Spurious Cross Sectional and Mindless Panel Regressions*

Donggyu Sul
University of Texas at Dallas
Incomplete

October 21, 2016

Abstract

It is widely known that the estimation results from estimating cross sectional and time series regressions are different. In addition, pooled OLS estimators are usually different from fixed effects estimators. This paper provides an intuitive explanation for why such anomalies are observed. We assume that a random variable can be decomposed into three components: A time invariant variable with non-zero mean; a common comoponent with zero mean; and an idiosyncratic time varying component with zero mean. The relationship among time invariant means is called the "static" relationship. The relationships among the common and idiosyncratic components are called the "common-dynamic" and "idio-dynamic" relationship, respectively. Whenever the static relationship is different from the dynamic relationships, the cross sectional regressions cannot identify or estimate the static relationship. To identify and estimate the static relationship consistently, we propose a modified between group estimator. When the panel data are cross sectionally dependent, the idio-dynamic relationship can be identified and estimated by augmenting the regression with the common factors. However when the parameters of interest are the common-dynamic relationship, the time series regression with cross sectional aggregates or PC estimates of the common factors should be estimated.

JEL Classification: C10, C21, C33

Keywords: *Cross Sectional Regressions, Panel Regressions, Static Relationship, Common-Dynamic Relationship, Idio-Dynamic Relationship.*

*Thanks for Bumsoo Kim at Korea University for providing the example of twin study and Masao Ogaki for Easterline paradox.

1 Introduction

Almost all recent empirical studies use panel data across all areas of economic, social, political, marketing, and financial sciences. Since panel data has two dimensions – cross section and time series, naturally several estimators – cross sectional, time series, pooled OLS, fixed effects, random effects and between group estimators – can be used to examine theories or establish empirical findings. Very often, however, researchers find that point estimates across estimators are quite different.

No theoretical study examines the underlying reasons for why all estimators provide different answers. This paper deals with this thorny task and provides intuitive explanations. Let's define the questions of the paper more precisely. Suppose that one is interested in finding a statistical relationship between two variables $\{y\}$ and $\{x\}$, both of which have two dimensions: Cross sectional and time series dimensions, which we denote as $i = 1, ..., N$ and $t = 1, ..., T$. Now let's ask the following two questions:

(i) Is the cross sectional regression of y on x different from the time series regression of y on x in general? If so, how should we interpret these results? What do the cross sectional and time series regressions estimate?

(ii) Alternatively, are the panel regressions of y on x (with random or fixed effects) different either from cross sectional and time series regressions? If so, how do we interpret these results?

The main thrust of this paper stems from viewing the panel data based on the following data generating process.

$$y_{it} = \mu_{yi} + y_{it}^{\dagger}, \quad x_{it} = \mu_{xi} + x_{it}^{\dagger}, \tag{1}$$

where both variables are decomposed into the time-invariant random variables, μ_{yi} and μ_{xi}, and the time-variant random variables, $y_{it}^{\dagger}$ and $x_{it}^{\dagger}$ with zero means. The time-invariant relationship will be called the '*static relationship*'. It is given by

$$\mu_{yi} = \alpha_{\mu} + \beta\mu_{xi} + v_i. \tag{2}$$

The time-varying variables can be further decomposed into common and idiosyncratic (or individual specific) components.

$$y_{it}^{\dagger} = h_{it}^{y} + y_{it}^{o}, \quad x_{it}^{\dagger} = h_{it}^{x} + x_{it}^{o}, \tag{3}$$

and the relationships between the common components and between the idiosyncratic components will be called '*common and idiosyncratic dynamic relationship*', respectively.

$$h_{it}^{y} = \alpha_h + \phi h_{it}^{x} + e_{it}, \quad y_{it}^{o} = \alpha_o + \gamma x_{it}^{o} + m_{it}. \tag{4}$$

There is a possibility that the static relationship (β) may not be equal to the dynamic relationship (γ or ϕ).

The econometric modelling in (2) and (3) delivers intuitive explanations for the reason why the cross sectional and pooled OLS (or random effects) estimators are so different from the time series and fixed effects estimators in general. The static relationship is the relationship between time invariant random variables, so this relationship can be thought as the relationship between the two random variables of y_{it} and x_{it} in the long run. In other words, the static relationship can be interpreted as the long run relationship if the long run value of each variable is well defined. The long run value of a random variable $\{x\}$ exists if and only if $\text{plim}_{T\to\infty}T^{-1}\sum x_{it} = \mu_{xi}$. Meanwhile the dynamic relationship is a temporal relationship between the two time varying components if the static relationship exists between the two variables. Figure 1 exhibits an example in which the difference between the static and dynamic relationship when the long run equilibrium for each variable is well defined: The static relationship is positive meanwhile the dynamic relationship is negative. If a positive or negative shock enters into the system, over time the shock is neutralized, and then the dynamic relationship becomes possibly negative. However the static relationship does not always imply the long run relationship and vice versa. If the long run value of each variable is not defined (the probability limit of the time series mean does not converge to a constant), then any meaningful static relationship does not exist.

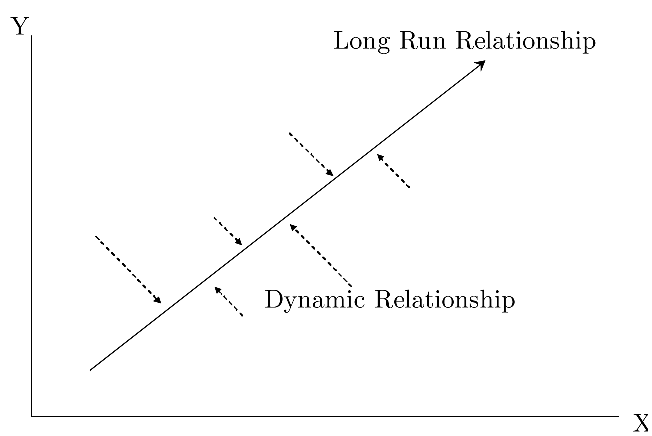

Figure 1: An Example of Static and Dynamic Relationships

When the time invariant components in (1) are dominant and there exists a meaningful static relationship, the cross sectional regressions may be estimating the static relationship. In this case, the time varying term $x_{it}^{\dagger}$ can be treated as the measurement error. As long as the measurement errors can be ignored (at least asymptotically), the cross sectional and pooled OLS regressions estimate the static relationship. Meanwhile the time series and the fixed effects panel regressions eliminate the time invariant terms completely, so that these estimators are purely capturing the dynamic relationship.

If the time variant terms are large enough, then it is unknown what the cross sectional regressions estimate even when the static relationship is well defined. The time series and panel fixed effects regressions estimate the dynamic relationship, but the pooled OLS and the cross sectional regressions estimate a mixture relationship between the static and dynamic relationship. However the between group estimator consistently estimates the static relationship with a large T but suffer from the bias with a small T. To attenuate this bias, we propose a modified between group estimator.

When random variables are cross sectionally dependent, the time varying components can be further decomposed into the common and idiosyncratic components in (3). Under cross sectional dependence, the cross sectional estimators contain all three relationships: The static relationship between the two time invariant terms, the dynamic relationship between the common components, and the dynamic relationship between the idiosyncratic components. We will show later, but the cross sectional regressions with the micro survey data are less influenced by the dynamic relationship among the common components. But the cross sectional regressions with cross sectionally aggregated data such as state level and cross-country data are affected by the relationship between the common components.

When the time varying components contain an integrated (or nonstationary) component, the static relationship is not well-defined since the time series averages are not well-defined in general. If the nonstationary variables are cointegrated, then the dynamic relationship becomes the long run relationship. The nonstationary time varying components become a very serious issue with cross sectionally integrated data. We will show later but in this case, the cross sectional regressions estimate the dynamic relationship between the common components if the idiosyncratic components are stationary. But if both idiosyncratic and common components are integrated, then the cross sectional regressions estimate the mixture relationship of the two dynamic relationships between the common and idiosyncratic components.

The rest of the paper consists of four sections. The next section provides basic econometric analysis on the five different estimators under somewhat strong regularity conditions, and then shows a couple of important empirical puzzles to demonstrate the economic meanings of the static and dynamic relationships. Also Section 2 deals with the identification and the estimation of the static relationship first, and then demonstrates how such identification procedure coincides the previous estimation methods in the literature. Section 3 provides how the cross sectional dependence and the level of cross sectional aggregation influence on the cross sectional estimators.

Section 4 provides Monte Carlo simulations and Section 5 concludes. Appendix includes technical proofs and derivations.

2 Static and Dynamic Relationship

In order to demonstrate the central issue of this paper, we exclude the common components on purpose in this section. Of course, in reality it is hard to find random variables which don't have any cross sectional dependence. For the time being, we will ignore the dependent term by assuming that the magnitude of these terms is relatively far smaller than time invariant and idiosyncratic components. That is, we set $\phi \in \varnothing$ but $\gamma \neq \beta$. As we will show shortly, neither the cross sectional estimator or the pooled regression estimate β or γ. Meanwhile the fixed effects (or within group) regressions estimate γ consistently. When $\beta \neq \gamma$, the cross sectional estimator is more or less similar to the pooled estimator but the within group estimator becomes distinctively different from both the cross sectional and pooled estimator.

2.1 Relationships Between the Five Estimators

Consider the following T cross sectional regressions, N time series regressions, two panel regressions and one between group regression.

$$\begin{array}{lll} \text{Cross section} & y_{it} = a_t + \mathbf{x}_{it}\mathbf{b}_t + u_{it} & \text{for each } t = 1, ..., T \\ \text{Time Series} & y_{it} = a_i + \mathbf{x}_{it}\mathbf{b}_i + e_{it} & \text{for each } i = 1, ..., N \\ \text{Pooled OLS} & y_{it} = a + \mathbf{x}_{it}\mathbf{b}_1 + \varepsilon_{it} & \\ \text{Fixed Effects} & y_{it} = a_i + \mathbf{x}_{it}\mathbf{b}_2 + \epsilon_{it} & \\ \text{Between Group} & \bar{y}_i = a + \bar{\mathbf{x}}_i\mathbf{b}_3 + \bar{\varepsilon}_i & \bar{y}_i = T^{-1}\sum^T y_{it},\ \bar{\mathbf{x}}_i{=}T^{-1}\sum^T \mathbf{x}_{it} \end{array} \tag{5}$$

where $\mathbf{x}_{it}$ is a $1 \times k$ vector of explanatory variables. Note that the true relationships are given by

$$\mu_{yi} = \boldsymbol{\mu}_{xi}\boldsymbol{\beta} + v_i, \quad y_{it}^o = \mathbf{x}_{it}^o\boldsymbol{\gamma} + m_{it}. \tag{6}$$

Here $\mathbf{x}_{it}^o$ and $\boldsymbol{\mu}_{xi}$ are assumed to be strictly exogenous and stationary. These assumptions are very strong, but they effectively deliver the core concept of this paper. The dynamic relationship can be heterogeneous either across t or i, but such heterogeneity does not influence the main thrust of this paper. However heterogeneity in $\boldsymbol{\gamma}$ can influence on the pooled estimators. This important issue is considered separately. See Sul (2015) for a more detailed discussion.

The relationship between μ_{yi} and $\boldsymbol{\mu}_{xi}$ can be interpreted as static or steady-state relationship in the sense that this relationship can be revealed when there are no time varying shocks, y_{it}^o and

x^o_{it}. It is important to note that the static relationship implies a 'long run' relationship rather than 'short run' relationship. To see this, taking time series mean yields

$$\frac{1}{T}\sum_{t=1}^{T} y_{it} = \mu_{yi} + O_p\left(T^{-1/2}\right) = \boldsymbol{\mu}_{xi}\boldsymbol{\beta} + v_i + O_p\left(T^{-1/2}\right),$$

as long as y^o_{it} and $\mathbf{x}^o_{it}$ have zero means and finite variances. In other words, when both $T^{-1}\sum_{t=1}^{T} y^o_{it}$ and $T^{-1}\sum_{t=1}^{T} x^o_{it}$ converge to zero in probability, the static or steady-state relationship can be well-defined as $T \to \infty$. Naturally the between group estimator estimates the static relationship consistently as $T \to \infty$ since the regressors and regressand are time series averages.

Given (6) the true DGP can be expressed as

$$\begin{aligned} y_{it} &= a_i + \mathbf{x}_{it}\boldsymbol{\gamma} + m_{it} &(7)\\ &= v_i + \mathbf{x}_{it}\boldsymbol{\beta} + k_{it} &(8)\end{aligned}$$

where

$$a_i = \boldsymbol{\mu}_{xi}\left(\boldsymbol{\beta} - \boldsymbol{\gamma}\right) + v_i, \quad k_{it} = \mathbf{x}^o_{it}\left(\boldsymbol{\gamma} - \boldsymbol{\beta}\right) + m_{it}.$$

When $\boldsymbol{\beta} = \boldsymbol{\gamma}$, the relationship between $\{y_{it}\}$ and $\{\mathbf{x}_{it}\}$ becomes unique. Otherwise, it is hard to define the relationship between the two variables if $\boldsymbol{\beta} \neq \boldsymbol{\gamma}$. Note that when $\boldsymbol{\beta} \neq \boldsymbol{\gamma}$, there is a correlation between fixed effects and regressors (a_i and $\mathbf{x}_{it}$) in (7), and the correlation coefficient on $\mathbf{x}_{it}$ represents the 'dynamic relationship' between $\{y\}$ and $\{\mathbf{x}\}$. Meanwhile, in (8) there is no correlation between the fixed effects and regressors (v_i and $\mathbf{x}_{it}$) but the regressors are correlated with the regression errors.

Here we provide more formal explanations of the five estimators. First we need the following assumption.

