Econometric Theory, **40**, 2024, 162–212.
doi:10.1017/S0266466622000391

CENTRAL LIMIT THEORY FOR COMBINED CROSS SECTION AND TIME SERIES WITH AN APPLICATION TO AGGREGATE PRODUCTIVITY SHOCKS

JINYONG HAHN
University of California, Los Angeles

GUIDO KUERSTEINER
University of Maryland

MAURIZIO MAZZOCCO
University of California, Los Angeles

Combining cross-sectional 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 datasets. 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 Murphy and Topel's (1985, *Journal of Business and Economic Statistics* 3, 370–379) formula.

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 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.

We thank Peter Phillips and referees in previous rounds for many helpful suggestions. Address correspondence to Guido Kuersteiner, Department of Economics, University of Maryland, Tydings Hall 3145, College Park, MD 20742, USA; e-mail: kuersteiner@econ.umd.edu.

Data combination generates theoretical challenges, even when only cross-sectional datasets or time-series datasets are combined. See Ridder and Moffitt (2007) for example. We focus on dependence between cross-sectional and time-series data stemming from 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. Andrews (2005) only considers cross sections and panels with fixed T with sufficient regularity conditions for consistent estimation. Our work in this paper and in our companion papers Hahn, Kuersteiner, and Mazzocco (2020, 2021) extends the analysis in Andrews (2005) to situations where consistent estimation is only possible by utilizing an additional time-series dataset. Examples in our companion papers draw on literatures for the structural estimation of rational expectations models as well as the program evaluation literature. Examples for the former include cross-sectional studies of the consumption-based asset pricing model in Runkle (1991), Shea (1995), and Vissing-Jorgensen (2002), and a recent example of the latter is Rosenzweig and Udry (2020).

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 (CLT) to derive the asymptotic distribution of parameter estimators obtained from combining the two data sources. Compared to our companion paper Hahn, Kuersteiner, and Mazzocco (2021; henceforth HKM21), in this paper, we impose a martingale difference structure as our fundamental moment condition on which our estimators are based. This implies that the theory in this paper is suitable for correctly specified maximum likelihood and conditional moment-based estimators. Our companion paper HKM21 does not impose a martingale structure for the time-series model, but rather allows for mixingale processes. Thus, certain forms of model misspecification can be handled by the theory of that paper. On the other hand, in this paper, we do not assume independence between time-series and cross-sectional data conditional on common factors, independence in the cross section conditional on common factors, or stationarity or homogeneity in the temporal direction of panel or time-series data, as is done in HKM21. Due to the strong assumptions in HKM21, the proof strategy for joint convergence is relatively simple in that paper. Conditionally on common factors, a CLT for independent cross sections combined with a stable CLT for stationary mixingales is sufficient to establish joint convergence of the time-series and cross-sectional components. In this paper where conditional independence of the cross-sectional sample is not assumed, a more complicated joint convergence argument establishing the joint limiting behavior of time-series and cross-sectional averages is required.

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

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

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 similar intuition and panel data.

To derive the new CLT and the asymptotic distribution of parameter estimates, we extend the model developed in Lee and Wolpin (2006, 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.

Another important literature that often uses micro data to calibrate model parameters is the literature on dynamic stochastic general equilibrium (DSGE) models in macro economics. The calibration approach was prominently advocated by Kydland and Prescott (1982), who also emphasize the importance of aggregate shocks to understand aggregate business cycle fluctuations. Other important contributions to the DSGE literature emphasizing the persistence of aggregate shocks are Long and Plosser (1983) and Smets and Wouters (2007), to name only two in a large literature. More recently, Schorfheide (2000) proposed a formal Bayesian approach that uses Bayesian priors to aid in the estimation of DSGE model parameters, and An and Schorfheide (2007) use micro data to inform the selection of prior distributions. To illustrate the contributions of this paper, we use the production function of Olley and Pakes (1996) as a micro foundation to estimate production function parameters from cross-sectional data. We then use the parameters estimated from the cross section to compute aggregate productivity shocks using time-series data. The estimated productivity shocks now form the basis for time-series estimates of the persistency parameter of aggregate shocks. The asymptotic theory developed in this paper provides the theoretical foundation to quantify the additional sampling uncertainty introduced by estimated, rather than observed aggregate shocks. The challenge specifically arises from the combination of two distinct, yet not independent data sources.

With the objective of creating a framework to perform inference in our general model, we first derive a joint functional stable CLT for cross-sectional and time-series data. The CLT 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 CLT 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 CLT, we then derive the asymptotic distribution of the parameter estimators that characterize our general model. To the best of our knowledge, this is the first paper that derives an asymptotic theory that combines cross-sectional and time-series data. In order to

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

deal with parameters estimated using two datasets of completely different nature, we adopt the notion of stable convergence. Stable convergence dates back to Rényi (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.

Stable convergence has found wide applications in many areas of statistics. Stable convergence was introduced into the econometrics literature by Phillips and Ouliaris (1990), who discuss the concept in detail and draw the connection to CLTs by McLeish (1975b) and Hall and Heyde (1980). The insight that panel limit theory often involves invariant sigma algebras generated by time series first appears in Phillips and Sul (2003) and is the basis for work by Andrews (2005) and Kuersteiner and Prucha (2013).

Another area in econometrics where stable convergence plays a prominent role is high-frequency financial models.[1] Important references to the high-frequency literature include Barndorff-Nielsen et al. (2008), Jacod, Podolskij, and Vetter (2010), and Li and Xiu (2016). Monographs treating the probability-theoretic foundation for this line of research are Jacod and Shiryaev (2002; henceforth JS) and Jacod and Protter (2012). In line with the literature on high-frequency financial econometrics, we base our definition of stable convergence in function spaces on JS. In the introduction to their book, JS outline two different strategies of proving CLTs. One is what they term the "martingale method," which dates back at least to Stroock and Varadhan (1979). The other is the approach followed by Billingsley (1968), which consists of establishing tightness, finite-dimensional convergence, and identification of the finite-dimensional limiting distribution. In this paper, we follow the second approach, whereas JS and the cited papers in the high-frequency literature rely on the martingale method. The reason for the difference in approach lies in the fact that the primitive parameters central to the martingale method, essentially a set of conditional moments called the triplets of characteristics of the approximating and limiting process in the language of JS, have no obvious analog in the models we study. The data generating processes (DGPs) producing our samples are not embedded in limiting diffusion processes. Even if our DGP could be represented by or approximated with such diffusions, the parameters of interest in our applications are not directly related to such approximations.

Our limit theory has a second connection to the high-frequency literature. Barndorff-Nielsen et al. (2008, Prop. 5) and Li and Xiu (2016, Lem. A3) develop methods to deduce joint stable convergence of two random sequences from stable convergence of the first random sequence and conditional convergence in law of the second random sequence. The assumptions made in these papers, namely that the noise terms are mean-zero conditional on the entire time series, are stronger than the assumptions we make in our paper. It is often unnatural to condition on the entire time series of common shocks, which we avoid in this paper.

[1] We thank one of the referees for bringing this literature to our attention.

While the formal derivation of the asymptotic distribution may appear complicated, the asymptotic formulas that we produce are straightforward to implement in some special cases and very similar to the standard formula of Murphy and Topel (1985).

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. However, 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 microeconomists 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, 2020; henceforth HKM20), we discuss in detail specific examples from the micro literature. In that paper, we focus on identification and provide an intuitive explanation of inference with combined cross-sectional and time-series data when aggregate factors are present. The purpose of this paper is to prove the asymptotic theory needed for inference rigorously.

The remainder of the paper is organized as follows. In Section 2, we present the Olley and Pakes model and illustrate how cross-sectional data can be used to help estimate the persistence of aggregate productivity shocks from time-series data. In Section 3, we introduce the general statistical model. Our main CLT is presented in Section 4. In Section 5, we discuss inference. Section 6 contains the analysis of the unit-root case. We collect proofs of our main results in Appendixes A–C and relegate additional derivations and results to the Supplementary Material.

2. AGGREGATE PRODUCTIVITY SHOCKS

In HKM20, we consider Olley and Pakes's (1996) method of estimating production functions, and show that a subset of parameters is consistently estimable from the cross section alone even in the presence of aggregate shocks. In this section, we argue that such a result is limited to only a subset of parameters. In general, access to both time-series and cross-sectional datasets is required to identify the full set of parameters, especially when the interest is in parameters that characterize the aggregate shock process.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

The general model introduced in Section 3 is flexible enough to both cover cases where the aggregate shocks are treated as nuisance quantities and cases where the aggregate shocks are treated as parameters or variables to be estimated. The example in this section can be understood to be the latter case. In our companion papers HKM20 and HKM21, the interested reader can find additional fully worked examples, some of them simpler and others more complex than the model discussed in this paper.

The essence of the example we discuss in this paper is as follows. The parameter of interest is the first-order autoregressive parameter $\alpha^{(A)}$ of an aggregate shock process ν_t. Aggregate shocks ν_t are unobserved but can be recovered from aggregate output Y_t^* and aggregate capital K_t^* through the relationship $\nu_t = Y_t^* - \beta_k K_t^*$. The parameter β_k is only identified from cross-sectional data. This approach is justified by the theory of Olley and Pakes (1996) and explained in more detail below. Inference for $\alpha^{(A)}$ then is complicated by two aspects: the estimator $\hat{\alpha}^{(A)}$ is based on estimated data $\hat{\nu}_t = Y_t^* - \hat{\beta}_k K_t^*$ and $\hat{\beta}_k$ is estimated on a cross-sectional dataset that is different from the aggregate data $\left(Y_t^*, K_t^*\right)$. Our paper provides the rigorous foundation for inference in this setting.

We now present a simplified version of Olley and Pakes's (1996) model. A profit-maximizing firm j produces a product $Y_{j,t}$ in period t, employing a production function that depends on the logarithm of labor $l_{j,t}$, the logarithm of capital $k_{j,t}$, and a productivity shock $\omega_{j,t}$. By denoting the logarithm of $Y_{j,t}$ by $\mathfrak{y}_{j,t}$, the production function takes the following form:

$$\mathfrak{y}_{j,t} = \beta_l l_{j,t} + \beta_k k_{j,t} + \omega_{j,t} + \eta_{j,t}, \tag{1}$$

where $\omega_{j,t}$ is a productivity shock and $\eta_{j,t}$ is a zero-mean measurement error with finite variance and i.i.d. over j and t. The intercept term is normalized to be zero. In each period, capital accumulates according to the equation $k_{j,t+1} = (1-\delta)\, k_{j,t} + i_{j,t}$, where δ is the rate at which capital depreciates. We abstract from age heterogeneity and exit decision. As in Olley and Pakes (1996), we assume that the optimal investment decision in period t is a function of the current stock of capital and of the productivity shock, i.e.,

$$i_{j,t} = i_t\left(\omega_{j,t}, k_{j,t}\right). \tag{2}$$

Olley and Pakes (1996) use the result that the investment decision (2) is strictly increasing in the productivity shock for every value of capital to invert (2), solve for the productivity shock, and obtain

$$\omega_{j,t} = h_t\left(i_{j,t}, k_{j,t}\right). \tag{3}$$

One can then replace the productivity shock in the production function using equation (3) to obtain

$$\mathfrak{y}_{j,t} = \beta_l l_{j,t} + \phi_t\left(i_{j,t}, k_{j,t}\right) + \eta_{j,t}, \tag{4}$$

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

where

$$\phi_t\left(i_{j,t},k_{j,t}\right)=\beta_k k_{j,t}+h_t\left(i_{j,t},k_{j,t}\right). \tag{5}$$

For simplicity, we assume that β_l and $\phi_t\left(i_{j,t},k_{j,t}\right)$ are known,[2] and work with $\mathfrak{y}^*_{j,t}\equiv\mathfrak{y}_{j,t}-\beta_l l_{j,t}$.

We now introduce an aggregate shock,[3] and assume that the productivity shock at t is the sum of an aggregate shock ν_t drawn from a distribution $F(\nu|\alpha)$ and of an idiosyncratic shock $\varepsilon_{j,t}$ independent of ν_t, i.e.,

$$\omega_{j,t}=\nu_t+\varepsilon_{j,t}. \tag{6}$$

Unlike HKM20, we assume that the firm observes $\omega_{j,t}$ but not ν_t and $\varepsilon_{j,t}$ separately.[4] We assume that ν_t and $\varepsilon_{j,t}$ are both Markov processes and in particular, that both ν_t and $\varepsilon_{j,t}$ are AR(1):

$$\nu_t=\alpha^{(A)}\nu_{t-1}+e_t^{(A)},$$
$$\varepsilon_{j,t}=\alpha^{(C)}\varepsilon_{j,t-1}+e_{j,t}^{(C)},$$

where the intercepts are zero for notational simplicity such that ν_t and $\varepsilon_{j,t}$ have mean zero.

It can be shown[5] that some of the parameters can be identified by using the cross-sectional generalized method of moments (GMM) estimator based on the moments

$$0=E\left[z_{j,t}\left(\mathfrak{y}^*_{j,t+1}-\left(\beta^*_{0,t+1}+\beta_k k_{j,t+1}+\alpha^{(C)}\left(\phi_t\left(i_{j,t},k_{j,t}\right)-\beta_k k_{j,t}\right)\right)\right)\right], \tag{7}$$

where $\beta^*_{0,t+1}\equiv\nu_{t+1}-\alpha^{(C)}\nu_t$ and the $z_{j,t}$ is an instrument uncorrelated with the error $e^{(C)}_{j,t+1}+\eta_{j,t+1}$. Note that identification of the parameters $\left(\beta^*_{0,t+1},\beta_k,\alpha^{(C)}\right)$ requires that the $z_{j,t}$ contain at least three components. The key cross-sectional parameter of interest for the application we have in mind is β_k, while the remaining

[2] Olley and Pakes (1996) identify the parameter β_l by

$$\beta_l=\frac{E\left[\left(l_{j,t}-E\left[l_{j,t}\middle|i_{j,t},k_{j,t}\right]\right)\left(\mathfrak{y}_{j,t}-E\left[\mathfrak{y}_{j,t}\middle|i_{j,t},k_{j,t}\right]\right)\right]}{E\left[\left(l_{j,t}-E\left[l_{j,t}\middle|i_{j,t},k_{j,t}\right]\right)^2\right]},$$

which can be consistently estimated by cross-sectional variation. In HKM20, it is shown that the cross-sectional variation identifies β_l even under the presence of aggregate shocks. We abstract away from the estimation of β_l because the above method of identification was critiqued for substantive economic reasons, for example, by Ackerberg, Caves, and Frazer (2015), and as such, researchers may prefer other methods of estimation.

[3] See Section I of the Supplementary Material for intuition of the moment condition for the simple case without any aggregate shock.

[4] In HKM20, we assume that ν_t and $\varepsilon_{j,t}$ are both Markov processes and that the firm observes the realization of the aggregate shock and, separately, of the idiosyncratic shock. This is an assumption of convenience to be consistent with Olley and Pakes's (1996) assumption that the problem solved by the firm is Markovian. To understand why, consider a case in which ν_t and $\varepsilon_{j,t}$ are both AR(1) processes. If we only use their sum as a state variable, the Markovian assumption is generally violated, because the sum of AR(1) processes is in general not an AR(1) but an ARMA(2,1) process. However, if we include ν_t and $\varepsilon_{j,t}$ as separate state variables—both observed by the firm—the Markovian structure is preserved.

[5] See Section I of the Supplementary Material for details that lead to (7).

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

cross-sectional parameters $\left(\beta^*_{0,t+1}, \alpha^{(C)}\right)$ are incidental to our final goal of estimating the degree of persistence of the aggregate shock.

The parameter $\alpha^{(A)}$ is not identified by the above procedure based on cross-sectional variation. On the other hand, $\alpha^{(A)}$ can be estimated consistently, possibly with the help of production function parameters estimated in the cross section, if aggregate time-series data with information about ν_t are available. For example, if an econometrician observes $\left\{\left(Y^*_t, K^*_t\right), t = 1, \ldots, \tau\right\}$, where τ is the time-series sample size and $Y^*_t \equiv \text{plim}_{n\to\infty} n^{-1} \sum_{j=1}^{n} \mathfrak{y}^*_{j,t}$ and $K^*_t \equiv \text{plim}_{n\to\infty} n^{-1} \sum_{j=1}^{n} k^*_{j,t}$, then equation (1) combined with (6) implies an aggregate relationship between output and capital of the form $Y^*_t - \beta_k K^*_t = \nu_t$. Implicit in this formulation is the assumption that firm-specific shocks $\varepsilon_{j,t}$ and $\eta_{j,t}$ average out in the aggregate. If ν_t is estimated by $\hat{\nu}_t = Y^*_t - \hat{\beta}_k K^*_t$, then the parameter $\alpha^{(A)}$ can be consistently estimated by an AR(1) regression of $Y^*_t - \hat{\beta}_k K^*_t$ on a constant and its own lagged value (using $\hat{\beta}_k$ estimated from the cross-sectional data) as long as τ is sufficiently large. The example illustrates how cross-sectional data and micro parameters recovered from it can be used to understand dynamic aggregate parameters.

