Estimating cointegrating relations from a cross section

Edith Madsen[1]

Department of Economics, University of Copenhagen, Denmark

November 2001

Abstract

This paper specifies a regression model describing homogeneous cointegrating relations between variables at the individual level. The model allows for correlation between the regressors and the regression error through shocks that are common to all cross-section units so the condition about strictly exogenous regressors is not imposed. It is shown that the estimator obtained by a cross-section regression performed at any point in time is a consistent estimator of the cointegrating parameters and its limiting distribution is normal. The model is extended to allow for the cointegrating parameters to differ randomly across units. In this case a cross-section regression will give a consistent estimator of the cointegrating parameter means.

Keywords: Dynamic models; Cointegrating relations; Cross-section regression

JEL classification: C31; C32

[1] Department of Economics, University of Copenhagen, Studiestræde 6, 1455 Kbh. K., Denmark
Email: edith.madsen@econ.ku.dk

1 Introduction

Recently the topic of unit roots and cointegration has found its way into the econometric analysis of panel data. Most of the research within this area has been concerned with asymptotic properties in non-stationary panel data models where both the cross-section dimension (N) and the time series dimension (T) are large. Some of the contributions to this new research area are reviewed in Banerjee (1999). It is different from most of the previous research performed in panel data models as this has been concentrated on asymptotic properties in stationary panel data models where the cross-section dimension is large and the time series dimension is small, i.e. asymptotics in models for so-called micropanels. Surveys of most of the research performed within this framework can be found in Hsiao (1986), Matyas & Sevestre (1992) and Baltagi (1995).

When performing an econometric analysis within the framework of a dynamic panel data model the best situation is to have observations of a lot of cross-section units over a long period of time. Although more and more such panel data sets become available there are still cases where only cross-section data is available, for example the U.S. Consumer Expenditure Survey and the U.K. Family Expenditure Survey. Obviously one cross section can not give any information on the specific dynamic properties. The question is what does a cross section obtained at some point in time reflect within the framework of a dynamic model. For instance, do cross-section estimates reflect long-run or short-run relations? In order to answer such questions an explicit dynamic model is necessary. The first contribution to this discussion is the paper by Grunfeld (1961). The paper considers a dynamic panel data model (a partial adjustment model) and investigates what a cross-section regression will reveal. In particular it gives conditions for a cross-section regression to reveal long-run relations. The paper was written long before long-run relations were given a natural interpretation in terms of cointegration within a non-stationary framework. Recently an unpublished paper by Adda & Robin (1998) has taken up the discussion again. The paper investigates estimators obtained by cross-section regression in the case where the variables are described by unit roots processes. The present paper is inspired by Adda & Robin (1998). Its purpose is to specify a dynamic model in which a cross-section regression will reveal the cointegrating relations.

The starting point is a regression model describing homogeneous cointegrating relations between variables at the individual level. Suppose that the $I(1)$ regressor is generated independently of the stationary regression error and that innovations are iid across cross-section units. In this case a cross-section regression performed at any point in time will give a consistent estimator (as $N \to \infty$) of the parameters in the cointegrating relations. However in the time series analysis of cointegrated variables the assumption about the regressor being strictly exogenous is usually regarded as being too strong. It is clear that in the case with unrestricted correlation between the $I(1)$ regressor and the stationary regression error a cross-section regression will give an inconsistent estimator (as $N \to \infty$) of the cointegrating parameters. Adda & Robin (1998) shows that the asymptotic bias will be small as the point in time where the cross section is obtained becomes large. However this result is not very useful in practice as it does not give any information on how to make inference on the cointegrating parameters. In similar models for panel data with large N and large T, the assumption about strictly exogenous regressors is not needed in order to obtain a consistent estimator of the cointegrating parameters. Phillips & Moon (1999) shows that a modified pooled estimator obtained by using techniques from the time series analysis to correct for the correlation between the regressor and the regression error is consistent (as $N \to \infty$ and $T \to \infty$). With one cross section it is clearly not possible to use such techniques. This paper shows that the assumption about strictly exogenous regressors can be weakened in order to allow for some degree of correlation between the regressor and the regression error. This is done by allowing for correlation to occur through shocks that are common to all cross-section units and therefore can be removed by subtracting the cross-section sample mean from all variables. When this type of endogeneity is allowed in the regression model described above the paper shows that the usual asymptotic results (as $N \to \infty$) well known from ordinary regression theory apply to the estimator obtained by a cross-section regression. In particular the estimator is a consistent estimator of the cointegrating parameters and its limiting distribution is normal. Similar results are shown for the model with randomly different cointegrating parameters. In this case a cross section regression will give a consistent estimator of the cointegrating parameter means. This result is shown in the paper by Pesaran & Smith (1995) under the stronger assumption that the regressors are strictly exogenous.

The paper is organized in the following way. Section 2 introduces the basic model and gives the underlying assumptions. Section 3 derives the asymptotic properties of the estimator obtained by a cross-section regression. Section 4 extends the model to allow for deterministic variables. Section 5 derives the asymptotic properties of the cross-section estimator in an extended version of the basic model with randomly different cointegrating parameters. Finally, Section 6 concludes the paper.

2 The basic model and assumptions

We consider the stochastic variables $Y_{it}(k_0 \times 1)$, $X_{1it}(k_1 \times 1)$ and $X_{2it}(k_2 \times 1)$ where $i = 1, ..., N$ and $t = 1, 2, ...$. For every cross-section unit $i = 1, ..., N$ we assume that the variables are generated by the following equations

$$Y_{it} = \gamma_1' X_{1it} + \gamma_2' X_{2it} + \eta_{0it} \quad (1)$$

$$X_{1it} = \eta_{1it} \quad (2)$$

$$X_{2it} = X_{2it-1} + \eta_{2it} \quad (3)$$

where γ_1 and γ_2 are $k_1 \times k_0$ and $k_2 \times k_0$ matrices of parameters, respectively, and where the innovation processes η_{0it}, η_{1it} and η_{2it} are (weakly) stationary for every cross-section unit $i = 1, ..., N$.

Under additional assumptions concerning the innovation processes η_{0it}, η_{1it} and η_{2it} that we will give below, the model expresses the following. Viewing the variables Y_{it}, X_{1it} and X_{2it} as time series, the variables Y_{it} and X_{2it} are $I(1)$ whereas the variable X_{1it} is stationary. Furthermore the two $I(1)$ variables Y_{it} and X_{2it} are cointegrated where the cointegrating relations are described by equation (1). Equation (2) can also be considered as a cointegrating relation expressing that the variable X_{1it} is trivially cointegrated since this variable is stationary. Altogether this implies that the cointegrating relations are the same for all cross-section units. Another way of expressing the time series behavior of the variables is by saying that the variable Y_{it} tracks a linear combination of the $I(1)$ variable X_{2it} as described by the columns in γ_2 up to a stationary deviation. Thus γ_2 describes a long-run or equilibrium relation between the two non-stationary variables Y_{it} and X_{2it} which is the same for all cross-section units. Notice that the parameter matrix γ_1 does not have a similar interpretation. In fact the parameter γ_1 is not identified in the time series sense as $Y_{it} - \gamma_2' X_{2it}$ plus any linear combination of the variable X_{1it} is stationary.

When viewing the variables Y_{it}, X_{1it} and X_{2it} as time series, the model is of the same kind as in the work by Park & Phillips (1989). In line with this work equation (1) can be considered as a regression equation where the common exogeneity condition is not imposed. That is for every cross-section unit

$i = 1, ..., N$ some correlation between the regressors X_{1it} and X_{2it} and the regression error η_{0it} is allowed. As shown in Park & Phillips (1989) it is possible to obtain a (super)consistent estimator of the long-run relation i.e. the parameter matrix γ_2 from a time series regression with or without including X_{1it} as a regressor. However in order to obtain a consistent estimator of the long-run relation from a cross-section regression it might be necessary to include the stationary variable X_{1it} as a regressor. We will return to this issue in the following.

In order to specify the behavior of the variables further we assume that the innovation processes η_{0it}, η_{1it} and η_{2it} can be decomposed in the following way

$$\eta_{0it} = \mu_{0i} + u_{0t} + v_{0it} \tag{4}$$

$$\eta_{1it} = \mu_{1i} + u_{1t} + v_{1it} \tag{5}$$

$$\eta_{2it} = \mu_{2i} + u_{2t} + v_{2it} \tag{6}$$

The terms μ_{0i}, μ_{1i} and μ_{2i} describe individual-specific effects that are constant over time. u_{0t}, u_{1t} and u_{2t} represent time-dependent innovations that are the same for all cross-section units. Finally v_{0it}, v_{1it} and v_{2it} describe individual-specific effects that are time-dependent. To be more explicit about the nature of the terms in the expressions (4)-(6) which define the innovation processes, we employ some assumptions concerning these terms.

Assumption 1 $\mu_i = (\mu'_{0i}, \mu'_{1i}, \mu'_{2i})'$ *where* $i = 1, ..., N$ *is a sequence of independent and identically distributed random variables with finite second moment.*

Note that it is not assumed that μ_i has mean zero. The term μ_{1i} allows for individual-specific effects in the level of the variable X_{1it} that are constant over time. The term μ_{2i} has a similar effect on changes in the variable X_{2it}. This means that μ_{2i} generates an individual-specific linear trend in the process X_{2it} when viewed as a time series. Finally the term μ_{0i} allows for individual-specific deviations from the long-run relation given by (1) that are constant over time.

Assumption 2 *For every cross-section unit* $i = 1, ..., N$ *the process* $v_{it} = (v'_{0it}, v'_{1it}, v'_{2it})'$ *is weakly stationary with mean zero. In addition the* v_{it}*'s are independent and identically distributed across cross-section units.*

The term v_{it} describes individual-specific effects that are time-dependent and allowed to be serially correlated. When v_{2it} in addition to being stationary is $I(0)$ when viewed as a time series there are individual-specific stochastic trends in the variable X_{2it} and by that in the variable Y_{it} when these are viewed as time series. This means that some of the individual-specific changes in the variable X_{2it} have permanent effect on the level of X_{2it} and by that on the level of Y_{it}. To illustrate this suppose v_{2it} is white noise. In this case there is an individual-specific random walk in the variable X_{2it} when viewed as a time series. On the other hand suppose v_{2it} is generated by $v_{2it} = w_{it} - w_{it-1}$ where w_{it} is white noise, i.e. v_{2it} is stationary but not $I(0)$. In this case the individual-specific effect on X_{2it} is described by white noise. In practice it might very well be the case that the time series behavior of the variable X_{2it} is described by both types of processes, meaning that some of the individual-specific changes in the variable X_{2it} have permanent effect and some have only transitory effect on the level of X_{2it}.

Assumption 3 *For every cross-section $i = 1, ..., N$ the following holds:*

i) μ_{0i} is independent of μ_{1i} and μ_{2i}

ii) v_{0it} is generated independently of v_{1it} and v_{2it}

The assumption above implies that no correlation between the regressors X_{1it} and X_{2it} and the regression error η_{0it} is allowed through the individual-specific terms. However the assumption about the regressors being uncorrelated with the regression error is usually regarded as being too strong in the literature on time series analysis of cointegrated variables. Therefore we will allow for some correlation between the regressors and the regression error and we do this through the innovations that are common for all cross-section units.

