Cross-Section Estimation of Long-Run Relations Using Time-Compressed Data

Nikolay Gospodinov* Serena Ng †

August 5, 2026
Preliminary and Incomplete
Please do not Quote

Abstract

Empirical investigations of long-run relations are sometimes based on cross-section regressions in averaged or long differenced data that effectively have the time dimension of a $T \times N$ panel compressed. We analyze a class of time-compressed data and show that they have magnified variability stemming from the fact that the cross-section variance of a non-stationary panel 'fans out' with time. Cross-section regressions in time compressed data can potentially yield estimates that are super-consistent and asymptotically normal, whether the regressors are stationary, non-stationary, or highly persistent. The fastest convergence rate of $\sqrt{NT}$ requires a compression scheme that not only magnifies the non-stationary signal, but also dilutes the regression noise. Omitted fixed effects preclude noise dilution but the estimates remain super-consistent. However, the fanning out effect can be weak when the data have a strong force for mean-reversion or convergence, a problem that seems relevant for temperature data. We consider three applications and find that the long-run relation between consumption and income, and between growth/inflation and demographic variables are reasonably well determined, but the estimated relation between growth and warming temperature is fragile.

Keywords: climate, aging, non-stationarity, long-differencing, Ricardian regressions.
JEL Classification: C2, Q5

*Federal Reserve Bank of Atlanta, Email: nikolay.gospodinov@atl.frb.org

†Dept. of Economics, Columbia University and NBER. 1126D-420 W. 118 St., New York, NY 10025, email: serena.ng at columbia.edu

Financial Support from the National Science Foundation (SES 2018368) is gratefully acknowledged. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

1 Introduction

Many long-run economic relations are of interest:- between consumption and output, productivity growth and real interest rates, public debt and GDP, purchasing power parity, inflation and interest rate, inflation and demographics, interest rates of different maturities. While it is common to study these relations using time series or panel regressions,[1] they have also been studied by employing long-averages of time series data in cross-section regressions. Examples include investigation of savings and investment in Feldstein and Horioka (1980), growth and its determinants considered in Feldstein and Horioka (1980); Barro (1991); Easterly and Levine (1997) and Mendelsohn et al. (1994). Long-differences of averaged data are increasingly used to estimate 'longer run' relation between economic and environmental variables, as in Dell et al. (2012); Burke and Emerick (2016). The data after averaging and differncing essentially have no time dimension, or what we will refer to as *time compressed*. What are the properties of the time compressed data and would non-stationarity matter? Is one method better than another? How effective are such data in recovering long-run relations, and more generally, do time series properties of the data affect cross-section regressions? This paper provides such an analysis.

Our point of departure is the result in Deaton and Paxson (1994) that income inequality will increase with age if individual incomes are random walks, with the implication that the cross-section variance within a cohort will 'fan-out' with age. Built on this insight, we study the properties of a class of time compression functions and obtain three results. First, the cross-section variance of data driven by simple stochastic trends will 'fan-out' (or magnified) over time. The maximum fanning out is obtained by 'point sampling' the last row of a $T \times N$ panel, while averaging the data over K periods has a shrinkage factor of one-third. Different forms of non-stationarity may change the fanning out profile, but significant variation in cross-section variance can be expected.

Second, the magnified signal from fanning out can be exploited in cross-section regressions. Theorem 1 shows that in the local-to-unity framework, the variance of time compressed data traverses smoothly from $O(1)$ to $O(K)$ as the largest autoregressive root moves towards the unit circle. The convergence rate of the least squares estimator self-adjusts with the degree of persistence. Even though cross-section regressions do not explicitly use time variations, fast convergence is a direct consequence of the strong signal in highly persistent data. Third, we show that super-consistent (faster than $\sqrt{N}$) and asymptotically normal estimates are usually possible. Theorem 2 shows that the fastest convergence rate requires that time compression performs both signal magnification and noise dilution. While K-averaging achieves this rate, point sampling does not because of the

[1] See, for example, Dees et al. (2007), Lunsford (2017), Yi and Zhang (2017), Jong-Won Yoon and Lee (2014), and Juselius and Takats (2021)

inability to dilute noise.

The main appeal of cross-section regressions in time compressed data is that standard normal inference can be used without knowing whether the regressors are I(1) or I(0), with no need for HAC standard errors. In contrast, the asymptotic distributions from time series or panel regressions are either discontinuous if the regressor has a unit root, and/or have nuisance parameters that complicate inference. The cross-section estimates are also robust to endogeneity and low-order misspecification biases because of the magnified signal in the regressors. Thus, while omitted fixed effects preclude the fastest convergence rate, super consistent and asymptotically normal estimates can still be attained at the cost of an inflated variance. Cross-section regressions based on time averaged data thus have more desirable theoretical properties than previously understood. In practice, fanning out may not be realized if the data exhibit a strong tendency for mean-reversion or convergence that work against the magnified signal. Checking the stability of the estimates with respect to K is effective in detecting this possibility. In the three applications considered, the problem does not seem to be of concern in regressions of consumption on income, and of growth or inflation on demographic variables. However, time compression schemes seem challenged to disentangle a warming trend from regular temperature fluctuations. In such cases, the long and short run estimates may well appear similar, a finding that has been attributed to a lack of adaptation in the literature.

The paper proceeds as follows. Section 2 begins with the key insight in Deaton and Paxson (1994) and Section 3 proceeds to analyze a class of time compression functions that include the important cases of point-sampling and K-averaging but exclude long-differencing. Section 4 considers estimation of a static model by cross-section regression in time compressed data under specific assumptions. Section 5 relaxes these assumptions. Section 6 considers mechanisms that can dampen the fanning out in the context of three applications. Section 7 concludes.

2 Fanning Out of Cross-Section Variations

This section studies the cross-section variability in a panel of N non-stationary series observed over T periods. We will let $\mathbb{E}_i[\cdot]$ and $\text{var}_i[\cdot]$ denote expectation and variance over the cross-section distribution of units, while $var(\cdot)$ is variance for a given i.

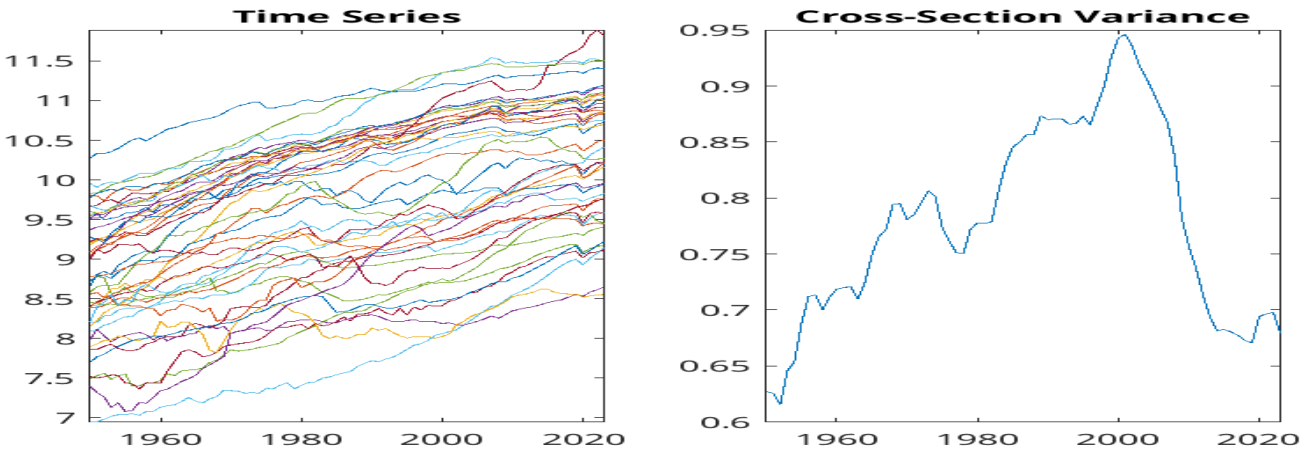
Our point of departure is the result in Deaton and Paxson (1994) that if utility is quadratic and the permanent income hypothesis (PIH) holds, consumption and income inequality will increase with age 'a'. The result arises because consumption (c) is a random walk under PIH, and can be characterized as $c_{ia} = c_{i,a-1} + u_{ia}$. Assuming in addition that $\text{cov}(c_{i,a-1}, u_{ia}) = 0$, the cross-section variance of a given set of N individuals in the same cohort at age a is related to the one at age $a-1$

by $\text{var}_{i|a}(c_{ia}) = \text{var}_{i|a-1}(c_{i,a-1}) + \sigma_i^2$. It follows that the cross-section variance of the cohort will increase with $'a'$, reflecting the cumulative difference in individual shocks as they aged. Ng (2008) shows that if a fraction κ of the panel is non-stationary, the cross-section variance will still fans out at rate κt. Deaton and Paxson (1994) noted that other preferences may increase or decrease consumption dispersion over time, and it can even be nullified if individuals are homogeneous. We now explore the fanning out effect outside of the PIH context, and under more general conditions.

2.1 Generalized Fanning out

Our working assumption is that the data are non-stationary but not necessarily a random walk. Consider log per-capita real GDP for 47 countries over the sample 1960-2020 taken from the Penn-World Table, Feenstra et al. (2015). The left panel of Figure 1 shows that the time series are non-stationary, and the right panel shows that the cross-section variance fans until 2000 when σ-convergence began. Many national accounts variables in this database have this property. Ratios such as export-to-GDP, import-to-GDP, debt-to-GDP in the Global Macroeconomic Database provided in Muller et al. (2025) also have cross-section variances that fan out over a long span.

Figure 1: log Real GDP Per-Capita: 47 Countries in PWT



To study the trending cross-section variance, we start with a simppe data generating process. For $t = 1, \ldots T$ with $x_{i0} = 0$, let $\Delta x_{it} = u_{it}$, $u_{it} = C_i(L)\epsilon_{it}$, where $\epsilon_{it} \sim (0, \sigma_i^2)$ is a white noise error with variance σ_i^2 that is independent of C_{ij} for all j, $C_i(L) = 1 + C_{i1}L + C_{i2}L^2 + \ldots$ is a polynomial in the lag operator L satisfying $\sum_{j=0}^{\infty} C_{ij}^2 < \infty$. Using $C_i(L) = C_i(1) + (1-L)C_i^*(L)$, where $C_i(1) = \sum_{j=0}^{\infty} C_{ij}$ and $C_{ij}^* = -\sum_{k=j+1}^{\infty} C_{ik}$, the Beveridge-Nelson Decomposition yields

$$x_{it} = C_i(1)\sum_{s=1}^{t}\epsilon_{is} + C_i^*(L)\epsilon_{it} = C_i(1)\sum_{s=1}^{t}\epsilon_{is} + w_{it}.$$

In the above, x_{it} is decomposed into a random walk (non-stationary or I(1)) component $C_i(1)\sum_{j=1}^{t}\epsilon_{ij}$, and a mean-reverting (stationary or I(0)) component $w_{it} = C_i^*(L)u_{it}$. For a given i,

$$\text{var}(x_{it}) = C_i(1)^2\sigma_i^2 t + \text{var}(w_t)$$

where $C_i(1)^2\sigma_i^2 = \omega_i^2$ is the long run variance of Δx_{it}. Thus unlike an I(0) process when $C_i(1) = 0$ and $\text{var}(x_{it}) = O(1)$, the variance of an I(1) process increases with calendar time t. It follows that the cross-section variance fans-out at rate t:

$$\text{var}_i(x_{it}) = \mathbb{E}_i[\omega_i^2]\cdot t + \text{var}_i[w_{i,t}] = O(t).$$

Fanning out requires $\mathbb{E}_i[\omega_i^2] \neq 0$, not that every $\omega_i^2 > 0$.

Introducing serial correlation in u_{it} changes the the pre-factor of the fanning out rate from the average short-run variance $\mathbb{E}_i[\sigma_i^2]$ to the average long-run variance $\mathbb{E}_i[\omega_i^2]$. This, however, is not inconsequential. Consider the IMA(1) model $\Delta x_{it} = \epsilon_{it} + \vartheta_i\epsilon_{it-1}$. With a long-run variance of $\omega_i^2 = (1+\vartheta_i)^2\sigma_i$, the cross-section variance at t is

$$\text{var}_i(x_{it}) = \mathbb{E}_i[(1+\vartheta_i)^2\sigma_i^2]\cdot t + \text{var}_i[\vartheta_i^2\sigma_i^2].$$

The closer is $\mathbb{E}_i[\vartheta_i]$ to -1, the weaker is the fanning out.[2] Two models with IMA(1) as reduced form are: i). the factor model $x_{it} = \lambda_i f_t + e_{it}$ where $f_t = f_{t-1} + u_t$ and w_{it} is stationary, $\text{var}_i(x_{it}) = \text{var}_i(\lambda_i)f_t^2 + \text{var}_i(w_{it})$; and ii). the local level model with $x_{it} = \mu_{it} + w_{it}$, $\mu_{it} = \mu_{it-1} + u_{it}$.

2.2 Convergence and Initial Conditions

The above simple non-stationary structure may not a good characterization of some socio-economic data. For example, the cross-section variance in Figure 1 stops increasing around year 2000 when emerging and populous economies have grown much faster than matured ones since the turn of the century. We now consider alternative forms of non-stationarity and show that some features can interact to generate a more complex cross-sectional variability.

i There may be situations when it is appropriate to assume $x_{i,-T_{int}} = 0$ instead of $x_{i0} = 0$. For example, if x_{it} was earnings, $-T_{init}$ may be age 0, and we observe data post-retirement indexed by $t \in [1, T]$. In this case, the effective cross-section variance is $O(\mathbf{t})$ where $\mathbf{t} = (t + T_{init})$.

ii Suppose $\Delta x_{it} = g_i + u_{it}$ where g_i is an exogenous drift. From $x_{it} = x_{i0} + g_i t + \sum_{s=1}^{t} u_{is}$,

$$\text{var}_i(x_{it}) = \text{var}_i(x_{i0}) + \mathbb{E}_i[\omega_i^2]t + \text{var}_i(g_i)t^2 = O(t^2).$$

The faster fanning out rate of t^2 arises because the deterministic trend dominates the stochastic trend. Nonetheless, heterogeneity in g_i is needed to achieve the faster rate.

[2] Perron and Ng (1996) refer to the case when ϑ is local to unity as nearly integrated, nearly white noise.

iii Suppose that $x_{i0} \neq 0$ and $\text{cov}_i(x_{i0}, g_i) < 0$. Then

$$\begin{aligned} \text{var}_i(x_{it}) &= \text{var}_i(x_{i0}) + \Big(2\text{cov}_i(x_{i0}, g_i) + \mathbb{E}_i[\omega_i^2]\Big)t + \text{var}_i(g_i)t^2 \\ &= \text{var}_i(x_{i0}) + c_1 t + c_2 t^2 = O(t^2). \end{aligned}$$

If $\mathbb{E}_i[\omega_i^2] < 2\text{cov}_i(x_{i0}, g_i)$, then $c_1 < 0$ and $\text{var}_i(x_{it})$ may trend down for some values of t. But this negative trend can only be temporary because $\text{var}_i(g_i)t^2$ grows with t. In such cases, the cross-section variance could exhibit a U-shape with a minimum at $t^* = -\frac{c_1}{2c_2}$.

iv Suppose $x_{it} = \lambda_i f_t + w_{it}$ and $f_t = g_f + f_{t-1} + u_t$, where g_f is the drift of the common factor f_t, and $u_t \sim (0, \sigma^2)$ has long-run variance ω^2. Since $f_t = f_0 + g_f t + \sum_{s=1}^{t} u_s$, it follows that

$$\text{var}_i(x_{it}) = \text{var}_i(x_{i0}) + \text{var}_i(\lambda_i) f_t^2 + \text{var}_i(w_{it}) = O(t^2).$$

But $f_t^2 = (f_0 + g_f t)^2 + O_p(t)$ is quadratic in t. Thus, if $g_f < 0$ and $f_0 \neq 0$, f_t^2 will be U-shaped and reaches a minimum at $t^* = -f_0/g_f$ before increasing. Even with heterogeneity in λ_i, the cross-section variance may not always increase with t.

In general, the fanning out profile will depend on data and the sample. There can be episodes when the cross-section variance will not increase, and could even be O(1). To obtain general results, we will assume that $\Delta x_{it} = u_{it}$ where x_{i0} is exogenous, with the understanding that the fanning out rate may be faster or slower than t obtained under alternative assumptions.

3 A Class of Time Compression Functions

This section shows that the fanning out effect is a feature of a class of time-compressed data. For each $i = 1, \dots T$ and $t = 1, \dots T$, let $\Delta x_{it} = u_{it}$ with $x_{i0} = 0$, let

$$\widetilde{X}_{iT_1} = \sum_{t=T_0+1}^{T_1} w(t) x_{it}.$$

be defined by a time compressed function $w(t)$ satisfying the following.

Assumption A: (i) $w(t) \geq 0$ for all $t \in [T_0 + 1, T_1]$; (ii) $\sum_{t=T_0+1}^{T_1} w(t) = 1$, and (iii) $w(t) > 0$ for some $t \in [T_0 + 1, T_1]$.

Assumption A.(i) and A.(ii) ensure that $\{w(t)\}$ is a proper probability distribution over the interval $[T_0 + 1, T_1]$. A.(iii) ensures that the weights are not concentrated in the early sample.

The class of functions include *concentrated* weights. An example is point sampling, which takes row T_1 from a $T \times N$ panel. Point sampling is a one parameter time compression scheme with $w(t) = 1_{t=T_1}$, and with weights $\phi_j = 1$ for all $j \leq T_1$ on the innovations.

$$\widetilde{X}_{iT_1} = x_{i,T_1} = x_{i0} + \sum_{j=1}^{T_1} u_{ij} = O(T_1).$$

Assumption A.(iii) requires that T_1 is not too close to the origin so that enough innovations are accumulated to generate fanning out. The preceding analysis implies that variability of $\widetilde{X}_{iT_1}$ is maximized by point sampling at $T_1 = T$. It is also immediate that $\text{var}(\widetilde{X}_{iT_1}) = O(1)$ if x is stationary, so the cross-section variance neither fans in nor out.

