Estimation with Aggregate Shocks

Jinyong Hahn* Guido Kuersteiner† Maurizio Mazzocco‡

UCLA University of Maryland UCLA

October 5, 2016

Abstract

Aggregate shocks affect most households' and firms' decisions. Using three stylized models we show that inference based on cross-sectional data alone generally fails to correctly account for decision making of rational agents facing aggregate uncertainty. We propose an econometric framework that overcomes these problems by explicitly parametrizing the agents' inference problem relative to aggregate shocks. Our framework and examples illustrate that the cross-sectional and time-series aspects of the model are often interdependent. Estimation of model parameters in the presence of aggregate shock requires, therefore, the combined use of cross-sectional and time series data. We provide easy-to-use formulas for test statistics and confidence intervals that account for the interaction between the cross-sectional and time-series variation.

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

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

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

1 Introduction

An extensive body of economic research suggests that aggregate shocks have important effects on households' and firms' decisions. Consider for instance the oil shock that hit developed countries in 1973. A large literature has provided evidence that this aggregate shock triggered a recession in the United States, where the demand and supply of non-durable and durable goods declined, inflation grew, the unemployment rate increased, and real wages dropped.

The profession has generally adopted one of the following three strategies to deal with aggregate shocks. The most common strategy is to assume that aggregate shocks have no effect on households' and firms' decisions and, hence, that aggregate shocks can be ignored. Almost all papers estimating discrete choice dynamic models or dynamic games are based on this premise. Examples include Keane and Wolpin (1997), Bajari, Bankard, and Levin (2007), and Eckstein and Lifshitz (2011). We show that, if aggregate shocks are an important feature of the data, ignoring them generally leads to inconsistent parameter estimates. The second approach is to add in a linear fashion time dummies to the model in an attempt to capture the effect of aggregate shocks on the estimation of the parameters of interest, as was done for instance in Runkle (1991) and Shea (1995). We clarify that, if the econometrician does not account properly for aggregate shocks, the parameter estimates will generally be inconsistent even if the actual realizations of the aggregate shocks are observed. The linear addition of time dummies, therefore, fails to solve the problem.[1] The last strategy is to fully specify how aggregate shocks affect individual decisions jointly with the rest of the structure of the economic problem. Using this approach, the econometrician can obtain consistent estimates of the parameters of interest. We are aware of only one paper that uses this strategy, Lee and Wolpin (2010). Their paper is primarily focused on the estimation of a specific empirical model, and they do not address the broader question of which statistical assumptions and what type of data requirements are needed more generally to obtain consistent estimators when aggregate shocks are present. Moreover, as we argue later on, in Lee and Wolpin's (2010) paper there are issues with statistical inference and efficiency.

[1] In the Euler equation context, Chamberlain (1984) uses examples to argue that, when aggregate shocks are present but disregarded, the estimated parameters are generally inconsistent. His examples make clear that generally time dummies do not solve the problems introduced by the presence of aggregate shocks.

The previous discussion reveals that there is no generally agreed upon econometric framework for statistical inference in models where aggregate shocks have an effect on individual decisions. The purpose of this paper is to provide such a general econometric framework. We show that inference based on cross-sectional data alone generally fails to correctly account for decision making of rational agents facing aggregate uncertainty. By parametrizing aggregate uncertainty and explicitly accounting for it when solving for the agents decision problem, we are able to offer an econometric framework that overcomes these problems. We advocate the combined use of cross-sectional and time series data, and we develop simple-to-use formulas for test statistics and confidence intervals that enable the combined use of time series and cross-sectional data.

We proceed in three steps. In Section 2, we introduce the general identification problem by examining a general class of models with the following two features. First, each model in this class is composed of two submodels. The first submodel includes all the cross-sectional features, whereas the second submodel is composed of all the time-series aspects. As a consequence, the parameters of the model can also be divided into two groups: the parameters that characterize the cross-sectional submodel and the parameters that enter the time-series submodel. The second feature is that the two submodels are linked by a vector of aggregates shocks and by the parameters that govern their dynamics. Individual decision making thus depends on aggregate shocks.

Given the interplay between the two submodels, aggregate shocks have complicated effects on the estimation of the parameters of interest. To better understand those effects, in the second step, we present three examples of the general framework that illustrate the complexities generated by the existence of the aggregate shocks. Section 3 considers a model of portfolio choices with aggregate shocks and shows that, if only cross-sectional variation is used, the estimates of the model parameters are biased and inconsistent. It also shows that to obtain unbiased and consistent estimates it is necessary to combine cross-sectional and time-series variation.

In Section 4, as a second example, we study the estimation of firms' production functions when aggregate shocks affect firms' decisions. This example shows that there are exceptional cases in which model parameters can be estimated using only repeated cross-sections if time dummies are skillfully used and not simply added as time intercepts. Specifically, we show that the method proposed by Olley and Pakes (1996) for the estimation of production functions can be modified with the proper inclusion of time dummies to account for the effect of aggregate shocks. The results of Section 4 are of independent interest since the estimation of firms' production functions is an important topic in industrial organization and aggregate shocks have significant effects in most markets.

In Section 5 we discuss as our last example a general equilibrium model of education and labor supply decisions. The portfolio example has the quality of being simply. But, because of its simplicity, it generates a one-directional relationship between the time-series and cross-sectional submodels: the variables and parameters of the time-series model affect the variables and parameters of the cross-sectional submodel, but the opposite is not true. As a result, the parameters of the time-series submodel can be estimated without knowing the cross-sectional parameters. However, this is not generally the case. In the majority of situations, the link between the two submodels is bi-directional. The advantage of the general-equilibrium example is that it produces a bi-directional relationship we can use to illustrate the complexity of the effect of aggregate shocks on parameter estimation.

The examples make clear that generally the best approach to consistently estimate the parameters of the investigated models is to combine cross-sectional data with a long time-series of aggregate variables.[2] As the last step, in Section 6 we provide easy-to-use formulas that can be employed to derive test statistics and confidence intervals for parameters estimated by combining those two data sources. The underlying asymptotic theory, which is presented in the companion paper Hahn, Kuersteiner, and Mazzocco (2016), is highly technical due to the complicated interactions that exists between the two-submodels. It is therefore surprising that the formulas necessary to perform inference take simple forms that are easy to adopt. We conclude the section, by illustrating using the general equilibrium model discussed in Section 5 how the formulas can be computed in concrete cases.

In addition to the econometric literature that deals with inferential issues, our paper also contributes to a growing literature whose objective is the estimation of general equilibrium models. Some examples of papers in this literature are Heckman and Sedlacek (1985), Heckman, Lochner, and Taber (1998), Lee (2005), Lee and Wolpin (2006), Gemici and Wiswall (2011), Gillingham,

[2] An alternative method would be to combine cross-sectional and time-series variation by using panel data. Panel data, however, are generally too short to achieve consistency, whereas long time-series data are easier to find for most of the variables that are of interest to economists. More on this at the end of Section 5.

Iskhakov, Munk-Nielsen, Rust, and Schjerning (2015). Aggregate shocks are a natural feature of general equilibrium models. Without them those models have the unpleasant implication that all aggregate variables can be fully explained by observables and, hence, that errors have no effects on those variables. Our general econometric framework makes this point clear by highlighting the impact of aggregate shocks on parameter estimation and the variation required in the data to estimate those models. More importantly, our results provide easy-to-use formulas that can be used to perform statistical inference in a general equilibrium context.

2 The General Identification Problem

This section introduces in a general form the identification problem generated by the existence of aggregate shocks. It follows closely Section 2 in our companion paper Hahn, Kuersteiner, and Mazzocco (2016). We consider a class of models with four main features. First, the model can be divided into two parts. The first part encompasses all the static aspects of the model and will be denoted with the term *cross-sectional submodel*. The second part includes the dynamic aspects of the aggregate variables and will be denoted with the term *time-series submodel*. Second, the two submodels are linked by the presence of a vector of aggregate shocks ν_t and by the parameters that govern their dynamics. Third, the vector of aggregate shocks may not be observed. If that is the case, it is treated as a set of parameters to be estimated. Lastly, the parameters of the model can be consistently estimated only if a combination of cross-sectional and time-series data are available.

We now formally introduce the general model. The variables that characterize the model can be divided into two vectors $y_{i,t}$ and z_s. The first vector $y_{i,t}$ includes all the variables that characterize the cross-sectional submodel, where i describes an individual decision-maker, a household or a firm, and t a time period in the cross-section.[3] The second vector z_s is composed of all the variables associated with the time-series model. Accordingly, the parameters of the general model can be divided into two sets, β and ρ. The first set of parameters β characterizes the cross-sectional submodel, in the sense that, if the second set ρ was known, β and ν_t can be con-

[3] Even if the time subscript t is not necessary in this subsection, we keep it here for notational consistency because later we consider the case where longitudinal data are collected.

sistently estimated using exclusively variation in the cross-sectional variables $y_{i,t}$. Similarly, the vector ρ characterizes the time-series submodel meaning that, if β and ν_t were known, those parameters can be consistently estimated using exclusively the time series variables z_s. There are two functions that relate the cross-sectional and time-series variables to the parameters. The function $f\left(y_{i,t}\right|\beta, \nu_t, \rho)$ restricts the behavior of the cross-sectional variables conditional on a particular value of the parameters. Analogously, the function $g\left(z_s\right|\beta, \rho)$ describes the behavior of the time-series variables for a given value of the parameters. An example is a situation in which (i) the variables $y_{i,t}$ for $i = 1, \ldots, n$ are i.i.d. given the aggregate shock ν_t, (ii) the variables z_s correspond to (ν_s, ν_{s-1}), (iii) the cross-sectional function $f\left(y_{i,t}\right|\beta, \nu_t, \rho)$ denotes the log likelihood of $y_{i,t}$ given the aggregate shock ν_t, and (iv) the time-series function $g\left(z_s\right|\beta, \rho) = g\left(\nu_s\right|\nu_{s-1}, \rho)$ is the log of the conditional probability density function of the aggregate shock ν_s given ν_{s-1}. In this special case the time-series function g does not depend on the cross-sectional parameters β.

We assume that our cross-sectional data consist of $\{y_{i,t},\ i = 1, \ldots, n\}$, and our time series data consist of $\{z_s,\ s = \tau_0 + 1, \ldots, \tau_0 + \tau\}$. For simplicity, we assume that $\tau_0 = 0$ in this section.

The parameters of the general model can be estimated by maximizing a well-specified objective function.[4] Since in our case the general framework is composed of two submodels, a natural approach is to estimate the parameters of interest by maximizing two separate objective functions, one for the cross-sectional model and one for the time-series model. We denote these criterion functions by $F_n(\beta, \nu_t, \rho)$ and $G_\tau(\beta, \rho)$. In the case of maximum likelihood these functions are simply $F_n(\beta, \nu_t, \rho) = \frac{1}{n}\sum_{i=1}^{n} f\left(y_{i,t}\right|\beta, \nu_t, \rho)$ and $G_\tau(\beta, \rho) = \frac{1}{\tau}\sum_{s=1}^{\tau} g\left(z_s\right|\beta, \rho)$. 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.[5]

In the class of models we consider, the identification of the parameters requires the joint use of cross-sectional and time-series data. Specifically, the objective function F of the cross-sectional model evaluated at the cross-sectional parameters β and aggregate shocks ν takes the same value for any feasible set of time-series parameters ρ. Similarly, the objective function G of the time-

[4] Our discussion is motivated by Newey and McFadden's (1994) unified treatment of maximum likelihood and GMM as extremum estimators.

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

series model evaluated at the time-series parameters and aggregate shocks takes the same value for any feasible set of cross-sectional parameters. In our class of models, however, all the parameters of interest can be consistently estimated if cross-sectional data are combined with time-series data.

The next three sections consider special cases of the type of models described above.

3 Example 1: Portfolio Choice

We now present a model of portfolio choices that illustrates the economic relevance of the general class of models introduced in Section 2.

Consider an economy that, in each period t, is populated by n households. These households are born at the beginning of period t, live for one period, and are replaced in the next period by n new families. The households living in consecutive periods do not overlap and, hence, make independent decisions. Each household is endowed with deterministic income and has preferences over a non-durable consumption good $c_{i,t}$. The preferences can be represented by Constant Absolute Risk Aversion (CARA) utility functions which take the following form: $U(c_{i,t}) = -e^{-\delta c_{i,t}}$. For simplicity, we normalize income to be equal to 1.

During the period in which households are alive, they can invest a share of their income in a risky asset with return $u_{i,t}$. The remaining share is automatically invested in a risk-free asset with a return r that does not change over time. At the end of the period, the return on the investment is realized and households consume the quantity of the non-durable good they can purchase with the realized income. The return on the risky asset depends on aggregate shocks. Specifically, it takes the following form: $u_{i,t} = \nu_t + \epsilon_{i,t}$, where ν_t is the aggregate shock and $\epsilon_{i,t}$ is an i.i.d. idiosyncratic shock. The idiosyncratic shock, and hence the heterogeneity in the return on the risky asset, can be interpreted as differences across households in transaction costs, in information on the profitability of different stocks, or in marginal tax rates. We assume that $\nu_t \sim N(\mu, \sigma_\nu^2)$, $\epsilon_{i,t} \sim N(0, \sigma_\epsilon^2)$, and hence that $u_{i,t} \sim N(\mu, \sigma^2)$, where $\sigma^2 = \sigma_\nu^2 + \sigma_\epsilon^2$.

