Central Limit Theory for Combined Cross-Section and Time Series

Jinyong Hahn* Guido Kuersteiner† Maurizio Mazzocco‡

UCLA University of Maryland UCLA

October 5, 2016

Abstract

Combining cross-section and time series data is a long and well established practice in empirical economics. We develop a central limit theory that explicitly accounts for possible dependence between the two data sets. We focus on common factors as the mechanism behind this dependence. Using our central limit theorem (CLT) we establish the asymptotic properties of parameter estimates of a general class of models based on a combination of cross-sectional and time series data, recognizing the interdependence between the two data sources in the presence of aggregate shocks. Despite the complicated nature of the analysis required to formulate the joint CLT, it is straightforward to implement the resulting parameter limiting distributions due to a formal similarity of our approximations with the standard Murphy and Topel's (1985) formula.

*UCLA, Department of Economics, 8283 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095, hahn@econ.ucla.edu

†University of Maryland, Department of Economics, Tydings Hall 3145, College Park, MD, 20742, kuersteiner@econ.umd.edu

‡UCLA, Department of Economics, 8283 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095, mmazzocc@econ.ucla.edu

1 Introduction

There is a long tradition in empirical economics of relying on information from a variety of data sources to estimate model parameters. In this paper we focus on a situation where cross-sectional and time-series data are combined. This may be done for a variety of reasons. Some parameters may not be identified in the cross section or time series alone. Alternatively, parameters estimated from one data source may be used as a first-step inputs in the estimation of a second set of parameters based on a different data source. This may be done to reduce the dimension of the parameter space or more generally for computational reasons.

Data combination generates theoretical challenges, even when only cross-sectional data sets or time-series data sets are combined. See Ridder and Moffitt (2007) for example. We focus on dependence between cross-sectional and time-series data produced by common aggregate factors. Andrews (2005) demonstrates that even randomly sampled cross-sections lead to independently distributed samples only conditionally on common factors, since the factors introduce a possible correlation. This correlation extends inevitably to a time series sample that depends on the same underlying factors.

The first contribution of this paper is to develop a central limit theory that explicitly accounts for the dependence between the cross-sectional and time-series data by using the notion of stable convergence. The second contribution is to use the corresponding central limit theorem to derive the asymptotic distribution of parameter estimators obtained from combining the two data sources.

Our analysis is inspired by a number of applied papers and in particular by the discussion in Lee and Wolpin (2006, 2010). Econometric estimation based on the combination of cross-sectional and time-series data is an idea that dates back at least to Tobin (1950). More recently, Heckman and Sedlacek (1985) and Heckman, Lochner, and Taber (1998) proposed to deal with the estimation of equilibrium models by exploiting such data combination. It is, however, Lee and Wolpin (2006, 2010) who develop the most extensive equilibrium model and estimate it using a similar intuition and panel data.

To derive the new central limit theorem and the asymptotic distribution of parameter estimates, we extend the model developed in Lee and Wolpin (2016, 2010) to a general setting that involves two submodels. The first submodel includes all the cross-sectional features, whereas the second submodel is composed of all the time-series aspects. The two submodels are linked by a vector of aggregate shocks and by the parameters that govern their dynamics. Given the interplay between the two submodels, the aggregate shocks have complicated effects on the estimation of the parameters of interest.

With the objective of creating a framework to perform inference in our general model, we first derive a joint functional stable central limit theorem for cross-sectional and time-series data. The central limit theorem explicitly accounts for the factor-induced dependence between the two samples even when the cross-sectional sample is obtained by random sampling, a special case covered by our theory. We derive the central limit theorem under the condition that the dimension of the cross-sectional data n as well as the dimension of the time series data τ go to infinity. Using our central limit theorem we then derive the asymptotic distribution of the parameter estimators that characterize our general model. To our knowledge, this is the first paper that derives an asymptotic theory that combines cross sectional and time series data. In order to deal with parameters estimated using two data sets of completely different nature, we adopt the notion of stable convergence. Stable convergence dates back to Reyni (1963) and was recently used in Kuersteiner and Prucha (2013) in a panel data context to establish joint limiting distributions. Using this concept, we show that the asymptotic distributions of the parameter estimators are a combination of asymptotic distributions from cross-sectional analysis and time-series analysis.

While the formal derivation of the asymptotic distribution may appear complicated, the asymptotic formulae that we produce are straightforward to implement and very similar to the standard Murphy and Topel's (1985) formula.

We also derive a novel result related to the unit root literature. We show that, when the time-series data are characterized by unit roots, the asymptotic distribution is a combination of a normal distribution and the distribution found in the unit root literature. Therefore, the asymptotic distribution exhibits mathematical similarities to the inferential problem in predictive regressions, as discussed by Campbell and Yogo (2006). However, the similarity is superficial in that Campbell and Yogo's (2006) result is about an estimator based on a single data source. But, similarly to Campbell and Yogo's analysis, we need to address problems of uniform inference. Phillips (2014) proposes a method of uniform inference for predictive regressions, which we adopt and modify to our own estimation problem in the unit root case.

Our results should be of interest to both applied microecomists and macroeconomists. Data combination is common practice in the macro calibration literature where typically a subset of parameters is determined based on cross-sectional studies. It is also common in structural microeconomics where the focus is more directly on identification issues that cannot be resolved in the cross-section alone. In a companion paper, Hahn, Kuersteiner, and Mazzocco (2016), we discuss in detail specific examples from the micro literature. In the companion paper, we also provide a more intuitive analysis of the joint use in estimation of cross-sectional and time-series data when aggregate factors are present, whereas in this paper the analysis is more technical and abstract.

The remainder of the paper is organized as follows. In Section 2, we introduce the general statistical model. In Section 3, we present the intuition underlying our main result, which is presented in Section 4.

2 Model

We assume that our cross-sectional data consist of $\{y_{i,t},\ i = 1, \ldots, n, t = 1, ..., T\}$, where the start time of the cross-section or panel, $t = 1$, is an arbitrary normalization of time. Pure cross-sections are handled by allowing for $T = 1$. Note that T is fixed and finite throughout our discussion while our asymptotic approximations are based on n tending to infinity. Our time series data consist of $\{z_s,\ s = \tau_0 + 1, \ldots, \tau_0 + \tau\}$ where the time series sample size τ tends to infinity jointly with n. The start point of the time series sample is fixed at an arbitrary time $\tau_0 \in (-\infty, \infty)$. The vector $y_{i,t}$ includes all information related to the cross-sectional submodel, where i is an index for individuals, households or firms, and t denotes the time period when the cross-sectional unit is observed. The second vector z_s contains aggregate data.

The technical assumptions for our CLT, detailed in Section 4, do not directly restrict the data, nor do they impose restrictions on how the data were sampled. For example, we do not assume that the cross-sectional sample was obtained by randomized sampling, although this is a special case that is covered by our assumptions. Rather than imposing restrictions directly on the data we postulate that there are two parametrized models that implicitly restrict the data. The function $f(y_{i,t} | \beta, \nu_t, \rho)$ is used to model $y_{i,t}$ as a function of cross-sectional parameters β, common shocks ν_t and time series parameters ρ. In the same way the function $g(z_s | \beta, \rho)$ restricts the behavior of some time series variables z_s.[1]

Depending on the exact form of the underlying economic model, the functions f and g may have different interpretations. They could be the likelihoods of $y_{i,t}$, conditional on ν_t, and z_s respectively. In a likelihood setting, f and g impose restrictions on $y_{i,t}$ and z_s because of the implied martingale properties of the score process. More generally, the functions f and g may be the basis for method of moments (the exactly identified case) or GMM (the overidentified case) estimation. In these situations parameters are identified from the conditions $E_C\left[f\left(y_{i,t}\right|\beta,\nu_t,\rho)\right]=0$ given the shock ν_t and $E_\tau\left[g\left(z_s|\beta,\rho\right)\right]=0$. The first expectation, E_C, is understood as being over the cross-section population distribution holding $\nu=(\nu_1,...,\nu_T)$ fixed, while the second, E_τ, is over the stationary distribution of the time-series data generating process. The moment conditions follow from martingale assumptions we directly impose on f and g. In our companion paper we discuss examples of economic models that rationalize these assumptions.

Whether we are dealing with likelihoods or moment functions, the central limit theorem is directly formulated for the estimating functions that define the parameters. We use the notation $F_n\left(\beta,\nu_t,\rho\right)$ and $G_\tau\left(\beta,\rho\right)$ to denote the criterion function based on the cross-section and time series respectively. When the model specifies a likelihood these functions are defined as $F_n\left(\beta,\nu,\rho\right)=\frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n}f\left(y_{i,t}\right|\beta,\nu_t,\rho)$ and $G_\tau\left(\beta,\rho\right)=\frac{1}{\tau}\sum_{s=1}^{\tau}g\left(z_s\right|\beta,\rho)$. When the model specifies moment conditions we let $h_n\left(\beta,\nu,\rho\right)=\frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n}f\left(y_{i,t}\right|\beta,\nu_t,\rho)$ and $k_\tau\left(\beta,\rho\right)=\frac{1}{\tau}\sum_{s=1}^{\tau}g\left(z_s\right|\beta,\rho)$. The GMM or moment based criterion functions are then given by $F_n\left(\beta,\nu,\rho\right)=-h_n\left(\beta,\nu,\rho\right)'W_n^C h_n\left(\beta,\nu,\rho\right)$ and $G_\tau\left(\beta,\rho\right)=-k_\tau\left(\beta,\rho\right)'W_\tau^\tau k_\tau\left(\beta,\rho\right)$ with $W_n^\times$ and W_τ^T two almost surely positive definite weight matrices. The use of two separate objective functions is helpful in our context because it enables us to discuss which issues arise if only cross-sectional variables or only time-series variables are used in the estimation.[2]

We formally justify the use of two data sets by imposing restrictions on the identifiability of

[1] The function g may naturally arise if the ν_t is an unobserved component that can be estimated from the aggregate time series once the parameters β and ρ are known, i.e., if $\nu_t\equiv\nu_t\left(\beta,\rho\right)$ is a function of $\left(z_t,\beta,\rho\right)$ and the behavior of ν_t is expressed in terms of ρ. Later, we allow for the possibility that g in fact is derived from the conditional density of ν_t given ν_{t-1}, i.e., the possibility that g may depend on both the current and lagged values of z_t. For notational simplicity, we simply write $g\left(z_s\right|\beta,\rho)$ here for now.

[2] Note that our framework covers the case where the joint distribution of $\left(y_{it},z_t\right)$ is modelled. Considering the two components separately adds flexibility in that data is not required for all variables in the same period.

parameters through the cross-section and time series criterion functions alone. We denote the probability limit of the objective functions by $F(\beta, \nu_t, \rho)$ and $G(\beta, \rho)$, in other words,

$$F(\beta, \nu, \rho) = \operatorname*{plim}_{n \to \infty} F_n(\beta, \nu, \rho),$$
$$G(\beta, \rho) = \operatorname*{plim}_{\tau \to \infty} G_\tau(\beta, \rho).$$

The true or pseudo true parameters are defined as the maximizers of these probability limits

$$(\beta(\rho), \nu(\rho)) \equiv \operatorname*{argmax}_{\beta, \nu} F(\beta, \nu, \rho), \tag{1}$$
$$\rho(\beta) \equiv \operatorname*{argmax}_{\rho} G(\beta, \rho), \tag{2}$$

and we denote with β_0 and ρ_0 the solutions to (1) and (2). The idea that neither F nor G alone are sufficient to identify both parameters is formalized as follows. If the function F is constant in ρ at the parameter values β and ν that maximize it then ρ is not identified by the criterion F alone. Formally we state that

$$\max_{\beta, \nu_t} F(\beta, \nu, \rho) = \max_{\beta, \nu_t} F(\beta, \nu, \rho_0) \quad \text{for all } \rho \in \Theta_\rho \tag{3}$$

It is easy to see that (3) is not a sufficient condition to restrict identification in a desirable way. For example (3) is satisfied in a setting where F does not depend at all on ρ. In that case the maximizers in (1) also do not depend on ρ and by definition coincide with β_0 and ν_0. To rule out this case we require that ρ_0 is needed to identify β_0 and ν_0. Formally, we impose the condition that

$$(\beta(\rho), \nu(\rho)) \neq (\beta_0, \nu_0) \quad \text{for all } \rho \neq \rho_0. \tag{4}$$

Similarly, we impose restrictions on the time series criterion functions that insure that the parameters β and ρ cannot be identified solely as the maximizers of G. Formally, we require that

$$\max_{\rho} G(\beta, \rho) = \max_{\rho} G(\beta_0, \rho) \quad \text{for all } \beta \in \Theta_\beta, \tag{5}$$
$$\rho(\beta) \neq \rho_0 \quad \text{for all } \beta \neq \beta_0.$$

To insure that the parameters can be identified from a combined cross-sectional and time-series data set we impose the following condition. Define $\theta \equiv (\beta', \nu')'$ and assume that (i) there exists a unique solution to the system of equations:

$$\left[\frac{\partial F(\beta,\nu,\rho)}{\partial\theta'}, \frac{\partial G(\beta,\rho)}{\partial\rho'}\right] = 0, \tag{6}$$

and (ii) the solution is given by the true value of the parameters. In summary, our model is characterized by the high level assumptions in (3), (4), (5) and (6).

3 Asymptotic Inference

Our asymptotic framework is such that standard textbook level analysis suffices for the discussion of consistency of the estimators. In standard analysis with a single data source, one typically restricts the moment equation to ensure identification, and imposes further restrictions such that the sample analog of the moment function converges uniformly to the population counterpart. Because these arguments are well known we simply impose as a high-level assumption that our estimators are consistent. The purpose of this section is to provide an overview over our results while a rigorous technical discussion is relegated to Section 4 which may be skipped by a less technically oriented reader.

3.1 Stationary Models

For expositional purposes, suppose that the time series z_t is such that its log of the conditional probability density function given z_{t-1} is $g\left(z_t\right| z_{t-1},\rho)$. To simplify the exposition in this section we assume that the time series model does not depend on the micro parameter β. Let $\widetilde{\rho}$ denote a consistent estimator.

We assume that the dimension of the time series data is τ, and that the influence function of $\widetilde{\rho}$ is such that

$$\sqrt{\tau}\left(\widetilde{\rho}-\rho\right) = \frac{1}{\sqrt{\tau}}\sum_{s=\tau_0+1}^{\tau_0+\tau}\varphi_s \tag{7}$$

with $E\left[\varphi_s\right] = 0$. Here, τ_0+1 denotes the beginning of the time series data, which is allowed to differ from the beginning of the panel data. Using $\widetilde{\rho}$ from the time series data, we can then consider maximizing the criterion $F_n\left(\beta,\nu_t,\rho\right)$ with respect to $\theta = (\beta,\nu_1,\ldots,\nu_T)$. Implicit in this representation is the idea that we are given a short panel for estimation of $\theta = (\beta,\nu_1,\ldots,\nu_T)$, where T denotes the time series dimension of the panel data. In order to emphasize that T is small, we use the term 'cross-section' for the short panel data set, and adopt asymptotics where T is fixed. The moment equation then is

$$\frac{\partial F_n\left(\widehat{\theta}, \widetilde{\rho}\right)}{\partial \theta} = 0$$

and the asymptotic distribution of $\widehat{\theta}$ is characterized by

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) \approx -\left(\frac{\partial^2 F\left(\theta, \widetilde{\rho}\right)}{\partial \theta \partial \theta'}\right)^{-1}\left(\sqrt{n}\frac{\partial F_n\left(\theta, \widetilde{\rho}\right)}{\partial \theta}\right).$$

Because $\sqrt{n}\left(\partial F_n\left(\theta, \widetilde{\rho}\right)/\partial \theta - \partial F_n\left(\theta, \rho\right)/\partial \theta\right) \approx \left(\partial^2 F\left(\theta, \rho\right)/\partial \theta \partial \rho'\right) \frac{\sqrt{n}}{\sqrt{\tau}}\sqrt{\tau}\left(\widetilde{\rho} - \rho\right)$ we obtain

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) \approx -A^{-1}\sqrt{n}\frac{\partial F_n\left(\theta, \rho\right)}{\partial \theta} - A^{-1}B\frac{\sqrt{n}}{\sqrt{\tau}}\left(\frac{1}{\sqrt{\tau}}\sum_{s=\tau_0+1}^{\tau_0+\tau}\varphi_s\right) \tag{8}$$

with

$$A \equiv \frac{\partial^2 F\left(\theta, \rho\right)}{\partial \theta \partial \theta'}, \quad B \equiv \frac{\partial^2 F\left(\theta, \rho\right)}{\partial \theta \partial \rho'}$$

We adopt asymptotics where $n, \tau \to \infty$ at the same rate, but T is fixed. We stress that a technical difficulty arises because we are conditioning on the factors $(\nu_1, \ldots, \nu_T)$. This is accounted for in the limit theory we develop through a convergence concept by Renyi (1963) called stable convergence, essentially a notion of joint convergence. It can be thought of as convergence conditional on a specified σ-field, in our case the σ-field $\mathcal{C}$ generated by $(\nu_1, \ldots, \nu_T)$. In simple special cases, and because T is fixed, the asymptotic distribution of $\frac{1}{\sqrt{\tau}}\sum_{s=\tau_0+1}^{\tau_0+\tau}\varphi_s$ conditional on $(\nu_1, \ldots, \nu_T)$ may be equal to the unconditional asymptotic distribution. However, as we show in Section 4, this is not always the case, even when the model is stationary.

Renyi (1963) and Aldous and Eagleson (1978) show that the concepts of convergence of the distribution conditional on any positive probability event in $\mathcal{C}$ and the concept of stable convergence are equivalent. Eagelson (1975) proves a stable CLT by establishing that the conditional characteristic functions converge almost surely. Hall and Heyde's (1980) proof of stable convergence on the other hand is based on demonstrating that the characteristic function converges weakly in L_1. As pointed out in Kuersteiner and Prucha (2013), the Hall and Heyde (1980) approach lends itself to proving the martingale CLT under slightly weaker conditions than what Eagleson (1975) requires. While both approaches can be used to demonstrate very similar stable and thus conditional limit laws, neither simplifies to conventional marginal weak convergence except in trivial cases. For this reason it is not possible to separate the cross-sectional inference problem from the time series problem simply by 'fixing' the common shocks $(\nu_1, \ldots, \nu_T)$. Such an approach would only be valid if the shocks $(\nu_1, \ldots, \nu_T)$ did not change in any state of the world, in other words if they were constants in a probabilistic sense. Only in that scenario would stable or conditional convergence be equivalent to marginal convergence. The inherent randomness of $(\nu_1, \ldots, \nu_T)$, taken into account by the rational agents in the models we discuss in our companion paper (Hahn, Kuersteiner and Mazzocco, 2016) is at the heart of our examples and is the essence of the inference problems we discuss in that paper. Thus, treating $(\nu_1, \ldots, \nu_T)$ as constants is not an option available to us. This is also the reason why time dummies are no remedy for the problems we analyze. A related idea might be to derive conditional (on $\mathcal{C}$) limiting results separately for the cross-section and time series dimension of our estimators. As noted before, such a result in fact amounts to demonstrating stable convergence, in this case for each dimension separately. Irrespective, this approach is flawed because it does not deliver joint convergence of the two components. It is evident from (8) that the continuous mapping theorem needs to be applied to derive the asymptotic distribution of $\widehat{\theta}$. Because both A and B are $\mathcal{C}$-measurable random variables in the limit the continuous mapping theorem can only be applied if joint convergence of $\sqrt{n}\partial F_n(\theta, \rho)/\partial\theta, \tau^{-1/2}\sum_{s=\tau_0+1}^{\tau_0+\tau}\varphi_s$ and any $\mathcal{C}$-measurable random variable is established. Joint stable convergence of both components delivers exactly that. Finally, we point out that it is perfectly possible to consistently estimate parameters, in our case $(\nu_1, \ldots, \nu_T)$, that remain random in the limit. For related results, see the recent work of Kuersteiner and Prucha (2015).