Whereas fanning out in point-sampling means that $\text{var}_i(x_{it})$ increases with calendar time t, as in Deaton and Paxson (1994), fanning-out using other schemes will be taken to mean that $\text{var}_i(\widetilde{X}_{iK}) = O(K)$ where $K = T_1 - T_0$ depends on at least two choice parameters, T_0 and T_1. Such functions include *spread-out* weights, the simplest being full-sample averaging with $T_0 = 0$, $T_1 = T$. But it also includes a three parameter scheme that gives a weight of zero to the first κ fraction of the sample, and the remaining $1 - \kappa$ fraction would receive a weight of $\frac{1}{(1-\kappa)K}$, with $K = T_1 - T_0$.

3.1 A Closer Look at K Averaging

Many well-cited cross-section regressions are based on time averaged data. For example, Feldstein and Horioka (1980) averaged the data over 15 years, while Barro (1991) averaged the data over 25 years. The WMO standard for long run temperature is 30-years.[3] It is thus useful to take a closer look at averaging data over K periods.

We will use $\bar{X}_{iK}$ to denote time compressed data by averaging. Averaging observations over time is commonly thought to remove high frequency noise, making it easier to visualize patterns over time. Consider first full-sample averaging where $\bar{X}_{i,T} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$. If x_{it} is stationary with long run variance ω_i^2, then by central limit theorem, $\sqrt{T}\bar{X}_{i,T} \xrightarrow{d} N(0, \omega_i^2)$. Time averaging of stationary data reduces noise relative to an arbitrary draw x_{it} at any t. It follows from $\text{var}(\bar{X}_{i,T}) = O(T^{-1})$ that the cross-section variance $\text{var}_i(\bar{X}_{iT}) = O(T^{-1})$ 'fans-in'.

Consider now the I(1) case when $\Delta x_{it} = u_{it}$, where $x_{i0} = 0$, $u_{it} \sim (0, \sigma_i^2)$ is stationary with long-run variance ω_i^2. Then by functional central limit theorem, $T^{-1/2}\bar{X}_{i,T} \Rightarrow \omega_i^2 \int_0^1 W(r)d(r)$, where $W(r) \sim N(0, r)$ is a standard Brownian motion. Since $E[\int_0^1 W(r)dr]^2 = 1/3$, for large T, we have $\text{var}(\bar{X}_{i,T}) = \omega_i^2 \frac{T}{3}$. In contrast to the stationary case when $\text{var}_i(\bar{X}_{iT})$ fans-in, now

$$\text{var}_i(\bar{X}_{iT}) = \mathbb{E}_i[\omega_i^2]\frac{T}{3}$$

[3] See `https://wmo.int/media/news/updated-30-year-reference-period-reflects-changing-climate`.

fans-out, but shrinks by 1/3 relative to point sampling at T.

To understand the reason for this shrinkage, consider averaging $K = T_1 - T_0$ observations between $T_0 + 1$ and T_1, which may be needed if complete data is not available for the full sample, or properties in the sub-sample are of interest. When applied sequentially to a single time series, K-averaging is a low-pass filter that passes all variations with peridocities longer than K and attenuates shorter ones.[4] We use K-averaging only once for each i to generate

$$\bar{X}_{iK} = \frac{1}{K}\sum_{t=T_0+1}^{T_1} x_{it} = x_{i,T_0} + \frac{1}{K}\sum_{t=T_0+1}^{T_1}\sum_{s=T_0+1}^{t} u_{is}.$$

Simplifying the double sum, we obtain

$$\bar{X}_{iK} \;=\; x_{i,T_0} + \sum_{j=1}^{K}\left(\frac{K-j+1}{K}\right)u_{i,T_0+j} = x_{i,T_0} + \sum_{j=1}^{K}\phi_j u_{i,T_0+j}.$$

where $\phi_j = \frac{K-j+1}{K}$ is the weight on innovation u_{i,T_0+j}, and $\sum_{j=1}^{K}\phi_j^2 = \frac{K(K+1)(2K+1)}{6K^2}$. Note that the relevant initial condition for K averaging is x_{i,T_0}, not $x_{i0} = 0$.

Lemma 1 *Suppose that $x_{it} = x_{it-1} + u_{it}$ where $u_{it} \sim (0, \sigma_i^2)$ is independent across i but possibly serially correlated with long-run variance ω_i^2, and $x_{i0} = 0$. Let $\bar{X}_{i,K}$ be the mean of unit i computed from K observations between T_0 and T_1. Let $\bar{a}_K = \frac{(K+1)(2K+1)}{6K}$. For fix K,*

$$var(\bar{X}_{iK}) = var_i(x_{i,T_0}) + (T_0 + \bar{a}_K)\omega_i^2.$$

for large K relative to T_0, $var_i(\bar{X}_{iK}) = \mathbb{E}_i[\omega_i^2]K/3$. □

The cross-section variance has two components: one from the variance of the conditional mean, which in this case is the initial condition, and one from the mean of the conditional variance due to the shocks. Assuming T_0 is small relative to K gives $\text{var}_i(\bar{X}_{iK}) = \mathbb{E}_i[\omega_i^2]K/3$ since $\bar{a}_K/K \approx 1/3$ for large K.

3.2 Properties of $\widetilde{X}_{iK}$

The results for point sampling and K-averaging draw attention to two distinct sets of weights of interest: $\{w(t)\}$ on the data, and $\{\phi_s\}$ on the innovations. This will still be the case as we now analyze the properties of $\widetilde{X}_{iK}$ for any $w(t)$ satisfying Assumption A.

[4] Let $\omega = 2\pi/K$. The discrete Fourier transform of the response is $H(w) = \frac{1}{K}\sum_{m=0}^{K-1} e^{-j\omega m} = \frac{1}{K}\frac{1-e^{-j\omega K}}{1-e^{-j\omega}}$. Using $\sin(\theta) = \frac{e^{j\theta}-e^{-j\theta}}{2j}$, the frequency response is the Dirichlet kernel, $H(\omega) \approx \frac{\sin(K\omega/2}{K\omega/2}$ whose first zero at $\omega = 2\pi/K$.

In the I(1) case, $x_{it} = x_{it-1} + u_{it} = x_{it} = x_{i0} + \sum_{s=T_0+1}^{t} u_{is}$. Changing the order of summation to sum over s first and then for $t > s$, we obtain

$$\begin{aligned}\widetilde{X}_{iK} &= x_{i,T_0} + \left(\sum_{t=T_0+1}^{T_1} w(t) \sum_{s=T_0+1}^{t} u_{is} \right) = x_{i,T_0} + \sum_{s=T_0+1}^{T_1} u_{is} \sum_{t=s}^{T_1} w(t) \\ &= x_{i,T_0} + \sum_{s=T_0+1}^{T_1} \phi_s u_{is}, \qquad \text{with} \qquad \phi_s = \sum_{t=s}^{T_1} w(t).\end{aligned}$$

Note that ϕ_s is the tail sum of the innovations at s, and the sequence $\{\phi_s\}$ is the survival function of the weight distribution w, starting at $\phi_{T_0+1} = \sum_{t=T_0+1}^{T_1} w(t) = 1$ and decreases to $\phi_{T_1} = w(T_1) \geq 0$.

$$\text{var}(\widetilde{X}_{iK}) = \text{var}_i(x_{i,T_0}) + \sigma_i^2 \sum_{s=T_0+1}^{T_1} \phi_s^2 \equiv \sigma_i^2 \Phi_K, \qquad \Phi_K = \widetilde{a}_K(w) K.$$

Condition A.(iii) ensures that $\sum_{s=T_0+1}^{T_1} \phi_s^2 > 0$ for any K. In general, $\Phi_K = O(K)$ but the precise magnitude of fanning out depends on $\widetilde{a}_K(w)$ which is scheme specific. Full sample averaging has $\widetilde{a}_K = 1/3$, while point sampling has $\widetilde{a}_K = 1$.

In the I(0) case, we have, by definition,

$$\text{var}(\widetilde{X}_{iK}) = \text{var}_i(x_{i,T_0}) + \sum_{s=T_0+1}^{T_1} \sum_{t=T_0+1}^{T_1} w(t)w(s)\text{cov}(x_{it}, x_{is}).$$

To make progress, consider the AR(1) case when $x_{it} = \rho x_{i,t-1} + u_{it} = \rho_i^t u_{i0} + \sum_{s=0}^{\infty} \rho_i^s u_{t-s}$ and $u_{it} \sim (0, \sigma_i^2)$. With $\text{cov}(x_{it}, x_{is}) = \frac{\sigma_i^2}{1-\rho_i^2} \rho_i^{|t-s|}$ and assuming ρ_i^t is small, we have

$$\widetilde{X}_{iK} = \sum_{s=T_0+1}^{T_1} \psi_s(\rho_i) u_{is}, \qquad \psi_s(\rho_i) = \sum_{t=s}^{T_1} w(t) \rho_i^{t-s}.$$

The function $\psi_s(\rho_i)$ plays the role of ϕ_s in the unit root case. We show in the Appendix that

$$\begin{aligned}\text{var}(\widetilde{X}_{iK}) &= \frac{\sigma_i^2}{1-\rho_i^2} \sum_{h=-(K-1)}^{(K-1)} \sum_{s,t,t-s=h} w(t)w(s)\rho_i^{|t-s|} \\ &= \frac{\sigma_i^2}{1-\rho_i^2} \sum_{h=-(K-1)}^{K-1} A_h \rho_i^{|h|} \equiv \frac{\sigma_i^2}{1-\rho_i^2} G_w \equiv \sigma_i^2 \Psi_K(\rho_i)\end{aligned}$$

where $A_h = \sum_{s,t,t-s=h}^{T_1} w(t)w(s)$ is the total weight placed on observations s and t separated h periods apart. Under Assumption A, $\sum_{h=-(K-1)}^{K-1} A_h = (\sum_{t=T_0+1}^{T_1} w_t)^2 = 1$, implying that $\{A_h\}$ forms a probability distribution over periods separated by h. Thus, $G_w(\rho_i) = \mathbb{E}_A|\rho_i^{|h|} \in [0,1]$ is bounded for any $\rho_i \in [0,1]$ and any $w(t)$. But $\text{var}(\widetilde{X}_{iK})$ increases with K when x is I(1).

To transition from a fix $\rho_i < 1$ to $\rho_i = 1$, it must be that $\Psi_K = \frac{G_w}{1-\rho_i^2}$ 'blows-up' with K via $\frac{1}{1-\rho_i^2} = \frac{1}{(1-\rho_i)(1+\rho_i)}$ as ρ_i approaches the unit circle. In the local-to-unity case when $\rho_i = 1 - \frac{c_i}{K}$, $G_w = \mathbb{E}_A[\rho_i^{|h|}] \approx 1 - \frac{c_i}{K}\mathbb{E}_A[|h|] = O(1)$ for any choice of $w(t)$ satisfying Assumption A and any finite h. It is $\frac{1}{1-\rho_i} \propto \frac{K}{c_i} = O(K)$ that allows Ψ_K to diverge with K oas ρ_i moves towards the unit circle.

Theorem 1 *Let $\widetilde{X}_{iK} = \sum_{t=T_0+1}^{T_1} w(t)x_{it}$ where $w(t)$ satisfies Assumption A. For a given K, it holds that*

a. $\psi_s(\rho_i) \xrightarrow{\rho_i \to 0} w(s)$ *and* $\psi_s(\rho_i) \xrightarrow{\rho_i \to 1} \phi_s$*;*

b. $\Psi_K(\rho_i) \to \Phi_K = O(K)$ *as* $\rho \to 1$.

Part (a) says that the tail sum ϕ_s interpolates continuously between the two extremes. Part (b) indicates that Ψ_K moves towards the unbounded Φ_K as ρ_i moves towards the unit circle. Indeed, if $\rho_i = 1 - \frac{c_i}{K^d}$, then $\frac{1}{1-\rho_i} \propto \frac{K^d}{c_i}$ for any $d > 1$. This allows $\text{var}(\bar{X}_i)$ to transition smoothly from being bounded to divergent as persistence increases. This result is important to preserve continuity of the asymptotic distribution in regressions that use $\widetilde{X}_{iK}$ as regressor.

3.3 Long Differencing

We use $\breve{X}_{iK}$ to denote data transformed by long differencing (LD) constructed as in Burke et al. (2015), Dell et al. (2012). Precisely, LD is based on the difference of two local average of $M = 2m+1$ observations, one centered at T_0, and one at T_1.

$$\breve{X}_{iK} = \bar{X}_{iM}(T_1) - \bar{X}_{iM}(T_0).$$

It thus requires three parameters, m, T_0 and T_1. Though quite widely used, the properties of LD data are not well understood.[5] If $m = 0$, the LD takes a long difference, $\breve{X}_{iK} = x_{i,T_1} - x_{i,T_0} = \sum_{s=T_0+1}^{T_1} u_{is}$ with $\breve{w}(t) = 1_{t=T_1} - 1_{t=T_0}$. It is immediate that $\text{var}_i(\breve{X}_{iK}) = O(K)$, the same as point sampling. When $m > 0$, $\breve{w}(t) = \frac{1}{M}1_{t\in[T_1-m,T_1+m]} - \frac{1}{M}1_{t\in[T_)-m,T_1+m}$ and its effects on the data are not so transparent. An example helps understand the weights. Suppose $T_0 = 4, T_1 = 10$, and $m = 1$. The observation and innovation weights are:

		3	4	5	6	7	8	9	10	11	12	
$\breve{w}(t)$	..	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0	..
$\breve{\phi}_s$	..	0	$\frac{1}{3}$	$\frac{2}{3}$	1	1	1	1	$\frac{2}{3}$	$\frac{1}{3}$	0	..

[5] Burke and Emerick (2016) simulate x_{it} as having a unit-specific deterministic trend and show that if K is sufficiently large, the L D estimates from a cross-section regression are more precise than the ones obtained from a $T \times N$ panel of raw data. Burke and Emerick (2016) simulated $x_{it} = \mu t + w_{it}$ and $y_{it} = \alpha_{is} + \beta_{\bar{X}}\bar{X}_i + \beta_w w_{it} + \varepsilon_{it}$. Their regression also has a quadratic term that is not important for the arguments to follow.

We see that LD fails Assumption A because not $\breve{w}(t)$ is not always positive and $\sum_t \breve{w}(t) = 0$, not 1. Furthermore, $\breve{\phi}_s$ is not smooth; it jumps from 0 to $1/M$ at $T_0 - m + 1$, and from $1/M$ to 0 at $T_1 + m + 1$. Perhaps the most notable feature of LD is that it puts a weight of 1 on the $K - M + 1$ innovations that are in the middle, while the observations at the two ends have smaller weights. This 'tent' pattern of $\widetilde{\phi}_s$ generalizes, and it can be shown that

$$\text{var}(\breve{X}_{iM}) = \sigma_i^2\left((K - M + 1) + 2\sum_{j=1}^{2m}(j/M)^2\right) = \sigma_i^2\left(K - \frac{(M^2-1)}{3M}\right) \approx \sigma_i^2(K - M/3). \quad (1)$$

Local averaging of M observations does not change the fanning out rate that is determined cross-section heterogeneity over K periods. However, when M is large, the effective rate of $K - M/3$ can be much slower than K. The choice of M is important.

The LD combines local-averaging (a low pass filter with zero at $\omega = \frac{2\pi}{M}$ known to attenuate high frequency variations) with differencing (a high-pass filter with zero at $\omega = 0$, known to attenuate low frequency variations).[6] It has a peak periodicity of roughly $p^* = \frac{2\pi}{\omega^*} = \frac{2MK}{M+K}$, which is the mean of the two boundary frequencies of $\frac{2\pi}{K}$ and $\frac{2\pi}{M}$. However, the use of 'medium' and 'longer' terminology in the literature does not align with the periodicity. Dell et al. (2012) report 'medium-term' effects using $m = 7$ and $K = 15$ which implies $p^* = 15$. In contrast, Burke and Emerick (2016) report 'longer-term effects' using $m = 2$ and $K = 20$, which implies $p^* = 8$. The LD is similar to the filter used in Kuznets (1958) to suggest evidence of a 20-year economic cycle. Howrey (1968) noted that the ten-year difference of two five-year moving-averages could spuriously induce a cycle with a 20 year periodicity even if the raw data were a white noise. However, the LD the transformation will only be used once at $t = T_0$ and T_1 to collapse the time variations into $\breve{X}_{iK}$.

We illustrate the extent of signal magnification by simulating $x_{it} = \mu_{it} + w_{it}$ where μ_{it} is non-stationary and w_{it} is stationary. Two models are considered.

1. local level model where $\mu_{it} = \mu_{i,t-1} + \epsilon_{it}$ with $q_i = \frac{\sigma_\epsilon^2}{\sigma_w^2}$:

2 factor model: $\mu_{it} = \lambda_i \mu_t + u_{it}$ where $\mu_t = \mu_{t-1} + \epsilon_t$, $\lambda_i \sim (0, \sigma_\lambda^2)$ with $q_i = \frac{\sigma_\lambda^2 \sigma_\epsilon^2}{\sigma_w^2}$.

where $e_{it} \sim (0, \tau_i^2)$, $\epsilon_{it} \sim (0, \sigma_{\epsilon,i}^2)$ and $w_{it} \sim (0, \sigma_w^2)$ are mutually uncorrelated. The local-level model provides a permanent-transitory decomposition of the data. Since $q_i = \frac{\sigma_\epsilon^2}{\sigma_w^2}$ for the local-level model with, signal magnification requires the average variability of μ_{it} to dominate those of the mean-reverting component w_{it}. The factor model provides a common-idiosyncratic representation of the data with non-stationary coming from the common trend μ_t. Both have an IMA(1) model

[6] The squared gain $H_{LD}(\omega)|^2 \propto \left(\frac{\sin(M\omega/2)}{M\sin(\omega/2)}\right)^2 4\sin^2(\frac{\omega K}{2})$ is a product of a function increasing from 0 at $\omega = 2\pi/K$ and another function that decreases to zero at $\omega = 2\pi/M$. The peak, ω^*. thus lies between the two zones.

