Cross-Section Regressions in Time Compressed Data: Estimating the Effect of Warming Temperature on Growth

Nikolay Gospodinov* Serena Ng †

March 20, 2026
Preliminary and Incomplete
Please do not Quote

Abstract

To learn about the effects of warming temperature on growth, it is quite common to transform the data by taking the difference of two centered moving average, thereby removing the time dimension. This paper studies the properties of time compression by K period averaging and their implications for regression. We establish that the cross-section variance of the averaged data 'fans out' with K if the data are I(1). We then show that a correctly specified cross-section regression in the averaged data can yield $\sqrt{N}K$ consistent and asymptotically normal estimates of many quantities of economic interest. With full sample averaging, a super-consistent rate of $\sqrt{NT}$ can be obtained without having to deal with unit-root asymptotics. Averaging belongs to a class of time-compressed functions that magnify the signal in the I(1) regressor while diluting the stationary regression noise, and still allow $\sqrt{N}$ consistent estimates of correctly specified models when the regressors are I(0). The result that such functions permit smooth transition of the asymptotic distribution from I(0) to I(1) is of general interest because it suggests that by doing cross-section regressions in time compressed data, standard normal inference can be used in the empirically relevant case when the regressors are highly persistent but not necessary non-stationary.

Keywords: climate change, non-stationarity, fanning out, long-difference 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

There is increased policy interest in learning about the effects of climate change on economic outcomes. In economic models, the short-run response is the effect that follows a shift in the climate with expectations, beliefs, and investment held fixed, while the long-run response is the effect after both expectations and investments have adjusted to the new climate. However, these model concepts do not have a direct mapping in regressions. Concerns were raised over early cross-section regressions that omitted variables could bias the estimate of the long-run effect, also known as long-run multiplier. More recent works use short T panels in fixed-effect estimation but are worried that such estimates would not reflect longer term effects. This motivates using long-differences (LD) of time averaged data to estimate longer-term impacts from cross-section regressions, thereby compressing the time dimension. Empirical results often find the short and longer-term estimates to be similar. See, for example, discussion in Kolstad and Moore (2020), Dell et al. (2014, p.778), Carleton et al. (2024, Section 3.1), Hsiang (2016).

This paper considers a class of time-compression functions of which using data from only one period and time averaging are members. Our point of departure is the result in Deaton and Paxson (1994) that income inequality will increase with age if individual incomes (in logs) are random walks, with the implication that the cross-section variance of a non-stationary panel will 'fan-out' over time. We consider a more general non-stationary setup and show that the cross-section variance of a panel with N units over a span of K periods will increase at rate K with a pre-factor that depends on $\mathbb{E}_i[\omega_i^2]$, the mean over units of the spectral density at frequency zero of innovations. This implies that if we can only take slice of the data from a $T \times N$ panel, taking data from the last sample period, T, will have the most variability.

We then consider time compression by averaging the data over time. Time averaging is commonly thought to remove high frequency noise, making it easier to visualize patterns at the cost of creating spurious dynamics.[1] This is true in the stationary case. It turns out that with non-stationary data, the cross-section variance still fans out at rate K, but shrinks by a factor of $1/3$ relative to point sampling at period K. While the shrinkage seems undesirable, using time averaged data in regressions has the effect of increasing signal in the I(1) regressor while reducing noise in the error variance. We show that the cointegrating vector estimated from a cross-section regression of a correctly specified static model is $\sqrt{N}K$ and asymptotically normal, and a rate of

[1] For example, the recession indicator of Sahm (2019), also known as Sahm rule, is based on a three-month moving average of the unemployment rate relative to its low in the prior 12 months. In the empirical growth literature, it is common to look at average growth rates over 5 or 10-years.

$\sqrt{NT}$ can be obtained with full sample averaging. It thus achieves fast convergence while retaining the convenience of standard normal inference. When the regressor is I(0), the estimate from the cross-section regression remains consistent and asymptotically normal, but converges at the standard rate of $\sqrt{N}$. Unlike time series regressions, the transition of the asymptotic distribution from I(0) to I(1) is smooth.