Here, for the purpose of illustration we consider the simple case where the dependence of the time series component on the factors $(\nu_1, \ldots, \nu_T)$ vanishes asymptotically. Let's say that the unconditional distribution is such that

$$\frac{1}{\sqrt{\tau}}\sum_{s=\tau_0+1}^{\tau_0+\tau}\varphi_s \to N(0, \Omega_\nu)$$

where Ω_ν is a fixed constant that does not depend on $(\nu_1, \ldots, \nu_T)$. Let's also assume that

$$\sqrt{n}\frac{\partial F_n(\theta, \rho)}{\partial\theta} \to N(0, \Omega_y)$$

conditional on $(\nu_1, \ldots, \nu_T)$. Unlike in the case of the time series sample, Ω_y generally does depend on $(\nu_1, \ldots, \nu_T)$ through the parameter θ.

We note that $\partial F(\theta, \rho)/\partial\theta$ is a function of $(\nu_1, \ldots, \nu_T)$. If there is overlap between $(1, \ldots, T)$ and $(\tau_0 + 1, \ldots, \tau_0 + \tau)$, we need to worry about the asymptotic distribution of $\frac{1}{\sqrt{\tau}} \sum_{s=\tau_0+1}^{\tau_0+\tau} \varphi_s$ conditional on $(\nu_1, \ldots, \nu_T)$. However, because in this example the only connection between y and φ is assumed to be through θ and because T is assumed fixed, the two terms $\sqrt{n}\partial F_n(\theta, \rho)/\partial\theta$ and $\tau^{-1/2} \sum_{t=\tau_0+1}^{\tau_0+\tau} \varphi_t$ are expected to be asymptotically independent in the trend stationary case and when Ω_ν does not depend on $(\nu_1, \ldots, \nu_T)$. Even in this simple setting, independence between the two samples does not hold, and asymptotic conditional or unconditional independence as well as joint convergence with $\mathcal{C}$-measurable random variables needs to be established formally. This is achieved by establishing $\mathcal{C}$-stable convergence in Section 4.2.

It follows that

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) \rightarrow N\left(0, A^{-1}\Omega_y A^{-1} + \kappa A^{-1} B \Omega_\nu B' A^{-1}\right), \tag{9}$$

where $\kappa \equiv \lim n/\tau$. This means that a practitioner would use the square root of

$$\frac{1}{n}\left(A^{-1}\Omega_y A^{-1} + \frac{n}{\tau} A^{-1} B \Omega_\nu B' A^{-1}\right) = \frac{1}{n} A^{-1}\Omega_y A^{-1} + \frac{1}{\tau} A^{-1} B \Omega_\nu B' A^{-1}$$

as the standard error. This result looks similar to Murphy and Topel's (1985) formula, except that we need to make an adjustment to the second component to address the differences in sample sizes.

The asymptotic variance formula is such that the noise of the time series estimator $\widetilde{\rho}$ can make quite a difference if κ is large, i.e., if the time series size τ is small relative to the cross section size n. Obviously. this calls for long time series for accurate estimation of even the micro parameter β. We also note that time series estimation has no impact on micro estimation if $B = 0$. This confirms the intuition that if ρ does not appear as part of the micro moment f, which is the case in Heckman and Sedlacek (1985), and Heckman, Lochner, and Taber (1998), cross section estimation can be considered separate from time series estimation.

In the more general setting of Section 4, Ω_ν may depend on $(\nu_1, \ldots, \nu_T)$. In this case the limiting distribution of the time series component is mixed Gaussian and dependent upon the limiting distribution of the cross-sectional component. This dependence does not vanish asymptotically even in stationary settings. As we show in Section 4 standard inference based on asymptotically pivotal statistics is available even though the limiting distribution of $\widehat{\theta}$ is no longer a sum of two independent components.

3.2 Unit Root Problems

When the simple trend stationary paradigm does not apply, the limiting distribution of our estimators may be more complicated. A general treatment is beyond the scope of this paper and likely requires a case by case analysis. In this subsection we consider a simple unit root model where initial conditions can be neglected. We use it to exemplify additional inferential difficulties that arise even in this relatively simple setting. In Section 4.3 we consider a slightly more complex version of the unit root model where initial conditions cannot be ignored. We show that more complicated dependencies between the asymptotic distributions of the cross-section and time series samples manifest. The result is a cautionary tale of the difficulties that may present themselves when nonstationary time series data are combined with cross-sections. We leave the development of inferential methods for this case to future work.

We again consider the model in the previous section, except with the twist that (i) ρ is the AR(1) coefficient in the time series regression of z_t on z_{t-1} with independent error; and (ii) ρ is at (or near) unity. In the same way that led to (8), we obtain

$$\sqrt{n}\left(\widehat{\theta}-\theta\right) \approx -A^{-1}\sqrt{n}\frac{\partial F_n\left(\theta,\rho\right)}{\partial\theta} - A^{-1}B\frac{\sqrt{n}}{\tau}\tau\left(\widetilde{\rho}-\rho\right)$$

For simplicity, again assume that the two terms on the right are asymptotically independent. The first term converges in distribution to a normal distribution $N\left(0, A^{-1}\Omega_y A^{-1}\right)$, but with $\rho = 1$ and i.i.d. AR(1) errors the second term converges to

$$\xi A^{-1}B\frac{W\left(1\right)^2 - 1}{2\int_0^1 W\left(r\right)^2 dr},$$

where $\xi = \lim \sqrt{n}/\tau$ and $W\left(\cdot\right)$ is the standard Wiener process, in contrast to the result in (9) when ρ is away from unity. The result is formalized in Section 4.3.

The fact that the limiting distribution of $\hat{\theta}$ is no longer Gaussian complicates inference. This discontinuity is mathematically similar to Campbell and Yogo's (2006) observation, which leads to a question of how uniform inference could be conducted. In principle, the problem here can be analyzed by modifying the proposal in Phillips (2014, Section 4.3). First, construct the $1-\alpha_1$ confidence interval for ρ using Mikusheva (2007). Call it $[\rho_L, \rho_U]$. Second, compute $\widehat{\theta}\left(\rho\right) \equiv \operatorname{argmax}_\theta F_n\left(\theta,\rho\right)$ for $\rho \in [\rho_L, \rho_U]$. Assuming that ρ is fixed, characterize the asymptotic variance

$\Sigma(\rho)$, say, of $\sqrt{n}\left(\widehat{\theta}(\rho)-\theta(\rho)\right)$, which is asymptotically normal in general. Third, construct the $1-\alpha_2$ confidence region, say $CI(\alpha_2;\rho)$, using asymptotic normality and $\Sigma(\rho)$. Our confidence interval for θ_1 is then given by $\bigcup_{\rho\in[\rho_L,\rho_U]} CI(\alpha_2;\rho)$. By Bonferroni, its asymptotic coverage rate is expected to be at least $1-\alpha_1-\alpha_2$.

4 Joint Panel-Time Series Limit Theory

In this section we first establish a generic joint limiting result for a combined panel-time series process and then specialize it to the limiting distributions of parameter estimates under stationarity and non-stationarity. The process we analyze consists of a triangular array of panel data $\psi^y_{n,it}$ observed for $i=1,...,n$ and $t=1,\ldots,T$ where $n\to\infty$ while T is fixed and $t=1$ is an arbitrary normalization of time at the beginning of the cross-sectional sample. It also consists of a separate triangular array of time series $\psi^\nu_{\tau,t}$ for $t=\tau_0+1,\ldots,\tau_0+\tau$ where τ_0 is fixed with $-\infty<-K\le\tau_0\le K<\infty$ for some bounded K and $\tau\to\infty$. Typically, $\psi^y_{n,it}$and $\psi^\nu_{\tau,t}$ are the scores of a cross-section and time series criterion function based on observed data y_{it} and z_t. We assume that $T\le\tau_0+\tau$. Throughout we assume that $\left(\psi^y_{n,it},\psi^\nu_{\tau,t}\right)$ is a martingale difference sequence relative to a filtration to be specified below. We derive the joint limiting distribution and a related functional central limit theorem for $\frac{1}{\sqrt{n}}\sum_{t=1}^T\sum_{i=1}^n\psi^y_{n,it}$ and $\frac{1}{\sqrt{\tau}}\sum_{t=\tau_0+1}^{\tau_0+\tau}\psi^\nu_{\tau,t}$.

We now construct the triangular array of filtrations similarly to Kuersteiner and Prucha (2013). We use the binary operator $\vee$ to denote the σ-field generated by the union of two σ-fields. Setting $\mathcal{C}=\sigma(\nu_1,...,\nu_T)$ we define

$$\begin{aligned}
\mathcal{G}_{\tau n,0}&=\mathcal{C} \\
\mathcal{G}_{\tau n,i}&=\sigma\left(z_{\min(1,\tau_0)},\left\{y_{j,\min(1,\tau_0)}\right\}_{j=1}^i\right)\vee\mathcal{C} \\
&\vdots \\
\mathcal{G}_{\tau n,n+i}&=\sigma\left(\left\{y_{j,\min(1,\tau_0)}\right\}_{j=1}^n,\left\{z_{\min(1,\tau_0)+1},z_{\min(1,\tau_0)}\right\},\left\{y_{j,\min(1,\tau_0)+1}\right\}_{j=1}^i\right)\vee\mathcal{C} \\
&\vdots \\
\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n+i}&=\sigma\left(\left\{y_{j,t-1},y_{j,t-2},\ldots,y_{j,\min(1,\tau_0)}\right\}_{j=1}^n,\left\{z_t,z_{t-1},\ldots,z_{\min(1,\tau_0)}\right\},\left\{y_{j,t}\right\}_{j=1}^i\right)\vee\mathcal{C}
\end{aligned} \tag{10}$$

We use the convention that $\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n}=\mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n+n}$. This implies that z_t and y_{1t} are added simultaneously to the filtration $\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n+1}$. Also note that $\mathcal{G}_{\tau n,i}$ predates the time series sample by at least one period, i.e. corresponds to the 'time zero' sigma field. To simplify notation define the function $q_n(t,i) = (t - \min(1,\tau_0))n + i$ that maps the two-dimensional index (t,i) into the integers and note that for $q = q_n(t,i)$ it follows that $q \in \{0,\ldots,\max(T,\tau)n\}$. The filtrations $\mathcal{G}_{\tau n,q}$ are increasing in the sense that $\mathcal{G}_{\tau n,q} \subset \mathcal{G}_{\tau n,q+1}$ for all q, τ and n. We note that $E\left[\psi^{\nu}_{\tau,t}\middle|\mathcal{G}_{\tau n,q_n(t-1,i)}\right] = 0$ for all i is not guaranteed because we condition not only on $z_{t-1}, z_{t-2}...$ but also on $\nu_1,...,\nu_T$, where the latter may have non-trivial overlap with the former.

The central limit theorem we develop needs to establish joint convergence for terms involving both $\psi^{y}_{n,it}$ and $\psi^{\nu}_{\tau,t}$ with both the time series and the cross-sectional dimension becoming large simultaneously. Let $[a]$ be the largest integer less than or equal a. Joint convergence is achieved by stacking both moment vectors into a single sum that extends over both t and i. Let $r \in [0,1]$ and define

$$\tilde{\psi}^{\nu}_{it}(r) = \frac{\psi^{\nu}_{\tau,t}}{\sqrt{\tau}} 1\{\tau_0 + 1 \leq t \leq \tau_0 + [\tau r]\} 1\{i = 1\}, \tag{11}$$

which depends on r in a non-trivial way. This dependence will be of particular interest when we specialize our models to the near unit root case. For the cross-sectional data define

$$\tilde{\psi}^{y}_{it} = \frac{\psi^{y}_{n,it}}{\sqrt{n}} \tag{12}$$

where $\tilde{\psi}^{y}_{it}$ is constant as a function of $r \in [0,1]$. In turn, this implies that functional convergence of the component (12) is the same as the finite dimensional limit. It also means that the limiting process is degenerate (i.e. constant) when viewed as a function of r. However, this does not matter in our applications as we are only interested in the sum

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{T} \sum_{i=1}^{n} \psi^{y}_{n,it} = \sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)} \sum_{i=1}^{n} \tilde{\psi}^{y}_{it} \equiv X^{y}_{n\tau}.$$

Define the stacked vector $\tilde{\psi}_{it}(r) = \left(\tilde{\psi}^{y\prime}_{it}, \tilde{\psi}^{\nu}_{it}(r)'\right)' \in \mathbb{R}^{k_\phi}$ and consider the stochastic process

$$X_{n\tau}(r) = \sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)} \sum_{i=1}^{n} \tilde{\psi}_{it}(r), \quad X_{n\tau}(0) = (X^{y\prime}_{n\tau}, 0)'. \tag{13}$$

We derive a functional central limit theorem which establishes joint convergence between the panel and time series portions of the process $X_{n\tau}(r)$. The result is useful in analyzing both trend stationary and unit root settings. In the latter, we specialize the model to a linear time series setting. The functional CLT is then used to establish proper joint convergence between stochastic integrals and the cross-sectional component of our model.

For the stationary case we are mostly interested in $X_{n\tau}(1)$ where in particular

$$\frac{1}{\sqrt{\tau}}\sum_{t=\tau_0+1}^{\tau_0+\tau}\psi_{\tau,t}^{\nu}=\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\tilde{\psi}_{it}^{\nu}(1).$$

The limiting distribution of $X_{n\tau}(1)$ is a simple corollary of the functional CLT for $X_{n\tau}(r)$. We note that our treatment differs from Phillips and Moon (1999), who develop functional CLT's for the time series dimension of the panel data set. In our case, since T is fixed and finite, a similar treatment is not applicable.

We introduce the following general regularity conditions. In later sections these conditions will be specialized to the particular models considered there.

Condition 1 *Assume that*

i) $\psi_{n,it}^{y}$ *is measurable with respect to* $\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n+i}$.

ii) $\psi_{\tau,t}^{\nu}$ *is measurable with respect to* $\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n+i}$ *for all* $i=1,...,n$.

iii) for some $\delta>0$ *and* $C<\infty$, $\sup_{it}E\left[\left\|\psi_{n,it}^{y}\right\|^{2+\delta}\right]\leq C$ *for all* $n\geq 1$.

iv) for some $\delta>0$ *and* $C<\infty$, $\sup_{t}E\left[\left\|\psi_{\tau,t}^{\nu}\right\|^{2+\delta}\right]\leq C$ *for all* $\tau\geq 1$.

v) $E\left[\psi_{n,it}^{y}\middle|\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n+i-1}\right]=0$.

vi) $E\left[\psi_{\tau,t}^{\nu}\middle|\mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n+i}\right]=0$ *for* $t>T$ *and all* $i=1,...,n$.

Remark 1 *Conditions 1(i), (iii) and (v) can be justified in a variety of ways. One is the subordinated process theory employed in Andrews (2005) which arises when* y_{it} *are random draws from a population of outcomes* y. *A sufficient condition for Conditions 1(v) to hold is that* $E\left[\psi(y|\theta,\rho,\nu_t)|\mathcal{C}\right]=0$ *holds in the population. This would be the case, for example, if* ψ *were the correctly specified score for the population distribution. See Andrews (2005, pp. 1573-1574).*

Condition 2 *Assume that:*

i) for any $s,r\in[0,1]$ *with* $r>s$,

$$\frac{1}{\tau}\sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]}\psi_{\tau,t}^{\nu}\psi_{\tau,t}^{\nu\prime}\xrightarrow{p}\Omega_{\nu}(r)-\Omega_{\nu}(s)\ \ \textit{as}\ \tau\to\infty$$

where $\Omega_\nu(r) - \Omega_\nu(s)$ is positive definite a.s. and measurable with respect to $\sigma(\nu_1, ..., \nu_T)$ for all $r \in (s, 1]$. Normalize $\Omega_\nu(0) = 0$.

ii) The elements of $\Omega_\nu(r)$ are bounded continuously differentiable functions of $r > s \in [0, 1]$. The derivatives $\dot{\Omega}_\nu(r) = \partial\Omega_\nu(r) / \partial r$ are positive definite almost surely.

iii) There is a fixed constant $M < \infty$ such that $\sup_{\|\lambda_\nu\|=1, \lambda_\nu \in \mathbb{R}^{k_\rho}} \sup_t \lambda_\nu' \partial\Omega_\nu(t) / \partial t \lambda_\nu \leq M$ a.s.

Condition 2 is weaker than the conditions of Billingsley's (1968) functional CLT for strictly stationary martingale difference sequences (mds). We do not assume that $E\left[\psi^\nu_{\tau,t}\psi^{\nu\prime}_t\right]$ is constant. Brown (1971) allows for time varying variances, but uses stopping times to achieve a standard Brownian limit. Even more general treatments with random stopping times are possible - see Gaenssler and Haeussler (1979). On the other hand, here convergence to a Gaussian process (not a standard Wiener process) with the same methodology (i.e. establishing convergence of finite dimensional distributions and tightness) as in Billingsley, but without assuming homoskedasticity is pursued. Heteroskedastic errors are explicitly used in Section 4.3 where $\psi^\nu_{\tau,t} = \exp((t - s)\gamma/\tau)\eta_s$. Even if η_s is iid$(0, \sigma^2)$ it follows that $\psi^\nu_{\tau,t}$ is a heteroskedastic triangular array that depends on τ. It can be shown that the variance kernel $\Omega_\nu(r)$ is $\Omega_\nu(r) = \sigma^2(1 - \exp(-2r\gamma)) / 2\gamma$ in this case. See equation (55).

Condition 3 *Assume that*

$$\frac{1}{n}\sum_{i=1}^{n} \psi^y_{n,it}\psi^{y\prime}_{n,it} \xrightarrow{p} \Omega_{ty}$$

where Ω_{ty} is positive definite a.s. and measurable with respect to $\sigma(\nu_1, ..., \nu_T)$.

Condition 2 holds under a variety of conditions that imply some form of weak dependence of the process $\psi^\nu_{\tau,t}$. These include, in addition to Condition 1(ii) and (iv), mixing or near epoch dependence assumptions on the temporal dependence properties of the process $\psi^\nu_{\tau,t}$. Assumption 3 holds under appropriate moment bounds and random sampling in the cross-section even if the underlying population distribution is not independent (see Andrews, 2005, for a detailed treatment).

4.1 Stable Functional CLT

This section details the probabilistic setting we use to accommodate the results that Jacod and Shiryaev (2002) (shorthand notation JS) develop for general Polish spaces. Let $(\Omega', \mathcal{F}', P')$ be a probability space with increasing filtrations $\mathcal{F}_t^n \subset \mathcal{F}$ and $\mathcal{F}_t^n \subset \mathcal{F}_{t+1}^n$ for any $t = 1, ..., k_n$ and an increasing sequence $k_n \to \infty$ as $n \to \infty$. Let $D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]$ be the space of functions $[0,1] \to \mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}$ that are right continuous and have left limits (see Billingsley (1968, p.109)). Let $\mathcal{C}$ be a sub-sigma field of $\mathcal{F}'$. Let $(\zeta, Z^n(\omega, t)) : \Omega' \times [0,1] \to \mathbb{R} \times \mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}$ be random variables or random elements in $\mathbb{R}$ and $D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]$, respectively defined on the common probability space $(\Omega', \mathcal{F}', P')$ and assume that ζ is bounded and measurable with respect to $\mathcal{C}$. Equip $D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]$ with the Skorohod topology, see JS (p.328, Theorem 1.14). By Billingsley (1968), Theorem 15.5 the uniform metric can be used to establish tightness for certain processes that are continuous in the limit.