$\Delta x_{it} = u_{it} + \vartheta_i u_{it-1}$ as reduced form, where $\vartheta_i = [-(q_i+2) \pm \sqrt{q_i^2 + 4q_i}]/2$ is closer to -1 the smaller is signal-noise ratio q. For the local level model, $q_i = \frac{\sigma_u^2}{\sigma_w^2}$, and for the factor model, $q_i = \frac{\lambda_i^2 \sigma_u^2}{\sigma_w^2}$. Thus, a strong common trend is not enough for signal magnification; it also requires the loadings to be sufficiently heterogeneous.

Table 1: Cross-Section Variances, $(T_0, T_1, T_2, T) = (10, 20, 40, 50)$

Local-Level Model

		PS:$\text{var}_i(x_{i,t})$			KA(0,T_0): $\text{var}_i(\bar{X}_{i,K})$			DA: $\text{var}_i(\tilde{X}_{i,K})$			LD(m, T_1): $\text{var}_i(\check{X}_{i,K})$		
σ_u	ϑ	T_0	T_1	T	$(0,T_0)$	$(0,T_1)$	$(0,T)$	$(0,T_0)$	$(0,T_1)$	$(0,T)$	$(0,T_2)$	$(4,T_2)$	$(9,T_2)$
3.000	-0.092	90.913	181.017	451.693	34.815	64.597	154.725	65.012	124.567	305.055	182.019	153.661	123.308
1.000	-0.382	10.981	20.993	51.078	3.958	7.222	17.208	7.399	13.926	33.930	21.992	17.270	13.794
0.500	-0.610	3.492	5.994	13.521	1.065	1.843	4.317	1.999	3.555	8.512	6.995	4.484	3.527
0.200	-0.819	1.397	1.797	3.005	0.255	0.337	0.707	0.487	0.652	1.396	2.799	0.904	0.653
0.100	-0.905	1.099	1.198	1.503	0.139	0.122	0.192	0.272	0.238	0.379	2.201	0.393	0.242
0.000	-1.000	1.000	0.999	1.003	0.100	0.050	0.020	0.200	0.100	0.040	2.002	0.222	0.106

Factor Model

σ_λ	ϑ	T_0	T_1	T	$(0,T_0)$	$(0,T_1)$	$(0,T)$	$(0,T_0)$	$(0,T_1)$	$(0,T)$	$(0,T_2)$	$(4,T_2)$	$(9,T_2)$
3.000	-0.092	93.578	186.698	462.614	35.494	66.817	158.891	66.374	129.000	310.905	180.487	151.912	123.125
1.000	-0.382	11.282	21.639	52.287	4.033	7.469	17.672	7.553	14.424	34.578	21.847	17.075	13.772
0.600	-0.554	4.700	8.431	19.463	1.516	2.721	6.374	2.848	5.257	12.473	9.151	6.289	5.025
0.400	-0.672	2.644	4.303	9.206	0.730	1.237	2.844	1.377	2.392	5.565	5.182	2.918	2.291
0.200	-0.819	1.410	1.825	3.053	0.258	0.347	0.726	0.494	0.673	1.421	2.799	0.896	0.652
0.000	-1.000	1.000	0.998	1.003	0.100	0.050	0.020	0.200	0.100	0.040	2.002	0.222	0.106

Note: (i) PS is point sampling; (ii) KA is K-averaging with $w(t) = 1/K$ for $t \in [T_a+1, T_b]$; (iii) DA averaging uses $w(t) = 0$ for $t \in [T_a+1, T_b-m]$ and $w(t) = 1/m$ for $t \in [T_b-m+1, T_b]$, $m = K/2$; (iv) LD is long-differencing of two local averages.

Table 1 reports the cross-section variance of four time compression schemes: (i) point sampling (PS), (ii) K-averaging (KA), (iii) an averaging scheme (DA) with $w_t = 0$ for $t \in [T_0+1, T_1-m]$ and $w_t = 1/m$ for $t \in [T_1-m+1, T_1]$ where $m = K/2$, $K = T_1 - T_0$, and (iv) long-differencing (LD). We evaluate LD at T_2 to accommodate m initial observations; it is thus comparable to PS at T_1. The cross-sectional variances in the first panel supports point sampling has maximum fanning out. The second panel for K sampling is in line with Lemma 1 that the fanning out is about 1/3 of point sampling. The third panel shows that by reducing the weights on the early observations, the fanning out effect of DA is larger than K-averaging that gives equal weight to all observations. The fourth panel shows that fanning out of LD is smaller the larger is M. The results also show that the signal magnification effect is smaller the closer is ϑ is to -1.

4 Cross-Section Regressions in Time Compressed Data

This section studies cross-section regression in time compressed data. Let $z_{it} = (x_{it}, y_{iy})$ be a vector of n variables characterized by $B_0 z_{it} = B_1 z_{i,t-1} + \epsilon_{it}$. To begin, we assume that $n = 2$, B_0 and B_1 are the same across units and specified to have the following structure:

$$x_{it} = x_{it-1} + u_{it} \tag{2}$$

$$y_{it} = \beta x_{it} + e_{it}, \qquad . \tag{3}$$

where $u_{it} \sim (0, \sigma_i^2)$ and $e_{it} \sim (0, \tau_i^2)$ are stationary and uncorrelated. Since x_{it} is I(1) and e_{it} is stationary by assumption, $(1, -\beta)$ is a cointegrating vector in the sense of Engle and Granger (1987). A time series regression using T observations of y and x of a particular unit i will converge at the fast rate of T, but the asymptotic distribution of $\hat{\beta}$ is, in general, non-standard. If we focus on the low frequency variations as in Mueller and Watson (2017), we have a small sample problem because few cycles with long periodicity will be observed over a short span. Panel regressions also have a fast convergence rate of $\sqrt{N}T$, and estimation can be quite involved if there is cross-section dependence. Instead, we consider least-squares estimation of a 'timeless', cross-section between-group regression in time compressed data $\widetilde{Y}_{iK}$ and $\widetilde{X}_{iK}$:

$$\widetilde{Y}_{iK} = c + \beta \widetilde{X}_{iK} + \widetilde{e}_{ik}. \tag{4}$$

We first consider the properties of $\hat{\beta}$ under the following assumptions.

Assumption B: (i) $u_{is} \perp e_{it}$ for all s, t and for all i; (ii). $u_{it} \sim (0, \sigma_i^2)$ is serially uncorrelated with $\mathbb{E}_i[\sigma_i^2] > 0$, (iii), $e_{it} \sim (0, \tau_i^2)$ is stationary; (iv.) $x_{i0} = 0$ for all i, (v.) $\{x_{it}, e_{it}\}$ and $(\bar{X}_{iK}, \bar{e}_{iK})$ are i.i.d. across i with finite fourth moments.

Assumption C: (i) $\Phi = \lim_{K\to\infty} \Phi_K/K = \widetilde{a}_K > 0$ and C.(ii) $A_0 = \sum_{t=T_0+1}^{T_1} w_t^2 = O(K^{-1})$.

We assume that u_{is} and e_{it} are independent across i, and for all s, t so that strict exogeneity holds. We do not require $\sigma_i^2 > 0$ for every i, only that $\mathbb{E}_i[\sigma_i^2] > 0$. The condition is needed for identification as it ensures $Q_{XX} = \text{var}_i(\widetilde{X}_{iK})$ is non-degenerate. Assumption B.iii allows e_{it} to be serially correlated but cannot be non-stationary to rule out spurious regressions. Cross-section dependence is not allowed for now and Assumption B.iv simplifies the analysis. Assumptions B.v allows us to use LLN and CLT for iid variables. In particular, $\frac{1}{N}\sum_i f(\bar{X}_{iK}, \bar{e}_{iK}) \xrightarrow{p} \mathbb{E}_i[f(\bar{X}_{iK}, \bar{e}_{iK})]$, giving

$$\frac{1}{N}\sum_{i=1}^{N} (\widetilde{X}_{iK} - \bar{\widetilde{X}}_K)^2 \xrightarrow{p} \text{var}_i(\widetilde{X}_{iK}) = \mathbb{E}_i[\sigma_i^2]\Phi_K \equiv Q_{XX},$$

and

$$\frac{1}{\sqrt{N}}\sum_i(\widetilde{X}_{iK}-\bar{\widetilde{X}}_K)\widetilde{e}_{iK}\xrightarrow{d}N(0,V_K),\quad \text{where}\quad V_K=\mathbb{E}_i[(\widetilde{X}_{iK}-\mu_X)^2\widetilde{e}_{iK}].$$

The properties of the least squares estimation is determined by

$$(\hat{\beta}-\beta) \;=\; \frac{N^{-1}\sum_{i=1}^N(\widetilde{X}_{iK}-\bar{\widetilde{X}}_K)\widetilde{e}_{iK}}{N^{-1}\sum_i(\widetilde{X}_{iK}-\bar{\widetilde{X}}_K)^2}.$$

Under Assumption B, strict exogeneity ensures that the estimator is $\sqrt{N}$ consistent when x is I(0) as the denominator of $\hat{\beta}-\beta$ converges to Q_{XX} which is $O(1)$, and the numerator scaled by $\sqrt{N}$ obeys a central limit theorem. But a faster convergence rate can be obtained when x is I(1) because Q_{XX} is then $O(K)$. We will refer to this as the *signal magnification* effect. It is scheme specific since $\Phi_K=\sum_{j=T_0+1}^{T_1}\phi_j^2$. Now $\text{var}(\widetilde{e}_{iK})=\text{var}(\sum_t w(t)e_{it})=\tau_i^2\sum_t w_t^2=\tau_i^2A_0(w)$. If the time compression scheme permits *noise dilution* in the sense that $\text{var}(\widetilde{e}_{iK})=o(K^{-1})$, an even faster rate can be obtained. Assumption C makes these requirements precise. Assumptions C.i ensures signal magnification/ Assumption C.ii ensures noise dilution. Obviously, if we point sampled $t=T_1+1$ and T_1 is small, then $\Phi_K=1$ and both C.i and C.ii would fail. If we point sample at a large T_1, C.i is satisfied, but C.ii fails because $\sum_t w(t)^2=1$. Time compression schemes with spread out weights satisfy both C.1 and C.ii, and of these, K-averaging is the simplest.[7]

Theorem 2 *Let $\widetilde{X}_{iK}=\sum_{t=T_0+1}^{T_1}w(t)x_{it}$, and let $\hat{\beta}$ be estimated from a correctly specified static regression $\widetilde{Y}_{iK}=c+\beta\widetilde{X}_{iK}+\widetilde{e}_{iK}$. Let Assumptions A and B hold and var$(x_{i,T_0})=o(1)$.*

a. With K fixed and whether x is I(1) or I(0), $\widetilde{\beta}\xrightarrow{p}\beta$ as $N\to\infty$. Furthermore, $\sqrt{N}(\hat{\beta}-\beta)\xrightarrow{d}N(0,Q_{XX}^{-1}VQ_X^{-1})$ where $V=var_i((\widetilde{X}_{iK}-\mu_X)\widetilde{e}_{iK})$.

b Let Assumption C.i hold but not C.ii. Then for large T_1, $\sqrt{NT_1}(\hat{\beta}-\beta)\xrightarrow{d}N(0,\mathcal{Q}_{XX}^{-1}\mathcal{V}_1\mathcal{Q}_{XX}^{-1})$, $\mathcal{V}_1=\lim_{T_1\to\infty}V$.

c Let Assumption C.ii hold but not C.i. Then for large K, $\sqrt{NK}(\hat{\beta}-\beta)\xrightarrow{d}N(0,\mathcal{Q}_{XX}^{-1}\mathcal{V}_2\mathcal{Q}_{XX}^{-1}$, $\mathcal{V}_2=\lim_{T_1\to\infty}V$.

d Let Assumption C.i and C.ii also hold, $\mathcal{Q}_{XX}=\lim_{K\to\infty}K^{-1}Q_{XX}$ and $\mathcal{V}_4=\lim_{K\to\infty}=V$. For large K $\sqrt{N}K(\hat{\beta}-\beta)\xrightarrow{d}N(0,\mathcal{Q}_{XX}^{-1}\mathcal{V}_2\mathcal{Q}_{XX}^{-1})$.

The statement on var(x_{i,T_0}) makes clear that limiting variation is due to a large K and not a large T_{init}. As stated in part (a), $\widetilde{\beta}$ is $\sqrt{N}$ consistent and asymptotically normal for fix K.

[7] To check C.ii, compute sum$(w.^2)$ which should be $O(1/K)$. To check C.ii, compute sum(cumsum$(w.^2)$) which should be $O(K)$.

The convergence rate of the estimator ultimately depends on the rate at which the variance of $\hat{\beta}$ tends to zero. Recall that $\text{var}(\widetilde{e}_{iK}) = \tau_i^2 A_0(w)$ and $\mathbb{E}_i[(X_i - \mu_X)^2] = Q_{XX}$. Further intuition can be gained if we consider homoskedasticity with $\tau_i^2 = \tau$, $\sigma_i^2 = \sigma^2$ for all i. Then $Q_{XX} = \sigma^2 \Phi_K = \sigma^2 \widetilde{a}_K K$, and $V = \tau^2 A_0 Q_{XX}$. Part (i) of the Theorem 2 becomes

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0, \frac{\tau^2}{\sigma^2} \frac{A_0(w)}{\widetilde{a}_K(w) K}\right). \tag{5}$$

The quantity $\frac{\tau^2}{\sigma^2}$ is the noise in the model relative to the signal in the data, while $A_0(w)$ (noise dilution) and $\widetilde{a}_K(w)$ (signal magnification) both depend on the compression scheme $w(t)$. Parts (b) and (c) are large K results for two different classes of $w(t)$. If there is signal magnification but not noise dilution as in point sampling at large T_1, then as indicated in part (b), $\hat{\beta}$ is $\sqrt{NT_1}$ consistent. Part (c) indicates that noise dilution alone can also accelerate convergence even in the absence of signal magnification. If both conditions of Assumption C are satisfied, the estimator is $\sqrt{N}K$ consistent. On the other hand, if $K = 1$, K-averaging and point sampling reduce to the standard cross-section result that $\hat{\beta}$ is $\sqrt{N}$ consistent and asymptotically normal.

Theorem 2 assumes that the cross-section variance of x fans out at a rate of K, but the rate can be slower or faster under alternative assumptions, as discussed above. In the case of an exogenous drift g_i for example, $\bar{X}_{iK} - \mu_X = \frac{K}{2}(g_i - \mathbb{E}_i(g)) + \sum_{s=T_0+1}^{t} u_{is} = O(K)$, $\bar{e}_{iK} = O(K^{-1/2})$. The meat of the sandwich variance is $V_K = O(K)$, and the asymptotic sandwich variance is $O(K^{-4})O(K) = O(K^{-3})$. It follows that $\hat{\beta}$ from the static regression in K-averaged data is $\sqrt{N}K^{3/2}$ consistent and asymptotically normal, faster than the $\sqrt{N}K$ rate when $g_i = 0$.[8] Another example is when $x_{i,T_{init}} = 0$, then with $\mathbb{T}_1 = T_{init} + T_1$, point sampling gives $Q_{XX} = \mathbb{T}_1 \mathbb{E}_i[\sigma_i^2]$ and $V_K \approx \frac{\mathbb{T}_1}{K} \mathbb{E}_i[\sigma_i^2 \tau_i^2]$. Thus, $Q_{XX}^{-1} V_K Q_{XX}^{-1} = O(\mathbb{T}_1^{-1})$, giving $\sqrt{N\mathbb{T}_1}$ consistent and asymptotically normal estimator as $\mathbb{T}_1 \to \infty$ with K fixed. When the fanning out rate is slower, the fastest convergence will not be attainable, but it cannot be slower than $\sqrt{N}$ provided that $\text{cov}_i(\widetilde{X}_{iK}, \widetilde{e}_{iK})$ is of lower order than that of $\text{var}_i(\widetilde{X}_{iK})$.

In spite of the possibility of many convergence rates, knowledge of the rate is not needed for inference, nor is the need for HAC standard errors. The reason is that the scores $(\widetilde{X}_{iK} - \mu_X)\widetilde{e}_{iK}$ are iid across i. Hence a CLT for iid data applies in spite of serial correlation structure of e_{it}, which is already absorbed in V. Asymptotic normality implies that the usual t statistic can be used for inference. It would be a simple Wald test to see if the estimates in non-overlapping samples are equal. Nonetheless, Assumption B.(iii) assumes $u_{it} \sim (0, \sigma^2)$ is serially uncorrelated. With serial correlation, σ_i^2 would be replaced by ω_i^2. In the IMA(1) case, $\omega_i^2 = (1 + \vartheta_i)^2 \sigma_i^2$. If $\vartheta_i = -1 + \frac{\delta_i}{\sqrt{T}}$

[8] The result parallels the convergence rate of $T^{3/2}$ for the coefficient on a linear trend. However, if $\text{var}_i(g_i) = 0$, the only source of fanning out is the stochastic trend and the convergence rate remains $\sqrt{N}K$. If $\mathbb{T}_1 > K$, $\text{var}(\bar{X}_{iK}) = \mathbb{T}_1 \mathbb{E}_i[\sigma_i^2]$.

as in the 'nearly integrated, nearly white noise' case considered in Perron and Ng (1996), then $\omega_i^2 \widetilde{a}_K(w)K = \delta_i^2 \widetilde{a}_K$. With $Q_{XX} = \mathbb{E}_i[\omega_i^2]\widetilde{a}_K K = \mathbb{E}_i[\delta_i^2]\widetilde{a}_K$, there will be no fanning out at all, and a $\sqrt{N}$ consistent estimate is all that can be achieved.

The LD is not in the class of functions considered, but some observations of its estimates can be made. The LD regression error is $\breve{e}_{iK} = \bar{e}_{iM}(T_1) - \bar{e}_{iM}(T_0)$ with variance $\text{var}_i(\breve{e}_{iK}) = 2\tau_i^2/M$, usually larger than $\text{var}_i(\bar{e}_{iK}) = \tau_i^2/K$ associated with K-averaging. Since $\text{var}(\breve{X}_{iK}) \approx \sigma_i^2(K - M/3)$ for LD, it has sandwich variance $\frac{2\tau^2}{M} Q_{XX}^{-1}$ under homoskedasticity. Compared to K-averaging:

$$\frac{\text{var(LD)}}{\text{var(KA)}} = \frac{2K(T_0 + a_K)}{M(K - M/3)} \approx \frac{2(3T_0 + K)}{3M(1 - M/(3K))} \approx \frac{2K}{3M} + \frac{2T_0}{M}.$$

This ratio only depends on the tuning parameters and not on σ_i^2 or ϑ_i. In practical designs when $\frac{K}{M} > \frac{3}{2}$, K averaging should have a smaller variance.