We also analyze the case when y is generated by a ADL(1,1) model. If the regressor is I(1), a static regression will yield $\sqrt{N}K$ consistent and asymptotically normal estimates of the long-run multiplier. If we estimate an ADL(1,1) model, the coefficients on x and its lag are $\sqrt{N}K$ consistent for any K while the coefficient on the lagged dependent variable has a Nickell bias that vanishes with K. In the I(0) case, the Nickell bias does not vanish but all other estimates are $\sqrt{N}$ consistent. Thus, estimates of coefficients relating to x are consistent and asymptotically normal whether or not x is stationary.

The smooth transition of the asymptotic distributions from I(0) to I(1) is possible not only with simple averaging, but for a class of time compressed functions with smooth weights on the innovations that can be characterized by a survival function. Importantly, the weights are continuous in the persistence parameter. Though the LD is not a member in this class, it also yields super consistent estimates in the I(1) case and is in fact more efficient than simple averaging. However, in the I(0) case and especially when the series is mildly persistent, it abandons information in all but the observations in a small neighbourhood, and discontinuous weights implicit in the LD prohibits an analysis of the stationary and non-stationary case within a unifying framework.

In what follows, we suppose that the ideal data consists of $\{y_{ist}, x_{ist}\}$ for $s = 1, \ldots, S$ locations, $i = 1, \ldots N$ units, and $t = 1, \ldots, T$ time periods. To simplify notation, we will use TS (TN) to refer to panel fixed effect regressions that do not involve units (locations). Then S and N are cross-section regressions. The SN derived from $S \times N$ panels of data averaged over all T periods are $\sqrt{SNT}$ consistent and asymptotically normal when the regressors are I(1) and the model is correctly specified. Section 2 begins with the key insight in Deaton and Paxson (1994). Section 3 analyzes time compression by K period averaging. Section 4 analyzes the implications for regressions. Section 5 summarizes the results for a class of time compressed functions. Section 6 analyzes long-differencing. Simulations are presented in Section 7. Section 8 concludes.

2 Cross-Section Variations at $t = K$ when x_{it} is Non-Stationary

Climate economists think of weather observations x_t like temperature as a draw from a climate distribution. If the climate distribution is time invariant with mean μ, we may write $x_t = \mu + w_t$, where $w_t \sim (0, \sigma^2)$ is stationary. But a constant μ is at odds with the data that appear to be trending.[2] The top panel of Figure 1 takes x_t to be the U.S temperature data from 1895-2022 along with the fit from a (local level) model that allows for a stochastic trend. We see that the data (in black) has a mean around 56^0F until the 1970s when it started to drift upwards, increasing to a mean of 59^0F in 2022. The bottom graph plots the series (in black) for New York. In dotted red along is the first principal component in a panel of demeaned data for the 50 states. For the state of New York, this component accounts for 83% of the total variation over time. Also plotted is the trend fitted to of the common component. Both plots indicate a slowly evolving trend.[3]

The effects of x_t on y_t at different horizons can be of interest, and one that has drawn attention is the long-run effect. This parameter is often estimated by a long-difference (LD) regressions. Given a $T \times N$ panel of data, an LD regression first takes the long-difference of the time averages of y and x, on the ground that the data transformation will better isolate the slowly evolving variations. Averaging and differencing collapse the time dimension to perform a cross-section regression.

We analyze the effects of time averaging when x is non-stationary. We will use two models to guide our analysis. Let u_{it} and w_{it} be mutually uncorrelated but possibly serially correlated with variances σ_u^2 and σ_w^2, respectively. Let q_i be the signal of the non-stationary component relative to the stationary component.

1. local level model with $q_i = \frac{\sigma_u^2}{\sigma_w^2}$:

$$\begin{aligned} x_{it} &= \mu_{it} + w_{it} \\ \mu_{it} &= \mu_{it-1} + u_{it} \end{aligned}$$

2 factor model: $q_i = \frac{\sigma_\lambda^2 \sigma_u^2}{\sigma_w^2}$.

$$\begin{aligned} x_{it} &= \lambda_i' \mu_t + w_{it} \\ \mu_t &= \mu_{t-1} + u_{it}, \end{aligned}$$

[2]Non-stationarity is modeled in, for example, Kolstad and Moore (2020, p.16), Nath et al. (2024, footnote 12), and Estrada et al. (2023). It is also considered in Kolstad and Moore (2020) and Hsiang (2016), Auffhammer (2018), Kolstad and Moore (2020), Carleton et al. (2024).

[3]The non-parametric approach of Mueller and Watson (2017) gives very similar estimates.

where σ_λ^2 denotes the variance of the loadings λ_i.

