\Theta} CP_{j,\gamma,n}(\theta,\alpha). \quad \textbf{(6.2)}$$

As discussed earlier in this paper, the true parameter β measures the strength of identification. In the definition of $AsySZ_{j,\gamma}$, the infimum is taken over all $\theta \in \Theta$

and, in particular, over $\beta \in R$. Thus, $AsySZ_{j,\gamma}(\alpha)$ approximates the finite-sample minimum coverage probability $\inf_{\theta\in\Theta} CP_{j,\gamma,n}(\theta,\alpha)$ irrespective of the strength of identification.

6.1. Confidence Intervals with Integrable Functions

The confidence intervals for both β and π are constructed in a two-step fashion. First, one determines the strength of identification by comparing $n^{1/4}|\hat{\beta}_n|$ to a positive number b_n. Second, one chooses critical values based on the asymptotic distribution of $n^{1/4}(\hat{\beta}_n - \beta)$ or $n^{1/4}\hat{\beta}_n(\hat{\pi}_n - \pi)$ at different levels of identification. Details are given subsequently. We require the sequence b_n to diverge to infinity but at a rate slower than $n^{1/4}$.

Assumption 6.1. $b_n^{-1} + n^{-1/4} b_n \to 0$.

Consider $\alpha \in (0,1)$. For $c \in R$, let $q_{I,\beta}(c,\pi_0,1-\alpha)$ be the $1-\alpha$ quantile of $|\tau_{I,\beta}(c,\pi_0) - c|$. Let $q_{I,\beta}(\infty,\pi_0,1-\alpha)$ be the $1-\alpha$ quantile of $|T_{I,\beta}(\pi_0)|$. Let

$$\hat{q}_{I,\beta}(\hat{\pi}_n, 1-\alpha) = \begin{cases} \sup_{c\in R_{[\pm\infty]}} \sup_{\pi\in\Pi} q_{I,\beta}(c,\pi,1-\alpha) & \text{if } n^{1/4}|\hat{\beta}_n| \le b_n \\ q_{I,\beta}(\infty,\hat{\pi}_n,1-\alpha) & \text{if } n^{1/4}|\hat{\beta}_n| > b_n \end{cases}. \quad \textbf{(6.3)}$$

We use $\hat{q}_{I,\beta}(\hat{\pi}_n, 1-\alpha)$ as the critical value to construct a confidence interval for β. This critical value is structured the same as that used in the robust confidence interval in Cheng (2008). The confidence interval for β is

$$CI_{I,\beta,n}(\alpha) = \left\{\beta : n^{1/4}|\hat{\beta}_n - \beta| \le \hat{q}_{I,\beta}(\hat{\pi}_n, 1-\alpha)\right\}. \quad \textbf{(6.4)}$$

Similarly, let $q_{I,\pi}(c,\pi_0,1-\alpha)$ be the $1-\alpha$ quantile of $|\tau_{I,\beta}(c,\pi_0)(\tau_{I,\pi}(c,\pi_0) - \pi_0)|$. Let $q_{I,\pi}(\infty,\pi_0,1-\alpha)$ be the $1-\alpha$ quantile of $|T_{I,\pi}(\pi_0)|$. Let

$$\hat{q}_{I,\pi}(\hat{\pi}_n, 1-\alpha) = \begin{cases} \sup_{c\in R_{[\pm\infty]}} \sup_{\pi\in\Pi} q_{I,\pi}(c,\pi,1-\alpha) & \text{if } n^{1/4}|\hat{\beta}_n| \le b_n \\ q_{I,\pi}(\infty,\hat{\pi}_n,1-\alpha) & \text{if } n^{1/4}|\hat{\beta}_n| > b_n \end{cases}. \quad \textbf{(6.5)}$$

The confidence interval for π is

$$CI_{I,\pi,n}(\alpha) = \left\{\pi \in \Pi : n^{1/4}|\hat{\beta}_n(\hat{\pi}_n - \pi)| \le \hat{q}_{I,\pi}(\hat{\pi}_n, 1-\alpha)\right\}. \quad \textbf{(6.6)}$$

Notice that the confidence interval of π is wide when $\hat{\beta}_n$ is small, reflecting circumstances in which π is only weakly identified.

The following theorem shows that these confidence intervals have the correct asymptotic size.

THEOREM 6.1. *Suppose Assumptions 2.1, 3.1, 4.1– 4.5 and 6.1 hold. Then for all* $\alpha \in (0,1)$,

(i) $AsySZ_{I,\beta}(\alpha) = \alpha$, *and*

(ii) $AsySZ_{I,\pi}(\alpha) = \alpha$.

6.2. Confidence Intervals with Asymptotically Homogeneous Functions

The confidence interval for π is constructed in the same way as in the previous section. The confidence interval for β has a different form because the test statistic for β, $n^{1/2}\kappa_{n,\hat{\pi}_n}(\hat{\beta}_n - \beta_n)$, does not necessarily converge in distribution when $n^{1/2}\kappa_{n,\pi_n}\beta_n \to c \in R$. In fact, $n^{1/2}\kappa_{n,\hat{\pi}_n}(\hat{\beta}_n - \beta_n)$ may diverge with positive probability because $n^{1/2}\kappa_{n,\hat{\pi}_n}\beta_n$ may diverge when $\hat{\pi}_n > \pi_n$, which happens with positive probability. We therefore construct a confidence interval for β based on the confidence interval for π, as discussed in detail subsequently.

The sequence b_n serves the same purpose as in the previous section, but the divergence rate of b_n is required to be different. The reason is that the drifting sequences of true values of β may drift to zero at a different rate for asymptotically homogeneous functions than for integrable functions and this rate may depend on π. The rate requirement on b_n is stated in the following assumption.

Assumption 6.2. For all $\pi \in \Pi$, $b_n^{-1} + n^{-1/2}\kappa_{n,\pi}^{-1} b_n \to 0$.

Remark. For typical asymptotically homogeneous functions the order function satisfies $\inf_\pi \kappa_{n,\pi} \geq \varepsilon > 0$. In the example considered earlier, the order function is $\kappa_{n,\pi} = n^{2\pi}$ and $\inf_\pi \kappa_{n,\pi} = n^{2\pi_a}$ with $\pi_a > 0$, so that $\liminf_{n\to\infty} \inf_\pi \kappa_{n,\pi} = \infty$. In such cases, Assumption 6.2 is satisfied as long as $b_n^{-1} + n^{-1/2} b_n \to 0$.

For $c \in R$, let $q_{H,\pi}(c, \pi_0, 1-\alpha)$ be the $1-\alpha$ quantile of $|\tau_{H,\beta}(c,\pi_0)(\tau_{H,\pi}(c,\pi_0) - \pi_0)|$. Let $q_{H,\pi}(\infty, \pi_0, 1-\alpha)$ be the $1-\alpha$ quantile of $|T_{H,\pi}(\pi_0)|$. Let

$$\hat{q}_{H,\pi}(\hat{\pi}_n, 1-\alpha) = \begin{cases} \alpha\kappa_{n,\hat{\pi}_n}^{-1}\kappa_{1,n,\hat{\pi}_n} \sup_{c\in R_{[\pm\infty]}} \sup_{\pi\in\Pi} q_{H,\pi}(c,\pi_0,1-\alpha) & \text{if } n^{1/2}|\kappa_{n,\hat{\pi}_n}\hat{\beta}_n| \leq b_n \\ q_{H,\pi}(\infty,\pi_0,1-\alpha) & \text{if } n^{1/2}|\kappa_{n,\hat{\pi}_n}\hat{\beta}_n| > b_n \end{cases}. \tag{6.7}$$

The confidence interval for π is

$$CI_{H,\pi,n}(\alpha) = \left\{\pi : n^{1/2}|\kappa_{1,n,\hat{\pi}_n}\hat{\beta}_n(\hat{\pi}_n - \pi)| \leq \hat{q}_{H,\pi}(\hat{\pi}_n, 1-\alpha)\right\}. \tag{6.8}$$

Let $q_{H,\beta}(\infty, \pi_0, 1-\alpha)$ be the $1-\alpha$ quantile of $|T_{H,\beta}(\pi_0)|$. Define the set $CI_n(\alpha) = \{\beta : n^{1/2}|\kappa_{n,\hat{\pi}_n}(\hat{\beta}_n - \beta_n)| \leq q_{H,\beta}(\infty, \hat{\pi}_n, 1-\alpha)\}$.

Then, the confidence interval for β is

$$CI_{H,\beta,n}(\alpha) = \begin{cases} cCI_n(\alpha) \cup \left\{\beta : \inf_{\pi\in CI_{H,\pi,n}(\alpha)} b_n^{-1} n^{1/2}|\kappa_{n,\pi}\beta| \leq 1\right\} & \text{if } n^{1/2}|\kappa_{n,\hat{\pi}_n}\hat{\beta}_n| \leq b_n \\ CI_n(\alpha) & \text{if } n^{1/2}|\kappa_{n,\hat{\pi}_n}\hat{\beta}_n| > b_n \end{cases}. \tag{6.9}$$

The following theorem shows that these confidence intervals have the correct asymptotic size.

THEOREM 6.2. *Suppose Assumptions 2.1, 3.1, 5.1–5.4, and 6.2 hold. Then for all* $\alpha \in (0, 1)$,

(i) $AsySZ_{H,\pi}(\alpha) = \alpha$, *and*
(ii) $AsySZ_{H,\beta}(\alpha) = \alpha$.

7. CONCLUSION

This work develops a local limit theory for NLS estimation under drifting parameter sequences that allow for the possibility of weak identification in a nonlinear cointegrating regression relationship. Such models are important empirically in situations where outcomes may be mildly impacted by certain stochastically nonstationary variables. One example is financial asset returns, which may be influenced in the long run by stochastic trends in economic fundamentals although these trend effects are nearly imperceptible in the short term. Another example is microeconomic behavior, which may be impacted in a minor way by common macroeconomic effects or aggregate economic fundamentals (e.g., Granger, 1987; Giacomini and Granger, 2004), whereas the dominant effects involve individual characteristics.

The model that is analyzed in this paper is a prototypical model of this type. The model allows for the following two features: (a) a regressor that is a nonlinear transformation of an integrated time series, so that the model is cointegrating; and (b) potentially weak cointegrating effects (in terms of a loading coefficient for these effects), so that the parameter in the nonlinear transformation is only weakly identified. We use the local limit theory derived here to construct confidence intervals for both the loading coefficient and the transformation parameter. The confidence intervals are shown to have correct asymptotic size irrespective of the strength of identification. The results of the paper can therefore be used to carry out robust inference on weakly cointegrated systems and to construct robust prediction intervals that allow for the presence of weak effects from stochastic trends.

NOTES

1. We thank an anonymous referee for suggesting these conditions.
2. If $b(x)$ is not symmetric and monotonic on R_+, we can let $b^*(x) = \sup_{x\in[-|x|,|x|]} b(x)$. Then b^* is symmetric and monotonic on R_+ and $b^*(x) \geq b(x)$ and thus can serve as the Liptschiz function in Assumption 5.1(b).
3. See, e.g., Kallenberg (2001, pp. 56–57).

REFERENCES

Andrews, D.W.K. & G. Soares (2010) Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* 78, 119–157.
Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.
Campbell, J.Y. & M. Yogo (2006) Efficient tests of stock return predictability. *Journal of Financial Economics* 8, 27–60.

Chang, Y. & J.Y. Park (2011) Endogeneity in nonlinear regressions with integrated time series. *Econometric Reviews* 30, 51–87.