Assumption 4 *The process $u_t = (u'_{0t}, u'_{1t}, u'_{2t})'$ is weakly stationary with mean zero. In addition $\sum_{s=-\infty}^{s=\infty} \Gamma_{2u}(s)$ is positive definite where $\Gamma_{2u}(s)$ is the autocovariance function of the process u_{2t}.*

The term u_t represents shocks that are common to all units. As mentioned above we allow for some degree of endogeneity between the regressors X_{1it} and X_{2it} and the dependent variable Y_{it} through this term since the components in u_t are not assumed to be uncorrelated. It is very important for the results in the next section that it is possible to remove the common shocks and by that the endogeneity between the regressors and the dependent variable by subtracting the cross-section sample mean from all variables. This means that no individual-specific reaction to common shocks is allowed in the model. All cross-section units must react in precisely the same manner to common shocks and this is a quite strong assumption. In addition the term u_t introduce the most simple form of dependency between the cross-section units. The assumption on the autocovariance function of u_{2t} ensures that all components of X_{2it} are $I(1)$ and that there are no cointegrating relations between these components, see Phillips (1986). This implies that the model defined by the equations (1)-(3) is indeed what it seems to be. In particular the parameter matrix γ_2 describes a long-run relation between the two non-stationary variables Y_{it} and X_{2it}.

Also an assumption concerning the initialization of the $I(1)$ variable X_{2it} is needed. The assumption is the following.

Assumption 5 *X_{2i0} where $i = 1, ..., N$ is a sequence of independent and identically distributed random variables with finite second moment.*

Finally, we need an assumption about the mutually dependency of the different terms generating the variables Y_{it}, X_{1it} and X_{2it}.

Assumption 6 *μ_i, v_{is}, u_t and X_{2i0} are mutually independent for all $i = 1, ..., N$ and all $s, t = 1, 2, ...$.*

Altogether we have a linear regression model describing cointegrating relations between the two non-stationary variables Y_{it} and X_{2it}. The terms in X_{2it} which cause the non-stationarity in the time series sense are first of all stochastic trends that are common to all cross-section units and an individual-specific linear trend. In addition to these terms there might be individual-specific stochastic trends as well. An important feature of the model is that endogeneity between the regressors X_{1it} and X_{2it} and the dependent variable Y_{it} is allowed through terms that are common to all cross-section units. This means that all stationary variables X_{1it} where correlation with the $I(1)$ regressor X_{2it} occur through individual-specific terms must be included as regressors in the model.

3 Cross-section regression

We consider the dynamic model specified in previous section as a model for the variables at the individual level. Suppose a cross section obtained at some point in time is available. The cross section is a sample consisting of observations of N cross-section units at time t where $t \in \mathbb{N}$. First of all the cross-section sample mean is subtracted from all variables and the corrected variables are defined as $Y_{it}^* = Y_{it} - \frac{1}{N}\sum_{i=1}^N Y_{it}$, $X_{1it}^* = X_{1it} - \frac{1}{N}\sum_{i=1}^N X_{1it}$ and $X_{2it}^* = X_{2it} - \frac{1}{N}\sum_{i=1}^N X_{2it}$. For notational convenience the stacked $(k_1 + k_2)$-dimensional stochastic variable X_{it}^* is defined as $X_{it}^* = (X_{1it}^{*\prime}, X_{2it}^{*\prime})'$ and the corresponding $(k_1 + k_2) \times k_0$ parameter matrix as $\gamma = (\gamma_1', \gamma_2')'$. Now the regression equation in (1) describing the cointegrating relations can be expressed in terms of the corrected variables in the following way

$$Y_{it}^* = \gamma' X_{it}^* + \eta_{0it}^* \qquad i = 1, ..., N \text{ and } t \in \mathbb{N} \tag{7}$$

where $\eta_{0it}^* = \eta_{0it} - \frac{1}{N}\sum_{i=1}^N \eta_{0it}$. The corresponding cross-section ordinary least square estimator denoted $\hat{\gamma}_{N,t}$ is defined as

$$\hat{\gamma}_{N,t} = \left(\sum_{i=1}^N X_{it}^* X_{it}^{*\prime}\right)^{-1} \left(\sum_{i=1}^N X_{it}^* Y_{it}^{*\prime}\right) \tag{8}$$

According to Assumption 3 the regressor X_{it}^* is independent of the regression error η_{0it}^* as the common shocks have been removed from the variables. This immediately implies that $\hat{\gamma}_{N,t}$ is an unbiased estimator of γ, i.e. $E\left(\hat{\gamma}_{N,t}\right) = \gamma$. The asymptotic behavior as $N \to \infty$ of the cross-section estimator $\hat{\gamma}_{N,t}$ is given in the proposition below.

Proposition 1 *Under Assumption 1-6 the following holds:*

$\hat{\gamma}_{N,t}$ *is a consistent estimator of* γ, *i.e.*

$$\hat{\gamma}_{N,t} \xrightarrow{P} \gamma \text{ as } N \to \infty \tag{9}$$

The limiting distribution of $\hat{\gamma}_{N,t}$ *is given by*

$$\sqrt{N}\left(\hat{\gamma}_{N,t} - \gamma\right) \xrightarrow{w} N\left(0, \Sigma_t^{-1} \otimes \Omega\right) \text{ as } N \to \infty \tag{10}$$

The variance in the limiting distribution can be estimated consistently by using the following results

$$\frac{1}{N}\sum_{i=1}^N X_{it}^* X_{it}^{*\prime} \xrightarrow{P} \Sigma_t \text{ as } N \to \infty \tag{11}$$

$$\frac{1}{N}\sum_{i=1}^N \left(Y_{it}^* - \hat{\gamma}_{N,t}' X_{it}^*\right)\left(Y_{it}^* - \hat{\gamma}_{N,t}' X_{it}^*\right)' \xrightarrow{P} \Omega \text{ as } N \to \infty \tag{12}$$

The proof of Proposition 1 is given in Appendix A.1. The proposition shows that by using ordinary regression methods it is possible to make asymptotic inference on the parameters in the cointegrating relations from a cross section obtained at any point in time. In particular it is possible to uncover the long-run relations between the two non-stationary variables Y_{it} and X_{2it} from a cross-section regression. The only difference between the result above and the well known result from ordinary regression is that the variance in the limiting distribution depends on the point in time where the cross section is obtained. Clearly this is because the regressor X_{2it} is non-stationary when viewed as a time series.

Now, consider the cross-section estimator of γ_2 defined as the submatrix of $\hat{\gamma}_{N,t}$ corresponding to the regressor X_{2it}. To be more specific let this estimator denoted $\hat{\gamma}_{2,N,t}$ be the last k_2 rows in $\hat{\gamma}_{N,t}$ and let Σ^{22t} be the lower $k_2 \times k_2$ diagonal block matrix of Σ_t^{-1}, i.e.

$$\Sigma^{22t} = \left(\Sigma_{22,t} - \Sigma_{21,t}\Sigma_{11,t}^{-1}\Sigma_{12,t}\right)^{-1} \tag{13}$$

where Σ_t is decomposed according to X_{1it} and X_{2it} as

$$\Sigma_t = \begin{bmatrix} \Sigma_{11,t} & \Sigma_{12,t} \\ \Sigma_{21,t} & \Sigma_{22,t} \end{bmatrix} \tag{14}$$

Then according to Proposition 1 the limiting distribution of $\hat{\gamma}_{2,N,t}$ is given by

$$\sqrt{N}\left(\hat{\gamma}_{2,N,t} - \gamma_2\right) \overset{w}{\to} N\left(0, \Sigma^{22t} \otimes \Omega\right) \text{ as } N \to \infty \tag{15}$$

The following assumption is used in the results given below.

Assumption 7 *For $a \in \mathbb{R}$ the diagonal matrix F_t is defined in the following way*

$$F_t = \begin{bmatrix} I_{k_1} & 0 \\ 0 & t^a I_{k_2} \end{bmatrix}$$

and the following condition is satisfied

$$\lim_{t\to\infty} \left(F_t \Sigma_t F_t\right) \text{ is positive definite} \tag{16}$$

The assumption implies that when the components of Σ_t are normalized correctly with respect to t then the limit of the normalized matrix $F_t \Sigma_t F_t$ as $t \to \infty$ is well-defined and positive definite. The results below concern the properties of the variance in the limiting distribution of $\hat{\gamma}_{2,N,t}$ as the point in time where the cross-section is obtained goes to infinity.

Result 1 *Let* $\Gamma_v(s)$ *be the autocovariance function of the weakly stationary process* $(v'_{1it}, v'_{2it})'$ *and assume that this process has absolutely summable autocovariances. Then the following holds:*

(a) *If Assumption 7 is satisfied with* $a = -1$ *then*

$$\Sigma^{22t} \otimes \Omega = O\left(t^{-2}\right) \tag{17}$$

(b) *If Assumption 7 is satisfied with* $a = -1/2$ *then*

$$\Sigma^{22t} \otimes \Omega = O\left(t^{-1}\right) \tag{18}$$

(c) *If Assumption 7 is satisfied with* $a = 0$ *then*

$$\lim_{t\to\infty}\left(\Sigma^{22t} \otimes \Omega\right) \text{ is well-defined} \tag{19}$$

The proof of these results are given in Appendix A.2. The assumption in (a) implies that $\mathrm{Var}(\mu_{2i})$ is positive definite. In the time series dimension this means that there is an individual-specific linear trend in the variable X_{2it}. From the time series analysis it is well known that asymptotically as $t \to \infty$ a linear trend will dominate possible $I(1)$ trends. In the cross-section dimension this means that the variation between the components of X_{2it} is of order t^2. Combining this with the property that the regression error η_{0it} is stationary implies that the convergence rate of the variance in the limiting distribution is at most of order t^{-2} meaning that the variance in the limiting distribution converges to zero faster than $t^{-2+\rho}$ for any $\rho > 0$ as $t \to \infty$. The assumption in (b) and (c) implies that $\mathrm{Var}(\mu_{2i}) = 0$ meaning that there is no individual-specific linear trend in the variable X_{2it} when viewed as a time series. A possible linear trend in X_{2it} is the same for all cross-section units and is therefore removed when the cross-section sample mean is subtracted. In addition the assumption in (b) implies that $\sum_{s=-\infty}^{\infty}\Gamma_{2v}(s)$ is positive definite where $\Gamma_{2v}(s)$ is the autocovariance function of v_{2it}. As mentioned earlier this means that there is an individual-specific $I(1)$ trend in all components of X_{2it} when viewed as a time series and these components are not cointegrated, see Phillips (1986). Thus the cross-section variation between the components of X_{2it} is of order t meaning that the variance in the limiting distribution still converges to zero but at a slower rate than in (a). The proof in Appendix A.2 shows that $\sum_{s=-\infty}^{\infty}\Gamma_{2v}(s)$ is the so-called long-run variance matrix of the process $v_{2i1} + ... + v_{2it}$ viewed as a time series and that $\frac{1}{t}\Sigma_{22,t}$ converges to this long-run variance as $t \to \infty$. On the other hand the assumption in (c) implies that $\sum_{s=-\infty}^{\infty} \Gamma_{2v}(s) = 0$ or in other words that the long-run variance of $v_{2i1} + ... + v_{2it}$ is zero. Again according to Phillips (1986) this implies that $v_{2i1} + ... + v_{2it}$ is stationary. More precisely it converges to its stationary distribution as $t \to \infty$.