Assumption A: (Existence of Long Run Values) *y^o_{it} and $\mathbf{x}^o_{it}$ have zero means and finite fourth moments over i for each t, and are weakly dependent and stationary over t with autocovariance sequence $\Gamma_i(h) = \mathrm{E}\left(\mathbf{x}^{o\prime}_{it}\mathbf{x}^o_{it+h}\right)$ satisfying $\sum_{h=1}^{\infty} h\left|\Gamma_i(h)\right| = \mathbf{M} < \infty$.*

Next, define

$$\begin{array}{c} \tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}, \quad \tilde{\mathbf{x}}^+_{it} = \mathbf{x}_{it} - \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{it} \quad \bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it} \\ \tilde{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \tilde{y}^+_{it} = y_{it} - \frac{1}{N}\sum_{i=1}^{N} y_{it} \quad \bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it} \\ \bar{\mathbf{x}}^*_i = \bar{\mathbf{x}}_i - \frac{1}{N}\sum_{i=1}^{N}\bar{\mathbf{x}}_i \quad \tilde{\mathbf{x}}^*_{it} = \mathbf{x}_{it} - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{x}_{it} \\ \bar{y}^*_i = \bar{y}_i - \frac{1}{N}\sum_{i=1}^{N}\bar{y}_i \quad \tilde{y}^*_{it} = y_{it} - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} y_{it} \end{array},$$

then we have

$$\tilde{\mathbf{x}}_{it} = \mathbf{x}^o_{it} - \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}^o_{it}, \qquad \bar{\mathbf{x}}_i = \boldsymbol{\mu}_{xi} + \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}^o_{it}, \qquad \tilde{\mathbf{x}}^+_{it} = \boldsymbol{\mu}_{xi} - \bar{\boldsymbol{\mu}}_x + \left(\mathbf{x}^o_{it} - \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}^o_{it}\right)$$

$$\bar{\mathbf{x}}^*_i = \boldsymbol{\mu}_{xi} - \bar{\boldsymbol{\mu}}_x + \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}^o_{it} - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{x}^o_{it}\right), \quad \tilde{\mathbf{x}}^*_{it} = \boldsymbol{\mu}_{xi} - \bar{\boldsymbol{\mu}}_x + \left(\mathbf{x}^o_{it} - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{x}^o_{it}\right).$$

Now, further define the five estimators; cross sectional ("cross"), time series ("time"), pooled OLS ("POLS"), within group ("WG"), and between group ("BG") estimators as

$$\begin{aligned}
\hat{\mathbf{b}}_{\text{cross,t}} &: = \hat{\mathbf{b}}_t = \left(\tilde{\mathbf{x}}^{+\prime}_t \tilde{\mathbf{x}}^+_t\right)^{-1} \tilde{\mathbf{x}}^{+\prime}_t \tilde{\mathbf{y}}^+_t, \\
\hat{\mathbf{b}}_{\text{time,i}} &: = \hat{\mathbf{b}}_i = \left(\tilde{\mathbf{x}}'_i \tilde{\mathbf{x}}_i\right)^{-1} \tilde{\mathbf{x}}'_i \tilde{\mathbf{y}}_i, \\
\hat{\mathbf{b}}_{\text{pols}} &: = \hat{\mathbf{b}}_1 = \left(\tilde{\mathbf{x}}^{*\prime} \tilde{\mathbf{x}}^*\right)^{-1} \tilde{\mathbf{x}}^{*\prime} \tilde{\mathbf{y}}^*, \\
\hat{\mathbf{b}}_{\text{wg}} &: = \hat{\mathbf{b}}_2 = \left(\tilde{\mathbf{x}}' \tilde{\mathbf{x}}\right)^{-1} \tilde{\mathbf{x}}' \tilde{\mathbf{y}}, \\
\hat{\mathbf{b}}_{\text{bw}} &: = \hat{\mathbf{b}}_3 = \left(\bar{\mathbf{x}}^{*\prime} \bar{\mathbf{x}}^*\right)^{-1} \bar{\mathbf{x}}^{*\prime} \bar{\mathbf{y}}^*.
\end{aligned}$$

where $\tilde{\mathbf{x}}^+_t = \left(\tilde{\mathbf{x}}^{+\prime}_{1t}, ..., \tilde{\mathbf{x}}^{+\prime}_{Nt}\right)'$, $\tilde{\mathbf{x}}_i = \left(\tilde{\mathbf{x}}'_{i1}, ..., \tilde{\mathbf{x}}'_{iT}\right)'$, $\tilde{\mathbf{x}}^* = \left(\tilde{\mathbf{x}}^{*\prime}_{11}, ..., \tilde{\mathbf{x}}^{*\prime}_{1T}, \tilde{\mathbf{x}}^{*\prime}_{21}, ..., \tilde{\mathbf{x}}^{*\prime}_{NT}\right)'$, $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}'_{11}, ..., \tilde{\mathbf{x}}'_{1T},$ $\tilde{\mathbf{x}}'_{21}, ..., \tilde{\mathbf{x}}'_{NT})'$, $\bar{\mathbf{x}}^* = (\bar{\mathbf{x}}^{*\prime}_1, ..., \bar{\mathbf{x}}^{*\prime}_N)'$, and $\{y\}$ is also defined in the same way.

Next, $\mathbf{x}^o_i = (\mathbf{x}^{o\prime}_{i1}, ..., \mathbf{x}^{o\prime}_{iT})'$, $\mathbf{x}^o_t = (\mathbf{x}^{o\prime}_{1t}, ..., \mathbf{x}^{o\prime}_{Nt})'$, $\mathbf{x}^o = (\mathbf{x}^{o\prime}_{11}, ..., \mathbf{x}^{o\prime}_{NT})'$, $\bar{\mathbf{x}}^o = (\bar{\mathbf{x}}^{o\prime}_1, ..., \bar{\mathbf{x}}^{o\prime}_N)'$, $\bar{\mathbf{x}}^o_i = T^{-1}\sum_{t=1}^{T}\mathbf{x}^o_{it}$, and $\{y^o\}$ is also defined in the same way. Let

$$\mathbf{Q}_{xoi} = \lim_T E\left(\frac{\mathbf{x}^{o\prime}_i \mathbf{x}^o_i}{T}\right) \quad \mathbf{Q}_{xot} = \lim_N E\left(\frac{\mathbf{x}^{o\prime}_t \mathbf{x}^o_t}{N}\right) \quad \mathbf{Q}_{xo} = \lim_N E\left(\frac{\mathbf{x}^{o\prime}\mathbf{x}^o}{TN}\right) \quad \boldsymbol{\Omega}_{xo} = \lim_N E\left(\frac{\bar{\mathbf{x}}^{o\prime}\bar{\mathbf{x}}^o}{N}\right)$$

$$\mathbf{Q}_{xyoi} = \lim_T E\left(\frac{\mathbf{x}^{o\prime}_i \mathbf{y}^o_i}{T}\right) \quad \mathbf{Q}_{xyot} = \lim_N E\left(\frac{\mathbf{x}^{o\prime}_t \mathbf{y}^o_t}{N}\right) \quad \mathbf{Q}_{xyo} = \lim_N E\left(\frac{\mathbf{x}^{o\prime}\mathbf{y}^o}{TN}\right) \quad \boldsymbol{\Omega}_{xyo} = \lim_N E\left(\frac{\bar{\mathbf{x}}^{o\prime}\bar{\mathbf{y}}^o}{N}\right)$$

$$\mathbf{Q}_{\mu x} = \lim_N E\frac{\left(\boldsymbol{\mu}_x - \frac{1}{N}\sum\boldsymbol{\mu}_{xi}\right)'\left(\boldsymbol{\mu}_x - \frac{1}{N}\sum\boldsymbol{\mu}_{xi}\right)}{N}$$

$$\mathbf{Q}_{\mu xy} = \lim_N E\frac{\left(\boldsymbol{\mu}_x - \frac{1}{N}\sum\boldsymbol{\mu}_{xi}\right)'\left(\mu_y - \frac{1}{N}\sum\mu_{yi}\right)}{N}$$

Then the probability limit for each estimator is given by

Theorem 1 (Asymptotic Consistency) *When the true data generating process is given by (6) and $\beta \neq \gamma$, the probability limits for five estimators under Assumption A are given by*

(*i*) $\text{plim}_{N\to\infty}\hat{\mathbf{b}}_{\text{cross},t} = \mathbf{W}_t\boldsymbol{\beta} + (\mathbf{I} - \mathbf{W}_t)\boldsymbol{\gamma} \neq \boldsymbol{\beta}$,

(*ii*) $\text{plim}_{T\to\infty}\hat{\mathbf{b}}_{\text{time},i} = \mathbf{Q}^{-1}_{xoi}\mathbf{Q}_{xyoi} = \boldsymbol{\gamma}$,

(*iii*) $\text{plim}_{N\to\infty}\hat{\mathbf{b}}_{\text{bw}} = \left(\mathbf{Q}_{\mu x} + T^{-1}\boldsymbol{\Omega}_{xo}\right)^{-1}\left(\mathbf{Q}_{\mu xy} + T^{-1}\boldsymbol{\Omega}_{xyo}\right)$

(*iv*) $\text{plim}_{N,T\to\infty}\hat{\mathbf{b}}_{\text{pols}} = \mathbf{W}\boldsymbol{\beta} + (\mathbf{I} - \mathbf{W})\boldsymbol{\gamma} \neq \boldsymbol{\beta}$

(*v*) $\text{plim}_{N,T\to\infty}\hat{\mathbf{b}}_{\text{wg}} = \mathbf{Q}^{-1}_{xo}\mathbf{Q}_{xyo} = \boldsymbol{\gamma}$

(*vi*) $\text{plim}_{N,T\to\infty}\hat{\mathbf{b}}_{\text{bw}} = \boldsymbol{\beta}$
where $\mathbf{W}_t = (\mathbf{Q}_{\mu xt}+\mathbf{Q}_{xot})^{-1}\mathbf{Q}_{\mu xt}$, *and* $\mathbf{W} = (\mathbf{Q}_{\mu x}+\mathbf{Q}_{xo})^{-1}\mathbf{Q}_{\mu x}$.

See Appendix A for a detailed proof. Here some heuristic explanations are provided. In general, the cross sectional regressions do not estimate either the static relationship or the dynamic relationship. The time varying components become nuisance terms and behave like measurement errors. As the magnitude of the time varying components is getting smaller, the value of $\mathbf{Q}_{xot}$ is getting smaller too, so that the weight matrix $\mathbf{W}_t$ is getting close to an identity matrix. Then the cross sectional and POLS regressions estimate the static relationship. Meanwhile the time series and within group regressions estimate only the dynamic relationship since the two estimators are built after taking off the time series means, or equivalently, after removing the static relationship. Lastly, the between group regression estimates the static relationship if the number of time series observations is sufficiently large. Hence in general, we can expect the following simple rule under the cross sectional independence.

$$\begin{array}{lllll} \text{plim}_{N\to\infty}\hat{\mathbf{b}}_{\text{cross},t} & \leq & \text{plim}_{N\to\infty}\hat{\mathbf{b}}_{\text{bw}} & \leq & \text{plim}_{T\to\infty}\hat{\mathbf{b}}_{\text{time},i} \text{ if } \boldsymbol{\beta}\leq\boldsymbol{\gamma}, \\ \text{plim}_{N\to\infty}\hat{\mathbf{b}}_{\text{cross},t} & > & \text{plim}_{N\to\infty}\hat{\mathbf{b}}_{\text{bw}} & > & \text{plim}_{T\to\infty}\hat{\mathbf{b}}_{\text{time},i} \text{ if } \boldsymbol{\beta}>\boldsymbol{\gamma}. \end{array} \tag{9}$$

The results in (9) and Theorem 1 provide important implication: The cross sectional regressions do not estimate the static relationship if $\boldsymbol{\beta}\neq\boldsymbol{\gamma}$. More importantly, the cross sectional results should be carefully interpreted whenever the variables of interest are time varying.

Remark 1: Interpretations on the Static and Dynamic Relationships It is important to address the fact that the economic interpretations on the static and dynamic relationships are very different. Let $k=1$ (single regressor), $\beta=1$ and $\gamma=-1$. Then eq. (6) can be rewritten as

$$\mu_{yi} = \mu_{xi} + v_i, \quad y_{it}^o = -x_{it}^o + m_{it}.$$

Further assume that both time varying components are small enough, so that they can be ignored.

$$y_{it} = \mu_{yi} + y_{it}^o \simeq \mu_{yi}, \quad x_{it} = \mu_{xi} + x_{it}^o \simeq \mu_{xi}.$$

One runs the following cross sectional regression at $t=0$ and obtains the point estimate of 0.95. Then the predicted value of y_{i0} becomes

$$\hat{y}_{i0} = 0.95 \times x_{i0}.$$

Typically this estimation result is interpreted as follows: If x_{i0} increases by one, y_{i0} will increase 0.95. This interpretation is, however, problematic. As we assumed before, the dynamic relationship between $\{y\}$ and $\{x\}$ is negative. Hence as $\{x\}$ increases by one, $\{y\}$ should decrease by one since $\gamma = -1$. The static or long run value of $\{x\}$ cannot change over time. This confusion arises especially when $\{x\}$ variables have time varying components.

Remark 1 and Theorem 1 state that whenever $\boldsymbol{\beta} \neq \boldsymbol{\gamma}$, the cross sectional regressions become spurious. However when the regressors are not time varying, the cross sectional estimates are well-defined. Consider drug experiments as an example. Let x_i be the amount of a drug dosage for the ith subject and y_i be the health outcome. Then the marginal effect, $\Delta y/\Delta x$, can be captured by running y_i on x_i and a constant. In this case, the static relationship β can be interpreted that as the drug dosage is increased by one, the health outcome increases by β, which is the typical interpretation. However when the regressors are time varying, the cross sectional results should be carefully interpreted. Here is another example.