Cheng, X. (2008) Robust Confidence Intervals in Nonlinear Regression under Weak Identification. Manuscript, Department of Economics, Yale University.

Cheng, X. (2010) Essays on weak identification and cointegrating rank selection. Ph.D. Dissertation, Yale University.

de Jong, R. (2002) Nonlinear Estimators with Integrated Regressors but without Exogeneity. Working paper, Michigan State University.

Giacomini, R. & C.W.J. Granger (2004) Aggregation of space-time processes. *Journal of Econometrics* 118, 7–26.

Granger, C.W.J. (1987) Implications of aggregation withc common factors. *Econometric Theory* 3, 208–222.

Hansen, B.E. (1996) Stochastic equicontinuity for unbounded dependent heterogeneous arrays. *Econometric Theory* 12, 347–359.

Ibragimov, R. & P.C.B. Phillips (2008) Regression asymptotics using martingale convergence methods. *Econometric Theory* 24, 888–947.

Jeganathan, P. (2008) Limit Theorems for Functionals of Sums That Converge to Fractional Brownian and Stable Motions. Cowles Foundation Discussion paper 1649, Yale University.

Kallenberg, O. (2001) *Foundations of Modern Probability*, 2nd ed. Springer-Verlag.

Kim, J. & D. Pollard (1990) Cube root asymptotics. *Annals of Statistics* 18, 191–219.

Kurtz, T.G. & P. Protter (1991) Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* 19, 1035–1070.

Marmer, V. (2008) Nonlinearity, nonstationarity, and spurious forecasts. *Journal of Econometrics* 142, 1–27.

Park, J.Y. & P.C.B. Phillips (1999) Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 15, 269–298.

Park, J.Y. & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69, 117–161.

Phillips, P.C.B. (1989) Partially identified econometric models. *Econometric Theory* 5, 181–240.

Phillips, P.C.B. & T. Magdalinos (2009) Unit root and cointegrating limit theory when initialization is in the infinite past. *Econometric Theory* 25, 1682–1715.

Pollard, D. (1990) *Empirical Process Theory and Applications*. Institute of Mathematical Statistics.

Stock, J.H. & J. Wright (2000) GMM with weak identification. *Econometrica* 68, 1055–1096.

van der Vaart, A. & J. Wellner (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag.

Wang, Q. & P.C.B. Phillips (2009) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710–738.

APPENDIX A: Auxiliary Lemmas

The following auxiliary lemmas are used in the proofs of the main lemmas and theorems. The first lemma is based on Lemma A.2 of PP and gives a convergence result to a stochastic integral.

LEMMA A.1. *Let Assumption 2.1 hold. For all $k \geq 1$, if $T : R^{d_x} \to R^k$ is regular (as defined in Definition 3.1 of PP, for which it is sufficient that the elements of T be piecewise continuous) then*

$$n^{-1/2} \sum_{t=1}^{n} T(n^{-1/2} X_t) u_t \to_p \int T(V(r)) dU(r) \quad \text{as } n \to \infty.$$

Proof of Lemma A.1. Lemma A.1 is the same as the second result in Lemma A.2 of PP except the convergence here is in probability instead of in distribution. The proof of the former is thus the same as the latter with only one modification. We only need to change the convergence $\rightarrow_d$ in equation (25) in the proof of the latter into $\rightarrow_p$. The change is valid by Theorem (2.2) in Kurtz and Protter (1991). ∎

Let $h_\pi(X_{n,t}, n, u_t) = h^*(X_{n,t}, n^{1/2}, \pi)u_t$ and let

$$\nu_n h_\pi = n^{-1/2} \sum_{t=1}^{n} h_\pi(X_{n,t}, n, u_t). \tag{A.1}$$

Let $\mathcal{F} = \{h_\pi : \pi \in \Pi\}$. Note that $\{\nu_n h : h \in \mathcal{F}\}$ is an empirical process indexed by h_π in $\mathcal{F}$. Define a semidistance d on $\mathcal{F}$ as follows:

$$d(h_\pi, h_{\pi'}) = \left|\pi - \pi'\right|. \tag{A.2}$$

Lemma A.2. is used in the proof of Lemma 5.2.

LEMMA A.2. *Suppose Assumptions 2.1, 3.1, 5.1, and 5.2 hold. Then the empirical process $\{\nu_n h_\pi : h_\pi \in \mathcal{F}\}$ is stochastically equicontinuous with respect to d.*

Proof of Lemma A.2. We proceed to show that $\{(\nu_n h_\pi)_{\pi\in\Pi}\}_{n\geq 1}$ is stochastically equicontinuous with respect to the pseudo distance:

$$d^h(h_\pi, h_{\pi'}) = \limsup_{n\to\infty} \left[n^{-1} \sum_{t=1}^{n} \mathrm{E}[h_\pi(X_{n,t}, n, u_t) - h_{\pi'}(X_{n,t}, n, u_t)]^2\right]^{1/2}. \tag{A.3}$$

The pseudo distance d^h is well defined because

$$\begin{aligned} d^h(h_\pi, h_{\pi'}) &= \sigma_u \limsup_{n\to\infty} \left[n^{-1} \sum_{t=1}^{n} \mathrm{E}[h^*(X_{n,t}, n^{1/2}, \pi') - h^*(X_{n,t}, n^{1/2}, \pi')]^2\right]^{1/2} \\ &\leq \sigma_u |\pi - \pi'| \limsup_{n\to\infty} \left[n^{-1} \sum_{t=1}^{n} \mathrm{E}b^2(X_{n,t})\right]^{1/2} \\ &= \bar{C}_b \left|\pi - \pi'\right| = \bar{C}_b d(h_\pi, h_{\pi'}), \end{aligned} \tag{A.4}$$

where the first equality holds by the definition of h_π and Assumptions 2.1(b) and (c), the inequality holds by Assumption 5.1(b), and $\bar{C}_b$ is a finite constant by Assumption 5.2(b).

Equation (A.4) also shows that d is a stronger pseudo distance than d^h and hence stochastic equicontinuity with respect to d^h implies stochastic equicontinuity with respect to d.

We use Theorem 2 in Hansen (1996) to show that $\{\nu_n h_\pi : \pi \in \Pi\}_{n\geq 1}$ is stochastically equicontinuous with respect to d^h. To invoke this theorem, we verify the following four conditions: (i) for all $\pi \in \Pi$, $\{h_\pi(X_{n,t}, n, u_t), \mathcal{F}_{n,t}\}$ is a martingale difference sequence; (ii) there exists $b^* : R^{d_x+1} \to R_+$ such that for all $\pi, \pi' \in \Pi$, $\big|h_\pi(X_{n,t}, n, u_t) - h_{\pi'}(X_{n,t}, n, u_t)\big| < b^*(X_{n,t}, u_t)|\pi - \pi'|$; (iii) $\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \mathrm{E}h_\pi^2(X_{n,t}, n, u_t) < \infty$; and

$$\text{(iv)} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \mathrm{E}[b^*(X_t, u_t)]^2 < \infty. \tag{A.5}$$

Condition (i) holds because

$$\begin{aligned} \mathrm{E}(h_\pi(X_{n,t},n,u_t)|\mathcal{F}_{n,t-1}) &= \mathrm{E}(h^*(X_{n,t},n^{1/2},\pi)u_t|\mathcal{F}_{n,t-1}) \\ &= h(X_{n,t},n^{1/2},\pi)\mathrm{E}(u_t|\mathcal{F}_{n,t-1}) = 0, \end{aligned} \tag{A.6}$$

where the second equality holds by Assumption 2.1(c) and the third equality holds by Assumption 2.1(b).

Condition (ii) holds with $b^*(X_{n,t},u_t) = b(X_{n,t})\,|u_t|$ because

$$\begin{aligned} \left|h_\pi(X_{n,t},n,u_t) - h_{\pi'}(X_{n,t},n,u_t)\right| &= |h^*(X_{n,t},n^{1/2},\pi) - h^*(X_{n,t},n^{1/2},\pi')|\,|u_t| \\ &\le b(X_{n,t})\,|u_t|. \end{aligned} \tag{A.7}$$

We now show that condition (iii) holds for large enough n. First we have

$$\begin{aligned} & n^{-1}\sum_{t=1}^n \mathrm{E}h_\pi^2(X_{n,t},n,u_t) \\ &\quad = \sigma_u^2 n^{-1}\sum_{t=1}^n \mathrm{E}h^{*2}(X_{n,t},n^{1/2},\pi) \\ &\quad = \sigma_u^2 n^{-1}\sum_{t=1}^n \mathrm{E}h^2(X_{n,t},\pi) + \sigma_u^2 n^{-1}\kappa^{-2}(n^{1/2},\pi)\sum_{t=1}^n \mathrm{E}R^2(X_{n,t},n^{1/2},\pi). \end{aligned} \tag{A.8}$$

In (A.8), the lim sup of the first term is finite by Assumption 5.2(a). To prove that the lim sup of the second term is finite, let $s_{\max} = \max_{r\in[0,1]} V(r)$ and $s_{\min} = \min_{r\in[0,1]} V(r)$. Let $K = [s_{\min} - 1, s_{\max} + 1]$.

By Definition 3.5 in PP, $R(X_{n,t},n^{1/2},\pi)$ is of smaller order than $\kappa(n^{1/2},\pi)$ in the sense of Definition 3.4 in PP. There are two cases. In case one, $R(X_{n,t},n^{1/2},\pi) = a(n^{1/2},\pi)$ $A(X_{n,t},\pi)$ with $a(n^{1/2},\pi) = o(\kappa(n^{1/2},\pi))$ and $\sup_{\pi\in\Pi} A(\cdot,\pi) \in \mathcal{T}_{LB}^0$, where $\mathcal{T}_{LB}^0$ is the set of exponentially locally bounded functions defined in PP. In this case, we have

$$\begin{aligned} n^{-1}\kappa^{-2}(n^{1/2},\pi)\sum_{t=1}^n \mathrm{E}R^2(X_{n,t},n^{1/2},\pi) &= o(1)n^{-1}\sum_{t=1}^n \mathrm{E}A^2(X_{n,t},\pi) \\ &\le o(1)\mathrm{E}\sup_{x\in K}||A^2(x,\pi)|| = o(1), \end{aligned} \tag{A.9}$$

where the inequality holds for large enough n by Assumption 2.1(a) and the second equality holds because $\sup_{\pi\in\Pi} A(\cdot,\pi) \in \mathcal{T}_{LB}^0$.

In case two, $R(X_{n,t},n^{1/2},\pi) = b(n^{1/2},\pi)A(X_{n,t},\pi)B(n^{1/2}X_{n,t},\pi)$, with $b(n^{1/2},\pi) = O(\kappa(n^{1/2},\pi))$ and $\sup_{\pi\in\Pi} B(\cdot,\pi) \in \mathcal{T}_B^0$, where $\mathcal{T}_B^0$ is the set of transformations that are bounded and vanish at infinity. We then have

$$\begin{aligned} n^{-1}\kappa^{-2}(n^{1/2},\pi)\sum_{t=1}^n \mathrm{E}R^2(X_{n,t},n^{1/2},\pi) &= O(1)n^{-1}\sum_{t=1}^n \mathrm{E}[A^2(X_{n,t},\pi)B^2(n^{1/2}X_{n,t},\pi)] \\ &\le O(1)[\mathrm{E}\sup_{x\in K}||A^4(x,\pi)||]^{1/2}[\mathrm{E}\sup_{x\in R}B^4(x,\pi)]^{1/2} \\ &= O(1), \end{aligned} \tag{A.10}$$

where the inequality holds for large enough n by Assumption 2.1(a) and the Cauchy–Schwarz inequality and the second equality holds because $\sup_{\pi\in\Pi} A(\cdot,\pi) \in \mathcal{T}_{LB}^0$ and $\sup_{\pi\in\Pi} B(\cdot,\pi) \in \mathcal{T}_B^0$.