It is clear that Result 1 depends on the regression error η_{0it} being stationary when viewed as a time series. As a possible extension of the model consider the case where the regression error contains an individual-specific linear trend when viewed as a time series. This means that the cointegrating relations between the two non-stationary variables Y_{it} and X_{2it} are trend-stationary. This extension of the model might be of interest in empirical applications. One example is the model for labor demand functions for industries used in Pesaran & Smith (1995). If the individual-specific linear trend in the regression error is independent of all terms in the regressors X_{1it} and X_{2it} then ordinary regression methods as given in Proposition 1 can be used to make asymptotic inference on the cointegrating parameters. The only difference is that the variance in the limiting distribution will be different as the cross-section variation of η_{0it} will be of order t^2 and hence Ω will be time-dependent.

As a special case of the model defined in Section 2 consider the following model

$$Y_{it} = \gamma_2' X_{2it} + \mu_{0i} + u_{0t} + v_{0it} \tag{20}$$

$$X_{2it} = X_{2i0} + \mu_2 t + u_{21} + ... + u_{2t} + v_{2i1} + ... + v_{2it} \tag{21}$$

The only individual-specific terms present in the $I(1)$ regressor X_{2it} are the initial value and cumulated process $v_{2i1} + ... + v_{2it}$. Consider the case where v_{2it} is white noise with $E(v_{2it} v_{2it}') = \Sigma_2$. From the proof of Result 1 given in Appendix A.2 it follows that $\Sigma^{22t} = (\text{Var}(X_{2i0}) + t\Sigma_2)^{-1}$ implying that for any $k > 0$ the following holds $\Sigma^{22t} > \Sigma^{22(t+k)}$, i.e. $\Sigma^{22t} - \Sigma^{22(t+k)}$ is positive definite[2]. Comparing with (15) this in turn implies that the cross-section estimator $\hat{\gamma}_{2,N,t+k}$ obtained at time $t+k$ is more efficient (asymptotically as $N \to \infty$) than the cross-section estimator $\hat{\gamma}_{2,N,t}$ obtained at time t. Therefore as regards estimators of the long-run parameters in this model there is a gain in efficiency by using a cross section obtained at a later point in time when v_{2it} is white noise. However this result relies on the time series properties of the process v_{2it} and it is easy to find examples where $\Sigma^{22t} - \Sigma^{22(t+k)}$ is not positive definite. This is for example the case when the cumulated process $v_{2i1} + ... + v_{2it}$ is the sum of a random

[2] $A_t = \text{Var}(X_{2i0}) + t\Sigma_2$ is positive definite. Then $\Sigma^{22t} - \Sigma^{22(t+k)} = A_t^{-1} - (A_t + k\Sigma_2)^{-1} = A_t^{-1}\left(A_t^{-1} + \frac{1}{k}\Sigma_2^{-1}\right)^{-1} A_t^{-1}$ is positive definite as $\left(A_t^{-1} + \frac{1}{k}\Sigma_2^{-1}\right)^{-1}$ is positive definite.

walk and a stationary VAR(1) process for suitable values of all parameters. Altogether this illustrates that the properties described in Result 1 are asymptotic properties as $t \to \infty$. The results depend on the time series properties of the individual-specific terms in X_{2it} and with just one cross-section it is not possible to get any information on these properties. Thus the results are not very useful in practice but are merely suppositions.

Once again consider the model above when v_{2it} is white noise. Using (21) it follows that for any $0 < k < t$ the regressor X_{2it} can be expressed as the sum of X_{2it-k} and an error term.

$$\begin{aligned} X_{2it} &= X_{2it-k} + w_{it} \\ w_{it} &= \mu_2 k + u_{2t-k+1} + \ldots + u_{2t} + v_{2it-k+1} + \ldots + v_{2it} \end{aligned}$$

As v_{2it} is white noise, dependency between X_{2it-k} and the error term w_{it} occurs only through terms that are common to all individuals. Note that this will not be the case if X_{2it} contains an individual-specific linear trend. Using the expression in (21) the cointegrating relation in (20) can be expressed as

$$Y_{it} = \gamma_2' X_{2it-k} + \gamma_2' w_{it} + \eta_{0it}$$

Using the same arguments as before it is clear that a cross-section regression of Y_{it}^* on X_{2it-k}^* (the variables corrected for their cross-section sample mean) will give an unbiased estimator of γ_2. Using (20) and (21) the following expression for ΔY_{it} is obtained

$$\Delta Y_{it} = \gamma_2' \Delta X_{2it} + \Delta \eta_{0it} = \gamma_2' \left(\mu_2 + u_{2t} + v_{2it} \right) + \Delta \eta_{0it}$$

Again as v_{2it} is white noise, the change in the dependent variable is independent of all lags of the regressor when both are corrected for their cross-section sample mean, i.e. ΔY_{it}^* and X_{2it-k}^* are independent. This is in fact just the conditions derived in Grunfeld (1961) for a cross-section regression of the type described above to give an unbiased estimator of long-run parameters within the framework of a partial adjustment model with stationary variables. Repeating the arguments in the proof of Proposition 1 in Appendix A.1 the estimator obtained by regressing Y_{it}^* on X_{2it-k}^* is also consistent (as $N \to \infty$) with a limiting distribution which is normal. Note that the error term $\gamma_2' w_{it} + \eta_{0it}$ is not stationary when viewed as a time series implying that the variance in the limiting distribution is different from the one in Proposition 1.

4 A model with deterministic variables

When specifying a model for variables at the individual level it is common to allow for differences between units by including some deterministic variables, for instance dummy variables to distinguish between cross-section units (individuals, households) with different demographic characteristics. In a regression model deterministic variables are usually included as additional explanatory variables without specifying how they affect other explanatory variables. Following this line we would just include deterministic variables as regressors in the regression equation (1) describing the cointegrating relations. However in the model defined in Section 2 the behavior of regressors X_{1it} and X_{2it} is specified to some extent and therefore we must also specify how the deterministic variables affect the regressors X_{1it} and X_{2it}. This is formalized by introducing a vector of deterministic variables $D_{it}(l \times 1)$ where $i = 1, ..., N$ and $t = 1, 2, ...$ and making the following extension of the model defined by the equations (1)-(3)

$$Y_{it} = \Phi_0' D_{it} + \gamma_1' X_{1it} + \gamma_2' X_{2it} + \eta_{0it} \quad (22)$$

$$X_{1it} = \Phi_1' D_{it} + \eta_{1it} \quad (23)$$

$$X_{2it} = \Phi_2' \Delta D_{it} + X_{2it-1} + \eta_{2it} \quad (24)$$

As before the processes η_{0it}, η_{1it} and η_{2it} are stationary for every cross-section unit $i = 1, ..., N$ and satisfy the assumptions in Section 2. Φ_0, Φ_1 and Φ_2 are $l \times k_0$, $l \times k_1$ and $l \times k_2$ matrices of parameters, respectively. In addition, the following assumption concerning the deterministic variable D_{it} is made.

Assumption 8 *The deterministic variable D_{it} satisfies the following conditions:*

i) $D_{i0} = 0$ for all $i = 1, ..., N$

ii) D_{it} and η_{is} where $i = 1, ..., N$ are orthogonal for all $t, s = 1, 2, ...$

iii) D_{it} and X_{2i0} where $i = 1, ..., N$ are orthogonal for all $t = 1, 2, ...$

Using that $D_{i0} = 0$ we obtain the following expression for X_{2it}

$$X_{2it} = \Phi_2' D_{it} + X_{2i0} + \eta_{2i1} + ... + \eta_{2it}$$

Thus, in this model the deterministic variable D_{it} appears in levels in all relations, both in (22)-(23) defining the stationary relations and in the relation (24) which causes the non-stationarity in the model.

Another formulation which seems straight forward is to include the level of D_{it} in the expression (24) implying that the deterministic variable gets cumulated in the non-stationary variable X_{2it}. However with the specification chosen above, the model for the variables Y_{it}, X_{1it} and X_{2it} corrected for the deterministic variable D_{it} is the one specified in Section 2. This follows by ii) and iii) in Assumption 8. To be more specific, let the variables Y_{it}, X_{1it} and X_{2it} corrected by cross-section ordinary least squares for D_{it} be denoted $Y_{it \cdot D}$, $X_{1it \cdot D}$ and $X_{2it \cdot D}$, respectively. These variables are generated by the model specified in Section 2. Then as in the previous section we correct these variables for their cross-section sample mean and denote the resulting variables $Y^*_{it \cdot D}$, $X^*_{1it \cdot D}$ and $X^*_{2it \cdot D}$, respectively, and the stacked variable $X^*_{it \cdot D}$ is defined as $X^*_{it \cdot D} = (X^{*\prime}_{1it \cdot D}, X^{*\prime}_{2it \cdot D})'$. The estimator obtained by cross-section regression of $Y^*_{it \cdot D}$ on $X^*_{it \cdot D}$ denoted $\tilde{\gamma}_{N,t}$ is defined as

$$\tilde{\gamma}_{N,t} = \left(\sum_{i=1}^{N} X^*_{it \cdot D} X^{*\prime}_{it \cdot D} \right)^{-1} \left(\sum_{i=1}^{N} X^*_{it \cdot D} Y^{*\prime}_{it \cdot D} \right) \tag{25}$$

Then using the same arguments as in the previous section it immediately follows that $\tilde{\gamma}_{N,t}$ is an unbiased estimator of γ and also the asymptotic results in Proposition 1 apply. This is formulated in the following result.

Result 2 *When the deterministic variable D_{it} satisfies Assumption 8 the cross-section estimator $\tilde{\gamma}_{N,t}$ has the asymptotic properties given in Proposition 1 except that* (11) *and* (12) *are replaced by*

$$\frac{1}{N} \sum_{i=1}^{N} X^*_{it \cdot D} X^{*\prime}_{it \cdot D} \xrightarrow{P} \Sigma_t \text{ as } N \to \infty \tag{26}$$

$$\frac{1}{N} \sum_{i=1}^{N} \left(Y^*_{it \cdot D} - \tilde{\gamma}'_{N,t} X^*_{it \cdot D} \right) \left(Y^*_{it \cdot D} - \tilde{\gamma}'_{N,t} X^*_{it \cdot D} \right)' \xrightarrow{P} \Omega \text{ as } N \to \infty \tag{27}$$

5 A model with heterogeneous cointegrating relations

We consider an extension of the model defined in the previous where the cointegrating relations are allowed to differ randomly across units. The variables are now generated by the following equations

$$Y_{it} = \gamma'_{1i} X_{1it} + \gamma'_{2i} X_{2it} + \eta_{0it} \tag{28}$$

$$X_{1it} = \eta_{1it} \tag{29}$$

$$X_{2it} = X_{2it-1} + \eta_{2it} \tag{30}$$

where the innovation processes η_{0it}, η_{1it} and η_{2it} are stationary for every cross-section unit $i = 1, ..., N$ and satisfy the assumptions in Section 2. The random coefficient matrices γ_{1i} and γ_{2i} satisfy the following assumption.

Assumption 9 *The* $(k_1 + k_2) \times k_0$ *dimensional stochastic variables* $\gamma_i = (\gamma'_{1i}, \gamma'_{2i})'$ *where* $i = 1, ..., N$ *define a sequence of independent and identically distributed random variables with finite fourth moment.*

In addition we need the terms generating the variables Y_{it}, X_{1it} and X_{2it} to satisfy a stronger moment condition and the random coefficients to be independent of all other terms as summarized in the assumptions below.

Assumption 10 X_{2i0}, μ_i, u_t, *and* v_{it} *all have finite fourth moments.*

Assumption 11 γ_i *is independent of* X_{2i0}, μ_i, u_t, *and* v_{it} *for all* $i = 1, ..., N$ *and all* $t = 1, 2, ...$.