Table 2: Estimates and Rejection Rates

DGP: $y_{it} = \beta_\mu \mu_{it} + \beta_w w_{it} + e_{it}$,
Regression: $\widetilde{Y}_{iK} = \beta \widetilde{X}_{iK} + \widetilde{e}_{iK}$
$(T, N) = (50, 50)$, $(T_0, T_1, T_2) = (10, 20, 40)$ and $k = 9$

β	ρ	ϑ	PS			KA				$LD_K(T_0, T_2, m)$		
			T_1	T_2	T	$(0,T)$	(T_0,T_1)	(T_0,T_2)	(T_1,T_2)	$m=0$	$m=4$	$m=9$
Estimates, T_{init}=0												
-1.00	0.00	-0.00	-0.999	-1.001	-1.000	-1.000	-1.000	-1.000	-1.000	-1.001	-1.000	-1.000
-0.99	0.00	-0.00	-0.989	-0.991	-0.990	-0.990	-0.990	-0.990	-0.990	-0.991	-0.990	-0.990
-0.97	0.00	-0.00	-0.974	-0.976	-0.975	-0.975	-0.975	-0.975	-0.975	-0.976	-0.975	-0.975
-0.95	0.00	-0.00	-0.949	-0.951	-0.950	-0.950	-0.950	-0.950	-0.950	-0.951	-0.950	-0.950
-1.00	0.50	-0.50	-0.999	-1.001	-1.000	-1.000	-1.000	-1.000	-1.000	-1.001	-1.000	-1.000
-0.99	0.50	-0.50	-0.989	-0.991	-0.990	-0.990	-0.990	-0.990	-0.990	-0.991	-0.990	-0.990
-0.97	0.50	-0.50	-0.974	-0.976	-0.975	-0.975	-0.975	-0.975	-0.975	-0.976	-0.975	-0.975
-0.95	0.50	-0.50	-0.949	-0.951	-0.950	-0.950	-0.950	-0.950	-0.950	-0.951	-0.950	-0.950
Rejection Rates of $H_0 : \beta = -1$												
-1.00	0.00	-0.00	0.073	0.075	0.075	0.074	0.071	0.074	0.073	0.072	0.076	0.081
-0.99	0.00	-0.00	0.077	0.076	0.079	0.138	0.082	0.119	0.113	0.071	0.085	0.096
-0.97	0.00	-0.00	0.085	0.091	0.100	0.438	0.137	0.346	0.317	0.079	0.135	0.184
-0.95	0.00	-0.00	0.114	0.147	0.170	0.918	0.321	0.824	0.782	0.097	0.292	0.464
-1.00	0.50	-0.50	0.073	0.075	0.075	0.074	0.072	0.074	0.072	0.072	0.076	0.081
-0.99	0.50	-0.50	0.077	0.076	0.079	0.138	0.082	0.118	0.114	0.071	0.085	0.096
-0.97	0.50	-0.50	0.085	0.091	0.100	0.439	0.137	0.346	0.317	0.079	0.135	0.185
-0.95	0.50	-0.50	0.115	0.147	0.170	0.918	0.321	0.824	0.782	0.097	0.292	0.463

x is generated by the local level model with $\sigma_u = 1$, w_{it} is AR(1) with persistence parameter ρ and $\sigma_w = 0.01$. The two values of ϑ correspond to $\rho = (0, 0.5)$. If $\vartheta = 0$, x is a pure random walk.

To verify the adequacy of the asymptotic results, we consider the finite-sample size and power

of a t-test based on different time compression schemes. We simulate data

$$\begin{aligned} y_{it} &= \beta_\mu \mu_{it} + \beta_w w_{it} + e_{it} \\ x_{it} &= \mu_{it} + w_{it}, \\ w_{is,t} &= \rho w_{i,t-1} + \varepsilon_{it}. \end{aligned}$$

We generate x using a local level model with $\sigma_i^2 = 1.0$, $\sigma_w^2 = 0.01$, $\tau^2 = 9$, $\beta_\mu = -1, \beta_w = 0.01$. The implied ϑ is based on two values of ρ. The results are presented in Table 2. The rejection rates of around 5% when $\beta = -1$ suggest that the normal distribution adequately approximtes the distribution of $\hat{\beta}$. Power is evaluated at close alternatives to the null value of -1. We see from the bottom panel that the power based on K-averaging is much higher than LD. At the alternative of -0.95, for example, K-averaging has a rejection rate of 0.824, compared to a rate of 0.464 for LD.

5 More General Specifications

Theorem 2 is built on Assumptions A, B and C. In general, the 'meat' of the sandwich variance is

$$V = \mathbb{E}_i[(X_i - \mu_X)^2 \widetilde{e}_{iK}^2] = \mathbb{E}_i[(X_i - \mu_X)^2]\mathbb{E}_i[\widetilde{e}_{iK}^2] + \text{cov}_i((\widetilde{X}_{iK} - \mu_X)^2, \widetilde{e}_{iK}^{*^2}). \tag{6}$$

The covariance term in V does not appear if the model is correctly specified. If this is not the case, $\widetilde{e}_{iK}$ will be replaced by $\widetilde{e}_{iK}^*$, and $\text{cov}_i(\widetilde{X}_{iK}, \widetilde{e}_{iK}^*)$ may not be zero. The next three subsections consider cases when $\text{cov}_i(\widetilde{X}_{iK}, \widetilde{e}_{iK}^*) \neq 0$ but super-consistency continues to hold. We then consider a case when $\text{cov}_i(\widetilde{X}_{iK}, \widetilde{e}_{iK}^*) = 0$ and yet the variance of $\widetilde{e}_{iK}^*$ precludes the fastest convergence rate.

5.1 Relaxing Strict Exogeneity

We assume that the DGP is the static model $y_{it} = \beta x_{it} + e_{it}$. Assumption B.i asserts $\text{cov}_i(\widetilde{X}_{iK}, \widetilde{e}_{iK}) = \sum_s \phi_s \sum_t w(t)\mathbb{E}_i[u_{it}e_{is}] = 0$. We now allow $\text{cor}(u_{it}, e_{is}) = \gamma_{i,t-s}$ with $\sum_h |\gamma_{i,h}| \leq \infty$. Contemporaneous correlation occurs when $\gamma_{0,i} = 0$. When $\gamma_{h,i} \neq 0$ for $h < 0$, u_{it} is correlated with future e_{is}, while $h > 0$ allows for feedback from past values of e_{it} to u_{it}.

Now $\widetilde{X}_{iK} = x_{i,T_0} + \sum_s \phi_s u_{is}$ and $\widetilde{e}_{iK} = \sum_t w(t)e_{it}$. Using $\text{cov}(u_{it}, e_{is}) = \gamma_{i,t-s}\sigma_i\tau_i$ for any i,

$$\begin{aligned} \text{cov}(\widetilde{X}_{iK}, \widetilde{e}_{iK}) &= \text{cov}\left(x_{i,T_0} + \sum_{s=T_0+1}^{T_1} \phi_s u_{is}, \sum_{t=T_0+1}^{T_1} w(t)e_{it} \right) \\ &= \sigma_i\tau_i \sum_{s=T_0+1}^{T_1} \sum_{t=T_0+1}^{T_1} \phi_s w(t)\gamma_{i,s-t} \\ &= \sigma_i\tau_i\Lambda_K(w,i) \end{aligned}$$

where $\Lambda_K(w,i) = \sum_{h=-(K-1)}^{K-1} \gamma_{ih} B_h(w)$ and $B_h(w) = \sum_{s-t=h} \phi_s w(t)$. The Appendix shows that $\Lambda_K(w,i) \to \Gamma_i B_\infty(w)$, where $\Gamma_i = \lim_{K\to\infty} \sum_{h=-(K-1)}^{K-1} \gamma_{ih}$ is the long-run cross-covariance, and $B_\infty(w) = \lim_{K\to\infty} B_h(w)$. Then

$$\text{cov}(\widetilde{X}_{iK}, \widetilde{e}_{iK}) \to \sigma_i \tau_i \Gamma_i B_\infty(w).$$

For K averaging where $w(t) = 1/K$ for all t, $B_h(w) \to 1/2$. For point sampling at large T_1 where $w(t) = 1_{t=T_1}$, $B_h(w) \to 1$. With a constant $\widetilde{b}_K$ that depends on the time compression scheme,

$$\mathbb{E}_i[\text{cov}(\bar{X}_{iK}, \bar{e}_{iK})] = \mathbb{E}_i[\sigma_i \tau_i \Gamma_i \widetilde{b}_K] = O(1).$$

Though the average covariance is $O(1)$ due to endogeneity, the cross-section variance is still $O(K)$. In consequence, endogeneity bias to vanishes with K. Allowing $\text{cov}(u_{it}, e_{is}) \neq 0$ only inflates variance, in stark contrast to the I(0) case when the estimators would be inconsistent.

Strict exogeneity may also fail because of feedback. Consider the dynamic model

$$\begin{aligned} x_{it} &= x_{it-1} + \kappa y_{it-1} + u_{it} \\ y_{it} &= \beta_0 x_{it} + \beta_1 x_{it-1} + e_{it}. \end{aligned}$$

Then $x_{it} = (1 + \kappa\beta_0)x_{it-1} + \kappa\beta_1 x_{it-2} + \kappa e_{it-1} + u_{it}$ is still nearly I(1) if $|\kappa(\beta_0 + \beta_1)|$ is small enough to keep $1 + \kappa(\beta_0 + \beta_1)$ in the $1/T$ neighbourhood of unity to preserve the fanning out effect. The effective innovation to x_{it} is now $u^*_{it} = \kappa e_{it-1} + u_{it}$. This implies that $\text{cov}(u^*_{it}, e_{i,t-1}) = \kappa\tau^2$. A static regression of $\widetilde{Y}_{iK}$ on $\widetilde{X}_{iK}$ now has two biases when K is fixed: one from omitting Δx_{iK}, and one from violating of strict exogeneity. But signal magnification makes the biases asymptotically negligible. In both cases when strict exogeneity is violated, we have a $\sqrt{N}K$ estimate of β from a static regression using K averaged data. and a $\sqrt{NT_1}$ consistent estimate using data point sampled at T_1.

5.2 Omitting Stationary Dynamics

In time series regressions, the bias from omitting stationary dynamics may also be of second order when x is I(1). We now show that this is the case in cross-section regressions. Suppose that

$$y_{it} = \alpha y_{it-1} + \beta_0 x_{it} + \beta_1 x_{it-1} + e_{it}. \tag{7}$$

Following Bewley (1979), (7) can be rearranged as

$$\begin{aligned} (1-\alpha)y_{it} &= -\alpha\Delta y_{it} + (\beta_0 + \beta_1)x_{it} - \beta_1 \Delta x_{it-1} + e_{it} \\ y_{it} &= \left(\frac{\beta_0 + \beta_1}{1-\alpha}\right) x_{it} - \left(\frac{\alpha}{1-\alpha}\Delta y_{it} + \frac{\beta_1}{1-\alpha}\Delta x_{it-1}\right) + \frac{e_{it}}{1-\alpha} \\ &= \beta x_{it} + e^*_{it} \end{aligned}$$

where the cointegrating coefficient

$$\beta = \frac{\beta_0 + \beta_1}{1-\alpha} = \lim_{h\to\infty} \frac{\partial y_{it+h}}{\partial x_{i,t+h}} = \frac{\lim_{h\to\infty} \partial y_{,it+h}/\partial u_{it}}{\lim_{h\to\infty} \partial x_{i,t+h}/\partial u_{it}}$$

characterizes the long run relation between y and x. It is often referred to in textbooks as the long-run multiplier.

A static regression of y on x that omits $\Delta y_{it}, \Delta x_{it-1}$ would have $\mathbb{E}_i[e^*_{it} x_{it}] \neq 0$. If x_{it} were stationary, a time series or a panel regression would yield inconsistent estimates. But as x_{it} is I(1), a time series regression would yield a super (T) consistent estimate because the omitted variable bias is of second order, but the distribution of $\hat{\beta}$ is non-standard. Here, the model in time compressed data

$$\bar{Y}_{iK} = \beta \bar{X}_{iK} - \underbrace{\left(\frac{\alpha}{1-\alpha}\Delta y_{iK} + \frac{\beta_1}{1-\alpha}\Delta x_{iK}\right)}_{\eta_{iK}/(1-\alpha)} + \frac{\bar{e}_{iK}}{1-\alpha}$$

where $\Delta y_{iK} = \frac{1}{K}(y_{i,T_1} - y_{i,T_0})$, $\Delta x_{iK} = \frac{1}{K}(x_{i,T_1} - x_{i,T_0})$. Importantly, the omitted η_{iK} is negligible when K is large.

Corollary 1 *Suppose that x_{it} is I(1), y_{it} is generated by the dynamic model given in (7) with $|\alpha| < 1$ and Assumption B holds. Let $\hat{\beta}$ be obtained by estimation of (4) using K-averaged data.*

i As $N \to \infty$ with K fixed, $\sqrt{N}\left(\hat{\beta} - \beta + \frac{cov_i(\bar{X}_{iK}, \eta_{iK})}{var_i(\bar{X}_{iK})}\right) \xrightarrow{d} N(0, Q_{XX}^{-1} W_K Q_{XX}^{-1})$ where $W_K = var_i((\bar{X}_{iK} - \mu_X)(\bar{e}^*_{iK} - \eta_{iK}))$.

ii For large K, $\sqrt{N}K(\hat{\beta} - \beta) \xrightarrow{d} N(0, \mathcal{Q}_{XX}^{-1} \mathcal{W} \mathcal{Q}_{XX}^{-1})$, where $\mathcal{W} = \lim_{K\to\infty} W_K$ and $\mathcal{Q}_{XX} = \lim_{K\to\infty} K^{-1} Q_{XX}$.

In the present setting, the omitted variable bias for a given K is bias $= \frac{\mathrm{cov}_i(\bar{X}_{iK}, \bar{e}^*_{iK} - \eta_{iK})}{\mathrm{var}_i(\bar{X}_{iK})}$. Since $\bar{X}_{iK}$ only depends on $\{u_{it}\}$ and $u_{it} \perp e_{is}$ by assumption, bias can only arise through $\mathrm{cov}_i(\bar{X}_{iK}, \eta_{iK})$. As $\bar{X}_{iK} = O_p(K^{1/2})$ and $\eta_{iK} = O_p(K^{-1/2})$, their product is $O(1)$. But upon scaling by $\mathrm{var}(\bar{X}_{iK})$, the bias is $O_p(K^{-1})$ that vanishes with K. We thus have a $\sqrt{N}K$ consistent estimate of the long-run relation β as K increases. Though $\bar{e}^*_{iK}$ does not contribute to the bias, $\mathrm{var}(\bar{e}^*_{iK}) > \mathrm{var}(\bar{e}_{iK})$. Omitting the lags yields a less efficient estimate of β as reflected in W_K.

It should be noted that β is distinct from long-run effect of a shock to x, defined as

$$\lim_{h\to\infty} \frac{\partial y_{it+h}}{\partial u_{it}} = \beta \times \lim_{h\to\infty} \frac{\partial x_{it+h}}{\partial u_{it}}.$$

If $\beta \neq 0$, the two coincide only if x_{it} is a random walk. In general, the long run effect depends on the dynamics of x_{it}. For example, in the IMA(1) case when $\Delta x_{it} = u_{it} + \vartheta_i u_{it-1}$, the long-run effect of u_{it} is $\beta(1+\vartheta_i) \neq \beta$. Consistent estimation of the long-run relation β is not enough to characterize

the long run effect. The same caveat applies to LD estimation. The LD transformation isolates variations with 'medium periodicities, but this data will not estimate the medium term effect after a shock, which also requires the dynamics of x_{it}, and thus require estimating the system $B_0 z_{it} = B_1 z_{i,t-1} + \epsilon_{it}$. Comparison of regression results with impulse response system analysis such as in Bilal and Kanzig (2024) should be made with caution.

5.3 Heterogeneous Coefficients

The above analysis shows that the strong cross section variability generated by I(1) data usually accelerates the convergence rate of $\hat{\beta}$. We now consider a counter example when the strong signal does not accelerate convergence.

Consider heterogeneous coefficient $\beta_i = \beta + v_i$ where $\mathbb{E}_i[v_i] = 0$, and $\text{var}_i(v_i) = \sigma_v^2$. The model $y_{it} = \beta_i x_{it} + e_{it}$ implies $y_{it} = x_{it}\beta + e_{it}^*$ where $e_{it}^* = v_i x_{it} + e_{it}$ so that $\text{cov}(x_{it}, e_{is}^*) = \mathbb{E}[v_i x_{it} x_{is}] \neq 0$. With $\bar{e}_{iK}^* = \bar{e}_{iK} + v_i \bar{X}_{iK}$, a static regression in K-averaged data yields

$$\hat{\beta} = \beta + \frac{\text{cov}_i(v_i, \bar{X}_{iK})}{\text{var}_i(\bar{X}_{iK})} = \beta + \frac{\mathbb{E}_i[v_i(\bar{X}_{iK} - \mu_X)^2]}{\text{var}_i(\bar{X}_{iK})}.$$

Suppose that $v_i \perp (\sigma_i^2, x_{i,T_0})$, then because $\mathbb{E}_i[v_i] = 0$,

$$\mathbb{E}_i[v_i(\bar{X}_{iK} - \mu_X)^2] = \mathbb{E}_i[v_i \bar{X}_{iK}^2] + \mathbb{E}_i[v_i(x_{i,T_0} - \mu_X)^2] = 0.$$

The orthogonality condition is satisfied, but the estimator is only $\sqrt{N}$ consistent. This result is best understood in the homoskedastic case when $\sigma^2 = \mathbb{E}_i[\sigma_i^2]$ and $\tau^2 = \mathbb{E}_i[\tau_i^2]$. Then $\text{var}(\bar{e}_{iK}^*) = \sigma_v^2 \mathbb{E}_i[\bar{X}_{iK}^2] + \text{var}(\bar{e}_{iK}) = O(K)$. K-averaging does not achieve noise dilution because parameter heterogeneity scales $\bar{e}_{iK}^*$ with $\bar{X}_{iK}$, and any signal magnified by time compression is offset by a magnified noise. Thus, even if x is a random walk with drift, the rate will remain $\sqrt{N}$. Pesaran and Smith (1995) considered a heterogeneous coefficient model where x is I(1) with a heterogeneous drift. They find that the coefficient estimate from a cross-section regression in full-sample averaged data is consistent. We show that the convergence rate is $\sqrt{N}$, whether or not there is a drift. It should be remarked that point sampling at T_1 will also be $\sqrt{N}$ consistent because it also has the problem that $\widetilde{e}_{i,T_1} = e_{i,T_1} + v_i X_{i,T_1}$ scales with X_{i,T_1}. The practical implication is that if a researcher wants to allow for heterogeneous coefficients, the usual rate of $\sqrt{N}$ should be expected.