The fact that $\alpha^{(A)}$ is estimated based on $\hat{\nu}_t$ rather than ν_t creates an estimated regressor problem that affects the limiting distribution of the estimator for $\alpha^{(A)}$. Unlike in classical estimated regressor problems, this paper considers the case where two samples that are not necessarily independent of each other are used to construct $\hat{\nu}_t$ and estimate $\alpha^{(A)}$.

3. MODEL AND PROBABILITY SPACE

In this section, we present a general modeling framework that includes the example in Section 2 as well as models considered in HKM20 and HKM21 as special cases. We assume that our cross-sectional data consist of $\left\{y_{i,t},\ i = 1, \ldots, n, t = 1, \ldots, T\right\}$,[6] 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. The need to keep T fixed is motivated by short cross-sectional panels and is critical for our theoretical development. Without this assumption, more complicated asymptotic approximations allowing for expanding parameter spaces as well as different limit theorems are required. The extensions are left for future work.

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 starting point of the time-series sample is either fixed at an arbitrary time τ_0 such that $-\infty < -K \leq \tau_0 \leq K < \infty$ for some bounded K and $\tau \to \infty$ or $\tau_0 = \tau_0(\tau)$ depends on τ such that $\tau_0(\tau) = -\upsilon\tau + \tau_{0,f} + T$ for $\upsilon \in [0,1]$ and $\tau_{0,f}$ a fixed constant. In the latter case, we use the shorthand notation τ_0 when no confusion arises. The fixed τ_0 scenario corresponds

[6]We do not consider models with estimated fixed effects in this paper because we assume that the parameter space is finite-dimensional.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

to a situation where a (hypothetical) time-series sample is observed into the infinite future. The specification $\tau_0(\tau)$ covers the case $\tau_0 = -\tau + T$. The case where τ_0 varies with the time-series sample size in the prescribed way can be used to model situations where the asymptotics are carried out "backward" in time (when $\upsilon = 1$) or where the panel data are located at a fixed fraction of the time-series sample as the time-series sample size tends to infinity $(0 < \upsilon < 1)$.When $\upsilon = 1$ such that $\tau_0 \to -\infty$, the end of the time-series sample is fixed at $\tau_{0,f} + T$. A hybrid case arises when $\upsilon \in (0,1)$ such that the starting point τ_0 of the time-series sample extends back in time simultaneously with the last observation in the sample $\tau_0 + \tau$ tending to infinity. For simplicity, we refer to both scenarios $\upsilon \in (0,1)$ and $\upsilon = 1$ as backward asymptotics. Backward asymptotics may be more realistic in cases where recent panel data are augmented with long historical records of time-series data. We show that under a mild additional regularity condition, both forward and backward asymptotics lead to the same limiting distribution. Conventional asymptotics are covered by setting $\upsilon = 0$. 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 parameterized models that implicitly restrict the data. The function $f\left(y_{i,t}\middle|\beta,\nu_t,\rho\right)$ is used to model $y_{i,t}$ as a function of cross-sectional parameters β and common shocks $\nu \equiv (\nu_1,\ldots,\nu_T)$ which are treated as parameters to be estimated, 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.[7]

Depending on the exact form of the underlying economic model, the functions f and g may have different interpretations. They could be the log-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 evaluated at the true parameter values. 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 $E_C\left[f\left(y_{i,t}\middle|\beta,\nu_t,\rho\right)\right] = 0$ given the shock ν_t and $E_\tau\left[g(z_s|\beta,\rho)\right] = 0$. The first expectation, E_C, is understood as being over the cross-sectional population distribution holding $\nu = (\nu_1,\ldots,\nu_T)$ fixed, whereas the second, E_τ, is over the distribution of the time-series DGP. The moment

[7]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 $\nu_t \equiv \nu_t(\beta,\rho)$ is a function of (z_t,β,ρ) 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 $g(z_s|\beta,\rho)$ here for now.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

conditions follow from martingale assumptions we directly impose on f and g. In our companion paper HKM20, we discuss examples of economic models that rationalize these assumptions.

Whether we are dealing with likelihoods or moment functions, the CLT is directly formulated for the estimating functions that define the parameters. We use the notation $F_n(\beta,\nu,\rho)$ and $G_\tau(\beta,\rho)$ to denote the criterion function based on the cross section and time series, respectively. When the model specifies a log-likelihood, these functions are defined as $F_n(\beta,\nu,\rho) = \frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n} f\left(y_{i,t}\middle|\beta,\nu_t,\rho\right)$ and $G_\tau(\beta,\rho) = \frac{1}{\tau}\sum_{s=\tau_0+1}^{\tau_0+\tau} g(z_s|\beta,\rho)$. When the model specifies moment conditions, we let $h_n(\beta,\nu,\rho) = \frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n} f\left(y_{i,t}\middle|\beta,\nu_t,\rho\right)$ and $k_\tau(\beta,\rho) = \frac{1}{\tau}\sum_{s=\tau_0+1}^{\tau_0+\tau} g(z_s|\beta,\rho)$. The GMM or moment-based criterion functions are then given by $F_n(\beta,\nu,\rho) = -h_n(\beta,\nu,\rho)' W_n^C h_n(\beta,\nu,\rho)$ and $G_\tau(\beta,\rho) = -k_\tau(\beta,\rho)' W_\tau^\tau k_\tau(\beta,\rho)$ with W_n^C and W_τ^τ 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.[8]

We formally justify the use of two datasets by imposing restrictions on the identifiability of parameters through the cross-sectional and time-series criterion functions alone. Let Θ be a compact set constructed from the product $\Theta = \Theta_\beta \times \Theta_\nu \times \Theta_\rho$, where Θ_β is a compact set that contains the true parameter value β_0, Θ_ν is a compact set that contains the true parameter value ν_0, and Θ_ρ is a compact set that contains the true parameter value ρ_0. 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

$$(\beta(\rho),\nu(\rho)) \equiv \operatorname*{argmax}_{\beta,\nu\in\Theta_\beta\times\Theta_\nu} F(\beta,\nu,\rho), \tag{8}$$

$$\rho(\beta) \equiv \operatorname*{argmax}_{\rho\in\Theta_\rho} G(\beta,\rho), \tag{9}$$

and we denote with β_0 and ρ_0 the solutions to (8) and (9). 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

$$\max_{\beta,\nu\in\Theta_\beta\times\Theta_\nu} F(\beta,\nu,\rho) = \max_{\beta,\nu\in\Theta_\beta\times\Theta_\nu} F(\beta,\nu,\rho_0) \quad \text{for all } \rho\in\Theta_\rho. \tag{10}$$

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

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

It is easy to see that (10) is not a sufficient condition to restrict identification in a desirable way. For example, (10) is satisfied in a setting where F does not depend at all on ρ. In that case, the maximizers in (8) 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

$$(\beta(\rho), \nu(\rho)) \neq (\beta_0, \nu_0) \quad \text{for all } \rho \neq \rho_0. \tag{11}$$

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

$$\max_{\rho \in \Theta_\rho} G(\beta, \rho) = \max_{\rho \in \Theta_\rho} G(\beta_0, \rho) \quad \text{for all } \beta \in \Theta_\beta, \tag{12}$$

$$\rho(\beta) \neq \rho_0 \quad \text{for all } \beta \neq \beta_0.$$

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

$$\left[\frac{\partial F(\beta, \nu, \rho)}{\partial \theta'}, \frac{\partial G(\beta, \rho)}{\partial \rho'}\right] = 0, \tag{13}$$

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 (10)–(12), and by the assumption that (13) only has one solution at the true parameter values.[9]

In order to accurately describe the theory that follows, we start with a precise definition of the probability space that we use. Let $(\Omega', \mathcal{G}', P')$ be a probability space with random sequences $\{z_t\}_{t=-\infty}^{\infty}$ and $\{y_{it}\}_{i=1,t=-\infty}^{\infty,\infty}$. The observed sample $\{z_t\}_{t=\tau_0+1}^{\tau_0+\tau}$ and $\{y_{i1}, \ldots, y_{iT}\}_{i=1}^{n}$ is a subset of these random sequences. The process we analyze consists of a triangular array of panel data $\psi_{n,it}^{y}$ where $\psi_{n,it}^{y}$ typically is a function of y_{it} and parameters, observed for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. We let $n \to \infty$ 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 $\psi_{\tau,s}^{\nu}$, where $\psi_{\tau,s}^{\nu}$ is a function of z_s and parameters for $s = \tau_0 + 1, \ldots, \tau_0 + \tau$. In typical applications, $\psi_{n,it}^{y}$ and $\psi_{\tau,s}^{\nu}$ are the influence functions of the cross-sectional and time-series estimators.

We now form the triangular array of filtrations similarly to Kuersteiner and Prucha (2013). The filtrations are a theoretical construct defined on the probability space in such a way that the observed sample is a strict subset of the random variables that generate the filtrations. We proceed by first collecting information about all of the common shocks, $(\nu_1, \ldots, \nu_T)$, then we pick the initial time-series

[9]Throughout this paper, we assume that the parameters are not identified using cross-sectional and time-series datasets alone. This seems to be the main reason for data combination. However, the proposed method of inference in this paper is effectively based on the moments in (13). This implies that our method of inference still works even if the parameters are identified from just one dataset.

realization $z_{\min\{1,\tau_0\}}$ in such a way that it predates or coincides with the initial period of the panel dataset. We then sequentially add the cross-sectional units for the same time period, starting from $i=1$ to $i=n$.[10] Subsequently, the same procedure is repeated by shifting the time index ahead by one period. The process ends once the time index reaches $\max(T,\tau)$. If $\tau > T$, which eventually happens in forward asymptotics, but may also arise in backward asymptotics, enlarge the filtration by adding random variables y_{it} even when $t > T$. These random variables are generated from the same distribution that produces the observed cross-sectional sample but are not actually included in the cross-sectional sample. This enlargement is used mostly for notational convenience.

Formally, the filtrations are defined as follows. We use the binary operator $\vee$ to denote the smallest σ-field that contains the union of two σ-fields. Setting $\mathcal{C} = \sigma(\nu_1, \ldots, \nu_T)$, we define

$$\begin{aligned}
\mathcal{G}_{\tau n,0} &= \mathcal{C}, \qquad (14)\\
\mathcal{G}_{\tau n,i} &= \sigma\left(z_{\min(1,\tau_0)}, \left\{y_{j,\min(1,\tau_0)}\right\}_{j=1}^{i}\right) \vee \mathcal{C},\\
&\vdots\\
\mathcal{G}_{\tau n,n+i} &= \sigma\left(\left\{y_{j,\min(1,\tau_0)}\right\}_{j=1}^{n}, \left\{z_{\min(1,\tau_0)+1}, z_{\min(1,\tau_0)}\right\}, \left\{y_{j,\min(1,\tau_0)+1}\right\}_{j=1}^{i}\right) \vee \mathcal{C},\\
&\vdots\\
\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n+i} &= \sigma\left(\left\{y_{j,t-1}, y_{j,t-2}, \ldots, y_{j,\min(1,\tau_0)}\right\}_{j=1}^{n}, \left\{z_t, z_{t-1}, \ldots, z_{\min(1,\tau_0)}\right\}, \left\{y_{j,t}\right\}_{j=1}^{i}\right) \vee \mathcal{C}.
\end{aligned}$$

As noted before, the filtration $\mathcal{G}_{\tau n,tn+i}$ is generated by a set of random variables that contain the observed sample as a strict subset. More specifically, we note that the cross section $y_{j,t}$ is only observed in the sample for a finite number of time periods, while the filtrations range over the entire expanding time-series sample period. 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. 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 index q orders the filtrations from smallest to largest. 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. However, they are not nested in the sense of Hall and Heyde's (1980) Condition (3.21) for two reasons. One is the fact that we are considering what essentially amounts to a panel structure generating the filtration. Kuersteiner and Prucha (2013) provide a detailed discussion of this aspect. A second reason has

[10]The filtrations are constructed based on the ordering of the cross-sectional sample. However, at the cost of slightly stronger moment conditions, the σ-fields can be constructed in a way that is invariant to reordering of the cross-sectional sample. We refer the interested reader to Kuersteiner and Prucha (2013), in particular Definition 1 and the related discussion for details. Note that the stronger moment conditions needed for the invariance property hold in situations where y_{it} is cross-sectionally independent conditional on $\{z_t, z_{t-1}, \ldots\} \vee \mathcal{C}$. The latter is a leading case in our examples.

to do with the possible “backward” asymptotics adopted in this paper. Even in a pure time-series setting, i.e., omitting $y_{i,t}$ from the definitions (14), backward asymptotics lead to a violation of Hall and Heyde’s Condition (3.21) because the definition of $\mathcal{G}_{\tau n,q}$ changes as q is held fixed but τ increases. The consequence of this is that unlike in Hall and Heyde (1980), stable convergence cannot be established for the entire probability space, but rather is limited to the invariant σ-field $\mathcal{C}$. This limitation is the same as in Kuersteiner and Prucha (2013) and also appears in Eagleson (1975), albeit due to very different technical reasons. The fact that the definition of $\mathcal{G}_{\tau n,q}$ changes for fixed q does not pose any problems for the proofs that follow, because the underlying proof strategy explicitly accounts for triangular arrays and does not use Hall and Heyde’s Condition (3.21).

To better understand the construction of $\mathcal{G}_{\tau n,q}$, we refer the reader to Kuersteiner and Prucha (2013, pp. 112–114) for a discussion of why the cross-sectional sample needs to be added to the filtration one at a time, why the ordering of the cross-sectional sample is irrelevant under certain regularity conditions, and why the nesting condition (3.21) of Hall and Heyde (1980) must fail in a panel context. Kuersteiner and Prucha (2013, Sect. 2.3) also provide a number of worked examples. The construction of $\mathcal{G}_{\tau n,q}$ proposed in this paper extends Kuersteiner and Prucha (2013) in two directions. On the one hand, an additional time-series component z_s is part of the generating mechanism for $\mathcal{G}_{\tau n,q}$. On the other hand, the filtration expands because two indices, τ and n, rather than just n in the case of Kuersteiner and Prucha (2013), tend to infinity. The construction of $\mathcal{G}_{\tau n,q}$ is specific to the type of CLT we prove and the fact that the joint process of $\psi^{y}_{n,it}$ and $\psi^{\nu}_{\tau,t}$ needs to satisfy a martingale difference property relative to the filtration $\mathcal{G}_{\tau n,q}$ for our proof to be valid. The fact that z_t and y_{1t} are added simultaneously to the filtration before the cross-sectional observations $y_{2t},\ldots,y_{nt}$ is necessitated by the possibility that z_t and y_{jt} are not independent. To understand this point, consider a hypothetical situation where z_t were added at the end of the cross-sectional sample together with y_{nt}. In such a scenario, it would no longer be credible to impose the moment condition $E\left[\psi^{\nu}_{\tau,t}|\mathcal{G}_{\tau n,q-1}\right]=0$ for $q=q_n(t,n)$ with $t>T$ because $\mathcal{G}_{\tau n,q-1}$ now would depend on $y_{1t},\ldots,y_{n-1t}$. These variables in turn may predict $\psi^{\nu}_{\tau,t}$. Similarly, the need to develop partial sums over the index i for the component $\psi^{y}_{n,it}$ requires that y_{it} be added one at a time to the filtration $\mathcal{G}_{\tau n,q}$, a point also explained in Kuersteiner and Prucha (2013).

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, in a later section, nonstationarity.