Example 1: (SAT scores and Parents Income)

Every year a college board reports the relationship between SAT scores and parents' income. The cross sectional regression result in 2013 is shown in the below.

$$\begin{aligned} \widehat{SAT}_i &= \underset{(49.44)}{-116.5} + \underset{(4.36)}{143.8} \times \ln\left(Income_i\right), \end{aligned} \tag{10}$$

where SAT_i stands for an individual average SAT score (total score is 2400), $Income_i$ is the parents' income of the ith student. The standard errors are reported in parentheses. Figure 2 plots the nonlinear relationship between the level of parents' income and SAT scores. Suppose that the time varying components are small enough to be ignored. Then can we interpret this result as follows: As the children's SAT score increases by 144 as $\ln(income)$ increases by 1? Alternatively, if one makes more income, then will his (her) children get better SAT score? The answer is no. If parents' income data are not time varying as we assumed, then the parents' income cannot either increase or decrease. Hence this result should be interpreted that the SAT scores of children with richer parents are higher than those of children with less rich parents. Here the income variable may capture or provide condensed information about the quality of a school zone, peer group effects, parents' education level, etc. Suppose that the true determinant were the quality of a school zone, then parents who want to increase their children's SAT score should not try to make more money but try to move in a better school zone. In fact, as parents work more to make more money, their children's SAT scores could decrease more likely because parents may pay less attention on their children's education. If so, the cross sectional regression result in (10) is underestimating the static relationship, and the true static relationship must be much higher than $\hat{b}_{\text{cross}}$.

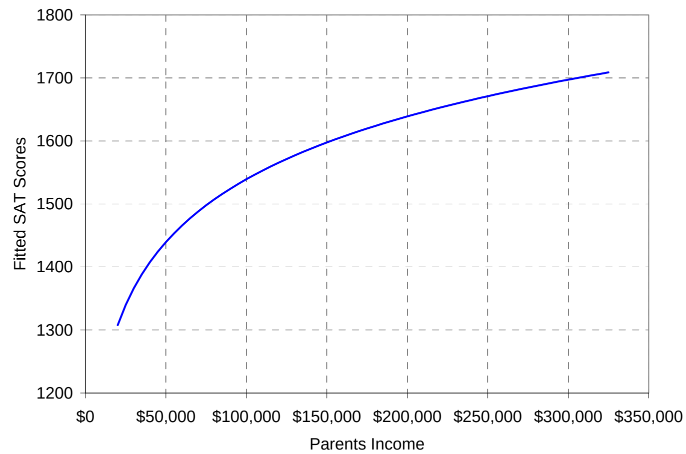

Figure 2: Fitted SAT Scores and Parents Income, 2013.

It is useful to note that the fixed effects regressions do not always estimate the dynamic relationship. Fixed effects have also been used in many cross sectional data analysis where the time dimension of the data is replaced by either family, region, or states. In this case, the dynamic relationship cannot be captured by running the fixed effects regressions. See Remark 2 for more detailed explanation.

Remark 2: (Pseudo Fixed Effects: Twin Studies) Twin data are often used in health economics to control for unobserved characteristics of mothers or families. Among many examples, see Almond, Chay and Lee (2005) and Royer (2009) for the relationship between low birth weight and infant mortality rates, and Behrman and Rosenzweig (2002), and Black, Devereus and Slavanes (2007) for the relationship between low birth weight and schooling outcomes. The typical data generating process or true model is assumed to be

$$y_{ij} = a + \mathbf{x}_{ij}\mathbf{b} + \mathbf{z}_j\boldsymbol{\delta} + v_{ij},$$

where j stands for the jth mother in the sample, and $\mathbf{z}_j$ is the vector of unobserved mother's characteristics which are generally correlated with the observed variables $\mathbf{x}_{ij}$. Running the pooled OLS regression without including $\mathbf{z}_j$ leads to an inconsistent estimator of $\mathbf{b}$ due to the missing variables. The twin data are used for controlling out the unobserved variable $\mathbf{z}_j$. To be specific, researchers have used usually the difference in the twin variables to isolate the characteristic of mothers. Let the subscript (i, i') denote a set of twins. Then by taking the difference between the twins, one can eliminate $\mathbf{z}_j$ variables from the regression. That is,

$$y_{ij} - y_{i'j} = \left(\mathbf{x}_{ij} - \mathbf{x}_{i'j}\right) \mathbf{b} + \left(v_{ij} - v_{i'j}\right).$$

This fixed effects estimator is not related to the dynamic relationship.

As Theorem 1 states, the consistency of the cross sectional and POLS estimators hinge on the relative variance ratio of the time varying to the time invariant components. As this ratio increases (the time varying components are getting bigger), the estimators become close to the dynamic relationship. Do the relative variance ratios change (or reduce) depending on the level of aggregation? In other words, are the relative variance ratios in the cross sectional survey regressions (individual level, such as PSID or NLYS) smaller than those in the cross sectional regressions with state level data? The answer to this question is dependent on the existence of common factors. If there is no common factor, or equivalently if the data is not cross sectionally dependent, then the level of aggregation is not related to the relative variance ratio. More formally we have:

Remark 3: (Impact of the Level of Aggregation on Cross Sectional Estimator under Cross Sectional Independence)

For an intuitive explanation, we assume the true data generating process for x_{it} is

$$x_{it} = \mu_{xi} + x_{it}^o, \quad y_{it} = \mu_{yi} + y_{it}^o, \text{ for } i = 1, ..., N.$$

Let n be the number of cross sectional units for each subgroup $j = 1, ..., J$. Here I assume that the number of cross sectional units for each j is identical for notational simplicity. The heterogeneous group number does not alter the result at all. Consider the following aggregated data.

$$x_{jt} = \frac{1}{n}\sum_{i=1}^{n} x_{it}, \quad y_{jt} = \frac{1}{n}\sum_{i=1}^{n} y_{it}, \text{ for } j = 1, ..., J.$$

Note that $n = N/J$. Consider running the following two regressions for a particular t.

$$y_{it} = a_1 + b_1 x_{it} + u_{it}, \quad (11)$$

$$y_{jt} = a_2 + b_2 x_{jt} + u_{jt}. \quad (12)$$

Then it is easy to show that as $n, J, N \to \infty$,

$$\text{plim}_{N\to\infty}\hat{b}_1 = \text{plim}_{n\to\infty}\hat{b}_2 = b_t = W_t\beta + (1 - W_t)\gamma, \quad (13)$$

but the convergence rate for $\hat{b}_1$ is $\sqrt{N}$ but that of $\hat{b}_2$ is $\sqrt{J}$. Hence $\hat{b}_1$ is more efficient than $\hat{b}_2$.

The proof for (13) is obvious and hence it is omitted. Here we provide a direct and heuristic explanation. It is easy to show that the convergence rate of $\hat{b}_1$ is $\sqrt{N}$ as long as the probability limit of $\hat{b}_1$ exists. Of course, the disaggregated cross sectional estimator, $\hat{b}_1$, is not equal to β or γ, but its limit exists. The cross sectional regression estimator with the aggregated data is less efficient simply due to the smaller number of observations, $J < N$. The variance of the regression error u_{jt} in (12) shrinks at an $O(\sqrt{n})$ rate but at the same time the variance of the demeaned regressor x_{jt}, $\tilde{x}_{jt} = x_{jt} - J^{-1}\sum_{j=1}^{J} x_{jt}$, also shrinks at the same rate. This is the reason why the aggregated data becomes less efficient. More importantly, the cross sectional aggregated data is not useful for estimating the static relationship. To identify and estimate the static relationship, one should use the time series averages, which will be discussed in detail in the next subsection.

2.2 Estimation of Static Relationships

As Theorem 1 states, with a large T, the between estimators $\hat{\mathbf{b}}_{\text{bw}}$ estimate the static relationships consistently. However, for a finite T dimension, the between estimators $\hat{\mathbf{b}}_{\text{bw}}$ become inconsistent. The order of inconsistency for large T can be expressed as follows

$$\text{plim}_{N\to\infty}\hat{\mathbf{b}}_{\text{bw}} = \boldsymbol{\beta} + T^{-1}\left(\mathbf{Q}_{\mu x} + T^{-1}\boldsymbol{\Omega}_{xo}\right)^{-1}\left(\boldsymbol{\Omega}_{xyo} - \boldsymbol{\Omega}_{xo}\boldsymbol{\beta}\right) = \boldsymbol{\beta} + O\left(T^{-1}\right) \tag{14}$$

To reduce the inconsistency up to a $O\left(T^{-2}\right)$ term, we can consider the following modified between group estimator.

$$\hat{\mathbf{b}}_{\text{bw}}^{+} = \left(\bar{\mathbf{x}}^{*\prime}\bar{\mathbf{x}}^{*} - \bar{\mathbf{x}}^{+\prime}\bar{\mathbf{x}}^{+}\right)^{-1}\left(\bar{\mathbf{x}}^{*\prime}\bar{\mathbf{y}}^{*} - \bar{\mathbf{x}}^{+\prime}\bar{\mathbf{y}}^{+}\right)$$

where

$$\tilde{\mathbf{x}}_{it}^{+} = \mathbf{x}_{it} - \frac{2}{T}\sum_{t=1}^{T/2}\mathbf{x}_{it} = \mathbf{x}_{it}^{o} - \frac{2}{T}\sum_{t=1}^{T/2}\mathbf{x}_{it}^{o}$$

and

$$\bar{\mathbf{x}}_{i}^{+} = \frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{+} = [-\mathbf{x}_{i1}^{o} - ... - \mathbf{x}_{i,T/2}^{o} + \mathbf{x}_{i,T/2+1}^{o} + ... + \mathbf{x}_{iT}^{o}]/T.$$

Then this modified estimators reduce the inconsistency to $O\left(T^{-2}\right)$. That is,

Theorem 2 (Asymptotic Properties of Between and Modified Between Estimators) *If the true data generating process is given by (6) and $\boldsymbol{\beta} \neq \boldsymbol{\gamma}$, Then the probability limits for the modified between group estimators are given by*

$$\text{plim}_{N\to\infty}\hat{\mathbf{b}}_{\text{bw}}^{+} = \boldsymbol{\beta} + O\left(T^{-2}\right).$$

See the Appendix for the detailed proof. Here we provide a heuristic explanation. The adjusted term, $\bar{\mathbf{x}}^{+\prime}\bar{\mathbf{x}}^{+}$, eliminates the contemporaneous variance for the time varying components completely and the covariance terms partially. Hence after the adjustment, the leftover term becomes sufficiently small and can be ignored. When the regressors are not serially correlated, there is no leftover term so that the biases do not exist at all.

It is also important to note that the existence of the bias of the between group estimator causes a somewhat serious problem for statistical inference. The standard t-statistics of the between group estimators do not converge to the standard normal distribution unless $N/T^2 \to 0$ as $N, T \to \infty$. Meanwhile the modified between group estimators require only $N/T^4 \to \infty$. We will provide more detailed discussions on this issue in the Monte Carlo section.

Remark 4: Estimation of Covariance Estimators

Get $\hat{\mathbf{b}}_{\text{bw}}^{+}$ and then construct the regression residuals

$$\widehat{\mathbf{u}}^{*} = \bar{\mathbf{y}}^{*} - \bar{\mathbf{x}}^{*}\hat{\mathbf{b}}_{\text{bw}}^{+}, \quad \widehat{\mathbf{u}}^{+} = \bar{\mathbf{y}}^{+} - \bar{\mathbf{x}}^{+}\hat{\mathbf{b}}_{\text{bw}}^{+}$$

Define

$$\eta_{iT} = \widehat{\bar{u}}_i^{*} - \widehat{\bar{u}}_i^{+}, \quad \mathbf{X}_{iT} = \bar{\mathbf{x}}_i^{*\prime}\bar{\mathbf{x}}_i^{*} - \bar{\mathbf{x}}_i^{+\prime}\bar{\mathbf{x}}_i^{+}.$$

Then the covariance matrix can be written as

$$\left(\sum_i^N \mathbf{X}_{iT}\right)^{-1}\left(\sum_i^N \hat{\eta}_{iT}^2 \mathbf{X}_{iT}\right)^{-1}\left(\sum_i^N \mathbf{X}_{iT}\right)^{-1}.$$

When there is no heteroskedasticity, the covariance matrix further becomes

$$\frac{1}{N}\sum_i^N \hat{\eta}_{iT}^2 \left(\sum_i^N \mathbf{X}_{iT}\right)^{-1},$$

and it converges in probability to the true covariance matrix.

Remark 5: Testing Homogeneity of $\mathcal{H}_0 : \boldsymbol{\gamma} = \boldsymbol{\beta}$

There are many methods available for testing the difference between fixed effects and random effects estimators. For example, Hausman (1978), Hausman and Taylor (1981), Arellano (1993), and Ahn and Low (1996) proposed various tests. All these tests can be used for testing the null hypothesis of $H_0 : \gamma = \beta$. Among them, the Hausman test is often used. Similar to Arellano (1993) and Ahn and Low (1996), we may run the following augmented pooled OLS given by

$$y_{it} = v + \mathbf{x}_{it}\boldsymbol{\gamma} + \bar{\mathbf{x}}_i\boldsymbol{\varpi} + \epsilon_{it}, \quad \text{for } \epsilon_{it} = (v_i - v) + m_{it} \tag{15}$$

where $\boldsymbol{\varpi} = \boldsymbol{\beta} - \boldsymbol{\gamma}$. As $N, T \to \infty$, the POLS estimators in (15) converge to their true values. Hence testing $\boldsymbol{\varpi} = \mathbf{0}$ becomes asymptotically equivalent to the Hausman test. Of course, the former should use the panel robust covariance matrix due to the serial correlation among ϵ_{it}. We examined the finite sample performance between the two tests, and found that the two tests are almost identical in terms of power and size. We do not report the results here.

2.3 Econometric Application

Many empirical panel regressions have the following form

$$y_{it} = \alpha + \mathbf{z}_i \mathbf{b}_1 + \mathbf{x}_{it} \mathbf{b}_2 + u_{it}, \quad u_{it} = \alpha_i - \alpha + e_{it}. \tag{16}$$

where parameters of interest are both $\mathbf{b}_1$ and $\mathbf{b}_2$. Usually $\mathbf{z}_i$ variables are time invariant individual characteristics such as gender, education attainment and race. Because of the time invariant variables in $\mathbf{z}_i$, the within group transformation for the fixed effects regression completely eliminates the $\mathbf{z}_i$ variables. Hence the within group transformation is not an option.