6 Empirical Issues

We have so far assumed that x_{it} only has one component that is a stochastic trend. In applications of empirical interest, x_{it} often has additional sources of variation. Subsection 1 considers fixed effects and cross-section depenence. Subsection 2 considers the case when the two components are observed. Subsection 3 considers coexistence of latent permanent and transitory components.

6.1 Fixed Effects and Cross-Section Dependence

A common critique of cross-section regressions is the bias from omitted time-invariant variables and cross-section dependence.[9] To investigate these issues in our setup, suppose that

$$y_{it} = \beta x_{it} + s_i f_t + e_{it}$$

where x_{it} is I(1), s_i is a unit-specific effect, f_t is stationary time variation common across i. This interactive fixed effect specification considered in Bai (2009) nests the additive fixed effect as special case. To make progress, let $f_t = \mu_f + f_t^0$ and $s_i = \mathbb{E}_i[s_i] + s_i^0$, where $\mu_f = \mathbb{E}[f_t]$ with $\mathbb{E}[f_t^0] = \mathbb{E}_i[s_i^0] = 0$. Because f_t^0 is stationary,

$$s_i \bar{f}_K = \underbrace{\mathbb{E}_i[s_i]\mu_f}_{\text{constant}} + s_i^0 \mu_f + \underbrace{s_i \bar{f}_K^0}_{O(K^{-1/2})}.$$

In this set up, the regression model in K-averaged data is

$$\begin{aligned} \bar{Y}_{iK} &= \beta \bar{X}_{iK} + s_i \bar{f}_K + \bar{e}_{iK} \\ &= c + \beta \bar{X}_{iK} + s_i^0 \mu_f + \bar{e}_{iK}^* \end{aligned}$$

where c absorbs the constant term in $s_i \bar{f}_K$ and $\bar{e}_{iK}^* = \bar{e}_{iK} + s_i \bar{f}_K^0$. The omitted variable bias is

$$\text{bias} = \frac{\text{cov}_i(\bar{X}_{iK}, s_i^0 \mu_f)}{\text{var}_i(\bar{X}_{iK})} = \frac{\mu_f \text{cov}_i(x_{i,T_0}, s_i^0)}{\text{var}_i(\bar{X}_{iK})}.$$

There is trivially no bias when $\mu_f = 0$ and $\hat{\beta}$ continues to be $\sqrt{N}K$ consistent.

More interesting is when $\mu_f \neq 0$, the important special case being $f_t = \mu_f = 1$ which makes s_i an individual fixed effect. If $x_{i0} = 0$, strict exogeneity ensures that $\text{cov}_i(x_{i,T_0}, s_i) = 0$ as is required for consistent estimation of β. The question is at what rate. Now $\bar{e}_{iK}^* = \bar{e}_{iK} + s_i = O(1)$ because s_i cannot be diluted away by averaging, and s_i will inflate V in the sandwich variance from $O(1)$ to $O(K)$,[10] making $\hat{\beta}$ $\sqrt{NK}$ (instead of $\sqrt{N}K$) consistent. However, the standard errors are valid if s_i^0 is independent across i. While the convergence rate using K-averaged data is downgraded upon loosing the benefit of noise dilution, this is not the case if we point sample at T_1 because it does not have the noise dilution property anyway. With $\widetilde{e}_{i,T_1} = e_{i,T_1} + s_i$ and $\text{var}_i(e_{i,T_1}^*) = O(1)$, the order is the same as $\text{var}(e_{i,T_1})$ when s_i was absent, and the convergence rate remains $\sqrt{NT_1}$. The LD treats differences away the fixed effect to identify β from the change rather than level of the non-stationary component, and the convergence rate should be $\sqrt{N}K$, a distinct advantage over K-averaging..

[9] See, for example, Deschenes and Greenstone (2007) and Mendelsohn and Masetti (2017) and Andrews (2005).

[10] The situation is different from Andrews (2005). In Andrew's case, common shocks create cross-section dependence and invalidates the usual standard errors for the $\sqrt{N}$ consistent estimate.

The above arguments break down if s_i is correlated with x_{it}, as strict exogeneity only rules out correlation between u_{it} and e_{is} to give a finite sample bias of $O(K^{-1})$. However, this bias can be mitigated in different ways. First, if we have a variable z_i correlated with s_i, we can use it to control for s_i. A second possibility is to note that the numerator of the bias $\mu_f \text{cov}(x_{i,T_0}, s_i^0) = O(1)$ does not grow with K or T_0, implying that the stronger is signal magnification, the faster is convergence. For example, under the $T_{init} = 0$ set up, $\text{var}_i(x_{i,T_0}) = O(\mathbb{T}_0)$ where $\mathbb{T} = T_{init} + T_0$, $\hat{\beta}$ is $\sqrt{N\mathbb{T}}$ consistent. Similarly, an exogenous drift will also give an estimator with fast enough convergence to reduce the bias from the omitted fixed effect. With K-averaged data, omitted fixed effects andy cross-section dependence impacts the convergence rate of the estimator which depends on the underlying profile of the cross-section variance, but it is at least $\sqrt{N}$ consistent if the model is otherwise well specified..

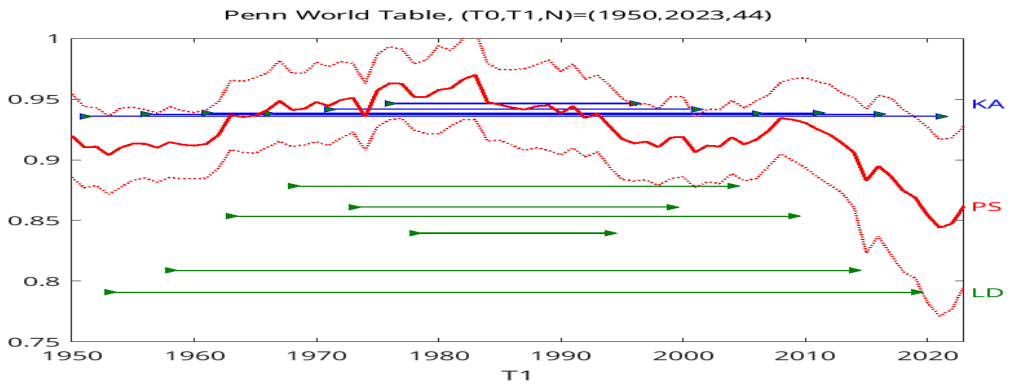
Consider a regression of log consumption on log real GDP, both in per capita terms, using data for N=47 countries from the Penn World Table from 1951 to 2023 for T=73 periods. We do not know the true value, hence as benchmark, we consider three panel estimators for non-stationary data with cross-section dependence, ie. $y_{it} = x_{it}'\beta + \lambda_i F_t + e_{it}$ where λ_i and F_t are unobserved. The first is the interactive fixed-effect estimator of Bai and Kao (2006). The point estimate of 0.931 is obtained by estimating two factors once but does not adjust for long-run covariance and endogeneity from possible correlation between F_t and x_{it}. The second is the continuously updated, bias-corrected cupBC estimator of Bai et al. (2009). The estimate of .933 removes asymptotic bias caused by unit roots, serial correlation, and endogeneity.[11] The third is the LSDV estimate of 0.926. It the is pooled least squares estimation in demeaned data which is consistent if x and F are strictly exogenous and has an asymptotic bias otherwise. By way of comparison, the K-averaging estimate over the full sample is 0.934, almost identical to cupBC. The point sampled estimate at T is 0.861, and the LD estimate with $(T_0, T_1, m) = (m+1, T-m, 2)$ is 0.787.

Under the assumptions of the analysis, $\hat{\beta}$ from K-averaged data should be fairly stable with respect to K. The point-sampled estimates should display more variation because there is no error-dilution, as should the LD estimates because differencing inflates the error variance. Figure 2 plots the estimates $(T_0 = a, T_1 = T - a)$ pairs for $a = 2 : 5 : 30$ along with their standard errors. We see that thr K-averaging estimates (in blue) do not vary much with $K = T_1 - T_0$, but the point-sampled (in red) and in particular the LD estimates (in green) are more sensitive to K. Note that these

[11] The two other estimators for non-stationary variables with cross-section dependence are the CEE-ARDL estimator in Chudik and Pesaran (2015) and the CEE-DL in Chudik et al. (2016). The former requires the factors to be serially uncorrelated and uncorrelated with the regressors. As discussed in Chudik et al. (2016), the latter the errors will be correlated with the regressors when there is feedback, and is sensitive to the stationarity assumption. However, both use cross-section average in observables to approximate the factors and bypass the need of iteratively estimating the factors and the parameters.

regressions use UM=1-emp/pop to control for stationary cross-section dependence, where emp and pop are employment and population. In other words, $\bar{f}_K$ is approximated by the K period average of UM. Without UM as control, the full sample K-averaged estimate is 0.903 and are more sensitive to the choice of T_0 and T_1. In practice, the sensitivity of the estimates to K appears to be a useful diagnostic of the adequacy of the model.

Figure 2: Regression of Log Consumption on Log GDP



The blue (K-averaged) and green (LD) lines are estimates for six choices of (T_0, T_1) pairs. The red line is the estimate from point sampling at every t, and its one-standard error confidence interval.

To be confident of the estimates, a calibrated monte-carlo exercise with 5000 replications is conducted by generating data using the cupBC estimate of 0.933, with the two estimated factors and the corresponding loadings held fixed. An AR(1) model is estimated for each of the demeaned idiosyncratic residuals, and synthetic data are generated by simulating errors for the AR(1) model. The exercise shows that all estimators have some bias, but it is particularly large for LD. Furthermore, the mean estimates are consistent with the findings in the application.

Results from Monte Carlo Exercise with 5000 replications

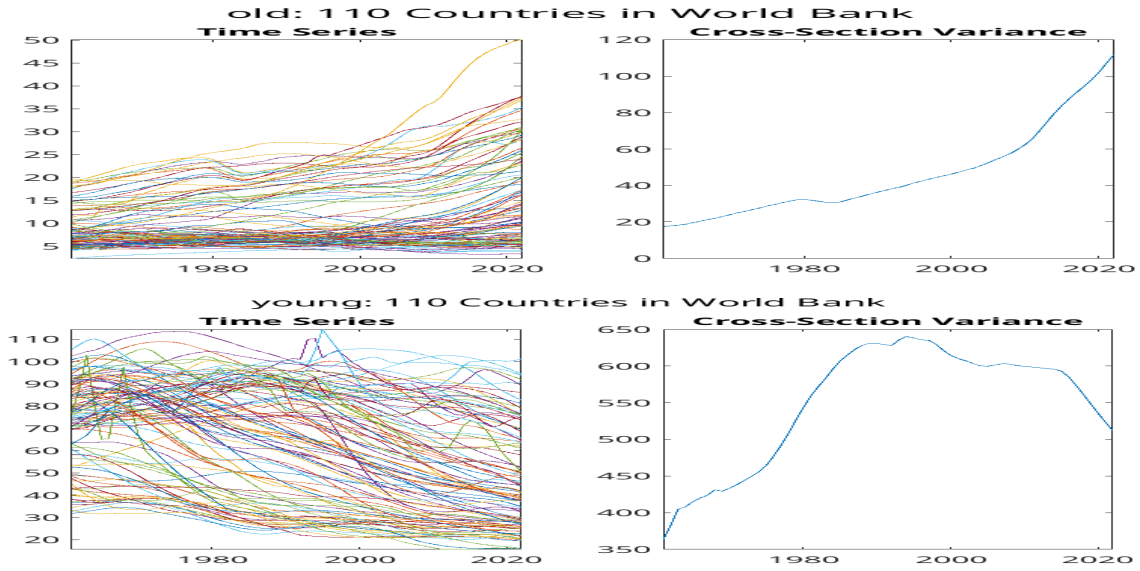
	cupBC	LSDV	Bai	PS	K-avg	LD
mean	0.962	0.930	0.961	0.883	0.936	0.809
median	0.962	0.930	0.961	0.883	0.936	0.809
sd	0.007	0.006	0.007	0.014	0.005	0.020
bias	0.012	-0.020	0.011	-0.067	-0.014	-0.140
rmse	0.007	0.006	0.007	0.014	0.005	0.020

6.2 Fanning-in vs Fanning-out

Consider the dependency ratio which is the sum of the population under 16 ('young') and over 65 ('old') relative to the working age population. Using data from the 110 countries in the World

Bank database from 1960-2022, the top panel of Figure 3 shows that 'old' is non-stationary; the time series data have been steadily increasing, while its cross-section variance fans out. The bottom panel shows that while the time series data of the 'young' are also non-stationary, the cross-section variance fans out until the turn of this century. The dependency ratio and population growth exhibit a similar feature as 'young'. A possible explanation is the global decline in fertility which creates convergence across units. The cross-section variance of the dependency ratio thus has a fanning-out component (due to the 'old') and a fanning-in (due to the 'young') component.

Figure 3: Share of Old and Young in Working Population



A high dependency ratio has been suggested as a possible explanation for higher inflation and a drag on growth.[12] The mixed evidence could be due to differences in the choice of the sample, as the relation between economic and demographic variables could be different for countries in different stages of economic development. We use a sub-sample of 35 countries with average real per-capita GDP in the past 10 years over 10,000 USD, and with maximum inflation in the sample under 100 percent. In this data, the cross-section variance of the 'young' fans-in over the entire sample, distinctly different from the fanning out cross-section variance of 'old'. If we use the dependency ratio as regressor, we find that the estimate is positive in the inflation regression and significant but varies significantly as K increases. The coefficient in the growth regression is more stable but is not significant. Because the dependency ratio has two components with different properties, we

[12]See, for example, Juselius and Takats (2021) for inflation. Gordon (2016) argues that an aging population creates a structural imbalance between high savings and low investment. A different view is gien in Acemoglu and Restrepo (2017).

Table 3: Growth/Inflation Regressions on Demographic Variables: K-averaging

T_0	T_1	inflation				growth			
		$\hat{\beta}_{old}$	t_{old}	$\hat{\beta}_{young}$	t_{young}	$\hat{\beta}_{old}$	t_{old}	$\hat{\beta}_{young}$	t_{young}
1963	2020	0.289	1.495	0.406	1.508	-0.154	-2.338	-0.048	-1.542
1968	2015	0.302	1.548	0.465	1.558	-0.172	-2.794	-0.048	-1.653
1973	2010	0.333	1.619	0.545	1.574	-0.176	-2.885	-0.055	-1.928
1978	2005	0.354	1.740	0.623	1.597	-0.164	-2.512	-0.062	-1.939
1983	2000	0.281	1.693	0.603	1.480	-0.152	-1.849	-0.065	-1.665

explore their distinctive economic impact.

Table 3 reports results from regressions of inflation (left) and growth (right) on the demographic variables for $(T_0 = a, T_1 = T - a)$ pairs with a=2:5:25. For inflation, the full sample estimate of the coefficient on 'old' is 0.289, and the coefficient on 'young' is 0.406, but neither is statistically significant at the 10% level. For the growth regression, the full sample estimate on 'old is -0.154 with a t statistic of -2.338 and all subsample samples are similar. The coefficient on 'young' is -0.048 and the subsample estimates are similar, but the estimates are only marginally significant.[13] The analysis suggests that the effect of the dependent population on inflation is positive but not statistically well determined, but there is a negative effect of both 'old' and 'young' on growth, though the estimate on 'old' is better determined. A possible explanation is that the data for 'old' has more variability because its cross-section variance fans out. When the cross-section variability of the two underlying components differ, using two covariates instead of the sum can be informative.

6.3 Permanent and Transitory Components

The dependency ratio is a sum of two components that we observe. Suppose now $x_{it} = \mu_{it} + w_{it}$ is a sum of two latent components where $\mu_{it} = \mu_{it-1} + u_{it}$ is non-stationary and w_{it} is a mean-reverting. If μ_{it} and w_{it} have the same effect on y_{it}, the presence of w_{it} would not change the analysis because $\Delta x_{it} = u_{it}$ already allows for the possibility that u_{it} can be serially correlated. But suppose that they have distinctive impact of y. For example, a warming temperature trend μ_{it} may affect economic outcome y_{it} differently from regular temperature fluctuations w_{it}. Thus consider

$$y_{it} = \beta_\mu \mu_{it} + \beta_w w_{it} + e_{it}.$$

[13] For the inflation regression, the LD estimate for the longest sample is -0.194 with a t statistic of -2.219, and estimates of the shorter subsamples are not statistically significant. The LD longest sample estimate of 'old' in the growth regression -0.184 with a t statistic of -4.421, but the subsample estimates are more variable, ranging from -0.051 to -0.451.