We develop asymptotic theory for the sums of some generic random vectors $\psi^{y}_{n,it}$ and $\psi^{\nu}_{\tau,t}$. Typically, $\psi^{y}_{n,it}$and $\psi^{\nu}_{\tau,t}$ are the scores or moment functions of a cross-sectional and time-series criterion function based on observed data y_{it} and z_t. Below, we introduce general regularity conditions for these generic random vectors

$\psi^y_{n,it}$ and $\psi^\nu_{\tau,t}$. Let k_θ be the dimension of the parameter θ, and let k_ρ be the dimension of the parameter ρ. With some abuse of notation, we also denote by k_θ the number of moment conditions used to identify θ when GMM estimators are used, with a similar convention applying to k_ρ. With this notation, $\psi^y_{n,it}$ takes values in R^{k_θ} and $\psi^\nu_{\tau,t}$ takes values in R^{k_ρ}. We assume that $T \le \tau_0 + \tau$. Throughout, we assume that $\left(\psi^y_{n,it}, \psi^\nu_{\tau,t}\right)$ is a type of a vector mixingale sequence relative to a filtration $\mathcal{G}_{\tau n,q}$. The concept of mixingales was introduced by Gordin (1969, 1973) and McLeish (1975a). We derive the joint limiting distribution and a related functional CLT for $\frac{1}{\sqrt{n}}\sum_{t=1}^{T}\sum_{i=1}^{n}\psi^y_{n,it}$ and $\frac{1}{\sqrt{\tau}}\sum_{t=\tau_0+1}^{\tau_0+\tau}\psi^\nu_{\tau,t}$.

The CLT 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

$$\tilde{\psi}^\nu_{it}(r) \equiv \frac{\psi^\nu_{\tau,t}}{\sqrt{\tau}} 1\{\tau_0 + 1 \le t \le \tau_0 + [\tau r]\} 1\{i = 1\}, \tag{15}$$

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

$$\tilde{\psi}^y_{it}(r) \equiv \tilde{\psi}^y_{it} \equiv \frac{\psi^y_{n,it}}{\sqrt{n}} 1\{1 \le t \le T\}, \tag{16}$$

where $\tilde{\psi}^y_{it}(r) = \tilde{\psi}^y_{it}$ is constant as a function of $r \in [0,1]$. In turn, this implies that functional convergence of the component (16) 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 sample averages

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

Define the stacked vector $\tilde{\psi}_{it}(r) = \left(\tilde{\psi}^y_{it}(r)', \tilde{\psi}^\nu_{it}(r)'\right)' \in \mathbb{R}^{k_\phi}$, where $\phi = (\theta, \rho)$ and k_ϕ is the dimension of ϕ. Consider the stochastic process

$$X_{n\tau}(r) = \sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\tilde{\psi}_{it}(r), \quad X_{n\tau}(0) = \left(X^{y\prime}_{n\tau}, 0\right)'. \tag{17}$$

We derive a functional CLT 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

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

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,

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

and we used the fact that $\sum_{i=1}^{n}\tilde{\psi}_{it}^{\nu}(1)=\tilde{\psi}_{1t}^{\nu}(1)=\frac{\psi_{\tau,t}^{\nu}}{\sqrt{\tau}}1\{\tau_0+1\leq t\leq\tau_0+\tau\}$ by (15). 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 CLTs for the time-series dimension of the panel dataset. In our case, since T is fixed and finite, a similar treatment is not applicable.

We introduce the following general regularity conditions for generic random vectors $\psi_{n,it}^{y}$ and $\psi_{\tau,t}^{\nu}$. Similarly, the CLT established in this section is for generic random vectors and empirical processes satisfying the regularity conditions. In later sections, these conditions will be specialized to the particular models considered there. To apply the general theory in this section to specific models, we evaluate the score or moment function at the true parameter value. In those instances, $\psi_{n,it}^{y}$ and $\psi_{\tau,t}^{\nu}$ will be interpreted as the score or moment function evaluated at the true parameter value. We use $\|.\|$ to denote the euclidean norm.

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* $i=1,\ldots,n$.
(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* $i=1,\ldots,n$.
(vii) $\left\|E\left[\psi_{\tau,t}^{\nu}\middle|\mathcal{G}_{\tau n,(t-\min(1,\tau_0)-1)n+i}\right]\right\|_2\leq\vartheta_t$, *for* $t<0$ *and all* $i=1,\ldots,n$, *where*

$$\vartheta_t\leq C\left(|t|^{1+\delta}\right)^{-1/2}$$

for the same δ *as in (iv) and some bounded constant* C *and where for a vector of random variables* $x=(x_1,\ldots,x_d)$ *and* $\|x\|_2=\left(\sum_{j=1}^{d}E\left[\left|x_j\right|^2\right]\right)^{1/2}$ *is the* L_2*-norm.*

Condition 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 Condition 1(v) to hold is that $E[\psi(y|\theta,\rho,\nu_t)|\mathcal{C}]=0$ holds in the population. This would be the case, for example, if ψ were the correctly specified score for the population distribution of $y_{i,t}$ given ν_t. See Andrews (2005, pp. 1573–1574).

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

Condition 1(i), (iii), and (v) imposes a martingale difference property for $\psi^y_{n,it}$ both in the time as well as in the cross-sectional dimensions. More specifically, $E\left[\psi^y_{n,it}|\psi^y_{n,i_1t},\psi^y_{n,i_2t_2},\ldots,\psi^y_{n,i_kt_k},\mathcal{C}\right]=0$ for any collection $(i_1,t),\ldots,(i_k,t_k)$ with $i_1<i,\ t_2,\ldots,t_k<t$ and $i_r\in\{1,\ldots,n\}$, for $2\le r\le k$. The CLT is established by letting the sums over the sample increase sequentially in line with the information contained in $\mathcal{G}_{\tau n,q}$ and thus mapping the sums over t and i into a single sum over an index set on the real line. This approach preserves the information structure in $\mathcal{G}_{\tau n,q}$ and in particular takes into account that in general $E\left[\psi^y_{n,it}|\psi^y_{n,js},\mathcal{C}\right]\neq 0$, for $s>t$ and all $j\in\{1,\ldots,n\}$. The score of a correctly specified likelihood is a leading example where the information structure takes the form implied by our conditions.

The construction in this paper is in contrast to some of the joint limit theory for panels that is based on martingale difference sequence (mds) assumptions in the time direction only and weak convergence of time-series averages of cross-sectional aggregates $\breve{\psi}^y_{n,t}=n^{-1/2}\sum_{i=1}^n\psi^y_{n,it}$, driven by T going to infinity. Our theory on the other hand depends on the cross-sectional sample size n tending to infinity, while T is kept fixed. The pure cross-sectional case is covered by allowing for $T=1$. Other examples include cases where the cross section is sampled at random conditional on $\mathcal{C}$ and $\psi^y_{n,it}$ is a conditional moment function in a rational agent model.

Condition 1(ii), (iv), and (vi) imposes a martingale property for $\psi^\nu_{\tau,t}$ in the time dimension. In addition, the condition also implies that $E\left[\psi^\nu_{\tau,t}|\psi^y_{n,is},\mathcal{C}\right]=0$ for any $i\in\{1,\ldots,n\}$ and $s<t$ and $t>T$. We note that this condition is weaker than assuming independence between $\psi^\nu_{\tau,t}$ and $\psi^y_{n,is}$, even conditionally on $\mathcal{C}$. While the examples in HKM20 do satisfy such a conditional independence restriction, it is not required for the CLT developed in this paper.

We note that $E\left[\psi^\nu_{\tau,t}\middle|\mathcal{G}_{\tau n,q_n(t-1,i)}\right]=0$, for all i, only holds for $t>T$ because we condition not only on $z_{t-1},z_{t-2},\ldots$ but also on $\nu_1,\ldots,\nu_T$, where the latter may have nontrivial overlap with the former. When τ_0 is fixed, the number of time periods t where $E\left[\psi^\nu_{\tau,t}\middle|\mathcal{G}_{\tau n,q_n(t-1,i)}\right]\neq 0$ is finite and thus can be neglected asymptotically. On the other hand, when τ_0 varies with τ, there is an asymptotically nonnegligible number of time periods where the condition may not hold. To handle this latter case, we impose an additional mixingale type condition that is satisfied for typical time-series models.

Condition 1(vii) is an additional restriction needed to handle situations where τ_0, the starting point of the time-series sample, is allowed to diverge to $-\infty$. We call this situation backward asymptotics. Since it is generally the case that $E\left[\psi^\nu_{\tau,t}|\mathcal{C}\right]\neq 0$ for $t\le T$ because $\mathcal{C}$ contains information about future realizations of $\psi^\nu_{\tau,t}$, we need a condition that limits this dependence as $t\to-\infty$. The following example illustrates that the condition naturally holds in linear time-series models.

Example 1. Assume that u_s is i.i.d. $N(0,1)$ and $z_s=\sum_{j=0}^\infty\rho^j u_{s-j}$ with $|\rho|<1$ is the stationary solution to $z_{s+1}=\rho z_s+u_{s+1}$. Use the convention that $\nu_s=z_s$. Then the score of the Gaussian likelihood is $\psi^\nu_{\tau,s}=z_s u_{s+1}$. Assuming that $\tau_0=-\tau+1, T=1$, and that z_s is independent of $y_{i,t}$ conditional on $\mathcal{C}=\sigma(\nu_1)$, it

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

is sufficient for this example to define $\mathcal{G}_{\tau n,(t-\min(1,\tau_0))n+i} = \sigma\left(\left\{z_t, z_{t-1}, \ldots, z_{\tau_0}\right\}\right) \vee \sigma(\nu_1)$. Then[11]

$$\left\| E\left[\psi^{\nu}_{\tau,s} \middle| \mathcal{G}_{\tau n,(s-\min(1,\tau_0)-1)n+i} \right] \right\|_2 = O\left(|\rho|^{|s|/2}\right) = o\left(|s|^{-(1+\delta)/2}\right).$$

The following conditions, Condition 2 for the time-series sample and Condition 3 for the cross-sectional sample, are put in place to ensure that, in combination with Condition 1, the variance of $X_{n\tau}(r)$ converges as n and τ tend to ∞ jointly. Conditions 2 and 3 correspond to Condition 3.19 in Hall and Heyde (1980, Thm. 3.2). We note in particular that the martingale structure in conjunction with uniform moment bounds imposed in Condition 1 is sufficient to guarantee that cross-covariance terms over different time periods converge to zero.

Condition 2. *Assume that:*
(i) For any $r \in [0,1]$,

$$\frac{1}{\tau} \sum_{t=\tau_0+1}^{\tau_0+[\tau r]} \psi^{\nu}_{\tau,t} \psi^{\nu\prime}_{\tau,t} \xrightarrow{p} \Omega_\nu(r), \text{ as } \tau \to \infty,$$

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

$\sup_{\|\lambda_\nu\|=1, \lambda_\nu \in \mathbb{R}^{k_\rho}} \sup_t \lambda_\nu' \dot{\Omega}_\nu(t) \lambda_\nu \leq M$ *a.s.*

Remark 1. Note that by construction $\Omega_\nu(0) = 0$.

Condition 2 is weaker than the conditions of Billingsley's (1968, Thm. 23.1) functional CLT for strictly stationary mds because we neither assume strict stationarity nor homoskedasticity. We do not assume that $E\left[\psi^{\nu}_{\tau,t}\psi^{\nu\prime}_{\tau,t}\right]$ 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 Gänssler and Häusler (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 (1968), but without assuming homoskedasticity, is pursued. Related results with heteroskedastic errors in the high-frequency and stochastic process literature can be found, for example, in Theorem IX.7.28 of JS.

Heteroskedastic errors are explicitly used in Section 6 where $\psi^{\nu}_{\tau,t} = \exp((t-s)\gamma/\tau)\eta_s$. Even if η_s is iid$\left(0,\sigma^2\right)$, it follows that $\psi^{\nu}_{\tau,t}$ is a heteroskedastic triangular

[11] Detailed derivations are in Section II of the Supplementary Material.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

array that depends on τ. It can be shown that the variance kernel $\Omega_\nu(r)$ is $\Omega_\nu(r) = \sigma^2(1-\exp(-2r\gamma))\big/2\gamma$ in this case. See equation (56).

CONDITION 3. *Assume that*

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

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

Condition 2 holds under a variety of conditions that imply some form of weak dependence of the process $\psi_{\tau,t}^{\nu}$. These include, in addition to Condition 1(ii) and (iv), mixing or near epoch dependence assumptions on the temporal dependence properties of the process $\psi_{\tau,t}^{\nu}$. Condition 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 JS develop for general Polish spaces. Let $(\Omega',\mathcal{G}',P')$ be a probability space with increasing filtrations $\mathcal{G}_{k_n,q}\subset\mathcal{G}'$ and $\mathcal{G}_{k_n,q}\subset\mathcal{G}_{k_n,q+1}$ for any $q=1,\ldots,k_n$ and an increasing sequence $k_n\to\infty$ as $n\to\infty$.[12] 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) with discontinuities synchronized across all elements in the vectors. Let $\mathcal{C}$ be a subsigma field of $\mathcal{G}'$. 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 $(\Omega',\mathcal{G}',P')$ and assume that ζ is bounded and measurable with respect to $\mathcal{C}$.

As in JS (p. 512), let $Z(\omega',x)=x$ be the canonical element on $D_{\mathbb{R}^{k_\theta}\times\mathbb{R}^{k_\rho}}[0,1]$ and let $Q(\omega',dx)$ be a version of the distribution of Z conditional on $\mathcal{C}$. Similarly, let $Q_n(\omega',dx)$ be a version of the conditional (on $\mathcal{C}$) distribution of Z^n. Following JS (Def. VI.1.1 and Thm. VI.1.14), we define the σ-field generated by all coordinate projections on $D_{\mathbb{R}^{k_\theta}\times\mathbb{R}^{k_\rho}}[0,1]$ as $\mathcal{D}_{\mathbb{R}^{k_\theta}\times\mathbb{R}^{k_\rho}}$. Then define the joint probability space $(\Omega,\mathcal{G},P)$ with $\Omega=\Omega'\times D_{\mathbb{R}^{k_\theta}\times\mathbb{R}^{k_\rho}}[0,1],\mathcal{G}=\mathcal{G}'\otimes\mathcal{D}_{\mathbb{R}^{k_\theta}\times\mathbb{R}^{k_\rho}}$, and

$$P(d\omega',dx)=P'(d\omega')\,Q(\omega',dx). \tag{18}$$

Following JS (p. 512, Def. 5.28), we say that Z^n converges $\mathcal{C}$-stably to Z if for all bounded, $\mathcal{C}$-measurable ζ and any continuous bounded functional $f:D_{\mathbb{R}^{k_\theta}\times\mathbb{R}^{k_\rho}}[0,1]\to\mathbb{R}$,

$$E\left[\zeta f(Z^n)\right]\to E[\zeta Q[f(Z)]], \tag{19}$$

[12]In our case, $k_n=\max(T,\tau)n$ where both $n\to\infty$ and $\tau\to\infty$ such that clearly $k_n\to\infty$.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

where $Q[f(Z)]$ is the expectation of $f(Z)$ conditional on $\mathcal{C}$. 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 (19) holds when Q is Wiener measure (for a definition and existence proof, see Billingsley (1968, Chap. 2, Sect. 9)). Our proof strategy is based on JS (Prop. VIII.5.33), which shows 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 Rényi (1963) and has found wide application in probability and statistics. Most relevant to the discussion here are the stable CLTs of Hall and Heyde (1980, Thm. 3.2) and Kuersteiner and Prucha (2013), who extend the result in Hall and Heyde (1980, Thm. 3.2) to panel data with fixed T. Related to our work, stable functional limit theorems were obtained previously for different settings by Rootzen (1983), Feigin (1985), and Dedecker and Merlevede (2002). For example, Dedecker and Merlevede (2002) established a related stable functional CLT for strictly stationary martingale differences, while we allow for heterogeneity and nonstationarity.

THEOREM 1. *Assume that Conditions 1–3 hold. Then it follows that for $\tilde{\psi}_{it}$ defined in (17), and as $\tau, n \to \infty$, and T fixed,*

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

where $B_y(r) = \Omega_y{}^{1/2} W_y(r), B_\nu(r) = \int_0^r \dot{\Omega}_\nu(s)^{1/2} dW_\nu(s)$, $\Omega(r) = \operatorname{diag}(\Omega_y, \Omega_\nu(r))$ 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, mutually independent, Brownian processes independent of Ω.

Proof. In Appendix A. □

Remark 2. Note that $W_y(r) = W_y(1)$ for each $r \in [0,1]$ by construction. Thus, $W_y(1)$ is simply a vector of standard Gaussian random variables, independent both of $W_\nu(r)$ and any random variable measurable with respect to $\mathcal{C}$.