The local-level model provides a permanent-transitory decomposition of the data and one can think of μ_{it} as a time varying intercept. The factor model provides a common-idiosyncratic representation of the data; non-stationary comes from the common trend μ_t. Both models imply that x is an I(1) process but have a different signal-to-noise ratio q_i. Both models have reduced form $\Delta x_{it} = e_{it} + \vartheta_i e_{it-1}$ where $\vartheta_i = \left[-(q_i+2) \pm \sqrt{q_i^2+4q_i}\right]/2$. Of note is that $\vartheta_i \to 0$ as $q_i \to \infty$ and $\vartheta_i \to -1$ if $q_i \to 0$. For the U.S. data in the top panel of 1, $\vartheta = -0.8133$. For the New York data in the bottom panel, $\vartheta = -0.8578$. The negative value of ϑ will be relevant in what follows.

2.1 The Fanning Out of Cross-Section Variations

Our first objective is to understand the cross-section variability of I(1) data in a panel of N (or S) non-stationary series observed over T period. We will use $\text{var}_i(x_{i,T_0})$ to denote cross-section variance of x_{it} at $t = T_0$.

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 t. The result arises because consumption (c) is a random walk under PIH and can be characterized as $c_{it} = c_{it-1} + e_{it}$. Assuming in addition that $\text{cov}(c_{i,t-1}, e_{it}) = 0$, the cross-section variance of a given set of N individuals in an age cohort at time t is related to the one in $t-1$ by $\text{var}_{i|t}(c_{it}) = \text{var}_{i|t-1}(c_{it-1}) + \sigma_i^2$. It follows that the cross-section variance of the cohort will increase with t (or age), reflecting the cumulative differenecs in individual shocks over time. Ng (2008) shows that if a fraction κ of the panel is non-stationary, the cross-section variance will still fan out with t, but at rate κ.

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. Specifically, we relax the random walk assumption to allow x_{it} to be a unit root process with serially correlated noise. Consider $\Delta x_{it} = C_i(L)e_{it}$, where $x_{i0} = 0$ and $e_{it} \sim (0, \sigma_i^2)$ is a white noise error, $C_i(L) = 1 + C_{i1}L + C_{i2}L^2 + \ldots$ is a polynomial in the lag operator L, possibly of infinite order with $\sum_{j=0}^{\infty} C_{ij}^2 < \infty$. 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

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

In the above, x_{it} is decomposed into a random walk (non-stationary) component $C_i(1)\sum_{j=1}^{t} e_{ij} + C_i^*(L)e_{it}$ and a mean-reverting (stationary) component $w_{it} = C_i^*(L)e_{it}$. For a given i, the variance of x_{it} at $t = T_0$ is

$$\text{var}(x_{i,T_0}) = C_i(1)^2\sigma_i^2 T_0 + \text{var}(w_{i,T_0}).$$

Provided that $C_i(1)^2 \neq 0$, this variance will diverge with T_0. The cross-section variance at $t = T_0$ is given by

$$\text{var}_i(x_{i,T_0}) = \mathbb{E}_i[C_i(1)^2\sigma_i^2] \cdot T_0 + \text{var}_i[w_{i,T_0}] = O(T).$$

In the random walk case, w_{it} is absent and the divergence of the cross-section variance with t is guaranteed. In that case, If we wish to select data from a particular t_* for analysis, putting $t_* = T$ will yield a cross-section sample with the most variation. The stronger is the signal in the non-stationary component, the stronger is the signal of the full sample average.[4] But the presence of ϑ can magnify or suppress the fanning out effect.

To gain intuition, consider the IMA(1,1) model $\Delta x_{it} = u_{it} + \vartheta_i u_{it-1}$, where $u_{it} \sim (0, \sigma_i^2)$. It has $C_i(1) = 1 + \vartheta_i$, $C_i^* = -\vartheta_i$ and $w_{it} = \vartheta_i e_{it-1}$. In this model, the variance of series i at $t = T_0$ is

$$\text{var}(x_{i,T_0}) = (1 + \vartheta_i)^2\sigma_i^2 T_0 + \vartheta_i^2\sigma_i^2.$$

The cross-section variance at T_0 is

$$\text{var}_i(x_{i,T_0}) = \mathbb{E}_i[(1 + \vartheta_i)^2\sigma_i^2]T_0 + \mathbb{E}_i[\vartheta_i^2\sigma_i^2].$$