Household i living in period t chooses the fraction of income to be allocated to the risk-free asset $\alpha_{i,t}$ by maximizing its life-time expected utility:

$$\max_{\widetilde{\alpha}} \; E\left[-e^{-\delta c_{i,t}}\right]$$
$$s.t. \quad c_{i,t} = \widetilde{\alpha}\left(1+r\right) + \left(1-\widetilde{\alpha}\right)\left(1+u_{i,t}\right), \tag{1}$$

where the expectation is taken with respect to the return on the risky asset. It is straightforward to show that the household's optimal choice of $\alpha_{i,t}$ is given by[6]

$$\alpha_{i,t}^{*} = \alpha = \frac{\delta\sigma^2 + r - \mu}{\delta\sigma^2}. \tag{2}$$

We will assume that the econometrician is mainly interested in estimating the risk aversion parameter δ.

We now consider an estimator that takes the form of a population analog of (2), and study the impact of aggregate shocks on the estimator's consistency when an econometrician works only with cross-sectional data. Our analysis reveals that such an estimator is inconsistent, due to the fact that cross-sectional data do not contain information about aggregate uncertainty. Our analysis makes explicit the dependence of the estimator on the probability distribution of the aggregate shock and points to the following way of generating a consistent estimator of δ. Using time series variation, one can consistently estimate the parameters pertaining to aggregate uncertainty. Those estimates can then be used in the cross-sectional model to estimate the remaining parameters.[7]

Without loss of generality, we assume that the cross-sectional data are observed in period $t = 1$. The econometrician observes data on the return of the risky asset $u_{i,t}$ and on the return of the risk-free asset r. We assume that in addition he also observes a noisy measure of the share of resources invested in the risky assets $\alpha_{i,t} = \alpha + e_{i,t}$, where $e_{i,t}$ is a zero-mean measurement error. We therefore have that $y_i = (u_{i1}, \alpha_{i1})$. We make the simplifying assumption that the aggregate shock is observable to econometricians and that the time-series variables only include the aggregate shock, i.e. $z_t = \nu_t$. Because $\mu = E\left[u_{i1}\right]$, $\sigma^2 = \text{Var}\left(u_{i1}\right)$, and $\alpha = E\left[\alpha_{i1}\right]$, if only cross-sectional

[6] See the appendix.

[7] Our model is a stylized version of many models considered in a large literature interested in estimating the parameter δ using cross-sectional variation. Estimators are often based on moment conditions derived from first order conditions (FOC) related to optimal investment and consumption decisions. Such estimators have similar problems, which we discuss in Appendix A.2.

variation is used for estimation, we would have the following method-of-moments estimators of those parameters:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} u_{i1} = \bar{u}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(u_{i1} - \bar{u}\right)^2, \qquad \text{and} \qquad \hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} \alpha_{i1}.$$

The econometrician can then use equation (2) to write the risk aversion parameter as $\delta = (\mu - r)/\left(\sigma^2 (1-\alpha)\right)$ and estimate it using the sample analog $\hat{\delta} = (\hat{\mu} - r)/\left(\hat{\sigma}^2 (1-\hat{\alpha})\right)$.

In the presence of the aggregate shock ν_t, those estimators can be written as

$$\begin{aligned}
\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} u_{i1} = \nu_1 + \frac{1}{n}\sum_{i=1}^{n}\epsilon_{i1} = \nu_1 + o_p(1),\\
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(u_{i1} - \bar{u}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\epsilon_{i1} - \bar{\epsilon}\right)^2 = \sigma_\epsilon^2 + o_p(1),\\
\hat{\alpha} &= \alpha + \frac{1}{n}\sum_{i=1}^{n} e_{i1} = \alpha + o_p(1),
\end{aligned}$$

which implies that δ will be estimated to be

$$\hat{\delta} = \frac{\nu_1 + o_p(1) - r}{\left(\sigma_\epsilon^2 + o_p(1)\right)\left(1 - \alpha + o_p(1)\right)} = \frac{\nu_1 - r}{\sigma_\epsilon^2 (1-\alpha)} + o_p(1). \tag{3}$$

Using Equation (3), we can study the properties of the proposed estimator $\hat{\delta}$. If there were no aggregate shock in the model, we would have $\nu_1 = \mu$, $\sigma_\nu^2 = 0$, $\sigma_\epsilon^2 = \sigma^2$ and, therefore, $\hat{\delta}$ would converge to δ, a nonstochastic constant, as n grows to infinity. It is therefore a consistent estimator of the risk aversion parameter. In the presence of the aggregate shock, however, the proposed estimator has different properties. If one conditions on the realization of the aggregate shock ν, the estimator $\hat{\delta}$ is inconsistent with probability 1, since it converges to $\frac{\nu_1 - r}{\sigma_\epsilon^2(1-\alpha)}$ and not to the true value $\frac{\mu - r}{(\sigma_\nu^2 + \sigma_\epsilon^2)(1-\alpha)}$. If one does not condition on the aggregate shock, as n grows to infinity, $\hat{\delta}$ converges to a random variable with a mean that is different from the true value of the risk aversion parameter. The estimator will therefore be biased and inconsistent. To see this, remember that $\nu_1 \sim N\left(\mu, \sigma_\nu^2\right)$. As a consequence, the unconditional asymptotic distribution of $\hat{\delta}$ takes the following form:

$$\hat{\delta} \rightarrow N\left(\frac{\mu - r}{\sigma_\epsilon^2 (1-\alpha)}, \left(\frac{1}{\sigma_\epsilon^2 (1-\alpha)}\right)^2 \sigma_\nu^2\right) = N\left(\delta + \delta\frac{\sigma_\nu^2}{\sigma_\epsilon^2}, \frac{\sigma_\nu^2}{\left(\sigma_\epsilon^2 (\alpha - 1)\right)^2}\right),$$

which is centered at $\delta + \delta\dfrac{\sigma_\nu^2}{\sigma_\epsilon^2}$ and not at δ, hence the bias.

We are not the first to consider a case in which the estimator converges to a random variable. Andrews (2005) and more recently Kuersteiner and Prucha (2013) discuss similar scenarios. Our example is remarkable because the nature of the asymptotic randomness is such that the estimator is not even asymptotically unbiased. This is not the case in Andrews (2005) or Kuersteiner and Prucha (2013), where in spite of the asymptotic randomness the estimator is unbiased.[8]

As mentioned above, there is a simple explanation for our result: cross-sectional variation is not sufficient for the consistent estimation of the risk aversion parameter if aggregate shocks affect individual decisions.[9] To make this point transparent, observe that, conditional on the aggregate shock, the assumptions of this section imply that y_i has the following distribution

$$y_i|\,\nu_1 \sim N\left(\left[\begin{array}{c} \nu_1 \\ \dfrac{\delta\left(\sigma_\nu^2+\sigma_\epsilon^2\right)+r-\mu}{\delta\left(\sigma_\nu^2+\sigma_\epsilon^2\right)} \end{array}\right],\left[\begin{array}{cc} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_e^2 \end{array}\right]\right), \tag{4}$$

Using (4), it is straightforward to see that the cross-sectional likelihood is maximized for any arbitrary choice of the time-series parameters $\rho = (\mu, \sigma_\nu^2)$, as long as one chooses δ that satisfies the following equation:

$$\frac{\delta\left(\sigma_\nu^2+\sigma_\epsilon^2\right)+r-\mu}{\delta\left(\sigma_\nu^2+\sigma_\epsilon^2\right)} = \alpha.$$

As a result, the cross-sectional parameters μ and σ_ν^2 cannot be consistently estimated by maximizing the cross-sectional likelihood and, hence, δ cannot be consistently estimated using only cross-sectional data.

A solution to the problem discussed in this section is to combine cross-sectional variables with time-series variables. In this case, one can consistently estimate (μ, σ_ν^2) by using the time-series

[8]Kuersteiner and Prucha (2013) also consider cases where the estimator is random and inconsistent. However, in their case this happens for different reasons: the endogeneity of the factors. The inconsistency considered here occurs even when the factors (i.e., aggregate shocks) are strictly exogenous.

[9]As discussed in the introductory section, a common practice to account for the effect of aggregate shocks is to include time dummies in the model. The portfolio example clarifies that the addition of time dummies does not solve the problem generated by the presence of aggregate shocks. The inclusion of time dummies is equivalent to the assumption that the aggregate shocks are known. But the previous discussion indicates that, using exclusively cross-sectional data, the estimator $\hat{\delta}$ is biased and inconsistent even if the aggregate shocks are known. An unbiased and consistent estimator of δ can only be obtained if the distribution of the aggregate shocks is known, which is feasible only by exploiting the variation contained in time-series data.

of aggregate data $\{z_t\}$. Consistent estimation of $(\delta, \sigma_\epsilon^2, \sigma_e^2)$ can then be achieved by plugging the consistent estimators of (μ, σ_ν^2) in the correctly specified cross-section likelihood (4).

The example presented in this section is a simplified version of the general class of models introduced in Section 2, since the relationship between the cross-sectional and time-series submodels is simple and one-directional. The variables and parameters of the time-series submodel affect the cross-sectional submodel, but the cross-sectional variables and parameters have no impact on the time-series submodel. As a consequence, the time-series parameters can be consistently estimated without knowing the cross-sectional parameters. In more complicated situations, such as general equilibrium models, where aggregate shocks are a natural feature, the relationship between the two submodels is generally bi-directional. In Section 5, we present a general-equilibrium with that type of relationship. But before considering that case, we study a situation in which the effect of aggregate shocks can be accounted for with the proper use of time dummies.

4 Example 2: Estimation of Production Function

In the previous section, we presented an example that illustrates the complicated nature of identification in the presence of aggregate shocks. The example highlights that generally there is no simple method for estimating the class of models considered in this paper. Estimation requires a careful examination of the interplay between the cross-sectional and time-series submodels. In this section, we consider an example showing that there are exceptions to this general rule. In the case we analyze, identification of a model with aggregate shocks can be achieved using only cross-sectional data provided that time dummies are skillfully employed. We will show that the naive practice of introducing additive time dummies is not sufficient to deal with the effects generated by aggregate shocks. But the solution is simpler than the general approach we adopted to identify the parameters of the portfolio model.

The example we consider here is a simplified version of the problem studied by Olley and Pakes (1996) and deals with an important topic in industrial organization: the estimation of firms' production functions. A profit-maximizing firm j produces in period t a product $y_{j,t}$ 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 $\omega_{j,t}$. It takes the following functional form:

$$y_{j,t} = \beta_0 + \beta_l l_{j,t} + \beta_k k_{j,t} + \omega_{j,t} + \eta_{j,t} \tag{5}$$

where $\eta_{i,t}$ is a measurement error.

Capital and labor are optimally chosen by the firm, jointly with the new investment in capital $i_{j,t}$, by maximizing a dynamic profit function subject to constraints that determine how capital accumulated over time.[10] In the model proposed by Olley and Pakes (1996), firms are heterogeneous in their age and can choose to exit the market. In this section, we will abstract from age heterogeneity and exit decisions because they make the model more complicated without adding more insight on the effect of aggregate shocks on the estimation of production functions.

A crucial feature of the model proposed by Olley and Pakes (1996) and of our example is that the investment decision in period t is a function of the current stock of capital and productivity shock, i.e.

$$i_{j,t} = i_t\left(\omega_{j,t}, k_{j,t}\right). \tag{6}$$

Olley and Pakes (1996) do not allow for aggregate shocks, but in this example we consider a situation in which the productivity shock at t is the sum of an aggregate shock ν_t and of an i.i.d. idiosyncratic shock $\varepsilon_{j,t}$, i.e.

$$\omega_{j,t} = \nu_t + \varepsilon_{j,t}. \tag{7}$$

We will assume that the firm observes the realization of the aggregate shock and, separately, of the i.i.d. shock.

We first describe how the production function (5) can be estimated when aggregate shocks are not present, the method proposed by Olley and Pakes (1996). We then discuss how that method has to be modified with the appropriate use of time dummies if aggregate shocks affect firms' decisions.