The interpretation of this model is as described in Section 2 with the difference that the cointegrating relations are allowed to differ across individuals. Using stacked notation the regression equation in (28) can be expressed as

$$Y_{it} = \gamma_i' X_{it} + \eta_{0it} = \gamma' X_{it} + w_{it} + \eta_{0it} \tag{31}$$

where $\gamma = E(\gamma_i)$ and $w_{it} = (\gamma_i - \gamma)' X_{it}$. Note that $Y_{it} - \gamma' X_{it}$ is not a cointegrating relation as w_{it} is not stationary unless $\gamma_i = E(\gamma_i)$ with probability 1, i.e. the cointegrating relations are homogeneous almost surely. Instead γ is the mean of the cointegrating parameters. Pesaran & Smith (1995) consider a special case of the model specified above where the most important difference is that in their model the common shocks are not present meaning that the regressors are assumed to be strictly exogenous. The estimator obtained by cross-section regression of the variable Y_{it} on X_{it} when both are corrected for their cross-section sample mean is defined in equation (8). It is important to note that the common shocks u_{1t} and u_{2t} are still present in w_{it} after the cross-section sample mean has been subtracted. This is because w_{it} contains a term where there is an interaction between the common shocks and the random coefficients. This will complicate the deviation of the asymptotic properties of the estimator $\hat{\gamma}_{N,t}$ given in the proposition below and it requires the existence of fourth moments of the regressors X_{1it} and X_{2it} and the regression error $w_{it} + \eta_{0it}$.

Proposition 2 *Under Assumption 1-6 and 9-11 the following holds:*

$\hat{\gamma}_{N,t}$ *is a consistent estimator of* $\gamma = E(\gamma_i)$, *i.e.*

$$\hat{\gamma}_{N,t} \overset{P}{\to} \gamma \text{ as } N \to \infty \tag{32}$$

When $u_{1t} \overset{a.s.}{=} 0$ *and* $u_{2t} \overset{a.s.}{=} 0$, *the limiting distribution of* $\hat{\gamma}_{N,t}$ *is given by*

$$\sqrt{N}\left(\hat{\gamma}_{N,t} - \gamma\right) \overset{w}{\to} N\left(0, \left(\Sigma_t^{-1} \otimes I_{k_0}\right)\Theta_t\left(\Sigma_t^{-1} \otimes I_{k_0}\right)\right) \text{ as } N \to \infty \tag{33}$$

When $u_{1t} \overset{a.s.}{=} 0$ *and* $u_{2t} \overset{a.s.}{=} 0$, *the variance in the limiting distribution can be estimated consistently by using the following results*

$$\frac{1}{N}\sum_{i=1}^{N} X_{it}^* X_{it}^{*\prime} \overset{P}{\to} \Sigma_t \text{ as } N \to \infty \tag{34}$$

$$\frac{1}{N}\sum_{i=1}^{N} \text{vec}\left(X_{it}^*\left(Y_{it}^* - \hat{\gamma}_{N,t}' X_{it}^*\right)'\right)\left(\text{vec}\left(X_{it}^*\left(Y_{it}^* - \hat{\gamma}_{N,t}' X_{it}^*\right)'\right)\right)' \overset{P}{\to} \Theta_t \text{ as } N \to \infty \tag{35}$$

The Proof of Proposition 2 is given in Appendix A.3. The proposition shows that in a model with heterogeneous cointegrating relations a cross-section regression performed at any point will give a consistent estimator of the cointegrating parameter means. This is shown without imposing the assumption that the regressors are strictly exogenous. Endogeneity is allowed through the common shocks. It is clear from the proof in Appendix A.3 that Assumption 11 about the random coefficients being independent of the regressors is crucial here. In order to derive the limiting distribution the assumption about the regressors being strictly exogenous is imposed. It might be possible to derive the limiting distribution without imposing this assumption and it will then involve a mixture of distributions but that goes beyond this paper.

6 Conclusion

This paper specifies a dynamic model in which a cross-section regression will reveal the cointegrating parameters. More specifically the paper specifies a regression model describing cointegrating relations between variables at the individual level and shows that ordinary regression methods can be used in order to make asymptotic inference on the cointegrating parameters from a cross section obtained at any point in time. An important feature of the model is that the assumption about strictly exogenous regressors is not imposed. The model allows for some degree of correlation between the regressors and the regression error namely through shocks that are common to all cross-section units. This in turn introduces the most simple type of dependency between the cross-section units and for instance individual-specific reactions to the common shocks are ruled out with this formulation. This is the common problem of how to model cross-section dependence in a satisfactory way.

One serious drawback of having just one observation of the cross-section units at some point in time is that it is not possible to test if the time series behavior of the variables is correctly specified. For instance it is not possible to determine whether the variables are in fact described by unit root processes. Obviously, observations of the cross-section units over time are needed in order to learn about the dynamic properties of the variables. The more observations over time the better. As mentioned in the introduction there is already papers concerning estimation of cointegrating parameters when both the cross-section dimension and the time series dimension are large. In between that and the situation considered in this paper is the situation where the cross-section dimension is large and the time dimension is fixed. So far that situation is more or less unexplored.

Acknowledgments

The author thanks M. Browning and H.C. Kongsted for useful discussions and comments.

A Appendices

A.1 Proof of Proposition 1

This appendix contains the proof of Proposition 1 in the main text. The results in the proposition are all based on the Lindeberg-Levy version of the Central Limit Theorem (CLT) and the Strong Law of Large Numbers (SLLN). The lemma given below appears to be useful in following.

Lemma 1 *Consider the following regression model*

$$Y_i = \beta' X_i + \varepsilon_i \quad \text{for } i = 1, ..., N \tag{36}$$

where the following assumptions are imposed

$$\begin{aligned} &X_i \text{ and } \varepsilon_i \text{ are iid with finite second moment} \\ &E(X_i) = 0 \text{ and } E(\varepsilon_i) = 0 \\ &X_i \text{ and } \varepsilon_i \text{ are independent} \end{aligned} \tag{37}$$

Define the variables corrected for sample mean in the following way

$$\begin{aligned} Y_i^* &= Y_i - \tfrac{1}{N} \textstyle\sum_{i=1}^N Y_i \\ X_i^* &= X_i - \tfrac{1}{N} \textstyle\sum_{i=1}^N X_i \\ \varepsilon_i^* &= \varepsilon_i - \tfrac{1}{N} \textstyle\sum_{i=1}^N \varepsilon_i \end{aligned} \tag{38}$$

The ordinary least square estimator $\hat{\beta}_N$ obtained by regressing Y_i^* on X_i^* can be expressed as

$$\hat{\beta}_N = \beta + \left(\sum_{i=1}^N X_i^* X_i^{*\prime} \right)^{-1} \left(\sum_{i=1}^N X_i^* \varepsilon_i^{*\prime} \right) \tag{39}$$

The limiting distribution of $\sqrt{N}\left(\hat{\beta}_N - \beta\right)$ is given by the following expression

$$\sqrt{N}\left(\hat{\beta}_N - \beta\right) \xrightarrow{w} N\left(0, \Sigma^{-1} \otimes \Omega\right) \text{ as } N \to \infty \tag{40}$$

where $\Sigma = \text{Var}(X_i)$ and $\Omega = \text{Var}(\varepsilon_i)$. In particular $\hat{\beta}_N - \beta \xrightarrow{P} 0$ as $N \to \infty$.

Proof of Lemma 1:

First of all we show that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i^*\varepsilon_i^{*\prime}$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i\varepsilon_i^{\prime}$ are asymptotically equivalent.

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i^*\varepsilon_i^{*\prime}-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i\varepsilon_i^{\prime}\\
=\ &\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\left(X_i-\frac{1}{N}\sum_{i=1}^{N}X_i\right)\left(\varepsilon_i-\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i\right)^{\prime}-X_i\varepsilon_i^{\prime}\right)\\
=\ &\frac{1}{\sqrt{N}}\left(-\frac{1}{N}\sum_{i=1}^{N}X_i\sum_{i=1}^{N}\varepsilon_i^{\prime}\right)\\
=\ &-\left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i\right)^{\prime}\xrightarrow{P}0\text{ as }N\to\infty
\end{aligned}$$

since $\frac{1}{N}\sum_{i=1}^{N}X_i\xrightarrow{P}E(X_i)=0$ as $N\to\infty$ by SLLN and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i$ converges in distribution by the Lindeberg-Levy CLT. This implies that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i^*\varepsilon_i^{*\prime}$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i\varepsilon_i^{\prime}$ have the same limiting distribution as given by the following expression

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i^*\varepsilon_i^{*\prime}\xrightarrow{w}N(0,\Sigma\otimes\Omega)\text{ as }N\to\infty \tag{41}$$

This follows by the Lindeberg-Levy CLT as $X_i\varepsilon_i^{\prime}$ is iid across i with $E(X_i\varepsilon_i^{\prime})=0$ and $\mathrm{Var}(X_i\varepsilon_i^{\prime})=\mathrm{Var}(X_i)\otimes\mathrm{Var}(\varepsilon_i)=\Sigma\otimes\Omega$ as X_i and ε_i are independent with finite second moment. Next we show that $\frac{1}{N}\sum_{i=1}^{N}X_i^*X_i^{*\prime}$ and $\frac{1}{N}\sum_{i=1}^{N}X_iX_i^{\prime}$ are asymptotically equivalent.

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}X_i^*X_i^{*\prime}-\frac{1}{N}\sum_{i=1}^{N}X_iX_i^{\prime}\\
=\ &\frac{1}{N}\sum_{i=1}^{N}\left(\left(X_i-\frac{1}{N}\sum_{i=1}^{N}X_i\right)\left(X_i-\frac{1}{N}\sum_{i=1}^{N}X_i\right)^{\prime}-X_iX_i^{\prime}\right)\\
=\ &-\left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)^{\prime}\xrightarrow{P}0\text{ as }N\to\infty
\end{aligned}$$

since $\frac{1}{N}\sum_{i=1}^{N}X_i\xrightarrow{P}E(X_i)=0$ as $N\to\infty$ by SLLN. This implies that $\frac{1}{N}\sum_{i=1}^{N}X_i^*X_i^{*\prime}$ and $\frac{1}{N}\sum_{i=1}^{N}X_iX_i^{\prime}$ have the same probability limit. Using SLLN on $\frac{1}{N}\sum_{i=1}^{N}X_iX_i^{\prime}$ the above implies the following

$$\frac{1}{N}\sum_{i=1}^{N}X_i^*X_i^{*\prime}\xrightarrow{P}E(X_iX_i^{\prime})=\mathrm{Var}(X_i)=\Sigma\text{ as }N\to\infty \tag{42}$$

Combining (41) and (42) the following can be obtained

$$\sqrt{N}\left(\hat{\beta}_N-\beta\right)=\left(\frac{1}{N}\sum_{i=1}^{N}X_i^*X_i^{*\prime}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i^*\varepsilon_i^{*\prime}\right)\xrightarrow{w}N\left(0,\Sigma^{-1}\otimes\Omega\right)\text{ as }N\to\infty$$

Especially $\hat{\beta}_N - \beta \xrightarrow{P} 0$ as $N \to \infty$. $\square$

Proof of Proposition 1:

Summing over t in equation (3) gives the following expression for X_{2it}

$$X_{2it} = X_{2i0} + \sum_{s=1}^{t} \eta_{2is}$$

Inserting the expressions for η_{0it}, η_{1it} and η_{2it} given in (4)-(6) we obtain the following

$$\begin{aligned} \eta_{0it} &= \mu_{0i} + u_{0t} + v_{0it} \\ X_{1it} &= \eta_{1it} = \mu_{1i} + u_{1t} + v_{1it} && (43) \\ X_{2it} &= X_{2i0} + \sum_{s=1}^{t} (\mu_{2i} + u_{2t} + v_{2it}) = X_{2i0} + t\mu_{2i} + \sum_{s=1}^{t} u_{2s} + \sum_{s=1}^{t} v_{2is} && (44) \end{aligned}$$

Subtracting the cross-section sample means in equations above gives

$$\begin{aligned} \eta^*_{0it} &= \mu_{0i} + v_{0it} - \frac{1}{N}\sum_{i=1}^{N} \mu_{0i} - \frac{1}{N}\sum_{i=1}^{N} v_{0it} \\ X^*_{1it} &= \mu_{1i} + v_{1it} - \frac{1}{N}\sum_{i=1}^{N} \mu_{1i} - \frac{1}{N}\sum_{i=1}^{N} v_{1it} \\ X^*_{2it} &= X_{2i0} + t\mu_{2i} + \sum_{s=1}^{t} v_{2is} - \frac{1}{N}\sum_{i=1}^{N} X_{2i0} - t\frac{1}{N}\sum_{i=1}^{N} \mu_{2i} - \sum_{s=1}^{t}\left(\frac{1}{N}\sum_{i=1}^{N} v_{2is}\right) \end{aligned}$$

We define the following variables

$$\begin{aligned} \tilde{\eta}_{0it} &= \mu_{0i} - E(\mu_{0i}) + v_{0it} \\ \tilde{X}_{1it} &= \mu_{1i} - E(\mu_{1i}) + v_{1it} \\ \tilde{X}_{2it} &= X_{2i0} - E(X_{2i0}) + t\mu_{2i} - tE(\mu_{2i}) + \sum_{s=1}^{t} v_{2is} \end{aligned}$$

For every $t \in \mathbb{N}$ these variables all define sequences that are iid across i with mean zero and finite second moment. In addition $\tilde{X}_{1it}$ and $\tilde{X}_{2it}$ are independent of $\tilde{\eta}_{0it}$. This follows by Assumption 1-3 and 5. We define the $(k_1 + k_2)$-dimensional stacked variable $\tilde{X}_{it} = \left(\tilde{X}'_{1it}, \tilde{X}'_{2it}\right)'$. For later use we need the following

$$\begin{aligned} \Omega &\equiv \operatorname{Var}(\tilde{\eta}_{0it}) = \operatorname{Var}(\mu_{0i}) + \operatorname{Var}(v_{0it}) \\ \Sigma_t &\equiv \operatorname{Var}\left(\tilde{X}_{it}\right) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12,t} \\ \Sigma'_{12,t} & \Sigma_{22,t} \end{bmatrix} && (45) \end{aligned}$$

where the Σ_{ij}'s are

$$\begin{aligned}
\Sigma_{11} &\equiv \operatorname{Var}\left(\tilde{X}_{1it}\right) = \operatorname{Var}\left(\mu_{1i}\right) + \operatorname{Var}\left(v_{1it}\right) \\
\Sigma_{12,t} &\equiv \operatorname{Cov}\left(\tilde{X}_{1it}, \tilde{X}_{2it}\right) = t \operatorname{Cov}\left(\mu_{1i}, \mu_{2i}\right) + \textstyle\sum_{s=1}^{t} E\left(v_{1it} v_{2is}'\right) \\
\Sigma_{22,t} &\equiv \operatorname{Var}\left(\tilde{X}_{2it}\right) = \operatorname{Var}\left(X_{2i0}\right) + t^2 \operatorname{Var}\left(\mu_{2i}\right) + \textstyle\sum_{s=1}^{t} \sum_{j=1}^{t} E\left(v_{2is} v_{2ij}'\right)
\end{aligned} \tag{46}$$

This follows by Assumptions 1-3 and 5-6. Finally note the following relations

$$\begin{aligned}
\eta_{0it}^{*} &= \tilde{\eta}_{0it} - \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \tilde{\eta}_{0it} \\
X_{1it}^{*} &= \tilde{X}_{1it} - \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \tilde{X}_{1it} \\
X_{2it}^{*} &= \tilde{X}_{2it} - \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \tilde{X}_{2it}
\end{aligned} \tag{47}$$

The estimator $\hat{\gamma}_{N,t}$ defined in (8) can be written as

$$\hat{\gamma}_{N,t} = \gamma + \left(\sum_{i=1}^{N} X_{it}^{*} X_{it}^{*\prime} \right)^{-1} \left(\sum_{i=1}^{N} X_{it}^{*} \eta_{0it}^{*\prime} \right) \tag{48}$$

According to the relations in (47) and by using Lemma 1 the estimator has a limiting distribution given by the following

$$\sqrt{N}\left(\hat{\gamma}_{N,t} - \gamma\right) \xrightarrow{w} N\left(0, \Sigma_t^{-1} \otimes \Omega\right) \text{ as } N \to \infty \tag{49}$$

In addition $\hat{\gamma}_{N,t} - \gamma \xrightarrow{P} 0$ as $N \to \infty$ meaning that $\hat{\gamma}_{N,t}$ is a consistent estimator of γ.

To show (11) we use the result in (42) from the proof of Lemma 1.

$$\frac{1}{N} \sum_{i=1}^{N} X_{it}^{*} X_{it}^{*\prime} \xrightarrow{P} E\left(\tilde{X}_{it} \tilde{X}_{it}'\right) = \operatorname{Var}\left(\tilde{X}_{it}\right) = \Sigma_t \text{ as } N \to \infty \tag{50}$$

Using the same arguments we also obtain

$$\frac{1}{N} \sum_{i=1}^{N} X_{it}^{*} \eta_{0it}^{*\prime} \xrightarrow{P} E\left(\tilde{X}_{it} \tilde{\eta}_{0it}'\right) \quad = \quad 0 \text{ as } N \to \infty \tag{51}$$

$$\frac{1}{N} \sum_{i=1}^{N} \eta_{0it}^{*} \eta_{0it}^{*\prime} \xrightarrow{P} E\left(\tilde{\eta}_{0it} \tilde{\eta}_{0it}'\right) \quad = \quad \operatorname{Var}\left(\tilde{\eta}_{0it}\right) = \Omega \text{ as } N \to \infty \tag{52}$$

Finally we show (12). Using that $\sum_{i=1}^{N} X_{it}^{*}\left(Y_{it}^{*}-\hat{\gamma}_{N,t}' X_{it}^{*}\right)'=0$ and the relation $Y_{it}^{*}=\gamma' X_{it}^{*}+\eta_{0it}^{*}$ we obtain the following

$$
\begin{aligned}
& \frac{1}{N}\sum_{i=1}^{N}\left(Y_{it}^{*}-\hat{\gamma}_{N,t}' X_{it}^{*}\right)\left(Y_{it}^{*}-\hat{\gamma}_{N,t}' X_{it}^{*}\right)' \\
= & \frac{1}{N}\sum_{i=1}^{N} Y_{it}^{*}\left(Y_{it}^{*\prime}-X_{it}^{*\prime}\hat{\gamma}_{N,t}\right) \\
= & \frac{1}{N}\sum_{i=1}^{N}\left(\gamma' X_{it}^{*}+\eta_{0it}^{*}\right)\left(X_{it}^{*\prime}\gamma+\eta_{0it}^{*\prime}-X_{it}^{*\prime}\hat{\gamma}_{N,t}\right) \\
= & \gamma'\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*}X_{it}^{*\prime}\left(\gamma-\hat{\gamma}_{N,t}\right)+\gamma'\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*}\eta_{0it}^{*\prime} \\
& +\frac{1}{N}\sum_{i=1}^{N}\eta_{0it}^{*}X_{it}^{*\prime}\left(\gamma-\hat{\gamma}_{N,t}\right)+\frac{1}{N}\sum_{i=1}^{N}\eta_{0it}^{*}\eta_{0it}^{*\prime}
\end{aligned}
$$

From (49) and (51) we have that as $N\to\infty$ the sequences $\hat{\gamma}_{N,t}-\gamma$ and $\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*}\eta_{0it}^{*\prime}$ both converges in probability to zero. According to (50) the sequence $\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*}X_{it}^{*\prime}$ has a well-defined probability limit. This gives

$$
\frac{1}{N}\sum_{i=1}^{N}\left(Y_{it}^{*}-\hat{\gamma}_{N,t}' X_{it}^{*}\right)\left(Y_{it}^{*}-\hat{\gamma}_{N,t}' X_{it}^{*}\right)'-\frac{1}{N}\sum_{i=1}^{N}\eta_{0it}^{*}\eta_{0it}^{*\prime}\xrightarrow{P}0 \text{ as } N\to\infty
$$

Combing this with (52) we obtain

$$
\frac{1}{N}\sum_{i=1}^{N}\left(Y_{it}^{*}-\hat{\gamma}_{N,t}' X_{it}^{*}\right)\left(Y_{it}^{*}-\hat{\gamma}_{N,t}' X_{it}^{*}\right)'\xrightarrow{P}\operatorname{Var}\left(\tilde{\eta}_{0it}\right)=\Omega \text{ as } N\to\infty \tag{53}
$$

which is the result in (12).

Altogether we have obtained the results stated in the proposition. □

A.2 Proof of Result 1

This appendix contains the proof of Result 1 in the main text. The result is based on the properties of a weakly stationary process given in the lemma below.

Let C be a $k\times k$ matrix. In the following the norm of C is defined as

$$
\|C\|=\max_{i,j}\left|C_{ij}\right| \tag{54}
$$

Lemma 2 *Let v_t be a weakly stationary k-dimensional vector process and $\Gamma_v(s)$ the corresponding autocovariance function. Assume that the autocovariances are absolutely summable i.e. $\sum_{s=0}^{\infty}\|\Gamma_v(s)\|<\infty$.*

In this case the following holds

$$\frac{1}{t}\sum_{s=1}^{t}\sum_{j=1}^{t}E\left(v_s v_j'\right)\rightarrow\sum_{s=-\infty}^{\infty}\Gamma_v\left(s\right)\;\; \textit{as } t\rightarrow\infty \tag{55}$$

For any $a>0$ *it holds that*

$$\frac{1}{t^a}\sum_{s=1}^{t}E\left(v_t v_s'\right)\rightarrow 0\;\; \textit{as } t\rightarrow\infty \tag{56}$$

Proof of Lemma 2:

Using that $\Gamma_v(s)$ is the autocovariance function of v_t we obtain the following

$$\sum_{s=1}^{t}\sum_{j=1}^{t}E\left(v_s v_j'\right)=\sum_{s=1}^{t}\sum_{j=1}^{t}\Gamma_v\left(s-j\right)=t\Gamma_v\left(0\right)+\sum_{s=1}^{t-1}\left(t-s\right)\left(\Gamma_v\left(s\right)+\Gamma_v\left(s\right)'\right)$$

It is easily follows that

$$\frac{1}{t}\sum_{s=1}^{t}\sum_{j=1}^{t}E\left(v_s v_j'\right)=\Gamma_v\left(0\right)+\sum_{s=1}^{t-1}\left(\Gamma_v\left(s\right)+\Gamma_v\left(s\right)'\right)-\sum_{s=1}^{t-1}\frac{s}{t}\left(\Gamma_v\left(s\right)+\Gamma_v\left(s\right)'\right)$$