Let

$$\mu_{yi} = \alpha + \mathbf{z}_i \boldsymbol{\beta}_1 + \boldsymbol{\mu}_{xi} \boldsymbol{\beta}_2 + v_i,$$

and

$$y_{it}^o = \alpha_i + \mathbf{x}_{it}^o \boldsymbol{\gamma} + \varepsilon_{it},$$

where we assume that $\mathrm{E}(\mathbf{z}_i \cdot v_j) = \mathrm{E}(\boldsymbol{\mu}_{xi} \cdot v_j) = 0$, but $\mathrm{E}(\mathbf{z}_i \cdot \boldsymbol{\mu}_{xi}) = \boldsymbol{\sigma}_{z\mu} \neq 0$ and $\boldsymbol{\beta}_2 \neq \boldsymbol{\gamma}$ where '$\cdot$' stands for the Hadamard product. Then the pooled estimators of $\boldsymbol{\beta}$ become inconsistent. To see this, we expand y_{it} as

$$y_{it} = \mu_{yi} + y_{it}^o = \alpha + \mathbf{z}_i \boldsymbol{\beta}_1 + (\boldsymbol{\mu}_{xi} + \mathbf{x}_{it}^o) \boldsymbol{\gamma} + v_i + \alpha_i + \boldsymbol{\mu}_{xi} (\boldsymbol{\gamma} - \boldsymbol{\beta}_2) + \varepsilon_{it}. \tag{17}$$

Hence, u_{it} in (16) becomes

$$u_{it} = v_i + \alpha_i + \boldsymbol{\mu}_{xi} (\boldsymbol{\gamma} - \boldsymbol{\beta}_2) + \varepsilon_{it}, \tag{18}$$

and

$$\mathrm{E}(\mathbf{z}_i \cdot u_{it}) = \mathrm{E}[\mathbf{z}_i \cdot \boldsymbol{\mu}_{xi} (\boldsymbol{\gamma} - \boldsymbol{\beta}_2)] = \boldsymbol{\sigma}_{z\mu} \cdot (\boldsymbol{\gamma} - \boldsymbol{\beta}_2) \neq 0, \tag{19}$$

$$\mathrm{E}(x_{it} \cdot u_{it}) = \mathrm{E}[\boldsymbol{\mu}_{xi} \cdot \boldsymbol{\mu}_{xi} (\boldsymbol{\gamma} - \boldsymbol{\beta}_2)] \neq 0. \tag{20}$$

Hence both estimators for $\mathbf{b}_1$ and $\mathbf{b}_2$ become inconsistent.

Hausman and Taylor (1981) considered exactly this case. Their critical assumption is that some variables in $\mathbf{z}_i$ and $\mathbf{x}_{it}$ are not correlated with the regression errors. To find such variables, Hausman and Taylor suggested three tests, which are basically to identify whether or not some of $\boldsymbol{\gamma}$ are equal to $\boldsymbol{\beta}_2$. If there is no such a variable, then there is no appropriate instrumental variable.

However, we can easily overcome this issue by running the following two regressions. The first regression is the fixed effects regression with only time varying variables, that is $\mathbf{x}_{it}$.

$$y_{it} = \alpha_i + \mathbf{x}_{it}\mathbf{b}_2 + \text{ error.}$$

These within group estimators are consistently estimating the dynamic relationship γ. That is, $\hat{b}_{2\text{wg}} \rightarrow^p \gamma$.

The second regression is the modified between group regression. For time invariant variables, regressors do not need to be modified. But for time varying variables, the regressors should be modified to reduce the small T bias. For example, the modified between group estimators are given by

$$\begin{bmatrix} \hat{\mathbf{b}}_{1,\text{bw}} \\ \hat{\mathbf{b}}_{2,\text{bw}}^{+} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{z}}'\tilde{\mathbf{z}} & \tilde{\mathbf{z}}'\bar{\mathbf{x}}^{*} \\ \tilde{\mathbf{z}}'\bar{\mathbf{x}}^{*} & \bar{\mathbf{x}}^{*\prime}\bar{\mathbf{x}}^{*} - \bar{\mathbf{x}}^{+\prime}\bar{\mathbf{x}}^{+} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{z}}'\bar{\mathbf{y}}^{*} \\ \bar{\mathbf{x}}^{*\prime}\bar{\mathbf{y}}^{*} - \bar{\mathbf{x}}^{+\prime}\bar{\mathbf{y}}^{+} \end{bmatrix}. \tag{21}$$

Then it is straightforward to show that

$$E\begin{bmatrix} \hat{\mathbf{b}}_{1,\text{bw}} - \boldsymbol{\beta} \\ \hat{\mathbf{b}}_{2,\text{bw}}^{+} - \boldsymbol{\beta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ O\left(T^{-2}\right) \end{bmatrix}.$$

The covariance matrix is given by

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\eta}_{iT}^{2}\begin{bmatrix} \tilde{\mathbf{z}}'\tilde{\mathbf{z}} & \tilde{\mathbf{z}}'\bar{\mathbf{x}}^{*} \\ \tilde{\mathbf{z}}'\bar{\mathbf{x}}^{*} & \bar{\mathbf{x}}^{*\prime}\bar{\mathbf{x}}^{*} - \bar{\mathbf{x}}^{+\prime}\bar{\mathbf{x}}^{+} \end{bmatrix}^{-1}.$$

In this section, we discussed the economic meaning of the static relationship and how to estimate it consistently and efficiently. All these discussions, however, are valid only when the panel data are free from cross sectional dependence. It is well known that the cross sectional regressions are not well-defined when the data are cross sectionally dependent (Andrews, 2005). In practice, many empirical studies have used simple common time fixed effects to handle cross sectional dependence. That is, the inclusion of the time fixed effects in (16) becomes

$$y_{it} = \alpha + \mathbf{z}_i\mathbf{b}_1 + \mathbf{x}_{it}\mathbf{b}_2 + \theta_t + u_{it}. \tag{22}$$

The next section discusses under what conditions the simple time fixed effects in (22) can effectively handle cross sectional dependence.

3 Static and Dynamic Relationship under Cross Sectional Dependence

In the previous section, the dynamic relationship between common components is not considered. This section studies how the existence of the common factors influences the cross sectional and panel regressions. In the recent common factor literature, cross sectional dependence has been modeled by a few common factors. See Bai and Ng (2008) for an excellent survey of the common factor literature. In the static common factor representation, the cross sectionally dependent data can be decomposed into two components: Common and idiosyncratic components. By adding the time invariant mean, y_{it} can be rewritten as

$$y_{it} = \mu_{yi} + h_{it}^{y} + y_{it}^{o}, \quad \text{for } h_{it}^{y} = \boldsymbol{\lambda}_{yi}' \mathbf{F}_{yt}, \tag{23}$$

and similarly

$$\mathbf{x}_{it} = \boldsymbol{\mu}_{xi} + \mathbf{h}_{it}^{x} + \mathbf{x}_{it}^{o}, \text{ for } \mathbf{h}_{it}^{x} = \boldsymbol{\lambda}_{\mathbf{x}i}' \mathbf{F}_{\mathbf{x}t}.$$

where $\boldsymbol{\lambda}_{yi}$ is a $(r \times 1)$ vector of the factor loadings, $\mathbf{F}_{yt}$ is a $(r \times 1)$ vector of the common factors, and y_{it}^{o} is the idiosyncratic component.

Naturally, we introduce three relationships between y_{it} and $\mathbf{x}_{it}$ as

$$\begin{array}{ll} \text{Static Relationship:} & \mu_{yi} = \boldsymbol{\mu}_{xi} \boldsymbol{\beta} + v_i, \\ \text{Idio-Dynamic Relationship:} & y_{it}^{o} = \mathbf{x}_{it}^{o} \boldsymbol{\gamma} + m_{it}, \\ \text{Common-Dynamic Relationship:} & h_{it}^{y} = \mathbf{h}_{it}^{x} \boldsymbol{\phi} + e_{it}. \end{array} \tag{24}$$

where h_{it}^{y} and h_{it}^{x} have both zero means. The common-dynamic relationship can be interpreted as the macro relationship since this relationship can be revealed with cross sectionally aggregated data. Meanwhile the idio-dynamic relationship is the micro relationship which is often of interest in treatment literatures.

We can write up the linear relationships between y_{it} and $\mathbf{x}_{it}$ as follows:

$$y_{it} = v_i + \mathbf{x}_{it} \boldsymbol{\beta} + k_{it}, \quad k_{it} = \mathbf{h}_{it}^{x} (\boldsymbol{\phi} - \boldsymbol{\beta}) + \mathbf{x}_{it}^{o} (\boldsymbol{\gamma} - \boldsymbol{\beta}) + m_{it} + e_{it}, \tag{25}$$

$$= a_i + \mathbf{x}_{it} \boldsymbol{\gamma} + w_{it}, \quad a_i = \boldsymbol{\mu}_{xi} (\boldsymbol{\beta} - \boldsymbol{\gamma}) + v_i, \; w_{it} = \mathbf{h}_{it}^{x} (\boldsymbol{\phi} - \boldsymbol{\gamma}) + m_{it} + e_{it}, \tag{26}$$

$$= a_i + \mathbf{x}_{it} \boldsymbol{\phi} + \varepsilon_{it}, \; \varepsilon_{it} = \mathbf{x}_{it}^{o} (\boldsymbol{\gamma} - \boldsymbol{\phi}) + m_{it} + e_{it}. \tag{27}$$

In the cross sectional regression for each t, the parameters of interest become $\boldsymbol{\beta}$. Hence the eq (25) becomes the true DGP for the cross sectional regression. Obviously, the regressors are correlated with the errors. In the panel regressions with fixed effects, the parameters of interest become $\boldsymbol{\gamma}$, which is the idio-dynamic relationship. However in eq. (26), the regressors are correlated with the common factors. Pesaran (2006), Bai (2009), and Greenaway-McGrevy, Han and Sul (2012, GHS) consider exactly the same problem of (26), and propose very similar estimators. The probability limits of the cross sectional, POLS and WG estimators are given by

$$
\begin{aligned}
\text{plim}_{N\to\infty}\hat{\mathbf{b}}_{\text{cross},t} &= \mathbf{W}_t\boldsymbol{\beta}+(\mathbf{I}-\mathbf{W}_t)\left[\mathbf{C}_t\boldsymbol{\gamma}+(\mathbf{I}-\mathbf{C}_t)\boldsymbol{\phi}\right] && (28)\\
\text{plim}_{N,T\to\infty}\hat{\mathbf{b}}_{\text{pols}} &= \mathbf{W}\boldsymbol{\beta}+(\mathbf{I}-\mathbf{W})\left[\mathbf{C}\boldsymbol{\gamma}+(\mathbf{I}-\mathbf{C})\boldsymbol{\phi}\right] && (29)\\
\text{plim}_{N,T\to\infty}\hat{\mathbf{b}}_{NT\text{wg}} &= \mathbf{C}\boldsymbol{\gamma}+(\mathbf{I}-\mathbf{C})\boldsymbol{\phi}, && (30)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{W}_t &= \left[\mathbf{Q}_{\mu x,t}+(\mathbf{Q}_{xo,t}+\mathbf{Q}_{ht})\right]^{-1}\mathbf{Q}_{\mu x,t}, \quad & \mathbf{C}_t &= \left[\mathbf{Q}_{xot}+\mathbf{Q}_{ht}\right]^{-1}\mathbf{Q}_{xot},\\
\mathbf{W} &= \left[\mathbf{Q}_{\mu x}+(\mathbf{Q}_{xo}+\mathbf{Q}_{h})\right]^{-1}\mathbf{Q}_{\mu x}, & \mathbf{C} &= \left[\mathbf{Q}_{xo}+\mathbf{Q}_{h}\right]^{-1}\mathbf{Q}_{xo},
\end{aligned}
$$

and

$$
\mathbf{Q}_{hi}=\lim{}_T E\left(\frac{\mathbf{h}_i'\mathbf{h}_i^o}{T}\right)\quad \mathbf{Q}_{ht}=\lim{}_N E\left(\frac{\mathbf{h}_t'\mathbf{h}_t}{N}\right)\quad \mathbf{Q}_{ho}=\lim{}_{NT} E\left(\frac{\mathbf{h}'\mathbf{h}}{TN}\right).
$$

However the general mathematical expressions like (28) through (30) do not provide insightful explanations. Depending on the type of the data, the variances, $\mathbf{W}$ and $\mathbf{C}$, can change. To see this, consider the cross sectional aggregated time series regression as an example. For notational convenience, we will consider the case of $r = 1$. The cross sectional aggregated data include only the common factor relationship. Most of all macro time series data are cross sectional aggregated data. Taking the cross sectional averages in (23) yields

$$
\bar{y}_t = \mu_y + \lambda_y F_{yt} + O_p\left(N^{-1/2}\right), \tag{31}
$$

where $\mu_y = \text{plim}_{N\to\infty}N^{-1}\sum\mu_{yi}$, $\lambda_y = \text{plim}_{N\to\infty}N^{-1}\sum\lambda_{yi}$. So for a large N, the third term becomes very tiny and can be ignored. We define $\bar{\mathbf{x}}_t$ similarly. Then the common-dynamic relationship, ϕ in (4), can be estimated by running $\bar{y}_t$ on $\bar{\mathbf{x}}_t$ with a constant. If there is no common factor, then the cross sectional average becomes just a constant.