Equations (A.9) and (A.10) imply that the lim sup of the second term in (A.8) is finite. Thus, condition (iii) holds.

Condition (iv) holds by $\mathrm{E}[b^*(X_{n,t},u_t)]^2 = \sigma^2 \mathrm{E}b^2(X_{n,t})$ and Assumption 5.2(b).

Therefore, Theorem 2 in Hansen (1996) applies, and Lemma A.2 is proved. ■

APPENDIX B: Proofs of the Theorems

Proof of Theorem 4.1.

(i) Part (i) is implied by $\hat{\pi}_n \to_p \pi_0$ because $\pi_n \to_p \pi_0$. Indeed, because $\hat{\pi}_n$ is the minimizer of $n^{-1/2}\beta_n^{-2}D_n(\pi,\pi_n)$, $\hat{\pi}_n \to_p \pi_0$ is implied by Lemma 4.2(i) and the argmax continuous mapping theorem (CMT) as long as the following two conditions hold: (a) $D_I(\cdot,\pi_0)$ is continuous, and (b) $D_I(\cdot,\pi_0)$ has a unique minimum π_0 a.s.

Condition (a) holds by Assumptions 4.2(a) and 4.3. Condition (b) holds because $D(\pi_0,\pi_0) = 0$ and for any $\pi \neq \pi_0$,

$$D_I(\pi,\pi_0) = \frac{\int_{-\infty}^{\infty} g^2(s,\pi_0)ds \int_{-\infty}^{\infty} g^2(s,\pi)ds - \left(\int_{-\infty}^{\infty} g(s,\pi)g(s,\pi_0)ds\right)^2}{\int_{-\infty}^{\infty} g^2(s,\pi)ds} \times L(0,1)$$
$$> 0, \tag{B.1}$$

by virtue of the Cauchy–Schwarz inequality and Assumption 4.4.

(ii) Part(ii) is implied by Lemmas 4.2(ii) and 4.3 and the argmax CMT. ■

Proof of Theorem 4.2.

(i) We first derive the asymptotic distribution of the stochastic process $n^{1/4}\hat{\beta}_n(\pi) : \pi \in \Pi$. We have

$$n^{1/4}\hat{\beta}_n(\cdot) = \frac{n^{-1/4}\sum_{t=1}^n u_t g(X_t,\cdot) + n^{1/4}\beta_n\left(n^{-1/2}\sum_{t=1}^n g(X_t,\cdot)g(X_t,\pi_n)\right)}{n^{-1/2}\sum_{t=1}^n g^2(X_t,\cdot)}$$
$$\to_d f_I(\cdot) := \frac{\sigma_u L(1,0)^{1/2} Z(\cdot) + cL(1,0)\int_{-\infty}^{\infty} g(s,\pi_0)g(s,\cdot)ds}{L(0,1)\int_{-\infty}^{\infty} g^2(s,\cdot)ds}, \tag{B.2}$$

where the convergence holds by the same arguments as those for Lemma 4.2(ii). The convergence $n^{1/4}\hat{\beta}_n(\cdot)$ holds jointly with the convergence of $nD_n(\cdot,\pi_n)$ in Lemma 4.2(ii) because $n^{1/4}\hat{\beta}_n(\cdot)$ and $nD_n(\cdot,\pi_n)$ are both composed of the same elements. Because $n^{1/4}\hat{\beta}_n(\hat{\pi}_n)$ is a continuous functional of $\left(n^{1/4}\hat{\beta}_n(\cdot), nD_n(\cdot,\pi_n)\right)$ with respect to the sup norm, the CMT applies and we have

$$n^{1/4}\hat{\beta}_n(\hat{\pi}_n) \to_d f_I(\tau_{I,\pi}(c,\pi_0)),$$

giving the desired result.

(ii) First we show that $\hat{\beta}_n$ is consistent. We have

$$\hat{\beta}_n(\pi)/\beta_n = o_p(1) + \frac{n^{-1/2}\sum_{t=1}^n g(X_t,\pi)g(X_t,\pi_n)}{n^{-1/2}\sum_{t=1}^n g^2(X_t,\pi)}$$
$$\to_p \frac{\int_{-\infty}^{\infty} g(s,\pi)g(s,\pi_0)ds}{\int_{-\infty}^{\infty} g^2(s,\pi)ds}, \tag{B.3}$$

uniformly over $\pi \in \Pi$, where the equality holds by Lemma 4.1 and $n^{-1/4}\beta_n^{-1} \to 0$ and the convergence holds by the same arguments as those for Lemma 4.2(i) in Appendix C. Thus, Theorem 4.1(i) and Assumption 4.2(a) imply that $\hat{\beta}_n/\beta_n := \hat{\beta}_n(\hat{\pi}_n)/\beta_n \to_p 1$.

The NLS estimators satisfy $\partial Q_n(\hat{\theta}_n)/\partial\theta = o_p(n^{-1/4})$, and a mean value expansion of $\partial Q_n(\hat{\theta}_n)/\partial\theta$ gives

$$o_p(n^{-1/4}) = \frac{\partial Q_n(\theta_n)}{\partial\theta} + \frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'}\left(\hat{\theta}_n - \theta_n\right), \tag{B.4}$$

where $\theta_n = (\beta_n, \pi_n)'$ and $\tilde{\theta}_n$ lies on the line segment joining θ_n and $\hat{\theta}_n$. Let $\Lambda_n = 2^{-1}\mathrm{diag}\left(n^{1/4}, n^{1/4}\beta_n^{-1}\right)$. Next we show

$$n^{1/2}\Lambda_n[\partial Q_n(\theta_n)/\partial\theta] \to_d \sigma_u L^{1/2}(1,0)\Sigma_{g\dot{g}}^{1/2} Z, \tag{B.5}$$

where $Z \sim N(0, I_2)$, and

$$\Lambda_n \frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'}\Lambda_n \to_p \Sigma_{g\dot{g}} L(1,0). \tag{B.6}$$

Under Assumptions 4.3 and 4.5(b), $\Sigma_{g\dot{g}}$ is invertible. Therefore, Theorem 4.2(ii) is implied by (B.4)–(B.6).

Result (B.5) is implied by Lemma 4.1 and the Cramér–Wold device applied to

$$n^{1/2}\Lambda_n[\partial Q_n(\theta_n)/\partial\theta] = n^{-1/2}\sum_{t=1}^n \begin{pmatrix} g(X_t,\pi_n) \\ \dot{g}(X_t,\pi_n)\end{pmatrix} u_t. \tag{B.7}$$

Equation (B.6) is implied by

$$\begin{aligned}
2^{-1}n^{1/2}\partial^2 Q_n(\tilde{\theta}_n)/\partial\beta^2 &= n^{-1/2}\sum_{t=1}^n g^2(X_t,\tilde{\pi}_n) \to_p L(1,0)\int_{-\infty}^{\infty} g^2(s,\pi)ds, \\
2^{-1}n^{1/2}\beta_n^{-1}\partial^2 Q_n(\tilde{\theta}_n)/\partial\beta\partial\pi &= n^{-1/2}\sum_{t=1}^n \dot{g}(X_t,\tilde{\pi}_n)\left(2\beta_n^{-1}\tilde{\beta}_n g(X_t,\tilde{\pi}_n) - g(X_t,\pi_n)\right) \\
&\quad - n^{-1/2}\beta_n^{-1}\sum_{t=1}^n \dot{g}(X_t,\tilde{\pi}_n)u_t \\
&\to_p L(1,0)\int_{-\infty}^{\infty}\dot{g}(s,\pi_0)g(s,\pi_0)ds, \\
2^{-1}n^{1/2}\beta_n^{-2}\partial^2 Q_n(\tilde{\theta}_n)/\partial\pi^2 &= n^{-1/2}\sum_{t=1}^n \ddot{g}(X_t,\tilde{\pi}_n)\left(\beta_n^{-2}\tilde{\beta}_n^2 g(X_t,\tilde{\pi}_n) - g(X_t,\pi_n)\right) \\
&\quad + n^{-1/2}\sum_{t=1}^n \dot{g}^2(X_t,\tilde{\pi}_n) \\
&\to_p L(1,0)\int_{-\infty}^{\infty}\dot{g}^2(s,\pi)ds,
\end{aligned} \tag{B.8}$$

where the convergence holds by Theorem 3.2 in PP, Assumptions 4.2 and 4.5, and Lemma 4.1. ■

Proof of Theorem 5.1. We show part (i) first. We have

(a) $D_H(\pi,\pi_0)$ is continuous in π because $h(v,\cdot)$ is continuous a.s. by Definition 3.5(ii) and Lemma A.8 in PP, and

$$\int h^2(V,\pi) = \int_{-\infty}^{\infty} h^2(s,\pi)L(1,s)ds > 0 \quad \text{a.s.}, \tag{B.9}$$

by Assumption 5.1(c); and

(b) $D_H(\pi,\pi_0)$ is uniquely minimized at $\pi=\pi_0$ a.s. because

$$\int h^2(V,\pi_0)\int h^2(V,\pi) \geq \left[\int h(V,\pi)h(V,\pi_0)\right]^2 \quad \text{a.s.,} \tag{B.10}$$

by the Cauchy–Schwarz inequality, where the equality holds if and only if $\int (h(V,\pi)-ah(V,\pi_0))^2 = 0$ a.s. for some $a \neq 0$, which holds if and only if $\pi=\pi_0$ by Assumption 5.1(d).

With Lemma 5.3(i) and conditions (a) and (b), we can apply the argmax CMT (see, e.g., van der Vaart and Wellner, (1996, Thm. 3.2.2, p. 286) and get $\hat{\pi}_n \to_d \pi_0$, which implies part (i) because π_0 is a constant.

Lemmas 5.3(ii) and 5.4 along with the argmax CMT yield part (ii). ∎

Proof of Theorem 5.2.

(i) We first derive the asymptotic distribution of the stochastic process $n^{1/2}\kappa_{n,\pi}\hat{\beta}_n(\pi):\pi\in\Pi$. We have

$$\begin{aligned} n^{1/2}\kappa_{n,\cdot}\hat{\beta}_n(\cdot) &= \frac{n^{-1/2}\kappa_{n,\cdot}^{-1}\sum_{t=1}^n u_t g(X_t,\cdot)}{n^{-1}\kappa_{n,\cdot}^{-2}\sum_{t=1}^n g^2(X_t,\cdot)} \\ &\quad + \frac{n^{1/2}\kappa_{n,\pi_n}\beta_n\left(n^{-1}\kappa_{n,\cdot}^{-1}\kappa_{n,\pi_n}^{-1}\sum_{t=1}^n g(X_t,\cdot)g(X_t,\pi_n)\right)}{n^{-1}\kappa_{n,\cdot}^{-2}\sum_{t=1}^n g^2(X_t,\cdot)} \\ &\to_p f_H(\cdot) := \frac{\int h(V,\pi)dU + c\int h(V,\pi)h(V,\pi_0)}{\int h^2(V,\pi)}, \end{aligned} \tag{B.11}$$

where the convergence holds by the same arguments as those used for Lemma 5.3(ii). The convergence $n^{1/2}\kappa_{n,\cdot}\hat{\beta}_n(\cdot)$ holds jointly with the convergence of $nD_n(\cdot,\pi_n)$ in Lemma 5.3(ii) because $n^{1/2}\kappa_{n,\cdot}\hat{\beta}_n(\cdot)$ and $nD_n(\cdot,\pi_n)$ are both composed of the same elements. Because $n^{1/2}\kappa_{n,\hat{\pi}_n}\hat{\beta}_n(\hat{\pi}_n)$ is a continuous functional of $\left(n^{1/2}\kappa_{n,\cdot}\hat{\beta}_n(\cdot), nD_n(\cdot,\pi_n)\right)$ with respect to the sup norm, the CMT applies and gives the desired result.