The limiting random variables $B_y(r)$ and $B_\nu(r)$ both depend on $\mathcal{C}$ and thus are mutually dependent. However, conditional on $\mathcal{C}$, the limiting random variables are independent because of the mutual independence of $W_y(r)$ and $W_\nu(r)$. 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 nonstochastic 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 $\dot{\Omega}_\nu(s)$ is obtained by utilizing a technique mentioned in Rootzen (1983, p. 10) but not utilized there, namely establishing finite-dimensional convergence using a stable

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

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 isometry properties of the stochastic integral. Rootzen (1983, p. 13) similarly utilizes characteristic functions to identify the limiting distribution in the case of standard Brownian motion. Similar representations have been obtained in the high-frequency time-series literature (see Jacod et al., 2010; Jacod and Protter, 2012). 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 6.

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 datasets 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 dataset 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 CLTs.

The following corollary is useful for possibly nonlinear but trend stationary models.

Corollary 1. *Assume that Conditions 1–3 hold. Then it follows that for $\tilde{\psi}_{it}$ defined in (17), and as $\tau, n \to \infty$ and T fixed,*

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

where $\Omega = \operatorname{diag}\left(\Omega_y, \Omega_\nu(1)\right)$ is $\mathcal{C}$-measurable and $W = \left(W_y(1), W_\nu(1)\right)$ is a vector of standard k_ϕ-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 $\mathcal{C}$. 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, Cor. 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 6 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.[13] To see that sequential convergence where first n or τ go to infinity, followed by the other index, is generally not useful in our setup, consider the following example. Consider the double-indexed process

$$X_{n\tau}(1) = \sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)} \sum_{i=1}^{n} \tilde{\psi}_{it}(1). \tag{20}$$

For each τ fixed, convergence in distribution of $X_{n\tau}$ as $n \to \infty$ follows from the CLT 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 CLT 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 $\tau^{-1/2}\sum_{t=\tau_0+1}^{\tau_0+\tau}\psi^{\nu}_{\tau,t}$ trivially converges in distribution (it does not change with n), but the distribution of $\tau^{-1/2}\sum_{t=\tau_0+1}^{\tau_0+\tau}\psi^{\nu}_{\tau,t}$ 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

This section provides the theoretical foundation for the inference methods proposed in Section 6 of HKM20. Let $\theta = (\beta, \nu_1, \ldots, \nu_T)$ 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 $g_{\rho,t}(\beta,\rho) = \partial g_t(\beta,\rho)/\partial\rho$. Furthermore, 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 $(\psi^{y}_{it}, \psi^{\nu}_{\tau,t})$ 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 $\nu_t = \nu_t(\beta_0,\rho_0)$ 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 $s^{\nu}_{ML}(\beta,\rho) = \tau^{-1/2}\sum_{t=\tau_0+1}^{\tau_0+\tau}\partial g(\nu_t(\beta,\rho)\,|\,\nu_{t-1}(\beta,\rho),$

[13] The discussion assumes that $0 < \kappa < \infty$. The cases where $\kappa = 0$ or $\kappa = \infty$ allow for a simpler treatment where either the time-series or cross-sectional sample can be ignored. In those situations, considerations of joint convergence play only a minor role.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

$\beta,\rho)/\partial\rho$ 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

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

and $s_{M}^{y}(\theta,\rho) = -(\partial h_n(\theta,\rho)/\partial\theta)' W_n^C n^{-1/2}\sum_{t=1}^{T}\sum_{i=1}^{n} f(y_{it}|\theta,\rho)$ 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

$$s^{\nu}\left(\hat{\beta},\hat{\rho}\right) = 0 \tag{21}$$

and cross-sectional data

$$s^{y}\left(\hat{\theta},\hat{\rho}\right) = 0. \tag{22}$$

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 $\phi = \left(\theta',\rho'\right)' \in \mathbb{R}^{k_\phi}, \theta \in \mathbb{R}^{k_\theta}$, 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. Assume that, for some $\varepsilon > 0$,*

$$\sup_{\phi:\|\phi-\phi_0\|\le\varepsilon}\left\|\frac{\partial s(\phi)}{\partial\phi'}D_{n\tau} - A(\phi)\right\| = o_p(1),$$

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 $A_{y,\theta} = \operatorname{plim} n^{-1/2}\partial s^{y}(\phi_0)/\partial\theta'$, $A_{y,\rho} = \operatorname{plim} n^{-1/2}\partial s^{y}(\phi_0)/\partial\rho'$, $A_{\nu,\theta} = \operatorname{plim}\tau^{-1/2}\partial s^{\nu}(\phi_0)/\partial\theta'$, and $A_{\nu,\rho} = \operatorname{plim}\tau^{-1/2}\partial s^{\nu}(\phi_0)/\partial\rho'$.

CONDITION 5. *For maximum likelihood criteria, the following holds:*
(i) For any $r \in [0,1]$, $\frac{1}{\tau}\sum_{t=\tau_0+1}^{\tau_0+[\tau r]} g_{\rho,t}g_{\rho,t}' \xrightarrow{p} \Omega_\nu(r)$ as $\tau \to \infty$ and where $\Omega_\nu(r)$ satisfies the same regularity conditions as in Condition 2.
(ii) $\frac{1}{n}\sum_{i=1}^{n} f_{\theta,it}f_{\theta,it}' \xrightarrow{p} \Omega_{ty}$ for all $t \in [1,\ldots,T]$ and where Ω_{ty} is positive definite a.s. and measurable with respect to $\sigma(\nu_1,\ldots,\nu_T)$. Let $\Omega_y = \sum_{t=1}^{T}\Omega_{ty}$.

CONDITION 6. *Let* $W^C = \text{plim}_n W_n^C$ *and* $W^\tau = \text{plim}_\tau W_\tau^\tau$, *and assume the limits to be positive definite and* $\mathcal{C}$*-measurable. Define* $h(\theta,\rho) = \text{plim}_n h_n(\beta,\nu_t,\rho)$ *and* $k(\beta,\rho) = \text{plim}_\tau k_\tau(\beta,\rho)$. *For moment-based criteria, the following holds:*
(i) $\frac{1}{\tau}\sum_{t=\tau_0+1}^{\tau_0+[\tau r]} g_t g_t' \xrightarrow{p} \Omega_g(r)$ *as* $\tau \to \infty$ *and where* $\Omega_g(r)$ *satisfies the same regularity conditions as in Condition 2.*
(ii) $\frac{1}{n}\sum_{i=1}^{n} f_{it} f_{it}' \xrightarrow{p} \Omega_{t,f}$ *for all* $t \in [1,\ldots,T]$. *Let* $\Omega_f = \sum_{t=1}^{T}\Omega_{t,f}$. *Assume that* Ω_f *is positive definite a.s. and measurable with respect to* $\sigma(\nu_1,\ldots,\nu_T)$.
Assume that, for some $\varepsilon > 0$:
(iii) $\sup_{\phi:\|\phi-\phi_0\|\le\varepsilon} \left\| (\partial k_\tau(\beta,\rho)/\partial\rho)' W_\tau^\tau - \partial k(\beta,\rho)'/\partial\rho W^\tau \right\| = o_p(1)$.
(iv) $\sup_{\phi:\|\phi-\phi_0\|\le\varepsilon} \left\| (\partial h_n(\theta,\rho)/\partial\theta)' W_n^C - (\partial h(\theta,\rho)/\partial\theta)' W^C \right\| = o_p(1)$.

It is easy to see that the regularity conditions laid out in Conditions 1, 4, and 5 are satisfied if the requirements in Footnote 32 of HKM20 are imposed on the estimating functions f and g defined in that paper. The following result establishes the joint limiting distribution of $\hat{\phi}$.

THEOREM 2. *In the case of likelihood-based estimators, assume that Conditions 1, 4, and 5 hold with* $\left(\psi_{it}^y, \psi_{\tau,t}^\nu\right) = \left(f_{\theta,it}, g_{\rho,t}\right)$. *In the case of moment-based estimators, assume that Conditions 1, 4, and 6 hold with* $\left(\psi_{it}^y, \psi_{\tau,t}^\nu\right) = \left(\frac{\partial h(\theta_0,\rho_0)'}{\partial\theta} W^C f_{it}, \frac{\partial k(\beta_0,\rho_0)'}{\partial\rho} W^\tau g_t\right)$. *Assume that* $\hat{\phi} - \phi_0 = o_p(1)$ *and that (21) and (22) 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.*

Proof. In Appendix B. □

Remark 3. The distribution of $\Omega^{1/2}W$ is given in Corollary 1. In particular, $\Omega = \text{diag}\left(\Omega_y, \Omega_\nu(1)\right)$. When 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.

The theorem provides formulas for the joint asymptotic variance covariance matrix of $\hat{\phi}$ in two scenarios. The first scenario obtains when f and g are either the scores of a likelihood function, or if they are estimating functions in a just identified set of moment conditions. The second scenario covers GMM estimators in a scenario where f and g are moment functions in an overidentified set of moment conditions. The methods reported in Section 6 of HKM20 use an exactly identified moment-based approach. There may be cases where one wants to estimate the cross-sectional model using a likelihood approach and the time-series model using a moment approach, or vice versa. These cases can be handled as a special case

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

of the second scenario, where f or g is an exactly identified moment condition, whereas the other one may be an overidentified moment condition.

5. 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. In this section, we illustrate how the rigorous technical results of Section 4 can be applied to statistical inference problems for specific examples.

5.1. Intuition

For expositional purposes, suppose that the time series z_s is such that the logarithm of its conditional probability density function given z_{s-1} is $g(z_s|\beta,\rho)$. To simplify the exposition in this section, we assume that the cross-sectional model does not depend on the macro parameter ρ. We denote the consistent first-stage estimator of $\theta = (\beta, \nu_1, \ldots, \nu_T)$ by $\widetilde{\theta}$.[14]

We assume that the dimension of the cross-sectional data is n. 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 dataset, and adopt asymptotics where T is fixed. Then, assume that $\widetilde{\theta}$ is a regular estimator with influence function φ_{it} such that

$$\sqrt{n}\left(\widetilde{\theta}-\theta\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}\varphi_{it} + o_p(1) \qquad \textbf{(23)}$$

with $E[\varphi_{it}] = 0$. Using $\widetilde{\theta}$ from the cross-sectional data, we can then consider maximizing the criterion $G_\tau(\theta,\rho) = \frac{1}{\tau}\sum_{s=\tau_0+1}^{\tau_0+\tau} g(z_s|\theta,\rho)$ with respect to ρ. Here, τ_0+1 denotes the beginning of the time-series data, which is allowed to differ from the beginning of the panel data. The moment equation then is

$$\frac{\partial G_\tau\left(\tilde{\theta},\hat{\rho}\right)}{\partial\rho} = 0,$$

[14]In order to emphasize the fact that θ is estimated using only the cross-sectional data, we use the symbol $\widetilde{\theta}$. In more complicated models, θ needs to be estimated using both cross-sectional and time-series data, and we reserve the notation $\hat{\theta}$ for the general joint estimator.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

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

$$\sqrt{\tau}\left(\widehat{\rho}-\rho\right)=-\left(\frac{\partial^2 G\left(\tilde{\theta},\rho\right)}{\partial\rho\partial\rho'}\right)^{-1}\left(\sqrt{\tau}\frac{\partial G_\tau\left(\tilde{\theta},\rho\right)}{\partial\rho}\right)+o_p(1).$$

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

$$\sqrt{\tau}\left(\widehat{\rho}-\rho\right)=-A_{\nu,\rho}^{-1}\sqrt{\tau}\frac{\partial G_\tau(\theta,\rho)}{\partial\rho}-A_{\nu,\rho}^{-1}A_{\nu,\theta}\frac{\sqrt{\tau}}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}\varphi_{it}\right)+o_p(1) \tag{24}$$

with

$$A_{\nu,\rho}\equiv\frac{\partial^2 G(\theta,\rho)}{\partial\rho\partial\rho'},\quad A_{\nu,\theta}\equiv\frac{\partial^2 G(\theta,\rho)}{\partial\rho\partial\theta'}.$$

Because both $A_{\nu,\rho}$ and $A_{\nu,\theta}$ are $\mathcal{C}$-measurable random variables in the limit, the continuous mapping theorem can only be applied if joint convergence of $\sqrt{\tau}\partial G_\tau(\theta,\rho)/\partial\theta, n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\varphi_{it}$ and any $\mathcal{C}$-measurable random variable is established. Joint stable convergence of both components delivers exactly that. We also 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 (2020).

Assume that the unconditional distribution is such that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}\varphi_{it}\xrightarrow{d}MN\left(0,\Omega_y\right),$$

where Ω_y generally does depend on $(\nu_1,\ldots,\nu_T)$ through the parameter θ and as a result the distribution is mixed normal in general. Let us also assume that

$$\sqrt{\tau}\frac{\partial G_\tau(\theta,\rho)}{\partial\rho}\xrightarrow{d}N\left(0,\Omega_\nu\right),$$

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

We note that in general φ_{it} 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 $\sqrt{\tau}\partial G_\tau(\theta,\rho)/\partial\rho$ 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 $\sqrt{\tau}\partial G_\tau(\theta,\rho)/\partial\rho$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}\varphi_{it}$ 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

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

variables needs to be established formally. This follows from $\mathcal{C}$-stable convergence established in Section 4.2 and is summarized in the following corollary.

COROLLARY 2. *Under the same conditions as in Theorem 2, it follows that*

$$\sqrt{\tau}\left(\widehat{\rho}-\rho\right) \xrightarrow{d} -A_{\nu,\rho}^{-1}\Omega_{\nu}^{1/2}(1)\,W_{\nu}(1) - \frac{1}{\sqrt{\kappa}}A_{\nu,\rho}^{-1}A_{y,\theta}\Omega_{y}^{1/2}W_{y}(1) \ (\mathcal{C}\text{-stably}).$$

For

$$\Omega_{\rho} = A_{\nu,\rho}^{-1}\Omega_{\nu}A_{\nu,\rho}^{-1} + \frac{1}{\kappa}A_{\nu,\rho}^{-1}A_{\nu,\theta}\Omega_{y}A_{\nu,\theta}'A_{\nu,\rho}^{-1},$$

it follows that

$$\sqrt{\tau}\Omega_{\rho}^{-1/2}\left(\widehat{\rho}-\rho\right) \xrightarrow{d} N(0,I) \ (\mathcal{C}\text{-stably}).$$

Corollary 2 follows directly from Theorem 2. It implies that

$$\sqrt{\tau}\left(\widehat{\rho}-\rho\right) \xrightarrow{d} MN\left(0, A_{\nu,\rho}^{-1}\Omega_{\nu}A_{\nu,\rho}^{-1} + \frac{1}{\kappa}A_{\nu,\rho}^{-1}A_{\nu,\theta}\Omega_{y}A_{\nu,\theta}'A_{\nu,\rho}^{-1}\right), \tag{25}$$

where $0 < \kappa \equiv \lim n/\tau < \infty$ and where the limiting distribution on the RHS of (25) is a mixed Gaussian distribution. This means that a practitioner would use the square root of

$$\frac{1}{\tau}\left(A_{\nu,\rho}^{-1}\Omega_{\nu}A_{\nu,\rho}^{-1} + \frac{1}{\kappa}A_{\nu,\rho}^{-1}A_{\nu,\theta}\Omega_{y}A_{\nu,\theta}'A_{\nu,\rho}^{-1}\right) \approx \frac{1}{\tau}A_{\nu,\rho}^{-1}\Omega_{\nu}A_{\nu,\rho}^{-1} + \frac{1}{n}A_{\nu,\rho}^{-1}A_{\nu,\theta}\Omega_{y}A_{\nu,\theta}'A_{\nu,\rho}^{-1}$$

as the standard error when formulating a t-ratio. 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 assumption that $0 < \kappa < \infty$ is used as a technical device to obtain an asymptotic approximation that accounts for estimation errors stemming both from the cross-sectional and time-series samples. Simulation results in HKM20 for data and sample sizes calibrated to actual macro data show that our approximation provides good control for estimator bias and test size. The knife edge case $\kappa = \infty$ corresponds to situations where the estimation of cross-sectional parameters can be neglected for inference about $\widehat{\rho}$, and where now $\sqrt{\tau}\left(\widehat{\rho}-\rho\right) \xrightarrow{d} MN\left(0, A_{\nu,\rho}^{-1}\Omega_{\nu}A_{\nu,\rho}^{-1}\right)$. The expansion in (24) also shows that the case $\kappa = 0$ leads to a scenario where uncertainty from the cross section dominates such that the rate of convergence of $\hat{\rho}$ now is $\sqrt{n}$ rather than $\sqrt{\tau}$ and where $\sqrt{n}\left(\widehat{\rho}-\rho\right) \xrightarrow{d} MN\left(0, A_{\nu,\rho}^{-1}A_{\nu,\theta}\Omega_{y}A_{\nu,\theta}'A_{\nu,\rho}^{-1}\right)$. However, in what follows, we focus on the case most relevant in practice where $0 < \kappa < \infty$.