In this section, two important points will be addressed based on the common factor representation. First, the role of the common factors on the cross sectional regressions is depending on the level of aggregation. For the very fine disaggregated data, such as PSID or NLSY, the role of the common factors is very limited so that they can be ignored. However, for state-level aggregated data, the common factors significantly influence the cross sectional and panel regressions. Second, if the common factors are integrated or nonstationary, then the static relationship may not be well-defined since the steady state value cannot be defined. However, depending on the level of aggregation, the influence of the nonstationary common factor on the cross sectional regression becomes different. These important issues are considered separately.

3.1 Level of Aggregation and Cross Sectional Dependence

The degree of the cross sectional dependence may be dependent on the level of aggregation. To show this relation, we take the following assumption.

Assumption B (The Number of Common Factors) *The number of common factors, r, is invariant over the level of the aggregation.*

The constant number of common factors is a typical assumption in the common factor literature. For more intuitive explanations and notational convienience, we set $\lambda_i = \lambda$ for the time being. We rewrite (23) as

$$y_{it} = \mu_{yi} + \lambda_y F_{yt} + y_{it}^o, \text{ for } i = 1, ..., N. \tag{32}$$

Here N stands for the total number of the entire individuals, hence y_{it} is not aggregated data. Assume that the relative cross sectional variance ratio between μ_{yi} and y_{it}^o is well-defined: Not too small nor too large either. That is,

$$V\left(\mu_{yi}\right)/V\left(y_{it}^o\right) = \zeta_t, \text{ with } 0 < \zeta_t < \infty \text{ for all } t. \tag{33}$$

where $V\left(\mu_{yi}\right) = \text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \left(\mu_{yi} - N^{-1}\sum_{i=1}^N \mu_{yi}\right)^2$ and $V\left(y_{it}^o\right)$ is similarly defined.

Next, consider the cross sectional aggregation based on J subgroups. Typically subgroups are formed by regional specification. Here we use states as an example. For each jth state, we assume that there are n_j individuals. Then the state level aggregated data for $\{y\}$ can be written as

$$y_{jt} = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{it} = \mu_{yj} + \lambda_y F_{yt} + y_{jt}^o. \tag{34}$$

Let's assume that the variance of the common factor relative to that of y_{jt}^o is finite. That is,

$$V\left(\lambda_y F_{yt}\right)/V\left(y_{jt}^o\right) = \xi_t, \quad \text{with } 0 < \xi_t < \infty \text{ for all } t. \tag{35}$$

Now, we are ready to provide the following theorem.

Theorem 3: (Level of Aggregation and Degree of Cross Sectional Dependence) *Under Assumption B, if (32), (34) and (35) are all true, then as $n_j \to \infty$,*

$$V\left(\lambda_y F_{yt}\right)/V\left(y_{it}^o\right) = \left[V\left(\lambda_y F_{yt}\right)/V\left(y_{jt}^o\right)\right]\left[V\left(y_{jt}^o\right)/V\left(y_{it}^o\right)\right] \to 0. \tag{36}$$

The proof of Theorem 3 is straightforward. Since

$$V\left(y_{jt}^o\right)/V\left(y_{it}^o\right) = O\left(n_j^{-1/2}\right) \to 0 \text{ as } n_j \to \infty,$$

the relative variance ratio between the common factors and the idiosyncratic components becomes zero as the level of aggregation is getting lower. Note that n_j becomes larger as the level of aggregation becomes higher. Hence this implies that compared to the variance of μ_{yi} or y_{it}^o, the common component $\lambda_y F_{yt}$ becomes very tiny as the level of disaggregation becomes lower. In other words, the cross sectional dependence in the micro survey data can be ignored asymptotically if Assumption B holds.

If the factor loadings in (23) share the same underlying population, then Assumption B implies that as the level of aggregation increases, the factor loadings become identical. To be specific, consider $r = 1$ but let λ_i is a random element defined on the same probability space $(\Omega_\lambda, \mathcal{B}_\lambda, \mathcal{P}_\lambda)$. Then it is easy to show that as $n_j \to \infty$,

$$\lambda_j - \lambda \to^p 0, \tag{37}$$

where $\lambda = \mathrm{E}\lambda_i$. A similar argument can be applied to the time invariant mean, μ_i. This result implies that the time fixed effects would be sufficient to control for the cross sectional dependence when the level of aggregation is high enough if Assumption B holds. Here is an empirical example.

Example 2: (State Level Aggregation)

Parker and Sul (2015) provide a new criterion for identifying whether or not a particular time series variable is a common factor in the conventional approximate factor model. Their method, however, can be used for detecting whether or not the panel data has the same factor loadings. Assume that y_{it} have single common factor. That is,

$$y_{jt} = \mu_{yj} + \lambda_{yj} F_{yt} + y_{jt}^o.$$

If $\lambda_{yj} = \lambda_y$ for all j,[1] then by pulling out the cross sectional average from y_{jt}, the common factor can be successfully eliminated. Hence the deviation from its cross sectional average, $\tilde{y}_{jt}$, does not have any significant common factor.

$$\tilde{y}_{jt} = y_{jt} - \frac{1}{J}\sum_{j=1}^{J} y_{jt} = \left(\mu_{yj} - \frac{1}{J}\sum_{j=1}^{J}\mu_{yj}\right) + \left(y_{jt}^o - \frac{1}{J}\sum_{j=1}^{J} y_{jt}^o\right).$$

As Parker and Sul (2015) suggest, the homogeneity restriction can be examined by testing for the number of common factors underlying $\tilde{y}_{jt}$. Table 1 shows that the state panel of crime data share the same common factor and their factor loadings are identical.[2] Bai

[1] The more general condition is given by $\lambda_{yj} = \lambda_y + O_p\left(T^{-1/2}\right)$.

[2] The FBI Uniform Crime Reports contain state-level, annual burglaries per 100,000 persons from 1965 to 2010 for the 50 United States. Natural logs are taken before first-differencing; then the series are demeaned and standardized.

and Ng (2002)'s IC_2 criteria is used for the estimation of the common factor number.

Table 1: Identical Common Factor in State Level Aggregation

Data	Estimation of the Factor Numbers			
	Assault	Burglary	Property	Police
with Δy_{jt}	1	1	1	1
with $\Delta y_{jt} - \frac{1}{J}\sum_{j=1}^{J} \Delta y_{jt}$	0	0	0	0

These empirical examples show that Assumption B is reasonable.

As Example 2 showed, cross sectional dependence in many state aggregated panel data can be eliminated successfully by taking off the cross sectional average or equivalently, including time fixed effects or year dummies in the regressions. However, when the parameters of interest are $\boldsymbol{\phi}$ or the common-dynamic relationship, the inclusion of the time and individual fixed effects to the regressions leads to the estimation of $\boldsymbol{\gamma}$ which is the idio-dynamic relationship. More formally, we have

Remark 6: (Effects of the Inclusion of Individual and Time Fixed Effects)

Consider the following panel regressions with and without fixed effects.

POLS	$y_{it} = \alpha + \mathbf{x}_{it}\mathbf{b}_1 + e_{it},$
State fixed effects	$y_{it} = \alpha_i + \mathbf{x}_{it}\mathbf{b}_2 + \varepsilon_{it},$
State and Year fixed effects	$y_{it} = \alpha_i + \theta_t + \mathbf{x}_{it}\mathbf{b}_3 + \epsilon_{it},$
Aggregated Time Series	$\bar{y}_t = \alpha + \bar{\mathbf{x}}_t\mathbf{b}_4 + \bar{e}_t$

where $e_{it} = (\alpha_i - \alpha) + \theta_t + \epsilon_{it}$ and $\varepsilon_{it} = \theta_t + \epsilon_{it}$. Assume that the regressand and regressors have a single common factor and that their factor loadings are identical across states. Further assume that the regression errors are stationary. Then the probability limit of each estimator is given by

$$\begin{aligned} \text{plim}_{N,T\to\infty}\hat{\mathbf{b}}_1 &= \mathbf{W}\boldsymbol{\beta} + (\mathbf{I}-\mathbf{W})\left[\mathbf{C}\boldsymbol{\gamma} + (\mathbf{I}-\mathbf{C})\,\boldsymbol{\phi}\right], \\ \text{plim}_{N,T\to\infty}\hat{\mathbf{b}}_2 &= \mathbf{C}\boldsymbol{\gamma} + (\mathbf{I}-\mathbf{C})\,\boldsymbol{\phi}, \\ \text{plim}_{N,T\to\infty}\hat{\mathbf{b}}_3 &= \boldsymbol{\gamma}, \\ \text{plim}_{N,T\to\infty}\hat{\mathbf{b}}_4 &= \boldsymbol{\phi}, \end{aligned} \tag{38}$$

Note that when the heterogeneity of individual fixed effects is small (that is, $\mu_{yi} = \mu_y + o_p(1)$ and $\boldsymbol{\mu}_{xi} = \boldsymbol{\mu}_x + o_p(1)$), the static weight, $\mathbf{W}$, becomes relatively small so that $\hat{\mathbf{b}}_1 \simeq \hat{\mathbf{b}}_2$.

Here we provide another interesting empirical example.

Example 3: (Determination of Crime)

In the empirical studies on the economics on crime, the most important task is the explanation of the sudden decline of the national crime after 1990s.[3] Figure 3 shows the national average of the log per capita property crimes and sworn police officers across 50 states from 1970 to 2005.[4] Evidently, property crime has declined since 1992 meanwhile the number of the sworn officers gradually increased. Nonetheless, all empirical studies in this literature have run the following panel regressions.

$$\Delta y_{it} = \alpha_i + \theta_t + \mathbf{X}_{it-1}\mathbf{b} + u_{it},$$

where y_{it} is the number of crimes in the ith state (or city) at time t. The determinant variables include many social and economic variables.

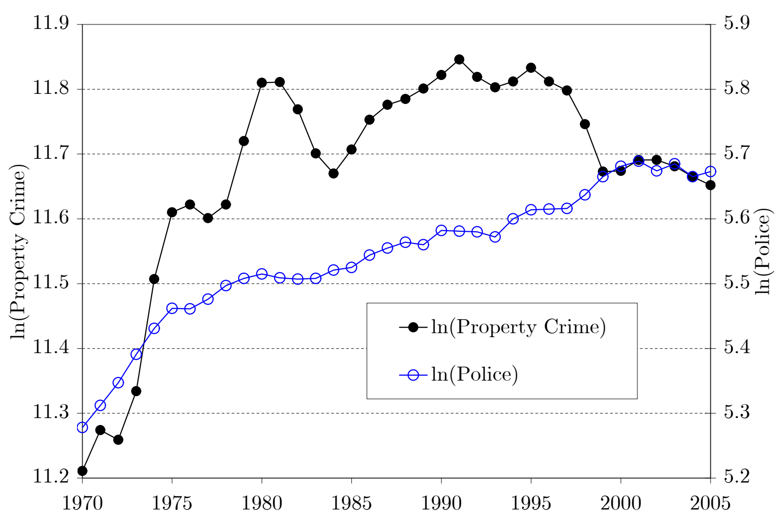

Figure 3: Trends in Crime and Police

The parameters of interest become the common-dynamic relationships (ϕ) rather than the idio-dynamic relationships (γ). However by excluding the common components, the most important information is excluded from the empirical analysis. To see this, we take the first difference of the log police and crime and run the following dynamic panel regressions. Note that we do not correct the Nickell bias since our purpose is to demonstrate the role of the time fixed effects (excluding the common components). We

[3] See Levitt (2004) for the excellent survey on this topic.

[4] The data are obtained from uniform crime report.

also do not include other control variables. Table 2 shows the empirical results as an example. When the time fixed effects (year dummies) are not included in the regressions, the point estimates of lagged crime and lagged police are significantly different from zero. Interestingly, the POLS estimates are similar to the WG estimates without including time fixed effects since as we mentioned in Remark 6 and Example 2, the heterogeneity of individual fixed effects is very limited in the state level aggregated data.

Table 2: Relationships between Police and Crime

Independent	Dependent Variables				
		Δy_{it}		$\Delta \bar{y}_t$	$\Delta \bar{y}_i$
Δx_{it-1}	0.292*	0.269*	-0.001		
Δy_{it-1}	0.244*	0.246*	-0.009		
$\Delta \bar{x}_{t-1}$				0.374*	
$\Delta \bar{y}_{t-1}$				1.295*	
$\Delta \bar{x}_i$					0.715*
fixed effects	No	Yes	Yes	n.a	n.a
time effects	No	No	Yes	n.a	n.a

As Figure 3 showed, the relationship between the two are in general positive. However, as the year effects are included, the relationship changed from the significantly positive to insignificantly negative. Hence the idio-dynamic relationship is close to zero. In other words, a pure idiosyncratic increase in sworn officers does not result in a decrease in crime. Next, we take the cross sectional averages of the crime and police, and run the time series regression. Since the two series are highly correlated as shown in Figure 3, the relationship becomes strongly positive. Lastly, we run the cross sectional regression with the time series averages of Δy_{it} and Δx_{it}, and find that the static relationship is positive. In other words, the states with more sworn officers have more property crimes. Again the results in Table 2 should not be considered seriously since other control variables are not included. More importantly, the two variables might be cointegrated with each other. A more careful empirical investigation is required.

It is important to note that Assumption B does not always hold. Especially among international aggregated data, Assumption B may not hold.. More formally, we have the following remark as a counter-example of Assumption B.