(ii) The NLS estimator $\hat{\pi}_n$ satisfies

$$\dot{Q}_n(\hat{\pi}_n) = o_p(1), \tag{B.12}$$

where $\dot{Q}$ denotes the first derivative of Q. Expand $\dot{Q}_n(\hat{\pi}_n)$ around π_0, and we have

$$o_p(1) = \dot{Q}_n(\pi_n) + \ddot{Q}_n(\tilde{\pi}_n)(\hat{\pi}_n - \pi_n), \tag{B.13}$$

where $\ddot{Q}$ denotes the second derivative of Q and $\tilde{\pi}_n$ lies between $\hat{\pi}_n$ and π_0.

To find the asymptotic distribution of $\hat{\pi}_n - \pi_n$, we need to find the asymptotic distribution of $\dot{Q}_n(\pi_0)$ and $\ddot{Q}_n(\pi_n)$. Let g_π, $\dot{g}_\pi$, and $\ddot{g}_\pi$ denote $g(X_t,\pi)$, $\dot{g}(X_t,\pi)$, and $\ddot{g}(X_t,\pi)$, respectively. Then

$$\begin{aligned} \dot{Q}_n(\pi_n) &= \frac{2n^{-1}\sum_{t=1}^n u_t g_{\pi_n}\left[\sum_{t=1}^n u_t\dot{g}_{\pi_n}\sum_{t=1}^n g_{\pi_n}^2 - \sum_{t=1}^n u_t g_{\pi_n}\sum_{t=1}^n g_\pi\dot{g}_\pi\right]}{\left(\sum_{t=1}^n g_{\pi_n}^2\right)^2} \\ &\quad + \frac{2n^{-1}\beta_n\left[\sum_{t=1}^n g_{\pi_n}^2\sum_{t=1}^n u_t\dot{g}_{\pi_n} - \sum_{t=1}^n g_{\pi_n}\dot{g}_{\pi_n}\sum_{t=1}^n u_t g_{\pi_n}\right]}{\sum_{t=1}^n g_{\pi_n}^2}, \end{aligned}$$

$$\ddot{Q}_n(\pi) = -2n^{-1}\frac{\sum_{t=1}^n Y_t g_\pi \sum_{t=1}^n Y_t \ddot{g}_\pi + \left(\sum_{t=1}^n Y_t \dot{g}_\pi\right)^2 + nQ_n(\pi)\sum_{t=1}^n [\dot{g}_\pi^2 + g_\pi \ddot{g}_\pi]}{\sum_{t=1}^n g_\pi^2}$$
$$-8n^{-1}\frac{\sum_{t=1}^n g_\pi \dot{g}_\pi \left[\sum_{t=1}^n Y_t g_\pi \sum_{t=1}^n Y_t \dot{g}_\pi + nQ_n(\pi)\sum_{t=1}^n g_\pi \dot{g}_\pi\right]}{\left(\sum_{t=1}^n g_\pi^2\right)^2}. \tag{B.14}$$

We have

$$n^{-1}\beta_n^{-1}\kappa_{n,\pi_n}^{-1}\kappa_{1,n,\pi}^{-1}\sum_{t=1}^n Y_t \dot{g}_\pi$$
$$= n^{-1}\beta_n^{-1}\kappa_{n,\pi_n}^{-1}\kappa_{1,n,\pi}^{-1}\sum_{t=1}^n u_t \dot{g}_\pi + n^{-1}\kappa_{n,\pi_n}^{-1}\kappa_{1,n,\pi}^{-1}\sum_{t=1}^n g_{\pi_n}\dot{g}_\pi. \tag{B.15}$$

The first term on the right of (B.15) is $o_p(1)$ uniformly over $\pi \in \Pi$ as $n^{-1/2}$ $\beta_n^{-1}\kappa_{n,\pi_n}^{-1} \to 0$ and

$$n^{-1/2}\kappa_{1,n,\pi}^{-1}\sum_{t=1}^n u_t \dot{g}_\pi \to_p \int h_1(V,\pi)dU, \tag{B.16}$$

uniformly over $\pi \in \Pi$ by Assumption 5.3 and the same procedure used in the proof of Lemma 5.1 in Appendix C. The second term in (B.15) converges a.s. to $\int h(V,\pi_0)h_1(V,\pi)$ uniformly over $\pi \in \Pi$ by Lemma A.6 and Theorem 3.3 in PP, $\pi_n \to \pi_0$, and the continuity of $h(v,\cdot)$. Thus,

$$(n\beta_n\kappa_{n,\pi_n}\kappa_{1,n,\pi})^{-1}\sum_{t=1}^n Y_t \dot{g}_\pi \to_p \int h(V,\pi_0)h_1(V,\pi), \tag{B.17}$$

uniformly over $\pi \in \Pi$. Similarly, we find

$$n^{-1}\kappa_{n,\pi}^{-1}\kappa_{1,n,\pi}^{-1}\sum_{t=1}^n g_\pi \dot{g}_\pi \to_p \int h(V,\pi)h_1(V,\pi),$$
$$(n\beta_n\kappa_{n,\pi_n}\kappa_{2,n,\pi})^{-1}\sum_{t=1}^n Y_t \ddot{g}_\pi \to_p \int h(V,\pi_0)h_2(V,\pi),$$
$$n^{-1}\kappa_{1,n,\pi}^{-2}\sum_{t=1}^n \dot{g}_\pi^2 \to_p \int h_1^2(V,\pi),$$
$$n^{-1}\kappa_{n,\pi}^{-1}\kappa_{2,n,\pi}^{-1}\sum_{t=1}^n g_\pi \ddot{g}_\pi \to_p \int h(V,\pi)h_2(V,\pi), \tag{B.18}$$

uniformly over $\pi \in \Pi$.

A by-product of the proof of Lemma 5.3(i) is that

$$\beta_n^{-2}n^{-1}\kappa_{n,\pi_n}^{-2}Q_n(\pi) \to Q(\pi) \quad \text{a.s.}, \tag{B.19}$$

uniformly over $\pi \in \Pi$, where $Q(\pi) = -\left[\int h_\pi(r)h_{\pi_0}(r)dr\right]^2 \Big/ \int h_\pi^2(r)dr$.

Equations (C.28), (B.14), (B.17), (B.18), and (B.19) and $\pi_n \to \pi_0$ and $\tilde{\pi}_n \to_p \pi_0$ together imply that

$$n^{-1/2}\kappa_{1,n,\pi_n}^{-1}\beta_n^{-1}\dot{Q}_n(\pi_n)$$
$$\to_p 2\int h_1(V,\pi_0)dU - \frac{2\int h(V,\pi_0)h_1(V,\pi_0)\int h(V,\pi_0)dU}{\int h^2(V,\pi_0)} \tag{B.20}$$

and

$$(\log n)^{-2} n^{-1} \beta_n^{-2} \kappa_{n,\pi_n}^{-2} \ddot{Q}_n(\tilde{\pi}_n) \to_p 2 \int h_1^2(V,\pi_0) - \frac{2[\int h(V,\pi_0) h_1(V,\pi_0)]^2}{\int h^2(V,\pi_0)}, \tag{B.21}$$

uniformly over $\pi \in \Pi$.

The asymptotic distribution of $\hat{\pi}_n$ follows easily from (B.13), (B.21), and (B.21).

(iii) First we show that $\hat{\beta}_n$ is consistent. We have

$$\begin{aligned} \kappa_{n,\pi} \hat{\beta}_n(\pi)/(\kappa_{n,\pi_n} \beta_n) &= \frac{(\beta_n^{-1} n^{-1/2} \kappa_{n,\pi_n}^{-1}) n^{-1/2} \kappa_{n,\pi}^{-1} \sum_{t=1}^n u_t g(X_t,\pi)}{n^{-1} \kappa_{n,\pi}^{-2} \sum_{t=1}^n g^2(X_t,\pi)} \\ &\quad + \frac{n^{-1} \kappa_{n,\pi}^{-1} \kappa_{n,\pi_n}^{-1} \sum_{t=1}^n g(X_t,\pi) g(X_t,\pi_n)}{n^{-1} \kappa_{n,\pi}^{-2} \sum_{t=1}^n g^2(X_t,\pi)} \\ &\to_p \frac{\int h(V,\pi) h(V,\pi_0)}{\int h^2(V,\pi)}, \quad \text{uniformly over } \pi \in \Pi, \end{aligned} \tag{B.22}$$

where the convergence holds by the same arguments as those for Lemma 5.3(i). Thus, Theorem 5.1(i) and the continuity of $\int h(V,\pi) h(V,\pi_0) / \int h^2(V,\pi)$ (Lemma 5.4) imply that

$$\kappa_{n,\hat{\pi}_n} \hat{\beta}_n(\hat{\pi}_n)/(\kappa_{n,\pi_n} \beta_n) \to_p 1. \tag{B.23}$$

By part (ii), $\hat{\pi}_n - \pi_n = O_p(n^{-1/2} \beta_n^{-1} \kappa_{1,n,\pi_n}^{-1}) = o_p(\kappa_{n,\pi_n} \kappa_{1,n,\pi_n}^{-1}) = o_p(1/\log n)$. Then we have

$$\kappa_{n,\hat{\pi}_n}/\kappa_{n,\pi_n} \to_p 1. \tag{B.24}$$

Thus $\hat{\beta}_n/\beta_n \to_p 1$.

Now we derive the asymptotic distribution of $\hat{\beta}_n$. We have

$$\begin{aligned} n^{1/2} \kappa_{n,\pi_n} (\hat{\beta}_n - \beta_n) &= \frac{n^{1/2} \kappa_{n,\pi_n} \sum_{t=1}^n g(X_t,\hat{\pi}_n) u_t}{\sum_{t=1}^n g^2(X_t,\hat{\pi}_n)} \\ &\quad - \beta_n \frac{n^{1/2} \kappa_{n,\pi_n} \sum_{t=1}^n g(X_t,\hat{\pi}_n) \dot{g}(X_t,\tilde{\pi}_n)}{\sum_{t=1}^n g^2(X_t,\hat{\pi}_n)} \\ &\quad \times \left(\hat{\pi}_n - \pi_n\right) \\ &\to_p \frac{\int h(V,\pi_0) dU}{\int h^2(V,\pi_0)} - \frac{\int h(V,\pi_0) h_1(V,\pi_0)}{\int h^2(V,\pi_0)} \times T_{H,\pi}(\pi_0), \end{aligned} \tag{B.25}$$

where the equality holds by a mean-value expansion of $g(X_t,\hat{\pi}_n)$ around π_n and the convergence holds by part (ii), (B.24), and the same arguments as those for Lemma 5.3(i). Thus, part (ii) is proved. ■

Proof of Theorem 6.1. The proof is similar to that of Theorem 1 in Andrews and Soares (2010). The proofs of parts (i) and (ii) are analogous, and therefore only the proof of part (i) is presented here.