The asymptotic variance formula is such that the noise of the cross-sectional estimator $\widetilde{\theta}$ can make quite a difference if κ is small, i.e., if the cross-sectional size n is small relative to the time-series size τ. Obviously, this calls for larger cross sections for accurate estimation of the time-series parameter ρ. We also note

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

that cross-sectional estimation asymptotically has no impact on macro estimation if $A_{\nu,\theta}=0$. One scenario where $A_{\nu,\theta}=0$ is the case where the model is additively separable in θ and ρ such that $G(\theta,\rho)=G_1(\theta)+G_2(\rho)$.

For completeness, we also present a result that focuses on the limiting distribution of the subset of parameters that are associated with the cross-sectional model. Here, we no longer impose the restriction that the cross-sectional model does not depend on time-series parameters. Cross-sectional parameters are the main object of interest in HKM20, whereas in Section 2 of this paper, we consider a model where the main parameter of interest is a time-series parameter.

COROLLARY 3. *Under the same conditions as in Theorem 2, it follows that*

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right)\xrightarrow{d}-A^{y,\theta}\Omega_y^{1/2}W_y(1)-\sqrt{\kappa}A^{y,\rho}\Omega_\nu^{1/2}(1)\,W_\nu(1)\ (\mathcal{C}\text{-stably}),\tag{26}$$

where

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

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

For

$$\Omega_\theta=A^{y,\theta}\Omega_yA^{y,\theta\prime}+\kappa A^{y,\rho}\Omega_\nu(1)A^{y,\rho\prime},\tag{27}$$

it follows that

$$\sqrt{n}\Omega_\theta^{-1/2}\left(\hat{\theta}-\theta_0\right)\xrightarrow{d}N(0,I)\ (\mathcal{C}\text{-stably}).\tag{28}$$

The corollary develops the asymptotic distribution of the estimators for the general case where neither the time-series nor the cross-sectional parameters are identified separately. We note that the exposition in Section 6 of HKM20 does impose the additional restriction that $A_{\nu,\theta}=0$, which significantly simplifies (27). When $A_{\nu,\theta}=0$, the distributional approximation reported in HKM20 (Sect. 6, eqn. (35)) corresponds to the result obtained in (28).[15]

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 (28) 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 r a $q\times 1$

[15]We also note that the the Supplementary Material of HKM20 contains explicit formulas for Ω_θ for the general equilibrium model considered in that paper. Section VI of the Supplementary Material of this paper contains similar explicit formulas for a version of the Olley and Pakes's model considered here.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

vector. These insights form the basis for the standard errors proposed in Section 6 of HKM20.

We note that when two datasets are combined, the variance estimate of the estimators have to reflect the estimation error of the nuisance parameters. The formula above boils down to the usual variance formula of the two-step estimator. In fact, the formula is somewhat simpler than the usual two-step formula, at least in the stationary scenario. The reason is that the covariance of the moments of the two steps is zero when the moments are based on cross-sectional and time-series data, respectively. This is generally not the case for two-step procedures based only on one sample.

An additional theoretical difficulty that arises in this paper is common factors that remain random in the limit and affect the limiting variance Ω_θ. Theoretically, we handle this difficulty by relying on the concept of stable convergence to establish the limit distribution of our estimators. While inference based on pivotal statistics, such as the t-ratio, is not affected by stable limits, caution needs to be exercised when interpreting standard errors and confidence intervals for $\hat{\theta}$. The reason is that these quantities remain data-dependent through their dependence on common shocks even in the limit and may not be comparable across different empirical studies. This point is emphasized on page 1390 of HKM20.

5.2. A Worked Example

We now discuss how the model in Section 2 fits into our theory and discuss how to obtain valid standard errors. The estimator introduced in Section 2 is defined in terms of moment (not likelihood)-based criterion functions. Using the Taylor series expansion-based intuition, we discuss how the asymptotic distribution can be understood. Our discussion in this section parallels and complements the material in Section 6 of HKM20.

Unlike the model in Section 3, the moment (7) in Section 2 does not identify all the $\nu_1, \ldots, \nu_T$, and it only identifies $\left(\beta^*_{0,t+1}, \beta_k, \alpha^{(C)}\right)$, where we define $\beta^*_{0,t+1} \equiv \nu_{t+1} - \alpha^{(C)}\nu_t$. Therefore, it is convenient to define a finite-dimensional parameter that is identified from the cross section as $\theta(\nu)$, which may depend on the aggregate shocks $\nu = (\nu_1, \ldots, \nu_T)$ instead of working with (β, ν). For the model in Section 2, $\theta(\nu)$ is equal to $\left(\nu_{t+1} - \alpha^{(C)}\nu_t, \beta_k, \alpha^{(C)}\right)$. The parameter β in Section 5.1 denotes the collection of cross-sectional parameters that do not depend on ν. Since $\theta(\nu) = \left(\nu_{t+1} - \alpha^{(C)}\nu_t, \beta_k, \alpha^{(C)}\right)$ in Section 2, only the parameters $\left(\beta_k, \alpha^{(C)}\right)$ do not depend on ν. Thus, the $\left(\beta_k, \alpha^{(C)}\right)$ in Section 2 plays the role of β in Section 5.1.

We consider the following GMM estimation functions in the cross-sectional and time-series samples. Following the notational convention in Section 3, we define $h_n(\theta) = \frac{1}{n}\sum_{t=1}^{T}\sum_{j=1}^{n} f\left(y_{j,t}\middle|\theta\right)$ with $\theta = \theta(\nu)$ and $k_\tau(\beta, \rho) = \frac{1}{\tau}\sum_{s=\tau_0+1}^{\tau_0+\tau} g(z_s|\beta, \rho)$, where in Section 2 the parameters are $\theta = \left(\nu_{t+1} - \alpha^{(C)}\nu_t, \beta_k, \alpha^{(C)}\right)$

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

and $\rho = \alpha^{(A)}$. The reason why it is sufficient to focus on $\left(\beta_k, \alpha^{(C)}\right)$ is that the main interest lies in $\alpha^{(A)}$, which can be identified in the time-series sample with knowledge of β_k alone. Also note that for this model, $y_{j,t} = \left(i_{j,t}, k_{j,t}, l_{j,t}, \mathfrak{y}^*_{j,t}, i_{j,t-1}, k_{j,t-1}, l_{j,t-1}, \mathfrak{y}^*_{j,t-1}\right)$ and $z_s = \left(Y^*_s, K^*_s\right)$ is the vector of aggregate observed data.

The cross-sectional moment function $f\left(y_{i,t}|\theta\right)$ can be specified as

$$f\left(y_{j,t}|\theta\right) = \left(\mathfrak{y}^*_{j,t} - \left(\beta^*_{0,t} + \beta_k k_{j,t} + \alpha^{(C)}\left(\phi_t\left(i_{j,t-1}, k_{j,t-1}\right) - \beta_k k_{j,t-1}\right)\right)\right) z_{j,t},$$

where $z_{j,t}$ can be chosen as the vector $z_{j,t} = \left(1, k_{j,t-1}, i_{j,t-1}\right)'$, for example.

Similarly, specialize the generically defined function $g\left(z_s|\beta,\rho\right)$ for the aggregate time-series model to the score of the conditional pseudo-likelihood for the aggregate shock process, denoted by $g\left(\nu_s(\beta)|\nu_{s-1}(\beta),\beta,\rho\right) \equiv g\left(z_s|\beta,\rho\right)$, and where the aggregate shock $\nu_s(\beta_k) = Y^*_s - \beta_k K^*_s$ depends on $z_s = \left(Y^*_s, K^*_s\right)$ through the parameter β_k. When β_k is evaluated at the true parameter value $\beta_{k,0}$, we use the shorthand notation $\nu_s \equiv \nu_s\left(\beta_{k,0}\right)$. For the AR(1) model we postulate for ν_s, the function $g\left(\nu_s(\beta)|\nu_{s-1}(\beta),\beta,\rho\right)$ can be written explicitly as $g\left(\nu_s(\beta)|\nu_{s-1}(\beta),\beta,\rho\right) = \left(\nu_s(\beta) - \alpha^{(A)}\nu_{s-1}(\beta)\right)\nu_{s-1}(\beta)$.

Differentiating the counterparts of F_n and G_τ discussed in Section 3, we can see that the GMM estimator for ϕ solves the two moment conditions

$$s^y_M(\theta) = -\left(\partial h_n(\theta,\rho)/\partial\theta\right)' W^C_n n^{-1/2} \sum_{t=1}^{T}\sum_{i=1}^{n} f\left(y_{it}|\theta,\rho\right) = 0, \tag{29}$$

$$s^\nu_M(\beta,\rho) = -\left(\partial k_\tau(\beta,\rho)/\partial\rho\right)' W^\tau_\tau \tau^{-1/2} \sum_{t=\tau_0+1}^{\tau_0+\tau} g\left(\nu_t(\beta,\rho)|\nu_{t-1}(\beta,\rho),\beta,\rho\right) = 0.$$

Proceeding as in HKM20, let $J_{n\tau}(\phi) = \left[n^{-1/2}s^y_M(\theta), \tau^{-1/2}s^\nu_M(\beta,\rho)\right]$ and $D_{n\tau} = \text{diag}\left(n^{-1/2}I_f, \tau^{-1/2}\right)$, where I_f is an identity matrix equal to the dimension of θ. A Taylor series expansion of $J_{n\tau}(\phi)$ around ϕ_0 leads to

$$0 = D^{-1}_{n\tau} J_{n\tau}(\phi_0) + A D^{-1}_{n\tau}\left(\hat{\phi} - \phi_0\right) + o_p(1), \tag{30}$$

where $A = \text{plim}\left(D^{-1}_{n\tau}\partial J_{n\tau}(\phi)/\partial\phi' D_{n\tau}\right)$. The elements of the matrix A for the example are obtained as

$$A = \begin{bmatrix} A_{y,\theta} & 0 \\ \frac{1}{\sqrt{\kappa}} A_{\nu,\theta} & A_{\nu,\rho} \end{bmatrix},$$

where the upper right corner of A is zero because the cross-sectional model does not depend on the time-series parameters ρ. This feature of the model implies that a plug in estimator using the first-step cross-sectional estimate $\tilde{\beta}_k$ for the time-series problem estimating $\alpha^{(A)}$ is equivalent to an estimator $\hat{\phi}$ obtained jointly on the two

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

samples.[16] The nonzero elements of the matrix A are defined as

$$A_{y,\theta} = -\operatorname{plim} n^{-1/2} \frac{\partial s_M^y(\theta)}{\partial \theta'},\ A_{\nu,\rho} = -\operatorname{plim} \tau^{-1/2} \frac{\partial s_M^\nu(\theta)}{\partial \rho'},\ A_{\nu,\theta} = -\operatorname{plim} \tau^{-1/2} \frac{\partial s_M^\nu(\theta)}{\partial \theta'}.$$

Let $\mathcal{C} = \sigma(\nu_1, \ldots, \nu_T)$ be the sigma field generated by the aggregate shocks of the cross-sectional sample. It can be shown[17]

$$D_{n\tau}^{-1} J_{n\tau}(\phi_0) \to_d N(0, \Omega)\ \mathcal{C}\text{-stably},$$

where $\Omega = \operatorname{diag}(\Omega_y, \Omega_\nu)$ is the asymptotic variance covariance matrix of the moment functions defined in (29).

By the continuous mapping theorem and (30), the limiting distribution of $\hat{\phi}$ then is characterized by $D_{n\tau}^{-1}\left(\hat{\phi} - \phi_0\right) \to_d N(0, V)$ $\mathcal{C}$-stably where $V = A^{-1}\Omega\left(A'\right)^{-1}$. Suppose that the "conventional" weight matrices are chosen so that $\operatorname{plim} W_n^C = \Omega_f^{-1}$ and $\operatorname{plim} W_\tau^\tau = \Omega_g^{-1}$, where Ω_f and Ω_g denote the asymptotic variance of $\frac{1}{\sqrt{n}} \sum_{t=1}^T \sum_{j=1}^n f\left(y_{j,t} \middle| \theta_0\right)$ and $\frac{1}{\sqrt{\tau}} \sum_{s=\tau_0+1}^{\tau_0+\tau} g(z_s | \beta_0, \rho_0)$.[18] We would then have $\Omega = \operatorname{diag}\left(A_{y,\theta}, A_{\nu,\rho}\right)$. With straightforward algebra, it can be shown that

$$V = \begin{bmatrix} A_{y,\theta}^{-1} & 0 \\ 0 & A_{\nu,\rho}^{-1} + \frac{1}{\kappa} A_{\nu,\rho}^{-1} A_{\nu,\theta} A_{y,\theta}^{-1} A_{\nu,\theta}' A_{\nu,\rho}^{-1} \end{bmatrix},$$

which shows that the two sets of estimators are asymptotically independent in our example. The form of the limiting variance for ρ confirms the intuitive derivation in Section 5.1. Note in particular that $A_{y,\theta}^{-1} = \Omega_y$ when GMM with the optimal weight matrix is used. In general, V is a random variable measurable with respect to $\mathcal{C}$. A further application of the continuous mapping theorem shows that $V^{-1/2} D_{n\tau}^{-1}\left(\hat{\phi} - \phi_0\right)$converges to a standard Gaussian random vector.

Standard errors can now be computed based on this distributional approximation. To this end, use the following estimator $\hat{V}$ for the asymptotic variance–covariance matrix V. Let $\hat{\phi} = \left(\hat{\theta}, \hat{\rho}\right)$ be the joint solution to the moment conditions (29). Note that $\hat{\phi} = \left(\hat{\beta}_{0,1}^*, \ldots, \hat{\beta}_{0,T}^*, \hat{\beta}_k, \hat{\alpha}^{(C)}, \hat{\alpha}^{(A)}\right)$. Obtain the residuals $\hat{u}_{j,t} = \mathfrak{y}_{j,t}^* - \left(\hat{\beta}_{0,t}^* + \hat{\beta}_k k_{j,t} + \hat{\alpha}^{(C)}\left(\phi_t\left(i_{j,t-1}, k_{j,t-1}\right) - \hat{\beta}_k k_{j,t-1}\right)\right)$ as well as $\hat{\nu}_s = Y_s^* - \hat{\beta}_k K_s^*$ and $\hat{e}_s^{(A)} = \hat{\nu}_s - \hat{\alpha}^{(A)} \hat{\nu}_{s-1}$ and form the matrices

$$\hat{\Omega}_f = \frac{1}{n} \sum_{t=1}^T \sum_{j=1}^n \hat{u}_{j,t}^2 z_{j,t} z_{j,t}',\ \hat{\Omega}_g = \frac{1}{\tau} \sum_{s=\tau_0+1}^{\tau_0+\tau} \left(\hat{e}_s^{(A)}\right)^2 \hat{\nu}_{s-1}^2. \tag{31}$$

[16] In HKM20 and in Section 4.2, we show that these simplifications are not generic features of the problem we study and that joint estimation is needed except in special cases.

[17] See Appendix B.