As $\text{cov}_(x, y) = \beta_\mu \text{var}(\mu) + \beta_w \text{var}(w)$, the estimand, β, in a regression of y on x will be a variance weighted average of β_μ and β_w. It is as if the regression error $e^*_{it} = (\beta_w - \beta_\mu) w_{it} + e_{it}$ has a measurement noise. The problem is analogous to the inability to identify the effect of permanent income on consumption from income data. With point-sampling at T_1, the weight on β_μ would depend exclusively on the cross-section volatility of $w_{i,t}$ at $t = T_1$. Here, a regression in K-averaged data helps because the estimand is

$$\beta_x = \frac{\text{cov}_i(\bar{Y}_{iK}, \bar{X}_{iK})}{\text{var}_i(\bar{X}_{iK})} = \frac{\beta_\mu \text{var}_i(\bar{\mu}_{iK}) + \beta_w \text{var}_i(\bar{w}_{iK})}{\text{var}_i(\bar{\mu}_{iK}) + \text{var}_i(\bar{w}_{iK})}.$$

In terms of regression noise, $\bar{e}^*_{iK} = (\beta_w + \beta_\mu)\bar{w}_{iK} + \bar{e}_{iK} = O(K^{-1})$, and in terms of signal, $\text{var}_i(\bar{\mu}_{iK}) = O(K)$. Thus, $\beta_x \to \beta_\mu$ as $K \to \infty$. A larger K better isolates β_μ by increasing the permanent-to-transitory ratio $\sigma^2_\mu / \sigma^2_w$. In applications when K is not large enough, OLS will estimate a weighted average of β_μ and β_w.

Table 4: Local-Level Model

DGP: $y_{it} = \beta_\mu \mu_{it} + \beta_w w_{it} + e_{it}$, $(\beta_\mu, \beta_w) = (-1, -2.5)$,
Cross-section regression: $\widetilde{Y}_{iK} = \beta_x \widetilde{X}_{iK} + \widetilde{e}_{iK}$,
$(T, N) = (50, 50)$, $(T_0, T_1, T_2) = (10, 20, 40)$

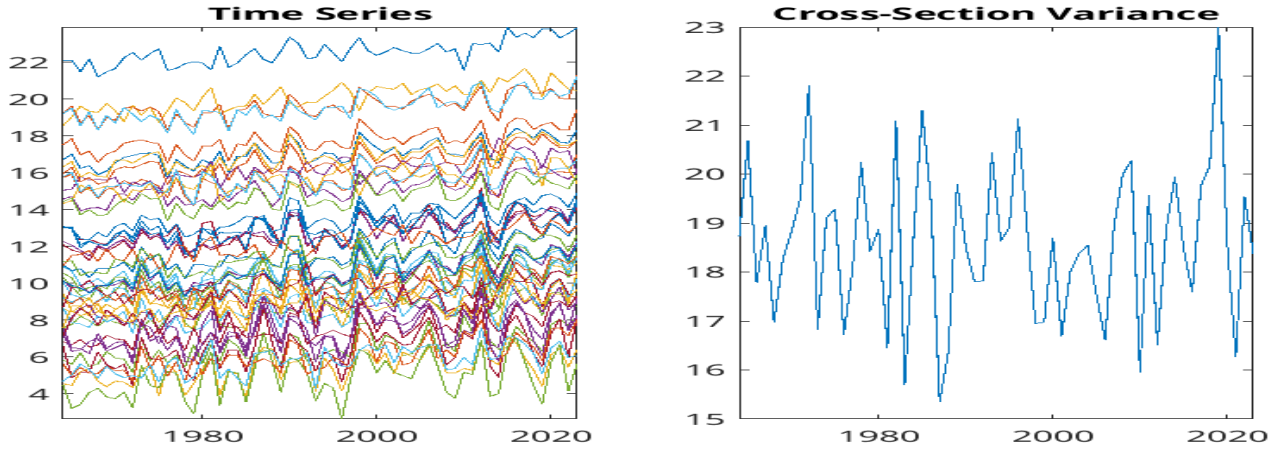
σ_u	ρ	ϑ	Oracle		PS			KA				$LD_K(T_0, T_2, m)$		
			$\hat{\beta}^*_\mu$	$\hat{\beta}^*_w$	T_1	T_2	T	$(1, T)$	(T_0, T_1)	(T_0, T_2)	(T_1, T_2)	$m = 0$	$m = 4$	$m = 9$
			Estimates of β_x											
4.00	0.00	-0.06	-1.000	-2.506	-1.004	-1.002	-1.002	-1.000	-1.001	-1.000	-1.000	-1.006	-1.001	-1.000
1.00	0.00	-0.38	-1.000	-2.502	-1.071	-1.037	-1.030	-1.002	-1.012	-1.003	-1.003	-1.093	-1.012	-1.006
0.50	0.50	-0.79	-1.001	-2.502	-1.315	-1.177	-1.145	-1.027	-1.139	-1.038	-1.041	-1.393	-1.152	-1.092
0.01	0.50	-1.00	-1.104	-2.501	-2.498	-2.495	-2.491	-2.491	-2.498	-2.494	-2.494	-2.491	-2.501	-2.501
			S.E.											
4.00	0.00	-0.06	0.007	0.250	0.027	0.019	0.017	0.004	0.010	0.005	0.005	0.031	0.011	0.008
1.00	0.00	-0.38	0.028	0.086	0.106	0.076	0.068	0.017	0.041	0.020	0.021	0.121	0.044	0.033
0.50	0.50	-0.79	0.079	0.070	0.195	0.149	0.134	0.041	0.094	0.048	0.051	0.217	0.100	0.077
0.01	0.50	-1.00	7.909	0.055	0.377	0.381	0.380	0.222	0.234	0.222	0.227	0.380	0.236	0.226

To illustrate this point, we conduct a monte carlo exercise with 10,000 draws of data from the local-level model for different values of σ_u and ρ with $(\beta_\mu, \beta_w) = (-1.0, -2.5)$. The estimates of β_x and corresponding standard errors are reported in Table 4. A panel regression using μ_{it} and w_{it} as covariates is infeasible because these variables are not observed by the econometrician, but they are observed in simulations. We label these estimates as ORACLE. Since w is stationary, a non-degenerate long-run relation can only exist between y and μ. Table 4 shows that $\hat{\beta}_x$ is often close to β_μ when ϑ is away from -1, but is contaminated by β_w when ϑ is close to -1. K-averaging is more capable of producing estimates $\hat{\beta}_x$ that are closer to β_μ than LD or point sampling when $\vartheta \neq -1$. When $\vartheta = -1$ and $\mathbb{E}_i[(1 + \vartheta)^2] = 0$, even K-averaging will not recover β_μ because the

cross-section variance no longer fans out.

The problem can be relevant to regressors such as temperature which can be characterized by $x_{it} = \mu_{it} + w_{it}$, where μ_{it} is climate and w_{it} is regular temperature fluctuations. Figure 4 plots the temperature data (in Celsius) for the 50 states in the U.S.[14] There is a slight upward trend in the time series, and yet the cross-section variance does not show much fanning out. Whether this is due to a small random walk component as in the local-level model, or a small common trend with relatively homogeneous factor loadings as in the case of the factor model, the implication in terms of the IMA(1) model is that the average long-run variance $\mathbb{E}_i[\omega_i^2]$ is small. When the fanning out effect is limited, regressions in time compressed I(1) data will not benefit from signal magnification, and omitted serial correlation and fixed effect will no longer be of second order importance.

Figure 4: Temperature in Celsius for 48 States in the U.S.



To illustrate, we use output growth as economic outcome, ie. $y_{it} = 100\Delta \log \text{GSP}_{it}$ where GSP is real state product per capita. We first use K-averaged data to estimate the linear model $\bar{Y}_{iK} = c + \beta_1 \bar{X}_{iK} +$ error with $T_0 = 0, T_1 = t$ for $t = 1, \ldots, T$. As noted in the consumption example, absent omitted variables, the estimates should be stable with respect to K. We first estimate a linear model and find that the estimates on temperature decline as K increases. Adding latitude of the state capital to control for omitted fixed effects helps stablize the estimates, suggesting that there omitted fixed effects might be relevant in this empirical setting.

Next, we consider possible non-linear effects by adding a quadratic term:

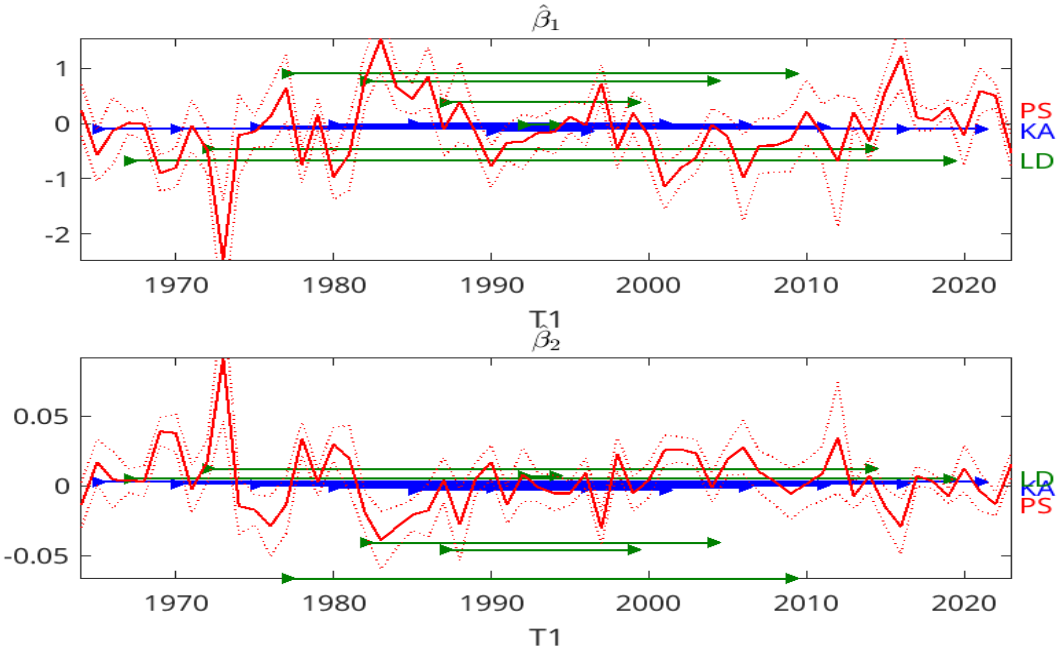
$$\bar{Y}_{iK} = b_0 + \beta_1 \bar{X}_{iK} + \beta_2 \bar{X}_{iK}^2 + \text{latitude}_i + \text{ error}.$$

[14]Data Source: The U.S. data are based on USGS Geographic Names Information System (GNIS), utilizing the North American Datum of 1983 (NAD 83) aligned with standard WGS 84 latitude-longitude configurations. Country codes via ISO 3166-1 alpha-3; Capital city coordinates via standard World Geodetic System (WGS 84) data listings.

If $x_{it} = x_{i0} + \sum_{s=1}^{t} u_{is}$ is a random walk, then $\mathbb{E}[x_{it}^2] = \mathbb{E}[x_{i0}^2] + t\sigma_i^2$ but x_{it}^2 is not I(2) because it remains non-stationary after differencing.[15] Inference would be difficult for time series regressions. But in this data, $\text{var}_i(x_{it}^2)$ clearly fans out even when though $\text{var}_i(x_{it})$ does not. Standard normal inference should be possible without knowing the order of integration, which is particularly appealing. Note, however, $\bar{X}_{iK}^2$ is the average of x_{it}^2, not $(\bar{X}_{ia})^2$ as in Burke and Emerick (2016).

The regression is estimated for different $(T_0 = a, T_1 = T_1 - a)$ pairs with a=2:5:30. Figure 5 shows that the K-averaging estimates (in blue) of β_1 are negative while the β_2 estimates are positive. However, the t statistics are -1.301 and 1.236, respectively. Evidence for a significant effect from the sub-sample estimates are even weaker. In this application, the LD estimates (in green) range from positive to negative, but they are also not well determined. Thus, even with methodologies that target the non-stationary variations, the estimate of β_μ is not well determined. The point sampled estimates are particularly informative in this case as they do not benefit from noise dilution. Figure 5 finds significant fluctuations in these estimates (in red) over time, suggesting that it is the stationary component that dominates in the data, consistent with the lack of fanning out in the cross-section variance. The estimate can well be similar to that of β_w if the estimation methodology is ineffective in isolating such variations.

Figure 5: Regressions of Growth on Temperature: 48 States



[15] In particular, $\text{var}(\Delta x_{it}^2) = 4x_{i,t-1}^2\sigma_i^2 + 2\sigma_i^4 = O(t)$. Ng (2008) refers to x_{it}^2 as a heteroskedastic random walk; see also Berenguer-Rico and Gonzalo (2014).

7 Conclusion and Discussion

This paper considers static regressions using time-compressed data. We first show that the cross-section variance of a class of time compressed data will fan out if the underlying data are non-stationary, and the magnified signals can be exploited in cross-section regressions of the form $y = x\beta + e$. We show conditions under which time compressed data can yield (super-)consistent and asymptotically normal estimates. Simple time averaging is particularly appealing because it amplifies the signal in the I(1) regressor and dampens the stationary regression noise. It also makes endogeneity and misspecification bias of second order. We investigate the long-run relation between consumption and income, between inflation/growth and demographic variables, and between temperature and growth. While the first two examples find estimates that are stable with respect to K, the last example is challenging because temperature data are dominated by short-term fluctuations, making it difficult even for time compression schemes to isolate the trending variations.

A-1 Appendix

In what follows, we will use $\mathbb{E}_i[\,\cdot\,]$ and $\text{Var}_i(\,\cdot\,)$ to denote expectation and variance over the cross-sectional distribution of units. To simplify notation, we write $\mathbb{E}[\cdot|i]$ as $\mathbb{E}[\cdot]$.

Since $\mathbb{E}[\int_0^1 W(r)dr]^2 = 1/3$, $\text{var}(T^{-1/2}\bar{X}) = \omega_v^2/3$ when x_t is I(1).

A-2 Proof of Lemma 1

If x_t is stationary, then $\sqrt{T}\bar{X} \xrightarrow{d} N(0, \omega_x^2)$ where $\omega_x^2 = 2\pi \sum_{j=-\infty}^{\infty} \gamma(j))$ is the long-run variance of x. If $\Delta x_t = v_t$, $v_t \sim (0, \sigma^2)$ and is serially correlated with long run variance ω_v^2, then $\frac{\bar{X}}{\sqrt{T}} = T^{-3/2} \sum_{t=1}^{T} x_t \Rightarrow \omega_v^2 \int_0^1 W(r)dr = A$. To understand the finite sample analog of this asymptotic analysis in a general setting, we consider averaging $K = T_1 - T_0$ observations.

From $x_{it} = x_{i,t-1} + u_t$, $x_{it} = x_{i,T_0} + \sum_{s=T_0+1}^{t} u_{is}$, we have $\bar{X}_{i,K} = \frac{1}{K}\sum_{t=T_0+1}^{T_1} x_{it} = x_{i,T_0} + \frac{1}{K}\sum_{t=T_0+1}^{T_1}\sum_{s=T_0+1}^{t} u_{is}$. Since u_{is} in the double sum for $t = s$ to $s = T_1$ appears $T_1 - s + 1$ times, and $T_1 - s + 1 = K - (s - T_0) + 1$, we can let $j = T_1 - s + 1$ so that $j = 1$ when $s = T_1$ and $j = K$ when $s = T_0 + 1$ to obtain

$$\begin{aligned}
\bar{X}_{iK} &= x_{i,T_0} + \sum_{s=T_0+1}^{T_1} \frac{T_1 - s + 1}{K} u_{is} = x_{i,T_0} + \sum_{j=1}^{K} \frac{(K - j + 1)}{K} u_{i,T_0+j} \\
&= x_{i,T_0} + \sum_{j=1}^{K} \phi_j u_{i,T_0+j}, \qquad \text{where} \quad \phi_j = \frac{K - j + 1}{K}.
\end{aligned}$$

Note that $\mu_X = \mathbb{E}_i[\bar{X}_{iK}] = \mathbb{E}_i[x_{i,T_0}]$, because u_{it} have mean zero. Furthermore, $\sum_{j=1}^{K} j^2 = \frac{K(K+1)(2K+1)}{6}$. We thus have $\text{var}(\bar{X}_{iK}) = T_0\omega_i^2 + a_K\omega_i^2$ where $a_K = \frac{(K+1)(2K+1)}{6}$. Therefore, $\text{var}_i(\bar{X}_{iK}) = \text{var}_i(x_{i,T_0}) + a_K\mathbb{E}_i[\omega_i^2]$, where $\text{var}_i(x_{i,T_0}) = \mathbb{E}_i[(x_{i,T_0} - \mu_X)^2] = T_0\mathbb{E}_i[\omega_i^2]$.

A-3 Proof of Theorem 2

We proof Theorem 1 for the case of K averaging. The OLS estimator satisfies:

$$\hat{\beta} - \beta = \frac{\frac{1}{N}\sum_i (\bar{X}_{iK} - \bar{\bar{X}})\,\bar{e}_{iK}}{\frac{1}{N}\sum_i (\bar{X}_{iK} - \bar{\bar{X}})^2}.$$

Part (i): K fixed By LLN, the denominator converges to the fixed constant:

$$Q_{XX} \equiv \text{var}_i(\bar{X}_{iK}) = a_K\,\mathbb{E}_i[\sigma_i^2] + \mathbb{E}_i[var(x_{i,T_0})].$$

Since $E[u_{is}e_{it}] = 0$ for all t, s by assumption, a CLT applied to $\frac{1}{\sqrt{N}}\sum_i (\bar{X}_{iK} - \bar{\bar{X}})\bar{e}_{iK}$ gives a mean zero process with variance $V_K = \mathbb{E}_i[(\bar{X}_{iK} - \mu_X)^2\bar{e}_{iK}]$. By law of iterated expectations,

$\mathbb{E}_i[(\bar{X}_{iK}-\mu_X)^2\bar{e}_{iK}^2] = \mathbb{E}_i\Big[(\bar{X}_{iK}-\mu_X)^2\mathbb{E}_i[\bar{e}_{iK}^2|\tau_i^2,\bar{X}_{iK}]\Big] = \mathbb{E}_i[(\bar{X}_{iK}-\mu_X)^2\tau_i^2/K$. Expanding $(\bar{X}_{iK}-\mu_X)^2 = (\sum_1 j\phi_j x_{i,T_0+j} + (x_{i,T_0}-\mu_X))^2$ and using $\mathbb{E}[u_{is}]=0$,

$$V_K = \mathbb{E}_i[(\bar{X}_{iK}-\mu_X)^2\bar{e}_{iK}] \;\; = \;\; \frac{1}{K}\Big[a_K\mathbb{E}_i[\sigma_i^2\tau_i^2] + \mathbb{E}_i[(x_{i,T_0}-\mu_X)^2\tau_i^2\Big]$$

Combining the results for the numerator and denominator gives the variance stated in the Proposition. We keep σ_i^2 and τ_i^2 inside a joint cross-section expectation since they may be correlated across units. The same applies to x_{i,T_0} and τ_i^2.