We use the results of JS to define a precise notion of stable convergence on $D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]$. JS (p.512, Definition 5.28) define stable convergence for sequences Z^n defined on a Polish space. We adopt their definition to our setting, noting that by JS (p.328, Theorem 1.14), $D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]$ equipped with the Skorohod topology is a Polish space. Also following their Definition VI1.1 and Theorem VI1.14 we define the σ-field generated by all coordinate projections as $\mathcal{D}_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}$.

Definition 1 *The sequence Z^n converges $\mathcal{C}$-stably if for all bounded ζ measurable with respect to $\mathcal{C}$ and for all bounded continuous functions f defined on $D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]$ there exists a probability measure μ on $(\Omega' \times D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1], \mathcal{C} \times \mathcal{D}_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}})$ such that*

$$E[\zeta f(Z^n)] \to \int_{\Omega' \times D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]} \zeta(\omega') f(x) \mu(d\omega', dx).$$

As in JS, let $Q(\omega', dx)$ be a distribution conditional on $\mathcal{C}$ such that $\mu(d\omega', dx) = P'(d\omega') Q(\omega', dx)$ and let $Q_n(\omega', dx)$ be a version of the conditional (on $\mathcal{C}$) distribution of Z^n. Then we can define the joint probability space $(\Omega, \mathcal{F}, P)$ with $\Omega = \Omega' \times D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]$, $\mathcal{F} = \mathcal{F}' \times \mathcal{D}_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}$ and $P = P(d\omega, dx) = P'(d\omega') Q(\omega', dx)$. Let $Z(\omega', x) = x$ be the canonical element on $D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]$. It follows that $\int \zeta(\omega') f(x) \mu(d\omega', dx) = E[\zeta Q f]$. We say that Z^n converges $\mathcal{C}$-stably to Z if for all bounded, $\mathcal{C}$-measurable ζ,

$$E[\zeta f(Z^n)] \to E[\zeta Q f]. \tag{14}$$

More specifically, if $W(r)$ is standard Brownian motion, we say that $Z^n \Rightarrow W(r)$ $\mathcal{C}$-stably where the notation means that (14) holds when Q is Wiener measure (for a definition and existence proof see Billingsley (1968, ch.9)). By JS Proposition VIII5.33 it follows that Z^n converges $\mathcal{C}$-stably iff Z^n is tight and for all $A \in \mathcal{C}$, $E[1_A f(Z^n)]$ converges.

The concept of stable convergence was introduced by Renyi (1963) and has found wide application in probability and statistics. Most relevant to the discussion here are the stable central limit theorem of Hall and Heyde (1980) and Kuersteiner and Prucha (2013) who extend the result in Hall and Heyde (1980) to panel data with fixed T. Dedecker and Merlevede (2002) established a related stable functional CLT for strictly stationary martingale differences.

Following Billingsley (1968, p. 120) let $\pi_{r_1,\dots,r_k} Z^n = \left(Z^n_{r_1}, \dots, Z^n_{r_k}\right)$ be the coordinate projections of Z^n. By JS VIII5.36 and by the proof of Theorems 5.7 and 5.14 on p. 509 of JS (see also, Rootzen (1983), Feigin (1985), Dedecker and Merlevede (2002, p. 1057)), $\mathcal{C}$-stable convergence for Z^n to Z follows if $E\left[\zeta f\left(Z^n_{r_1}, \dots, Z^n_{r_k}\right)\right] \to E\left[\zeta f\left(Z_{r_1}, \dots, Z_{r_n}\right)\right]$ and Z^n is tight under the measure P. We note that the first condition is equivalent to stable convergence of the finite dimensional vector of random variables $Z^n_{r_1}, \dots, Z^n_{r_k}$ defined on $\mathbb{R}^k$ and is established with a multivariate stable central limit theorem.

Theorem 1 *Assume that Conditions 1, 2 and 3 hold. Then it follows that for $\tilde{\psi}_{it}$ defined in (13), and as $\tau, n \to \infty$ and T fixed,*

$$X_{n\tau}(r) \Rightarrow \begin{bmatrix} B_y(1) \\ B_\nu(r) \end{bmatrix} \quad (\mathcal{C}\text{-stably})$$

where $B_y(r) = \Omega_y{}^{1/2} W_y(r)$, $B_\nu(r) = \int_0^r \dot{\Omega}_\nu(s)^{1/2} dW_\nu(s)$ and $\Omega(r) = \operatorname{diag}(\Omega_y, \Omega_\nu(r))$ is $\mathcal{C}$-measurable, $\dot{\Omega}_\nu(s) = \partial\Omega_\nu(s)/\partial s$ and $(W_y(r), W_\nu(r))$ is a vector of standard k_ϕ-dimensional Brownian processes independent of Ω.

Proof. In Appendix A. ■

Remark 2 *Note that the component $W_y(1)$ of $W(r)$ does not depend on r. Thus, $W_y(1)$ is simply a vector of standard Gaussian random variables, independent both of $W_\nu(r)$ and any random variable measurable w.r.t $\mathcal{C}$.*

The limiting random variables $B_y(r)$ and $B_\nu(r)$ both depend on $\mathcal{C}$ and are thus mutually dependent. The representation $B_y(1) = \Omega_y{}^{1/2} W_y(1)$, where a stable limit is represented as the product of an independent Gaussian random variable and a scale factor that depends on $\mathcal{C}$, is common in the literature on stable convergence. Results similar to the one for $B_\nu(r)$ were obtained by Phillips (1987, 1988) for cases where $\dot{\Omega}_\nu(s)$ is non-stochastic and has an explicitly functional form, notably for near unit root processes and when convergence is marginal rather than stable. Rootzen (1983) establishes stable convergence but gives a representation of the limiting process in terms of standard Brownian motion obtained by a stopping time transformation. The representation of $B_\nu(r)$ in terms of a stochastic integral over the random scale process $\dot{\Omega}_\nu(s)$ seems to be new. It is obtained by utilizing a technique mentioned in Rootzen (1983, p. 10) but not utilized there, namely establishing finite dimensional convergence using a stable martingale CLT. This technique combined with a tightness argument establishes the characteristic function of the limiting process. The representation for $B_\nu(r)$ is then obtained by utilizing results for characteristic functions of affine diffusions in Duffie, Pan and Singletion (2000). Rootzen (1983, p.13) similarly utilizes characteristic functions to identify the limiting distribution in the case of standard Brownian motion, a much simpler scenario than ours. Finally, the results of Dedecker and Merlevede (2002) differ from ours in that they only consider asymptotically homoskedastic and strictly stationary processes. In our case, heteroskedasticity is explicitly allowed because of $\dot{\Omega}_\nu(s)$. An important special case of Theorem 1 is the near unit root model discussed in more detail in Section 4.3.

More importantly, our results innovate over the literature by establishing joint convergence between cross-sectional and time series averages that are generally not independent and whose limiting distributions are not independent. This result is obtained by a novel construction that embeds both data sets in a random field. A careful construction of information filtrations $\mathcal{G}_{\tau n,n+i}$ allows to map the field into a martingale array. Similar techniques were used in Kuersteiner and Prucha (2013) for panels with fixed T. In this paper we extend their approach to handle an additional and distinct time series data-set and by allowing for both n and τ to tend to infinity jointly. In addition to the more complicated data-structure we extend Kuersteiner and Prucha (2013) by considering functional central limit theorems.

The following corollary is useful for possibly non-linear but trend stationary models.

Corollary 1 *Assume that Conditions 1, 2 and 3 hold. Then it follows that for $\tilde{\psi}_{it}$ defined in (13), and as $\tau, n \to \infty$ and T fixed,*

$$X_{n\tau}(1) \xrightarrow{d} B := \Omega^{1/2} W \ (\mathcal{C}\text{-stably})$$

where $\Omega = \operatorname{diag}(\Omega_y, \Omega_\nu(1))$ is $\mathcal{C}$-measurable and $W = (W_y(1), W_\nu(1))$ is a vector of standard d-dimensional Gaussian random variables independent of Ω. The variables $\Omega_y, \Omega_\nu(.), W_y(.)$ and $W_\nu(.)$ are as defined in Theorem 1.

Proof. In Appendix A. ■

The result of Corollary 1 is equivalent to the statement that $X_{n\tau}(1) \xrightarrow{d} N(0, \Omega)$ conditional on positive probability events in $\mathcal{C}$. As noted earlier, no simplification of the technical arguments are possible by conditioning on $\mathcal{C}$ except in the trivial case where Ω is a fixed constant. Eagleson (1975, Corollary 3), see also Hall and Heyde (1980, p. 59), establishes a simpler result where $X_{n\tau}(1) \xrightarrow{d} B$ weakly but not ($\mathcal{C}$-stably). Such results could in principle be obtained here as well, but they would not be useful for the analysis in Sections 4.2 and 4.3 because the limiting distributions of our estimators not only depend on B but also on other $\mathcal{C}$-measurable scaling matrices. Since the continuous mapping theorem requires joint convergence, a weak limit for B alone is not sufficient to establish the results we obtain below.

Theorem 1 establishes what Phillips and Moon (1999) call diagonal convergence, a special form of joint convergence. To see that sequential convergence where first n or τ go to infinity, followed by the other index, is generally not useful in our set up, consider the following example. Assume that $d = k_\phi$ is the dimension of the vector $\tilde{\psi}_{it}$. This would hold for just identified moment estimators and likelihood based procedures. Consider the double indexed process

$$X_{n\tau}(1) = \sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)} \sum_{i=1}^{n} \tilde{\psi}_{it}(1). \tag{15}$$

For each τ fixed, convergence in distribution of $X_{n\tau}$ as $n \to \infty$ follows from the central limit theorem in Kuersteiner and Prucha (2013). Let X_τ denote the "large n, fixed τ" limit. For each n fixed, convergence in distribution of $X_{n\tau}$ as $\tau \to \infty$ follows from a standard martingale central limit theorem for Markov processes. Let X_n be the "large τ, fixed n" limit. It is worth pointing out that the distributions of both X_n and X_τ are unknown because the limits are trivial in one direction.

For example, when τ is fixed and n tends to infinity, the component $\tau^{-1/2}\sum_{t=\tau_0+1}^{\tau_0+\tau}\psi_{\tau,t}^{\nu}$ trivially converges in distribution (it does not change with n) but the distribution of $\tau^{-1/2}\sum_{t=\tau_0+1}^{\tau_0+\tau}\psi_{\tau,t}^{\nu}$ is generally unknown. More importantly, application of a conventional CLT for the cross-section alone will fail to account for the dependence between the time series and cross-sectional components. Sequential convergence arguments thus are not recommended even as heuristic justifications of limiting distributions in our setting.

4.2 Trend Stationary Models

Let $\theta = (\beta, \nu_1, ..., \nu_T)$ and define the shorthand notation $f_{it}(\theta,\rho) = f(y_{it}|\theta,\rho)$, $g_t(\beta,\rho) = g(\nu_t|\nu_{t-1},\beta,\rho)$, $f_{\theta,it}(\theta,\rho) = \partial f_{it}(\theta,\rho)/\partial\theta$ and $g_{\rho,t}(\beta,\rho) = \partial g_t(\beta,\rho)/\partial\rho$. Also let $f_{it} = f_{it}(\theta_0,\rho_0)$, $f_{\theta,it} = f_{\theta,it}(\theta_0,\rho_0)$, $g_t = g_t(\beta_0,\rho_0)$ and $g_{\rho,t} = g_{\rho,t}(\beta_0,\rho_0)$. Depending on whether the estimator under consideration is maximum likelihood or moment based we assume that either $(f_{\theta,it}, g_{\rho,t})$ or (f_{it}, g_t) satisfy the same Assumptions as $\left(\psi_{it}^{y}, \psi_{\tau,t}^{\nu}\right)$ in Condition 1. We recall that $\nu_t(\beta,\rho)$ is a function of (z_t,β,ρ), where z_t are observable macro variables. For the CLT, the process $\nu_t = \nu_t(\rho_0,\beta_0)$ is evaluated at the true parameter values and treated as observed. In applications, ν_t will be replaced by an estimate which potentially affects the limiting distribution of ρ. This dependence is analyzed in a step separate from the CLT.

The next step is to use Corollary 1 to derive the joint limiting distribution of estimators for $\phi = (\theta', \rho')'$. Define $s_{ML}^{\nu}(\beta,\rho) = \tau^{-1/2}\sum_{t=\tau_0+1}^{\tau_0+\tau}\partial g(\nu_t(\beta,\rho)|\nu_{t-1}(\beta,\rho),\beta,\rho)/\partial\rho$ and $s_{ML}^{y}(\theta,\rho) = n^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{n}\partial f(y_{it}|\theta,\rho)/\partial\theta$ for maximum likelihood, and

$$s_M^{\nu}(\beta,\rho) = -\left(\partial k_\tau(\beta,\rho)/\partial\rho\right)' W_\tau^\tau \tau^{-1/2}\sum_{t=\tau_0+1}^{\tau_0+\tau} g(\nu_t(\beta,\rho)|\nu_{t-1}(\beta,\rho),\beta,\rho)$$

and $s_M^{y}(\theta,\rho) = -\left(\partial h_n(\theta,\rho)/\partial\theta\right)' W_n^C n^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{n} f(y_{it}|\theta,\rho)$ for moment based estimators. We use $s^{\nu}(\beta,\rho)$ and $s^{y}(\theta,\rho)$ generically for arguments that apply to both maximum likelihood and moment based estimators. The estimator $\hat{\phi}$ jointly satisfies the moment restrictions using time series data

$$s^{\nu}\left(\hat{\beta},\hat{\rho}\right) = 0. \tag{16}$$

and cross-sectional data

$$s^{y}\left(\hat{\theta},\hat{\rho}\right) = 0. \tag{17}$$

Defining $s(\phi) = \left(s^y(\phi)', s^\nu(\phi)'\right)'$ the estimator $\hat{\phi}$ satisfies $s\left(\hat{\phi}\right) = 0$. A first order Taylor series expansion around ϕ_0 is used to obtain the limiting distribution for $\hat{\phi}$. We impose the following additional assumption.

Condition 4 *Let $\phi = (\theta', \rho')' \in \mathbb{R}^{k_\phi}$, $\theta \in \mathbb{R}^{k_\theta}$, and $\rho \in \mathbb{R}^{k_\rho}$. Define $D_{n\tau} = \operatorname{diag}\left(n^{-1/2}I_y, \tau^{-1/2}I_\nu\right)$, where I_y is an identity matrix of dimension k_θ and I_ν is an identity matrix of dimension k_ρ. Let $W^C = \operatorname{plim}_n W_n^C$ and $W^\tau = \operatorname{plim}_\tau W_\tau^\tau$ and assume the limits to be positive definite and $\mathcal{C}$-measurable. Define $h(\theta, \rho) = \operatorname{plim}_n h_n(\beta, \nu_t, \rho)$ and $k(\beta, \rho) = \operatorname{plim}_\tau k_\tau(\beta, \rho)$. Assume that for some $\varepsilon > 0$,*

i) $\sup_{\phi:\|\phi-\phi_0\|\le\varepsilon} \left\|(\partial k_\tau(\beta, \rho)/\partial\rho)' W_\tau^\tau - \partial k(\beta, \rho)'/\partial\rho W^\tau\right\| = o_p(1)$,

ii) $\sup_{\phi:\|\phi-\phi_0\|\le\varepsilon} \left\|(\partial h_n(\theta, \rho)/\partial\theta)' W_n^C - (\partial h(\theta, \rho)/\partial\theta)' W^C\right\| = o_p(1)$,

iii)s $\sup_{\phi:\|\phi-\phi_0\|\le\varepsilon} \left\|\frac{\partial s(\phi)}{\partial \phi'} D_{n\tau} - A(\phi)\right\| = o_p(1)$ *where $A(\phi)$ is $\mathcal{C}$-measurable and $A = A(\phi_0)$ is full rank almost surely. Let $\kappa = \lim n/\tau$,*

$$A = \begin{bmatrix} A_{y,\theta} & \sqrt{\kappa}A_{y,\rho} \\ \frac{1}{\sqrt{\kappa}}A_{\nu,\theta} & A_{\nu,\rho} \end{bmatrix}$$

with $A_{y,\theta} = \operatorname{plim} n^{-1}\partial s^y(\phi_0)/\partial\theta'$, $A_{y,\rho} = \operatorname{plim} n^{-1}\partial s^y(\phi_0)/\partial\rho'$, $A_{\nu,\theta} = \operatorname{plim} \tau^{-1}\partial s^\nu(\phi_0)/\partial\theta'$ and $A_{\nu,\rho} = \operatorname{plim} \tau^{-1}\partial s^\nu(\phi_0)/\partial\rho'$.

Condition 5 *For maximum likelihood criteria the following holds:*

i) for any $s, r \in [0,1]$ with $r > s$, $\frac{1}{\tau}\sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]} g_{\rho,t}g_{\rho,t}' \xrightarrow{p} \Omega_\nu(r) - \Omega_\nu(s)$ as $\tau \to \infty$ and where $\Omega_\nu(r)$ satisfies the same regularity conditions as in Condition 2(ii).

ii) $\frac{1}{n}\sum_{i=1}^{n} f_{\theta,it}f_{\theta,it}' \xrightarrow{p} \Omega_{ty}$ for all $t \in [1, ..., T]$ and where Ω_{ty} is positive definite a.s. and measurable with respect to $\sigma(\nu_1, ..., \nu_T)$. Let $\Omega_y = \sum_{t=1}^{T} \Omega_{ty}$.

Condition 6 *For moment based criteria the following holds:*

i) for any $s, r \in [0,1]$ with $r > s$, $\frac{1}{\tau}\sum_{t,q=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]} g_t g_q' \xrightarrow{p} \Omega_g(r) - \Omega_g(s)$ as $\tau \to \infty$ and where $\Omega_g(r)$ satisfies the same regularity conditions as in Condition 2(ii).

ii) $\frac{1}{n}\sum_{i=1}^{n} f_{it}f_{ir}' \xrightarrow{p} \Omega_{t,rf}$ for all $t, s \in [1, ..., T]$ Let $\Omega_f = \sum_{t,r=1}^{T} \Omega_{t,sf}$. Assume that Ω_f is positive definite a.s. and measurable with respect to $\sigma(\nu_1, ..., \nu_T)$.

Condition 6 accounts for the possibility of misspecification of the model. In that case, the martingale difference property of the moment conditions may not hold, necessitating the use of robust standard errors through long run variances.

The following result establishes the joint limiting distribution of $\hat{\phi}$.