The main problem in the estimation of the production function (5) is that the productivity shock is correlated with labor and capital, but not observed by the econometrician. To deal with that issue, Olley and Pakes (1996) use the result that the investment decision (6) is strictly increasing in the productivity shock for every value of capital to invert the corresponding function

[10] For details of the profit function and the accumulation equation for capital, see Olley and Pakes (1996).

and solve for the productivity shock, which implies

$$\omega_{j,t} = h_t\left(i_{j,t}, k_{j,t}\right). \tag{8}$$

One can then replace the productivity shock in the production function using the equation (8) to obtain

$$y_{j,t} = \beta_l l_{j,t} + \phi_t\left(i_{j,t}, k_{j,t}\right) + \eta_{j,t}, \tag{9}$$

where

$$\phi_t\left(i_{j,t}, k_{j,t}\right) = \beta_0 + \beta_k k_{j,t} + h_t\left(i_{j,t}, k_{j,t}\right). \tag{10}$$

The parameter β_l and the function ϕ can then be estimated by regressing $y_{j,t}$ on $l_{j,t}$ and a polynomial in $i_{j,t}$ and $k_{j,t}$. Equivalently, β_l is identified by

$$\beta_l = \frac{E\left[\left(l_{j,t} - E\left[l_{j,t}\middle| i_{j,t}, k_{j,t}\right]\right)\left(y_{j,t} - E\left[y_{j,t}\middle| i_{j,t}, k_{j,t}\right]\right)\right]}{E\left[\left(l_{j,t} - E\left[l_{j,t}\middle| i_{j,t}, k_{j,t}\right]\right)^2\right]}. \tag{11}$$

To identify the parameter on the logarithm of capital β_k observe that

$$E\left[y_{i,t+1} - \beta_l l_{j,t+1}\middle| k_{j,t+1}\right] = \beta_0 + \beta_k k_{j,t+1} + E\left[\omega_{j,t+1}\middle| \omega_{j,t}\right] = \beta_0 + \beta_k k_{j,t+1} + g\left(\omega_{j,t}\right), \tag{12}$$

where the first equality follows from $k_{j,t+1}$ being determined conditional on $\omega_{j,t}$. The shock $\omega_{j,t} = h_t\left(i_{j,t}, k_{j,t}\right)$ is not observed but, using equation (10), can be written in the following form:

$$\omega_{j,t} = \phi_t\left(i_{j,t}, k_{j,t}\right) - \beta_0 - \beta_k k_{j,t}, \tag{13}$$

where ϕ_t is known from the first-step estimation. Substituting for $\omega_{j,t}$ into the function $g(.)$ in equation (12) and letting $\xi_{j,t+1} = \omega_{j,t+1} - E\left[\omega_{j,t+1}\middle| \omega_{j,t}\right]$, equation (12) can be written as follows:

$$y_{i,t+1} - \beta_l l_{j,t+1} = \beta_k k_{j,t+1} + g\left(\phi_t - \beta_k k_{j,t}\right) + \xi_{j,t+1} + \eta_{j,t}. \tag{14}$$

where β_0 has been included in the function $g(.)$. The parameter β_k can then be estimated by using the estimates of β_l and ϕ_t obtained in the first step and by minimizing the sum of squared residuals in the previous equation employing a kernel or a series estimator for the function g.

We now consider the case in which aggregate shocks affect the firm's decisions and analyze how the model parameters can be identified using only cross-sectional variation. The introduction of aggregate shocks changes the estimation method in two main ways. First, the investment decision is affected by the aggregate shock and takes the following form:

$$i_{j,t} = i_t\left(\nu_t, \varepsilon_{j,t}, k_{j,t}\right).$$

where ν_t and $\varepsilon_{j,t}$ enter as independent arguments because the firm observes them separately. Second, all expectation are conditional on the realization of the aggregate shock since in the cross-section there is no variation in that shock and only its realization is relevant.

It is straightforward to show that, if the investment function is strictly increasing in the productivity shock $\omega_{j,t}$ for all capital levels, it is also strictly increasing in ν_t and $\varepsilon_{j,t}$ for all $k_{j,t}$. Using this result, we can invert $i_t(.)$ to derive $\varepsilon_{j,t}$ as a function of the aggregate shock, investment, and the stock of capital, i.e.

$$\varepsilon_{j,t} = h_t\left(\nu_t, i_{j,t}, k_{j,t}\right).$$

The production function can therefore be rewritten in the following form:

$$\begin{aligned} y_{j,t} &= \beta_0 + \beta_l l_{j,t} + \beta_k k_{j,t} + \nu_t + \varepsilon_{j,t} + \eta_{j,t} \\ &= \beta_l l_{j,t} + \left[\beta_0 + \beta_k k_{j,t} + \nu_t + h_t\left(\nu_t, i_{j,t}, k_{j,t}\right)\right] + \eta_{j,t} \\ &= \beta_l l_{j,t} + \phi_t\left(\nu_t, i_{j,t}, k_{j,t}\right) + \eta_{j,t}. \end{aligned} \tag{15}$$

If β_l is estimated using repeated cross-sections and the method developed for the case with no aggregate shocks, the estimated coefficient will generally be biased because the econometrician does not account for the aggregate shock and its correlation with the firm's choice of labor. There is, however, a small variation of the method proposed earlier that produces unbiased estimates of β_l, as long as $\varepsilon_{j,t}$ is independent of $\eta_{j,t}$. The econometrician should regress $y_{j,t}$ on $l_{j,t}$ and a polynomial in $i_{j,t}$ and $k_{j,t}$ where the polynomial is interacted with time dummies. It is this atypical use of time dummies that enables the econometrician to account for the effect of aggregate shocks on firms' decisions. The β_l can therefore be identified by

$$\beta_l = \frac{E\left[\left(l_{j,t} - E\left[l_{j,t} \middle| i_{j,t}, k_{j,t}, \nu_t = \bar{\nu}\right]\right)\left(y_{j,t} - E\left[y_{j,t} \middle| i_{j,t}, k_{j,t}, \nu_t = \bar{\nu}\right]\right)\right]}{E\left[\left(l_{j,t} - E\left[l_{j,t} \middle| i_{j,t}, k_{j,t}, \nu_t = \bar{\nu}\right]\right)^2\right]}. \tag{16}$$

Observe that the expectation operator in the previous equation is defined with respect to a probability distribution function that includes the randomness of the aggregate shock ν_t. But, when one uses cross-sectional variation, ν_t is fixed at its realized value. As a consequence, the distribution is only affected by the randomness of ε_{it}.

For the estimation of β_k, observe that

$$\begin{aligned} & E\left[y_{i,t+1} - \beta_l l_{j,t+1} \right| k_{j,t+1}, \nu_{t+1} = \bar{\nu}'] \\ & = \beta_0 + \beta_k k_{j,t+1} + E\left[\nu_{t+1} + \varepsilon_{j,t+1} \right| , \nu_{t+1} = \bar{\nu'}, \nu_t = \bar{\nu}, \varepsilon_{j,t}] \\ & = \beta_0 + \beta_k k_{j,t+1} + \bar{\nu}' + E\left[\varepsilon_{j,t+1} \right| \nu_t = \bar{\nu}, \varepsilon_{j,t}] \\ & = \beta_0 + \beta_k k_{j,t+1} + \bar{\nu}' + g_t\left(\varepsilon_{j,t} \right) \end{aligned} \tag{17}$$

where the first equality follows from $k_{j,t+1}$ being known if ν_t and $\varepsilon_{j,t}$ are known and the last equality follows from the inclusion of the aggregate shock $\nu_t = \bar{\nu}$ in the function $g_t(.)$.

The only variable of equation (17) that is not observed is $\varepsilon_{j,t}$. But remember that

$$\varepsilon_{j,t} = h_t\left(\nu_t, i_{j,t}, k_{j,t}\right) = \phi_t\left(\nu_t, i_{j,t}, k_{j,t}\right) - \beta_0 - \beta_k k_{j,t} - \nu_t.$$

We can therefore use the above expression to substitute for $\varepsilon_{j,t}$ in equation (17) and obtain

$$\begin{aligned} & E\left[y_{i,t+1} - \beta_l l_{j,t+1} \right| k_{j,t+1}, \nu_{t+1} = \bar{\nu}'] \\ & = \beta_0 + \beta_k k_{j,t+1} + \bar{\nu}' + g_t\left(\phi_t\left(\nu_t, i_{j,t}, k_{j,t}\right) - \beta_0 - \beta_k k_{j,t} - \bar{\nu} \right) \\ & = \beta_k k_{j,t+1} + g_{t,t+1}\left(\phi_t - \beta_k k_{j,t} \right), \end{aligned}$$

where in the last equality β_0, $\bar{\nu}$, and $\bar{\nu}'$ have been included in the function $g_{t,t+1}(.)$. Hence, if one defines $\xi_{j,t+1} = \varepsilon_{j,t+1} - E\left[\varepsilon_{j,t+1}\right| \nu_t = \bar{\nu}, \varepsilon_{j,t}]$, the parameter β_k can be estimated using the following equation:

$$y_{i,t+1} - \beta_l l_{j,t+1} = \beta_k k_{j,t+1} + g_{t,t+1}\left(\phi_t - \beta_k k_{j,t}\right) + \xi_{j,t+1} + \eta_{j,t+1}. \tag{18}$$

But notice that the approach without aggregate shocks cannot be applied directly to equation (18) because the function $g(.)$ depends on time t and $t+1$ aggregate shocks. With aggregate shocks a different function $g(.)$ must be estimated for each period. This can be achieved by replacing $g(.)$ with a polynomial interacted with time dummies.

The previous discussion indicates that firms' production functions can be estimated using only cross-sectional data as long as the functions ϕ and g are estimated period by period. In practice, both functions are often estimated by low degree polynomials. Our analysis indicates that if the coefficients of these polynomials are interacted with time dummies the estimation of production functions will generally be robust to the presence of aggregate shocks.

We conclude by drawing attention to three important features of the example considered in this section. First, in order to deal with the effect of aggregate shocks, we had to carefully examine the meaning of seemingly straightforward objects such as the expectation operator E. We also had to impose assumptions on the information set of the firms, namely that the firm observes the current aggregate shock. Lastly, the time dummies must be interacted with the polynomials. The standard practice of simply adding time dummies as separate intercepts for each time period does not solve the issues introduced by aggregate shocks.

5 Example 3: A General Equilibrium Model

In this section, we consider as a third example a general equilibrium model of education and labor supply decisions in which aggregate shocks influence individual choices. This example provides additional insight on the effect of aggregate shocks on the estimation of model parameters because, differently from the portfolio example, it considers a case in which the relationship between the cross-sectional and time-series models is bi-directional: the cross-sectional parameters cannot be identified without knowledge of the time-series parameters and the time-series parameters cannot be identified without knowing the cross-sectional parameters. In principle, we could have used as a general example a model proposed in the general equilibrium literature such as the model developed in Lee and Wolpin (2006). We decided against this alternative because in those models the effect of the aggregate shocks and the relationship between the cross-sectional and time-series submodels is complicated and therefore difficult to describe. Instead, we have decided to develop a model that is sufficiently general to generate an interesting relationship between the two submodels, but at the same time is sufficiently stylized for this relationship to be easy to describe and understand.

The main objective of the model we develop is to evaluate the effect of aggregate shocks on the education decisions of young individuals and on their subsequent labor supply decisions when of working-age. For that purpose, we consider an economy in which in each period $t \in T$ a young and a working-age generation overlap. Each generation is composed of a measure N_t of individuals who are endowed with preferences over a non-durable consumption good and leisure. The preferences of individual i are represented by a Cobb-Douglas utility function $U^i(c,l) = \left(c^\sigma l^{1-\sigma}\right)^{1-\gamma_i} \Big/ (1-\gamma_i)$, where the risk aversion parameter γ_i is a function of the observable variables $x_{i,t}$, the unobservable variables $\xi_{i,t}$, and a vector of parameters ς, i.e. $\gamma_i = \gamma\left(x_{i,t}, \xi_{i,t} \middle| \varsigma\right)$. Both young and working-age individuals are endowed with a number of hours $\mathcal{T}$ that can be allocated to leisure or to a productive activity. In each period t, the economy is hit by an aggregate shock ν_t whose conditional probability $P\left(\nu_{t+1} \middle| \nu_t\right)$ is given by $\log \nu_{t+1} = \varrho \log \nu_t + \eta_t$. We will assume that η_t is normally distributed with mean 0 and variance ω^2. The aggregate shock affects the labor market in a way that will be established later on.

In each period t, young individuals are endowed with an exogenous income $y_{i,t}$ and choose the type of education to acquire. They can choose either a flexible type of education F or a rigid type of education R. Working-age individuals with flexible education are affected less by adverse aggregate shocks, but they have lower expected wages. The two types of education have identical cost $C_e < y_{i,t}$ and need the same amount of time to acquire $\mathcal{T}_e < \mathcal{T}$. Since young individuals have typically limited financial wealth, we assume that there is no saving decision when young and that any transfer from parents or relatives is included in non-labor income $y_{i,t}$. We also abstract from student loans and assume that all young individuals can afford to buy one of the two types of education. As a consequence, the part of income $y_{i,t}$ that is not spent on education will be consumed. At each t, working-age individuals draw a wage offer $w_{i,t}^F$ if they have chosen the flexible education when young and a wage offer $w_{i,t}^R$ otherwise. They also draw a productivity shock $\varepsilon_{i,t}$ which determines how productive their hours of work are in case they choose to supply labor. We assume that the productivity shock is known to the individual, but not to the econometrician. Given the wage offer and the productivity shock, working-age individuals choose how much to work $h_{i,t}$ and how much to consume. If a working-age individual decides to supply $h_{i,t}$ hours of work, the effective amount of labor hours supplied is given by $\exp\left(\varepsilon_{i,t}\right) h_{i,t}$. We will also assume that $E\left[\exp\left(\varepsilon_{i,t}\right)\right] = 1$.