Part (ii): $K\to\infty$ With large K, $a_K = (K+1)(2K+1)/(6K) \approx K/3$ and term relating to the initial condition becomes negligible. Now $Q_{XX}^{-2}V_K = O(K^{-2})$. Multiplying the estimator by K $\sqrt{N}K(\hat{\beta}-\beta)\xrightarrow{d}N(0,\mathcal{Q}_{XX}^{-1}\mathcal{V}\mathcal{Q}_{XX}^{-1})$. For this model, we obtain a non-degenerate limit, $\mathcal{V} = \mathbb{E}_i[\sigma_i^2\tau_i^2]/3$ and $\mathcal{Q}_{XX} = \mathbb{E}_i[\sigma_i^2]/3$, giving an asymptotic variance of $3\mathbb{E}_i[\sigma_i^2\tau_i^2]/(\mathbb{E}_i[\sigma_i^2])^2$.

A-4 Proof of Corollary 1:

Now the DGP is the ADL(1,1) model. We use the Bewley transformation to obtain

$$y_{it} = \beta x_{it} - \alpha^*\Delta y_{it} - \beta^*\Delta x_{it} + + e_{it}^*,$$

where $\beta = \frac{(\beta_0+\beta_1)}{1-\alpha}$, $\alpha^* = \frac{\alpha}{1-\alpha}$, $\beta_1^* = \frac{\beta_1}{1-\alpha}$, and $e_{it}^* = \frac{e_{it}}{1-\alpha}$. Averaging over K gives

$$\bar{Y}_{iK} = \beta - \underbrace{(\alpha^*\Delta y_{iK} + \beta_1^*\Delta x_{iK})}_{\eta_{iK}} + \bar{e}_{iK}^*,$$

where $y_{iK} = \Delta y_{iK} = \frac{(y_{i,T_1}-y_{i,T_0})}{K} = O_p(K^{-1/2})$ and $x_{iK} = \frac{(x_{i,T_1}-x_{i,T_0})}{K} = O_p(K^{-1/2})$ only depend on the endpoint observations.. The static regression is $\bar{Y}_{iK} = c + \beta\bar{X}_{iK} + \text{error}$, where error$=-\eta_i+\bar{e}_{iK}$. Since x depends on u and $\mathbb{E}[u_{is}e_{it}]$ for all s and for every i, $\text{cov}(\bar{X}_{iK},\bar{e}_{iK}) = 0$. But $\eta_{iK} = O_p(K^{-1/2})$, which is the same order as $\text{var}_i(\bar{e}_{iK})$. By Cauchy Schwarz inequality,

$$|\text{cov}_i(\bar{X}_{iK},\eta_{iK})| \le \sqrt{\text{var}_i(\bar{X}_{iK})}\sqrt{\text{var}_i(\eta_{iK})} = O(K^{1/2})O(K^{-1/2}) = O(1).$$

The effects of the omitted variables do not vanish with K. However, $\text{var}_i(\bar{X}_{iK}) = O(K)$. Thus, for fixed K,

$$\text{plim}\hat{\beta} = \beta - \frac{\text{cov}_i(\bar{X}_{iK},\eta_{iK})}{\text{var}_i(\bar{X}_{iK})}$$

and

$$\sqrt{N}\left(\hat{\beta}-\beta+\frac{\text{cov}_i(\bar{X}_{iK},\eta_{iK})}{\text{var}_i(\bar{X}_{iK})}\right)\xrightarrow{d}N(0,Q_{XX}^{-1}W_KQ_{XX}^{-1}),$$

where $W_K = \text{var}_i((\bar{X}_{iK} - \mu_X)(\bar{e}_{iK} - \eta_{iK})$. Though $\bar{e}_{iK}$ does not contribute to bias, it contributes to variance. For large K,

$$\frac{\text{cov}_i(\bar{X}_{iK}, \eta_{iK})}{\text{var}_i(\bar{X}_{iK})} = \frac{O(1)}{O(K)} = O(K^{-1}) \to 0$$

and with $\mathcal{W} = \lim_{K\to\infty} W_K$ and $\mathcal{Q}_{XX} = \lim_{K\to\infty} KQ_{XX}$, we have

$$\sqrt{N}K(\hat{\beta} - \beta) \xrightarrow{d} N(0, \mathcal{Q}_{XX}^{-1}\mathcal{W}\mathcal{Q}_{XX}^{-1}).$$

A-5 Proof of Theorem 1

Since $x_{it} = x_{it-1} + u_{it}$, $x_{it} = x_{i,T_0} + \sum_{s=T_0+1}^{t} u_{is}$. The weighted average $\bar{X}_i = \sum_t w(t)x_{it}$ with $w_t > 0$ and $\sum_t w(t) = 1$. More generally,

$$\bar{X}_{iK} = \sum_{t=T_0+1}^{T_1} w(t)\left(x_{i,T_0} + \sum_{s=T_0+1}^{t} u_{is}\right) = x_{i,T_0} + \sum_{t=T_0}^{T_1} w(t) \sum_{s=T_0+1}^{t} u_{is}.$$

Reversing the order of summation to sum over s first, then over all $t \geq s$, we have $\bar{X}_{iK} = x_{i,T_0} + \sum_{s=T_0+1}^{T_1} u_{is} \sum_{t=s}^{T_1} w(t)$. Defining $\phi_s \equiv \sum_{t=s}^{T_1} w(t)$ yields:

$$\widetilde{X}_i = x_{i,T_0} + \sum_{s=T_0+1}^{T_1} \phi_s\, u_{is}, \qquad \phi_s \equiv \sum_{t=s}^{T_1} w(t), \tag{8}$$

where ϕ_s is the tail sum of w at s, and the sequence $\{\phi_s\}$ is the survival function of the weight distribution w. Since $\phi_s \in [0, 1]$ for any non-negative w such that $\sum_{t=T_0+1}^{T_1} w(t) = 1$, we have $\phi_s^2 \leq \phi_s$ for each s. Consider now

$$\sum_{s=T0+1}^{T_1} \phi_s = \sum_{s=T_0+1}^{T_1} \sum_{t=s}^{T_1} w(t) = \sum_{t=T_0+1}^{T_1} w(t) \sum_{s=T_0+1}^{t} 1 = \sum_{t=T_0+1}^{T_1} w(t)(t - T_0) = K\mathbb{E}_w\left[\frac{t - T_0}{K}\right].$$

For any non-degenerate w, $\mathbb{E}_w[(t - T_0/K] = O(1)$. Thus $\sum_s \phi_s = O(K)$ and $\sum_{s=T_0+1}^{T_1} \phi_s^2 = O(K)$.

In the stationary case, we have, by definition,

$$\text{var}(\widetilde{X}_{iK}) \quad = \quad \sum_{s=T_0+1}^{T_1} \sum_{t=T_0+1}^{T_1} w(t)w(s)\text{cov}(x_{it}, x_{is}).$$

Consider $x_{it} = \rho x_{i,t-1} + u_{it} = \sum_{s=0}^{\infty} \rho_i^s u_{t-s}$ where $u_{it} \sim (0, \sigma_i^2)$, we have $\bar{X}_{iK} = \sum_s \psi_s(\rho_i)u_{is}$ with $\psi_s(\rho_i) = \sum_{t=s}^{T_1} w(t)\rho_i^{t-s}$ and $\text{cov}(x_{it}, x_{is}) = \frac{\sigma_i^2}{1-\rho_i^2}\rho_i^{|t-s|}$. Thus rearranging the double sum by letting

$h = s - t$ which runs from $-(K-1)$ to $K-1$:

$$\begin{aligned}
\text{var}(\widetilde{X}_{iK}) &= \frac{\sigma_i^2}{1-\rho_i^2} \sum_{h=-(K-1)}^{(K-1)} \sum_{t=-s=h} w(t)w(s)\rho_i^{|t-s|} \\
&= \frac{\sigma_i^2}{1-\rho_i^2} \sum_{h=(K-1)}^{K-1} A_h \rho_i^{|h|} = \frac{\sigma_i^2}{1-\rho_i^2} G(\rho_i) \\
&\equiv \sigma_i^2 \Psi_K(\rho)
\end{aligned}$$

where $A_h = \sum_{t-s=h} w(t)w(s)$, and $\Psi_K(\rho) = \frac{1}{1-\rho_i^2} G(\rho_i)$. But $\{A_h\}$ forms a probability distribution over periods separated h periods because $A_h > 0$ and $\sum_h A_h = \sum_h \sum_{t-s=h} w(t)w(s) = \sum_t \sum_s w(t)w(s) = (\sum_t w(t))(\sum_s w(s)) = (\sum_t w(t))^2 = 1$. It follows that, $G(\rho_i) = \mathbb{E}_A[\rho_i^{|h|}]$. Since $0 \leq \rho_i^{|h|} \leq 1$ for any $|\rho_i| \leq 1$, $0 \leq G(\rho_i) \leq 1$ for any $\rho_i \in [0,1]$ and any w. In the local to unity framework, $\rho_i^{|h|} = (1 - c_i/K)^{|h|} \approx 1 - \frac{c_i|h|}{K}$. Thus, $G(\rho_i) = \mathbb{E}_w[1 - \frac{|h|a_i}{K}] = 1 - \frac{c_i}{K}\mathbb{E}_w[|h|]$. For simple averaging, $\mathbb{E}_A[|h|] \approx K/3$ and thus $G(\rho_i) \approx 1 = O(1)$. The order of Ψ_K is fundamentally determined by ρ_i and not about the choice of $w(t)$, so long as it satisfies $w(t) \geq 0$ and $\sum_t w(t) = 1$.

For simple averaging, the result is immediate since $\text{var}(\bar{X}_{iK}) = \frac{1}{K}\frac{\sigma_i^2}{(1-\rho_i)^2}$. In the local-to-unity case when $\rho_i = 1 - c_i/K$, $\sigma_i^2 = K^2\sigma^2/a_i^2$, implying $\Psi_K = O(K)$. Thus $\text{var}_i(\bar{X}_{iK}$ behaves as if $\rho_i = 1$, and $Q_{XX} = O(K)$ in the local-to-unity setting.

A-5.1 Proof of Equation 5

The meat of the sandwich variance is $V = \text{var}_i(\widetilde{X}_{iK} - \mu_X \widetilde{e}_{iK})$. We want to show that under homoskedasticity ($\sigma_i^2 = \sigma^2, \tau_i^2 = \tau^2 \; \forall i$), the cross-sectional variance is a product of unconditional moments.

Under strict exogeneity, $\mathbb{E}_i[(\widetilde{X}_{iK} - \mu_X)\widetilde{e}_{iK}] = 0$ and $V = \mathbb{E}_i[(\widetilde{X}_{iK} - \mu_X)^2\widetilde{e}_{iK}^2]$. Since $\widetilde{X}_{iK}$ depends on u_{it} and $\widetilde{e}_{iK}$ depends on e_{it}, and the shocks are independent across i,

$$\mathbb{E}_i[(\widetilde{X}_{iK} - \mu_X)^2\widetilde{e}_{iK}] = \mathbb{E}_i[(\widetilde{X}_{iK} - \mu_X)^2]\mathbb{E}_i[\widetilde{e}_{iK}^2] = Q_{XX}\mathbb{E}_i[\widetilde{e}_{iK}^2].$$

But for a given i, $\mathbb{E}[\widetilde{e}_{iK}] = 0$. Thus, $\mathbb{E}_i[\widetilde{e}_{iK}^2] = \text{var}_i(\widetilde{e}_{iK}) = \mathbb{E}_i[\text{var}(\widetilde{e}_{iK})] + \text{var}_i(\mathbb{E}[\widetilde{e}_{iK}))$. The second term is zero, and the first term $\mathbb{E}_i[\text{var}(\widetilde{e}_{iK})] = \mathbb{E}_i[\tau_i^2 A_0(w)]$. Under homoskdasticty, this is $\tau_2 A_0(w)$. Combining the result, $V = Q_{XX}\tau^2 A_0(w) = \sigma^2\Phi_K\tau^2 A_0(w)$. The sandwich variance $Q_{XX}^{-1}VQ_{XX}^{-1}$ is thus $\frac{\tau^2}{\sigma^2}\frac{A_0(w)}{\Phi_K}$.

When strict exogeneity fails, $\text{cov}_i(\widetilde{X}_{iK}, \widetilde{e}_{iK}^*) \neq 0$. Using $\mathbb{E}_i[A^2B^2] = \mathbb{E}_i[A^2]\mathbb{E}_i[B^2] + \text{cov}_i(A^2, B_2)$ with $A = \widetilde{X}_{iK} - \mu_X$ and $B = \widetilde{e}_{iK}^*$, we have, as shown in (6): $V = \sigma_{e_*}^2 Q_{XX} + \text{cov}_i((\widetilde{X}_{iK} - \mu_X)^2, (\widetilde{e}_{iK}^*)^2)$.

A-5.2 Relaxing Strict Exogeneity

Under strict exogeneity in Assumption A.i , $\text{cov}_i(\widetilde{X}_{iK}, \widetilde{e}_{iK}) = \sum_s \phi_s \sum_t w(t)\mathbb{E}_i[u_{it}e_{is}] = 0$. We now allow a correlation between e_{is} and u_{it}. Let $\text{cor}(u_{it}, e_{is}) = \gamma_{i,t-s}$ such that $\sum_{h=0}^{p} |\gamma_{i,h}|$ is bounded. In terms of covariance, $\text{cov}(u_{it}, e_{is}) = \gamma_{i,t-s}\sigma_i\tau_i$. From $\widetilde{X}_{iK} = x_{i,T_0} + \sum_s \phi_s u_{is}$ and $\widetilde{e}_{iK} = \sum_t w(t)e_{it}$,

$$\begin{aligned}
\text{cov}(\widetilde{X}_{iK}, \widetilde{e}_{iK}) &= \text{cov}\left(x_{i,T_0} + \sum_{s=T_0+1}^{T_1} \phi_s u_{is}, \sum_{t=T_0+1}^{T_1} w(t)e_{it}\right) \\
&= \sigma_i\tau_i \sum_{s=T_0+1}^{T_1} \sum_{t=T_0+1}^{T_1} \phi_s w(t)\gamma_{i,s-t} \\
&= \sigma_i\tau_i \sum_{h=-(K-1)}^{K-1} \gamma_{ih}B_h(w), \quad \text{where} \quad B_h(w) = \sum_{s-t=h} \phi_s w(t)
\end{aligned}$$

where the last equality follows from grouping by lag $h = s - t$ from $-(K-1)$ to $(K-1)$, for $s, t \in [T_0 + 1, T_1]$. Note that $B_h(w)$ does not depend on i and let $B_\infty(w) = \lim_{K\to\infty} B_h(w)$. Since $\sum_h |\gamma_{ih}| < \infty$ and $|B_h(w)| \leq \max_s |\phi_s| \sum_t |w(t)| = O(1)$, dominated convergence gives

$$\Lambda_K(w, i) \to \Gamma_i B_\infty(w)$$

For K averaging, $w(t) = 1/K$ and $\phi_s = \frac{(T_1 - s + 1)}{K}$. Thus

$$B_h(w) = \sum_{s-t=h} \phi_s w(t) = \frac{1}{K} \sum_{\substack{s-t=h \\ T_0+1\leq s,t\leq T_1}} \frac{T_1 - s + 1}{K} = \frac{1}{K} \sum_{j=h+1}^{K} \frac{(K - j + 1)}{K}.$$

When $j = h+1$, the term is $K-h$. When $j = h+2$, it is $K-h-1$, when $j = K$, it is 1. Changing index to let $m = K - j + 1$, we see that as j increases from $h+1$ to K, m decreases from $K-h$ to 1. Thus,

$$\frac{1}{K} \sum_{j=h+1}^{K} \frac{(K - j + 1)}{K} = \sum_{m=1}^{K-h} m = \frac{1}{K^2} \frac{(K-h)(K-h+1)}{2} \to 1/2.$$

For point sampling at T_1, $w(t) = 1_{t=T_1}$ and $\phi_s = \sum_{t>s} w(t) = 1$ for all $s \leq T_1$. But with any h, the only contributing pair is $(s, t) \in [T_1 + h, T_1]$ with $\phi_{T_1+h} = 1$, and $w(T_1) = 1$. Thus, $B_h(w) = B_\infty(w) = 1$. With $\widetilde{b}_K = 1/2$ for K averaging, and $\widetilde{b}_K = 1$ for sampling at T_1, we have $\text{cov}(\widetilde{X}_{iK}, \widetilde{e}_{iK}) = \sigma_i\tau_i\Lambda_K(w, i) \to \sigma_i\tau_i B_\infty(w) = O(1)$.