Theorem 2 *Assume that Conditions 1, 4, and either 5 with* $\left(\psi_{it}^{y}, \psi_{\tau,t}^{\nu}\right) = \left(f_{\theta,it}, g_{\rho,t}\right)$ *in the case of likelihood based estimators or 6 with* $\left(\psi_{it}^{y}, \psi_{\tau,t}^{\nu}\right) = \left(f_{it}, g_{t}\right)$ *in the case of moment based estimators hold. Assume that* $\hat{\phi} - \phi_0 = o_p(1)$ *and that (16) and (17) hold. Then,*

$$D_{n\tau}^{-1}\left(\hat{\phi} - \phi_0\right) \xrightarrow{d} -A^{-1}\Omega^{1/2}W \ \ (\mathcal{C}\text{-stably})$$

where A is full rank almost surely, $\mathcal{C}$-measurable and is defined in Condition 4. The distribution of $\Omega^{1/2}W$ is given in Corollary 1. In particular, $\Omega = \operatorname{diag}\left(\Omega_y, \Omega_\nu(1)\right)$. Then the criterion is maximum likelihood Ω_y and $\Omega_\nu(1)$ are given in Condition 5. When the criterion is moment based, $\Omega_y = \frac{\partial h(\theta_0,\rho_0)'}{\partial \theta} W^{C} \Omega_f W^{C\prime} \frac{\partial h(\theta_0,\rho_0)}{\partial \theta}$ *and* $\Omega_\nu(1) = \frac{\partial k(\beta_0,\rho_0)'}{\partial \rho} W^{\tau} \Omega_g(1) W^{\tau\prime} \frac{\partial k(\beta_0,\rho_0)}{\partial \rho}$ *with* Ω_f *and* $\Omega_g(1)$ *defined in Condition 6.*

Proof. In Appendix A. ■

Corollary 2 *Under the same conditions as in Theorem 2 it follows that*

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} -A^{y,\theta}\Omega_y^{1/2}W_y(1) - \sqrt{\kappa}A^{y,\rho}\Omega_\nu^{1/2}(1)\,W_\nu(1) \ \ (\mathcal{C}\text{-stably}). \tag{18}$$

where

$$\begin{aligned} A^{y,\theta} &= A_{y,\theta}^{-1} + A_{y,\theta}^{-1}A_{y,\rho}\left(A_{\nu,\rho} - A_{\nu,\theta}A_{y,\theta}^{-1}A_{y,\rho}\right)^{-1}A_{\nu,\theta}A_{y,\theta}^{-1} \\ A^{y,\rho} &= -A_{y,\theta}^{-1}A_{y,\rho}\left(A_{\nu,\rho} - A_{\nu,\theta}A_{y,\theta}^{-1}A_{y,\rho}\right)^{-1}. \end{aligned}$$

For

$$\Omega_\theta = A^{y,\theta}\Omega_y A^{y,\theta\prime} + \kappa A^{y,\rho}\Omega_\nu(1)\,A^{y,\rho\prime}$$

it follows that

$$\sqrt{n}\Omega_\theta^{-1/2}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N(0, I) \ \ (\mathcal{C}\text{-stably}). \tag{19}$$

Note that Ω_θ, the asymptotic variance of $\sqrt{n}\left(\hat{\theta}-\theta_0\right)$ conditional on $\mathcal{C}$, in general is a random variable, and the asymptotic distribution of $\hat{\theta}$ is mixed normal. However, as in Andrews (2005), the result in (19) can be used to construct an asymptotically pivotal test statistic. For a consistent estimator $\hat{\Omega}_\theta$ the statistic $\sqrt{n}\hat{\Omega}_\theta^{-1/2}\left(R\hat{\theta}-r\right)$ is asymptotically distribution free under the null hypothesis $R\theta - r = 0$ where R is a conforming matrix of dimension $q \times k_\theta$ and and r a $q \times 1$ vector.

4.3 Unit Root Time Series Models

In this section we consider the special case where ν_t follows an autoregressive process of the form $\nu_{t+1} = \rho\nu_t + \eta_t$. As in Hansen (1992), Phillips (1987, 1988, 2014) we allow for nearly integrated processes where $\rho = \exp\left(\gamma/\tau\right)$ is a scalar parameter localized to unity such that

$$\nu_{\tau,t+1} = \exp\left(\gamma/\tau\right)\nu_{\tau,t} + \eta_{t+1} \tag{20}$$

and the notation $\nu_{\tau,t}$ emphasizes that $\nu_{\tau,t}$ is a sequence of processes indexed by τ. We assume that

$$\tau^{-1/2}\nu_{\tau,\min(1,\tau_0)} = \nu_0 = V\left(0\right) \text{ a.s.}$$

where ν_0 is a potentially nondegenerate random variable. In other words, the initial condition for (20) is $\nu_{\tau,\min(1,\tau 0)} = \tau^{1/2}\nu_0$. We explicitly allow for the case where $\nu_0 = 0$, to model a situation where the initial condition can be ignored. This assumption is similar, although more parametric than, the specification considered in Kurtz and Protter (1991). We limit our analysis to the case of maximum likelihood criterion functions. Results for moment based estimators can be developed along the same lines as in Section 4.2 but for ease of exposition we omit the details. For the unit root version of our model we assume that ν_t is observed in the data and that the only parameter to be estimated from the time series data is ρ. Further assuming a Gaussian quasi-likelihood function we note that the score function now is

$$g_{\rho,t}\left(\beta,\rho\right) = \nu_{\tau,t-1}\left(\nu_{\tau,t} - \nu_{\tau,t-1}\rho\right). \tag{21}$$

The estimator $\hat{\rho}$ solving sample moment conditions based on (21) is the conventional OLS estimator given by

$$\hat{\rho} = \frac{\sum_{t=\tau_0+1}^{\tau}\nu_{\tau,t-1}\nu_{\tau,t}}{\sum_{t=\tau_0+1}^{\tau}\nu_{\tau,t-1}^2}.$$

We continue to use the definition for $f_{\theta,it}(\theta,\rho)$ in Section 4.2 but now consider the simplified case where $\theta_0 = (\beta, \nu_0)$. We note that in this section, ν_0 rather than $\nu_{\tau,\min(1,\tau 0)}$ is the common shock used in the cross-sectional model. The implicit scaling of $\nu_{\tau,\min(1,\tau_0)}$ by $\tau^{-1/2}$ is necessary in the cross-sectional specification to maintain a well defined model even as $\tau \to \infty$.

Consider the joint process $(V_{\tau n}(r), Y_{\tau n})$ where $V_{\tau n}(r) = \tau^{-1/2}\nu_{\tau[\tau r]}$, and

$$Y_{\tau n} = \sum_{t=1}^{T}\sum_{i=1}^{n}\frac{f_{\theta,it}}{\sqrt{n}}.$$

Note that

$$\int_0^r V_{\tau n} dW_{\tau n} = \tau^{-1}\sum_{t=\tau_0+1}^{\tau_0+[\tau r]}\nu_{\tau,t-1}\eta_t$$

with $W_{\tau n}(r) = \tau^{-1/2}\sum_{t=\tau_0+1}^{\tau_0+[\tau r]}\eta_t$. We define the limiting process for $V_{\tau n}(r)$ as

$$V_{\gamma,V(0)}(r) = e^{\gamma r}V(0) + \int_0^r \sigma e^{\gamma(r-s)}dW_\nu(s) \tag{22}$$

where W_ν is defined in Theorem 1. When $\nu_0 = 0$, Theorem 1 directly implies that $e^{-\gamma[r\tau]/\tau}V_{\tau n}(r) \Rightarrow \int_0^r \sigma e^{-s\gamma}dW_\nu(s)$ $\mathcal{C}$-stably noting that in this case $\Omega_\nu(s) = \sigma^2(1-\exp(-2s\gamma))2\gamma$ and $\dot{\Omega}_\nu(s)^{1/2} = \sigma e^{-s\gamma}$. The familiar result (cf. Phillips 1987) that $V_{\tau n}(r) \Rightarrow \int_0^r \sigma e^{\gamma(r-s)}dW_\nu(s)$ then is a consequence of the continuous mapping theorem. The case in (22) where ν_0 is a $\mathcal{C}$-measurable random variable now follows from $\mathcal{C}$-stable convergence of $V_{\tau n}(r)$. In this section we establish joint $\mathcal{C}$-stable convergence of the triple $\left(V_{\tau n}(r), Y_{\tau n}, \int_0^r V_{\tau n}dW_{\tau n}\right)$.

Let $\phi = (\theta', \rho)' \in \mathbb{R}^{k_\phi}$, $\theta \in \mathbb{R}^{k_\theta}$, and $\rho \in \mathbb{R}$. The true parameters are denoted by θ_0 and $\rho_{\tau 0} = \exp(\gamma_0/\tau)$ with $\gamma_0 \in \mathbb{R}$ and both θ_0 and γ_0 bounded. We impose the following modified assumptions to account for the the specific features of the unit root model.

Condition 7 *Define $\mathcal{C} = \sigma(\nu_0)$. Define the σ-fields $\mathcal{G}_{n,|\min(1,\tau_0)|n+i}$ in the same way as in (10) except that here $\tau = \kappa n$ such that dependence on τ is suppressed and that ν_t is replaced with η_t as in*

$$\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n+i} = \sigma\left(\left\{y_{jt-1}, y_{jt-2}, \ldots, y_{j\min(1,\tau_0)}\right\}_{j=1}^{n}, \left\{\eta_t, \eta_{t-1}, \ldots, \eta_{\min(1,\tau_0)}\right\}, (y_{j,t})_{j=1}^{i}\right) \vee \mathcal{C}$$

Assume that

i) $f_{\theta,it}$ is measurable with respect to $\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n+i}$.

ii) η_t *is measurable with respect to* $\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n+i}$ *for all* $i = 1, ..., n$

iii) for some $\delta > 0$ *and* $C < \infty$, $\sup_{it} E\left[\|f_{\theta,it}\|^{2+\delta}\right] \leq C$

iv) for some $\delta > 0$ *and* $C < \infty$, $\sup_t E\left[\|\eta_t\|^{2+\delta}\right] \leq C$

v) $E\left[f_{\theta,it}|\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n+i-1}\right] = 0$

vi) $E\left[\eta_t|\mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n+i}\right] = 0$ *for* $t > T$ *and all* $i = \{1, ..., n\}$.

vii) For any $1 > r > s \geq 0$ *fixed let* $\Omega_{\tau,\eta}^{r,s} = \tau^{-1}\sum_{t=\min(1,\tau_0)+[\tau s]+1}^{\tau_0+[\tau r]} E\left[\eta_t^2|\mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n+n}\right]$. *Then,* $\Omega_{\tau,\eta}^{r,s} \rightarrow_p (r-s)\,\sigma^2$.

viii) Assume that $\frac{1}{n}\sum_{i=1}^n f_{\theta,it}f'_{\theta,it} \xrightarrow{p} \Omega_{ty}$ *where* Ω_{ty} *is positive definite a.s. and measurable with respect to* $\mathcal{C}$. *Let* $\Omega_y = \sum_{t=1}^T \Omega_{ty}$.

Conditions 7(i)-(vi) are the same as Conditions 1 (i)-(vi) adapted to the unit root model. Condition 7(vii) replaces Condition 2. It is slightly more primitive in the sense that if η_t^2 is homoskedastic, Condition 7(vii) holds automatically and convergence of $\tau^{-1}\sum_{t=\min(1,\tau_0)+[\tau s]+1}^{\tau_0+[\tau r]} \eta_t^2 \rightarrow (r-s)\,\sigma^2$ follows from an argument given in the proofs rather than being assumed. On the other hand, Condition 7(vii) is somewhat more restrictive than Condition 2 in the sense that it limits heteroskedasticity to be of a form that does not affect the limiting distribution. In other words, we essentially assume $\tau^{-1}\sum_{t=\min(1,\tau_0)+[\tau s]+1}^{\tau_0+[\tau r]} \eta_t^2$ to be proportional to $r-s$ asymptotically. This assumption is stronger than needed but helps to compare the results with the existing unit root literature.

For Condition 7(viii) we note that typically $\Omega_{ty}(\phi) = E\left[f_{\theta,it}f'_{\theta,it}\right]$ and $\Omega_{ty} = \Omega_{ty}(\phi_0)$ where $\phi_0 = (\beta'_0, \nu'_0, \rho_{\tau 0})$. Thus, even if $\Omega_{ty}(.)$ is non-stochastic, it follows that Ω_{ty} is random and measurable with respect to $\mathcal{C}$ because it depends on ν_0 which is a random variable measurable w.r.t $\mathcal{C}$.

The following results are established by modifying arguments in Phillips (1987) and Chan and Wei (1987) to account for $\mathcal{C}$-stable convergence and by applying Theorem 1.

Theorem 3 *Assume that Conditions 7 hold. As* $\tau, n \rightarrow \infty$ *and* T *fixed with* $\tau = \kappa n$ *for some* $\kappa \in (0,\infty)$ *it follows that*

$$\left(V_{\tau n}(r), Y_{\tau n}, \int_0^s V_{\tau n} dW_{\tau n}\right) \Rightarrow \left(V_{\gamma,V(0)}(r), \Omega_y^{1/2}W_y(1), \int_0^s \sigma V_{\gamma,V(0)} dW_\nu\right) \text{ ($\mathcal{C}$-stably)}$$

in the Skorohod topology on $D_{R^d}[0,1]$.

Proof. In Appendix A. ■

We now employ Theorem 3 to analyze the limiting behavior of $\hat{\theta}$ when the common factors are generated from a linear unit root process. To derive a limiting distribution for $\hat{\phi}$ we impose the following additional assumption.

Condition 8 *Let* $\hat{\theta} = \arg\max \sum_{t=1}^{T} \sum_{i=1}^{n} f\left(y_{it}|\theta, \hat{\rho}\right)$. *Assume that* $\left(\hat{\theta} - \theta_0\right) = O_p\left(n^{-1/2}\right)$.

Condition 9 *Let* $\tilde{s}_{it}^{y}\left(\phi\right) = f_{\theta,it}\left(\phi\right)/\sqrt{n}$ *and* $\tilde{s}_{it}^{\nu}\left(\phi\right) = 1\left\{i = 1\right\} g_{\rho,t}\left(\phi\right)/\tau$. *Assume that* $\tilde{s}_{it}^{y}\left(\phi\right) : \mathbb{R}^{k_\theta} \rightarrow \mathbb{R}^{k_\theta}$, $\tilde{s}_{it}^{\nu}\left(\phi\right) : \mathbb{R} \rightarrow \mathbb{R}$ *and define* $D_{n\tau} = \operatorname{diag}\left(n^{-1/2} I_y, \tau^{-1}\right)$, *where* I_y *is an identity matrix of dimension* k_θ. *Let* $\kappa = \lim n/\tau^2$. *Let* $A_{y,\theta}\left(\phi\right) = \sum_{t=1}^{T} E\left[\partial s_{it}^{y}\left(\phi\right)/\partial\theta'\right]$,

$$A_{y,\rho}\left(\phi\right) = \sum_{t=1}^{T} E\left[\partial s_{it}^{y}\left(\phi\right)/\partial\rho\right]$$

and define $A^{y}\left(\phi\right) = \left[\begin{array}{cc} A_{y,\theta}\left(\phi\right) & \sqrt{\kappa} A_{y,\rho}\left(\phi\right) \end{array}\right]$ *where* $A\left(\phi\right)$ *is a* $k_\theta \times k_\phi$ *dimensional matrix of non-random functions* $\phi \rightarrow \mathbb{R}$. *Assume that* $A_{y,\theta}\left(\phi_0\right)$ *is full rank almost surely. Assume that for some* $\varepsilon > 0$,

$$\sup_{\phi:\|\phi-\phi_0\|\leq\varepsilon} \left\| \sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)} \sum_{i=1}^{n} \frac{\partial \tilde{s}_{it}^{y}\left(\phi\right)}{\partial\phi'} D_{n\tau} - A^{y}\left(\phi\right) \right\| = o_p\left(1\right)$$

We make the possibly simplifying assumption that $A\left(\phi\right)$ only depends on the factors through the parameter θ.

Theorem 4 *Assume that Conditions 7, 8 and 9 hold. It follows that*

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} -A_{y,\theta}^{-1}\Omega_y^{1/2} W_y\left(1\right) - \sqrt{\kappa} A_{y,\theta}^{-1} A_{y,\rho} \left(\int_0^1 V_{\gamma,V(0)}^2 dr\right)^{-1} \left(\int_0^1 \sigma V_{\gamma,V(0)} dW_\nu\right) \text{ ($\mathcal{C}$-stably).}$$

Proof. In Appendix A. ■

The result in Theorem 4 is an example that shows how common factors affecting both time series and cross-section data can lead to non-standard limiting distributions. In this case, the initial condition of the unit root process in the time series dimension causes dependence between the components of the asymptotic distribution of $\hat{\theta}$ because both Ω_y and $V_{\gamma,V(0)}$ in general depend on ν_0. Thus, the situation encountered here is generally more difficult than the one considered in Stock and Yogo (2006) and Phillips (2014). In addition, because the limiting distribution of $\hat{\theta}$ is not mixed asymptotically normal, simple pivotal test statistics as in Andrews (2005) are not readily available contrary to the stationary case.

5 Summary

We develop a new limit theory for combined cross-sectional and time-series data sets. We focus on situations where the two data sets are interdependent because of common factors that affect both. The concept of stable convergence is used to handle this dependence when proving a joint Central Limit Theorem. Our analysis is cast in a generic framework of cross-section and time-series based criterion functions that jointly, but not individually, identify the parameters. Within this framework, we show how our limit theory can be used to derive asymptotic approximations to the sampling distribution of estimators that are based on data from both samples. We explicitly consider the unit root case as an example where particularly difficult to handle limiting expressions arise. Our results are expected to be helpful for the econometric analysis of rational expectation models involving individual decision making as well as general equilibrium settings. We investigate these topics, and related implementation issues, in a companion paper.

References

[1] Andrews, D.W.K. (2005): "Cross-Section Regression with Common Shocks," Econometrica 73, pp. 1551-1585.

[2] Aldous, D.J. and G.K. Eagleson (1978): "On mixing and stability of limit theorems," The Annals of Probability 6, 325–331.

[3] Billingsley, P. (1968): "Convergence of Probability Measures," John Wiley and Sons, New York.

[4] Brown, B.M. (1971): "Martingale Central Limit Theorems," Annals of Mathematical Statistics 42, pp. 59-66.

[5] Campbell, J.Y., and M. Yogo (2006): "Efficient Tests of Stock Return Predictability," Journal of Financial Economics 81, pp. 27–60.

[6] Chan, N.H. and C.Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," Annals of Statistics 15, pp.1050-1063.

[7] Dedecker, J and F. Merlevede (2002): "Necessary and Sufficient Conditions for the Conditional Central Limit Theorem," Annals of Probability 30, pp. 1044-1081.

[8] Duffie, D., J. Pan and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump Diffusions," Econometrica, pp. 1343-1376.

[9] Durrett, R. (1996): "Stochastic Calculus," CRC Press, Boca Raton, London, New York.

[10] Eagleson, G.K. (1975): "Martingale convergence to mixtures of infinitely divisible laws," The Annals of Probability 3, 557–562.

[11] Feigin, P. D. (1985): "Stable convergence of Semimartingales," Stochastic Processes and their Applications 19, pp. 125-134.

[12] Hahn, J., and G. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large," Econometrica 70, pp. 1639–57.

[13] Hahn, J., G. Kuersteiner and M. Mazzocco (2016): "Estimation with Aggregate Shocks," manuscript.

[14] Hall, P., and C. Heyde (1980): "Martingale Limit Theory and its Applications," Academic Press, New York.

[15] Hansen, B.E. (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," Econometric Theory 8, pp. 489-500.

[16] Heckman, J.J., L. Lochner, and C. Taber (1998), "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," Review of Economic Dynamics 1, pp. 1-58.

[17] Heckman, J.J., and G. Sedlacek (1985): "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," Journal of Political Economy, 93, pp. 1077-1125.

[18] Jacod, J. and A.N. Shiryaev (2002): "Limit Theorems for stochastic processes," Springer Verlag, Berlin.

[19] Kuersteiner, G.M., and I.R. Prucha (2013): "Limit Theory for Panel Data Models with Cross Sectional Dependence and Sequential Exogeneity," Journal of Econometrics 174, pp. 107-126.

[20] Kuersteiner, G.M and I.R. Prucha (2015): "Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity," CESifo Working Paper No. 5445.

[21] Kurtz and Protter (1991): "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations," Annals of Probability 19, pp. 1035-1070.

[22] Lee, D., and K.I. Wolpin (2006): "Intersectoral Labor Mobility and the Growth of Service Sector," Econometrica 47, pp. 1-46.

[23] Lee, D., and K.I. Wolpin (2010): "Accounting for Wage and Employment Changes in the U.S. from 1968-2000: A Dynamic Model of Labor Market Equilibrium," Journal of Econometrics 156, pp. 68–85.