By the definition of $AsySZ_{I,\beta}$, there exists a sequence θ_n such that

$$AsySZ_{I,\beta}(\alpha) = \liminf_{n\to\infty} P_{\theta_n}(\beta_n \in CI_{I,\beta,n}(\alpha))$$
$$= \liminf_{n\to\infty} P_{\theta_n}(n^{1/4}|\hat{\beta}_n - \beta_n| \le \hat{q}_{I,\beta}(\hat{\pi}_n, 1-\alpha)). \quad \textbf{(B.26)}$$

Let $\{u_n\}$ be a subsequence of $\{n\}$ such that $AsySZ_{I,\beta}(\alpha) = \lim_{n\to\infty} P_{\theta_{u_n}}(u_n^{1/4}|\hat{\beta}_{u_n} - \beta_{u_n}| \le \hat{q}_{I,\beta}(\hat{\pi}_{u_n}, 1-\alpha))$. Such a subsequence always exists. Because the euclidean space is complete, there exists a subsequence $\{a_n\}$ of $\{u_n\}$ such that $(a_n^{1/4}\beta_{a_n}, \pi_{a_n}) \to (c, \pi_0)$ where $c \in R_{[\pm\infty]}$ and $\pi_0 \in \Pi$. Then

$$AsySZ_{I,\beta}(\alpha) = \lim_{n\to\infty} P_{\theta_{a_n}}(a_n^{1/4}|\hat{\beta}_{a_n} - \beta_{a_n}| \le \hat{q}_{I,\beta}(\hat{\pi}_{a_n}, 1-\alpha)). \quad \textbf{(B.27)}$$

If $c \in R$, then by Theorem 4.2(i) and Assumption 6.1, $a_n^{1/4}|\hat{\beta}_{a_n}| = O_p(1) < b_n$ with probability approaching one. Thus, $\hat{q}_{I,\beta}(\hat{\pi}_{a_n}, 1-\alpha) = \sup_{c' \in R_\infty} \sup_{\pi\in\Pi} q_{I,\beta}(c', \pi, 1-\alpha)$ with probability approaching one. By Theorem 4.2(i), $a_n^{1/4}(\hat{\beta}_{a_n} - \beta_{a_n}) \to_d \tau_{I,\beta}(c, \pi_0) - c$. (Theorem 4.2 is in terms of $\{n\}$, but all the proofs go through with $\{n\}$ replaced with a subsequence $\{a_n\}$ of $\{n\}$.) The distribution of $\tau_{I,\beta}(c, \pi_0) - c$ is continuous and strictly increasing because $Z \sim N(0,1)$ and the local time $L(1,0) > 0$ with probability one. Thus, with probability approaching one

$$AsySZ_{I,\beta}(\alpha) = \lim_{n\to\infty} P_{\theta_{a_n}}(a_n^{1/4}|\hat{\beta}_{a_n} - \beta_{a_n}| \le \hat{q}_{I,\beta}(\hat{\pi}_{a_n}, 1-\alpha))$$
$$\ge \lim_{n\to\infty} P_{\theta_{a_n}}(a_n^{1/4}|\hat{\beta}_{a_n} - \beta_{a_n}| \le q_{I,\beta}(c, \pi, 1-\alpha))$$
$$= 1-\alpha. \quad \textbf{(B.28)}$$

If $c = \pm\infty$, by Theorem 4.2(ii), $a_n^{1/4}(\hat{\beta}_{a_n} - \beta_{a_n}) \to_d T_{I,\beta}(\pi_0)$. Then

$$AsySZ_{I,\beta}(\alpha)$$
$$\ge \lim_{n\to\infty} P_{\theta_{a_n}}(a_n^{1/4}|\hat{\beta}_{a_n} - \beta_{a_n}| \le q_{I,\beta}(\infty, \hat{\pi}_n, 1-\alpha)) P_{\theta_{a_n}}(a_n^{1/4}\hat{\beta}_{a_n} > b_{a_n})$$
$$+ \lim_{n\to\infty} P_{\theta_{a_n}}(a_n^{1/4}|\hat{\beta}_{a_n} - \beta_{a_n}| \le \sup_{\pi\in\Pi} q_{I,\beta}(\infty, \pi, 1-\alpha)) P_{\theta_{a_n}}(a_n^{1/4}\hat{\beta}_{a_n} \le b_{a_n})$$
$$\ge \lim_{n\to\infty} P_{\theta_{a_n}}(a_n^{1/4}|\hat{\beta}_{a_n} - \beta_{a_n}| \le q_{I,\beta}(\infty, \hat{\pi}_n, 1-\alpha)) P_{\theta_{a_n}}(a_n^{1/4}\hat{\beta}_{a_n} > b_{a_n})$$
$$+ (1-\alpha) \lim_{n\to\infty} P_{\theta_{a_n}}(a_n^{1/4}\hat{\beta}_{a_n} > b_{a_n}), \quad \textbf{(B.29)}$$

where the second inequality holds because $\sup_{\pi\in\Pi} q_{I,\beta}(\infty, \pi, 1-\alpha) > q_{I,\beta}(\infty, \pi_0, 1-\alpha)$ and $T_{I,\beta}(\pi_0)$ has a continuous distribution for the same reason that $\tau_{I,\beta}(c, \pi_0) - c$ does.

By (B.28) and (B.29), we can conclude that

$$AsySZ_{I,\beta}(\alpha) \ge 1-\alpha, \quad \textbf{(B.30)}$$

if

$$\lim_{n\to\infty} P_{\theta_{a_n}}(a_n^{1/4}|\hat{\beta}_{a_n} - \beta_{a_n}| \le q_{I,\beta}(\infty, \hat{\pi}_n, 1-\alpha)) \ge 1-\alpha. \quad \textbf{(B.31)}$$

Equation (B.31) holds if $q_{I,\beta}(\infty,\hat{\pi}_n,1-\alpha)\rightarrow_p q_{I,\beta}(\infty,\pi_0,1-\alpha)$, which holds because (a) $T_{I,\beta}(\hat{\pi}_n)\rightarrow_p T_{I,\beta}(\pi_0)$ by Theorem 4.1(i) and Assumption 4.2(a), and (b) $T_{I,\beta}(\pi_0)$ has a continuous and strictly increasing cumulative distribution function (cdf).

It is left to show that

$$AsySZ_{I,\beta}(\alpha)\le 1-\alpha. \tag{B.32}$$

Consider $\theta=(\beta,\pi)\in(R/\{0\})\times\Pi$. Then by definition,

$$AsySZ_{I,\beta}(\alpha)\le \liminf_{n\to\infty} P_\theta(\beta\in CI_{I,\beta,n}(\alpha)). \tag{B.33}$$

Because $\beta\neq 0$, $n^{1/4}b_n^{-1}\beta$ diverges to ∞ or $-\infty$ by Assumption 6.2. Without loss of generality, suppose $n^{1/4}b_n^{-1}\beta\rightarrow\infty$. Then by Theorem 4.2(ii), $n^{1/4}|\hat{\beta}_n|>b_n$ with probability approaching one. Thus,

$$\begin{aligned}\liminf_{n\to\infty} P_\theta\left(\beta\in CI_{I,\beta,n}(\alpha)\right)&=\liminf_{n\to\infty} P_\theta\left(n^{1/4}|\hat{\beta}_n-\beta_n|\le q_{I,\beta}(\infty,\hat{\pi}_n,1-\alpha)\right)\\&=1-\alpha,\end{aligned} \tag{B.34}$$

where the second equality holds by Theorem 4.2(ii), $q_{I,\beta}(\infty,\hat{\pi}_n,1-\alpha)\rightarrow_p q_{I,\beta}(\infty,\pi_0,1-\alpha)$ (shown earlier), and the continuity of the cdf of $T_{I,\beta}(\pi_0)$.

Combining (B.30), (B.33), and (B.34), we obtain part (i). ■

Proof of Theorem 6.2.

(i) The proof is essentially the same as that of Theorem 6.1(i) and is omitted for brevity.

(ii) Similar to the proof of Theorem 6.1(i), we show

$$AsySZ_{H,\beta}(\alpha)\ge 1-\alpha \quad\text{and}\quad AsySZ_{H,\beta}(\alpha)\le 1-\alpha. \tag{B.35}$$

The proof of $AsySZ_{H,\beta}(\alpha)\le 1-\alpha$ is essentially the same as that of (B.32) in the proof of Theorem 6.1(i) and thus is omitted for brevity. Next we show $AsySZ_{H,\beta}(\alpha)\ge 1-\alpha$.

As in (B.27), we find a subsequence $\{a_n\}$ of $\{n\}$ and a sequence $\{\theta_n\}$ such that $(a_n^{1/2}\kappa_{a_n,\pi_{a_n}}\beta_{a_n},\pi_{a_n})\rightarrow(c,\pi_0)$ and

$$AsySZ_{H,\beta}(\alpha)=\lim_{n\to\infty}\Pr_{\theta_{a_n}}\left(\beta_{a_n}\in CI_{H,\beta,n}(\alpha)\right). \tag{B.36}$$

If $c=\pm\infty$, the same arguments as those for (B.29) and (B.31) can be used to show that $AsySZ_{H,\beta}(\alpha)\ge 1-\alpha$. If $c\in R$, then $a_n^{1/2}\kappa_{a_n,\hat{\pi}_{a_n}}\hat{\beta}_{a_n}=O_p(1)<b_{a_n}$ with probability approaching one by Theorem 5.2(i). Thus,

$$\begin{aligned}AsySZ_{H,\beta}(\alpha)&\ge\lim_{n\to\infty}\Pr_{\theta_{a_n}}\left(\inf_{\pi\in CI_{H,\pi,a_n}(\alpha)} b_{a_n}^{-1}a_n^{1/2}|\kappa_{a_n,\pi}\beta_{a_n}|\le 1\right)\\&\ge\lim_{n\to\infty}\Pr_{\theta_{a_n}}\left(\inf_{\pi\in CI_{H,\pi,a_n}(\alpha)} b_{a_n}^{-1}a_n^{1/2}|\kappa_{a_n,\pi}\beta_{a_n}|\le 1\ \&\ \pi_{a_n}\in CI_{H,\pi,a_n}(\alpha)\right)\\&\ge\lim_{n\to\infty}\Pr_{\theta_{a_n}}\left(b_{a_n}^{-1}a_n^{1/2}|\kappa_{a_n,\pi_{a_n}}\beta_{a_n}|\le 1\ \&\ \pi_{a_n}\in CI_{H,\pi,a_n}(\alpha)\right)\\&=\lim_{n\to\infty}\Pr_{\theta_{a_n}}\left(\pi_{a_n}\in CI_{H,\pi,a_n}(\alpha)\right)\\&\ge 1-\alpha,\end{aligned} \tag{B.37}$$

where the first inequality holds by the definition of $CI_{H,\beta,n}(\alpha)$, the equality holds because $b_{a_n}^{-1} \to 0$ and $a_n^{1/2}\kappa_{a_n,\pi_{a_n}}\beta_{a_n} \to c \in R$, and the last inequality holds by part (i). Therefore, $AsySZ_{H,\beta}(\alpha) \geq 1-\alpha$, and part (ii) is proved. ∎

APPENDIX C: Proofs of the Main Lemmas

Proofs of Lemma 4.1. (i) and (ii). Parts (i) and (ii) are analogous. Thus, it suffices to prove part (i) only. (The proof of this part is inspired by the techniques from Wang and Phillips, 2009.) Part (i) holds if we can show that $x_t \equiv X_t/\sqrt{t}$ has a uniformly bounded density $f_t(x)$ conditional on $\mathcal{F}_0$, the σ-field generated by $(\varepsilon_{-1}, \varepsilon_{-2}, \ldots)$, because by Assumption 4.2, for every $\pi \in \Pi$ and every n,

$$\begin{aligned}
n^{-1/2}\sum_{t=1}^{n} \mathrm{E}\left(g^2(X_t,\pi)|\mathcal{F}_0\right) &= n^{-1/2}\sum_{t=1}^{n}\int_{-\infty}^{\infty} g^2\left(\sqrt{t}x,\pi\right) f_t(x)\,dx \\
&= n^{-1/2}\sum_{t=1}^{n} t^{-1/2}\int_{-\infty}^{\infty} g^2(x,\pi)\, f_t\left(x/\sqrt{t}\right) dx \\
&\leq A\int_{-\infty}^{\infty} g^2(x,\pi)\,dx \times n^{-1/2}\sum_{t=1}^{n} t^{-1/2},
\end{aligned} \tag{C.1}$$

for some constant A. The expression after $\leq$ in the preceding equation is bounded uniformly in n because $\lim_{n\to\infty} n^{-1/2}\sum_{t=1}^{n} t^{-1/2} < \infty$.