Convergence Under K-averaging when there is a drift

i. Suppose x_{it} is a unit root process with drift g_i, ie $\Delta x_{it} = g_i + u_{it}$ with $\text{cov}(g_i, x_{i,T_0}) = 0$. Then for $t \in [T_0 + 1, T_1]$, $x_{it} = x_{i,T_0} + g_i(t - T_0) + \sum_{s=T_0+1}^{t} u_{is}$ has both a deterministic and

a stochastic trend. Since $\frac{1}{K}\sum_{t=T_0+1}^{T_1}(t-T_0)=\frac{K+1}{2}$, $\bar{X}_{iK}=x_{i,T_0}+g_i\frac{K+1}{2}+\sum_{s=T_0+1}^{T_1}u_{is}$,

$$\text{var}_i(\bar{X}_{iK}) \quad = \quad \text{var}_i(x_{i,T_0})+\bar{K}^2\text{var}_i(g_i)+a_K\mathbb{E}_i[\sigma_i^2]$$

where $\bar{K}=(K+1)/2$. The fanning out rate of K^2 is faster when $g_i\neq 0$ and is heterogeneous.

ii Suppose now $\text{cov}_i(x_{i,T_0},g_i)<0$. With convergence,[16]

$$\text{var}_i(\bar{X}_{iK}) \quad = \quad \underbrace{\text{var}_i(x_{i,T_0})+2\bar{K}\text{cov}_i(x_{i,T_0},g_i)+\bar{K}^2\text{var}_i(g_i)+\bar{K}\mathbb{E}_i[\sigma_i^2]}_{\text{var}_i(\bar{x}_{i,T_0+\bar{K}})}+(\bar{a}_K-\bar{K})\mathbb{E}_i[\sigma_i^2].$$

The first three terms represent the cross-section variance at the midpoint of $\bar{K}$ which is strictly positive, while the last term comes from noise accumulation since $\bar{K}$. Like point sampling, the cross-section variance of the K-averaged data also shows a U-shape and the minimum is generally at a point later than under point sampling.

iii If $x_{it}=\lambda_i f_t+w_{it}$, $f_t=g_f+f_{t-1}+u_{it}$, w_{it} is stationary. Then after cross-section demeaning,

$$\text{var}_i(\bar{X}_{iK})=\text{var}_i(\lambda_i)\bar{f}_K^2+\text{var}_i(\bar{w}_{iK}).$$

Now $\bar{w}_{iK}=O(K^{-1/2})$. With $\bar{f}_K=f_0+g_fK/2+O(K^{1/2}$, we have

$$\bar{f}_K^2=(f_0+g_fK/2)^2+O_p(K)$$

. If $g_f<0$, $\text{var}_i(\bar{X}_{iK})$ can be U-shaped. The deterministic part reaches zero at $K^*=-2f_0/g_f=2t^*$, where t^* is the turning point under point sampling. At K^*, $\text{var}_i(\bar{X}_{iK})=O_p(K^*)$. As with point sampling, a U-shaped cross-section requires $f_0\neq 0$. It also does not mean that $\text{var}_i(\bar{X}_{iK})$ tends to zero, but that recovery can be expected in time.

Long Differencing

Define $\widetilde{X}_{iK}=\bar{X}_{iM}(T_1)-\bar{X}_{iM}(T_0)$, where

$$\widetilde{X}_{iK}=\frac{1}{K}\sum_{t=T_1-m}^{T_1+m}\sum_{s=T_0+1}^{t}u_{is}-\frac{1}{K}\sum_{t=T_0-m}^{T_0+m}\sum_{s=T_0+1}^{T_1}u_{is}..$$

The weight function ϕ_s on the innovations has the following structure:

- j/M for $j=1,\ldots,2m$ as s moves from T_0-m+1 to T_0+m;
- 1 for $s\in(T_0+m,T_1+m]$, or $K-M+1$ innovations

[16] We need to evaluate $K^{-2}\sum_s\sum_\ell\text{cov}_i(x_{i,T_0+s},x_{i,T_0+\ell}$. For $[s,\ell]\in[1,K]$, $K^{-2}\sum_s\sum_\ell 1=1$, $K^{-2}\sum_s\sum_{ell}(s+\ell)=2\bar{K}$, $K^{-2}\sum_s\sum_\ell s\cdot\ell=\bar{K}^2$, and $K^{-2}\sum_s\sum_\ell\min(s,\ell)=\bar{a}_K$.

- $\frac{(2m-j+1)}{M}$ for $j = 1, \ldots, 2m$ as s moves from $T_1 - m$ to $T_1 + m$.

The square weights has the form $(K - M + 1) + \frac{2}{M^2}\sum_{j=1}^{2m} j^2$, and $\sum_{j=1}^{2m} = m(2m+1)(4m+1)/3$ so that $d_M = \frac{2}{M^2}\sum_{j=1}^{2m} j^2 = \frac{(M-1)(2M-1)}{2M}$ and $\text{var}(\widetilde{X}_{iK}) = \sigma_i^2(K - \frac{M^2-1}{3M})$. Thus, in the correctly specified static regression, $Q_{XX} \approx \mathbb{E}_i[\sigma_i^2]K$. Furthermore, $\text{var}_i(\breve{e}_{iK}) = 2\mathbb{E}_i[\tau_i^2]/M$. The sandwich variance for LD is $Q_{XX}^{-1}V_K Q_{XX}^{-1} = O(K^{-2})$, implying that the LD has a convergence rate of $\sqrt{N}K$, as in K-averaging.

Now suppose the DGP is the ADL model. For $j = 0, 1$,

$$\bar{Y}_{iM}(T_j) = \alpha\bar{Y}_{iM,-1}(T_j) + \beta_0\bar{X}_{iM}(T_j) + \beta_1\bar{X}_{iM,-1}(T_j) + \bar{e}_{iM}(T_j).$$

Let $\Delta x_{i,T_j-m} = x_{i,T_j-m-1}$ for $j = 0, 1$, and $\Delta y_{i,T_j-m}$ is similarly defined. Then

$$\breve{Y}_{iK} = \beta\breve{X}_{iK} \underbrace{-\frac{\alpha}{1-\alpha}\frac{\Delta y_{i,T_1-m} - \Delta y_{i,T_0,m}}{M} - \frac{\beta}{1-\alpha}\frac{\Delta x_{i,T_1-m} - \Delta x_{i,T_0,m}}{M}}_{\breve{\eta}_{iK}} + \frac{\breve{e}_{iK}}{1-\alpha}$$

Nickell bias is bias $= -\frac{\text{cov}(\breve{X}_{iK}, \breve{\eta}_{iK})}{\text{var}_i\breve{X}_{iK})}$. Since x is a random walk, $\Delta x_{iM}(T_j) = O_p(M^{-1/2})$ and similarly for Δy_{T_j}. Hence, $\breve{\eta}_{iK} = O_p(M^{-1/2})$. With $\text{var}_i(\breve{X}_{iK}) = O(K)$, the bias is $O((KM)^{-1/2}) \to 0$ as $K \to \infty$. Hence β can also be consistently estimated by LD, but the bias is larger than the $K-$averaging bias which is $O(K^{-1})$. In the homoskedastic case I(1) case,

	Q_{XX}	$\mathbb{E}_i[\text{var}(\bar{e}_{iK})]$
K-averaging	$\mathbb{E}_i(\sigma_i^2)(T_0 + a_K)$	$\frac{1}{K}\mathbb{E}_i(\tau_i^2)$
LD	$\mathbb{E}_i(\sigma_i^2)(K - \frac{M^2-1}{3M})$	$\frac{2}{M}\mathbb{E}_i(\tau_i^2)$

The variance ratio of LD to K averaging is

$$\text{ratio} = \frac{2K(T_0 + a_K)}{M(K - M/3 - 1/(3M))} \approx \frac{2(3T_0 + K)}{3M(1 - M/(3K))} \approx \frac{2K}{3M} + \frac{2T_0}{M}$$

where the first approximation uses $a_K \approx K/3$ and the second assumes $M/(3K)$ is small.

When x is I(0). Consider again $x_{it} = \rho x_{it-1} + u_{it}$ with $\rho < 1$. Since $\bar{X}_{iM}(T_j) = \frac{1}{3}(x_{i,T_j-1} + x_{i,T_j} + x_{i,T_j+1})$, the weight on innovation u_{is} in $\bar{X}_{iM}(T_j)$ for $s \leq T_j - 1$ is:

$$\frac{1}{3}\left(\rho^{T_j-1-s} + \rho^{T_j-s} + \rho^{T_j+1-s}\right) = \frac{\rho^{T_j-1-s}(1 + \rho + \rho^2)}{3}.$$

Denoting $\kappa = (1 + \rho + \rho^2)/3$, the weights on u_{is} in $\breve{X}_{iK} = \bar{X}_{iM}(T_1) - \bar{X}_{iM}(T_0)$ are:

s	$\cdots$	3	4	5	6	7	8	9	10	11	$\cdots$
$w_0(s)$	$\cdots$	ϕ	$\frac{1+\rho}{3}$	$\frac{1}{3}$	0	0	0	0	0	0	$\cdots$
$w_1(s)$	$\cdots$	$\phi\rho^6$	$\phi\rho^5$	$\phi\rho^4$	$\phi\rho^3$	$\phi\rho^2$	$\phi\rho$	ϕ	$\frac{1+\rho}{3}$	$\frac{1}{3}$	$\cdots$
$w(s)$	$\cdots$	$\phi(\rho^6 - 1)$	$\phi\rho^5 - \frac{1+\rho}{3}$	$\phi\rho^4 - \frac{1}{3}$	$\phi\rho^3$	$\phi\rho^2$	$\phi\rho$	ϕ	$\frac{1+\rho}{3}$	$\frac{1}{3}$	$\cdots$
$\psi_s(\rho)$	$\cdots$	-0.574	-0.482	-0.297	0.073	0.146	0.292	0.583	0.500	0.333	$\cdots$

In the above, $\psi_s(\rho) = \sum_{t \geq s} w(t)\rho^{t-s}$, where $w(s) = w_1 - w_0$. For $\rho = 1$, this reduces to a weight of 1 for $s = 6, 7, 8, 9$ obtained earlier. For $\rho < 1$, innovations in the m periods from the T_1 window receive weight ρ_i^m in $\bar{X}_{iM}(T_1)$, and the weight becomes negligible when ρ_i^m is small. The variance of $\breve{X}_{iK}$ is almost entirely driven by innovations within $O(1/(1-\rho_i))$ periods of each window, leaving a lot of information unused. The last row shows $\psi_s(\rho)$ evaluated at $\rho = 0.5$. When $\rho_i = 0$ for example, $\text{var}(\widetilde{X}_{iK}) = 2\sigma_i^2/M$. More generally, $\text{var}(\breve{X}_{iM}(T_j)) \approx \sigma_i^2/M$ changes with M, not K. Accordingly, the Q_{XX} is $O(M)$ for any K. In a sense, the effective sample size of an LD regression is M when x is stationary. Consider again the static regression when the DGP is dynamic. When x is I(0), $\Delta x_{iM}(T_j) = O_p(M^{-1})$ and $\breve{\eta}_{iK} = O_p(M^{-1})$. With $\text{var}_i(\breve{X}_{iK}) = O(M^{-1})$, bias is $O(M^{-1/2})$. But M is fixed and the bias does not vanish.

We also use Table 2 to verify the theoretical predictions. When $\rho = 0$, $\vartheta = 0$, and x is a pure random walk. The fanning out effect is smaller when $\vartheta = -0.5$, obtained by setting $\rho = 0.5$. To verify some results presented earlier, we evaluate KA$(T_0 : T_2)$ and LD$(T_0, T_2, m = 9)$ with $T_0 = 10, T_2 = 40$, $K = 30$ for two setings. In the first case with $T_{init} = 0$, $\mathbb{T}_0 = T_0 = 10$. With $a_K = (K+1)(3K+1)/(6K) = 10.52$, the Q_{XX} for K-averaging is $10 + a_{30}$=10+10.52=20.52, compared to 20.65 averaged over simulations. For LD with $M = 2m + 1 = 19$, $Q_{XX} = (K - \frac{M^2-1}{3M}) = 23.68$ compared to 23.75 in simulations. Since the design assumes homoskedasticity, V_K simplifies to regression-error variance times Q_{XX}. For K averaging, $\text{var}_i(\bar{e}_{iK}) = \tau^2/K$, so $V_K = 20.52 \times 2/30 = 1.368$, compared to 1.285 in simulations. For LD, $\text{var}_i(\breve{e}_{iK} = 2\tau^2/M$, so $V_K = 23.68 \times 4/19 = 4.985$, compared to 4.613 in simulations. In the second case when $T_{init} = 50$ and thus $\mathbb{T}_0 = 60$, we have $Q_{XX} = 70.52$ for K-averaging, compared to 70.037 in simulations) and 23.6 for LD (compared to 23.55 in simulations. For V_K, it is 4.701 for K-averaging, compared to a White's estimate of 4.285 in simulations. For LD, it is 4.98 (compared to a White's estimate of 4.528 in simulations).

Inn the first case, LD and KA have comparable Q_{XX}, but V_K is more favorable for K-averaging. Comparing the variance of $\hat{\beta}$ of LD to KA, we have a ratio of 2.74. KA is more efficient because it dilutes noise, LD magnifies by differencing. In the second case, LD and KA have comparable V_K, Q_{XX} is larger for KA. The variance of $\hat{\beta}$ of LD relative to KA is 9.4; KAis more efficient in this case because it uses information in the initial condition while the LD discards. The higher variance of LD in both configurations translates power differences.

References

Acemoglu, D. and Restrepo, P. 2017, Secular Stagnation? The Effect of Aging on Economic Growth in the Age of Automation, *American Economic Review* **107:5**, 174–179.

Andrews, D. K. 2005, Cross-Section Regression with Common Shocks, *Econometrica* **73:5**, 1551–1586.

Bai, J. 2009, Panel Data Models with Interactive Fixed Effects, *Econometrica* **77**, 1229–1279.

Bai, J. and Kao, C. 2006, On the Estimation and Inference of a Panel Cointegration Model with Cross-Section Dependence, *in* B. Baltagi (ed.), *Contributions to Economic Analysis*, Elsevier, pp. 3–30.

Bai, J., Kao, C. and Ng, S. 2009, Panel Cointegration with Global Stochastic Trends.

Barro, R. 1991, Economic Growth in Cross-Section of Countries, *Quarterly Journal of Economics* pp. 407–443.

Berenguer-Rico, V. and Gonzalo, J. 2014, Summability of Stochastic Processes- A Generalization of Integration for Non-Linear Proccess, *Journal of Econometrics* **178:2**, 331–341.

Bewley, R. A. 1979, The Direct Estimation of the Equilibrium Response in a Linear Model, *Economics Letters* **7**, 357–61.

Bilal, A. and Kanzig, D. 2024, The Macroeconomic Impact of Climate Change: Global vs Local Temperature, NBER Working paper 32450.

Burke, M. and Emerick, K. 2016, Adaptation to Climate Change: Evidence from US Agriculture, *American Economic Journal: Economic Policy* **8:3**, 106–140.

Burke, M., Hsiang, S. and Miguel, E. 2015, Global Non-Linear Effect of Temperature on Economic Production, *Nature* **527-7577**, 235–239.

Chudik, A. and Pesaran, M. H. 2015, Common Correlated Effects Estimation of Hetgerogeneous Dynamic Panel Data Models with Weakly Exogeneous Regressors, *Journal of Econometrics* **188:2**, 393–420.

Chudik, A., Mohaddes, K., Pesaran, M. H. and Raissi, M. 2016, Long-Run Effects in Large Heterogeneous Panel Data Models with Cross-Sectionally Correlated Errors, *Essays in Honor of Aman Ullah*, Emerald Group Publishing Limited.

Deaton, A. S. and Paxson, C. 1994, Intertemporal Choice and Inequality, *Journal of Political Economy* **CII**, 437–467.

Dees, S., Holly, S., Pesaran, M. H. and Smith, L. V. 2007, Long Run Macroeconomic Relations in the Global Economy, *Economics - The Open-Access, Open-Assessment E-Journal (2007-2020), Kiel Institute for the World Economy* **1**, 1–20.

Dell, M., Jones, B. and Olken, B. 2012, Temperature Shocks and Economic Growth: Evidence from the Last Half Century, *American Economic Journal: Macroeconomics* **4:3**, 66–95.

Deschenes, O. and Greenstone, M. 2007, The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather, *American Economic Review* **97:1**, 354–84.

Easterly, W. and Levine, R. 1997, Africa's Growth Tragedy: Policies and Ethnic Divisions, *The Quarterly Journal of Economics* **112:4**, 1203–1250.

Engle, R. F. and Granger, C. 1987, Cointegration and Error-Correction: Representation, Estimation, and Testing, *Econometrica* **55**, 251–276.

Feenstra, R. C., Inklaar, R. and Timmer, M. P. 2015, The Next Generation of the Penn World Table, *American Economic Review* **105:10**, 3150–3182.

Feldstein, M. and Horioka, C. 1980, Domestic Saving and International Capital Flows, *Economic Journal* **90:358**, 314–329.

Gordon, R. J. 2016, *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*, Princeton University Press, Princeton.

Howrey, P. 1968, A Spectrum Analysis of the Long-Swing Hypothesis, *International Economic Review* **9:2**, 228–252.

Jong-Won Yoon, J. K. and Lee, J. 2014, Impact of Demographic Changes on Inflation and the Macroeconomy. IMF Working Papers 2014, 210.

Juselius, M. and Takats, E. 2021, Inflation and Demography Through Time, *Journal of Economic Dynamics and Control* **128**, 104136.

Kuznets, S. 1958, Low Swings in the Growth of Population and in Related Economic Variables, *Proceedings of the American Philosophical Society* **102**, 25–52.

Lunsford, K. G. 2017, Productivity Growth and Real Interest Rates in the Long Run, *Federal Reserve Bank of Cleveland Economic Commentary*.

Mendelsohn, R. and Masetti, E. 2017, The Use of Cross-Sectional Analysis to Measure Climate Impacts on Agriculture: Theory and Evidence, *Review of Environmental Economics and Policy* **11:2**, 280–298.

Mendelsohn, R., Nordhaus, W. D. and Shaw, D. 1994, The Impact of Global Warming on Agricutlure: A Ricardian Analysis, *American Economic Review* **89:4**, 753–771.

Mueller, U. K. and Watson, M. W. 2017, Low Frequency Econometrics, *Advances in Economics and Econometrics*, Vol. 2, Cambridge University Press, p. p53.

Muller, K., Xu, C., Lehbib, M. and Chen, Z. 2025, The Global Macro Database: A New International Macroeconomic Dataset.

Ng, S. 2008, A Simple Test for Non-Stationarity in Mixed Panels, *Journal of Business and Economic Statistics* **26:1**, 113–127.

Perron, P. and Ng, S. 1996, Useful Modifications to Unit Root Tests with Dependent Errors and their Local Asymptotic Properties, *Review of Economic Studies* **63**, 435–465.

Pesaran, M. H. and Smith, R. 1995, Estimating Long-Run Relationships from Dynamic Heterogeneous Panels, *Journal of Econometrics* **68**, 79–113.

Yi, K. M. and Zhang, J. 2017, Understanding Global Trends in Real Interset Rates, *Federal Reserve Bank of Chicago: Economic Perspectives*.