[24] Mikusheva, A. (2007): "Uniform Inference in Autoregressive Models," Econometrica 75, pp. 1411–1452.

[25] Murphy, K. M. and R. H. Topel (1985): "Estimation and Inference in Two-Step Econometric Models," Journal of Business and Economic Statistics 3, pp. 370 – 379.

[26] Phillips, P.C.B. (1987): "Towards a unified asymptotic theory for autoregression," Biometrika 74, pp. 535-547.

[27] Phillips, P.C.B. (1988): "Regression theory for near integrated time series," Econometrica 56, pp. 1021-1044.

[28] Phillips, P.C.B. and S.N. Durlauf (1986): "Multiple Time Series Regression with Integrated Processes," Review of Economic Studies 53, pp. 473-495.

[29] Phillips, P.C.B. and H.R. Moon (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," Econometrica 67, pp. 1057-1111.

[30] Phillips, P.C.B. (2014): "On Confidence Intervals for Autoregressive Roots and Predictive Regression," Econometrica 82, pp. 1177–1195.

[31] Renyi, A (1963): "On stable sequences of events," *Sankya Ser. A*, 25, 293-302.

[32] Ridder, G., and R. Moffitt (2007): "The Econometrics of Data Combination," in Heckman, J.J. and E.E. Leamer, eds. Handbook of Econometrics, Vol 6, Part B. Elsevier B.V.

[33] Rootzen, H. (1983): "Central limit theory for martingales via random change of time," Department of Statistics, University of North Carolina, Technical Report 28.

[34] Tobin, J. (1950): "A Statistical Demand Function for Food in the U.S.A.," Journal of the Royal Statistical Society, Series A. Vol. 113, pp.113-149.

[35] Wooldridge, J.M. and H. White (1988): "Some invariance principles and central limit theorems for dependent heterogeneous processes," Econometric Theory 4, pp.210-230.

Appendix

A Proofs for Section 4

A.1 Proof of Theorem 1

To prove the functional central limit theorem we follow Billingsley (1968) and Dedecker and Merlevede (2002). The proof involves establishing finite dimensional convergence and a tightness argument. For finite dimensional convergence fix $r_1 < r_2 < \cdots < r_k \in [0,1]$. Define the increment

$$\Delta X_{n\tau}(r_i) = X_{n\tau}(r_i) - X_{n\tau}(r_{i-1}). \tag{23}$$

Since there is a one to one mapping between $X_{n\tau}(r_1), ..., X_{n\tau}(r_k)$ and $X_{n\tau}(r_1), \Delta X_{n\tau}(r_2), ..., \Delta X_{n\tau}(r_k)$ we establish joint convergence of the latter. The proof proceeds by checking that the conditions of Theorem 1 in Kuersteiner and Prucha (2013) hold. Let $k_n = \max(T,\tau)n$ where both $n \to \infty$ and $\tau \to \infty$ such that clearly $k_n \to \infty$ (this is a diagonal limit in the terminology of Phillips and Moon, 1999). Let $d = k_\theta + k_\rho$. To handle the fact that $X_{n\tau} \in \mathbb{R}^d$ we use Lemmas A.1 - A.3 in Phillips and Durlauf (1986). Define $\lambda_j = \left(\lambda'_{j,y}, \lambda'_{j,\nu}\right)'$and let $\lambda = (\lambda_1, \ldots, \lambda_k) \in \mathbb{R}^{dk}$ with $\|\lambda\| = 1$. Define $t^* = t - \min(1, \tau_0)$.

For each n and τ_0 define the mapping $q(t,i) : \mathbb{N}_+^2 \to \mathbb{N}_+$ as $q(i,t) := t^* n + i$ and note that $q(i,t)$ is invertible, in particular for each $q \in \{1, ..., k_n\}$ there is a unique pair t, i such that $q(i,t) = q$. We often use shorthand notation q for $q(i,t)$. Let

$$\ddot{\psi}_{q(i,t)} \equiv \sum_{j=1}^{k} \lambda'_j \left(\Delta\tilde{\psi}_{it}(r_j) - E\left[\Delta\tilde{\psi}_{it}(r_j) \,|\mathcal{G}_{\tau n, t^* n+i-1}\right]\right) \tag{24}$$

where

$$\Delta\tilde{\psi}_{it}(r_j) = \tilde{\psi}_{it}(r_j) - \tilde{\psi}_{it}(r_{j-1});\ \Delta\tilde{\psi}_{it}(r_1) = \tilde{\psi}_{it}(r_1). \tag{25}$$

Note that $\Delta\tilde{\psi}_{it}(r_j) = \left(\Delta\tilde{\psi}^y_{it}(r_j), \Delta\tilde{\psi}^\nu_t(r_j)\right)'$with

$$\Delta\tilde{\psi}^y_{it}(r_j) = \begin{cases} 0 & \text{for } j > 1 \\ \tilde{\psi}^y_{it} & \text{for } j = 1 \end{cases} \tag{26}$$

and

$$\Delta\tilde{\psi}_{it}^{\nu}(r_j) = \begin{cases} \tilde{\psi}_{\tau,t}^{\nu}(r_j) & \text{if } [\tau r_{j-1}] < t \leq [\tau r_j] \text{ and } i = 1 \\ 0 & \text{otherwise} \end{cases}. \tag{27}$$

Using this notation and noting that $\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)} \sum_{i=1}^{n} \ddot{\psi}_{q(i,t)} = \sum_{q=1}^{k_n} \ddot{\psi}_q$, we write

$$\begin{aligned} &\lambda_1' X_{n\tau}(r_1) + \sum_{j=2}^{k} \lambda_j' \Delta X_{n\tau}(r_j) \\ &= \sum_{q=1}^{k_n} \ddot{\psi}_q + \sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)} \sum_{i=1}^{n} \sum_{j=1}^{k} \lambda_j' E\left[\Delta\tilde{\psi}_{it}(r_j) \middle| \mathcal{G}_{\tau n, t^* n+i-1}\right] \end{aligned} \tag{28}$$

First analyze the term $\sum_{q=1}^{k_n} \ddot{\psi}_q$. Note that $\psi_{n,it}^{y}$ is measurable with respect to $\mathcal{G}_{\tau n, t^* n+i}$ by construction. Note that by (24), (26) and (27) the individual components of $\ddot{\psi}_q$ are either 0 or equal to $\tilde{\psi}_{it}(r_j) - E\left[\tilde{\psi}_{it}(r_j) | \mathcal{G}_{\tau n, t^* n+i-1}\right]$ respectively. This implies that $\ddot{\psi}_q$ is measurable with respect to $\mathcal{G}_{\tau n, q}$, noting in particular that $E\left[\tilde{\psi}_{it}(r_j) | \mathcal{G}_{\tau n, t^* n+i-1}\right]$ is measurable w.r.t $\mathcal{G}_{\tau n, t^* n+i-1}$ by the properties of conditional expectations and $\mathcal{G}_{\tau n, t^* n+i-1} \subset \mathcal{G}_{\tau n, q}$. By construction, $E\left[\ddot{\psi}_r | \mathcal{G}_{\tau n, q-1}\right] = 0$. This establishes that for $S_{nq} = \sum_{s=1}^{q} \ddot{\psi}_s$,

$$\{S_{nq}, \mathcal{G}_{\tau n, q}, 1 \leq q \leq k_n, n \geq 1\}$$

is a mean zero martingale array with differences $\ddot{\psi}_q$.

To establish finite dimensional convergence we follow Kuersteiner and Prucha (2013) in the proof of their Theorem 2. Note that, for any fixed n and given q, and thus for a corresponding unique vector (t, i), there exists a unique $j \in \{1, \ldots, k\}$ such that $\tau_0 + [\tau r_{j-1}] < t \leq \tau_0 + [\tau r_j]$. Then,

$$\begin{aligned} \ddot{\psi}_{q(i,t)} &= \sum_{l=1}^{k} \lambda_l' \left(\Delta\tilde{\psi}_{it}(r_l) - E\left[\Delta\tilde{\psi}_{it}(r_l) | \mathcal{G}_{\tau n, t^* n+i-1}\right]\right) \\ &= \lambda_{1,y}' \left(\tilde{\psi}_{it}^{y} - E\left[\tilde{\psi}_{it}^{y} | \mathcal{G}_{\tau n, t^* n+i-1}\right]\right) 1\{j=1\} \\ &+ \lambda_{j,\nu}' \left(\tilde{\psi}_{\tau,t}^{\nu}(r_j) - E\left[\tilde{\psi}_{\tau,t}^{\nu}(r_j) | \mathcal{G}_{\tau n, t^* n+i-1}\right]\right) 1\{[\tau r_{j-1}] < t \leq [\tau r_j]\} 1\{ i=1\} \end{aligned}$$

where all remaining terms in the sum are zero because of by (24), (26) and (27). For the subsequent inequalities, fix $q \in \{1, ..., k_n\}$ (and the corresponding (t, i) and j) arbitrarily. Introduce the shorthand notation $1_j = 1\{j = 1\}$ and $1_{ij} = 1\{[\tau r_{j-1}] < t \leq [\tau r_j]\} 1\{ i = 1\}$.

First, note that for $\delta \geq 0$, and by Jensen's inequality applied to the empirical measure $\frac{1}{4}\sum_{i=1}^{4} x_i$ we have that

$$
\begin{aligned}
&\left|\ddot{\psi}_q\right|^{2+\delta} \\
&= 4^{2+\delta}\left|\frac{1}{4}\lambda_{1,y}'\left(\tilde{\psi}_{it}^{y} - E\left[\tilde{\psi}_{it}^{y}|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right)1_j + \frac{1}{4}\lambda_{j,\nu}'\left(\tilde{\psi}_{\tau,t}^{\nu}\left(r_j\right) - E\left[\tilde{\psi}_{\tau,t}^{\nu}\left(r_j\right)|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right)1_{ij}\right|^{2+\delta} \\
&\leq 4^{2+\delta}\left(\frac{1}{4}\left\|\lambda_{1,y}\right\|^{2+\delta}\left\|\tilde{\psi}_{it}^{y}\right\|^{2+\delta} + \frac{1}{4}\left\|\lambda_{1,y}\right\|^{2+\delta}\left\|E\left[\tilde{\psi}_{it}^{y}|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right\|^{2+\delta}\right)1_j \\
&+ 4^{2+\delta}\left(\frac{1}{4}\left\|\lambda_{j,\nu}\right\|^{2+\delta}\left\|\tilde{\psi}_{\tau,t}^{\nu}\left(r_j\right)\right\|^{2+\delta} + \frac{1}{4}\left\|\lambda_{j,\nu}\right\|^{2+\delta}\left\|E\left[\tilde{\psi}_{\tau,t}^{\nu}\left(r_j\right)|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right\|^{2+\delta}\right)1_{ij} \\
&= 2^{2+2\delta}\left(\left\|\lambda_{1,y}\right\|^{2+\delta}\left\|\tilde{\psi}_{it}^{y}\right\|^{2+\delta} + \left\|\lambda_{1,y}\right\|^{2+\delta}\left\|E\left[\tilde{\psi}_{it}^{y}|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right\|^{2+\delta}\right)1_j \\
&+ 2^{2+2\delta}\left(\left\|\lambda_{j,\nu}\right\|^{2+\delta}\left\|\tilde{\psi}_{\tau,t}^{\nu}\left(r_j\right)\right\|^{2+\delta} + \left\|\lambda_{j,\nu}\right\|^{2+\delta}\left\|E\left[\tilde{\psi}_{\tau,t}^{\nu}\left(r_j\right)|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right\|^{2+\delta}\right)1_{ij}.
\end{aligned}
$$

We further use the definitions in (11) such that by Jensen's inequality and for $i = 1$ and $t \in [\tau_0 + 1, \tau_0 + \tau]$

$$
\begin{aligned}
&\left\|\tilde{\psi}_{\tau,t}^{\nu}\left(r_j\right)\right\|^{2+\delta} + \left\|E\left[\tilde{\psi}_{\tau,t}^{\nu}\left(r_j\right)|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right\|^{2+\delta} \\
&\leq \frac{1}{\tau^{1+\delta/2}}\left(\left\|\psi_{\tau,t}^{\nu}\right\|^{2+\delta} + \left(E\left[\left\|\psi_{\tau,t}^{\nu}\right\||\mathcal{G}_{\tau n,t^*n+i-1}\right]\right)^{2+\delta}\right) \\
&\leq \frac{1}{\tau^{1+\delta/2}}\left(\left\|\psi_{\tau,t}^{\nu}\right\|^{2+\delta} + E\left[\left\|\psi_{\tau,t}^{\nu}\right\|^{2+\delta}|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right)
\end{aligned}
$$

while for $i > 1$ or $t \notin [\tau_0 + 1, \tau_0 + \tau]$,

$$\left\|\tilde{\psi}_{it}^{\nu}\right\| = 0.$$

Similarly, for $t \in [1, ..., T]$

$$
\begin{aligned}
&\left\|\tilde{\psi}_{it}^{y}\right\|^{2+\delta} + \left\|E\left[\tilde{\psi}_{it}^{y}|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right\|^{2+\delta} \\
&\leq \frac{1}{n^{1+\delta/2}}\left(\left\|\psi_{it}^{y}\right\|^{2+\delta} + E\left[\left\|\psi_{it}^{y}\right\|^{2+\delta}|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right)
\end{aligned}
$$

while for $t \notin [1, ..., T]$

$$\left\|\tilde{\psi}_{it}^{y}\right\| = 0.$$

Noting that $\left\|\lambda_{j,y}\right\| \leq 1$ and $\left\|\lambda_{j,\nu}\right\| < 1$,

$$
\begin{aligned}
E\left[\left|\ddot{\psi}_q\right|^{2+\delta}\middle|\mathcal{G}_{\tau n,q-1}\right] \leq{}& \frac{2^{3+2\delta}1\left\{i = 1, t \in [\tau_0 + 1, \tau_0 + \tau]\right\}}{\tau^{1+\delta/2}}E\left[\left\|\psi_{\tau,t}^{\nu}\right\|^{2+\delta}\middle|\mathcal{G}_{\tau n,t^*n+i-1}\right] \\
&+ \frac{2^{3+2\delta}1\left\{t \in [1, ..., T]\right\}}{n^{1+\delta/2}}E\left[\left\|\psi_{it}^{y}\right\|^{2+\delta}\middle|\mathcal{G}_{\tau n,t^*n+i-1}\right], \qquad (29)
\end{aligned}
$$

where the inequality in (29) holds for $\delta \geq 0$. To establish the limiting distribution of $\sum_{q=1}^{k_n} \ddot{\psi}_q$ we check that

$$\sum_{q=1}^{k_n} E\left[\left|\ddot{\psi}_q\right|^{2+\delta}\right] \to 0, \tag{30}$$

$$\sum_{q=1}^{k_n} \ddot{\psi}_q^2 \xrightarrow{p} \sum_{t\in\{1,..,T\}} \lambda_{1,y}'\Omega_{yt}\lambda_{1,y} + \sum_{j=1}^{k} \lambda_{j,\nu}'\Omega_{\nu}\left(r_j - r_{j-1}\right)\lambda_{j,\nu}, \tag{31}$$

and

$$\sup_n E\left[\left(\sum_{q=1}^{k_n} E\left[\ddot{\psi}_q^2\middle|\mathcal{G}_{\tau n,q-1}\right]\right)^{1+\delta/2}\right] < \infty, \tag{32}$$

which are adapted to the current setting from Conditions (A.26), (A.27) and (A.28) in Kuersteiner and Prucha (2013). These conditions in turn are related to conditions of Hall and Heyde (1980) and are shown by Kuersteiner and Prucha (2013) to be sufficient for their Theorem 1.

To show that (30) holds note that from (29) and Condition 1 it follows that for some constant $C < \infty$,

$$\begin{aligned}
\sum_{q=1}^{k_n} E\left[\left|\ddot{\psi}_q\right|^{2+\delta}\right] &\leq \frac{2^{3+2\delta}}{\tau^{1+\delta/2}} \sum_{t=\tau_0+1}^{\tau_0+\tau} E\left[\left\|\psi_{\tau,t}^{\nu}\right\|^{2+\delta}\middle|\mathcal{G}_{\tau n,t^*n}\right] \\
&+ \frac{2^{3+2\delta}}{n^{1+\delta/2}} \sum_{t=1}^{T}\sum_{i=1}^{n} E\left[\left\|\psi_{it}^{y}\right\|^{2+\delta}\middle|\mathcal{G}_{\tau n,t^*n}\right] \\
&\leq \frac{2^{3+2\delta}\tau C}{\tau^{1+\delta/2}} + \frac{2^{3+2\delta}nTC}{n^{1+\delta/2}} = \frac{2^{3+2\delta}C}{\tau^{\delta/2}} + \frac{2^{3+2\delta}TC}{n^{\delta/2}} \to 0
\end{aligned}$$

because $2^{3+2\delta}C$ and T are fixed as $\tau, n \to \infty$.