Remark 7 (Violation of Assumption B: Hierarchical Factor Structure)

Even though Assumption B looks reasonable, there are some cases where Assumption B does not hold. For example, consider real GDP per capita across countries. The level of the aggregation in real GDP per capita data is the highest possible. But as Phillips and Sul (2007, 2009) show, real GDP per capita do not share the same common factors. Of course, real GDP per capita for each country may be defined on the heterogeneous probability spaces. Here we consider alternative factor structure: The hierarchical model. Assume that a random variable $y_{ij,t}$ has the following hierarchical factor representation.

$$y_{ij,t} = \mu_{yij} + \delta_{Wij} F_{wt} + \delta_{Fij} F_{jt} + y_{ij,t}^{o} \tag{39}$$

where $j = 1, ..., J$, $i = 1, ..., n$ and j, i stand for country and individual, respectively. The common factor, F_{wt}, is the global factor meanwhile F_{jt} is a country-specific common factor. The total number of common factors becomes $1 + J$ for the individual level. Here we assume that in each country, the total number of individuals is the same. Next, consider the country level aggregation.

$$y_{jt} = \mu_j + \delta_{Wj} F_{wt} + \delta_{Fj} F_{jt} + o_p(1)$$

where y_{jt}, μ_j, δ_{Wj} and δ_{Fj} are the cross sectional averages of y_{ijt}, μ_{yij}, δ_{Wij} and δ_{Fij}, respectively. The last $o_p(1)$ term is the cross sectional average of $y_{ij,t}^{o}$ over i, so that as $n \to \infty$, this term converges to zero in probability. Now $\delta_{Fj} F_{jt}$ becomes the idiosyncratic term for the country level aggregation, so that the total number of common factors becomes one.

Remark 7 shows that under the hierarchical factor structure, the level of aggregation is not related to the degree of the cross sectional dependence. It is not easy to test whether or not a factor structure satisfies Assumption B simply because all available micro individual level data does not have enough time series observations. Also, if Assumption B holds, then the common factors are not visible at all in the micro disaggregated data. Nonetheless, in our limited experience, Assumption B seems to hold within a nation but not across nations.

Example 4: Gravity Model

Serlenga and Shin (2007) consider the following gravity model

$$y_{it} = \alpha_i + \mathbf{x}_{it}\boldsymbol{\gamma} + \mathbf{z}_i\mathbf{b} + \delta_i' F_{wt} + \varepsilon_{it}$$

where y_{it} is the total trade volume for the ith country to the numeraire country, $\mathbf{x}_{it}$ are a vector of control variables including GDP, real exchange rate etc, and one of $\mathbf{z}_i$ variables is the distance between the country i and the numeraire. They consider two models. In the first model, they impose the restriction of $\lambda_i = 1$ so that they just run a simple time and individual dummy variables regression. In the second model, they use the cross sectional averages of y_{it} and $\mathbf{x}_{it}$ as a proxy of W_t, and run CCE regressions. By using IV estimation, they find that

$$\begin{array}{ll} \text{Year Fixed Effects} & \text{Total Trade Volume} = \underset{(0.48)}{0.38} \times \text{Distance} + F_{wt} + \alpha_i + ... + \text{error}, \\ \text{CCE} & \text{Total Trade Volume} = \underset{(0.21)}{\text{-0.44}} \times \text{Distance} + \lambda_i' F_{wt} + \alpha_i + ... + \text{error}. \end{array}$$

Evidently, using simple time dummies results in wrong estimation results. This implies that even after eliminating cross sectional dependence from total trade volumes across countries, the distance between two countries becomes a determinant of trade.

When the degree of the cross sectional dependence is not small enough to be ignored, the cross sectional estimators are in general inconsistent as Andrews (2005) points out. What Andrews considered was the case where the dynamic relationship among the common components is different from the static relationship. That is the case of $\boldsymbol{\phi} \neq \boldsymbol{\beta}$. However, regardless of the level of the cross sectional dependence, as long as $\boldsymbol{\beta} \neq \boldsymbol{\gamma}$ or $\boldsymbol{\beta} \neq \boldsymbol{\phi}$, the cross sectional estimators become inconsistent.

When the static relationship is of interest, then the time invariant mean can be estimated by using the estimated common factors. The asymptotic properties of this factor augmented approach are well studied by Bai (2009) and Greenaway-McGrevy, Han and Sul (2012). We will show their finite sample properties in Monte Carlo simulation section later.

Next, we consider the case where the common factors are integrated or nonstationary.

3.2 Cross Sectional Regressions with Nonstationary Common Factors

When the time varying components are integrated (or follow random walks), the long run values do not exist since the time series averages are not converging to fixed constants even when the number of time series observations goes to infinity. Since the long run averages cannot be defined, naturally the static relationship does not exist. In this case, the long run relationship can be defined as the cointegrating relationship between time varying components. In fact, it is unknown that what the cross sectional regressions with nonstationary components estimate. In this subsection, this issue is discussed based on the level of aggregation. To be specific, we consider two different cross sectional regressions: Micro survey (individual level) data and city, state or national level aggregated data. As we discussed before, Assumption B holds for the individual micro survey data but not for international data.

Micro Survey Data As we discussed before, the household level data may not be cross sectionally dependent. Moreover, it is not realistic to say that the household income is nonstationary simply because the life span of an individual is finite. As a person's age approaches to the retirement age, it is hard to say that his (her) income will follow a random walk process. Meanwhile city or state level aggregate income data are distinctly nonstationary. City and state level aggregate data are simple cross sectional averages across individuals.[5] Assumption B reconciles these two facts. Small nonstationary common factors in the individual level become more distinct as the level of aggregation becomes higher. Rewrite (32) by allowing heterogeneous factor loadings as

$$y_{it} = \mu_{yi} + \lambda_{yi} F_{yt} + y_{it}^o,$$

where for simplicity we assume just single factor structure, but the factor, F_{yt}, is nonstationary or follows a random walk. Here we treat the common factor F_{yt} as an exogenous macro factor, the influence of which on each individual is very limited. In fact, new young individuals enter the set of $\{y\}$ and in the same year, some individuals leave the set. The common factor F_{yt} may not be influenced much by these new entries and exits for every t.[6] Then modelling the stochastic process of y_{it} in the individual level – heterogeneous economic agents – should be focused on μ_{yi} and y_{it}^o. Meanwhile the average individuals or aggregated data become purely dependent on the common factor F_{yt}. Hence modelling a representative economic agent's behavior should be very different from modelling heterogeneous economic agents' behaviors. We will discuss this issue later in Example 6 again. Nonetheless, there is no direct empirical evidence whether or not the idiosyncratic components are stationary simply because of the lack of time dimension and data. However indirectly we can conjecture this by using state level aggregation data. For example, Evans and Karras (1996) examine carefully whether or not per capita real state incomes share the same nonstationary common factor. After controlling the common time fixed effects, Evans and Karras (1996) show that the idiosyncratic components in the state level are stationary. Also Phillips and Sul (2009) show that the per capita real incomes are relatively converging across states. Hence it is not reasonable to assume that the micro survey data contains a very tiny amount of a nonstationary common factor.

[5] It is important to address that the firm level data should be considered as the aggregate data as long as a firm consists of many individuals.

[6] Of course, a mass hysteria death toll by a disease – for example, a pest outbreak in Europe in the 14 Century – can influence on the common factor directly. We exclude such a case here.

Cross Sectionally Aggregate Data Once individual data are cross sectionally aggregated, the fraction of the common factors is no longer ignorable. As Andrews (2005) points out, for a given t, the probability limit of the cross sectional estimators can be either a constant or random even when the number of cross sectional units becomes infinity. However if the common factors are nonstationary, then as t goes to infinity, the fraction of the common components becomes larger since $\mathrm{E}\left(F_{yt}^2\right)$ goes to infinity. Let J be the number of cross sectional units of the cross sectionally aggregated data y_{jt}. Then the cross sectional estimators have the following properties.

$$\lim_{t\to\infty} \mathrm{plim}_{n\to\infty} \hat{\mathbf{b}}_{\text{cross},t} = \begin{cases} \boldsymbol{\phi} & \text{if } y_{jt}^o \sim I(0) \\ \bar{\mathbf{C}}\gamma + \left(\mathbf{I} - \bar{\mathbf{C}}\right)\boldsymbol{\phi} & \text{if } y_{jt}^o \sim I(1) \end{cases}. \tag{40}$$

where $\bar{\mathbf{C}} = \lim_{t\to\infty} \mathbf{C}_t$. The proof for (40) becomes obvious, so it is omitted. From (28), the cross sectional estimators are dependent on the weight matrices $\mathbf{W}_t$ and $\mathbf{C}_t$. When the idiosyncratic terms are stationary, the cross sectional estimators become heavily dependent on the variance of the common component, $\mathbf{Q}_{ht}$, which becomes dominant as t increases. Naturally as $t \to \infty$, $\mathbf{W}_t \to 0$ and $\mathbf{C}_t \to 0$ as well. However when the idiosyncratic terms are nonstationary, the cross sectional estimators become dependent on the variance ratio between the common and idiosyncratic components. See Example 6 for an interesting empirical illustration. Now let's evaluate the economic meaning of "$t \to \infty$". In contrast to the individual level data, the time dimension with the state or city level aggregated data is well-defined since there is no limited life span for states or cities. Then the economic meaning of "$t \to \infty$" may imply any given t as long as a state or city is not newly born.

Now we are ready to examine the following two famous puzzles.

Example 5 (Easterlin Paradox)

Easterlin (1974) found that with various micro survey data for each major industrialized country, the happiness index is positively correlated with the level of personal income, but with national time series aggregated data, the positive correlation is not found any more. This result has been interpreted as meaning that in the short run, an increase in income leads to more happiness, but in the long run more income does not lead to more happiness. Denote H_{it} and $\ln Y_{it}$ as the ith person's happiness index and disposable income at time t. Suppose that both variables can be modelled as

$$H_{it} = \mu_{Hi} + \lambda_{Hi} F_{Ht} + H_{it}^o, \quad \ln Y_{it} = \mu_{Y,i} + \lambda_{Yi} F_{Yt} + y_{it}^o.$$

Here for simplicity, we assume that the number of common factors is one. The time invariant term for H_{it}, μ_{Hi}, may indicate personality or how much a person is optimistic.

Meanwhile the time invariant term for $\ln Y_{it}$, μ_{Yi}, may include a person's education level, type of jobs and other personal characteristics. Since there is no empirical evidence for the relationship between μ_{Hi} and μ_{Yi}, here we assume the static relationship, β, is equal to zero. As Theorem 3 states, the role of the common factor on the micro survey data is limited as long as Assumption B holds. Then the cross sectional results are mixed in with the static and dynamic relationship among the idiosyncratic components. If $\beta = 0$, then the cross sectional estimates is under-estimate the dynamic relationship between H_{it}^o and y_{it}^o. That is, as personal income increases more, the actual happiness may increase further than that reported. Next, consider cross sectional aggregated data. Let

$$H_t = \frac{1}{N}\sum_{i=1}^{N} H_{it} \ \& \ \ln Y_t = \ln\left(\frac{1}{N}\sum_{i=1}^{N} y_{it}\right).$$

Then it is easy to show that $H_t = \bar{\lambda}_H F_{Ht}$ and $\ln Y_t = \bar{\lambda}_Y F_{Yt}$ where $\bar{\lambda}_H$ and $\bar{\lambda}_Y$ are the cross sectional averages of λ_{Hi} and λ_{Yi}. If F_{Ht} is not correlated with F_{Yi}, there should not be any positive correlation with national time series aggregated data.

Example 6 (Consumption Puzzle)

The famous consumption or Kuznets' puzzle (1946) is based on the following two simple regressions.

$$\ln C_{it} = a_t + b_t \ln Y_{it} + u_{it}, \tag{41}$$

$$\ln \bar{C}_t = a + b \ln \bar{Y}_t + u_t, \tag{42}$$

where C_{it} is the expenditure, Y_{it} is the disposable income for the i-th household at time t, meanwhile $\bar{C}_t$ and $\bar{Y}_t$ stand for the aggregated consumption and disposable income (per capita real consumption and disposable income), respectively. The cross sectional estimates of b_t with micro survey data range from 0.4 to 0.5, and a_t is significantly positive, meanwhile the aggregate time series estimates of b is more or less 0.9 but a is close to zero. The cross sectional household survey results had been interpreted as meaning that the wealthy households save more, but the poor households spend more than what they make. For an intuitive explanation, assume that $\ln Y_{it}$ and $\ln C_{it}$ have single factors of $\ln \bar{Y}_t$ and $\ln \bar{C}_t$, respectively. Then without loss of generality, the individual log disposable income and consumption at year t, $\ln Y_{it}$ and $\ln C_{it}$, can be rewritten as

$$\ln Y_{it} = \mu_{Yi} + \lambda_{Yi} \ln \bar{Y}_t + \ln Y_{it}^o, \quad \ln C_{it} = \mu_{Ci} + \lambda_{Ci} \ln \bar{C}_t + \ln C_{it}^o,$$

where the common factors $\ln \bar{Y}_t$ and $\ln \bar{C}_t$ are integrated but assumed to be cointegrated with each other. The time invariant term μ_{Ci} may include the information about personal characteristics such as time preference and risk aversion. As we discussed before, the variance of the factor loadings may be small enough to be ignored in the micro survey level. In other words, by using the micro survey data it is very hard to capture the behavior of the common factors (or the macro behavior). For an example, Dynan (2000) could not find any habit formation behavior from the PSID data. Under this assumption, we estimate the weight parameter W_t from the log food consumption of PSID data from 1968 and 1972. The total number of households is 2,952 after eliminating missing or zero observations. The estimated value is around 0.8 ($Q_{\mu x} = 0.260$, $Q_{xo} = 0.067$). Meanwhile the fixed effects estimate is around 0.93. From these estimates, we can conjecture the static relationship, β, is around 0.25 which is much lower than the micro survey results. However it is important to address that the cross sectional regressions are not estimating the short run behavior. To show this, we construct a panel data of 70 countries from World development Indicator based on the data availability from 1970 to 2000. We ran 31 cross sectional, pooled OLS and fixed effects regressions. The average of 31 cross sectional, POLS and the fixed effects estimates are almost identical and around 0.915 as we claimed in (40). The underlying reason is rather straightforward: Taking the cross sectional averages yields a consistent estimate of the common factors. Since the consumption and income share the same common stochastic trend, running the cross sectional regressions with cross sectionally aggregated data yields the dynamic relationship among the common components.