[18] See (S.41) and (S.42) in Section VI of the Supplementary Material.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

Similarly, obtain

$$\frac{\partial \hat{k}(\beta,\rho)}{\partial \theta} = \frac{1}{\tau}\sum_{s=\tau_0+1}^{\tau_0+\tau} \frac{\partial g\left(z_s|\hat{\beta},\hat{\rho}\right)}{\partial \theta}, \quad \frac{\partial \hat{k}(\beta,\rho)}{\partial \rho} = \frac{1}{\tau}\sum_{s=\tau_0+1}^{\tau_0+\tau} \frac{\partial g\left(z_s|\hat{\beta},\hat{\rho}\right)}{\partial \rho}$$

$$\frac{\partial \hat{h}(\theta)}{\partial \theta} = \frac{1}{n}\sum_{t=1}^{T}\sum_{j=1}^{n} \frac{\partial f\left(y_{j,t}|\hat{\theta}\right)}{\partial \theta}$$

and

$$\hat{A}_{y,\theta} = \frac{\partial \hat{h}(\theta)'}{\partial \theta}\hat{\Omega}_f^{-1}\frac{\partial \hat{h}(\theta)}{\partial \theta'}, \quad \hat{A}_{\nu,\rho} = \frac{\partial \hat{k}(\beta,\rho)'}{\partial \rho}\hat{\Omega}_g^{-1}\frac{\partial \hat{k}(\beta,\rho)}{\partial \rho'},$$

$$\hat{A}_{\nu,\theta} = \frac{\partial \hat{k}(\beta,\rho)'}{\partial \rho}\hat{\Omega}_g^{-1}\frac{\partial \hat{h}(\theta)}{\partial \theta'}.$$

The asymptotic variance–covariance matrix then can be estimated as

$$\hat{V} = \begin{bmatrix} \hat{A}_{y,\theta}^{-1} & 0 \\ 0 & \hat{A}_{\nu,\rho}^{-1} + \frac{1}{\kappa}\hat{A}_{\nu,\rho}^{-1}\hat{A}_{\nu,\theta}\hat{A}_{y,\theta}^{-1}\hat{A}_{\nu,\theta}'\hat{A}_{\nu,\rho}^{-1} \end{bmatrix}.$$

Now let ϕ_j be the jth element of ϕ with estimator $\hat{\phi}_j$. Then, a t-ratio for $\hat{\phi}_j$ based on the asymptotic approximation for $\hat{\phi}$ can be constructed as $\hat{\phi}_j \Big/ \left(d_{j,j}\sqrt{\hat{V}_{j,j}}\right)$, where $\hat{V}_{j,j}$ is the jth diagonal element of $\hat{V}$ and $d_{j,j}$ is the jth diagonal element of $D_{n\tau}^{-1}$.

Focusing on the time-series parameter $\alpha^{(A)}$, one obtains the following t-ratio:

$$\hat{\alpha}^{(A)} \Big/ \sqrt{\frac{1}{\tau}\hat{A}_{\nu,\rho}^{-1} + \frac{1}{n}\hat{A}_{\nu,\rho}^{-1}\hat{A}_{\nu,\theta}\hat{A}_{y,\theta}^{-1}\hat{A}_{\nu,\theta}'\hat{A}_{\nu,\rho}^{-1}},$$

which corresponds to the asymptotic variance formula obtained in (25).

6. UNIT-ROOT TIME-SERIES MODELS

6.1. 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 6.2, 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-sectional and time-series samples manifest. The result is a cautionary tale of the difficulties that may present themselves when nonstationary time-series data

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

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 unity. Using intuition similar to (24), we obtain

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

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

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

where $\xi \equiv \lim \sqrt{n}\big/\tau$ and $W(\cdot)$ is the standard Wiener process. In contrast to the result in (28) when ρ is away from unity, $\sqrt{n}/\tau$ rather than n/τ is assumed to converge to a constant. Because $\tilde{\rho}$ is superconsistent under the unit-root scenario, from a theoretical point of view, the correction term is relevant only in cases where n is much larger than τ such that $\xi > 0$ in the limit. The result is formalized in Section 6.2.

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, Sect. 4.3).[19] First, construct the $1-\alpha_1$ confidence interval for ρ using Mikusheva (2007). Call it $[\rho_L, \rho_U]$. Second, compute $\widehat{\theta}(\rho) \equiv \text{argmax}_\theta F_n(\theta,\rho)$ 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 $1-\alpha_1-\alpha_2$.

There are some cases where standard asymptotics obtain for certain parameters in nonstationary scenarios (see, for example, Inoue and Kilian, 2020). We expect that such results will carry over to the case of joint cross-sectional and time-series inference, in which case the results in Section 4.2 could be applied. We leave the detailed technical analysis of these cases for future research.

[19] A rigorous proof of the validity of the proposed uniform inference procedure is beyond the scope of this paper and left for future research.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

6.2. Unit-Root Limit Theory

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) and 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

$$\nu_{\tau,t+1} = \exp\left(\gamma/\tau\right)\nu_{\tau,t} + \eta_{t+1} \tag{32}$$

and the notation $\nu_{\tau,t}$ emphasizes that $\nu_{\tau,t}$ is a sequence of processes indexed by τ. We assume that $\tau_0 = 0$ is fixed and

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

where $V(0)$ is a potentially nondegenerate random variable. In other words, the initial condition for (32) is $\nu_{\tau,\,\min(1,\tau_0)} = \tau^{1/2}V(0)$. We explicitly allow for the case where $V(0) = 0$, to model a situation where the initial condition can be ignored. This assumption is similar to, 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

$$g_{\rho,t}(\beta,\rho) = \nu_{\tau,t-1}\left(\nu_{\tau,t} - \nu_{\tau,t-1}\rho\right). \tag{33}$$

The estimator $\hat{\rho}$ solving sample moment conditions based on (33) is the conventional ordinary least squares 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 $\theta_0 = (\beta, V(0))$. We note that in this section, $V(0)$ 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 $\left(V_{\tau n}(r), s_{ML}^y\right)$ where $V_{\tau n}(r) \equiv \tau^{-1/2}\nu_{\tau[\tau r]}$, and

$$s_{ML}^y \equiv s_{ML}(\theta_0,\rho_0) \equiv \sum_{t=1}^{T}\sum_{i=1}^{n}\frac{f_{\theta,it}}{\sqrt{n}}.$$

Note that

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

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

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

$$V_{\gamma, V(0)}(r) = e^{\gamma r} V(0) + \int_0^r \sigma e^{\gamma(r-s)} dW_\nu(s), \tag{34}$$

where W_ν is defined in Theorem 1. When $V(0) = 0$, 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 (34) where $V(0)$ 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 $\left(V_{\tau n}(r), s_{ML}^y, \int_0^r V_{\tau n}(u) dW_{\tau n}(u)\right)$.

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 specific features of the unit-root model.

Condition 7. *Define $\mathcal{C} = \sigma(V(0))$. Define the σ-fields $\mathcal{G}_{n,(t-\min(1,\tau_0))n+i}$ in the same way as in (14) except that here $\tau = \kappa n$ such that dependence on τ is suppressed and that ν_t is replaced with η_t as in*

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

Assume that:

(i) *$f_{\theta,it}$ is measurable with respect to $\mathcal{G}_{n,(t-\min(1,\tau_0))n+i}$.*
(ii) *η_t is measurable with respect to $\mathcal{G}_{n,(t-\min(1,\tau_0))n+i}$, for all $i = 1,\ldots,n$.*
(iii) *For some $\delta > 0$ and $C < \infty$, $\sup_{it} E\left[\left\|f_{\theta,it}\right\|^{2+\delta}\right] \le C$.*
(iv) *For some $\delta > 0$ and $C < \infty$, $\sup_t E\left[\|\eta_t\|^{2+\delta}\right] \le C$.*
(v) *$E\left[f_{\theta,it} | \mathcal{G}_{n,(t-\min(1,\tau_0))n+i-1}\right] = 0$.*
(vi) *$E\left[\eta_t | \mathcal{G}_{n,(t-\min(1,\tau_0)-1)n+i}\right] = 0$, for $t > T$ and all $i = \{1,\ldots,n\}$.*
(vii) *For any $1 > r > s \ge 0$ fixed, let $\Omega_{\tau,\eta}^{r,s} = \tau^{-1} \sum_{t=\min(1,\tau_0)+[\tau s]+1}^{\tau_0+[\tau r]} E\left[\eta_t^2 | \mathcal{G}_{n,(t-\min(1,\tau_0))n}\right]$. Then, $\Omega_{\tau,\eta}^{r,s} \to_p (r-s)\sigma^2$.*
(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 $\Omega_y = \sum_{t=1}^T \Omega_{ty}$.*

Condition 7(i)–(vi) is the same as Condition 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 $\tau^{-1} \sum_{t=\min(1,\tau_0)+[\tau s]+1}^{\tau_0+[\tau r]} \eta_t^2 \to (r-s)\sigma^2$ 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.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

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 $\phi_0=\left(\beta_0',V_0(0),\rho_{\tau 0}\right)$. Thus, even if $\Omega_{ty}(.)$ is nonstochastic, it follows that Ω_{ty} is random and measurable with respect to $\mathcal{C}$ because it depends on $V(0)$, which is a random variable measurable with respect to $\mathcal{C}$.

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

THEOREM 3. *Assume that Condition 7 hold. With $\tau_0=0$ and as $\tau,n\to\infty$ and T fixed with $\tau=\kappa n$, for some $\kappa\in(0,\infty)$, it follows that*

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

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 $\kappa=\lim n/\tau^2$. Let $A_{y,\theta}(\phi)=\sum_{t=1}^T E\left[\partial f_{\theta,it}/\partial\theta'\right]$, $A_{y,\rho}(\phi)=\sum_{t=1}^T E\left[\partial f_{\theta,it}/\partial\rho\right]$, and define $A^y(\phi)=\left[\begin{array}{cc}A_{y,\theta}(\phi) & \sqrt{\kappa}A_{y,\rho}(\phi)\end{array}\right]$ where $A(\phi)$ is a $k_\theta\times k_\phi$-dimensional matrix of nonrandom functions $\phi\to\mathbb{R}$. Assume that $A_{y,\theta}(\phi_0)$ is full rank almost surely. Assume that, for some $\varepsilon>0$,*

$$\sup_{\phi:\|\phi-\phi_0\|\le\varepsilon}\left\|\frac{\partial\tilde{s}^y(\phi)}{\partial\phi'}D_{n\tau}-A^y(\phi)\right\|=o_p(1).$$

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

THEOREM 4. *Assume that Conditions 7–9 hold. It follows that*

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

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

Proof. In Appendix A. □

Note that the term $\left(\int_0^1 V^2_{\gamma,V(0)}(r)\,dr\right)^{-1}$ corresponds to $A^{-1}_{\nu,\rho}$ in the stationary case when the time-series model does not depend on cross-sectional parameters. The result in Theorem 4 is an example that shows how common factors affecting both time-series and cross-sectional data can lead to nonstandard 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 $V(0)$. Thus, the situation encountered here is generally more difficult than the one considered in Campbell 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.

7. SUMMARY

We develop a new limit theory for combined cross-sectional and time-series datasets. We focus on situations where the two datasets are interdependent because of common factors that affect both. The concept of stable convergence is used to handle this dependence when proving a joint CLT. Our analysis is cast in a generic framework of cross-sectional- 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 HKM20. The question of efficient inference in the context of our model is an interesting topic for future research, but beyond the scope of the current paper.

APPENDIX A: Proofs for the Stable Functional CLT

The proof of the functional CLT is given in Appendix A.2. Our proof is self-contained but follows the strategy of Billingsley (1968) for the case of conventional weak convergence adapted to our setting of stable convergence. The proof consists of three steps: (a) establishing finite-dimensional convergence through a stable CLT, (b) establishing tightness of the empirical process, and (c) providing a stochastic process representation of the limiting distribution. Tightness is established by extending techniques developed by Billingsley (1968) to our setting. Finally, the characterization of the limiting distribution is based on a proof strategy in Rootzen (1983), which we again adapt to our setting.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

A.1. Auxiliary Results

For ease of reference, we present two results that are used in the proof of Theorem 1. The first result is Theorem 1 of Kuersteiner and Prucha (2013).

THEOREM 5 (Kuersteiner and Prucha, 2013, Thm. 1). *Let* $\{S_{nq}, \mathcal{F}_{nq}, 1 \le q \le k_n, n \ge 1\}$ *be a zero-mean, square integrable martingale array with differences* X_{ni}. *Let* $\mathcal{F}_0 = \cap_{n=1}^{\infty} \mathcal{F}_{n0}$ *with* $\mathcal{F}_{n0} \subseteq \mathcal{F}_{n1}$ *for each n and* $E[X_{n1}|\mathcal{F}_{n0}] = 0$, *and let* η^2 *be an a.s. finite random variable measurable w.r.t.* $\mathcal{F}_0$. *If* $\max_q |X_{nq}| \xrightarrow{p} 0$, $\sum_{q=1}^{k_n} X_{nq}^2 \xrightarrow{p} \eta^2$, *and* $E\left(\max_q X_{nq}^2\right)$ *is bounded in n, then*

$$S_{nk_n} = \sum_{v=1}^{k_n} X_{nv} \xrightarrow{d} Z \; (\mathcal{F}_0\text{-stably}),$$

where the random variable Z has characteristic function $E\left[\exp\left(-\frac{1}{2}\eta^2 t^2\right)\right]$. *In particular,* $S_{nk_n} \xrightarrow{d} \eta\xi$ *(*$\mathcal{F}_0$*-stably), where* $\xi \sim N(0,1)$ *is independent of* η *(possibly after redefining all variables on an extended probability space).*

The second result is Theorem 8.3 of Billingsley (1968). To state the theorem, the following notation is needed (see Billingsley, 1968, pp. 19–20, 55). Let C be the space of continuous functions on $[0,1]$ with the uniform metric. Let $\mathcal{C}$ be the class of Borel sets in C. Let P_n be a sequence of probability measures on $(C, \mathcal{C})$.

THEOREM 6 (Billingsley, 1968, Thm. 8.3). *The sequence* $\{P_n\}$ *is tight if these two conditions are satisfied:*

(i) For each positive c, there exists an a such that $P_n(x : |x(0)| > a) \le c, n \ge 1$.

(ii) For each positive c and ε, *there is a* δ, *with* $0 < \delta < 1$ *and an integer* n_0 *such that*

$$\frac{1}{\delta} P_n\left(x : \sup_{t \le s \le t+\delta} |x(s) - x(t)| \ge c\right) \le \varepsilon, \; n \ge n_0,$$

for all t.

A.2. Proof of Theorem 1

We first establish that the stable functional CLT follows from establishing finite-dimensional convergence and tightness. To see this, note that JS (p. 512, Defn. 5.28) define stable convergence for sequences Z^n defined on a Polish space as in (19). We adopt the definition in JS to our setting, noting that by JS (p. 328, Thm. 1.14), $D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]$ equipped with the Skorohod topology is a Polish space. Following Billingsley (1968, p. 120), let $r_1 < r_2 < \cdots < r_k$ be an arbitrary finite partition of $[0,1]$ and $\pi_{r_1,\ldots,r_k} Z^n = \left(Z_{r_1}^n, \ldots, Z_{r_k}^n\right)$ be the coordinate projections of Z^n. By Proposition VIII.5.33(iv) of JS, $\mathcal{C}$-stable convergence of Z^n to Z is equivalent to the following two statements: (1) Z^n is tight and (2) for any bounded continuous function H with domain $D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]$ and for all $A \in \mathcal{C}$ and letting 1_A be the indicator function of the set A, it follows that $E\left[1_A H\left(Z^n\right)\right]$ converges. Part (1) is established by noting that by Billingsley (1968, Thm. 15.5), convergence under the uniform metric implies tightness for partial sum processes in $D_{\mathbb{R}^{k_\theta} \times \mathbb{R}^{k_\rho}}[0,1]$ that have

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

stochastically bounded initial conditions (see Billingsley, 1968, Cond. 15.17, which is satisfied in our case). Part (2) is established as follows. First, prove stable convergence of the finite-dimensional vector of random variables $Z^n_{r_1},\dots,Z^n_{r_k}$ defined on $\mathbb{R}^k$ using a multivariate stable CLT. This is shown in Step (a) of the proof below. Second, use an argument based on the proof of Theorems VIII5.7 and VIII5.14 of JS and JS (p. 509) as follows: For P defined in (18), define the new probability measure $\tilde{P}(d\omega) = P(d\omega)\,1_A(\omega)$ such that $E\left[1_A H\left(Z^n\right)\right] = \tilde{E}\left[H\left(Z^n\right)\right]$, where $\tilde{E}$ is the expectation with respect to $\tilde{P}$. Then, by the stable coordinatewise CLT, the finite-dimensional distributions of Z^n converge under $\tilde{P}$. Furthermore, for any compact subset K of $D_{\mathbb{R}^{k_\theta}\times\mathbb{R}^{k_\rho}}[0,1]$, it follows that $\tilde{P}\left(Z^n\in K\right)\le P\left(Z^n\in K\right)$ such that it is sufficient to show that Z^n is tight under the measure P. This shows that a tightness argument following Billingsley (1968, Thm. 8.3) is sufficient to establish Part (2) above. This is done in Step (b) of the proof below, where we show that Billingsley (1968, Thm. 8.3(ii)) holds.

We now proceed by first establishing finite-dimensional stable convergence in Step (a) and tightness in Step (b). Finally, in Step (c), we give a stochastic process representation for the limiting distribution.