Next, consider the probability limit of $\sum_{q=1}^{k_n} \ddot{\psi}_q^2$. We have

$$\sum_{q=1}^{k_n} \ddot{\psi}_q^2$$

$$= \frac{1}{\tau}\sum_{j=1}^{k}\sum_{t=\tau_0+1}^{\tau_0+\tau}\left(\lambda_{j,\nu}'\left(\psi_{\tau,t}^{\nu} - E\left[\psi_{\tau,t}^{\nu}|\mathcal{G}_{\tau n,t^*n}\right]\right)\right)^2 1_{ij} \tag{33}$$

$$+ \frac{2}{\sqrt{\tau n}} \sum_{\substack{t\in\{\tau_0,\ldots,\tau_0+\tau\} \\ \cap\{\min(1,\tau_0),..,T\}}} \lambda_{1,\nu}'\left(\psi_{\tau,t}^{\nu} - E\left[\psi_{\tau,t}^{\nu}|\mathcal{G}_{\tau n,t^*n}\right]\right)\left(\psi_{1t}^{y} - E\left[\psi_{1t}^{y}|\mathcal{G}_{\tau n,t^*n}\right]\right)'\lambda_{1,y}1\left\{t \leq \tau_0 + [\tau r_1]\right\}1_{ij}1_j \tag{34}$$

$$+ \frac{1}{n}\sum_{t\in\{\min(1,\tau_0),..,T\}}\sum_{i=1}^{n}\left(\lambda_{1,y}'\left(\psi_{n,it}^{y} - E\left[\psi_{n,it}^{y}|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right)\right)^2 1_j \tag{35}$$

where for (33) we note that $E\left[\psi^{\nu}_{\tau,t}|\mathcal{G}_{\tau n,t^*n}\right]=0$ when $t>T$. This implies that

$$\begin{aligned}
&\frac{1}{\tau}\sum_{j=1}^{k}\sum_{t=\tau_0+1}^{\tau_0+\tau}\left(\lambda'_{j,\nu}\left(\psi^{\nu}_{\tau,t}-E\left[\psi^{\nu}_{\tau,t}|\mathcal{G}_{\tau n,t^*n}\right]\right)\right)^2 1_{ij}\\
&=\frac{1}{\tau}\sum_{j=1}^{k}\sum_{t\in\{\tau_0,\dots,\tau_0+\tau\}\cap\{\min(1,\tau_0+1),..,T\}}\left(\lambda'_{j,\nu}\left(\psi^{\nu}_{\tau,t}-E\left[\psi^{\nu}_{\tau,t}|\mathcal{G}_{\tau n,t^*n}\right]\right)\right)^2 1_{ij}\\
&+\frac{1}{\tau}\sum_{j=1}^{k}\sum_{t=\max\{\tau_0,T\}}^{\tau_0+\tau}\left(\lambda'_{j,\nu}\psi^{\nu}_{\tau,t}\right)^2 1_{ij}
\end{aligned}$$

where by Condition 2

$$\frac{1}{\tau}\sum_{j=1}^{k}\sum_{t=\max\{\tau_0,T\}}^{\tau_0+\tau}\left(\lambda'_{j,\nu}\psi^{\nu}_{\tau,t}\right)^2 1\left\{\tau_0+[\tau r_{j-1}]<t\leq\tau_0+[\tau r_j]\right\}\xrightarrow{p}\sum_{j=1}^{k}\lambda'_{j,\nu}\left(\Omega_\nu\left(r_j\right)-\Omega_\nu\left(r_{j-1}\right)\right)\lambda_{j,\nu}$$

and

$$E\left[\left\|\frac{1}{\tau}\sum_{j=1}^{k}\sum_{\substack{t\in\{\tau_0,\dots,\tau_0+\tau\}\\ \cap\{\min(1,\tau_0+1),..,T\}}}\left(\lambda'_{j\nu}\left(\psi^{\nu}_{\tau,t}-E\left[\psi^{\nu}_{\tau,t}|\mathcal{G}_{\tau n,t^*n}\right]\right)\right)^2 1_{ij}\right\|^{1+\delta/2}\right] \tag{36}$$

$$\begin{aligned}
&=\frac{1}{\tau^{1+\delta/2}}E\left[\left(\sum_{j=1}^{k}\sum_{\substack{t\in\{\tau_0,\dots,\tau_0+\tau\}\\ \cap\{\min(1,\tau_0+1),..,T\}}}\left\|\lambda'_{\nu}\left(\psi^{\nu}_{\tau,t}-E\left[\psi^{\nu}_{\tau,t}|\mathcal{G}_{\tau n,t^*n}\right]\right)\right\|^2 1_{ij}\right)^{1+\delta/2}\right]\\
&\leq\frac{\left(T+|\tau_0|\right)^{\delta/2}k^{\delta/2}}{\tau^{1+\delta/2}}\sum_{j=1}^{k}\sum_{t\in\{\tau_0,\dots,\tau_0+\tau\}\cap\{\min(1,\tau_0+1),..,T\}}E\left[\left\|\lambda'_{\nu}\left(\psi^{\nu}_{\tau,t}-E\left[\psi^{\nu}_{\tau,t}|\mathcal{G}_{\tau n,t^*n}\right]\right)\right\|^{2+\delta}\right]\\
&\leq\frac{2^{2+\delta}\left(T+|\tau_0|\right)^{1+\delta/2}k^{1+\delta/2}}{\tau^{1+\delta/2}}\sup_t E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^{2+\delta}\right]\to 0,
\end{aligned}$$

where the first inequality follows from noting that the set $\{\tau_0,...,\tau_0+\tau\}\cap\{\min(1,\tau_0),..,T\}$ has at most $T+|\tau_0|$ elements and from using Jensen's inequality on the counting measure. The second inequality follows from Hölder's inequality. Finally, we use the fact that $\left(T+|\tau_0|\right)/\tau\to 0$ and $\sup_t E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^{2+\delta}\right]\leq C<\infty$ by Condition 1(iv).

Next consider (34) where

$$E\left[\left|\frac{2}{\sqrt{\tau n}}\sum_{\substack{t\in\{\tau_0,\dots,\tau_0+\tau\}\\ \cap\{\min(1,\tau_0),..,T\}}}\lambda_{1,\nu}'\left(\psi_{\tau,t}^{\nu}-E\left[\psi_{\tau,t}^{\nu}|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right)\left(\psi_{1t}^{y}-E\left[\psi_{1t}^{y}|\mathcal{G}_{\tau n,t^*n}\right]\right)'\lambda_{1,y}1\left\{t\le\tau_0+[\tau r_1]\right\}\right|\right]$$

$$\le\frac{2}{\sqrt{\tau n}}\sum_{\substack{t\in\{\tau_0,\dots,\tau_0+\tau\}\\ \cap\{\min(1,\tau_0),..,T\}}}\left\{\left(E\left[\left|\lambda_{1\nu}'\left(\psi_{\tau,t}^{\nu}-E\left[\psi_{\tau,t}^{\nu}|\mathcal{G}_{\tau n,t^*n}\right]\right)\right|^2\right]\right)^{1/2}\right.$$

$$\left.\times\left(E\left[\left|\left(\psi_{1t}^{y}-E\left[\psi_{1t}^{y}|\mathcal{G}_{\tau n,t^*n}\right]\right)'\lambda_{1y}\right|^2\right]\right)^{1/2}\right\}1\left\{\tau_0<t\le\tau_0+[\tau r_j]\right\}\tag{37}$$

$$\le\frac{2^3}{\sqrt{\tau n}}\sup_t\left(E\left[\left\|\psi_{\tau,t}^{\nu}\right\|\right]\right)^{1/2}\sum_{\substack{t\in\{\tau_0,\dots,\tau_0+\tau\}\\ \cap\{\min(1,\tau_0),..,T\}}}\left(E\left[\left|\left(\psi_{1t}^{y}-E\left[\psi_{1t}^{y}|\mathcal{G}_{\tau n,t^*n}\right]\right)'\lambda_{1,y}\right|^2\right]\right)^{1/2}\tag{38}$$

$$\le\frac{2^4T+|\tau_0|}{\sqrt{\tau n}}\sup_t\left(E\left[\left\|\psi_{\tau,t}^{\nu}\right\|\right]\right)^{1/2}\left(\sup_{i,t}E\left[\left\|\psi_{n,it}^{y}\right\|^2\right]\right)^{1/2}\to 0\tag{39}$$

where the first inequality in (37) follows from the Cauchy-Schwartz inequality, (38) uses Condition 1(iv), and the last inequality uses Condition 1(iii). Then we have in (38), by Condition 1(iii) and the Hölder inequality that

$$E\left[\left|\left(\psi_{1t}^{y}-E\left[\psi_{1t}^{y}|\mathcal{G}_{\tau n,t^*n}\right]\right)'\lambda_y\right|^2\right]\le 2E\left[\left\|\psi_{1t}^{y}\right\|^2\right]$$

such that (39) follows.

We note that (39) goes to zero as long as $T/\sqrt{\tau n}\to 0$. Clearly, this condition holds as long as T is held fixed, but holds under weaker conditions as well.

Next the limit of (35) is, by Condition 1(v) and Condition 3,

$$\frac{1}{n}\sum_{t\in\{1,..,T\}}\sum_{i=1}^{n}\left(\lambda_{1,y}'\left(\psi_{n,it}^{y}-E\left[\psi_{n,it}^{y}|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right)\right)^2\xrightarrow{p}\sum_{t\in\{1,..,T\}}\lambda_{1,y}'\Omega_{yt}\lambda_{1,y}.$$

This verifies (31). Finally, for (32) we check that

$$\sup_n E\left[\left(\sum_{q=1}^{k_n}E\left[\left|\ddot{\psi}_q\right|^2\middle|\mathcal{G}_{\tau n,q-1}\right]\right)^{1+\delta/2}\right]<\infty.\tag{40}$$

First, use (29) with $\delta=0$ to obtain

$$\sum_{q=1}^{k_n}E\left[\left|\ddot{\psi}_q\right|^2\middle|\mathcal{G}_{\tau n,q-1}\right]\le\frac{2^3}{\tau}\sum_{t=\tau_0}^{\tau_0+\tau}E\left[\left\|\psi_{\tau,t}^{\nu}\right\|^2\middle|\mathcal{G}_{\tau n,t^*n}\right]$$

$$+\frac{2^3}{n}\sum_{t\in\{1,..,T\}}\sum_{i=1}^{n}E\left[\left\|\psi_{n,it}^{y}\right\|^2\middle|\mathcal{G}_{\tau n,t^*n+i-1}\right].\tag{41}$$

Applying (41) to (40) and using the Hölder inequality implies

$$E\left[\left(\sum_{q=1}^{k_n} E\left[\left|\ddot{\psi}_q\right|^2 \middle| \mathcal{G}_{\tau n,q-1}\right]\right)^{1+\delta/2}\right]$$
$$\leq 2^{\delta/2} E\left[\left(\frac{2^3}{\tau}\sum_{t=\tau_0+1}^{\tau_0+\tau} E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^2 \middle| \mathcal{G}_{\tau n,t^*n+i-1}\right]\right)^{1+\delta/2}\right]$$
$$+ 2^{\delta/2} E\left[\left(\frac{2^3}{n}\sum_{t\in\{\tau_0,\ldots,\tau_0+\tau\}\cap\{1,..,T\}}\sum_{i=1}^{n} E\left[\left\|\psi^{y}_{n,it}\right\|^2 \middle| \mathcal{G}_{\tau n,t^*n+i-1}\right]\right)^{1+\delta/2}\right]$$

By Jensen's inequality, we have

$$\left(\frac{1}{\tau}\sum_{t=\tau_0+1}^{\tau_0+\tau} E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^2 \middle| \mathcal{G}_{\tau n,t^*n+i-1}\right]\right)^{1+\delta/2} \leq \frac{1}{\tau}\sum_{t=\tau_0+1}^{\tau_0+\tau} E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^2 \middle| \mathcal{G}_{\tau n,t^*n+i-1}\right]^{1+\delta/2}$$

and

$$E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^2 \middle| \mathcal{G}_{\tau n,t^*n+i-1}\right]^{1+\delta/2} \leq E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^{2+\delta} \middle| \mathcal{G}_{\tau n,t^*n+i-1}\right]$$

so that

$$E\left[\left(\frac{2^3}{\tau}\sum_{t=\tau_0+1}^{\tau_0+\tau} E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^2 \middle| \mathcal{G}_{\tau n,t^*n+i-1}\right]\right)^{1+\delta/2}\right] \leq \frac{2^{3+3\delta/2}}{\tau}\sum_{t=\tau_0+1}^{\tau_0+\tau} E\left[E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^{2+\delta} \middle| \mathcal{G}_{\tau n,t^*n+i-1}\right]\right]$$
$$\leq 2^{3+3\delta/2}\sup_t E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^{2+\delta}\right] < \infty. \tag{42}$$

and similarly, for all $\tau > T$ (which holds eventually)

$$E\left[\left(\frac{2^3}{n}\sum_{t\in\{1,..,T\}}\sum_{i=1}^{n} E\left[\left\|\psi^{y}_{n,it}\right\|^2 |\mathcal{G}_{\tau n,t^*n+i-1}\right]\right)^{1+\delta/2}\right] \tag{43}$$
$$\leq \frac{2^{3+3\delta/2}(Tn)^{\delta/2}}{n^{1+\delta/2}}\sum_{t=1}^{T}\sum_{i=1}^{n} E\left[\left\|\psi^{y}_{n,it}\right\|^{2+\delta}\right]$$
$$\leq 2^{3+3\delta/2}T^{1+\delta/2}\sup_{i,t} E\left[\left\|\psi^{y}_{n,it}\right\|^{2+\delta}\right] < \infty$$

By combining (42) and (43) we obtain the following bound for (40),

$$E\left[\left(\sum_{q=1}^{k_n} E\left[\left|\ddot{\psi}_q\right|^{2+\delta} |\mathcal{G}_{\tau n,q-1}\right]\right)^{1+\delta/2}\right]$$
$$\leq 2^{3+3\delta/2}\sup_t E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^{2+\delta}\right] + 2^{3+3\delta/2}T^{1+\delta/2}\sup_{i,t} E\left[\left\|\psi^{y}_{n,it}\right\|^{2+\delta}\right] < \infty.$$

This establishes that (30), (31) and (32) hold and thus establishes the CLT for $\sum_{q=1}^{k_n} \ddot{\psi}_q$.

It remains to be shown that the second term in (28) can be neglected. Consider

$$\begin{aligned}
&\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)} \sum_{i=1}^{n} \sum_{j=1}^{k} \lambda_j' E\left[\Delta\tilde{\psi}_{it}\left(r_j\right)|\mathcal{G}_{\tau n,t^*n+i-1}\right] \\
&= \tau^{-1/2} \sum_{t=\tau_0}^{\tau_0+\tau} \sum_{j=1}^{k} \lambda_{j,\nu}' E\left[\psi_{\tau,t}^{\nu}|\mathcal{G}_{\tau n,t^*n+i-1}\right] 1\left\{\tau_0 + [\tau r_{j-1}] < t \leq \tau_0 + [\tau r_j]\right\} \\
&+ n^{-1/2} \sum_{t=1}^{T} \sum_{i=1}^{n} \lambda_{1,y}' E\left[\psi_{n,it}^{y}|\mathcal{G}_{\tau n,t^*n+i-1}\right].
\end{aligned}$$

Note that

$$E\left[\psi_{\tau,t}^{\nu}|\mathcal{G}_{\tau n,t^*n+i-1}\right] = 0 \text{ for } t > T$$

and

$$E\left[\psi_{n,it}^{y}|\mathcal{G}_{\tau n,t^*n+i-1}\right] = 0.$$

This implies, using the convention that a term is zero if it is a sum over indices from a to b with $a > b$, that

$$\begin{aligned}
&\tau^{-1/2} \sum_{t=\tau_0}^{\tau_0+\tau} \lambda_{\nu}' E\left[\psi_{\tau,t}^{\nu}|\mathcal{G}_{\tau n,t^*n+i-1}\right] \\
&= \tau^{-1/2} \sum_{t=\tau_0}^{T} \sum_{j=1}^{k} \lambda_{j,\nu}' E\left[\psi_{\tau,t}^{\nu}|\mathcal{G}_{\tau n,t^*n+i-1}\right] 1\left\{\tau_0 + [\tau r_{j-1}] < t \leq \tau_0 + [\tau r_j]\right\}.
\end{aligned}$$

By a similar argument used to show that (36) vanishes, and noting that T is fixed while $\tau \to \infty$, it follows that

$$E\left[\left\|\tau^{-1/2} \sum_{t=\tau_0}^{T} \lambda_{j,\nu}' E\left[\psi_{\tau,t}^{\nu}|\mathcal{G}_{\tau n,t^*n}\right]\right\|^{1+\delta/2}\right] \to 0$$

as $\tau \to \infty$. The Markov inequality then implies that

$$\tau^{-1/2} \sum_{t=\tau_0}^{\tau_0+\tau} \sum_{i=1}^{n} \sum_{j=1}^{k} \lambda_j' E\left[\Delta\tilde{\psi}_{it}\left(r_j\right)|\mathcal{G}_{\tau n,t^*n+i-1}\right] = o_p\left(1\right).$$

and consequently that

$$\lambda_1' X_{n\tau}\left(r_1\right) + \sum_{j=2}^{k} \lambda_j' \Delta X_{n\tau}\left(r_j\right) = \sum_{q=1}^{k_n} \ddot{\psi}_q + o_p\left(1\right). \tag{44}$$

We have shown that the conditions of Theorem 1 of Kuersteiner and Prucha (2013) hold by establishing (30), (31), (32) and (44). Applying the Cramer-Wold theorem to the vector

$$Y_{nt} = \left(X_{n\tau}\left(r_1\right)', \Delta X_{n\tau}\left(r_2\right)' ..., \Delta X_{n\tau}\left(r_k\right)'\right)'$$

it follows from Theorem 1 in Kuersteiner and Prucha (2013) that for all fixed $r_1, .., r_k$ and using the convention that $r_0 = 0$,

$$E\left[\exp\left(i\lambda' Y_{nt}\right)\right] \to E\left[\exp\left(-\frac{1}{2}\left(\sum_{t\in\{1,..,T\}} \lambda_{1,y}'\Omega_{yt}\lambda_{1,y} + \sum_{j=1}^{k} \lambda_{j,\nu}'\left(\Omega_\nu\left(r_j\right) - \Omega_\nu\left(r_{j-1}\right)\right)\lambda_{j,\nu}\right)\right)\right]. \tag{45}$$

When $\Omega_\nu\left(r\right) = r\Omega_\nu$ for all $r \in [0,1]$ and some Ω_ν positive definite and measurable w.r.t $\mathcal{C}$ this result simplifies to

$$\sum_{j=1}^{k} \lambda_{j,\nu}'\left(\Omega_\nu\left(r_j\right) - \Omega_\nu\left(r_{j-1}\right)\right)\lambda_{j,\nu} = \sum_{j=1}^{k} \lambda_{j,\nu}'\Omega_\nu\lambda_{j,\nu}\left(r_j - r_{j-1}\right).$$

The second step in establishing the functional CLT involves proving tightness of the sequence $\lambda' X_{n\tau}\left(r\right)$. By Lemma A.3 of Phillips and Durlauf (1986) and Proposition 4.1 of Wooldridge and White (1988), see also Billingsley (1968, p.41), it is enough to establish tightness componentwise. This is implied by establishing tightness for $\lambda' X_{n\tau}\left(r\right)$ for all $\lambda \in \mathbb{R}^d$ such that $\lambda'\lambda = 1$. In the following we make use of Theorems 8.3 and 15.5 in Billingsley (1968). We need to show that for the 'modulus of continuity'

$$\omega\left(X_{n\tau}, \delta\right) = \sup_{|t-s|<\delta}\left|\lambda'\left(X_{n\tau}\left(s\right) - X_{n\tau}\left(t\right)\right)\right| \tag{46}$$

where $t, s \in [0,1]$ it follows that

$$\lim_{\delta\to 0}\limsup_{n,\tau} P\left(\omega\left(X_{n\tau}, \delta\right) \geq \varepsilon\right) = 0.$$

Define

$$X_{n\tau,y}\left(r\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{T}\sum_{i=1}^{n}\psi_{n,it}^{y},\ X_{n\tau,\nu}\left(r\right) = \frac{1}{\sqrt{\tau}}\sum_{t=\tau_0+1}^{\tau_0+[\tau r]}\psi_{\tau,t}^{\nu}.$$

Since

$$\left|\lambda'\left(X_{n\tau}\left(s\right) - X_{n\tau}\left(t\right)\right)\right| \leq \left|\lambda_y'\left(X_{n\tau,y}\left(s\right) - X_{n\tau,y}\left(t\right)\right)\right| + \left|\lambda_\nu'\left(X_{n\tau,\nu}\left(s\right) - X_{n\tau,\nu}\left(t\right)\right)\right|$$

and noting that $\left|\lambda_y' \left(X_{n\tau,y}(s) - X_{n\tau,y}(t)\right)\right| = 0$ uniformly in $t, s \in [0,1]$ because of the initial condition $X_{n\tau}(0)$ given in (13) and the fact that $X_{n\tau,y}(t)$ is constant as a function of t. It follows that

$$\omega(X_{n\tau}, \delta) \leq \sup_{|s-t|<\delta} |\lambda_\nu' (X_{n\tau,\nu}(s) - X_{n\tau,\nu}(t))| \tag{47}$$

such that

$$P(\omega(X_{n\tau}, \delta) \geq 3\varepsilon) \leq P(\omega(X_{n\tau,\nu}, \delta) \geq 3\varepsilon).$$