The assumption about the autocovariances being absolutely summable implies that for any $\varepsilon>0$ we can find j such that

$$\left\|\Gamma_v\left(j+1\right)\right\|+\left\|\Gamma_v\left(j+2\right)\right\|+...<\varepsilon/4 \tag{57}$$

For this given j we can find T such that for all $t\geq T$ the following holds

$$\frac{1}{t}\left\|\Gamma_v\left(1\right)\right\|+\frac{2}{t}\left\|\Gamma_v\left(2\right)\right\|+...+\frac{j}{t}\left\|\Gamma_v\left(s\right)\right\|<\varepsilon/4$$

Using these results it follows that for all $t\geq T$ we have the following

$$\begin{aligned}\left\|\sum_{s=1}^{\infty}\frac{s}{t}\left(\Gamma_v\left(s\right)+\Gamma_v\left(s\right)'\right)\right\| &\leq \sum_{s=1}^{\infty}\frac{s}{t}\left\|\left(\Gamma_v\left(s\right)+\Gamma_v\left(s\right)'\right)\right\|\leq 2\sum_{s=1}^{\infty}\frac{s}{t}\left\|\Gamma_v\left(s\right)\right\| \\ &= 2\sum_{s=1}^{j}\frac{s}{t}\left\|\Gamma_v\left(s\right)\right\|+2\sum_{s=j+1}^{\infty}\frac{s}{t}\left\|\Gamma_v\left(s\right)\right\| \\ &\leq 2\sum_{s=1}^{j}\frac{s}{t}\left\|\Gamma_v\left(s\right)\right\|+2\sum_{s=j+1}^{\infty}\left\|\Gamma_v\left(s\right)\right\|\leq\varepsilon\end{aligned}$$

Altogether this gives the result in (55). That is

$$\frac{1}{t}\sum_{s=1}^{t}\sum_{j=1}^{t}E\left(v_s v_j'\right)\rightarrow\Gamma_v\left(0\right)+\sum_{s=1}^{\infty}\left(\Gamma_v\left(s\right)+\Gamma_v\left(s\right)'\right)=\sum_{s=-\infty}^{\infty}\Gamma_v\left(s\right)\;\;\text{as } t\rightarrow\infty \tag{58}$$

where the limit on the right hand side is well-defined as $\sum_{s=1}^{\infty}\Gamma_v\left(s\right)$ is well-defined by the assumption about the process v_t having absolutely summable autocovariances.

The result in (56) is simply a consequence of the assumption that $\sum_{s=0}^{\infty} \Gamma_v(s)$ is well-defined as the limit of $\sum_{s=0}^{t} \Gamma_v(s)$ as $t \to \infty$ i.e.

$$\sum_{s=1}^{t} E\left(v_t v_s'\right) = \sum_{s=0}^{t-1} \Gamma_v(s) \to \sum_{s=0}^{\infty} \Gamma_v(s) \text{ as } t \to \infty \tag{59}$$

□

Proof of Result 1:

Let Σ^{22t} be the lower $k_2 \times k_2$ diagonal block matrix of Σ_t^{-1}. Comparing with (45) in Appendix A.1 we obtain the following

$$\Sigma^{22t} = \left(\Sigma_{22,t} - \Sigma_{12,t}' \Sigma_{11}^{-1} \Sigma_{12,t}\right)^{-1}$$

According to (46) also in Appendix A.1 we have the following expressions

$$\begin{aligned} \Sigma_{22,t} &= \operatorname{Var}(X_{2i0}) + t^2 \operatorname{Var}(\mu_{2i}) + \textstyle\sum_{s=1}^{t} \sum_{j=1}^{t} E\left(v_{2is} v_{2ij}'\right) \\ \Sigma_{12,t} &= t \operatorname{Cov}(\mu_{1i}, \mu_{2i}) + \textstyle\sum_{s=1}^{t} E\left(v_{1it} v_{2is}'\right) \end{aligned} \tag{60}$$

For $a \in \mathbb{R}$ the diagonal matrix F_t is defined as

$$F_t = \begin{bmatrix} I_{k_1} & 0 \\ 0 & t^a I_{k_2} \end{bmatrix}$$

The condition in Assumption 7 concerns the limit as $t \to \infty$ of the matrix $F_t \Sigma_t F_t$ which can be decomposed in the same way as Σ_t as follows

$$F_t \Sigma_t F_t = \begin{bmatrix} \Sigma_{11} & t^a \Sigma_{12,t} \\ t^a \Sigma_{12,t}' & t^{2a} \Sigma_{22,t} \end{bmatrix} \tag{61}$$

(a) Comparing with (60) we obtain the following expressions

$$\begin{aligned} \frac{1}{t^2} \Sigma_{22,t} &= \operatorname{Var}(\mu_{2i}) + \frac{1}{t^2} \sum_{s=1}^{t} \sum_{j=1}^{t} E\left(v_{2is} v_{2ij}'\right) + \frac{1}{t^2} \operatorname{Var}(X_{2i0}) \\ \frac{1}{t} \Sigma_{12,t} &= \operatorname{Cov}(\mu_{1i}, \mu_{2i}) + \frac{1}{t} \sum_{s=1}^{t} E\left(v_{1it} v_{2is}'\right) \end{aligned}$$

Using Lemma 2 above we obtain the following

$$\begin{aligned} &\frac{1}{t^2} \Sigma_{22,t} \to \operatorname{Var}(\mu_{2i}) \text{ as } t \to \infty \\ &\frac{1}{t} \Sigma_{12,t} \to \operatorname{Cov}(\mu_{1i}, \mu_{2i}) \text{ as } t \to \infty \end{aligned}$$

Setting $a = -1$ in (61) this implies that as $t \to \infty$

$$F_t \Sigma_t F_t = \begin{bmatrix} \Sigma_{11} & \frac{1}{t}\Sigma_{12,t} \\ \frac{1}{t}\Sigma_{12,t}' & \frac{1}{t^2}\Sigma_{22,t} \end{bmatrix} \to \begin{bmatrix} \Sigma_{11} & \operatorname{Cov}(\mu_{1i}, \mu_{2i}) \\ \operatorname{Cov}(\mu_{2i}, \mu_{1i}) & \operatorname{Var}(\mu_{2i}) \end{bmatrix}$$

As Assumption 7 is satisfied with $a = -1$ the limit on the right hand side in the expression above is positive definite which in turn implies that $\operatorname{Var}(\mu_{2i}) - \operatorname{Cov}(\mu_{2i}, \mu_{1i}) \Sigma_{11}^{-1} \operatorname{Cov}(\mu_{1i}, \mu_{2i})$ is positive definite. This means that as $t \to \infty$

$$t^2 \Sigma^{22t} = \left(\frac{1}{t^2}\Sigma_{22,t} - \frac{1}{t}\Sigma_{12,t}' \Sigma_{11}^{-1} \frac{1}{t}\Sigma_{12,t} \right)^{-1} \to \left(\operatorname{Var}(\mu_{2i}) - \operatorname{Cov}(\mu_{2i}, \mu_{1i}) \Sigma_{11}^{-1} \operatorname{Cov}(\mu_{1i}, \mu_{2i}) \right)^{-1}$$

Therefore that $\lim_{t \to \infty} \left(t^2 \Sigma^{22t} \otimes \Omega \right)$ is well-defined which gives (17) in Result 1. Note that (16) in Assumption 7 also implies that $\operatorname{Var}(\mu_{2i})$ is positive definite.

(b) Setting $a = -1/2$ in (61) leads us to consider the following

$$\begin{aligned} \frac{1}{t}\Sigma_{22,t} &= t \operatorname{Var}(\mu_{2i}) + \frac{1}{t} \sum_{s=1}^{t} \sum_{j=1}^{t} E\left(v_{2is} v_{2ij}' \right) + \frac{1}{t} \operatorname{Var}(X_{2i0}) \\ \frac{1}{t^{1/2}}\Sigma_{12,t} &= t^{1/2} \operatorname{Cov}(\mu_{1i}, \mu_{2i}) + \frac{1}{t^{1/2}} \sum_{s=1}^{t} E\left(v_{1it} v_{2is}' \right) \end{aligned}$$

According to Lemma 2 the limits as $t \to \infty$ of these expressions are well-defined if and only if $\operatorname{Var}(\mu_{2i}) = 0$. In the case where $\operatorname{Var}(\mu_{2i}) = 0$ the results in Lemma 2 give the following as $t \to \infty$

$$\begin{aligned} &\frac{1}{t}\Sigma_{22,t} \to \sum_{s=-\infty}^{\infty} \Gamma_{2v}(s) \\ &\frac{1}{t^{1/2}}\Sigma_{12,t} \to 0 \end{aligned}$$

where $\Gamma_{2v}(s)$ is the autocovariance function corresponding to v_{2it}. Thus we have

$$F_t \Sigma_t F_t = \begin{bmatrix} \Sigma_{11} & \frac{1}{t^{1/2}}\Sigma_{12,t} \\ \frac{1}{t^{1/2}}\Sigma_{12,t}' & \frac{1}{t}\Sigma_{22,t} \end{bmatrix} \to \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \sum_{s=-\infty}^{\infty} \Gamma_{2v}(s) \end{bmatrix}$$

Again as Assumption 7 is satisfied with $a = -1/2$ the limit on the right hand side in the expression above is positive definite which in particular means that $\sum_{s=-\infty}^{\infty} \Gamma_{2v}(s)$ is positive definite. This gives that as $t \to \infty$

$$t \Sigma^{22t} = \left(\frac{1}{t}\Sigma_{22,t} - \frac{1}{t^{1/2}}\Sigma_{12,t}' \Sigma_{11}^{-1} \frac{1}{t^{1/2}}\Sigma_{12,t} \right)^{-1} \to \left(\sum_{s=-\infty}^{\infty} \Gamma_{2v}(s) \right)^{-1}$$

which gives (18) in Result 1. Note that the limit of $\frac{1}{t}\Sigma_{22,t}$ is the so-called long-run variance of the process $v_{2i1} + \ldots + v_{2it}$.

(c) When $a = 0$ we have $F_t = I_{k_1+k_2}$ the assumption that (16) is satisfied implies that $\lim_{t\to\infty}(\Sigma_t)$ is well-defined which in turn implies that $\lim_{t\to\infty}\left(\Sigma^{22t}\right)$ is well-defined. This gives (19) in Result 1. Consider the expressions in (60). One necessary condition for these expressions to have well-defined limits as $t\to\infty$ is that $\text{Var}(\mu_{2i}) = 0$. Another necessary condition is that $\frac{1}{t}\sum_{s=1}^{t}\sum_{j=1}^{t}E\left(v_{2is}v_{2ij}'\right)\to 0$ as $t\to\infty$ that is $\sum_{s=-\infty}^{\infty}\Gamma_{2v}(s) = 0$. Otherwise the limit as $t\to\infty$ of $\sum_{s=1}^{t}\sum_{j=1}^{t}E\left(v_{2is}v_{2ij}'\right)$ is not well-defined. $\square$

A.3 Proof of Proposition 2

This appendix contains the proof of Proposition 2 in the main text. The consistency result is as before based on the Lindeberg-Levy CLT and SLLN. When deriving the limiting distribution the lemma below is useful.