The economy is populated by two types of firms to whom the working-age individuals supply labor. The first type of firm employs only workers with education F, whereas the second type of firm employs only workers with education R. Both use the same type of capital K. The labor demand functions of the two types of firms are exogenously given and take the following form:

$$\ln H_t^{D,F} = \alpha_0 + \alpha_1 \ln w_t^F,$$

and

$$\ln H_t^{D,R} = \alpha_0 + \alpha_1 \ln w_t^R + \ln \nu_t,$$

where $H^{D,E}$ is the total demand for *effective labor*, with $E = F, R$, $\alpha_0 > 0$, and $\alpha_1 < 0$. We assume that the two labor demands have identical slopes for simplicity. These two labor demand functions allow us to introduce in the model the feature that workers with a more flexible education are affected less by aggregate shocks such as business cycle shocks. The wage for each education group is determined by the equilibrium in the corresponding labor market. It will therefore generally depend on the aggregate shock.

The description of the model implies that there is only one source of uncertainty in the economy, the aggregate shock, and two sources of heterogeneity across individuals, the risk aversion parameter and the productivity shock.

We can now introduce the problem solved by an individual of the young generation. In period t, young individual i chooses consumption, leisure, and the type of education that solve the following problem:

$$\max_{c_{i,t}, l_{i,t}, c_{i,t+1}, l_{i,t+1}, S} \frac{\left(c_{i,t}^{\sigma} l_{i,t}^{1-\sigma}\right)^{1-\gamma_i}}{1-\gamma_i} + \beta \int \frac{\left(c_{i,t+1}^{\sigma} l_{i,t+1}^{1-\sigma}\right)^{1-\gamma_i}}{1-\gamma_i} dP\left(\nu_{t+1} \middle| \nu_t\right) \tag{19}$$

$$s.t. \quad c_{i,t} = y_{i,t} - C_e \quad \text{and} \quad l_{i,t} = \mathcal{T} - \mathcal{T}_e$$

$$c_{i,t+1} = w_{i,t+1}^S \left(\nu_{t+1}\right) \exp\left(\varepsilon_{i,t+1}\right) \left(\mathcal{T} - l_{i,t+1}\right) \quad \text{for every } \nu_{t+1}.$$

Here, $w_{i,t+1}^S(\nu_{t+1})$ denotes the wage rate of individual i in the second period, which depends on the realization of the aggregate shock ν_{t+1} and the education choice $S = F, R$. The wage rate is per unit of the effective amount of labor hours supplied and is determined in equilibrium. The problem solved by a working-age individual takes a simpler form. Conditional on the realization of the aggregate shock ν_t and on the type of education S chosen when young, in period t, individual i of the working-age generation chooses consumption and leisure that solve the following problem:

$$\max_{c_{i,t}, l_{i,t}} \frac{\left(c_{i,t}^{\sigma} l_{i,t}^{1-\sigma}\right)^{1-\gamma_i}}{1-\gamma_i} \tag{20}$$

$$s.t. \ c_{i,t} = w_{i,t}^S \left(\nu_t\right) \exp\left(\varepsilon_{i,t}\right) \left(\mathcal{T} - l_{i,t}\right).$$

We will now solve the model starting from the problem of a working-age individual. Using the first order conditions of problem (20), it is straightforward to show that the optimal choice of consumption, leisure, and hence labor supply for a working-age individual takes the following form:

$$c^*_{i,t} = \sigma w_t(\nu_t, S)\exp(\varepsilon_{i,t})\mathcal{T}, \tag{21}$$

$$l^*_{i,t} = (1-\sigma)\mathcal{T}, \tag{22}$$

$$h^*_{i,t} = \mathcal{T} - l_{i,t}(\nu_{i,t}) = \sigma\mathcal{T}.$$

The supply of effective labor is therefore equal to $\sigma\exp(\varepsilon_{i,t})\mathcal{T}$. Given the optimal choice of consumption and leisure, conditional on the aggregate shock, the value function of a working-age individual with education S can be written as follows:

$$V_{i,t}(S,\nu_t) = \frac{\left[\left(\sigma w^S_{i,t}(\nu_t)\exp(\varepsilon_{i,t})\mathcal{T}\right)^{\sigma}\left((1-\sigma)\mathcal{T}\right)^{1-\sigma}\right]^{1-\gamma_i}}{1-\gamma_i}, \quad \text{S = F, R.}$$

Given the value functions of a working-age individual, we can now characterize the education choice of a young individual. This individual will choose education F if the expectation taken over the aggregate shocks of the corresponding value function is greater than the analogous expectation conditional on choosing education R:

$$\int V_{i,t}(F,\nu_{t+1})\,dP(\nu_{t+1}|\nu_t) \geq \int V_{i,t}(R,\nu_{t+1})\,dP(\nu_{t+1}|\nu_t). \tag{23}$$

Before we can determine which factors affect the education choice, we have to derive the equilibrium in the labor market. We show in the appendix that the labor market equilibrium is characterized by the following two wage equations:

$$\ln w^F_{i,t} = \frac{\ln n^F_t + \ln\sigma + \ln\mathcal{T} - \alpha_0}{\alpha_1} + \varepsilon_{i,t}, \tag{24}$$

$$\ln w^R_{i,t} = \frac{\ln n^R_t + \ln\sigma + \ln\mathcal{T} - \alpha_0 - \ln\nu_t}{\alpha_1} + \varepsilon_{i,t}, \tag{25}$$

where $w^F_{i,t}$ and $w^R_{i,t}$ are the individual wages observed in sectors F and R and n^F_t and n^R_t are the measures of individuals that choose education F and R. We can now replace the equilibrium wages inside inequality (23) and analyze the education decision of a young individual. To simplify the discussion, we will assume that $\varepsilon_{i,t}$ is independent of $\xi_{i,t}$, thereby eliminating sample selection issues in the wage equations. In the appendix, we show that, if the risk aversion parameter γ is greater than or equal to 1, a young individual chooses the flexible type of education if the following inequality is satisfied:

$$\frac{\sigma\omega^2}{2\alpha_1} \geq \frac{1}{1-\gamma\left(x_{i,t},\xi_{i,t}\right|\varsigma)} \log\left(\frac{n_t^F}{n_t^R}\right) + \frac{\varrho \log \nu_t}{1-\gamma\left(x_{i,t},\xi_{i,t}\right|\varsigma)}. \tag{26}$$

If $\gamma < 1$, the inequality is reversed and the young individual chooses the flexible education if

$$\frac{\sigma\omega^2}{2\alpha_1} < \frac{1}{1-\gamma\left(x_{i,t},\xi_{i,t}\right|\varsigma)} \log\left(\frac{n_t^F}{n_t^R}\right) + \frac{\varrho \log \nu_t}{1-\gamma\left(x_{i,t},\xi_{i,t}\right|\varsigma)}. \tag{27}$$

These inequalities provide some insight about the educational choice of young individuals.[11] They are more likely to choose the flexible education which insures them against aggregate shocks if the variance of the aggregate shock is larger, if they are more risk averse, if the aggregate shock at the time of the decision is lower as long as $\varrho > 0$, and if the elasticity of the wage for the rigid education with respect to the aggregate shock is less negative.

Similarly to the first example, we can classify some of the variables and some of the parameters as belonging to the cross-sectional submodel and the remaining to the time-series submodel. The cross-sectional variables include consumption $c_{i,t}$, leisure $l_{i,t}$, individual wages $w_{i,t}^F$ and $w_{i,t}^R$, the variable determining the educational choice $D_{i,t}$, the amount of time an individual can divide between leisure and productive activities $\mathcal{T}$, and the variables that enter the risk aversion parameter $x_{i,t}$. The time-series variables are composed of the aggregate shock ν_t, the measure of young individuals choosing the two types of education n^F and n^R, and the aggregate equilibrium wages in the two sectors $w_t^F = E\left[w_{it}^F\right]$ and $w_t^R = E\left[w_{it}^F\right]$.[12] We want to stress the difference between individual wages and aggregate wages. Individual wages are typically observed in panel data or repeated cross-sections whose time dimension is generally short, whereas aggregate wages are available in longer time-series of aggregate data. The cross-sectional parameters consist of the relative taste for consumption σ and the parameters of the wage equations α_0 and α_1, whereas the time-series parameters include the two parameters governing the evolution of the aggregate shock ϱ and ω^2, the parameters defining the risk aversion ς, and the discount factor β. The discount

[11] Equations (26) and (27) can also be used to illustrate the problems of achieving consistent estimation using cross-sectional variation alone. See the appendix for details.

[12] The expectation operator E corresponds to the expectation taken over the distribution of cross sectional variables.

factor is notoriously difficult to estimate. For this reason, in the rest of the section we will assume it is known.

We can now consider the estimation of the parameters of interest. Parameters in this model can be consistently estimated exploiting both cross-sectional and time-series data. We assume that the (repeated) cross-sectional data include $\bar{n}_1$ and $\bar{n}_2$ i.i.d. observations $\left(w_{i,t}^F, c_{i,t}^*, l_{i,1}^*\right)$ for individuals with $S = F, R$ from two periods $t = 1, 2$. We also assume that the time-series data span $t = 1, \ldots, \tau$, and consists of $\left(n_t^F, w_t^F, n_t^R, w_t^R\right)$.

We first discuss how α_1 can be consistently estimated with a large number of individuals using the wage equation for the flexible education (24). Using equation (24), we can consistently estimate α_1 by $\widehat{\alpha}_1$ which solves

$$\frac{1}{\bar{n}_1}\sum_{i=1}^{\bar{n}_1} \ln w_{i,1}^F - \frac{1}{\bar{n}_2}\sum_{i=1}^{\bar{n}_2} \ln w_{i,2}^F = \frac{1}{\widehat{\alpha}_1}\left(n_1^F - n_2^F\right).$$

Observe that this can be done because ε_t and γ are assumed to be independent of each other, which implies that there is no sample selectivity problem. Second, the consumption and leisure choices of working-age individuals (21) and (22) can be used to consistently estimate σ by $\widehat{\sigma}$ which solves

$$\frac{1}{\bar{n}_1}\sum_{i=1}^{\bar{n}_1} \frac{c_{i,1}^*}{l_{i,1}^*} = w_1^F \frac{\widehat{\sigma}}{1-\widehat{\sigma}}.$$

Third, with α_1 consistently estimated, it is straightforward to show that the aggregate shock in period t can be consistently estimated for $t = 1, \ldots, \tau$ using the following equation:

$$\widehat{\ln \nu_t} = \widehat{\alpha}_1\left(\ln w_t^F - \ln w_t^R\right) - \left(\ln n_t^F - \ln n_t^R\right). \tag{28}$$

The parameters ϱ and ω^2 can then be consistently estimated by the time-series regression of the following equation:

$$\widehat{\log \nu_{t+1}} = \varrho \widehat{\log \nu_t} + \eta_t. \tag{29}$$

This is the step where we use the time-series variation. The only parameters left to estimate are the ς defining the individual risk aversion $\gamma\left(x_{i,t}, \xi_{i,t} \middle| \varsigma\right)$. If the distribution of ξ is parametrically specified, those parameters can be consistently estimated by MLE using cross-sectional variation on the educational choices and inequalities (26) and (27). Note that we were able to consistently estimate ς only because ϱ and ω^2 had been previously estimated using time-series variation. Hence, the bi-directional relationship between the cross-sectional and time-series submodels.

The previous discussion illustrates the importance of combining cross-sectional data with a long time-series of aggregate data. One alternative would be to exploit panel data in which the time-series dimension of the panel is sufficiently long. This alternative has, however, a potential drawback. In Appendix C, we argue that a "long panel" approach is mathematically equivalent to time-series analysis where T goes to infinity, while n stays fixed. The standard errors in the long panel analysis should, therefore, be based exclusively on the time series variation. This discussion has an implication for the standard errors computed in Lee and Wolpin's (2006) (see their footnote 37). Donghoon Lee, in private communication, kindly informed us that their standard errors do not account for the noise introduced by the estimation of the time series parameters, i.e. the standard errors reported in the paper assume that the time-series parameters are known or, equivalently, fixed at their estimated value. We also note that since almost all panel data sets have limited time-series dimension, using this alternative approach would lead to imprecise estimates.

When cross-sectional data are combined with long time-series of aggregate data, the standard formulas for the computation of test statistics and confidence intervals are no longer valid. In the next section, we provide new easy-to-use formulas that can be employed for coefficients estimated by combining those two data sources. The formal derivation of those formulas is contained in the companion paper Hahn, Kuersteiner, and Mazzocco (2016).

6 Standard Errors

In this section, after the derivation of the formulas required for the computation of test statistics and confidence intervals of coefficients estimated using a combination of time-series and cross-sectional data, we will explain how they can be employed in concrete cases using, as an example, the general equilibrium model developed in the previous section.

6.1 Formulas for Test Statistics and Confidence Intervals

The asymptotic theory underlying the estimators obtained from the combination of the two data sources considered in this paper is complex. It is based on a new central limit theorem that requires a novel martingale representation. Given its complexity, the theory is presented in a separate paper (Hahn, Kuersteiner and Mazzocco (2016)). However, the mechanical implementation of the formulas required for the computation of test statistics and confidence intervals is straightforward. In the rest of this subsection we provide the step-by-step description of how those formulas can be calculated.