Before we end this section, we want to issue a warning on the use of the micro survey or the micro panel data. As we showed in Example 6, the macro behavior is hardly evident in the micro survey or panel data since the role of the common factors in the micro survey or panel data is very limited. In fact, the cross sectionally aggregated food consumption in Example 6 from 1968 and 1972 does not look similar to the national food consumption data. Around 4,000 household data sounds like a huge number, but compared to the national population, the number of PSID observations is very tiny. Therefore empirical evidence based on micro survey or panel data should be carefully interpreted.

Now we turn our focus on the finite sample performance of the suggested BG estimator in the next section.

4 Monte Carlo Studies

We consider two simulation designs: Cross sectional independent and dependent designs. The first design is under cross sectional independence and is based on the following data generating process (DGP):

$$\begin{aligned} \mu_{yi} &= \beta_1 z_i + \beta_2 \mu_{xi} + v_i, \quad z_i = \mu_{xi} + u_i, \quad v_i \sim iidN\left(0, \sigma_v^2\right), \quad u_i \sim iidN\left(0,1\right). \\ y_{it}^o &= \gamma x_{it}^o + m_{it}, \quad x_{it}^o = \rho_i x_{it-1}^o + \varepsilon_{it}, \quad m_{it} \sim iidN\left(0,1\right), \quad \varepsilon_{it} \sim iidN\left(0,1\right), \; \rho_i \in U\left[0, 0.5\right]. \end{aligned} \tag{43}$$

Under this DGP, we run the following pooled regression,

$$y_{it} = \alpha + b_1 z_i + b_2 x_{it} + \epsilon_{it},$$

and estimate the ordinary and modified BG estimators. The ordinary BG estimates are denoted as $\hat{b}_{s,\text{bw}}$ meanwhile the modified ones are denoted as $\hat{b}_{s,\text{bw}}^{+}$ for $s = 1, 2$.

We consider two different values of β under (43): We set $\beta_1 = \beta_2 = 1$ in the first case but $\beta_1 = \beta_2 = 0$ in the second case. In both cases, we set $\gamma = -1$, $T = 5, 10, 20$, and $N = 100, 200, 500, 1000$. All innovations except for v_i are generated from $iidN\left(0,1\right)$. The variance of v_i, therefore, captures the weight on the static relationship. Higher σ_v^2 leads to more weight on β. We set $\sigma_v^2 \in [2, 3]$. The total number of simulations is 2,000 in all cases.

Table 3 shows the performance of the various estimators when $\beta_1 = \beta_2 = 1$, but $\gamma = -1$. As Theorem 1 states, the POLS estimator for b_2 does not estimate either β_2 or γ. As we showed in (17) and (18), the probability limit of $\hat{b}_{1,\text{pols}}$ is dependent on the size of $(\beta_2 - \gamma)$ and the variance σ_v^2. Also the POLS estimator for b_1 does not estimate β_1 either since $\mathrm{E} z_i \mu_{xi} \neq 0$ and $b_2 \not\to^p \beta_2$. Meanwhile the WG estimator always estimates the dynamic relationship, γ, consistently regardless of the value of σ_v^2. The ordinary BG estimator becomes biased with finite T. The biases increase as σ_v^2 increases. With a small T, such as $T = 5$, the biases are not small at all. As T increases, the biases decrease as we showed in (14). However even with a moderately large T such as $T = 20$, the ordinary BG estimators suffer still from somewhat significant biases. The modified BG estimator, however, shows little bias even with a small T. This result confirms that the asymptotic result in Theorem 2 works very well even in the finite sample.

Table 4 shows the variances of the four estimators. Of course, the WG estimator shows the lowest variance regardless of the values of N and T since its convergence rate is $\sqrt{NT}$. However the variances of the POLS estimators are much larger than those of the WG estimator, but are similar to those of the BG estimators. As Theorem 1 states, the POLS estimators do not estimate the dynamic relationship but a value somewhere in between the dynamic and static relationships. Hence the convergence rate of the POLS estimator becomes $\sqrt{N}$ rather than $\sqrt{NT}$. The variances of the modified BG estimators are not too different from the ordinary BG estimators, but they are usually larger than those of the ordinary BG estimators. However the difference goes away very quickly as N increases. From these results, we can say that the modified BG estimator performs very well.

Table 5 summarizes the comparison between ordinary and modified BG estimators in terms of the size distortion, biases and variances when $\beta_1 = \beta_2 = 0$ and $\gamma = -1$. In the last comparison, $\beta_1 = \beta_2 = 1$. Hence in this case, the bias of the ordinary BG estimator becomes less than before, but the size distortion increases as $N \to \infty$ with fixed T. This is because, as we showed before, the asymptotic normality of the ordinary BG estimator requires $N/T^2 \to 0$. Meanwhile, the modified BG estimator exhibits very little size distortion. In terms of the bias, the modified BG estimator exhibits less bias as we showed in Table 1. Interestingly, when the distance between static and dynamic relationships, $|\beta - \gamma|$, gets smaller, the variance of the modified BG estimator is in fact smaller than that of the ordinary BG estimator.

In the next data generating process there is a somewhat strong degree of cross section dependence.

$$y_{it} = \mu_{yi} + \lambda_{yi} F_{xt} + y_{it}^o, \quad x_{it} = \mu_{xi} + \lambda_{xi} F_{xt} + x_{it}^o,$$

$$\mu_{yi} = \beta \mu_{xi} + v_i, \quad h_{y,it} = \phi h_{x,it} + e_{it}, \quad y_{it}^o = \gamma x_{it}^o + m_{it},$$

and

$$\begin{bmatrix} F_{yt} \\ F_{xt} \\ x_{it}^o \end{bmatrix} = \begin{bmatrix} \rho_{1i} & 0 & 0 \\ 0 & \rho_{2i} & 0 \\ 0 & 0 & \rho_{3i} \end{bmatrix} \begin{bmatrix} F_{yt} \\ F_{xt} \\ x_{it}^o \end{bmatrix} + \begin{bmatrix} e_{yt} \\ e_{xt} \\ \varepsilon_{it} \end{bmatrix},$$

where $v_i \sim iidN\left(0, \sigma_v^2\right)$ and other innovations are generated from $iidN\,(0,1)$. We set $\beta = \phi = 0$, and $\gamma = 2$. Here we estimate β by using the ordinary and modified BG estimators after eliminating the estimated common components. The factor number is one and assumed to be known. To estimate the dynamic relationship between common components, we use two estimators: Principal Component (PC) estimator and simple cross sectional averages: The common-dynamic relationship is estimated with the estimated common components. We set $T = 20, 30, 50$ and $N = 50, 100, 200$.

Table 6 shows the results. First, in terms of the size distortion, the finite sample performances for the de-augmented, modified BG estimators are always better than those of the de-augmented ordinary BG estimators, especially with large N. This is because the ordinary BG estimators require $N/T^2 \to 0$ for asymptotic normality but the required condition for the modified BG estimators is only $N/T^4 \to 0$. Also in terms of bias and variance, the modified between group estimators are slightly better than the ordinary between group estimator. Second, the simple cross sectional average (denoted as $\hat{\phi}_1$) performs very poorly. The bias, size distortion, and variances are all larger compared to those of the cross sectional average of the estimated common component (denoted as $\hat{\phi}_2$). Note that both $\hat{\phi}_1$ and $\hat{\phi}_2$ are time series estimates. However, as N increases, the common components are more accurately estimated so that the size distortion of $\hat{\phi}_2$ decreases. Also the variance and bias of $\hat{\phi}_2$ also decrease as N increases.

5 Conclusion

The starting point of this paper was decomposing a random variable into three components: A time invariant component with non-zero mean; a common, time-varying component with zero mean and finite variance; and an idiosyncratic, time varying component with zero mean and finite variance. The relationship among time invariant components is called a 'static' relationship. The relationship among the common components is called a 'common-dynamic' relationship, and represents the 'macro' relationship since most macro data are cross sectional aggregated data. The relationship among the idiosyncratic components is called an 'idio-dynamic' relationship, which is usually the relationship of interest to microeconomists.

Whenever the static relationship is different to the dynamic relationships, the cross sectional regressions cannot identify or estimate the static relationship. We showed that many empirical puzzles arise because the static relationship is different from either the common-dynamic or idio-dynamic relationship. To identify the static relationship, we need panel data information: By using the modified BG estimation, the static relationship can be identified and estimated consistently. In other words, the results from the cross sectional regressions are spurious if the static relationship is different from the dynamic relationship.

When the panel data are cross sectionally dependent, the idio-dynamic relationship can be estimated by augmenting the panel regression with the estimated common factors. The simplest version of the factor augmented regressions is the inclusion of the time and individual fixed effects. However when the parameters of interest are the common-dynamic relationship, the factor augmented regressions should not be used. Instead, the time series regression with cross sectional aggregates or PC estimates of the common factors must be run to identify the dynamic relationship among the common components. In other words, the panel regressions with individual and time fixed effects are valid only when the idio-dynamic relationship becomes of interest. Otherwise, the panel regression becomes mindless.

References

[1] Ahn, S.C., and S. Low. (1996). "A reformulation of the Hausman test for regression models with pooled cross-section-time-series data." *Journal of Econometrics* 71(1), 309–319.

[2] Almond, D., K.Y. Chay, and D.S. Lee. (2005). "The costs of low birth weight." *The Quarterly Journal of Economics* 120(3), 1031–1083.

[3] Alvarez, J., & Arellano, M. (2003). "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators." *Econometrica*, 71(4), 1121-1159.

[4] Andrews, D.W.K. (2005). "Cross-section Regression with Common Shocks." *Econometrica* 73(5), 1551–1585.

[5] Arellano, M. (1993). "On the testing of correlated effects with panel data." *Journal of Econometrics* 59(1), 87–97.

[6] Bai, J. (2009). "Panel data models with interactive fixed effects." *Econometrica* 77(4), 1229-1279.

[7] Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191-221.

[8] Bai, J., & Ng, S. (2008). *Large dimensional factor analysis.* Now Publishers Inc.

[9] Behrman, J. R., & Rosenzweig, M. R. (2002). Does increasing women's schooling raise the schooling of the next generation?. *American Economic Review*, 323-334.

[10] Black, S. E., Devereux, P. J., & Salvanes, K. G. (2007). From the cradle to the labor market? The effect of birth weight on adult outcomes. *The Quarterly Journal of Economics*, 122(1), 409-439.

[11] Dynan, K.E. (2000) Habit Formation in Consumer Preference: Evidence from Panel Data, *American Economic Review*, 90(3), 391-406.

[12] Easterlin, R. A. (1974). *Does economic growth improve the human lot? Some empirical evidence.* Nations and households in economic growth, 89.

[13] Evans, P., & Karras, G. (1996). Convergence revisited. *Journal of Monetary Economics,* 37(2), 249-265.

[14] Forbes, K. J. (2000). A Reassessment of the Relationship between Inequality and Growth. *American Economic Review*, 869-887.

[15] Greenaway-McGrevy, R., Han, C., & Sul, D. (2012). Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics*, 169(1), 48-53.

[16] Kuznets, S. (1946). National Product Since 1869 (assisted by L. Epstein and E. Zenks), New York: National Bureau of Economic Research.

[17] Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 1251-1271.

[18] Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, 1377-1398.

[19] Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4), 967-1012.

[20] Parker J. and D. Sul (2015). "Identification of unknown common factors: Leaders and followers." forthcoming in *Journal of Business, Economics and Statistics*, University of Texas at Dallas.

[21] Phillips, P.C.B. and D. Sul (2014). "Absolute Convergence." *mimeo,* University of Texas at Dallas.

[22] Phillips, P. C., & Sul, D. (2007). Transition modeling and econometric convergence tests. *Econometrica,* 75(6), 1771-1855.

[23] Phillips, P. C., & Sul, D. (2009). Economic transition and growth. *Journal of Applied Econometrics*, 24(7), 1153-1185.

[24] Royer, H. (2009). Separated at girth: US twin estimates of the effects of birth weight. *American Economic Journal: Applied Economics*, 1(1), 49-85.

[25] Serlenga, L. and Y. Shin. (2007) Gravity models of intra-EU trade: application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics* 22(2), 361-381.

[26] Sul, D (2015). Pooling is sometimes harmful. *mimeo*, University of Texas at Dallas.