Lemma 3 *Let x_i, y_i, z_i and v_i where $i = 1,\ldots,N$ be sequences of iid scalar variables with mean zero and finite fourth moments then as $N\to\infty$*

$$Z_N = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \frac{1}{N}\sum_{i=1}^{N}x_i\right)\left(y_i - \frac{1}{N}\sum_{i=1}^{N}y_i\right)\left(z_i - \frac{1}{N}\sum_{i=1}^{N}z_i\right)\left(v_i - \frac{1}{N}\sum_{i=1}^{N}v_i\right)\xrightarrow{P}E\left(x_iy_iz_iv_i\right)$$

Proof of Lemma 3:

As all variables are assumed to have mean zero and finite fourth moments SLLN gives the following

$$\begin{aligned}
m_x &\equiv \frac{1}{N}\sum_{i=1}^{N}x_i \xrightarrow{P} E(x_i) = 0 \text{ as } N\to\infty\\
m_{xy}^2 &\equiv \frac{1}{N}\sum_{i=1}^{N}x_iy_i \xrightarrow{P} E(x_iy_i) \text{ as } N\to\infty\\
m_{xyz}^3 &\equiv \frac{1}{N}\sum_{i=1}^{N}x_iy_iz_i \xrightarrow{P} E(x_iy_iz_i) \text{ as } N\to\infty\\
m_{xyzv}^4 &\equiv \frac{1}{N}\sum_{i=1}^{N}x_iy_iz_iv_i \xrightarrow{P} E(x_iy_iz_iv_i) \text{ as } N\to\infty
\end{aligned}$$

where the indexes can be any combination of x, y, z and v. Using the notation introduced above and the results concerning the probability limits we find

$$\begin{aligned}
Z_N = & \frac{1}{N}\sum_{i=1}^{N} x_i y_i z_i v_i - m_{xyz}^3 m_v - m_{xyv}^3 m_z - m_{xzv}^3 m_y - m_{yzv}^3 m_x \\
& + m_{xy}^2 m_z m_v + m_{xz}^2 m_y m_v + m_{xv}^2 m_y m_z + m_{yz}^2 m_x m_v + m_{yv}^2 m_x m_z + m_{zv}^2 m_x m_y - 3 m_x m_y m_z m_v \\
\xrightarrow{P} & \; E\left(x_i y_i z_i v_i\right) \text{ as } N \to \infty
\end{aligned}$$

□

Remark 1 The result in Lemma 3 also holds when for example $x_i = y_i$. In that case

$$\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \frac{1}{N}\sum_{i=1}^{N} x_i\right)^2 \left(z_i - \frac{1}{N}\sum_{i=1}^{N} z_i\right)\left(v_i - \frac{1}{N}\sum_{i=1}^{N} v_i\right) \xrightarrow{P} E\left(x_i^2 z_i v_i\right) \text{ as } N \to \infty$$

Proof of Proposition 2:

From (50) and (51) in Appendix A.1 we have the following results

$$\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*} X_{it}^{*\prime} \xrightarrow{P} \Sigma_t \text{ as } N \to \infty \tag{62}$$

$$\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*} \eta_{0it}^{*\prime} \xrightarrow{P} 0 \text{ as } N \to \infty \tag{63}$$

The estimator $\hat{\gamma}_{N,t}$ can be expressed as follows

$$\hat{\gamma}_{N,t} = \gamma + \left(\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*} X_{it}^{*\prime}\right)^{-1}\left(\left(\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*} w_{it}^{*\prime}\right) + \left(\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*} \eta_{0it}^{*\prime}\right)\right) \tag{64}$$

where $w_{it} = \left(\gamma_i - \gamma\right)' X_{it}$ and $w_{it}^{*} = w_{it} - \frac{1}{N}\sum_{i=1}^{N} w_{it}$. Thus $\hat{\gamma}_{N,t}$ is a consistent estimator of γ if $\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*} w_{it}^{*\prime} \xrightarrow{P} 0$ as $N \to \infty$. To show this we use the results derived below. As γ_i is independent of X_{2i0}, μ_i and v_{it} by Assumption 11, SLLN gives the following as $N \to \infty$

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left(\gamma_{2i} - \gamma_2\right)' X_{2i0} \xrightarrow{P} E\left(\left(\gamma_i - \gamma\right)' X_{2i0}\right) = E\left(\gamma_i - \gamma\right)' E\left(X_{2i0}\right) = 0 \\
&\frac{1}{N}\sum_{i=1}^{N}\left(\gamma_{2i} - \gamma_2\right)' \eta_{2it} = \frac{1}{N}\sum_{i=1}^{N}\left(\gamma_{2i} - \gamma_2\right)'\left(\mu_{2i} + u_{2t} + v_{2it}\right) \xrightarrow{P} E\left(\gamma_{2i} - \gamma_2\right)'\left(E\left(\mu_{2i}\right) + u_{2t} + E\left(v_{2it}\right)\right) = 0 \\
&\frac{1}{N}\sum_{i=1}^{N}\left(\gamma_{1i} - \gamma_1\right)' \eta_{1it} = \frac{1}{N}\sum_{i=1}^{N}\left(\gamma_{1i} - \gamma_1\right)'\left(\mu_{1i} + u_{1t} + v_{1it}\right) \xrightarrow{P} 0
\end{aligned}$$

Comparing with the expressions for X_{1it} and X_{2it} in (43) and (44) in Appendix A.1 the above implies that

$$\frac{1}{N}\sum_{i=1}^{N} w_{it} = \frac{1}{N}\sum_{i=1}^{N}\left(\gamma_i - \gamma\right)' X_{it} \xrightarrow{P} 0 \text{ as } N \to \infty \tag{65}$$

Again by using the expressions for X_{1it} and X_{2it} in (43) and (44), SLLN gives the following as $N \to \infty$

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} X_{1it} &= \frac{1}{N}\sum_{i=1}^{N}\left(\mu_{1i}+u_{1t}+v_{1it}\right) \xrightarrow{P} E\left(\mu_{1i}\right)+u_{1t} \\
\frac{1}{N}\sum_{i=1}^{N} X_{2it} &= \frac{1}{N}\sum_{i=1}^{N}\left(X_{2i0}+t\mu_{2i}+u_{21}+...+u_{2t}+v_{2i1}+...+v_{2it}\right) \\
&\xrightarrow{P} E\left(X_{2i0}\right)+tE\left(\mu_{2i}\right)+u_{21}+...+u_{2t}
\end{aligned}$$

Thus, $\frac{1}{N}\sum_{i=1}^{N} X_{it}$ has a well-defined (stochastic) probability limit as $N \to \infty$. By using this and (65) it follows that $\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*} w_{it}^{*\prime}$ and $\frac{1}{N}\sum_{i=1}^{N} X_{it} w_{it}'$ are asymptotically equivalent, i.e.

$$\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*} w_{it}^{*\prime} - \frac{1}{N}\sum_{i=1}^{N} X_{it} w_{it}' = -\left(\frac{1}{N}\sum_{i=1}^{N} X_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N} w_{it}\right) \xrightarrow{P} 0 \text{ as } N \to \infty$$

This implies that $\frac{1}{N}\sum_{i=1}^{N} X_{it}^{*} w_{it}^{*\prime}$ and $\frac{1}{N}\sum_{i=1}^{N} X_{it} w_{it}'$ have the same probability limit as $N \to \infty$. Therefore showing that $\frac{1}{N}\sum_{i=1}^{N} X_{it} w_{it}' \xrightarrow{P} 0$ as $N \to \infty$ will imply that $\hat{\gamma}_{N,t}$ is a consistent estimator of γ. This is done by showing that $\frac{1}{N}\sum_{i=1}^{N} X_{it} w_{it}'$ converges to zero in mean square. Using that X_{it} and γ_i are independent by Assumption 11 and both identically distributed across i we find that

$$E\left(\frac{1}{N}\sum_{i=1}^{N} X_{it} w_{it}'\right) = E\left(\frac{1}{N}\sum_{i=1}^{N} X_{it} X_{it}'\left(\gamma_i-\gamma\right)\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(X_{it}X_{it}'\right)E\left(\gamma_i-\gamma\right) = 0 \qquad (66)$$

$$\begin{aligned}
&\operatorname{Var}\left(\frac{1}{N}\sum_{i=1}^{N} X_{it} w_{it}'\right) \\
=\;& \frac{1}{N^2}E\left(\operatorname{vec}\left(\sum_{i=1}^{N} X_{it}X_{it}'\left(\gamma_i-\gamma\right)\right)\left(\operatorname{vec}\left(\sum_{i=1}^{N} X_{it}X_{it}'\left(\gamma_i-\gamma\right)\right)\right)'\right) \\
=\;& \frac{1}{N^2}E\left(\sum_{i=1}^{N}\left(\left(I_{k_0}\otimes X_{it}X_{it}'\right)\operatorname{vec}\left(\gamma_i-\gamma\right)\right)\sum_{i=1}^{N}\left(\operatorname{vec}\left(\gamma_i-\gamma\right)'\left(I_{k_0}\otimes X_{it}X_{it}'\right)\right)\right) \\
=\;& \frac{1}{N^2}E\left(\sum_{i=1}^{N}\left(I_{k_0}\otimes X_{it}X_{it}'\right)\operatorname{Var}\left(\gamma_i-\gamma\right)\left(I_{k_0}\otimes X_{it}X_{it}'\right)\right) \\
=\;& \frac{1}{N}E\left(\left(I_{k_0}\otimes X_{it}X_{it}'\right)\operatorname{Var}\left(\gamma_i-\gamma\right)\left(I_{k_0}\otimes X_{it}X_{it}'\right)\right)
\end{aligned}$$

The third equality sign in the expression above follows by using the law of iterated expectations and that $(\gamma_i - \gamma)$ is iid across i with mean zero according to Assumption 9. The fourth equality sign follows by using that X_{it} is identically distributed across i according to the assumptions in Section 2. By Assumption 10 the variable X_{it} have finite fourth moment that is $E\left(\left(I_{k_0}\otimes X_{it}X_{it}'\right)\operatorname{Var}\left(\gamma_i-\gamma\right)\left(I_{k_0}\otimes X_{it}X_{it}'\right)\right) < \infty$ implying that

$$\operatorname{Var}\left(\frac{1}{N}\sum_{i=1}^{N} X_{it} w_{it}'\right) \to 0 \text{ as } N \to \infty$$

Altogether this shows that $\frac{1}{N}\sum_{i=1}^{N}X_{it}w_{it}'$ converges to zero in mean square implying that $\hat{\gamma}_{N,t}\xrightarrow{P}\gamma$ as $N\to\infty$.