Now we show that $x_t \equiv X_t/\sqrt{t}$ has a uniformly bounded density $f_t(x)$ conditional on the σ-field $\mathcal{F}_0$. Write X_t as

$$\begin{aligned}
x_t &= t^{-1/2}\left(\sum_{j=1}^{t}\sum_{k=1}^{\infty}\varphi_k\varepsilon_{j-k}\right) \\
&= t^{-1/2}\left(\sum_{j=0}^{t-1}\tilde{\varphi}_j\varepsilon_j + \sum_{j=-\infty}^{-1}\tilde{\varphi}_j\varepsilon_j\right) \\
&=: S_t + R_{0,t},
\end{aligned} \tag{C.2}$$

where $\tilde{\varphi}_j = \begin{cases} \sum_{l=0}^{t-1-j}\varphi_l \text{ for } j = 0,\ldots,t-1 \\ \sum_{l=-j}^{t-1-j}\varphi_l \text{ for } j = -1,-2,\ldots \end{cases}$. Because $R_{0,t}$ is measurable with respect to $\mathcal{F}_0$, it suffices to show that $S_t = \sum_{j=0}^{t-1}\tilde{\varphi}_j\varepsilon_j/\sqrt{t}$ has a uniformly bounded density $f_t(x)$. Let $c_t(u)$ be the characteristic function of S_t. Because of the duality between characteristic functions and density functions, it suffices to show that

$$\sup_t \int_{-\infty}^{\infty} |c_t(u)|\,du < \infty. \tag{C.3}$$

We use the following two facts: I. for sufficiently large t there exist bounds $\underline{\varphi} < |\tilde{\varphi}_j| \leq \overline{\varphi}$ for all $t/2 \leq j \leq t-1$ with $0 < \underline{\varphi} < \overline{\varphi} < \infty$; this fact is assured by virtue of $\sum_{j=1}^{\infty}\varphi_j \neq 0$ and $\sum_{j=1}^{\infty}|\varphi_j| < \infty$ (Assumption 4.1); and II.

$$c(u) := \mathrm{E}e^{iu\varepsilon_1} \leq \begin{cases} e^{-au^2} & \text{if } |u| < \delta_0 \\ \eta & \text{if } |u| > \delta_0 \end{cases},$$

for some $\delta_0, a > 0$ and some $\eta \in (0,1)$, which follows as in Wang and Phillips (2009, proof of Cor. 2.2).

Using I and II, we prove (C.3) as follows:

$$
\begin{aligned}
&\int_{-\infty}^{\infty} |c_t(u)|du \\
&\quad = \int_{-\infty}^{\infty} \Pi_{j=0}^{t-1} |\mathrm{E}e^{iu\tilde{\varphi}_j\varepsilon_j/\sqrt{t}}|du \\
&\quad \le \int_{-\infty}^{\infty} \Pi_{j=[t/2]}^{t-1} |\mathrm{E}e^{iu\tilde{\varphi}_j\varepsilon_j/\sqrt{t}}|du \\
&\quad = \int_{|u|<\sqrt{t}\delta_0/\overline{\varphi}} \Pi_{j=[t/2]}^{t-1} |\mathrm{E}e^{iu\tilde{\varphi}_j\varepsilon_j/\sqrt{t}}|du + \int_{|u|\ge\sqrt{t}\delta_0/\overline{\varphi}} \Pi_{j=[t/2]}^{t-1} |\mathrm{E}e^{iu\tilde{\varphi}_j\varepsilon_j/\sqrt{t}}|du \\
&\quad \le \int_{-\infty}^{\infty} \Pi_{j=[t/2]}^{t-1} e^{-au^2\tilde{\varphi}_j^2/t} du + \int_{|u|\ge\sqrt{t}\delta_0/\overline{\varphi}} \Pi_{j=[t/2]}^{t-1} |\mathrm{E}e^{iu\tilde{\varphi}_j\varepsilon_j/\sqrt{t}}|du \\
&\quad \le \int_{-\infty}^{\infty} e^{-au^2\underline{\varphi}^2/2} du + \left(\eta \vee e^{-a\delta_0^2\underline{\varphi}^2/\overline{\varphi}^2}\right)^{t-1/2} \left(\sqrt{t}/\underline{\varphi}\right) \int_{-\infty}^{\infty} |\mathrm{E}e^{iu\varepsilon_1}|du \\
&\quad \le \int_{-\infty}^{\infty} e^{-au^2\underline{\varphi}^2/2} du + \underline{\varphi}^{-1} \int_{-\infty}^{\infty} |c(u)|du < \infty. \qquad \textbf{(C.4)}
\end{aligned}
$$

The first inequality in expression (C.4) holds because $\left|e^{iux}\right| \le 1$ for any u and any x. The second inequality holds by fact II. The third inequality holds by fact I (for the first term) and facts I and II and a change of variable (for the second term). The fourth inequality holds for t sufficiently large because $\eta \vee e^{-a\delta_0^2\underline{\varphi}^2/\overline{\varphi}^2} < 1$, and the last inequality holds by Assumption 4.1.

(iii). The proof of part (iii) applies Theorem 10.2 in Pollard (1990). Lemma 4.1 is proved once we verify the three conditions of this theorem: (i) $(\Pi, |\cdot|)$ is totally bounded, where $|\cdot|$ is the euclidean norm on R, (ii) for any $\{\pi_1,\ldots,\pi_J\} \subset \Pi$, finite-dimensional convergence holds: $(\nu_n(\pi_1),\ldots,\nu_n(\pi_J)) \to_d (\nu(\pi_1),\ldots,\nu(\pi_J))$, and (iii) $\{\nu_n(\pi) : \pi \in \Pi\}$ is stochastically equicontinuous with respect to $|\cdot|$.

Condition (i) holds because Π is a compact subset of R. Condition (ii) holds by Theorem 3.2 in PP applied to the linear combination

$$
\sum_{j=1}^{J} \alpha_j \nu_n(\pi_j) = n^{-1/4} \sum_{t=1}^{n} \left\{ \sum_{j=1}^{J} \alpha_j g(X_t, \pi_j) \right\} u_t,
$$

for arbitrary scalars $\left\{\alpha_j : j = 1,\ldots,J\right\}$, yielding

$$
\begin{aligned}
\sum_{j=1}^{J} \alpha_j \nu_n(\pi_j) &\to_d \left\{\sigma_u^2 L(1,0)\right\}^{1/2} \times N\left(0, \int_{-\infty}^{\infty} \left(\sum_{j=1}^{J} \alpha_j g\left(s,\pi_j\right)\right)^2 ds\right) \\
&:= \left\{\sigma_u^2 L(1,0)\right\}^{1/2} \sum_{j=1}^{J} \alpha_j Z(\pi_j) := \sum_{j=1}^{J} \alpha_j \nu(\pi_j),
\end{aligned}
$$

where $\alpha' = (\alpha_1,\ldots,\alpha_J)$, and $\nu(\pi_j) := \sigma_u L(1,0)^{1/2} Z\left(\pi_j\right)$, where $Z(\pi)$ is a Gaussian process with covariance kernel

$$
\mathrm{E}(Z(\pi_a) Z(\pi_b)) = k_Z(\pi_a,\pi_b) = \int_{-\infty}^{\infty} g(s,\pi_a) g(s,\pi_b) ds.
$$

Now we show condition (iii). The proof uses Theorem 2 in Hansen (1996) and is similar to that of Lemma A.2. For brevity, we borrow arguments from the proof of Lemma A.2 and make adjustments when necessary. To keep notation consistent with Lemma A.2, let

$$h_\pi(X_t,n,u_t)=n^{1/4}g(X_t,\pi)u_t \quad \text{and} \quad \nu_n h_\pi = n^{-1/2}\sum_{t=1}^{n} h_\pi(X_t,n,u_t)\equiv \nu_n(\pi). \tag{C.5}$$

Notice that here the first argument of h_π is X_t instead of the rescaled version $X_{n,t}$ as in Lemma A.2. Let the set of functions $\mathcal{F}$ and the semidistance d be defined the same way as in the paragraph preceding Lemma A.2. Let another semidistance d^h be defined as in (A.3). The semidistance d^h is well defined because for any $\pi,\pi'\in\Pi$,

$$\begin{aligned} d^h\left(h_\pi,h_{\pi'}\right) &= \limsup_{n\to\infty}\sigma_u\left[n^{-1/2}\sum_{t=1}^{n}\mathrm{E}\left[g(X_t,\pi)-g(X_t,\pi')\right]^2\right]^{1/2}\\ &\le \limsup_{n\to\infty}\sigma_u\left[\mathrm{E}n^{-1/2}\sum_{t=1}^{n}T(X_t)^2\right]^{1/2}\left|\pi-\pi'\right|\\ &= \bar{C}_b d\left(h_\pi,h_{\pi'}\right), \end{aligned} \tag{C.6}$$

where $\bar{C}_b=\sigma_u\left[\sup_n \mathrm{E}n^{-1/2}\sum_{t=1}^{n}T(X_t)^2\right]^{1/2}<\infty$, the first equality holds by Assumptions 2.1(b) and (c), and the inequality holds by Assumption 4.2(a). The second equality holds by part (ii).