The first term is the between-unit variance of the unit-specific conditional mean, and the second term is the average within-unit variance. While the first term arises from dispersion in the shocks, the second term is due to parameter heterogeneity. If x_{it} has unit-specific mean μ_i, then $\mathbb{E}[\bar{X}_{iK}] = \mu$ and $\text{var}_i(\mu_i)$ contributes to cross section heterogeneity. The within-unit term contributes additionally through time series variation within units. Throughout, we assume that $\mathbb{E}[\bar{X}_{iK}] = x_{iT_0}$ for any given i, so the first term is due to cross-section variance in the initial condition, $\text{var}_i(x_{i,T_0})$. The second term in the IMA(1) fans out, being $K\mathbb{E}_i[(1 + \vartheta_i)^2\sigma_i^2]$. While a positive ϑ_i increases absolute cross-section variability, a negative one decreases it. Because the effect of the innovation is

[4] If $\Delta x_{is,t} = \delta_{is} + u_{is,t}$ and thus $x_{is,t} = x_{is,0}\delta_{is}t + \sum_{j=1}^{t} u_{is.t-j}$, the deterministic trend dominates the stochastic trend in the averaged data, and $T^{-1}\bar{X}_{is,T} \xrightarrow{d} N(0, 1/2)$ implying that if $x_{is,t}$ is a random walk with drift, and $\text{var}_{T|is}(T^{-1}\bar{X}_{is,T}) = \frac{\sigma_{is}^2}{2}$.

quickly reversed, the fanning out of the cross-section variance is weakened as $\mathbb{E}_i[\vartheta_i]$ moves towards -1. The mean-reversion becomes extreme when $\vartheta_i = -1$ for all i since $\Delta x_{it} = \Delta e_{it}$ can be reduced to $x_{it} = e_{it}$. By implication, the fanning out effect of the stochastic trend can be significantly dampened when ϑ_i is close to -1. Perron and Ng (1996) refer to this case as nearly integrated, nearly white noise. Fanning out of cross-section variance across units can be expected only if such cases are ruled out. The issue is relevant because for two plots in Figure 1, the ϑ for the U.S. series is -0.8133 and for the New York series is -0.8578.

Figure 2 plots the population-weighted cross-section variance of three measures of temperature from 1895 to 2023. The first, labeled V and plotted in black, is the cross-section variance of the temperature series itself. The second, labeled V_1 and plotted in blue, is cross-section variance in the non-stationary component based on the local-level model (μ_{it}) estimated for each state. The third, labeled V_2 and plotted in red, is the cross-section variance of the stationary component in the local-level model. We see that while V_1 trends up starting around 1980, V_2 fluctuates around a constant mean of around 1. The timing that V_1 trends up is consistent with the reported warming in temperature in the U.S. While V is dominated by the stationary (high frequency) fluctuations in temperature, they are not strong enough to offset the underlying (low-frequency) trend. Of interest is that V_2, the cross-section variance of the non-stationary component, trends up but the cross-section variance of the stationary component does not.

Remark If x_{it} is a unit root process with drift δ_i; i.e., $\Delta x_{it} = \delta_i + u_{it}$. Then $x_{it} = \delta_i t + \sum_{j=0}^{t} u_{it-j}$ is driven by both a deterministic and a stochastic trend. The fanning out effect continues to hold if $\delta_i = \delta \forall i$. In that case, $\text{var}_i(x_{it}) = \mathbb{E}_i[(x_{it} - \delta t)^2] = \sigma_i^2 t$, which is the same as the no drift case. But if there is variation in δ_i across i, the cross-section variance at t becomes $\text{var}_{i|t}(x_{it}) = \mathbb{E}_i[(x_{it} - \mathbb{E}_i(\delta)t)^2] = \text{var}_i(\delta_i)t^2 + \sigma_i^2 t$. The fanning out effect is amplified by heterogeneity in δ_i.

The fanning out effect suggests that if we perform 'time compression' by picking one time period in a $T \times N$ panel of data for analysis, or what will subsequently be referred to as 'point sampled', the variability is maximized by selecting the last time period, T. Then $\text{var}_i(x_{iT})$ will be of order T. But information in the remaining $T-1$ periods will be ignored. The next subsection considers an alternative 'time compression scheme.'