Now assume that $u_{1t}=0$ and $u_{2t}=0$ almost surely. Note that this implies that $w_{it}=(\gamma_i-\gamma)'X_{it}$ where $i=1,...,N$ defines a sequence of iid variables. First of all we show that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_{it}^{*}(w_{it}^{*}+\eta_{0it}^{*})'$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{X}_{it}(w_{it}+\tilde{\eta}_{0it})'$ are asymptotically equivalent. Using the relations in (47) in Appendix A.1 the following is obtained

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_{it}^{*}(w_{it}^{*}+\eta_{0it}^{*})'-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{X}_{it}(w_{it}+\tilde{\eta}_{0it})'\\
=&\ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\tilde{X}_{it}-\frac{1}{N}\sum_{i=1}^{N}\tilde{X}_{it}\right)\left(w_{it}-\frac{1}{N}\sum_{i=1}^{N}w_{it}+\tilde{\eta}_{0it}-\frac{1}{N}\sum_{i=1}^{N}\tilde{\eta}_{0it}\right)'-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{X}_{it}(w_{it}+\tilde{\eta}_{0it})'\\
=&\ -\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{X}_{it}\right)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}w_{it}\right)'-\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{X}_{it}\right)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\eta}_{0it}\right)'\xrightarrow{P}0\text{ as }N\to\infty
\end{aligned}$$

since $\frac{1}{N}\sum_{i=1}^{N}\tilde{X}_{it}\xrightarrow{P}E\left(\tilde{X}_{it}\right)=0$ as $N\to\infty$ by SLLN and both $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\eta}_{0it}$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}w_{it}$ converges in distribution by the Lindeberg-Levy CLT as both $\tilde{\eta}_{0it}$ and w_{it} where $i=1,...,N$ define sequences of iid variables with finite second moment. This implies that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_{it}^{*}(w_{it}^{*}+\eta_{0it}^{*})'$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{X}_{it}(w_{it}+\tilde{\eta}_{0it})'$ have the same limiting distribution as given by the expression below

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_{it}^{*}(w_{it}^{*}+\eta_{0it}^{*})'\xrightarrow{w}N(0,\Theta_t)\ \text{ as }N\to\infty \tag{67}$$

where $\Theta_t=\text{Var}\left(\tilde{X}_{it}(w_{it}+\tilde{\eta}_{0it})'\right)$. This follows by the Lindeberg-Levy CLT as $\tilde{X}_{it}(w_{it}+\tilde{\eta}_{0it})'$ is iid across i with mean

$$E\left(\tilde{X}_{it}(w_{it}+\tilde{\eta}_{0it})'\right)=E\left(\tilde{X}_{it}X_{it}'(\gamma_i-\gamma)\right)+E\left(\tilde{X}_{it}\tilde{\eta}_{0it}'\right)=0$$

and variance

$$\begin{aligned}
&\text{Var}\left(\tilde{X}_{it}(w_{it}+\tilde{\eta}_{0it})'\right)\\
=&\ \text{Var}\left(\tilde{X}_{it}w_{it}'\right)+\text{Var}\left(\tilde{X}_{it}\tilde{\eta}_{0it}'\right)\\
=&\ \text{Var}\left(\tilde{X}_{it}X_{it}'(\gamma_i-\gamma)\right)+\text{Var}\left(\tilde{X}_{it}\tilde{\eta}_{0it}'\right)\\
=&\ E\left(\text{vec}\left(\tilde{X}_{it}X_{it}'(\gamma_i-\gamma)\right)\left(\text{vec}\left(\tilde{X}_{it}X_{it}'(\gamma_i-\gamma)\right)\right)'\right)+\text{Var}\left(\tilde{X}_{it}\right)\otimes\text{Var}(\tilde{\eta}_{0it})\\
=&\ E\left(\left(I_{k_0}\otimes\tilde{X}_{it}X_{it}'\right)\text{vec}(\gamma_i-\gamma)(\text{vec}(\gamma_i-\gamma))'\left(I_{k_0}\otimes X_{it}\tilde{X}_{it}'\right)\right)+\Sigma_t\otimes\Omega\\
=&\ E\left(\left(I_{k_0}\otimes\tilde{X}_{it}X_{it}'\right)\text{Var}(\gamma_i-\gamma)\left(I_{k_0}\otimes X_{it}\tilde{X}_{it}'\right)\right)+\Sigma_t\otimes\Omega
\end{aligned}$$

The first equality sign in the expression for the variance above results from $\tilde{X}_{it}w_{it}'$ and $\tilde{X}_{it}\tilde{\eta}_{0it}'$ being uncorrelated as $\tilde{\eta}_{0it}$ is independent of $\tilde{X}_{it}$ and w_{it} and has mean zero. From the last equality sign in the expression above it is clear that the variance of $\tilde{X}_{it}(w_{it}+\tilde{\eta}_{0it})'$ is well-defined as X_{it} has finite fourth moments by Assumption 10. Now combining (62) and (67) the following is obtained

$$\begin{aligned}\sqrt{N}\left(\hat{\gamma}_{N,t}-\gamma\right) = & \left(\frac{1}{N}\sum_{i=1}^{N}X_{it}^{*}X_{it}^{*\prime}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_{it}^{*}\left(w_{it}^{*}+\eta_{0it}^{*}\right)' \\ \overset{w}{\to} & N\left(0,\left(\Sigma_t^{-1}\otimes I_{k_0}\right)\Theta_t\left(\Sigma_t^{-1}\otimes I_{k_0}\right)\right) \text{ as } N\to\infty\end{aligned}$$

which is the result (33) in Proposition 2.

Finally to show (35) we deduce the following results from Lemma 3

$$\frac{1}{N}\sum_{i=1}^{N}\mathrm{vec}\left(X_{it}^{*}X_{it}^{*\prime}\right)\left(\mathrm{vec}\left(X_{it}^{*}X_{it}^{*\prime}\right)\right)' = O_P(1) \tag{68}$$

$$\frac{1}{N}\sum_{i=1}^{N}\mathrm{vec}\left(X_{it}^{*}X_{it}^{*\prime}\right)\left(\mathrm{vec}\left(X_{it}^{*}\eta_{0it}^{*\prime}\right)\right)' = O_P(1) \tag{69}$$

$$\frac{1}{N}\sum_{i=1}^{N}\mathrm{vec}\left(X_{it}^{*}X_{it}^{*\prime}\right)\left(\mathrm{vec}\left(X_{it}^{*}w_{it}^{*\prime}\right)\right)' = O_P(1) \tag{70}$$

This follows as all the variables $\tilde{X}_{it}$, $\tilde{\eta}_{0it}$ and w_{it} are iid across i with mean zero and finite fourth moment. Using these results we obtain the following

$$\begin{aligned}
& \frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}\left(Y_{it}^{*}-\hat{\gamma}_{N,t}^{\prime}X_{it}^{*}\right)^{\prime}\right)\left(\text{vec}\left(X_{it}^{*}\left(Y_{it}^{*}-\hat{\gamma}_{N,t}^{\prime}X_{it}^{*}\right)^{\prime}\right)\right)^{\prime} \\
& -\frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\left(\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\right)^{\prime} \\
= & \frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}X_{it}^{*\prime}\left(\gamma-\hat{\gamma}_{N,t}\right)+X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\left(\text{vec}\left(X_{it}^{*}X_{it}^{*\prime}\left(\gamma-\hat{\gamma}_{N,t}\right)+X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\right)^{\prime} \\
& -\frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\left(\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\right)^{\prime} \\
= & \frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}X_{it}^{*\prime}\left(\gamma-\hat{\gamma}_{N,t}\right)\right)\left(\text{vec}\left(X_{it}^{*}X_{it}^{*\prime}\left(\gamma-\hat{\gamma}_{N,t}\right)\right)\right)^{\prime} \\
& +\frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}X_{it}^{*\prime}\left(\gamma-\hat{\gamma}_{N,t}\right)\right)\left(\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\right)^{\prime} \\
& +\frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\left(\text{vec}\left(X_{it}^{*}X_{it}^{*\prime}\left(\gamma-\hat{\gamma}_{N,t}\right)\right)\right)^{\prime} \\
= & \left(\left(\gamma-\hat{\gamma}_{N,t}\right)^{\prime}\otimes I_{k_1+k_2}\right)\frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}X_{it}^{*\prime}\right)\left(\text{vec}\left(X_{it}^{*}X_{it}^{*\prime}\right)\right)^{\prime}\left(\left(\gamma-\hat{\gamma}_{N,t}\right)\otimes I_{k_1+k_2}\right) \\
& +\left(\left(\gamma-\hat{\gamma}_{N,t}\right)^{\prime}\otimes I_{k_1+k_2}\right)\frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}X_{it}^{*\prime}\right)\left(\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\right)^{\prime} \\
& +\frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\left(\text{vec}\left(X_{it}^{*}X_{it}^{*\prime}\right)\right)^{\prime}\left(\left(\gamma-\hat{\gamma}_{N,t}\right)\otimes I_{k_1+k_2}\right)\overset{P}{\to}0 \text{ as } N\to\infty
\end{aligned}$$

According to the above what remains is to show that $\frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\left(\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\right)^{\prime}\overset{P}{\to}$ Θ_t as $N\to\infty$. This is done by using the result in Lemma 3 once again and we obtain the following

$$\begin{aligned}
& \frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\left(\text{vec}\left(X_{it}^{*}w_{it}^{*\prime}+X_{it}^{*}\eta_{0it}^{*\prime}\right)\right)^{\prime}\overset{P}{\to} \\
& E\left(\text{vec}\left(\tilde{X}_{it}w_{it}^{\prime}\right)\left(\text{vec}\left(\tilde{X}_{it}w_{it}^{\prime}\right)\right)^{\prime}\right)+E\left(\text{vec}\left(\tilde{X}_{it}\tilde{\eta}_{0it}^{\prime}\right)\left(\text{vec}\left(\tilde{X}_{it}\tilde{\eta}_{0it}^{\prime}\right)\right)^{\prime}\right) \\
& +E\left(\text{vec}\left(\tilde{X}_{it}w_{it}^{\prime}\right)\left(\text{vec}\left(\tilde{X}_{it}\tilde{\eta}_{0it}^{\prime}\right)\right)^{\prime}\right)+E\left(\text{vec}\left(\tilde{X}_{it}\tilde{\eta}_{0it}^{\prime}\right)\left(\text{vec}\left(\tilde{X}_{it}w_{it}^{\prime}\right)\right)^{\prime}\right) \\
= & E\left(\text{vec}\left(\tilde{X}_{it}w_{it}^{\prime}\right)\left(\text{vec}\left(\tilde{X}_{it}w_{it}^{\prime}\right)\right)^{\prime}\right)+E\left(\text{vec}\left(\tilde{X}_{it}\tilde{\eta}_{0it}^{\prime}\right)\left(\text{vec}\left(\tilde{X}_{it}\tilde{\eta}_{0it}^{\prime}\right)\right)^{\prime}\right) \\
= & \text{Var}\left(\tilde{X}_{it}\left(w_{it}+\tilde{\eta}_{0it}\right)^{\prime}\right)=\Theta_t
\end{aligned}$$

where we have also used that $\tilde{X}_{it}w_{it}^{\prime}$ and $\tilde{X}_{it}\tilde{\eta}_{0it}^{\prime}$ are uncorrelated.

Altogether we have shown the results stated in Proposition 2. □

References

Adda, J. and J-M. Robin, 1998, Estimation from cross-sections of integrated time-series, Working Paper n 9802, C.E.P.R.E.M.A.P., France.

Baltagi, B., 1995, Econometric analysis of panel data (Wiley, New York).

Banerjee, A., 1999, Panel data unit roots and cointegration: An overview, Oxford Bulletin of Economics and Statistics, Special Issue, 607-630.

Grunfeld, Y., 1961, The interpretation of cross section estimates in a dynamic model, Econometrica 29, 397-405.

Hsiao, C., 1986, Analysis of panel data (Cambridge University Press, Cambridge).

Matays, L. and P. Sevestre, 1992, The econometrics of panel data (Kluwer Academic Publishers, Boston, Massachusetts).

Park, J.Y. and P.C.B. Phillips, 1989, Statistical inference in regressions with integrated processes: Part 2, Econometric Theory 5, 95-131.

Pesaran, M.H. and R. Smith, 1995, Estimating long-run relationships from dynamic heterogeneous panels, Journal of Econometrics 68, 79-113.

Phillips, P.C.B., 1986, Understanding spurious regressions in econometrics, Journal of Econometrics 33, 311-40.

Phillips, P.C.B. and H.R. Moon, 1999, Linear regression limit theory for nonstationary panel data, Econometrica 67, 1057-1111.