We verify the four conditions of Hansen (1996) listed in the proof of Lemma A.2. Condition (i) holds by the same argument as in Lemma A.2. Condition (ii) is established with $b^*(X_t,n,u_t)=n^{1/4}T(X_t)|u_t|$ because, by Assumption 4.2(a),

$$\begin{aligned} \left|h_\pi(X_t,n,u_t)-h_{\pi'}(X_t,n,u_t)\right| &= n^{1/4}\left|g(X_t,\pi)-g\left(X_t,\pi'\right)\right||u_t|\\ &\le n^{1/4}T(X_t)|u_t|\left|\pi-\pi'\right|. \end{aligned} \tag{C.7}$$

Condition (iii) holds because for every $\pi\in\Pi$,

$$\begin{aligned} \limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n}\mathrm{E}h_\pi(X_t,n,u_t)^2 &= \limsup_{n\to\infty}\sigma_u^2 n^{-1/2}\sum_{t=1}^{n}\mathrm{E}g(X_t,\pi)^2\\ &<\infty, \end{aligned} \tag{C.8}$$

where the equality holds by Assumptions 2.1(b) and (c) and the inequality holds by part (i). Condition (iv) holds by exactly the same arguments as those for (C.6). Therefore, Hansen (1996) applies, and $\nu_n(\pi):\pi\in\Pi$ is stochastically equicontinuous. ■

Proof of Lemma 4.2. Observe first that

$$\begin{aligned} D_n(\pi,\pi_n) &= n^{-1}\sum_{t=1}^{n}[\hat\beta_n^2(\pi)g^2(X_t,\pi)-\hat\beta_n^2(\pi_0)g^2(X_t,\pi_n)]\\ &\quad -2n^{-1}\sum_{t=1}^{n}[\hat\beta_n(\pi)g(X_t,\pi)Y_t-\hat\beta_n(\pi_n)g(X_t,\pi_n)Y_t]\\ &= \frac{n^{-1}\left(\sum_{t=1}^{n}Y_t g(X_t,\pi_n)\right)^2}{\sum_{t=1}^{n}g^2(X_t,\pi_n)}-\frac{n^{-1}\left(\sum_{t=1}^{n}Y_t g(X_t,\pi)\right)^2}{\sum_{t=1}^{n}g^2(X_t,\pi)}. \end{aligned} \tag{C.9}$$

(i) By Assumption 4.2 in this paper and Lemma A.6 in PP, $g^2(X_t,\pi)$ and $g(X_t,\pi)$ $g(X_t,\pi')$: $(\pi,\pi') \in \Pi^2$ are I-regular. By Theorem 3.2 in PP we have

$$n^{-1/2}\sum_{t=1}^{n} g^2(X_t,\pi) \to_p L(1,0)\int_{-\infty}^{\infty} g^2(s,\pi)ds,$$
$$n^{-1/2}\sum_{t=1}^{n} g(X_t,\pi)g(X_t,\pi') \to_p L(1,0)\int_{-\infty}^{\infty} g(s,\pi)g(s,\pi')ds, \tag{C.10}$$

uniformly over $(\pi,\pi') \in \Pi^2$. Also, by Lemma 4.1,

$$n^{-1/2}\beta_n^{-1}\sum_{t=1}^{n} g(X_t,\pi)u_t \to_p 0, \quad \text{uniformly over } \pi \in \Pi. \tag{C.11}$$

Equations (C.10) and (C.11) combined give us the probability limit of the second term in (C.9):

$$\begin{aligned} & n^{1/2}\beta_n^{-2}\frac{n^{-1}\left(\sum_{t=1}^{n} Y_t g(X_t,\pi)\right)^2}{\sum_{t=1}^{n} g^2(X_t,\pi)} \\ &= \frac{\left(n^{-1/2}\beta_n^{-1}\sum_{t=1}^{n} u_t g(X_t,\pi) + n^{-1/2}\sum_{t=1}^{n} g(X_t,\pi)g(X_t,\pi_n)\right)^2}{n^{-1/2}\sum_{t=1}^{n} g^2(X_t,\pi)} \\ &\to_p \frac{\left[\int_{-\infty}^{\infty} g(s,\pi)g(s,\pi_0)ds\right]^2}{\int_{-\infty}^{\infty} g^2(s,\pi)ds} \times L(1,0), \quad \text{uniformly over } \pi \in \Pi. \end{aligned} \tag{C.12}$$

The probability limit of the first term in (C.9) is a special case of the second term. Therefore, part (i) is proved.

(ii) In part (ii), because $n^{-1/4}\beta_n^{-1} \to c^{-1}$, the covariance term $n^{-1/2}\beta_n^{-1}\sum_{t=1}^{n}$ $g(X_t,\pi)u_t$ does not vanish in the limit. Thus, we need the joint asymptotic distribution of the stochastic processes $n^{-1/2}\sum_{t=1}^{n} g^2(X_t,\pi)$, $n^{-1/2}\sum_{t=1}^{n} g(X_t,\pi)g(X_t,\pi')$, and $\nu_n(\pi) : (\pi,\pi') \in \Pi^2$. Equation (C.10) implies that the sequence of stochastic processes $\{\nu_n^g(\pi,\pi') : (\pi,\pi') \in \Pi^2\}$ converges weakly to $\nu^g(\pi,\pi') : (\pi,\pi') \in \Pi^2$, where

$$\nu_n^g(\pi,\pi') = \begin{pmatrix} n^{-1/2}\sum_{t=1}^{n} g^2(X_t,\pi) \\ n^{-1/2}\sum_{t=1}^{n} g(X_t,\pi)g(X_t,\pi') \end{pmatrix},$$
$$\nu^g(\pi,\pi') = \begin{pmatrix} L(1,0)\int_{-\infty}^{\infty} g^2(s,\pi)ds \\ L(1,0)\int_{-\infty}^{\infty} g(s,\pi)g(s,\pi')ds \end{pmatrix}. \tag{C.13}$$

It follows from equation (46) and surrounding arguments in PP that joint convergence applies and we have

$$\begin{pmatrix} \nu_n^g(\cdot,\cdot) \\ \nu_n(\cdot) \end{pmatrix} \to_d \begin{pmatrix} \nu^g(\cdot,\cdot) \\ \nu(\cdot) \end{pmatrix}. \tag{C.14}$$

Then, by the CMT,

$$
\begin{aligned}
nD_n(\pi,\pi_n) \to_d\ & \frac{\left\{cL\,(1,0)\int_{-\infty}^{\infty} g^2(s,\pi_0)ds + L\,(1,0)^{1/2}\,Z\,(\pi_0)\right\}^2}{L\,(1,0)\int_{-\infty}^{\infty} g^2(s,\pi_0)ds} \\
& - \frac{\left\{cL\,(1,0)\int_{-\infty}^{\infty} g(s,\pi_0)g(s,\pi)ds + L\,(1,0)^{1/2}\,Z\,(\pi)\right\}^2}{L\,(1,0)\int_{-\infty}^{\infty} g^2(s,\pi)ds} \\
& = \left\{cL\,(1,0)^{1/2}\left(\int_{-\infty}^{\infty} g^2(s,\pi_0)ds\right)^{1/2} + \frac{Z\,(\pi_0)}{\left(\int_{-\infty}^{\infty} g^2(s,\pi_0)ds\right)^{1/2}}\right\}^2 \\
& - \left\{cL\,(1,0)^{1/2}\frac{\int_{-\infty}^{\infty} g(s,\pi_0)g(s,\pi)ds}{\left(\int_{-\infty}^{\infty} g^2(s,\pi)ds\right)^{1/2}} + \frac{Z\,(\pi)}{\left(\int_{-\infty}^{\infty} g^2(s,\pi)ds\right)^{1/2}}\right\}^2,
\end{aligned}
$$

and part (ii) holds. ■

Proof of Lemma 4.3. Assumptions 4.2(a) and 4.3 imply that every sample path of $D(c,\pi,\pi_0)$ is continuous in π. Because Π is compact, every sample path of $D(c,\pi,\pi_0)$ achieves its minimum on Π.

We now show that the minimizer of $D(c,\pi,\pi_0)$ is unique with probability one using the technique in the proof of Lemma 3.2 in Cheng (2008), which is based on Kim and Pollard (1990). First, observe that minimizing $D(c,\cdot,\pi_0)$ is equivalent to maximizing $A^2(\pi)$ where

$$
A(\pi) \equiv \frac{cL^{1/2}(1,0)\int_{-\infty}^{\infty} g(s,\pi_0)g(s,\pi)ds}{\left[\int_{-\infty}^{\infty} g^2(s,\pi)ds\right]^{1/2}} + \frac{Z\,(\pi)}{\left(\int_{-\infty}^{\infty} g^2(s,\pi)ds\right)^{1/2}}. \tag{C.15}
$$

Because $L^{1/2}(1,0)$ and Z are independent, conditional on $L^{1/2}(1,0)$, $A(\pi)$ is a Gaussian process. By the proof of Lemma 3.2 in Cheng (2008), we only need to show that for all $\pi_1 \neq \pi_2$,

$$
\operatorname{Var}(A(\pi_1)-A(\pi_2)|L^{1/2}(1,0)) > 0 \quad \text{and} \quad \operatorname{Var}(A(\pi_1)+A(\pi_2)|L^{1/2}(1,0)) > 0, \quad \text{a.s.} \tag{C.16}
$$

Now

$$
\begin{aligned}
A(\pi_1)-A(\pi_2) = {} & cL^{1/2}(1,0)\int_{-\infty}^{\infty} g(s,\pi_0)\left[q\,(s,\pi_1)-q\,(s,\pi_2)\right]ds \\
& + \left[W\,(\pi_1)-W\,(\pi_2)\right],
\end{aligned}
$$

where

$$
q\,(s,\pi) = \frac{g(s,\pi)}{\left[\int_{-\infty}^{\infty} g^2(a,\pi)da\right]^{1/2}}, \qquad W\,(\pi) = \frac{Z\,(\pi)}{\left(\int_{-\infty}^{\infty} g^2(s,\pi)ds\right)^{1/2}}.
$$

The first inequality in (C.16) holds because $L(1,0)$ is independent of $Z\,(\pi)$ and so

$$
\operatorname{Var}(A(\pi_1)-A(\pi_2)|L^{1/2}(1,0)) = \operatorname{Var}\left[W\,(\pi_1)-W\,(\pi_2)\right] > 0, \tag{C.17}
$$

where the inequality holds by Assumption 4.4 and the fact that

$$\mathrm{Var}\left[W(\pi_1) - W(\pi_2)\right] = 2\sigma_u^2 \left\{ 1 - \frac{\int_{-\infty}^{\infty} g(s,\pi_1) g(s,\pi_2) da}{\left[\int_{-\infty}^{\infty} g^2(s,\pi_1) ds \int_{-\infty}^{\infty} g^2(s,\pi_2) ds\right]^{1/2}} \right\} > 0$$

for $\pi_1 \neq \pi_2$. The second inequality in (C.16) holds because

$$\begin{aligned} \mathrm{Var}(A(\pi_1) + A(\pi_2) | L^{1/2}(1,0)) &= \mathrm{Var}[W(\pi_1) + W(\pi_2)] \\ &= 2\sigma_u^2 \left\{ 1 + \frac{\int_{-\infty}^{\infty} g(s,\pi_1) g(s,\pi_2) da}{\left[\int_{-\infty}^{\infty} g^2(s,\pi_1) ds \int_{-\infty}^{\infty} g^2(s,\pi_2) ds\right]^{1/2}} \right\} > 0, \end{aligned}$$

again by Assumption 4.4. ■

Proof of Lemma 5.1. Lemma 5.1 is a direct application of Theorem 3.23 of Kallenberg (2001, p. 57). The moment condition in that theorem holds because $\left(\int \mathrm{E}b^2(V)\right) < \infty$ by Assumption 5.2(c) and

$$\begin{aligned} \mathrm{E}\left(\int \left(h(V,\pi) - h(V,\pi')\right) dU\right)^2 &= \int \mathrm{E}\left(h(V,\pi) - h(V,\pi')\right)^2 \\ &\leq \left(\int \mathrm{E}b^2(V)\right)\left(\pi - \pi'\right)^2, \end{aligned} \tag{C.18}$$

where the equality holds by the fundamental property of the stochastic integral and the inequality holds by Assumptions 5.1(a) and (b) (also see the remark following Assumption 5.1). ■

Proof of Lemma 5.2. Because $g(x,\pi)$ is H-regular on π (Assumption 5.1(a)), we have for each $\pi \in \Pi$

$$\begin{aligned} \nu_n h_\pi &\equiv n^{-1/2} \kappa^{-1}(n^{1/2}, \pi) \sum_{t=1}^{n} g(X_t, \pi) u_t \\ &= n^{-1/2} \sum_{t=1}^{n} h(X_{n,t}, \pi) u_t + n^{-1/2} \kappa^{-1}(n^{1/2}, \pi) \sum_{t=1}^{n} R(X_{n,t}, \pi, n^{1/2}) u_t \\ &= n^{-1/2} \sum_{t=1}^{n} h(X_{n,t}, \pi) u_t + o_p(1), \end{aligned} \tag{C.19}$$

where the last equality holds by Lemma A.5(ii) in PP.