3 Time Compression by Averaging over K Periods

Let $\bar{X}_{i,T} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$ be the full sample average computed over all T observations. If x_{it} is stationary with long run variance ω_i^2, then $\sqrt{T}\bar{X}_{i,T} \xrightarrow{d} N(0,\omega_i^2)$. Then, $\frac{\text{var}(\bar{X}_{i,T})}{\text{var}(x_{it})} = \frac{\sigma_i^2}{T\sigma_i^2}$. Time averaging of stationary data reduces noise relative to an arbitrary draw x_{it}. But if $\Delta x_{it} = u_{it}$ with $u_{it} \sim (0,\sigma_i^2)$ is stationary and has long-run variance ω_i^2, it follows that $T^{-1/2}\bar{X}_{i,T} \Rightarrow \omega_i^2 \int_0^1 W(r)d(r)$, where $W(r) \sim N(0,r)$ is a standard Brown motion and $E[\int_0^1 W(r)dr]^2 = 1/3$. Thus, if x_{it} is a random walk, then for large T,

$$\text{var}(\bar{X}_{i,T}) = \omega_i^2 \frac{T}{3}.$$

In contrast to the statiionary case when the variance of $\bar{X}_{iT}$ decreases with T, the variance of $\bar{x}_{iT}$ now diverges with T. It follows that the cross-section of $\bar{X}_{iT}$ also fans out with T, but not by as much as x_{iT}.

To better understand the shrinkage factor of 1/3, consider averaging $K = T_1 - T_0$ observations for some $T_1 > T_0$. Then,

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

Because x_{it} has a unit root, computing $\bar{X}_{i,T_1-T_0}$ is akin to having a heterogeneous initial condition of x_{i,T_0}, and $\mathbb{E}[\bar{X}_{i,K}] = x_{i,T_0}$. But $\bar{X}_{i,K}$ is a triangular weighted sum of future shocks with more weights to the earlier periods. Since $\text{var}(u_{it}) = \sigma_i^2$ and $\sum_{j=1}^{K}\phi_j^2 = \frac{K(K+1)(2K+1)}{6K^2} = \frac{(K+1)(2K+1)}{6K}$,[5]

$$\text{var}(\bar{X}_{iK}) = T_0\omega_i^2 + c_K\omega_i^2, \qquad c_K = \frac{(K+1)(2K+1)}{6K}.$$

Lemma 1 *Suppose that $x_{it} = x_{it-1} + u_{it}$ where $u_{it} \sim (0,\sigma_i^2)$ is independent across i but serially correlated with long-run variance ω_i^2, and $u_{i0} = 0$. Let $\bar{X}_{i,K}$ be the mean of unit i computed from K observations between T_0 and T_1. For fixed $K = T_1 - T_0$, the cross-section variance is $var_i(\bar{X}_{i,K}) = (T_0 + c_k)\mathbb{E}_i[\omega_i^2]$, and for large K, $var_i(\bar{X}_{iK}) = \mathbb{E}_i[\omega_i^2]K/3$.* □

By the decomposition of total variance, the cross-section variance has two components: one from the variance of the conditional mean, which in this case is the initial condition, $\text{var}_i(x_{i,T_0}) = \sigma_i^2 T_0$,

[5] Changing the index in $\sum_{j=1}^{K}\phi_j^2$ to $k = K - j - 1$, gives $\sum_{j=1}^{K}\phi_i^2 = \frac{1}{K^2}\sum_{k=1}^{K}k^2 = \frac{K(K+1)(2K+1)}{6K^2}$.

and one from the mean of the conditional variance due to the shocks. We thus have $\text{var}_i(\bar{X}_{ik}) = K(T_0 + c_K)\mathbb{E}_i[\sigma_i^2]$ for fixed K. Now $c_K \approx K/3$ for large K, and assuming T_0 is small relative to K gives a large K cross-section variance of $\mathbb{E}_i[\omega_i^2]K/3$.