(a) Finite-Dimensional Convergence. For finite-dimensional convergence, fix $r_1 < r_2 < \cdots < r_k \in [0,1]$. We use the notational convention $r_0 = 0$ below. Define the increment

$$\Delta X_{n\tau}(r_i) = X_{n\tau}(r_i) - X_{n\tau}\left(r_{i-1}\right). \tag{35}$$

Since there is a one-to-one mapping between $X_{n\tau}(r_1),\dots,X_{n\tau}(r_k)$ and $X_{n\tau}(r_1)$, $\Delta X_{n\tau}(r_2),\dots,\Delta X_{n\tau}(r_k)$, we establish joint convergence of the latter. The proof proceeds by checking that the conditions of Theorem 1 in Kuersteiner and Prucha (2013) hold. For the convenience of the reader, Kuersteiner and Prucha (2013, Thm. 1) is stated in Appendix A.1. 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 $d = k_\phi = k_\theta + k_\rho$. 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 $\lambda = (\lambda_1,\dots,\lambda_k)\in\mathbb{R}^{dk}$ 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 $q(i,t)\equiv t^*n+i$ and note that $q(i,t)$ is invertible, in particular for each $q\in\{1,\dots,k_n\}$, there is a unique pair t,i such that $q(i,t) = q$. We often use shorthand notation q for $q(i,t)$. Let

$$\ddot{\psi}_{q(i,t)} \equiv \sum_{j=1}^{k}\lambda'_j\left(\Delta\tilde{\psi}_{it}\left(r_j\right) - E\left[\Delta\tilde{\psi}_{it}\left(r_j\right)|\mathcal{G}_{\tau n,t^*n+i-1}\right]\right), \tag{36}$$

where

$$\Delta\tilde{\psi}_{it}\left(r_j\right) = \tilde{\psi}_{it}\left(r_j\right) - \tilde{\psi}_{it}\left(r_{j-1}\right);\ \Delta\tilde{\psi}_{it}(r_1) = \tilde{\psi}_{it}(r_1). \tag{37}$$

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

$$\Delta\tilde{\psi}^y_{it}\left(r_j\right) = \begin{cases}\tilde{\psi}^y_{it}, & \text{for } j=1,\\ 0, & \text{otherwise,}\end{cases} \tag{38}$$

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

and

$$\Delta\tilde{\psi}_{it}^{\nu}\left(r_j\right) = \begin{cases} \tilde{\psi}_{\tau,t}^{\nu}\left(r_j\right), & \text{if } \left[\tau r_{j-1}\right] < t-\tau_0 \leq \left[\tau r_j\right] \text{ and } i=1, \\ 0, & \text{otherwise.} \end{cases} \tag{39}$$

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

$$\begin{aligned} &\lambda_1' X_{n\tau}\left(r_1\right) + \sum_{j=2}^{k}\lambda_j'\Delta X_{n\tau}\left(r_j\right) \\ &= \sum_{q=1}^{k_n}\ddot{\psi}_q + \sum_{t=\min(1,\tau_0+1)}^{\max(T,\tau_0+\tau)}\sum_{i=1}^{n}\sum_{j=1}^{k}\lambda_j' E\left[\Delta\tilde{\psi}_{it}\left(r_j\right)\middle|\mathcal{G}_{\tau n,t^*n+i-1}\right]. \end{aligned} \tag{40}$$

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 (36), (38), and (39), the individual components of $\ddot{\psi}_q$ are either 0 or equal to $\tilde{\psi}_{it}\left(r_j\right) - E\left[\tilde{\psi}_{it}\left(r_j\right)|\mathcal{G}_{\tau n,t^*n+i-1}\right]$, respectively. This implies that $\ddot{\psi}_q$ is measurable with respect to $\mathcal{G}_{\tau n,q}$, noting in particular that $E\left[\tilde{\psi}_{it}\left(r_j\right)|\mathcal{G}_{\tau n,t^*n+i-1}\right]$ 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, $E\left[\ddot{\psi}_q|\mathcal{G}_{\tau n,q-1}\right] = 0$. This establishes that for $S_{nq} = \sum_{s=1}^{q}\ddot{\psi}_s$,

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

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. To establish the limiting distribution of $\sum_{q=1}^{k_n}\ddot{\psi}_q$, we check that

$$\sum_{q=1}^{k_n} E\left[\left|\ddot{\psi}_q\right|^{2+\delta}\right] \to 0, \tag{41}$$

$$\sum_{q=1}^{k_n}\ddot{\psi}_q^2 \xrightarrow{p} \sum_{t\in\{1,\ldots,T\}}\lambda_{1,y}'\Omega_{yt}\lambda_{1,y} + \sum_{j=1}^{k}\lambda_{j,\nu}'\Omega_{\nu}\left(r_j - r_{j-1}\right)\lambda_{j,\nu}, \tag{42}$$

and

$$\sup_n E\left[\left(\sum_{q=1}^{k_n} E\left[\ddot{\psi}_q^2\middle|\mathcal{G}_{\tau n,q-1}\right]\right)^{1+\delta/2}\right] < \infty, \tag{43}$$

which are adapted to the current setting from Conditions (A.26)–(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.) We check these conditions in Sections III.1–III.3 of the Supplementary Material, which establishes that (41)–(43) hold and thus establishes the CLT for $\sum_{q=1}^{k_n}\ddot{\psi}_q$. In Section III.4 of the Supplementary Material, we also show that the second term in (40) can be neglected.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

We therefore have

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

We have shown that the conditions of Theorem 1 of Kuersteiner and Prucha (2013) hold by establishing (41)–(44). Applying the Cramer–Wold theorem to the vector

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

and Theorem 1 in Kuersteiner and Prucha (2013), we obtain that for all fixed $r_1, \ldots, r_k$ and using the convention that $r_0 = 0$,

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

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

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

(b) Tightness. The second step in establishing the functional CLT involves proving tightness of the sequence $\lambda' X_{n\tau}(r)$. 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 $\lambda' X_{n\tau}(r)$ for all $\lambda \in \mathbb{R}^d$ such that $\lambda'\lambda = 1$. In the following, we make use of Theorem 8.3 in Billingsley (1968) (see Appendix A.1 for a statement of Theorem 8.3). The fact that tightness in our case can be established using Criterion (8.5) in Billingsley (1968, Thm. 8.3) follows from Billingsley (1968, Thm. 15.5) and the proof of Billingsley (1968, Thm. 8.3).

Recall the definition

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

and note that

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

where $\left|\lambda_y'\left(X_{n\tau,y}(s) - X_{n\tau,y}(t)\right)\right| = 0$ uniformly in $t, s \in [0,1]$ because of the initial condition $X_{n\tau}(0)$ given in (17) and the fact that $X_{n\tau,y}(t)$ is constant as a function of t. Thus, to show tightness, we only need to consider $\lambda_\nu' X_{n\tau,\nu}(t)$.

Billingsley (1968, pp. 58–59) constructs a continuous approximation to $\lambda' X_{n\tau}(r)$. We denote it as $\lambda_\nu' X_{n\tau,\nu}^c(r)$ and define it analogously to Billingsley (1968, eqn. (8.15)) as

$$\lambda_\nu' X_{n\tau,\nu}^c(r) = \frac{1}{\sqrt{\tau}} \sum_{t=\tau_0+1}^{\tau_0+[\tau r]} \lambda_\nu' \psi_{\tau,t}^{\nu} + \frac{(\tau r - [\tau r])}{\sqrt{\tau}} \lambda_\nu' \psi_{\tau,[\tau r]+1}^{\nu}.$$

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

First, note that, for $\varepsilon > 0$, $\sup_{\tau,r} |\tau r - [\tau r]| \leq 1$ such that

$$P\left(\max_{r\in[0,1]} \left|\frac{(\tau r - [\tau r])}{\sqrt{\tau}} \lambda_\nu' \psi_{\tau,[\tau r]+1}^\nu\right| > \varepsilon\right) \leq \frac{\sup_t E\left[\left\|\psi_{\tau,t}^\nu\right\|\right]}{\varepsilon\sqrt{\tau}} \to 0$$

and consequently, $X_{n\tau,\nu}^c(r) = X_{n\tau,\nu}(r) + o_p(1)$ uniformly in $r \in [0,1]$. To establish tightness, we need to establish that the "modulus of continuity"

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

where $t, s \in [0,1]$ satisfies

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

Let $S_{\tau,k} = \frac{1}{\sqrt{\tau}} \sum_{t=\tau_0+1}^{\tau_0+k} \lambda_\nu' \psi_{\tau,t}^\nu$. By the inequalities in Billingsley (1968, p. 59), it follows that, for k such that $k/\tau < t < (k+1)/\tau$,

$$\sup_{t\leq s\leq t+\delta/2} \left|\lambda'\left(X_{n\tau}^c(s) - X_{n\tau}^c(t)\right)\right| \leq 2 \max_{0\leq i\leq \tau} \left|S_{\tau,k+i} - S_{\tau,k}\right|.$$

By Billingsley (1968, Thm. 8.4) and the comments in Billingsley (1968, p. 59) to establish tightness, it is enough to show that, translated to our notation, for each positive ϵ, there exists a positive $c > 1$ and an integer τ_0 such that, for $\tau \geq \tau_0$ and for all k, it follows that

$$P\left(\max_{s\leq\tau} \left|S_{\tau,k+s} - S_{\tau,k}\right| > c\right) \leq \frac{\epsilon}{c^2}. \tag{47}$$

We note that we normalized the scaling factor $\sigma = 1$ relative to the expression in Billingsley (1968) (see also Billingsley, 1968, p. 58). Using properties of lim sup and lim, Condition 47 is implied by Condition 48, which states that

$$\lim_{c\to\infty} \limsup_{\tau\to\infty} c^2 P\left(\max_{s\leq\tau} \left|S_{\tau,k+s} - S_{\tau,k}\right| > c\right) = 0 \tag{48}$$

holds for all $k \in \mathbb{N}$. A proof of (48) is given in Section III.5 of the Supplementary Material.

(c) Characterization of the limit distribution. 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 $r_1, \ldots, r_k$ is arbitrary implies that, for $\lambda \in \mathbb{R}^d$ with $\lambda = \left(\lambda_y', \lambda_\nu'\right)'$,

$$E\left[\exp\left(i\lambda' X_{n\tau}(r)\right)\right] \to E\left[\exp\left(-\frac{1}{2}\left(\lambda_y'\Omega_y\lambda_y + \lambda_\nu'\Omega_\nu(r)\lambda_\nu\right)\right)\right] \tag{49}$$

with $\Omega_y = \sum_{t\in\{1,\ldots,T\}} \Omega_{yt}$. The final step of the argument consists in representing the limiting process in terms of stochastic integrals over isonormal Gaussian processes.[20] By the law of iterated expectations and the fact that by Assumptions (2) and (3) the matrices

[20] We are grateful to an anonymous referee for suggesting a simplified method of proof for this step.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

Ω_y and $\Omega_\nu(r)$ are $\mathcal{C}$-measurable, it follows that

$$E\left[\exp\left(-\frac{1}{2}\left(\lambda_y'\Omega_y\lambda_y+\lambda_\nu'\Omega_\nu(r)\lambda_\nu\right)\right)\right]$$
$$=E\left[E\left[\exp\left(-\frac{1}{2}\lambda_y'\Omega_y\lambda_y\right)\middle|\mathcal{C}\right]E\left[\exp\left(-\frac{1}{2}\lambda_\nu'\Omega_\nu(r)\lambda_\nu\right)\middle|\mathcal{C}\right]\right]. \tag{50}$$

Let $W(r)=\left(W_y(r),W_\nu(r)\right)$ 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 first term on the RHS of (50) satisfies

$$E\left[\exp\left(-\frac{1}{2}\lambda_y'\Omega_y\lambda_y\right)\middle|\mathcal{C}\right]=E\left[\exp\left(i\lambda_y'\Omega_y^{1/2}W_y(1)\right)\middle|\mathcal{C}\right] \tag{51}$$

by the properties of the standard Gaussian characteristic function. To analyze the second conditional expectation $E\left[\exp\left(-\frac{1}{2}\lambda_\nu'\Omega_\nu(r)\lambda_\nu\right)|\mathcal{C}\right]$, note that by Kallenberg (1997, p. 210), it follows from the isometry of the stochastic integral that there exists a standard Brownian process $W_\nu(r)$ such that

$$E\left[\left(\int_0^r\lambda_\nu'\left(\dot{\Omega}_\nu(t)\right)^{1/2}dW_\nu(t)\right)^2\middle|\mathcal{C}\right]=\int_0^r\lambda_\nu'\left(\dot{\Omega}_\nu(t)\right)\lambda_\nu dt=\lambda_\nu'\Omega_\nu(r)\lambda_\nu.$$

By linearity of the stochastic integral, conditional on $\mathcal{C}$, $\int_0^r\lambda_\nu'\left(\dot{\Omega}_\nu(t)\right)^{1/2}dW_\nu(t)$ is a centered Gaussian process with conditional (on $\mathcal{C}$) characteristic function

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

Combining (50)–(52) gives

$$E\left[\exp\left(-\frac{1}{2}\left(\lambda_y'\Omega_y\lambda_y+\lambda_\nu'\Omega_\nu(r)\lambda_\nu\right)\right)\right]$$
$$=E\left[E\left[\exp\left(i\lambda_y'\Omega_y^{1/2}W_y(1)\right)\middle|\mathcal{C}\right]E\left[\exp\left(i\int_0^r\lambda_\nu'\left(\dot{\Omega}_\nu(t)\right)^{1/2}dW_\nu(t)\right)\middle|\mathcal{C}\right]\right]. \tag{53}$$

Since by construction, $\left(W_y(r),W_\nu(r)\right)$ are mutually independent conditional on $\mathcal{C}$, it follows that the RHS of (53) can be written as

$$E\left[E\left[\exp\left(i\lambda_y'\Omega_y^{1/2}W_y(1)\right)\middle|\mathcal{C}\right]E\left[\exp\left(i\int_0^r\lambda_\nu'\left(\dot{\Omega}_\nu(t)\right)^{1/2}dW_\nu(t)\right)\middle|\mathcal{C}\right]\right]$$
$$=E\left[E\left[\exp\left(i\lambda_y'\Omega_y^{1/2}W_y(1)+i\int_0^r\lambda_\nu'\left(\dot{\Omega}_\nu(t)\right)^{1/2}dW_\nu(t)\right)\middle|\mathcal{C}\right]\right]$$
$$=E\left[\exp\left(i\lambda_y'\Omega_y^{1/2}W_y(1)+i\int_0^r\lambda_\nu'\left(\dot{\Omega}_\nu(t)\right)^{1/2}dW_\nu(t)\right)\right],$$

where the last equality follows from the law of iterated expectations.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

A.3. 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 $\Omega_\nu(1)^{1/2} W_\nu(1)$.

APPENDIX B: Proofs for Trend Stationary Models

B.1. 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, and where we assume exact identification to simplify the notation. The overidentified case follows by the same arguments and with straight forward adjustments to the notation. Using the notation developed before, we define

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

analogously to (16) and

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

analogously to (15). Stack the moment vectors

$$\tilde{s}_{it}(\phi) \equiv \tilde{s}_{it}(\theta,\rho) = \left(\tilde{s}_{it}^y(\theta,\rho)', \tilde{s}_{it}^\nu(\beta,\rho)'\right)' \qquad \textbf{(54)}$$

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 (21) and (22) 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 Condition 4(i) and (ii) that

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

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 $\|\bar{\phi}-\phi_0\| \le \left\|\hat{\phi}-\phi_0\right\|$ 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

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

where $\tilde{s}_{it,\rho}^{\nu}$ denotes moment conditions associated with ρ. From Condition 4 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)

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

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

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

B.2. Proof of Corollary 3

Partition

$$A(\phi_0) = \begin{bmatrix} A_{y,\theta} & \sqrt{\kappa}A_{y,\rho} \\ \frac{1}{\sqrt{\kappa}}A_{\nu,\theta} & A_{\nu,\rho} \end{bmatrix}$$

with inverse

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

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

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

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

where the right-hand side has a mixed normal distribution,

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

and

$$A(\phi_0)^{-1}\Omega A(\phi_0)'^{-1}=\begin{bmatrix} A^{y,\theta}\Omega_y A^{y,\theta\prime}+\kappa A^{y,\rho}\Omega_\nu(1)A^{y,\rho\prime} & \frac{1}{\sqrt{\kappa}}A^{y,\theta}\Omega_y A^{\nu,\theta\prime}+\sqrt{\kappa}A^{y,\rho}\Omega_\nu(1)A^{\nu,\rho\prime} \\ \frac{1}{\sqrt{\kappa}}A^{\nu,\theta}\Omega_y A^{y,\theta\prime}+\sqrt{\kappa}A^{\nu,\rho}\Omega_\nu(1)A^{y,\rho\prime} & \frac{1}{\kappa}A^{\nu,\theta}\Omega_y A^{\nu,\theta\prime}+A^{\nu,\rho}\Omega_\nu(1)A^{\nu,\rho\prime}\end{bmatrix}.$$

The form of the matrices Ω_y and Ω_ν follows 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 $s_M^y(\theta,\rho)$ and $s_M^\nu(\beta,\rho)$, and Condition 4(i) and (ii).