The computation starts with the explicit characterization of the "moments" that identify the parameters. Let $\theta = (\beta, \nu_1, ..., \nu_T)$ and denote with $f_{\theta,i}(\theta, \rho)$ and $g_{\rho,t}(\beta, \rho)$ the i-th and t-th moments used in the identification of the cross-sectional and time-series parameters. Suppose that our estimator can be written as the solution to the following system of equations:

$$\sum_{i=1}^{n} f_{\theta,i}\left(\hat{\theta}, \hat{\rho}\right) = 0, \tag{30}$$

$$\sum_{t=\tau_0+1}^{\tau_0+\tau} g_{\rho,t}\left(\hat{\beta}, \hat{\rho}\right) = 0. \tag{31}$$

Formulas can then be calculated using the following steps:

1. Let $\phi = (\theta', \rho')'$ be the vector of parameters.

2. Let

$$\mathbf{A} = \begin{bmatrix} \hat{A}_{y,\theta} & \hat{A}_{y,\rho} \\ \hat{A}_{\nu,\theta} & \hat{A}_{\nu,\rho} \end{bmatrix},$$

with

$$\hat{A}_{y,\theta} = n^{-1} \sum_{i=1}^{n} \frac{\partial f_{\theta,i}\left(\hat{\theta}, \hat{\rho}\right)}{\partial \theta'}, \qquad \hat{A}_{y,\rho} = n^{-1} \sum_{i=1}^{n} \frac{\partial f_{\theta,i}\left(\hat{\theta}, \hat{\rho}\right)}{\partial \rho'},$$

$$\hat{A}_{\nu,\theta} = \tau^{-1} \sum_{t=\tau_0+1}^{\tau_0+\tau} \frac{\partial g_{\rho,t}\left(\hat{\beta}, \hat{\rho}\right)}{\partial \theta'}, \qquad \hat{A}_{\nu,\rho} = \tau^{-1} \sum_{t=\tau_0+1}^{\tau_0+\tau} \frac{\partial g_{\rho,t}\left(\hat{\beta}, \hat{\rho}\right)}{\partial \rho'}.$$

3. Let

$$\hat{\Omega}_y = \frac{1}{n} \sum_{i=1}^{n} f_{\theta,i}\left(\hat{\theta}, \hat{\rho}\right) f_{\theta,i}\left(\hat{\theta}, \hat{\rho}\right)'$$

and

$$\hat{\Omega}_v = \frac{1}{n} \sum_{i=1}^{n} g_{\rho,t}\left(\hat{\theta}, \hat{\rho}\right) g_{\rho,t}\left(\hat{\theta}, \hat{\rho}\right)'$$

4. Let

$$W = \begin{bmatrix} \frac{1}{n}\hat{\Omega}_y & 0 \\ 0 & \frac{1}{\tau}\hat{\Omega}_\nu \end{bmatrix}$$

5. Calculate

$$\mathbf{V} = \mathbf{A}^{-1} W \left(\mathbf{A}'\right)^{-1}$$

and use it as the "variance" (not the asymptotic variance) of the estimator. For instance, if one is interested in the 95% confidence interval of the first component of ϕ, it can be written as $\hat{\phi}_1 \pm 1.96\sqrt{\mathbf{V}_{1,1}}$

6.2 Formulas Applied to the General Equilibrium Model

To apply the five steps described in the previous subsection to the general equilibrium model, we only have to derive the moment conditions used in its estimation $f_{\theta,i}(\theta,\rho)$ and $g_{\rho,t}(\beta,\rho)$.

For simplicity of notation, we assume that $\bar{n}_1 = \bar{n}_2 = n$. Also, we denote by $F_{i,t}$ a dummy variable that takes the value 1 if the flexible type of education is chosen and 0 otherwise. From the discussion in Section 5, it follows directly that the moments employed in the estimation of α_1 and σ take the following form:

$$\sum_i \left(\ln w_{i,1}^F - \ln w_{i,2}^F - \frac{1}{\alpha_1}\left(n_1^F - n_2^F\right) \right) = 0$$

and

$$\sum_i \left(\frac{c_{i,1}^*}{l_{i,1}^*} - w_1^F \frac{\sigma}{1-\sigma} \right) = 0.$$

For the estimation of the parameters ρ and ω^2, equation (29) implies that the OLS estimator of ϱ and the corresponding estimator for ω^2 solve:

$$\frac{1}{\tau}\sum_t \widehat{\log \nu_t}\left(\widehat{\log \nu_{t+1}} - \widehat{\varrho}\widehat{\log \nu_t}\right) = 0$$

and

$$\frac{1}{\tau}\sum_t \left(\widehat{\log \nu_{t+1}} - \widehat{\varrho}\widehat{\log \nu_t}\right)^2 = \widehat{\omega}^2.$$

Replacing for $\widehat{\log \nu_{t+1}}$ and $\widehat{\log \nu_t}$ using equation(28), we obtain the following two moment conditions:

$$\sum_t \begin{pmatrix} \alpha_1\left(\ln w_t^F - \ln w_t^R\right) \\ -\left(\ln n_t^F - \ln n_t^R\right) \end{pmatrix} \left(\begin{pmatrix} \alpha_1\left(\ln w_{t+1}^F - \ln w_{t+1}^R\right) \\ -\left(\ln n_{t+1}^F - \ln n_{t+1}^R\right) \end{pmatrix} - \varrho \begin{pmatrix} \alpha_1\left(\ln w_t^F - \ln w_t^R\right) \\ -\left(\ln n_t^F - \ln n_t^R\right) \end{pmatrix} \right) = 0$$

$$\sum_t \left(\left(\begin{pmatrix} \alpha_1\left(\ln w_{t+1}^F - \ln w_{t+1}^R\right) \\ -\left(\ln n_{t+1}^F - \ln n_{t+1}^R\right) \end{pmatrix} - \varrho \begin{pmatrix} \alpha_1\left(\ln w_t^F - \ln w_t^R\right) \\ -\left(\ln n_t^F - \ln n_t^R\right) \end{pmatrix} \right)^2 - \omega^2 \right) = 0$$

To derive the moments employed in the estimation of the risk aversion parameters, we need an exact expression for $\gamma\left(x_{i,t},\xi_{i,t}|\,\varsigma\right)$ and select a distribution for $\xi_{i,t}$, which require some assumptions. We will consider the case $1-\gamma\geq 0$. Similar arguments can be used if $1-\gamma<0$. To derive a close-form for $\gamma\left(x_{i,t},\xi_{i,t}|\,\varsigma\right)$, inequality (26) can alternatively be written as[13]

$$\gamma\left(x_{i,t},\xi_{i,t}|\,\varsigma\right)\leq 1-\frac{2\alpha_1}{\sigma\omega^2}\left(\log\left(\frac{n_t^F}{n_t^R}\right)+\varrho\log\nu_t\right).$$

Suppose we consider the following parameterization for $\gamma\left(x_{i,t},\xi_{i,t}|\,\varsigma\right)$:

$$\gamma\left(x_{i,t},\xi_{i,t}|\,\varsigma\right)=G\left(x_{i,t}'\varsigma+\xi_{i,t}\right),$$

where G is some monotonically increasing function bounded above by 1 and $\xi_{i,t}\sim N\left(0,1\right)$. We can then conclude that an individual chooses the flexible education ($F_{i,t}=1$) if and only if

$$x_{i,t}'\varsigma+\xi_{i,t}\leq G^{-1}\left(1-\frac{2\alpha_1}{\sigma\omega^2}\left(\log\left(\frac{n_t^F}{n_t^R}\right)+\varrho\log\nu_t\right)\right)$$

or, equivalently,

$$\xi_{i,t}\leq G^{-1}\left(1-\frac{2\alpha_1}{\sigma\omega^2}\left(\log\left(\frac{n_t^F}{n_t^R}\right)+\varrho\log\nu_t\right)\right)-x_{i,t}'\varsigma.$$

The probability that $F_{i,t}=1$ is therefore given by the following expression:

$$\Phi\left[G^{-1}\left(1-\frac{2\alpha_1}{\sigma\omega^2}\left(\log\left(\frac{n_t^F}{n_t^R}\right)+\varrho\log\nu_t\right)\right)-x_{i,t}'\varsigma\right],$$

where Φ and ϕ denote the CDF and PDF of a $N\left(0,1\right)$.

Despite the complicated nature of the probability, the whole expression is linear in ς implying that it is a special case of a textbook probit. The First Order Condition (FOC) derived from the

[13]For $1-\gamma\geq 0$, the inequality (26) can alternatively be written as

$$\frac{\sigma\omega^2}{2\alpha_1}\geq\frac{\log\left(\frac{n_t^F}{n_t^R}\right)+\varrho\log\nu_t}{1-\gamma\left(x_{i,t},\xi_{i,t}|\,\varsigma\right)}$$

Using $1-\gamma\left(x_{i,t},\xi_{i,t}|\,\varsigma\right)\geq 0$, we get

$$1-\gamma\left(x_{i,t},\xi_{i,t}|\,\varsigma\right)\geq\frac{2\alpha_1}{\sigma\omega^2}\left(\log\left(\frac{n_t^F}{n_t^R}\right)+\varrho\log\nu_t\right)$$

or

$$\gamma\left(x_{i,t},\xi_{i,t}|\,\varsigma\right)\leq 1-\frac{2\alpha_1}{\sigma\omega^2}\left(\log\left(\frac{n_t^F}{n_t^R}\right)+\varrho\log\nu_t\right).$$

maximization of that probability with respect to ς takes the form

$$\sum_i \frac{F_{i,t} - \Phi\left[G^{-1}\left(1 - \frac{2\alpha_1}{\sigma\omega^2}\left(\log\left(\frac{n_t^F}{n_t^R}\right) + \varrho\log\nu_t\right)\right) - x'_{i,t}\varsigma\right]}{\Phi\left[G^{-1}\left(1 - \frac{2\alpha_1}{\sigma\omega^2}\left(\log\left(\frac{n_t^F}{n_t^R}\right) + \varrho\log\nu_t\right)\right) - x'_{i,t}\varsigma\right]\left\{1 - \Phi\left[G^{-1}\left(1 - \frac{2\alpha_1}{\sigma\omega^2}\left(\log\left(\frac{n_t^F}{n_t^R}\right) + \varrho\log\nu_t\right)\right) - x'_{i,t}\varsigma\right]\right\}} \times$$

$$\times\phi\left[G^{-1}\left(1 - \frac{2\alpha_1}{\sigma\omega^2}\left(\log\left(\frac{n_t^F}{n_t^R}\right) + \varrho\log\nu_t\right)\right) - x'_{i,t}\varsigma\right]x_{i,t} = 0$$

Using the equation

$$\ln\nu_t = \alpha_1\left(\ln w_t^F - \ln w_t^R\right) - \left(\ln n_t^F - \ln n_t^R\right),$$

and defining

$$Z_t\left(\alpha_1, \sigma, \varrho, \omega^2\right) \equiv G^{-1}\left(1 - \frac{2\alpha_1}{\sigma\omega^2}\left(\log\left(\frac{n_t^F}{n_t^R}\right) + \varrho\left(\begin{array}{c}\alpha_1\left(\ln w_t^F - \ln w_t^R\right)\\ -\left(\ln n_t^F - \ln n_t^R\right)\end{array}\right)\right)\right),$$

we can rewrite the first order condition as follows:

$$\sum_i \frac{F_{i,t} - \Phi\left[Z_t\left(\alpha_1, \sigma, \varrho, \omega^2\right) - x'_{i,t}\varsigma\right]}{\Phi\left[Z_t\left(\alpha_1, \sigma, \varrho, \omega^2\right) - x'_{i,t}\varsigma\right]\left\{1 - \Phi\left[Z_t\left(\alpha_1, \sigma, \varrho, \omega^2\right) - x'_{i,t}\varsigma\right]\right\}}\phi\left[Z_t\left(\alpha_1, \sigma, \varrho, \omega^2\right) - x'_{i,t}\varsigma\right]x_{i,t} = 0,$$

which corresponds to the moments used in the estimation of the risk aversion parameters.