Let the random process $(\nu h_\pi : \pi \in \Pi) := (\int h(V,\pi) dU : \pi \in \Pi)$. Then Lemma A.1 and (C.19) give

$$(\nu_n h_{\pi_1}, \ldots, \nu_n h_{\pi_k})' \to_p (\nu h_{\pi_1}, \ldots, \nu h_{\pi_k})'. \tag{C.20}$$

For all $\delta > 0$, by Assumption 3.1, there exists $\pi_1, \pi_2, \ldots, \pi_{k(\delta)}$, $k(\delta) < \infty$ such that

$$\sup_{\pi \in \Pi} \inf_{j \leq k(\delta)} \left|\pi - \pi_j\right| < \delta. \tag{C.21}$$

Then we have

$$\begin{aligned}
&\sup_{\pi\in\Pi} |\nu_n h_\pi - \nu h_\pi| \\
&\quad = \max_{j\le k(\delta)} \sup_{\pi\in\Pi:|\pi-\pi_j|\le\delta} |\nu_n h_\pi - \nu_n h_{\pi_j} + \nu_n h_{\pi_j} - \nu h_{\pi_j} + \nu h_{\pi_j} - \nu h_\pi| \\
&\quad \le \sup_{\pi\in\Pi:|\pi-\pi'|\le\delta} |\nu_n h_\pi - \nu_n h_{\pi'}| + \max_{j\le k(\delta)} |\nu_n h_{\pi_j} - \nu h_{\pi_j}| + \sup_{\pi\in\Pi:|\pi-\pi'|\le\delta} |\nu h_\pi - \nu h_{\pi'}| \\
&\quad \equiv A_n(\delta) + B_n(\delta) + C_n(\delta). \qquad \textbf{(C.22)}
\end{aligned}$$

Fix an $\varepsilon > 0$. By Lemma A.2, for all $\zeta > 0$, there exists a $\delta_A > 0$ small enough such that

$$\limsup_{n\to\infty} \Pr(A_n(\delta_A) > \varepsilon/3) \le \zeta. \qquad \textbf{(C.23)}$$

By Lemma 5.1 and the remark there, νh_π is continuous with probability one. Because Π is compact, νh_π is uniformly continuous with probability one. Thus, $\lim_{\delta\to 0} C_n(\delta) = 0$ a.s. This implies the existence of a $\delta_C > 0$ small enough such that

$$\Pr(C_n(\delta_C) > \varepsilon/3) \le \zeta. \qquad \textbf{(C.24)}$$

Let $\delta_{\min} = \min\{\delta_A, \delta_C\}$. By (C.20),

$$\limsup_{n\to\infty} \Pr(B_n(\delta_{\min}) > \varepsilon/3) = 0. \qquad \textbf{(C.25)}$$

Combining (C.23), (C.24), and (C.25), we get

$$\begin{aligned}
&\limsup_{n\to\infty} \Pr\Big(\sup_{\pi\in\Pi} |\nu_n h_\pi - \nu h_\pi| > \varepsilon\Big) \\
&\quad \le \limsup_{\zeta\to 0} \Big[\limsup_{n\to\infty} \Pr(A_n(\delta_{\min}) > \varepsilon/3) + \limsup_{n\to\infty} \Pr(B_n(\delta_{\min}) > \varepsilon/3) \\
&\qquad\qquad + \limsup_{n\to\infty} \Pr(C_n(\delta_{\min}) > \varepsilon/3)\Big] \\
&\quad \le \limsup_{\zeta\to 0} 2\zeta = 0. \qquad \textbf{(C.26)}
\end{aligned}$$

Therefore, $\sup_{\pi\in\Pi} |\nu_n h_\pi - \nu h_\pi| \to_p 0$, and Lemma 5.2 is proved. ■

Proof of Lemma 5.3. As in the proof of Lemma 4.2, we have

$$D_n(\pi,\pi_n) = \frac{n^{-1}\left(\sum_{t=1}^n Y_t g(X_t,\pi_n)\right)^2}{\sum_{t=1}^n g^2(X_t,\pi_n)} - \frac{n^{-1}\left(\sum_{t=1}^n Y_t g(X_t,\pi)\right)^2}{\sum_{t=1}^n g^2(X_t,\pi)}. \qquad \textbf{(C.27)}$$

The denominator of the second term on the right side of (C.27) converges a.s. when properly scaled:

$$n^{-1}\kappa^{-2}(n^{1/2},\pi) \sum_{t=1}^n g^2(X_t,\pi) \to \int h^2(V,\pi) \quad \text{a.s.,} \qquad \textbf{(C.28)}$$

uniformly over π, by Theorem 3.3 in PP. We prove part (i) and part (ii) next using the preceding equations.

(i) We have

$$\begin{aligned}
&\beta_n^{-1} n^{-1} \kappa^{-1}(n^{1/2}, \pi) \kappa^{-1}(n^{1/2}, \pi_n) \sum_{t=1}^{n} Y_t g(X_t, \pi) \\
&\quad = n^{-1} \kappa^{-1}(n^{1/2}, \pi) \kappa^{-1}(n^{1/2}, \pi_n) \sum_{t=1}^{n} g(X_t, \pi) g(X_t, \pi_n) \\
&\qquad + \beta_n^{-1} n^{-1} \kappa^{-1}(n^{1/2}, \pi) \kappa^{-1}(n^{1/2}, \pi_n) \sum_{t=1}^{n} u_t g(X_t, \pi) \\
&\quad \to \int h(V, \pi) h(V, \pi_0) \quad \text{a.s., uniformly over } \pi \in \Pi, \qquad \textbf{(C.29)}
\end{aligned}$$

where the convergence holds by Theorem 3.3 in PP, Lemma 5.1, and Assumption 5.1(a). Part (i) is implied by (C.27)–(C.29).

(ii) We have

$$\begin{aligned}
&n^{-1/2} \kappa^{-1}(n^{1/2}, \pi) \sum_{t=1}^{n} Y_t g(X_t, \pi) \\
&\quad = (\beta_n n^{1/2} \kappa(n^{1/2}, \pi_n)) n^{-1} \kappa^{-1}(n^{1/2}, \pi) \kappa^{-1}(n^{1/2}, \pi_n) \\
&\qquad \times \sum_{t=1}^{n} g(X_t, \pi) g(X_t, \pi_n) + n^{-1/2} \kappa^{-1}(n^{1/2}, \pi) \sum_{t=1}^{n} u_t g(X_t, \pi) \\
&\quad \to_p c \int h(V, \pi) h(V, \pi_0) + \int h(V, \pi) dU, \qquad \textbf{(C.30)}
\end{aligned}$$

uniformly over $\pi \in \Pi$, where the convergence holds by Theorem 3.3 in PP, Lemma 5.1, and Assumption 5.1(a). Part (ii) is implied by (C.27), (C.28), and (C.30). ■

Proof of Lemma 5.4. Let

$$A(c, \pi) = \frac{c \int h(V, \pi) h(V, \pi_0) + \int h(V, \pi) dU}{\left[\int h^2(V, \pi)\right]^{1/2}}. \qquad \textbf{(C.31)}$$

First we show that $A^2(c, \pi)$ has a continuous sample path with probability one. This is done by showing that (a) the denominator and the numerator are continuous with probability one, and (b) the denominator is strictly positive with probability one. Condition (a) holds by Lemma A.8 in PP and Lemma 5.1. Condition (b) holds because

$$\int h^2(V, \pi) = \int_{-\infty}^{\infty} h^2(s, \pi) L(1, s) ds > 0 \quad \text{a.s.}, \qquad \textbf{(C.32)}$$

where the equality holds by the occupation time formula (e.g., PP) and the inequality holds by Assumption 5.1(c).

To show that $A^2(c, \pi)$ has a unique maximum, it suffices to show that with probability one, no sample path of $A(c, \pi)$ achieves its maximum or minimum at two distinct points in Π, and no sample path has maximum and minimum with the same absolute value.

The procedure used in Lemma 3.2 in Cheng (2008) applies here if we can write $A(c, \pi)$ in terms of continuous Gaussian processes. We can achieve this goal by splitting $U(r)$ into $V(r)$ and a standard Brownian motion, $Z(r)$, independent of $V(r)$, following Phillips (1989):

$$U(r) = a_1 \sigma_u V(r) + a_2 Z(r), \qquad \textbf{(C.33)}$$

where $a_1 = \rho\sigma_u/\sigma_v$ and $a_2 = \sigma_u\sqrt{1-\rho^2}$. Such a $Z(r)$ exists by Assumption 2.1(a). Using (C.33) in $A(c,\pi)$ we get

$$A(c,\pi) = \frac{c\int h(V,\pi)h(V,\pi_0) + a_1\int h(V,\pi)dV + a_2\int h(V,\pi)dZ}{\left[\int h^2(V,\pi)\right]^{1/2}}. \tag{C.34}$$

Because Z is a standard Brownian motion independent of V, conditioning on a sample path of V, $A(c,\pi)$ is a continuous Gaussian process indexed by $\pi \in \Pi$, with covariance kernel:

$$H(c,\pi,\pi') = \frac{a_2^2\int h(V,\pi)h(V,\pi')}{\left(\int h^2(V,\pi)\right)^{1/2}\left(\int h^2(V,\pi')\right)^{1/2}}. \tag{C.35}$$

Subsequently we show that $A^2(c,\pi)|V = v$ has a unique maximum with probability one for all sample paths v of V. This implies that with probability one, $A^2(c,\pi)$ has unique maximum, i.e., Lemma 5.4.

We proceed to show that $A^2(c,\pi)|V = v$ has a unique maximum. We apply the procedure in the proof of Lemma 3.2 in Cheng (2008). By Cheng's argument, it suffices to show that for $\pi \neq \pi'$,

$$\begin{aligned}&\operatorname{Var}\left(\frac{a_2\int h(V,\pi)dZ}{\left[\int h^2(V,\pi)\right]^{1/2}} - \frac{a_2\int h(V,\pi')dZ}{\left[\int h^2(V,\pi')\right]^{1/2}}\,\middle|\, V = v\right) > 0 \quad \text{and}\\ &\operatorname{Var}\left(\frac{a_2\int h(V,\pi)dZ}{\left[\int h^2(V,\pi)\right]^{1/2}} + \frac{a_2\int h(V,\pi')dZ}{\left[\int h^2(V,\pi')\right]^{1/2}}\,\middle|\, V = v\right) > 0.\end{aligned} \tag{C.36}$$

The preceding inequalities are equivalent to

$$H(c,\pi,\pi) + H(c,\pi',\pi') \pm 2H(c,\pi,\pi') > 0 \tag{C.37}$$

or, equivalently,

$$2 \pm \frac{\int h(V,\pi)h(V,\pi')}{\left(\int h^2(V,\pi)\right)^{1/2}\left(\int h^2(V,\pi')\right)^{1/2}} > 0, \tag{C.38}$$

which holds by the Cauchy–Schwarz inequality and Assumption 5.1(d). ■