APPENDIX C: Proofs for Section 6

C.1. Proof of Theorem 3

We first establish the joint stable convergence of $\left(V_{\tau n}(r),s_{ML}^y\right)$. Recall that

$$\tau^{-1/2}\nu_{\tau,t}=\exp((t-\min(1,\tau_0))\gamma/\tau)V(0)+\tau^{-1/2}\sum_{s=\min(1,\tau_0)+1}^{t}\exp((t-s)\gamma/\tau)\eta_s$$

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

$$\tau^{-1/2}\nu_{\tau,\tau_0+[\tau r]}=\exp((t-\min(1,\tau_0))\gamma/\tau)V(0)+\exp([\tau r]\gamma/\tau)\tilde{V}_{\tau n}(r).$$

We establish joint stable convergence of $\left(\tilde{V}_{\tau n}(r),s_{ML}^y\right)$ 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 Theorem VIII. 5.33(ii) of JS) and an argument used in Kuersteiner and Prucha (2013, p. 119), stable convergence of $\left(\tilde{V}_{\tau n}(r),s_{ML}^y\right)$, implies that

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

also converges jointly and $\mathcal{C}$-stably. Subsequently, this argument will simply be referred to as the "continuous mapping theorem." In addition, $\exp(([\tau r]-\min(1,\tau_0))\gamma/\tau)V(0)\xrightarrow{p}\exp(r\gamma)V(0)$, which is measurable with respect to $\mathcal{C}$. Together these results imply joint stable convergence of $\left(V_{\tau n}(r),s_{ML}^y\right)$. We thus turn to $\left(\tilde{V}_{\tau n}(r),s_{ML}^y\right)$. To apply Theorem 1, we need to show that $\psi_{\tau,s}=\exp(-s\gamma/\tau)\eta_s$ satisfies Conditions 1(iv) and 2. Since

$$|\exp(-s\gamma/\tau)\eta_s|^{2+\delta}=|\exp(-s/\tau)|^{\gamma(2+\delta)}|\eta_s|^{2+\delta}=e^{-\gamma s(2+\delta)/\tau}|\eta_s|^{2+\delta}\le|\eta_s|^{2+\delta} \tag{55}$$

such that

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

Condition 1(iv) holds. Note that $E\left[|\eta_t|^{2+\delta}\right] \leq C$ holds since we impose Condition 7. Next, note that $E\left[\exp(-2s\gamma/\tau)\,\eta_s^2\right] = \sigma^2 \exp(-2s\gamma/\tau)$. Then, it follows from the proof of Chan and Wei (1987, eqn. (2.3))[21] that

$$\tau^{-1} \sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]} (\psi_{\tau,s})^2 = \tau^{-1} \sum_{t=\tau_0+[\tau s]+1}^{\tau_0+[\tau r]} \exp(-2\gamma t/\tau)\,\eta_t^2 \xrightarrow{p} \sigma^2 \int_s^r \exp(-2\gamma t)\,dt. \tag{56}$$

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 (53) and Theorem 1, we have that

$$\left(\tilde{V}_{\tau n}(r), s_{ML}^y\right) \Rightarrow \left(\sigma \int_0^r e^{-s\gamma}\,dW_\nu(s), \Omega_y^{1/2} W_y(1)\right) \ \mathcal{C}\text{-stably},$$

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

$$\left(V_{\tau n}(r), s_{ML}^y\right) \Rightarrow \left(\exp(r\gamma)\,V(0) + \sigma \int_0^r e^{(r-s)\gamma}\,dW_\nu(s), \Omega_y^{1/2} W_y(1)\right) \ \mathcal{C}\text{-stably}. \tag{57}$$

Note that $\sigma\int_0^r e^{(r-s)\gamma}\,dW_\nu(s)$ is the same term as in Phillips (1987), whereas the limit given in (57) is the same as in Kurtz and Protter (1991, p. 1043).

We now square (32) and sum both sides as in Chan and Wei (1987, eqn. (2.8)) or Phillips (1987) to write

$$\tau^{-1} \sum_{s=\tau_0+1}^{\tau+\tau_0} \nu_{\tau s-1}\eta_s = \frac{e^{-\gamma/\tau}}{2}\tau^{-1}\left(\nu_{\tau,\tau+\tau_0}^2 - \nu_{\tau,\tau_0}^2\right) \tag{58}$$
$$+ \frac{\tau e^{-\gamma/\tau}}{2}\left(1-e^{2\gamma/\tau}\right)\tau^{-2}\sum_{s=\tau_0+1}^{\tau+\tau_0}\nu_{\tau s-1}^2 - \frac{e^{-\gamma/\tau}}{2}\tau^{-1}\sum_{s=\tau_0+1}^{\tau+\tau_0}\eta_s^2.$$

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

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

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

[21] See Section V of the Supplementary Material for details.

https://doi.org/10.1017/S0266466622000391 Published online by Cambridge University Press

$$E\left[\left|U_{\tau,\tau}\right|^{2+\delta}\right] \leq \tau^{-1}\sum_{t=\tau_0+1}^{\tau+\tau_0} E\left[\left|E\left[\eta_s^2|\mathcal{G}_{n,t^*n}\right]\right|^{1+\delta/2}\right] \leq \sup_t E\left[|\eta_t|^{2+\delta}\right] < \infty \tag{59}$$

such that $U_{\tau,\tau}^2$ is uniformly integrable. The bound in (59) also means that by Theorem 2.23 of Hall and Heyde (1980), it follows that $E\left[\left|U_{\tau,\tau}^2 - \tau^{-1}\sum_{s=\tau_0+1}^{\tau+\tau_0}\eta_t^2\right|\right] \to 0$ 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

$$\tau^{-1}\nu_{\tau,\tau+\tau_0}^2 = V_{\tau n}(1)^2, \tag{60}$$

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

and

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

such that by the continuous mapping theorem and (57) it follows that

$$\tau^{-1}\sum_{s=\tau_0+1}^{\tau+\tau_0}\nu_{\tau s-1}\eta_s \Rightarrow \frac{1}{2}\left(V_{\gamma,V(0)}(1)^2 - V(0)^2\right) - \gamma\int_0^1 V_{\gamma,V(0)}(r)^2\,dr - \frac{\sigma^2}{2}. \tag{61}$$

An application of Ito's calculus to $V_{\gamma,V(0)}(r)^2/2$ shows that the RHS of (61) is equal to $\sigma\int_0^1 V_{\gamma,V(0)}dW_\nu$, which also appears in Kurtz and Protter (1991, eqn. (3.10)). However, note that the results in Kurtz and Protter (1991) do not establish stable convergence and thus do not directly apply here. When $V(0) = 0$, these expressions are the same as in Phillips (1987, eqn. (8)). It then is a further consequence of the continuous mapping theorem that

$$\left(V_{\tau n}(r), s_{ML}^y, \tau^{-1}\sum_{s=\tau_0+1}^{\tau+\tau_0}\nu_{\tau s-1}\eta_s\right)$$
$$\Rightarrow \left(V_{\gamma,V(0)}(r), \Omega_y^{1/2}W_y(1), \sigma\int_0^1 V_{\gamma,V(0)}(r)\,dW_\nu(r)\right) \; (\mathcal{C}\text{-stably}).$$

C.2. Proof of Theorem 4

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

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

Defining

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

and partitioning $A^{y}_{\tau n}(\phi) = \left(A^{y,\theta}_{\tau n}(\phi), A^{y,\rho}_{\tau n}(\phi) \right)$ where $A^{y,\theta}_{\tau n}(\phi)$ and $A^{y,\rho}_{\tau n}(\phi)$ contain the partial derivatives with respect to θ and ρ, we have as before for some $\left\| \tilde{\phi} - \phi \right\| \leq \left\| \hat{\phi} - \phi \right\|$ that for

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

we have

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

Using the representation

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

it follows from the continuous mapping theorem and Theorem 3 that

$$\left(V_{\tau n}(r), s^{y}_{ML}, A^{y}_{\tau n}(\phi_0), \int_0^1 V_{\tau n}(r)^2 \, dr, \tau^{-1} \sum_{s=\tau_0+1}^{\tau+\tau_0} \nu_{\tau s-1} \eta \right) \tag{62}$$

$$\Rightarrow \left(V(r), \Omega^{1/2}_{y} W_y(1), A^{y}(\phi_0), \int_0^1 V_{\gamma,V(0)}(r)^2 \, dr, \int_0^s \sigma V_{\gamma,V(0)} dW_\nu \right) \text{ ($\mathcal{C}$-stably).}$$

The partitioned inverse formula implies that

$$A(\phi_0)^{-1} = \begin{bmatrix} A^{-1}_{y,\theta} & A^{-1}_{y,\theta} A_{y,\rho} \left(\int_0^1 V_{\gamma,V(0)}(r)^2 \, dr \right)^{-1} \\ 0 & -\left(\int_0^1 V_{\gamma,V(0)}(r)^2 \, dr \right)^{-1} \end{bmatrix}. \tag{63}$$

By Condition 9, (62), and the continuous mapping theorem, it follows that

$$D^{-1}_{n\tau} \left(\hat{\phi} - \phi_0 \right) \Rightarrow -A(\phi_0)^{-1} \begin{bmatrix} \Omega^{1/2}_{y} W_y(1) \\ \int_0^s \sigma V_{\gamma,V(0)} dW_\nu \end{bmatrix}. \tag{64}$$

The result now follows immediately from (63) and (64).

SUPPLEMENTARY MATERIAL

Hahn, J., G. Kuersteiner, and M. Mazzocco (2022) Supplement to "Central Limit Theory for Combined Cross Section and Time Series with an Application to Aggregate Productivity Shocks," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466622000391.

REFERENCES

Ackerberg, D. A., K. Caves & G. Frazer (2015) Identification properties of recent production function estimators. *Econometrica* 83, 2411–2451.

An, S. & F. Schorfheide (2007) Left Bayesian analysis of DSGE models. *Econometric Reviews* 26, 113–172.

Andrews, D.W.K. (2005) Cross-section regression with common shocks. *Econometrica* 73, 1551–1585.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, & N. Shephard (2008) Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.

Brown, B.M. (1971) Martingale central limit theorems. *Annals of Mathematical Statistics* 42, 59–66.

Campbell, J.Y. & M. Yogo (2006) Efficient tests of stock return predictability. *Journal of Financial Economics* 81, 27–60.

Chan, N.H. & C.Z. Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.

Dedecker, J. & F. Merlevede (2002) Necessary and sufficient conditions for the conditional central limit theorem. *Annals of Probability* 30, 1044–1081.

Eagleson, G.K. (1975) Martingale convergence to mixtures of infinitely divisible laws. *The Annals of Probability* 3, 557–562.

Feigin, P.D. (1985) Stable convergence of semimartingales. *Stochastic Processes and their Applications* 19, 125–134.

Gänssler, P. & E. , Häusler (1979) Remarks on the functional central limit theorem for martingales. *Z. fuer Wahrscheinlichkeitstheorie* 50, 237–243.

Gordin, M.I. (1969) The central limit theorem for stationary processes. *Doklady Akademii Nauk SSSR* 188, 739–741.

Gordin, M.I. (1973) Abstracts of Communication, T.1: A-K, International Conference on Probability Theory, Vilnius.

Hahn, J., G. Kuersteiner, & M. Mazzocco (2020) Estimation with aggregate shocks. *Review of Economic Studies* 87(3), 1365–1398.

Hahn, J., G. Kuersteiner, & M. Mazzocco (2021) Joint time series and cross-section limit theory under mixingale assumptions. *Econometric Theory*, published online 11 August 2020. https://doi.org/10.1017/S0266466620000316.

Hall, P. & C. Heyde (1980) *Martingale Limit Theory and Its Applications*. Academic Press.

Hansen, B.E. (1992) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.

Heckman, J.J., L. Lochner, & 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, 1–58.

Heckman, J.J. & G. Sedlacek (1985) Heterogeneity, aggregation, and market wage functions: An empirical model of self-selection in the labor market. *Journal of Political Economy* 93, 1077–1125.

Inoue, A. & L. Kilian (2020) The uniform validity of impulse response inference in autoregressions. *Journal of Econometrics* 215(2), 450–472.

Jacod, J., M. Podolskij, & M. Vetter (2010) Limit theorems for moving averages of discretized processes plus noise. *Annals of Statistics* 38, 1478–1545.

Jacod, J. & P. Protter (2012) *Discretization of Processes*. Springer.

Jacod, J. & A.N. Shiryaev (2002) *Limit Theorems for Stochastic Processes, 2nd Edition*. Springer.

Kallenberg, O. (1997) *Foundations of Modern Probability*. Springer.

Kuersteiner, G.M. & I.R. Prucha (2013) Limit theory for panel data models with cross sectional dependence and sequential exogeneity. *Journal of Econometrics* 174, 107–126.

Kuersteiner, G.M. & I.R. Prucha (2020) Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica* 88, 2109–2146.

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

Kydland, F.E. & E.C. Prescott (1982) Time to build and aggregate fluctuations. *Econometrica* 50, 1345–1370.

Lee, D. & K.I. Wolpin (2006) Intersectoral labor mobility and the growth of service sector. *Econometrica* 47, 1–46.

Lee, D. & 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, 68–85.

Li, J. & D. Xiu (2016) Generalized method of integrated moments for high-frequency data. *Econometrica* 84, 1613–1633.

Long, J.B. & C.I. Plosser (1983) Real business cycles. *Journal of Political Economy* 91, 39–69.

McLeish, D.L. (1975a) A maximal inequality and dependent strong laws. *Annals of Probability* 3, 829–839.

McLeish, D.L. (1975b) Invariance principles for dependent variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 32, 165–178.

Mikusheva, A. (2007) Uniform inference in autoregressive models. *Econometrica* 75, 1411–1452.

Murphy, K.M. & R.H. Topel (1985) Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics* 3, 370–379.

Olley, S. & A. Pakes (1996) The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64, 1263–1297.

Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.

Phillips, P.C.B. (1988) Regression theory for near integrated time series. *Econometrica* 56, 1021–1044.

Phillips, P.C.B. (2014) On confidence intervals for autoregressive roots and predictive regression. *Econometrica* 82, 1177–1195.

Phillips, P.C.B. & S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495.

Phillips, P.C.B. & H.R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.

Phillips, P.C.B. & S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–193.

Phillips, P.C.B. & D. Sul (2003) Dynamic panel estimation and homogeneity testing under cross section dependence. *The Econometrics Journal* 6, 217–259.

Rényi, A. (1963) On stable sequences of events. *Sankhya: The Indian Journal of Statistics, Series A* 25, 293–302.

Ridder, G. & R. Moffitt (2007) The econometrics of data combination. In J.J. Heckman and E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 6 (Part B). Elsevier.

Rootzen, H. (1983) Central limit theory for martingales via random change of time. In L. Holst and A. Gut (eds.), *Essays in Honour of Carl Gustav Essen*, pp. 154–190. Uppsala University.

Rosenzweig, M.R. & C. Udry (2020) External validity in a stochastic world: Evidence from low-income countries. *Review of Economic Studies* 87(1), 343–381.

Runkle, D.E. (1991) Liquidity constraints and the permanent-income hypothesis: Evidence from panel data. *Journal of Monetary Economics* 27, 73–98.

Schorfheide, F. (2000) Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics* 15, 645–670.

Shea, J. (1995) Union contracts and the life-cycle/permanent-income hypothesis. *American Economic Review* 85, 186–200.

Smets, F. & R. Wouters (2007) Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review* 97, 586–606.

Stroock, D.W. & S.R.S. Varadhan (1979) *Multidimensional Diffusion Processes*. Springer.

Tobin, J. (1950) A statistical demand function for food in the U.S.A. *Journal of the Royal Statistical Society, Series A* 113, 113–149.

Vissing-Jørgensen, A. (2002) Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy* 110, 825–853.

Wooldridge, J.M. & H. White (1988) Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4, 210–230.