To analyze the term in (47) use Billingsley (1968, Theorem 8.4) and the comments in Billingsley (1968, p. 59). Let $S_s = \sum_{t=\tau_0+1}^{\tau_0+s} \lambda_\nu' \psi_{\tau,t}^\nu$. To establish tightness it is enough to show that for each $\varepsilon > 0$ there exists $c > 1$ and τ' such that if $\tau > \tau'$

$$P\left(\max_{s\leq\tau} |S_{k+s} - S_k| > c\varepsilon\sqrt{\tau}\right) \leq \frac{\varepsilon}{c^2} \tag{48}$$

hold for all k. Note that for each k fixed, $M_s = S_{s+k} - S_k$ and $\mathcal{F}_s = \mathcal{G}_{\tau n,(s+k-\min(1,\tau_0)n+1)}$, $\{M_s, \mathcal{F}_s\}$ is a martingale. By a maximal inequality, see Hall and Heyde (1980, Corollary 2.1), it follows that for each k

$$\begin{aligned}
P\left(\max_{s\leq\tau} |S_{k+s} - S_k| > c\varepsilon\sqrt{\tau}\right) &= P\left(\max_{s\leq\tau} |S_{k+s} - S_k|^p > (c\varepsilon)^p \tau^{p/2}\right) \\
&\leq \frac{1}{(c\varepsilon)^p \tau^{p/2}} E\left[\left|\sum_{t=k+1}^{k+\tau} \lambda_\nu' \psi_{\tau,t}^\nu\right|^p\right] \\
&\leq \frac{2^p \tau^{p/2}}{(c\varepsilon)^p \tau^{p/2}} \sup_t E\left[\left\|\psi_{\tau,t}^\nu\right\|^p\right] = \frac{\varepsilon}{c^2} \frac{2^p}{c^{p-2}\varepsilon^{1+p}} \sup_t E\left[\left\|\psi_{\tau,t}^\nu\right\|^p\right]
\end{aligned} \tag{49}$$

by an inequality similar to (36). Note that the bound for (49) does not depend on k. Now choose $c = 2^{p/(p-2)} \left(\sup_t E\left[\left\|\psi_{\tau,t}^\nu\right\|^p\right]\right)^{1/(p-2)} / \varepsilon^{(1+p)/(p-2)}$ such that (48) follows. We now identify the limiting distribution using the technique of Rootzen (1983). Tightness together with finite dimensional convergence in distribution in (45), Condition 2 and the fact that the partition $r_1, ..., r_k$ is arbitrary implies that for $\lambda \in \mathbb{R}^d$ with $\lambda = \left(\lambda_y', \lambda_\nu'\right)'$

$$E\left[\exp\left(i\lambda' X_{n\tau}(r)\right)\right] \to E\left[\exp\left(-\frac{1}{2}\left(\lambda_y'\Omega_y\lambda_y + \lambda_\nu'\Omega_\nu(r)\lambda_\nu\right)\right)\right] \tag{50}$$

with $\Omega_y = \sum_{t\in\{1,..,T\}} \Omega_{yt}$. Let $W(r) = (W_y(r), W_\nu(r))$ be a vector of mutually independent standard Brownian motion processes in $\mathbb{R}^d$, independent of any $\mathcal{C}$-measurable random variable. We note that the RHS of (50) is the same as

$$E\left[\exp\left(-\frac{1}{2}\left(\lambda_y'\Omega_y\lambda_y + \lambda_\nu'\Omega_\nu(r)\lambda_\nu\right)\right)\right] = E\left[\exp\left(i\lambda_y'\Omega_y^{1/2} W_y(1) + i\int_0^r \lambda_\nu' \left(\dot{\Omega}_\nu(t)\right)^{1/2} dW_\nu(t)\right)\right]. \tag{51}$$

The result in (51) can be deduced in the same way as in Duffie, Pan and Singleton (2000), in particular p.1371 and their Proposition 1. Conjecture that $X_t = \int_0^t (\partial\Omega_\nu(t)/\partial t)^{1/2} dW_\nu(t)$. By Condition 2(iii) and the fact that $\partial\Omega_\nu(t)/\partial t$ does not depend on X_t it follows that the conditions of Durrett (1996, Theorem 2.8, Chaper 5) are satisfied. This means that the stochastic differential equation $X_t = \int_0^t (\partial\Omega_\nu(t)/\partial t)^{1/2} dW_\nu(s)$ with initial condition $X_0 = 0$ has a strong solution $\left(X, W_t, \mathcal{F}_t^W \vee \mathcal{C}\right)$ where $\mathcal{F}_t^W$ is the filtration generated by $W_\nu(t)$. Then, X_t is a martingale (and thus a local martingale) w.r.t the filtration $\mathcal{F}_t^W$. For $\mathcal{C}$-measurable functions $\alpha(t) : [0,1] \to \mathbb{R}$ and $\beta(t) : [0,1] \to \mathbb{R}^{k_\rho}$ define the transformation

$$\Psi_r = \exp\left(\alpha(r) + \beta(r)' X_r\right).$$

The terminal conditions $\alpha(r) = 0$ and $\beta(r) = i\lambda_\nu$ are imposed such that

$$\Psi_r = \exp\left(\alpha(r) + \beta(r)' X_r\right) = \exp\left(i\lambda_\nu' X_r\right).$$

The goal is now to show that

$$E[\Psi_r|\mathcal{C}] = \Psi_0 = \exp\left(a(0) + \beta(0)' X_0\right) = \exp(a(0))$$

where the initial condition $X_0 = 0$ was used. In other words, we need to find $\alpha(t)$ and $\beta(t)$ such that Ψ_r is a martingale. Following the proof of Proposition 1 in Duffie, Pan and Singleton (2000) and letting $\eta_t = \Psi_t \beta(t)' \sigma(X_t)$, $\mu_\psi(t) = \frac{1}{2}\beta(t)' \sigma(X_t)\sigma(X_t)\beta(t) + \dot{\alpha}(t) + \dot{\beta}(t) X_t$ use Ito's Lemma to obtain

$$\Psi_r = \Psi_0 + \int_0^r \Psi_s \mu_\psi(s)\, ds + \int_0^r \eta_s dW_s.$$

It follows that for Ψ_r to be a martingale we need $\mu_\psi(t) = 0$ which implies the differential equations $\dot{\beta}(t) = 0$ and $\dot{\alpha}(t) = 1/2\lambda_\nu' (\partial\Omega_\nu(t)/\partial t)\lambda_\nu$. Using the terminal condition $\alpha(r) = 0$ it follows that

$$\alpha(r) - \alpha(0) = \int_0^r \dot{\alpha}(t)\, dt = \int_0^r \frac{1}{2}\lambda_\nu' \partial\Omega_\nu(t)/\partial t \lambda_\nu dt = \lambda_\nu' (\Omega_\nu(r) - \Omega_\nu(0))\lambda_\nu = \frac{1}{2}\lambda_\nu' \Omega_\nu(r)\lambda_\nu$$

or $\alpha(0) = -\frac{1}{2}\lambda_\nu' \Omega_\nu(r)\lambda_\nu$ and

$$E\left[\exp\left(i\int_0^r \lambda_\nu' \left(\dot{\Omega}_\nu(t)\right)^{1/2} dW_\nu(t)\right)\middle|\mathcal{C}\right] = \exp\left(-\frac{1}{2}\lambda_\nu' \Omega_\nu(r)\lambda_\nu\right) \text{ a.s.} \tag{52}$$

which implies (51) after taking expectations on both sides of (52). To check the regularity conditions in Duffie et al (2000, Definition A) note that $\gamma_t = 0$ because $\lambda(x) = 0$. Thus, (i) holds automatically. For (ii) we have $\eta_t = \Psi_t \beta(t)' \sigma(X_t)$ such that

$$\eta_t \eta_t' = -\Psi_t^2 \lambda_\nu' \partial\Omega_\nu(t)/\partial t \lambda_\nu = \exp\left(2\alpha(t) + 2\beta(t)' X_t\right) \lambda_\nu' \partial\Omega_\nu(t)/\partial t \lambda_\nu$$

and, noting that $\beta(t) = i\lambda_\nu$ and therefore $\left|\exp\left(2\beta(t)' X_t\right)\right| \leq 1$ it follows that

$$\begin{aligned}
|\eta_t \eta_t'| &\leq \sup_{\|\lambda_\nu\|=1, \lambda_\nu \in \mathbb{R}^{k_\rho}} \sup_t \lambda_\nu' \partial\Omega_\nu(t)/\partial t \lambda_\nu \left|\exp(2\alpha(t))\right| \left|\exp\left(2\beta(t)' X_t\right)\right| \\
&\leq \sup_{\|\lambda_\nu\|=1, \lambda_\nu \in \mathbb{R}^{k_\rho}} \sup_t \lambda_\nu' \partial\Omega_\nu(t)/\partial t \lambda_\nu \left|\exp(2\alpha(t))\right| \\
&\leq \sup_{\|\lambda_\nu\|=1, \lambda_\nu \in \mathbb{R}^{k_\rho}} \sup_t \lambda_\nu' \partial\Omega_\nu(t)/\partial t \lambda_\nu \left|\exp\left(2 \sup_{\|\lambda_\nu\|=1, \lambda_\nu \in \mathbb{R}^{k_\rho}} \sup_t \lambda_\nu' \partial\Omega_\nu(t)/\partial t \lambda_\nu\right)\right|
\end{aligned}$$

such that condition (ii) holds by Condition 2(iii) where $\sup_{\|\lambda_\nu\|=1, \lambda_\nu \in \mathbb{R}^{k_\rho}} \sup_t \lambda_\nu' \partial\Omega_\nu(t)/\partial t \lambda_\nu \leq M$ a.s. Finally, for (iii) one obtains similarly that

$$|\Psi_t| \leq |\exp(\alpha(t))| \left|\exp\left(\beta(t)' X_t\right)\right| \leq |\exp(\alpha(t))| \leq \exp(M)\, a.s.$$

such that the inequality follows.

A.2 Proof of Corollary 1

We note that finite dimensional convergence established in the proof of Theorem 1 implies that

$$E\left[\exp\left(i\lambda' X_{n\tau}(1)\right)\right] \to E\left[\exp\left(-\frac{1}{2}\left(\lambda_y' \Omega_y \lambda_y + \lambda_\nu' \Omega_\nu(1) \lambda_\nu\right)\right)\right].$$

We also note that because of (52) it follows that

$$E\left[\exp\left(i \int_0^1 \lambda_\nu' \left(\dot{\Omega}_\nu(t)\right)^{1/2} dW_\nu(t)\right)\right] = E\left[\exp\left(-\frac{1}{2} \lambda_\nu' \Omega_\nu(1) \lambda_\nu\right)\right]$$

which shows that $\int_0^1 \left(\dot{\Omega}_\nu(t)\right)^{1/2} dW_\nu(t)$ has the same distribution as $\Omega_\nu(1)^{1/2} W_\nu(1)$.

A.3 Proof of Theorem 2

Let $s_{it}^y(\theta,\rho) = f_{\theta,it}(\theta,\rho)$ and $s_t^\nu(\rho,\beta) = g_{\rho,t}(\rho,\beta)$ in the case of maximum likelihood estimation and $s_{it}^y(\theta,\rho) = f_{it}(\theta,\rho)$ and $s_t^\nu(\rho,\beta) = g_t(\rho,\beta)$ in the case of moment based estimation. Using the notation developed before we define

$$\tilde{s}_{it}^y(\theta,\rho) = \begin{cases} \frac{s_{it}^y(\theta,\rho)}{\sqrt{n}} & \text{if } t \in \{1,...,T\} \\ 0 & \text{otherwise} \end{cases}$$

analogously to (12) and

$$\tilde{s}^{\nu}_{it}(\beta,\rho)=\begin{cases}\frac{s^{\nu}_{t}(\beta,\rho)}{\sqrt{\tau}} & \text{if } t\in\{\tau_0+1,...,\tau_0+\tau\} \text{ and } i=1\\ 0 & \text{otherwise}\end{cases}$$

analogously to (11). Stack the moment vectors in

$$\tilde{s}_{it}(\phi):=\tilde{s}_{it}(\theta,\rho)=\left(\tilde{s}^{y}_{it}(\theta,\rho)',\tilde{s}^{\nu}_{it}(\beta,\rho)'\right)' \tag{53}$$

and define the scaling matrix $D_{n\tau}=\operatorname{diag}\left(n^{-1/2}I_y,\tau^{-1/2}I_\nu\right)$ where I_y is an identity matrix of dimension k_θ and I_ν is an identity matrix of dimension k_ρ. For the maximum likelihood estimator, the moment conditions (16) and (17) can be directly written as

$$\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\tilde{s}_{it}\left(\hat{\theta},\hat{\rho}\right)=0.$$

For moment based estimators we have by Conditions 4(i) and (ii) that

$$\sup_{\|\phi-\phi_0\|\le\varepsilon}\left\|\left(s^{y}_{M}(\theta,\rho)',s^{\nu}_{M}(\beta,\rho)'\right)'-\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\tilde{s}_{it}(\theta,\rho)\right\|=o_p(1).$$

It then follows that for the moment based estimators

$$0=s\left(\hat{\phi}\right)=\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\tilde{s}_{it}\left(\hat{\theta},\hat{\rho}\right)+o_p(1).$$

A first order mean value expansion around ϕ_0 where $\phi=(\theta',\rho')'$ and $\hat{\phi}=\left(\hat{\theta}',\hat{\rho}'\right)'$ leads to

$$o_p(1)=\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\tilde{s}_{it}(\phi_0)+\left(\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\frac{\partial\tilde{s}_{it}\left(\bar{\phi}\right)}{\partial\phi'}D_{n\tau}\right)D_{n\tau}^{-1}\left(\hat{\phi}-\phi_0\right)$$

or

$$D_{n\tau}^{-1}\left(\hat{\phi}-\phi_0\right)=-\left(\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\frac{\partial\tilde{s}_{it}\left(\bar{\phi}\right)}{\partial\phi'}D_{n\tau}\right)^{-1}\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\tilde{s}_{it}(\phi_0)+o_p(1)$$

where $\bar{\phi}$ satisfies $\left\|\bar{\phi}-\phi_0\right\|\le\left\|\hat{\phi}-\phi_0\right\|$ and we note that with some abuse of notation we implicitly allow for $\bar{\phi}$ to differ across rows of $\partial\tilde{s}_{it}\left(\bar{\phi}\right)/\partial\phi'$. Note that

$$\frac{\partial\tilde{s}_{it}\left(\bar{\phi}\right)}{\partial\phi'}=\begin{bmatrix}\partial\tilde{s}^{y}_{it}(\theta,\rho)/\partial\theta' & \partial\tilde{s}^{y}_{it}(\theta,\rho)/\partial\rho'\\ \partial\tilde{s}^{\nu}_{it,\rho}(\beta,\rho)/\partial\theta' & \partial\tilde{s}^{\nu}_{it,\rho}(\beta,\rho)/\partial\rho'\end{bmatrix}$$

where $\tilde{s}^{\nu}_{it,\rho}$ denotes moment conditions associated with ρ. From Condition 4(iii) and Theorem 1 it follows that (note that we make use of the continuous mapping theorem which is applicable because Theorem 1 establishes stable and thus joint convergence)

$$D_{n\tau}^{-1}\left(\hat{\phi}-\phi_0\right)=-A\left(\phi_0\right)^{-1}\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\tilde{s}_{it}\left(\phi_0\right)+o_p\left(1\right)$$

It now follows from the continuous mapping theorem and joint convergence in Corollary 1 that

$$D_{n\tau}^{-1}\left(\hat{\phi}-\phi_0\right)\xrightarrow{d}-A\left(\phi_0\right)^{-1}\Omega^{1/2}W\ \left(\mathcal{C}\text{-stably}\right)$$

A.4 Proof of Corollary 2

Partition

$$A\left(\phi_0\right)=\left[\begin{array}{cc}A_{y,\theta} & \sqrt{\kappa}A_{y,\rho}\\ \frac{1}{\sqrt{\kappa}}A_{\nu,\theta} & A_{\nu,\rho}\end{array}\right]$$

with inverse

$$\begin{aligned}A\left(\phi_0\right)^{-1}&=\left[\begin{array}{cc}A_{y,\theta}^{-1}+A_{y,\theta}^{-1}A_{y,\rho}\left(A_{\nu,\rho}-A_{\nu,\theta}A_{y,\theta}^{-1}A_{y,\rho}\right)^{-1}A_{\nu,\theta}A_{y,\theta}^{-1} & -\sqrt{\kappa}A_{y,\theta}^{-1}A_{y,\rho}\left(A_{\nu,\rho}-A_{\nu,\theta}A_{y,\theta}^{-1}A_{y,\rho}\right)^{-1}\\ -\frac{1}{\sqrt{\kappa}}\left(A_{\nu,\rho}-A_{\nu,\theta}A_{y,\theta}^{-1}A_{y,\rho}\right)^{-1}A_{\nu,\theta}A_{y,\theta}^{-1} & \left(A_{\nu,\rho}-A_{\nu,\theta}A_{y,\theta}^{-1}A_{y,\rho}\right)^{-1}\end{array}\right]\\ &=\left[\begin{array}{cc}A^{y,\theta} & \sqrt{\kappa}A^{y,\rho}\\ \frac{1}{\sqrt{\kappa}}A^{\nu,\theta} & A^{\nu,\rho}\end{array}\right].\end{aligned}$$

It now follows from the continuous mapping theorem and joint convergence in Corollary 1 that

$$D_{n\tau}^{-1}\left(\hat{\phi}-\phi_0\right)\xrightarrow{d}-A\left(\phi_0\right)^{-1}\Omega^{1/2}W\ \left(\mathcal{C}\text{-stably}\right)$$

where the right hand side has a mixed normal distribution,

$$A\left(\phi_0\right)^{-1}\Omega^{1/2}W\sim MN\left(0,A\left(\phi_0\right)^{-1}\Omega A\left(\phi_0\right)^{\prime-1}\right)$$

and

$$A\left(\phi_0\right)^{-1}\Omega A\left(\phi_0\right)^{\prime-1}=\left[\begin{array}{cc}A^{y,\theta}\Omega_y A^{y,\theta\prime}+\kappa A^{y,\rho}\Omega_\nu\left(1\right)A^{y,\rho\prime} & \frac{1}{\sqrt{\kappa}}A^{y,\theta}\Omega_y A^{\nu,\theta\prime}+\sqrt{\kappa}A^{y,\rho}\Omega_\nu\left(1\right)A^{\nu,\rho\prime}\\ \frac{1}{\sqrt{\kappa}}A^{\nu,\theta}\Omega_y A^{y,\theta\prime}+\sqrt{\kappa}A^{\nu,\rho}\Omega_\nu\left(1\right)A^{y,\rho\prime} & \frac{1}{\kappa}A^{\nu,\theta}\Omega_y A^{\nu,\theta\prime}+A^{\nu,\rho}\Omega_\nu\left(1\right)A^{\nu,\rho\prime}\end{array}\right]$$

The form of the matrices Ω_y and Ω_ν follow from Condition 5 in the case of the maximum likelihood estimator. For the moment based estimator, Ω_y and Ω_ν follow from Condition 6, the definition of $s^y_M\left(\theta,\rho\right)$ and $s^\nu_M\left(\beta,\rho\right)$ and Conditions 4(i) and (ii).