Assuming without loss of generality that only the cross sections at $t = 1$ and $t = 2$ are used in the estimation, the previous discussion implies that the vectors of moments $f_{\theta,i}$ and $g_{\rho,t}$ take the following form:

$$f_{\theta,i}(\theta, \rho) = \left[\begin{array}{c}\ln w_{i,1}^F - \ln w_{i,2}^F - \frac{1}{\alpha_1}\left(n_t^F - n_{t-1}^F\right)\\ \frac{c_{i,1}^*}{l_{i,1}^*} - w_1^F\frac{\sigma}{1-\sigma}\\ \frac{F_{i,1} - \Phi\left[Z_1\left(\alpha_1,\sigma,\varrho,\omega^2\right) - x'_{i,1}\varsigma\right]}{\Phi\left[Z_1(\alpha_1,\sigma,\varrho,\omega^2) - x'_{i,1}\varsigma\right]\left\{1 - \Phi\left[Z_1(\alpha_1,\sigma,\varrho,\omega^2) - x'_{i,1}\varsigma\right]\right\}}\phi\left[Z_1\left(\alpha_1, \sigma, \varrho, \omega^2\right) - x'_{i,1}\varsigma\right]x_{i,1}\end{array}\right],$$

and

$$g_{\rho,t}(\beta, \rho) = \left[\begin{array}{c}\left(\begin{array}{c}\alpha_1\left(\ln w_t^F - \ln w_t^R\right)\\ -\left(\ln n_t^F - \ln n_t^R\right)\end{array}\right)\left(\left(\begin{array}{c}\alpha_1\left(\ln w_{t+1}^F - \ln w_{t+1}^R\right)\\ -\left(\ln n_{t+1}^F - \ln n_{t+1}^R\right)\end{array}\right) - \varrho\left(\begin{array}{c}\alpha_1\left(\ln w_t^F - \ln w_t^R\right)\\ -\left(\ln n_t^F - \ln n_t^R\right)\end{array}\right)\right)\\ \left(\left(\begin{array}{c}\alpha_1\left(\ln w_{t+1}^F - \ln w_{t+1}^R\right)\\ -\left(\ln n_{t+1}^F - \ln n_{t+1}^R\right)\end{array}\right) - \varrho\left(\begin{array}{c}\alpha_1\left(\ln w_t^F - \ln w_t^R\right)\\ -\left(\ln n_t^F - \ln n_t^R\right)\end{array}\right)\right)^2 - \omega^2\end{array}\right],$$

and the parameter estimates of the general equilibrium model are the solution to the following system of equations:

$$n^{-1/2}\sum_{i=1}^{n} f_{\theta,i}\left(\hat{\theta},\hat{\rho}\right)=0,$$

$$\tau^{-1/2}\sum_{t=\tau_0+1}^{\tau_0+\tau} g_{\rho,t}\left(\hat{\beta},\hat{\rho}\right)=0.$$

We are now ready to describe the five steps required in the computation of test statistics and confidence intervals for the general equilibrium model. As a first step, let $\theta=\beta=(\alpha_1,\sigma,\varsigma)$ and $\rho=(\varrho,\omega^2)$. Observe that the aggregate shock is not in the set of estimated parameters, since the general equilibrium model implies that $\ln\nu_t=\alpha_1\left(\ln w_t^F-\ln w_t^R\right)-\left(\ln n_t^F-\ln n_t^R\right)$. In the second, third, and fourth steps compute the matrices A, $\hat{\Omega}_y$, $\hat{\Omega}_v$, and W using the vectors of moments $f_{\theta,i}$ and $g_{\rho,t}$ derived above. In the last step, calculate the variance matrix $V=A^{-1}W\left(\mathbf{A}'\right)^{-1}$.

7 Summary

Using a general econometric framework and three examples we have shown that generally, when aggregate shocks are present, model parameters cannot be identified using cross-sectional variation alone. Identification of those parameters requires the combination of cross-sectional and time-series data. When those two data sources are jointly used, standard formulas for the computation of test statistics and confidence intervals are no longer valid. We provide new easy-to-use formulas that account for the interaction between the time-series and cross-sectional data. Our results are expected to be helpful for the econometric analysis of rational expectations models involving individual decision making as well general equilibrium models.

References

[1] Ackerberg, D.A., K. Caves, and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators," Econometrica 83, 2411–2451.

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

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

[4] Arellano, M., R. Blundell, and S. Bonhomme (2014): "Household Earnings and Consumption: A Nonlinear Framework," unpublished working paper.

[5] Bajari, Patrick, Benkard, C. Lanier, and Levin, Jonathan (2007): "Estimating Dynamic Models of Imperfect Competition," Econometrica 75, 5, pp. 1331-1370.

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

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

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

[9] Chamberlain, G. (1984): "Panel Data," in Handbook of Econometrics, eds. by Z. Griliches and M. Intriligator. Amsterdam: North Holland, pp. 1247-1318.

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

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

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

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

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

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

[16] Gagliardini, P., and C. Gourieroux (2011): "Efficiency in Large Dynamic Panel Models with Common Factor," unpublished working paper.

[17] Gemici, A., and M. Wiswall (2014): "Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors," International Economic Review 55, 23–56.

[18] Gillingham, K., F. Iskhakov, A. Munk-Nielsen, J. Rust, and B. Schjerning (2015): "A Dynamic Model of Vehicle Ownership, Type Choice, and Usage," unpublished working paper.

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

[20] Hahn, J., and W.K. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," Econometrica 72, pp. 1295–1319.

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

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

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

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

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

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

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

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

[29] Lee, D. (2005): “An Estimable Dynamic General Equilibrium Model of Work, Schooling and Occupational Choice,” International Economic Review 46, pp. 1-34.

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

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

[32] Magnac, T. and D. Thesmar (2002): “Identifying Dynamic Discrete Decision Processes,” Econometrica 70, pp. 801–816.

[33] Magnus, J.R. and H. Neudecker (1988): “Matrix Differential Calculus with Applications in Statistics and Econometrics,” Wiley.

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

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

[36] Olley, G.S., and A. Pakes (1996): “The Dynamics of Productivity in the Telecommunications Equipment Industry,” Econometrica 64, pp. 1263 – 1297.

[37] Peccati, G. and M.S. Taqqu (2007): "Stable convergence of generalized L^2stochastic integrals and the principle of conditioning," Electronic Journal of Probability 12, pp.447-480.

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

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

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

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

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

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

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

[45] Runkle, D.E. (1991): "Liquidity Constraints and the Permanent-Income Hypothesis: Evidence from Panel Data," Journal of Monetary Economics 27, pp. 73–98.

[46] Shea, J. (1995):"Union Contracts and the Life-Cycle/Permanent-Income Hypothesis," American Economic Review 85, pp. 186–200.

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

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

A Discussion for Section 3

A.1 Proof of (2)

The maximization problem is equivalent to

$$\max_{\alpha} \; -e^{-\delta(\alpha(1+r)+(1-\alpha))} E\left[e^{-\delta(1-\alpha)u_{i,t}}\right].$$

Since $-\delta(1-\alpha)u_{i,t} \sim N\left(-\delta(1-\alpha)\mu, \delta^2(1-\alpha)^2\sigma^2\right)$, we have

$$E\left[e^{-\delta(1-\alpha)u_{i,t}}\right] = e^{-\delta(1-\alpha)\mu + \frac{\delta^2(1-\alpha)^2\sigma^2}{2}},$$

and the maximization problem can be rewritten as follows:

$$\max_{\alpha} \; -e^{-\delta\left(\alpha(1+r)+(1-\alpha)(1+\mu)-\frac{\delta(1-\alpha)^2\sigma^2}{2}\right)}.$$

Taking the first order condition, we have,

$$0 = -\delta\left(r - \mu + \sigma^2\delta - \alpha\sigma^2\delta\right)$$

from which we obtain the solution

$$\alpha = \frac{1}{\sigma^2\delta}\left(r - \mu + \sigma^2\delta\right).$$

A.2 Euler Equation and Cross Section

Our model in Section 3 is a stylized version of many models considered in a large literature interested in estimating the parameter δ using cross-sectional variation. Estimators are often based on moment conditions derived from first order conditions (FOC) related to optimal investment and consumption decisions. We illustrate the problems facing such estimators.

Assume a researcher has a cross-section of observations for individual consumption and returns $c_{i,t}$ and $u_{i,t}$. The population FOC of our model[14] takes the simple form $E\left[e^{-\delta c_{i,t}}(r - u_{i,t})\right] = 0$. A

[14] We assume $\delta \neq 0$ and rescale the equation by $-\delta^{-1}$.

just-identified moment based estimator for δ solves the sample analog $n^{-1}\sum_{i=1}^{n} e^{-\hat{\delta}c_{i,t}}\left(r-u_{i,t}\right)=0$. It turns out that the probability limit of $\hat{\delta}$ is equal to $\left(\nu_t-r\right)/\left(\left(1-\alpha\right)\sigma_\epsilon^2\right)$, i.e., $\hat{\delta}$ is inconsistent.

We now compare the population FOC a rational agent uses to form their optimal portfolio with the empirical FOC an econometrician using cross-sectional data observes:

$$n^{-1}\sum_{i=1}^{n} e^{-\delta c_{i,t}}\left(r-u_{i,t}\right)=0.$$

Noting that $u_{i,t}=\nu_t+\epsilon_{i,t}$ and substituting into the budget constraint

$$c_{i,t}=1+\alpha r+\left(1-\alpha\right)u_{i,t}=1+\alpha r+\left(1-\alpha\right)\nu_t+\left(1-\alpha\right)\epsilon_{i,t}$$

we have

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} e^{-\delta c_{i,t}}\left(r-u_{i,t}\right) &= n^{-1}\sum_{i=1}^{n} e^{-\delta\left(1+\alpha r+\left(1-\alpha\right)\nu_t\right)-\delta\left(1-\alpha\right)\epsilon_{i,t}}\left(r-\nu_t-\epsilon_{i,t}\right) \\ &= e^{-\delta\left(1+\alpha r+\left(1-\alpha\right)\nu_t\right)}\left(\left(r-\nu_t\right)n^{-1}\sum_{i=1}^{n} e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}-n^{-1}\sum_{i=1}^{n} e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}\epsilon_{i,t}\right). \end{aligned} \tag{32}$$

Under suitable regularity conditions including independence of $\epsilon_{i,t}$ in the cross-section it follows that

$$n^{-1}\sum_{i=1}^{n} e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}=E\left[e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}\right]+o_p\left(1\right)=e^{\frac{\delta^2\left(1-\alpha\right)^2\sigma_\epsilon^2}{2}}+o_p\left(1\right) \tag{33}$$

and

$$n^{-1}\sum_{i=1}^{n} e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}\epsilon_{i,t}=E\left[e^{-\delta\left(1-\alpha\right)\epsilon_{i,t}}\epsilon_{i,t}\right]+o_p\left(1\right)=-\delta\left(1-\alpha\right)\sigma_\epsilon^2 e^{\frac{\delta^2\left(1-\alpha\right)^2\sigma_\epsilon^2}{2}}+o_p\left(1\right). \tag{34}$$

Taking limits as $n\to\infty$ in (32) and substituting (33) and (34) then shows that the method of moments estimator based on the empirical FOC asymptotically solves

$$\left(\left(r-\nu_t\right)+\delta\left(1-\alpha\right)\sigma_\epsilon^2\right)e^{\frac{\delta^2\left(1-\alpha\right)^2\sigma_\epsilon^2}{2}}=0. \tag{35}$$

Solving for δ we obtain

$$\operatorname{plim}\hat{\delta}=\frac{\nu_t-r}{\left(1-\alpha\right)\sigma_\epsilon^2}.$$

This estimate is inconsistent because the cross-sectional data set lacks cross sectional ergodicity, or in other words does not contain the same information about aggregate risk as is used by rational agents. Therefore, the empirical version of the FOC is unable to properly account for aggregate risk and return characterizing the risky asset. The estimator based on the FOC takes the form of an implicit solution to an empirical moment equation, which obscures the effects of cross-sectional non-ergodicity. A more illuminative approach uses our modelling strategy in Section 2.

On the other hand, it is easily shown using properties of the Gaussian moment generating function that the population FOC is proportional to

$$E\left[e^{-\delta(1-\alpha)u_{i,t}}\left(r-u_{i,t}\right)\right]=\left(r-\mu+\delta\left(1-\alpha\right)\sigma^{2}\right)e^{-\delta(1-\alpha)\mu+\frac{\delta^{2}(1-\alpha)^{2}\sigma^{2}}{2}}=0. \tag{36}$$

The main difference between (33) and (34) lies in the fact that σ_v^2 is estimated to be 0 in the sample and that $\nu_t \neq \mu$ in general. Note that (36) implies that consistency may be achieved with a large number of repeated cross sections, or a panel data set with a long time series dimension. However, this raises other issues discussed later in Section C.