Table 3: Biases of pols, bw, wg and modified bw estimators

True Relationship: $\beta_1 = \beta_2 = 1, \gamma_1 = -1$

σ_v^2	N	T	$\hat{b}_{1,\text{pols}}$	$\hat{b}_{2,\text{pols}}$	$\hat{b}_{1,\text{bw}}$	$\hat{b}_{2,\text{bw}}$	$\hat{b}^+_{1,\text{bw}}$	$\hat{b}^+_{2,\text{bw}}$	$\hat{b}_{2,\text{wg}}$
3	100	5	1.203	0.594	1.075	0.851	0.995	1.012	-1.000
3	200	5	1.196	0.607	1.068	0.862	0.990	1.018	-1.000
3	500	5	1.196	0.606	1.070	0.859	0.993	1.012	-1.001
3	1000	5	1.196	0.606	1.070	0.858	0.994	1.011	-1.000
3	100	10	1.197	0.603	1.035	0.927	1.000	0.997	-1.000
3	200	10	1.199	0.604	1.040	0.922	1.006	0.989	-0.999
3	500	10	1.197	0.605	1.040	0.919	1.007	0.986	-1.000
3	1000	10	1.197	0.608	1.040	0.922	1.006	0.989	-1.000
3	100	20	1.199	0.597	1.019	0.957	0.999	0.998	-1.000
3	200	20	1.200	0.602	1.022	0.958	1.002	0.997	-1.001
3	500	20	1.197	0.606	1.021	0.958	1.002	0.997	-1.000
3	1000	20	1.197	0.606	1.021	0.957	1.002	0.996	-1.000
2	100	5	1.362	0.276	1.150	0.701	0.981	1.038	-1.000
2	200	5	1.355	0.289	1.144	0.711	0.980	1.037	-1.000
2	500	5	1.355	0.289	1.145	0.708	0.985	1.028	-1.001
2	1000	5	1.355	0.289	1.145	0.707	0.986	1.026	-1.000
2	100	10	1.358	0.282	1.080	0.836	1.005	0.986	-1.000
2	200	10	1.357	0.287	1.085	0.832	1.013	0.976	-0.999
2	500	10	1.356	0.288	1.085	0.829	1.014	0.972	-1.000
2	1000	10	1.356	0.291	1.085	0.832	1.014	0.974	-1.000
2	100	20	1.359	0.278	1.044	0.908	1.000	0.997	-1.000
2	200	20	1.358	0.284	1.047	0.908	1.004	0.995	-1.001
2	500	20	1.355	0.289	1.046	0.908	1.004	0.993	-1.000
2	1000	20	1.355	0.290	1.046	0.907	1.004	0.992	-1.000

Table 4: Variances$\times 10^2$ of POLS, BG, WG and modified BG estimators
True Relationship: $\beta_1 = \beta_2 = 1, \gamma_1 = -1,$

σ_v^2	N	T	$\hat{b}_{1,\text{pols}}$	$\hat{b}_{2,\text{pols}}$	$\hat{b}_{1,\text{bw}}$	$\hat{b}_{2,\text{bw}}$	$\hat{b}^+_{1,\text{bw}}$	$\hat{b}^+_{2,\text{bw}}$	$\hat{b}_{1,\text{wg}}$
3	100	5	1.109	1.805	1.186	2.193	1.320	2.703	0.259
3	200	5	0.512	0.891	0.567	1.122	0.640	1.378	0.136
3	500	5	0.218	0.349	0.235	0.428	0.259	0.518	0.054
3	1000	5	0.106	0.175	0.113	0.215	0.126	0.264	0.027
3	100	10	1.023	1.752	1.142	2.256	1.193	2.434	0.110
3	200	10	0.521	0.847	0.551	1.050	0.569	1.129	0.056
3	500	10	0.194	0.327	0.211	0.421	0.220	0.453	0.023
3	1000	10	0.097	0.162	0.105	0.206	0.110	0.223	0.011
3	100	20	1.009	1.567	1.108	2.091	1.130	2.183	0.050
3	200	20	0.488	0.822	0.531	1.083	0.543	1.129	0.025
3	500	20	0.188	0.316	0.208	0.417	0.213	0.437	0.011
3	1000	20	0.088	0.146	0.096	0.195	0.098	0.205	0.005
2	100	5	1.195	1.654	1.320	2.327	1.789	4.053	0.259
2	200	5	0.545	0.788	0.621	1.165	0.853	1.978	0.136
2	500	5	0.233	0.316	0.260	0.457	0.341	0.755	0.054
2	1000	5	0.115	0.160	0.126	0.229	0.167	0.387	0.027
2	100	10	1.066	1.520	1.235	2.376	1.380	2.883	0.110
2	200	10	0.561	0.757	0.599	1.095	0.652	1.315	0.056
2	500	10	0.206	0.286	0.227	0.442	0.250	0.531	0.023
2	1000	10	0.103	0.144	0.113	0.217	0.125	0.263	0.011
2	100	20	1.045	1.316	1.145	2.118	1.204	2.352	0.050
2	200	20	0.509	0.698	0.551	1.107	0.580	1.225	0.025
2	500	20	0.196	0.272	0.216	0.427	0.229	0.477	0.011
2	1000	20	0.093	0.125	0.100	0.199	0.105	0.222	0.005

Table 5: Comparison between ordinary and modified between group estimators

$\beta_1 = \beta_2 = 0$, but $\gamma = -1$

σ_v^2	N	T	Size (5%)×10²				Bias×10²				Variance×10²			
			$b_{1,\mathrm{bw}}$	$b_{2,\mathrm{bw}}$	$b_{1,\mathrm{bw}}^+$	$b_{2,\mathrm{bw}}^+$	$b_{1,\mathrm{bw}}$	$b_{2,\mathrm{bw}}$	$b_{1,\mathrm{bw}}^+$	$b_{2,\mathrm{bw}}^+$	$b_{1,\mathrm{bw}}$	$b_{2,\mathrm{bw}}$	$b_{1,\mathrm{bw}}^+$	$b_{2,\mathrm{bw}}^+$
3	100	5	6.8	7.6	5.1	5.7	-3.0	6.4	1.0	-1.8	1.2	2.5	1.2	2.6
	200	5	8.6	13	5.2	5.8	-3.8	7.5	0.1	-0.3	0.7	1.6	0.6	1.2
	500	5	13	22	5.9	5.1	-3.6	7.1	0.2	-0.6	0.3	0.9	0.2	0.5
	1000	5	20	36	4.4	5.0	-3.5	7.0	0.3	-0.7	0.2	0.7	0.1	0.2
	100	10	5.7	7.0	5.4	6.1	-2.4	4.6	-0.7	1.1	1.1	2.3	1.1	2.3
	200	10	6.6	7.5	6.1	5.0	-1.9	3.8	-0.2	0.4	0.6	1.2	0.5	1.1
	500	10	6.6	9.5	5.1	5.5	-2.0	3.8	-0.3	0.4	0.2	0.6	0.2	0.4
	1000	10	9.5	14	5.1	5.3	-1.9	4.0	-0.3	0.7	0.1	0.4	0.1	0.2
	100	20	6.5	6.2	6.5	5.5	-1.2	2.0	-0.2	0.0	1.1	2.1	1.1	2.2
	200	20	5.5	6.5	5.1	6.1	-0.9	2.0	0.1	0.0	0.5	1.1	0.5	1.1
	500	20	5.7	6.5	5.0	5.4	-1.1	2.1	-0.1	0.2	0.2	0.5	0.2	0.4
	1000	20	5.3	5.9	3.9	4.5	-1.0	2.0	0.0	0.0	0.1	0.2	0.1	0.2
2	100	5	11	19	5.5	6.8	-6.8	14	1.8	-3.2	1.6	4.0	1.5	3.3
	200	5	19	34	5.0	6.5	-7.6	15	0.6	-1.2	1.1	3.3	0.7	1.5
	500	5	37	65	6.0	6.0	-7.4	15	0.6	-1.4	0.8	2.5	0.3	0.6
	1000	5	60	92	4.5	6.5	-7.3	15	0.7	-1.5	0.6	2.3	0.1	0.3
	100	10	8.3	11	5.5	6.5	-4.7	9.1	-1.0	1.7	1.3	2.9	1.2	2.5
	200	10	9.1	14	5.7	5.7	-4.2	8.4	-0.5	1.1	0.7	1.7	0.6	1.2
	500	10	15	28	4.9	6.3	-4.2	8.3	-0.6	1.1	0.4	1.1	0.2	0.5
	1000	10	26	48	5.4	6.5	-4.2	8.5	-0.6	1.4	0.3	0.9	0.1	0.2
	100	20	7.4	6.6	6.4	5.6	-2.5	4.6	-0.2	0.1	1.2	2.3	1.2	2.3
	200	20	6.0	8.3	5.4	6.2	-2.1	4.4	0.0	0.1	0.6	1.2	0.6	1.2
	500	20	7.6	12	5.1	5.7	-2.3	4.6	-0.2	0.4	0.3	0.6	0.2	0.4
	1000	20	9.9	17	3.7	4.5	-2.3	4.5	-0.1	0.2	0.1	0.4	0.1	0.2

Table 6: Finite Sample Performance of De-augmented Factor Estimators

$\beta = \phi = 0$, and $\gamma = 2$

σ_v^2	N	T	Size (5%×10^2)				Bias×10^2				Variance×10^2			
			$\hat{b}_{\text{bw}}$	$\hat{b}_{\text{bw}}^{+}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{b}_{\text{bw}}$	$\hat{b}_{\text{bw}}^{+}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{b}_{\text{bw}}$	$\hat{b}_{\text{bw}}^{+}$	$\hat{\phi}_1$	$\hat{\phi}_2$
3	50	20	5.7	5.1	46	9.3	2.7	1.1	92	1.2	2.2	2.1	123	8.5
	100	20	4.6	4.0	45	7.2	2.2	0.5	90	0.5	1.1	1.0	117	6.3
	200	20	6.5	5.6	44	7.7	2.1	0.4	89	0.6	0.6	0.6	118	6.8
	50	30	6.9	6.5	63	9.9	1.8	0.6	92	1.3	2.4	2.4	109	5.6
	100	30	6.4	5.7	63	5.9	1.4	0.1	92	0.2	1.1	1.1	108	4.0
	200	30	5.4	5.3	62	5.7	1.4	0.2	94	0.1	0.6	0.5	109	3.9
	50	50	6.5	6.2	84	9.3	0.9	0.1	96	1.3	2.3	2.3	107	3.2
	100	50	4.9	4.7	85	5.8	1.3	0.5	96	0.3	1.0	1.0	105	2.2
	200	50	5.3	5.3	86	5.8	0.6	-0.1	96	0.0	0.5	0.5	103	2.3
2	50	20	6.6	5.2	46	9.3	5.3	1.8	92	1.2	2.5	2.3	123	8.5
	100	20	6.9	3.9	45	7.2	4.7	1.0	90	0.5	1.3	1.1	117	6.3
	200	20	9.5	5.1	44	7.7	4.6	0.9	89	0.6	0.8	0.6	118	6.8
	50	30	7.3	6.0	63	9.9	3.6	0.9	92	1.3	2.5	2.5	109	5.6
	100	30	8.0	5.5	63	5.9	3.1	0.3	92	0.2	1.2	1.2	108	4.0
	200	30	7.9	5.1	62	5.7	3.1	0.4	94	0.1	0.7	0.6	109	3.9
	50	50	6.5	5.6	84	9.3	2.0	0.2	96	1.3	2.3	2.3	107	3.2
	100	50	5.0	4.3	85	5.8	2.4	0.6	96	0.3	1.0	1.0	105	2.2
	200	50	5.8	4.9	86	5.8	1.7	-0.1	96	0.0	0.6	0.6	103	2.3

6 Technical Appendix

The following lemma is required for the proof of Theorem 1.

Lemma 1 *For any number or nonsingular matrix* $\mathbf{A}$ *and* $\mathbf{B}$*, we can decompose* $(\mathbf{A}+\mathbf{B})^{-1}(\mathbf{C}+\mathbf{D}) = \boldsymbol{\alpha}\mathbf{A}^{-1}\mathbf{C}+(\mathbf{I}-\boldsymbol{\alpha})\mathbf{B}^{-1}\mathbf{D}$ *where* $\boldsymbol{\alpha}=\mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}$.

Proof: Note that

$$\begin{aligned}
(\mathbf{A}+\mathbf{B})^{-1}(\mathbf{C}+\mathbf{D}) &= (\mathbf{A}+\mathbf{B})^{-1}\mathbf{C}+(\mathbf{A}+\mathbf{B})^{-1}\mathbf{D} \\
&= (\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}\mathbf{A}^{-1}\mathbf{C}+(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}\mathbf{B}^{-1}\mathbf{D} \\
&= \boldsymbol{\alpha}\mathbf{A}^{-1}\mathbf{C}+(\mathbf{I}-\boldsymbol{\alpha})\mathbf{B}^{-1}\mathbf{D}.
\end{aligned}$$

Proof of Theorem 1 Define $\text{plim}_{N\to\infty}\left[\mathbf{Q}_{\mu xt}^{-1}\mathbf{Q}_{\mu xyt}\right]=\boldsymbol{\beta}$ and $\text{plim}_{N\to\infty}\left[\mathbf{Q}_{xot}^{-1}\mathbf{Q}_{xyot}\right]=\boldsymbol{\gamma}$. Then from Lemma 1, it is straightforward to show that

$$\text{plim}_{N\to\infty}\hat{\mathbf{b}}_{\text{cross,t}} = (\mathbf{Q}_{\mu xt}+\mathbf{Q}_{xot})^{-1}\left(\mathbf{Q}_{\mu xyt}+\mathbf{Q}_{xyot}\right) = \mathbf{W}_t\boldsymbol{\beta}+(\mathbf{I}-\mathbf{W}_t)\boldsymbol{\gamma}.$$

The remainder of proofs follow by using Lemma 1. Hence they are omitted here.

Proof of Theorem 2 It is enough to show that

$$\text{E}\left(\frac{\bar{\mathbf{x}}^{*\prime}\bar{\mathbf{x}}^{*}}{N}-\frac{\bar{\mathbf{x}}^{+\prime}\bar{\mathbf{x}}^{+}}{N}\right)=O\left(T^{-2}\right).$$

Note that

$$\begin{aligned}
&\text{plim}_{N\to\infty}\frac{1}{N}\sum^{N}\left[(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})-\left(\bar{\mathbf{x}}_i^{o+}-\bar{\mathbf{x}}^{o+}\right)'\left(\bar{\mathbf{x}}_i^{o+}-\bar{\mathbf{x}}^{o+}\right)\right] \\
=\ &\mathbf{Q}_{\mu x}+\frac{\Omega_{xo}}{T}-\frac{\Omega_{xo}^{+}}{T}=\mathbf{Q}_{\mu x}+O\left(T^{-2}\right), \text{ say.}
\end{aligned}$$

Under Assumption A,

$$\begin{aligned}
\frac{\Omega_{xo}-\Omega_{xo}^{+}}{T} &= E\frac{1}{N}\frac{1}{T^2}\sum^{N}\left(\sum_{t=1}^{T/2}\mathbf{x}_{it}^{o}\right)'\left(\sum_{t=T/2+1}^{T}\mathbf{x}_{it}^{o}\right)<\frac{4}{T^2}\sum_{s=1}^{T/2-1}s\left|\Gamma(s)\right| \\
&< \frac{4}{T^2}\mathbf{M}=O\left(T^{-2}\right).
\end{aligned}$$

Note that when $\mathbf{x}_{it}^{o}$ is not serially correlated, $\Omega_{xo}=\Omega_{xo}^{+}$ so that there is no bias.