A.5 Proof of Theorem 3

We first establish the joint stable convergence of $(V_{\tau n}(r), Y_{\tau n})$. Recall that

$$\tau^{-1/2}\nu_{\tau,t} = \exp\left(\left(t - \min\left(1,\tau_0\right)\right)\gamma/\tau\right)\nu_0 + 1\left\{t > \min\left(1,\tau_0\right)\right\}\tau^{-1/2}\sum_{s=\min(1,\tau_0)}^{t}\exp\left(\left(t-s\right)\gamma/\tau\right)\eta_s$$

and $V_{\tau n}(r) = \tau^{-1/2}\nu_{\tau[\tau_0+\tau r]}$. Define $\tilde{V}_{\tau n}(r) = \tau^{-1/2}\sum_{s=\min(1,\tau_0)}^{[\tau r]}\exp\left(-s\gamma/\tau\right)\eta_s$. It follows that

$$\tau^{-1/2}\nu_{\tau[\tau r]} = \exp\left(\left(t - \min\left(1,\tau_0\right)\right)\gamma/\tau\right)\nu_0 + 1\left\{t > \min\left(1,\tau_0\right)\right\}\exp\left(\left[\tau r\right]\gamma/\tau\right)\tilde{V}_{\tau n}(r).$$

We establish joint stable convergence of $\left(\tilde{V}_{\tau n}(r), Y_{\tau n}\right)$ and use the continuous mapping theorem to deal with the first term in $\tau^{-1/2}\nu_{\tau[\tau r]}$. By the continuous mapping theorem (see Billingsley (1968, p.30)), the characterization of stable convergence on $D[0,1]$ (as given in JS, Theorem VIII 5.33(ii)) and an argument used in Kuersteiner and Prucha (2013, p.119), stable convergence of $\left(\tilde{V}_{\tau n}(r), Y_{\tau n}\right)$ implies that

$$\left(\exp\left(\left[\tau r\right]\gamma/\tau\right)\tilde{V}_{\tau n}(r), Y_{\tau n}\right)$$

also converges jointly and $\mathcal{C}$-stably. Subsequently, this argument will simply be referred to as the 'continuous mapping theorem'. In addition $\exp\left(\left(\left[\tau r\right] - \min\left(1,\tau_0\right)\right)\gamma/\tau\right)\nu_0 \to^p \exp\left(r\gamma\right)\nu_0$ which is measurable with respect to $\mathcal{C}$. Together these results imply joint stable convergence of $(V_{\tau n}(r), Y_{\tau n})$. We thus turn to $\left(\tilde{V}_{\tau n}(r), Y_{\tau n}\right)$. To apply Theorem 1 we need to show that $\psi_{\tau,s} = \exp\left(-s\gamma/\tau\right)\eta_s$ satisfies Conditions 1 iv) and 2. Since

$$\left|\exp\left(-s\gamma/\tau\right)\eta_s\right|^{2+\delta} = \left|\exp\left(-s/\tau\right)\right|^{\gamma(2+\delta)}\left|\eta_s\right|^{2+\delta} \leq e^{|\gamma|(2+\delta)}\left|\eta_s\right|^{2+\delta} \tag{54}$$

such that

$$E\left[\left|\exp\left(-s\gamma/\tau\right)\eta_s\right|^{2+\delta}\right] \leq C$$

and Condition 1 iv) holds. Note that $E\left[\left|\eta_t\right|^{2+\delta}\right] \leq C$ holds since we impose Condition 7. Next, note that $E\left[\exp\left(-2s\gamma/\tau\right)\eta_s^2\right] = \sigma^2\exp\left(-2s\gamma/\tau\right)$. Then, it follows from the proof of Chan and Wei (1987, Equation 2.3)[3] that

$$\tau^{-1}\sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]}\left(\psi_{\tau,s}\right)^2 = \tau^{-1}\sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]}\exp\left(-2\gamma t/\tau\right)\eta_t^2 \to^p \sigma^2\int_s^r \exp\left(-2\gamma s\right)dt. \tag{55}$$

[3] See Appendix B for details.

In this case, $\Omega_\nu(r) = \sigma^2(1-\exp(-2r\gamma))/2\gamma$ and $\left(\dot{\Omega}_\nu(r)\right)^{1/2} = \sigma\exp(-\gamma r)$. By the relationship in (51) and Theorem 1 we have that

$$\left(\tilde{V}_{\tau n}(r), Y_{\tau n}\right) \Rightarrow \left(\sigma\int_0^r e^{-s\gamma}dW_\nu(s), \Omega_y W_y(1)\right) \ \mathcal{C}\text{-stably}$$

which implies, by the continuous mapping theorem and $\mathcal{C}$-stable convergence that

$$(V_{\tau n}(r), Y_{\tau n}) \Rightarrow \left(\exp(r\gamma)\nu_0 + \sigma\int_0^r e^{(r-s)\gamma}dW_\nu(s), \Omega_y W_y(1)\right) \ \mathcal{C}\text{-stably.} \tag{56}$$

Note that $\sigma\int_0^r e^{(r-s)\gamma}dW_\nu(s)$ is the same term as in Phillips (1987) while the limit given in (56) is the same as in Kurtz and Protter (1991,p.1043).

We now square (20) and sum both sides as in Chan and Wei (1987, Equation (2.8) or Phillips, (1987) to write

$$\tau^{-1}\sum_{s=\tau_0+1}^{\tau+\tau_0}\nu_{\tau s-1}\eta_s = \frac{e^{-\gamma/\tau}}{2}\tau^{-1}\left(\nu^2_{\tau,\tau+\tau_0} - \nu^2_{\tau,\tau_0}\right) + \frac{\tau e^{-\gamma/\tau}}{2}\left(1-e^{2\gamma/\tau}\right)\tau^{-2}\sum_{s=\tau_0+1}^{\tau+\tau_0}\nu^2_{\tau s-1} - \frac{e^{-\gamma/\tau}}{2}\tau^{-1}\sum_{s=\tau_0+1}^{\tau+\tau_0}\eta_s^2. \tag{57}$$

We note that $e^{-\gamma/\tau}\to 1$, $\tau e^{-\gamma/\tau}\left(1-e^{2\gamma/\tau}\right)\to -2\gamma$. Furthermore, note that for all $\alpha,\varepsilon > 0$ it follows by the Markov and triangular inequalities and Condition 7iv) that

$$\begin{aligned} &P\left(\left|\tau^{-1}\sum_{t=\tau_0+1}^{\tau+\tau_0} E\left[\eta_s^2 1\left\{|\eta_t| > \tau^{1/2}\alpha\right\}|\mathcal{G}_{\tau n,t^*n}\right]\right| > \varepsilon\right) \\ &\leq \frac{1}{\tau\varepsilon}\sum_{t=\tau_0+1}^{\tau+\tau_0} E\left[\eta_s^2 1\left\{|\eta_t| > \tau^{1/2}\alpha\right\}\right] \leq \frac{\sup_t E\left[|\eta_t|^{2+\delta}\right]}{\alpha^\delta\tau^{\delta/2}} \to 0 \text{ as } \tau\to\infty. \end{aligned}$$

such that Condition 1.3 of Chan and Wei (1987) holds. Let $U^2_{\tau,k} = \tau^{-1}\sum_{t=\tau_0+1}^{k+\tau_0} E\left[\eta_s^2|\mathcal{G}_{\tau n,t^*n}\right]$. Then, by Holder's and Jensen's inequality

$$E\left[|U_{\tau,\tau}|^{2+\delta}\right] \leq \tau^{-1}\sum_{t=\tau_0+1}^{\tau+\tau_0} E\left[\left|E\left[\eta_s^2|\mathcal{G}_{\tau n,t^*n}\right]\right|^{1+\delta/2}\right] \leq \sup_t E\left[|\eta_t|^{2+\delta}\right] < \infty \tag{58}$$

such that $U^2_{\tau,\tau}$ is uniformly integrable. The bound in (58) also means that by Theorem 2.23 of Hall and Heyde it follows that $E\left[\left|U^2_{\tau,\tau} - \tau^{-1}\sum_{s=\tau_0+1}^{\tau+\tau_0}\eta_t^2\right|\right]\to 0$ and thus by Condition 7 vii) and by Markov's inequality

$$\tau^{-1}\sum_{s=\tau_0+1}^{\tau+\tau_0}\eta_t^2 \xrightarrow{p} \sigma^2.$$

We also have

$$\tau^{-1}\nu^2_{\tau,\tau+\tau_0} = V_{\tau n}(1)^2, \tag{59}$$

$$\tau^{-1}\nu^2_{\tau,\tau_0} \xrightarrow{p} V(0)^2$$

and

$$\tau^{-2}\sum_{s=\tau_0+1}^{\tau+\tau_0}\nu^2_{\tau s-1} = \tau^{-1}\sum_{s=1}^{\tau} V^2_{\tau n}\left(\frac{s}{\tau}\right) = \int_0^1 V^2_{\tau n}(r)\,dr$$

such that by the continuous mapping theorem and (56) it follows that

$$\tau^{-1}\sum_{s=\tau_0+1}^{\tau+\tau_0}\nu_{\tau s-1}\eta_s \Rightarrow \frac{1}{2}\left(V_{\gamma,V(0)}(1)^2 - V(0)^2\right) - \gamma\int_0^1 V_{\gamma,V(0)}(r)^2\,dr - \frac{\sigma^2}{2}. \tag{60}$$

An application of Ito's calculus to $V_{\gamma,V(0)}(r)^2/2$ shows that the RHS of (60) is equal to $\sigma\int_0^1 V_{\gamma,V(0)}dW_\nu$ which also appears in Kurtz and Protter (1991, Equation 3.10). However, note that the results in Kurtz and Protter (1991) do not establish stable convergence and thus don't directly apply here. When $V(0) = 0$ these expressions are the same as in Phillips (1987, Equation 8). It then is a further consequence of the continuous mapping theorem that

$$\left(V_{\tau n}(r), Y_{\tau n}(r), \tau^{-1}\sum_{s=\tau_0+1}^{\tau+\tau_0}\nu_{\tau s-1}\eta_s\right) \Rightarrow \left(V_{\gamma,V(0)}(r), Y(r), \sigma\int_0^1 V_{\gamma,V(0)}dW_\nu\right)\ (\mathcal{C}\text{-stably}).$$

A.6 Proof of Theorem 4

For $\tilde{s}_{it}(\phi) = \left(\tilde{s}^y_{it}(\theta,\rho)', \tilde{s}^\nu_{it,\rho}(\rho)\right)'$ we note that in the case of the unit root model

$$\frac{\partial\tilde{s}_{it}(\phi)}{\partial\phi'} = \begin{bmatrix} \partial\tilde{s}^y_{it}(\theta,\rho)/\partial\theta' & \partial\tilde{s}^y_{it}(\theta,\rho)/\partial\rho' \\ 0 & \partial\tilde{s}^\nu_{it,\rho}(\rho)/\partial\rho' \end{bmatrix}.$$

Defining

$$A^y_{\tau n}(\phi) = \left(\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\frac{\partial\tilde{s}^y_{it}(\phi)}{\partial\phi'}D_{n\tau}\right)$$

we have as before for some $\left\|\tilde{\phi}-\phi\right\| \le \left\|\hat{\phi}-\phi\right\|$ that for

$$A_{\tau n}(\phi) = \begin{bmatrix} & A^y_{\tau n}(\phi) \\ 0 & -\tau^{-2}\sum_{t=\tau_0}^{\tau_0+\tau}\nu^2_{\tau,t} \end{bmatrix},$$

we have

$$D_{n\tau}^{-1}\left(\hat{\phi}-\phi_0\right)=-A_{\tau n}\left(\tilde{\phi}\right)^{-1}\sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\tilde{s}_{it}\left(\phi_0\right)$$

Using the representation

$$\tau^{-2}\sum_{t=\tau_0}^{\tau_0+\tau}\nu_{\tau,t}^2=\int_0^1 V_{\tau n}\left(r\right)^2 dr,$$

it follows from the continuous mapping theorem and Theorem 3 that

$$\begin{aligned}&\left(V_{\tau n}\left(r\right),Y_{\tau n},A_{\tau n}^{y}\left(\phi_0\right),\int_0^1 V_{\tau n}\left(r\right)^2 dr,\tau^{-1}\sum_{s=\tau_0+1}^{\tau+\tau_0}\nu_{\tau s-1}\eta\right)\\ &\Rightarrow\left(V\left(r\right),\Omega_y^{1/2}W_y\left(1\right),A^{y}\left(\phi_0\right),\int_0^1 V_{\gamma,V(0)}\left(r\right)^2 dr,\int_0^s\sigma V_{\gamma,V(0)}dW_\nu\right)\ (\mathcal{C}\text{-stably}).\end{aligned}\tag{61}$$

The partitioned inverse formula implies that

$$A\left(\phi_0\right)^{-1}=\left[\begin{array}{cc}A_{y,\theta}^{-1} & A_{y,\theta}^{-1}A_{y,\rho}\left(\int_0^1 V_{\gamma,V(0)}\left(r\right)^2 dr\right)^{-1}\\ 0 & -\left(\int_0^1 V_{\gamma,V(0)}\left(r\right)^2 dr\right)^{-1}\end{array}\right]\tag{62}$$

By Condition 9, (61) and the continuous mapping theorem it follows that

$$D_{n\tau}^{-1}\left(\hat{\phi}-\phi_0\right)\Rightarrow-A\left(\phi_0\right)^{-1}\left[\begin{array}{c}\Omega_y^{1/2}W_y\left(1\right)\\ \int_0^s\sigma V_{\gamma,V(0)}dW_\nu\end{array}\right].\tag{63}$$

The result now follows immediately from (62) and (63).

B Proof of (55)

Lemma 1 *Assume that Conditions 7, 8 and 9 hold. For $r,s\in[0,1]$ fixed and as $\tau\to\infty$ it follows that*

$$\left|\tau^{-1}\sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]}\left(\left(\psi_{\tau,s}\right)^2-e^{(-2\gamma t/\tau)}E\left[\eta_t^2|\mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n}\right]\right)\right|\xrightarrow{p}0$$

Proof. By Hall and Heyde (1980, Theorem 2.23) we need to show that for all $\varepsilon>0$

$$\tau^{-1}\sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]}e^{-2\gamma t/\tau}E\left[\eta_t^2 1\left\{\left|\tau^{-1/2}e^{-\gamma t/\tau}\eta_t\right|>\varepsilon\right\}\middle|\mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n}\right]\xrightarrow{p}0.\tag{64}$$

By Condition 7iv) it follows that for some $\delta > 0$

$$E\left[\tau^{-1}\sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]} e^{-2\gamma t/\tau} E\left[\eta_t^2 1\left\{\left|\tau^{-1/2}e^{-\gamma t/\tau}\eta_t\right| > \varepsilon\right\}|\mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n}\right]\right]$$

$$\leq \tau^{-(1+\delta/2)}\sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]} \frac{\left(e^{-\gamma t/\tau}\right)^{2+\delta}}{\varepsilon^\delta} E\left[|\eta_t|^{2+\delta}\right]$$

$$\leq \sup_t E\left[|\eta_t|^{2+\delta}\right]\frac{[\tau r]-[\tau s]}{\tau^{1+\delta/2}\varepsilon^\delta}e^{(2+\delta)|\gamma|} \to 0.$$

This establishes (64) by the Markov inequality. Since $\tau^{-1}\sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]} e^{(-2\gamma t/\tau)}E\left[\eta_t^2|\mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n}\right]$ is uniformly integrable by (54) and (58) it follows from Hall and Heyde (1980, Theorem 2.23, Eq 2.28) that

$$E\left[\left|\tau^{-1}\sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]}\left((\psi_{\tau,s})^2 - e^{(-2\gamma t/\tau)}E\left[\eta_t^2|\mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n}\right]\right)\right|\right] \to 0.$$

The result now follows from the Markov inequality. ■

Lemma 2 *Assume that Conditions 7, 8 and 9 hold. For $r, s \in [0,1]$ fixed and as $\tau \to \infty$ it follows that*

$$\tau^{-1}\sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]} e^{(-2\gamma t/\tau)}E\left[\eta_t^2|\mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n}\right] \to^p \sigma^2\int_s^r \exp\left(-2\gamma t\right)dt.$$

Proof. The proof closely follows Chan and Wei (1987, p. 1060-1062) with a few necessary adjustments. Fix $\delta > 0$ and choose $s = t_0 \leq t_1 \leq ... \leq t_k = r$ such that

$$\max_{i\leq k}\left|e^{-2\gamma t_i} - e^{-2\gamma t_{i-1}}\right| < \delta.$$

This implies

$$\left|\int_s^r e^{-2\gamma t}dt - \sum_{i=1}^k e^{-2\gamma t_i}\left(t_i - t_{i-1}\right)\right| \leq \sum_{i=1}^k\int_{t_{i-1}}^{t_i}\left|e^{-2\gamma t} - e^{-2\gamma t_i}\right|dt \leq \delta. \quad (65)$$

Let $I_i = \{l : [\tau t_{i-1}] < l \leq [\tau t_i]\}$. Then,

$$
\begin{aligned}
&\tau^{-1} \sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]} e^{-2\gamma t/\tau} E\left[\eta_t^2 | \mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n}\right] - \sigma^2 \int_s^r e^{-2\gamma t} dt \\
&= \tau^{-1} \sum_{i=1}^{k} \sum_{l \in I_i} e^{-2\gamma l/\tau} E\left[\eta_l^2 | \mathcal{G}_{\tau n,(l-\min(1,\tau_0)-1)n}\right] - \sigma^2 \int_s^r e^{-2\gamma t} dt \\
&= \tau^{-1} \sum_{i=1}^{k} \sum_{l \in I_i} \left(e^{-2\gamma l/\tau} - e^{-2\gamma[\tau t_{i-1}]/\tau}\right) E\left[\eta_l^2 | \mathcal{G}_{\tau n,(l-\min(1,\tau_0)-1)n}\right] \\
&+ \sum_{i=1}^{k} e^{-2\gamma[\tau t_{i-1}]/\tau} \left(\tau^{-1} \sum_{l \in I_i} E\left[\eta_l^2 | \mathcal{G}_{\tau n,(l-\min(1,\tau_0)-1)n}\right] - \sigma^2 (t_i - t_{i-1})\right) \\
&+ \sum_{i=1}^{k} e^{-2\gamma[\tau t_{i-1}]/\tau} \sigma^2 (t_i - t_{i-1}) - \sigma^2 \int_s^r e^{-2\gamma t} dt \\
&= I_n + II_n + III_n.
\end{aligned}
$$

For III_n we have that $e^{-2\gamma[\tau t_{i-1}]/\tau} \to e^{-2\gamma t_{i-1}}$ as $\tau \to \infty$. In other words, there exists a τ' such that for all $\tau \geq \tau'$, $\left|e^{-2\gamma[\tau t_{i-1}]/\tau} - e^{-2\gamma t_{i-1}}\right| \leq \delta$ and by (65)

$$
|III_n| \leq 2\delta.
$$

We also have by Condition 7vii) that

$$
\tau^{-1} \sum_{l \in I_i} E\left[\eta_l^2 | \mathcal{G}_{\tau n,(l-\min(1,\tau_0)-1)n}\right] \to \sigma^2 (t_i - t_{i-1})
$$

as $\tau \to \infty$ such that by $\max_{i \leq k} \left|e^{2\gamma[\tau t_{i-1}]/\tau}\right| \leq e^{2|\gamma|}$

$$
|II_n| \leq e^{2|\gamma|} \left|\tau^{-1} \sum_{l \in I_i} E\left[\eta_l^2 | \mathcal{G}_{\tau n,(l-\min(1,\tau_0)-1)n}\right] - \sigma^2 (t_i - t_{i-1})\right| = o_p(1).
$$

Finally, there exists a τ' such that for all $\tau \geq \tau'$ it follows that

$$
\begin{aligned}
\max_{i \leq k} \max_{l \in I_i} \left|e^{-2\gamma l/\tau} - e^{-2\gamma[\tau t_{i-1}]/\tau}\right| &\leq \max_{i \leq k} \left|e^{-2\gamma[\tau t_i]/\tau} - e^{-2\gamma[\tau t_{i-1}]/\tau}\right| \\
&\leq 2 \max_{i \leq k} \left|e^{-2\gamma[\tau t_i]/\tau} - e^{-2\gamma t_i}\right| \\
&+ \max_{i \leq k} \left|e^{-2\gamma t_i} - e^{-2\gamma t_{i-1}}\right| \\
&\leq 2\delta + \delta = 3\delta.
\end{aligned}
$$

We conclude that

$$|I_n| \leq 3\delta \left| \tau^{-1} \sum_{i=1}^{k} \sum_{l \in I_i} E\left[\eta_l^2 | \mathcal{G}_{\tau n,(l-\min(1,\tau_0)-1)n} \right] \right| = 3\delta\sigma^2 \left(1 + o_p\left(1\right)\right).$$

The remainder of the proof is identical to Chan and Wei (1987, p. 1062). ■