B Details of Section 5

B.1 Proof of Inequalities (26) and (27)

In the proof we will drop the i subscripts for notational purposes. We can rewrite (23) as follows:

$$\begin{aligned}&\int\frac{\left[\left(\sigma w_{t+1}^{F}\left(\nu_{t+1}\right)\exp\left(\varepsilon_{t+1}\right)T\right)^{\sigma}\left(\left(1-\sigma\right)T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma}dP\left(\nu_{t+1}\left|\nu_{t}\right.\right)\\ &\geq\int\frac{\left[\left(\sigma w_{t+1}^{R}\left(\nu_{t+1}\right)\exp\left(\varepsilon_{t+1}\right)T\right)^{\sigma}\left(\left(1-\sigma\right)T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma}dP\left(\nu_{t+1}\left|\nu_{t}\right.\right),\end{aligned}$$

As a consequence, education F is chosen if

$$\psi\left(\gamma,\nu_{t}\right)\equiv\int\left[\left(w_{t+1}^{F}\left(\nu_{t+1}\right)\right)^{\sigma}\right]^{1-\gamma}dP\left(\nu_{t+1}\left|\nu_{t}\right.\right)-\int\left[w_{t+1}^{R}\left(\nu_{t+1}\right)^{\sigma}\right]^{1-\gamma}dP\left(\nu_{t+1}\left|\nu_{t}\right.\right)\geq0 \tag{37}$$

We rewrite the value function of an old individual with education F

$$V_{t}\left(F,\nu_{t}\right)=\frac{\left[\left(\left(\frac{n^{F}\sigma T\exp\left(\varepsilon_{t}\right)}{e^{\alpha_{0}}}\right)^{1/\alpha_{1}}\sigma T\right)^{\sigma}\left(\left(1-\sigma\right)T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma}$$

Likewise, the value function of an old individual with education R takes a similar form:

$$V_{t}\left(R,\nu_{t}\right)=\frac{\left[\left(\left(\frac{n^{R}\sigma T\exp\left(\varepsilon_{t}\right)}{e^{\alpha_{0}}\nu_{t}}\right)^{1/\alpha_{1}}\sigma T\right)^{\sigma}\left(\left(1-\sigma\right)T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma}$$

We can now describe the solution to the problem of a young worker. Given the assumptions, optimal consumption and leisure in the first period can be easily computed to be:

$$c_t^* = y_t - C_e,$$
$$l_t^* = T - T_e.$$

This implies that current consumption and leisure are independent of the education choice and of the aggregate shock. As a consequence, the current utility will also be independent of the education choice and of the aggregate shock. The education choice will therefore only depend on the utility when old. Specifically, the individual will choose education F if

$$\int V_t\left(F, \nu_{t+1}\right) dP\left(\nu_{t+1}\left|\nu_t\right.\right) \geq \int V_t\left(R, \nu_{t+1}\right) dP\left(\nu_{t+1}\left|\nu_t\right.\right).$$

Write

$$V_t\left(F, \nu_{t+1}\right) = \frac{\left[\left(\left(\frac{n^R \sigma T}{e^{\alpha_0}}\right)^{1/\alpha_1} \sigma T\right)^{\sigma} \left(\left(1-\sigma\right)T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma} \exp\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\varepsilon_{t+1}\right)$$

$$V_t\left(R, \nu_{t+1}\right) = \frac{\left[\left(\left(\frac{n^R \sigma T}{e^{\alpha_0}}\right)^{1/\alpha_1} \sigma T\right)^{\sigma} \left(\left(1-\sigma\right)T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma} \exp\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\varepsilon_{t+1}\right)\left(\nu_{t+1}^{-\sigma(1/\alpha_1)(1-\gamma)}\right)$$

We see that education F is chosen if and only if

$$\frac{\left[\left(\left(\frac{n^F \sigma T}{e^{\alpha_0}}\right)^{1/\alpha_1} \sigma T\right)^{\sigma} \left(\left(1-\sigma\right)T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma} \exp\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\varepsilon_{t+1}\right) \geq$$
$$\frac{\left[\left(\left(\frac{n^R \sigma T}{e^{\alpha_0}}\right)^{1/\alpha_1} \sigma T\right)^{\sigma} \left(\left(1-\sigma\right)T\right)^{1-\sigma}\right]^{1-\gamma}}{1-\gamma} \exp\left(\sigma\left(1/\alpha_1\right)\left(1-\gamma\right)\varepsilon_{t+1}\right) E_t\left[\nu_{t+1}^{-\sigma(1/\alpha_1)(1-\gamma)}\right]$$

(We made use of the assumption that ε_{t+1} is known to the workers.) This is equivalent to

$$\left(n^F\right)^{\sigma(1-\gamma)/\alpha_1} \geq \left(n^R\right)^{\sigma(1-\gamma)/\alpha_1} E_t\left[\nu_{t+1}^{-\sigma(1-\gamma)(1/\alpha_1)}\right] \qquad \text{if} \qquad 1-\gamma \geq 0 \tag{38}$$

and to

$$\left(n^F\right)^{\sigma(1-\gamma)/\alpha_1} < \left(n^R\right)^{\sigma(1-\gamma)/\alpha_1} E_t\left[\nu_{t+1}^{-\sigma(1-\gamma)(1/\alpha_1)}\right] \qquad \text{if} \qquad 1-\gamma < 0 \tag{39}$$

Because $\log \nu_{t+1} = \rho \log \nu_t + \eta_t$, or

$$\nu_{t+1} = \nu_t^{\rho} \exp(\eta_t)$$

we can write

$$E_t\left[\nu_{t+1}^{-\sigma(1-\gamma)(1/\alpha_1)}\right] = E_\eta\left[\left(\nu_t^{\rho} \exp(\eta_t)\right)^{-\sigma(1-\gamma)(1/\alpha_1)}\right] = \nu_t^{-\rho\sigma(1-\gamma)(1/\alpha_1)} E\left[\exp\left(-\sigma(1-\gamma)(1/\alpha_1)\eta_t\right)\right]$$

where $E_\eta[\cdot]$ denotes the integral with respect to η_t alone. The assumption that $\eta_t \sim N(0, \omega^2)$ allows us to write

$$E\left[\exp\left(-\sigma(1-\gamma)(1/\alpha_1)\eta_t\right)\right] = \exp\left(\frac{\left(\sigma\omega(1-\gamma)(1/\alpha_1)\right)^2}{2}\right)$$

recognizing that the expectation on the left is nothing but the moment generating function of $N(0, \omega^2)$ evaluated at $-\sigma(1-\gamma)(1/\alpha_1)$. Therefore, we have

$$E_t\left[\nu_{t+1}^{-\sigma(1-\gamma)(1/\alpha_1)}\right] = \nu_t^{-\rho\sigma(1-\gamma)(1/\alpha_1)} \exp\left(\frac{\left(\sigma\omega(1-\gamma)(1/\alpha_1)\right)^2}{2}\right) \tag{40}$$

Consider first the case $1-\gamma \geq 0$. Combining (38) and (40), we can rewrite the decision as

$$\left(n^F\right)^{\sigma(1-\gamma)/\alpha_1} \geq \left(n^R\right)^{\sigma(1-\gamma)/\alpha_1} \nu_t^{-\rho\sigma(1-\gamma)(1/\alpha_1)} \exp\left(\frac{\left(\sigma\omega(1-\gamma)(1/\alpha_1)\right)^2}{2}\right).$$

Taking logs, we obtain

$$\frac{\sigma(1-\gamma)}{\alpha_1} \log n^F \geq \frac{\sigma(1-\gamma)}{\alpha_1} \log n^R - \rho\frac{\sigma(1-\gamma)}{\alpha_1} \log \nu_t + \frac{\left(\sigma\omega(1-\gamma)\right)^2}{2\alpha_1^2}$$

Dividing by σ and multiplying by $\alpha_1 < 0$, we conclude that the decision is equivalent to

$$(1-\gamma)\log n^F \leq (1-\gamma)\log n^R - \rho(1-\gamma)\log \nu_t + \frac{\sigma(1-\gamma)^2\omega^2}{2\alpha_1}$$

or

$$-\rho(1-\gamma)\log \nu_t + \frac{\sigma(1-\gamma)^2\omega^2}{2\alpha_1} \geq (1-\gamma)\log\left(\frac{n^F}{n^R}\right).$$

which proves inequality (26). If $1-\gamma < 0$, following the same steps, we have

$$-\rho(1-\gamma)\log \nu_t + \frac{\sigma(1-\gamma)^2\omega^2}{2\alpha_1} < (1-\gamma)\log\left(\frac{n^F}{n^R}\right),$$

which proves inequality (27).

Remark 1 *Equations 26 and 27 hints at the problem of identification based on cross section variation alone. This is because the cross section variation does not identify ρ, which implies that γ is not identified as a consequence. Suppose that γ has a multinomial distribution, i.e., it has a finite support $\gamma_1^*, \ldots, \gamma_M^*$. Let ρ^* denote the true value of ρ. Let $\gamma_m(\rho)$ be defined by*

$$-\rho \log \nu_t + (1 - \gamma_m(\rho)) \frac{\sigma\omega^2}{2\alpha_1} = -\rho^* \log \nu_t + (1 - \gamma_m^*) \frac{\sigma\omega^2}{2\alpha_1}$$

More precisely, let

$$\gamma_m(\rho) = 1 - \frac{\rho \log \nu_t - \rho^* \log \nu_t + (1 - \gamma_m^*) \frac{\sigma\omega^2}{2\alpha_1}}{\frac{\sigma\omega^2}{2\alpha_1}}$$

We then have

$$-\rho \log \nu_t + (1 - \gamma_m(\rho)) \frac{\sigma\omega^2}{2\alpha_1} = -\rho^* \log \nu_t + (1 - \gamma_m^*) \frac{\sigma\omega^2}{2\alpha_1}$$

for range of possible values of ρ, and the education choice implied by $(\rho, \gamma_1(\rho), \ldots, \gamma_M(\rho))$ is identical to the one implied by the true value $(\rho^*, \gamma_1^*, \ldots, \gamma_M^*)$ of the parameter. The parameter $(\rho^*, \gamma_1^*, \ldots, \gamma_M^*)$ is not identified.

B.2 Proof of (24) and (25)

Note that individual heterogeneity is completely summarized by the vector $\chi_t \equiv (\varepsilon_t, \gamma)$, which means that we can define the labor supply $h_t^F(\chi)$ and $h_t^R(\chi)$ for each type χ of workers. We assume that the mass of individuals such that $(\varepsilon_t, \gamma) \in A$ for some $A \subset R^2$ is given by $N_t \int_A G(d\chi)$, where G is a joint CDF. For simplicity, we assume that G is such that the first and second components are independent of each other. We also assume that $\int \exp(\varepsilon_t) G(d\chi) = 1$.

The labor markets for both types of education must be in equilibrium. To determine the equilibrium, remember that the individual educational choice is summarized by the function $\psi(\gamma, \nu_{t-1})$, which was defined in inequality (37) and describes the values of the risk aversion parameter and of the aggregate shocks for which an individual chooses a particular education. Specifically, an individual choose education F if $\psi(\gamma, \nu_{t-1}) > 0$. We can now introduce the equilibrium condition for education F. It takes the following form:

$$H_t^{D,F} = N_t \int_{E=F} h_t^F(\chi) G(d\chi) = N_t \sigma \mathcal{T} \int_{\psi(\gamma, \nu_{t-1}) \geq 0} \exp(\varepsilon_t) G(d\chi)$$

$$H_t^{D,R} = N_t \int_{E=R} h_t^F(\chi) G(d\chi) = N_t \sigma \mathcal{T} \int_{\psi(\gamma, \nu_{t-1}) < 0} \exp(\varepsilon_t) G(d\chi)$$

Using independence between γ and ε as well as $\int \exp(\varepsilon_t) G(d\chi) = 1$, we can write

$$\begin{aligned} \int_{\psi(\gamma,\nu_t)\geq 0} \exp(\varepsilon_t) G(d\chi) &= \left(\int_{\psi(\gamma,\nu_{t-1})\geq 0} G(d\chi)\right)\left(\int \exp(\varepsilon_t) G(d\chi)\right) \\ &= \int_{\psi(\gamma,\nu_{t-1})\geq 0} G(d\chi) \\ &= \text{Fraction of workers in Sector } F \end{aligned}$$

so we can write $H_t^{D,F} = n_t^F \sigma T$, where n^F is the mass/measure of individuals that chose education F. Taking logs, we have:

$$\ln H_t^{D,F} = \ln n_t^F + \ln \sigma + \ln T,$$

Substituting for $H_t^{D,F}$, we obtain the following equilibrium condition:

$$\alpha_0 + \alpha_1 \ln w_t^F = \ln n_t^F + \ln \sigma + \ln T,$$

Solving for $\ln w_t^F$, we have the log equilibrium wage:

$$\left(z_t^F \equiv\right) \quad \ln w_t^F = \frac{\ln n_t^F + \ln \sigma + \ln T - \alpha_0}{\alpha_1}.$$

This wage is for the unit of effective labor. Because the worker i provides $\sigma \exp(\varepsilon_t) T$ of effective labor, his recorded earning is $\sigma \exp(\varepsilon_t) T \exp\left(\frac{\ln n_t^F + \ln \sigma + \ln T - \alpha_0}{\alpha_1}\right)$. Because he works for σT hours, his wage for the labor is $\exp(\varepsilon_t) \exp\left(\frac{\ln n_t^F + \ln \sigma + \ln T - \alpha_0}{\alpha_1}\right)$; we will assume that the cross section "error" consist of n IID copies of ε_t, i.e.,the observed log equilibrium individual wage follows:

$$\ln w_{it}^F = \frac{\ln n_t^F + \ln \sigma + \ln T - \alpha_0}{\alpha_1} + \varepsilon_{it}.$$

Similarly, the equilibrium condition for education R has the following form:

$$H_t^{D,R} = n_t^R \sigma T,$$

where n^R is the mass/measure of individuals that chose education R. Substituting for $H_t^{D,R}$ and solving for $\ln w_t^R$, we obtain the following equilibrium wage for R:

$$\left(z_t^R \equiv\right) \quad \ln w_t^R = \frac{\ln n_t^R + \ln \sigma + \ln T - \alpha_0 - \ln \nu_t}{\alpha_1}.$$

By the same reasoning, the observed log equilibrium wage would look like

$$\ln w_{it}^R = \frac{\ln n_t^R + \ln \sigma + \ln T - \alpha_0 - \ln \nu_t}{\alpha_1} + \varepsilon_{it}.$$

C Long Panels?

Our proposal requires access to two data sets, a cross-section (or short panel) and a long time series of aggregate variables. One may wonder whether we may obtain an estimator with similar properties by exploiting panel data sets in which the time series dimension of the panel data is large enough.

One obvious advantage of combining two sources of data is that time series data may contain variables that are unavailable in typical panel data sets. For example the inflation rate potentially provides more information about aggregate shocks than is available in panel data. We argue with a toy model that even without access to such variables, the estimator based on the two data sets is expected to be more precise, which suggests that the advantage of data combination goes beyond availability of more observable variables.

Consider the alternative method based on one long panel data set, in which both n and T go to infinity. Since the number of aggregate shocks ν_t increases as the time-series dimension T grows, we expect that the long panel analysis can be executed with tedious yet straightforward arguments by modifying ideas in Hahn and Kuersteiner (2002), Hahn and Newey (2004) and Gagliardini and Gourieroux (2011), among others.

We will now illustrate potential problem with the long panel approach with a simple artificial example. Suppose that the econometrician is interested in the estimation of a parameter γ that characterizes the following system of linear equations:

$$
\begin{aligned}
q_{i,t} &= x_{i,t}\frac{\gamma}{\omega} + \nu_t + \varepsilon_{i,t} \qquad i = 1, \ldots, n;\ t = 1, \ldots, T, \\
\nu_t &= \omega \nu_{t-1} + u_t.
\end{aligned}
$$

The variables $q_{i,t}$ and $x_{i,t}$ are observed and it is assumed that $x_{i,t}$ is strictly exogenous in the sense that it is independent of the error term $\varepsilon_{i,t}$, including all leads and lags. For simplicity, we also assume that u_t and $\varepsilon_{i,t}$ are normally distributed with zero mean and that $\varepsilon_{i,t}$ is i.i.d. across both i and t. We will denote by δ the ratio γ/ω.

In order to estimate γ based on the panel data $\{(q_{i,t}, x_{i,t}), i = 1, \ldots, n;\ t = 1, \ldots, T\}$, we can adopt a simple two-step estimator of γ. In a first step, the parameter δ and the aggregate shocks ν_t are estimated using an Ordinary Least Square (OLS) regression of $q_{i,t}$ on $x_{i,t}$ and time dummies.

In the second step, the time-series parameter ω is estimated by regressing $\widehat{\nu}_t$ on $\widehat{\nu}_{t-1}$, where $\widehat{\nu}_t$, $t = 1, \ldots, T$, are the aggregate shocks estimated in the first step using the time dummies. An estimator of γ can then be obtained as $\widehat{\delta}\widehat{\omega}$.

The following remarks are useful to understand the properties of the estimator $\widehat{\gamma} = \widehat{\delta}\widehat{\omega}$. First, even if ν_t were observed, for $\widehat{\omega}$ to be a consistent estimator of ω we would need T to go to infinity, under which assumption we have $\widehat{\omega} = \omega + O_p\left(T^{-1/2}\right)$. This implies that it is theoretically necessary to assume that our data source is a "long" panel, i.e., $T \to \infty$. Similarly, $\hat{\nu}_t$ is a consistent estimator of ν_t only if n goes to infinity. As a consequence, we have $\hat{\nu}_t = \nu_t + O_p\left(n^{-1/2}\right)$. This implies that it is in general theoretically necessary to assume that $n \to \infty$.[15] Moreover, if n and T both go to infinity, $\widehat{\delta}$ is a consistent estimator of δ and $\widehat{\delta} = \delta + O_p\left(n^{-1/2}T^{-1/2}\right)$. All this implies that

$$\widehat{\gamma} = \widehat{\delta}\widehat{\omega} = \left(\delta + O_p\left(\frac{1}{\sqrt{nT}}\right)\right)\left(\omega + O_p\left(\frac{1}{\sqrt{T}}\right)\right) = \delta\omega + O_p\left(\frac{1}{\sqrt{T}}\right) = \gamma + O_p\left(\frac{1}{\sqrt{T}}\right).$$

The $O_p\left(n^{-1/2}T^{-1/2}\right)$ estimation noise of $\widehat{\delta}$, which is dominated by the $O_p\left(T^{-1/2}\right)$, is the term that would arise if ω were not estimated. The term reflects typical findings in long panel analysis (i.e., large n, large T), where the standard errors are inversely proportional to the square root of the number $n \times T$ of observations. The fact that it is dominated by the $O_p\left(T^{-1/2}\right)$ term indicates that the number of observations is effectively equal to T, i.e., the long panel should be treated as a time series problem for all practical purposes.

This conclusion has two interesting implications. First, the sampling noise due to cross-section variation should be ignored and the "standard" asymptotic variance formulae should generally be avoided in the panel data analysis when aggregate shocks are present. We note that Lee and Wolpin's (2006, 2010) standard errors use the standard formula that ignores the $O_p\left(T^{-1/2}\right)$ term. Second, since in most cases the time-series dimension T of a panel data set is relatively small, despite the theoretical assumption that it grows to infinity, estimators based on panel data will generally be more imprecise than may be expected from the "large" number $n \times T$ of observations.[16]

[15] For $\widehat{\omega}$ to have the same distribution as if ν_t were observed, we need n to go to infinity faster than T or equivalently that $T = o(n)$. See Heckman and Sedlacek (1985, p. 1088).

[16] This raises an interesting point. Suppose there is an aggregate time series data set available with which consistent estimation of γ is feasible at the standard rate of convergence. Also suppose that the number of observations there, say τ, is a lot larger than T. If this were the case, we should probably speculate that the panel data analysis is strictly dominated by the time series analysis from the efficiency point of view.

Central Limit Theory for Combined Cross-Section and Time Series

Jinyong Hahn* — UCLA
Guido Kuersteiner† — University of Maryland
Maurizio Mazzocco‡ — UCLA

October 7, 2016

Abstract

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

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

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

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

1 Introduction

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

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

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

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

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

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

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

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

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

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

2 Model

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

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

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

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

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

[1] The function g may naturally arise if the ν_t is an unobserved component that can be estimated from the aggregate time series once the parameters β and ρ are known, i.e., if $\nu_t\equiv\nu_t\left(\beta,\rho\right)$ is a function of (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 $g\left(z_s\right|\beta,\rho)$ here for now.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3 Asymptotic Inference

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

3.1 Stationary Models

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

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

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

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

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

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

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

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

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

with

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

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

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

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

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

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

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

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

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

It follows that

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

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

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

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

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

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

3.2 Unit Root Problems

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

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

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

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

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

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

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

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

4 Joint Panel-Time Series Limit Theory

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Condition 1 *Assume that*

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

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

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

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

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

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

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

Condition 2 *Assume that:*

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

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

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

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

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

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

Condition 3 *Assume that*

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

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

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

4.1 Stable Functional CLT

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

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

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

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

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

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

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

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

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

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

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

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

Proof. In Appendix A. ■

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

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

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

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

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

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

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

Proof. In Appendix A. ■

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

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

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

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

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

4.2 Trend Stationary Models

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

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

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

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

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

and cross-sectional data

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Proof. In Appendix A. ■

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

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

where

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

For

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

it follows that

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

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

4.3 Unit Root Time Series Models

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

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

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

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

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

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

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

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

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

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

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

Note that

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

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

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

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

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

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

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

Assume that

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

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

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

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

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

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

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

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* $\Omega_y = \sum_{t=1}^{T}\Omega_{ty}$.

Conditions 7(i)-(vi) are the same as Conditions 1 (i)-(vi) adapted to the unit root model. Condition 7(vii) replaces Condition 2. It is slightly more primitive in the sense that if η_t^2 is homoskedastic, Condition 7(vii) holds automatically and convergence of $\tau^{-1}\sum_{t=\min(1,\tau_0)+[\tau s]+1}^{\tau_0+[\tau r]}\eta_t^2 \to (r-s)\sigma^2$ 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. In other words, we essentially assume $\tau^{-1}\sum_{t=\min(1,\tau_0)+[\tau s]+1}^{\tau_0+[\tau r]}\eta_t^2$ to be proportional to $r-s$ asymptotically. This assumption is stronger than needed but helps to compare the results with the existing unit root literature.

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

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

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

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

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

Proof. In Appendix A. ■

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

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

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

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

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 non-random functions* $\phi \to \mathbb{R}$. *Assume that* $A_{y,\theta}(\phi_0)$ *is full rank almost surely. Assume that for some* $\varepsilon > 0$,

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

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

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

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

Proof. In Appendix A. ■

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

5 Summary

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix

A Proofs for Section 4

A.1 Proof of Theorem 1

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

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

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

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

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

where

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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

where by Condition 2

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

and

$$\begin{aligned}
&E\left[\left\|\frac{1}{\tau}\sum_{j=1}^{k}\sum_{\substack{t\in\{\tau_0,\dots,\tau_0+\tau\}\\ \cap\{\min(1,\tau_0+1),..,T\}}}\left(\lambda'_{j\nu}\left(\psi^{\nu}_{\tau,t}-E\left[\psi^{\nu}_{\tau,t}|\mathcal{G}_{\tau n,t^* n}\right]\right)\right)^2 1_{ij}\right\|^{1+\delta/2}\right] \qquad (36)\\
&=\frac{1}{\tau^{1+\delta/2}}E\left[\left(\sum_{j=1}^{k}\sum_{\substack{t\in\{\tau_0,\dots,\tau_0+\tau\}\\ \cap\{\min(1,\tau_0+1),..,T\}}}\left\|\lambda'_{\nu}\left(\psi^{\nu}_{\tau,t}-E\left[\psi^{\nu}_{\tau,t}|\mathcal{G}_{\tau n,t^* n}\right]\right)\right\|^2 1_{ij}\right)^{1+\delta/2}\right] \\
&\leq\frac{\left(T+|\tau_0|\right)^{\delta/2}k^{\delta/2}}{\tau^{1+\delta/2}}\sum_{j=1}^{k}\sum_{t\in\{\tau_0,\dots,\tau_0+\tau\}\cap\{\min(1,\tau_0+1),..,T\}}E\left[\left\|\lambda'_{\nu}\left(\psi^{\nu}_{\tau,t}-E\left[\psi^{\nu}_{\tau,t}|\mathcal{G}_{\tau n,t^* n}\right]\right)\right\|^{2+\delta}\right] \\
&\leq\frac{2^{2+\delta}\left(T+|\tau_0|\right)^{1+\delta/2}k^{1+\delta/2}}{\tau^{1+\delta/2}}\sup_t E\left[\left\|\psi^{\nu}_{\tau,t}\right\|^{2+\delta}\right]\to 0,
\end{aligned}$$

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

Next consider (34) where

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

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

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

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

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

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

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

such that (39) follows.

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

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

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

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

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

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

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

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

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

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

By Jensen's inequality, we have

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

and

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

so that

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

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

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

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

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

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

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

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

Note that

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

and

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

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

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

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

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

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

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

and consequently that

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

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

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

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

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

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

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

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 Theorems 8.3 and 15.5 in Billingsley (1968). We need to show that for the 'modulus of continuity'

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

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

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

Define

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

Since

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

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

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

such that

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

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

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

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

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

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

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

with $\Omega_y = \sum_{t\in\{1,..,T\}} \Omega_{yt}$. Let $W(r) = \left(W_y(r), W_\nu(r)\right)$ 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 RHS of (50) is the same as

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

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

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

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

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

The goal is now to show that

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

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

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

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

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

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

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

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

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

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

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

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

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

such that the inequality follows.

A.2 Proof of Corollary 1

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

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

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

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

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

A.3 Proof of Theorem 2

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

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

analogously to (12) and

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

analogously to (11). Stack the moment vectors in

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

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

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

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

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

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

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

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

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

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

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

A.4 Proof of Corollary 2

Partition

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

with inverse

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

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

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

where the right hand side has a mixed normal distribution,

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

and

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

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

A.5 Proof of Theorem 3

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

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

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

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

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

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

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

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

such that

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

and Condition 1 iv) holds. Note that $E\left[|\eta_t|^{2+\delta}\right] \le C$ 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, Equation 2.3)[3] that

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

[3]See Appendix B for details.

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

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

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

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

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

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

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

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

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

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

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

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

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

We also have

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

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

and

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

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

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

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

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

A.6 Proof of Theorem 4

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

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

Defining

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

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

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

we have

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

Using the representation

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

it follows from the continuous mapping theorem and Theorem 3 that

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

The partitioned inverse formula implies that

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

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

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

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

B Proof of (55)

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

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

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

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

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

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

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

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

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

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

The result now follows from the Markov inequality. ■

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

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

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

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

This implies

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

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

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

$$= \tau^{-1} \sum_{i=1}^{k} \sum_{l \in I_i} e^{-2\gamma l/\tau} E\left[\eta_l^2 | \mathcal{G}_{\tau n,(l-\min(1,\tau_0)-1)n}\right] - \sigma^2 \int_s^r e^{-2\gamma t} dt$$

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

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

$$+ \sum_{i=1}^{k} e^{-2\gamma[\tau t_{i-1}]/\tau} \sigma^2 (t_i - t_{i-1}) - \sigma^2 \int_s^r e^{-2\gamma t} dt$$

$$= I_n + II_n + III_n.$$

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

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

We also have by Condition 7vii) that

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

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

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

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

$$\max_{i \leq k} \max_{l \in I_i} \left|e^{-2\gamma l/\tau} - e^{-2\gamma[\tau t_{i-1}]/\tau}\right| \leq \max_{i \leq k} \left|e^{-2\gamma[\tau t_i]/\tau} - e^{-2\gamma[\tau t_{i-1}]/\tau}\right|$$

$$\leq 2 \max_{i \leq k} \left|e^{-2\gamma[\tau t_i]/\tau} - e^{-2\gamma t_i}\right|$$

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

$$\leq 2\delta + \delta = 3\delta.$$

We conclude that

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

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