If x_{it} is a random walk, $T_0 = 0$, $T_1 = T = K$ in the case of full sample averaging, then $c_K \approx T/3$ and $\text{var}(\bar{X}_{iT}) = \mathbb{E}[\sigma_i^2]T/3$ as obtained earlier. The shrinkage factor of 1/3 still appears because averaging gives unequal weights ϕ_j to the shocks. Thus, Lemma 1 shows that introducing serial correlation in u_{it} only changes the result from the short-run variance σ_i^2 to the long-run variance ω_i^2. There can exist long-run variances that dampen the fanning out. Suppose that x_{it} is a IMA(1,1); i.e., $x_{it} = x_{it-1} + u_{it} + \vartheta u_{it-1}$. If ϑ_i is negative, $\omega_i^2 = (1+\vartheta_i)^2\sigma_i^2$; the fanning effect will significantly reduced if ϑ is close to -1. The MA(1) errors with a negative ϑ is of interest because the local level model and common factor model that fit temperature have an MA(1) model as reduced form. Both models have the potential for the cross-section variance to fan out, but for different reasons. In the local-level model, the driving force of fanning out is the N individual unit-root processes each dominating the respective serially correlated noise, and hence the signal-to-noise ratio is $q_i = \sigma_u^2/\sigma_w^2$. In the factor model, we only have a single stochastic trend. The fanning out is realized only if the trend is sufficiently strong and is complemented by heterogeneity in the loadings λ_i to dominate the idiosyncratic variations. Hence, the signal-to-noise ratio for the factor model is $q_i = \sigma_u^2\sigma_\lambda^2/\sigma_w^2$.

To illustrate the fanning out of the cross-section variance, we generate data from the local level and the factor models. However, both models have a non-stationary component, and no averaging should have three times larger variance than averaged data. Both models imply that

$$R_{10} = \frac{\text{var}_i(x_{i,T_1})}{\text{var}_i(x_{i,T_0})} = \frac{T_1}{T_0}.$$

We simulate x_{ist} for $i = 1, \ldots, N, s = 1, \ldots, S$, and $t = 1, \ldots, T$ with $(T_0, T_1, T_2) = (25, 50, 150)$, $(N, S) = (50, 100)$. For the local level model, we simulate w_{ist} and u_{ist} as two independent mean-zero Gaussian errors with $\sigma_w = 1$ and $\sigma_u = (1, 0.5, 0.2, 0.1, 0)$. For the factor model, we let $\lambda_{is} \sim N(0, \sigma_\lambda^2)$, $w_{sit} \sim N(0, 1)$ and $e_t \sim N(0, 1)$. We set $\sigma_\lambda = (1, 0.6, 0.4, 0.2, 0)$. When $\sigma_\lambda > 0$, the model has heterogeneous loadings. When $\sigma_\lambda = 0$, we set $\lambda = 0.5$ to arrive at a factor model with a non-stationary common factor with homogeneous loadings. In each case, the signal-to-noise ratio q_i is used to compute the implied ϑ_i. In each replication, we compute the variance of x using data T, N, S, and SN. The ratios (denoted R_{10}, R_{21} and R_{20}, respectively) are computed for T_1 relative to T_0, T_2 to T_1, and T_2 to T_0.

The top panel of Table 1 reports the ratios averaged over replications. The true variance ratios are 2, 3 and 6. Results for the local-level model are in the top panel. We see that the cross-sectional

Table 1: Variance ratios for (T_0, T_1, T_2)=(25,50,150), (S,N)=(100,50)

		T			N			S			SN		
		R_{10}	R_{21}	R_{20}	R_{10}	R_{21}	R_{20}	R_{10}	R_{21}	R_{20}	R_{10}	R_{21}	R_{20}
		2	3	6	2	3	6	2	3	6	2	3	6
σ_u	ϑ	Local level model											
1.000	-0.382	1.774	2.753	4.884	1.965	2.993	5.882	1.966	2.976	5.851	1.961	2.963	5.809
0.500	-0.610	1.503	2.334	3.509	1.869	2.873	5.370	1.864	2.865	5.343	1.861	2.853	5.310
0.200	-0.819	1.148	1.493	1.714	1.508	2.328	3.511	1.498	2.342	3.508	1.498	2.334	3.497
0.100	-0.905	1.045	1.143	1.194	1.204	1.652	1.989	1.197	1.669	1.999	1.198	1.667	1.997
0.000	-1.000	1.004	0.988	0.992	0.994	0.997	0.990	0.998	0.995	0.992	0.998	1.001	0.998
σ_λ	ϑ	Factor model											
1.000	-0.382	1.957	2.930	5.733	1.987	3.036	6.031	1.948	3.094	6.027	1.963	3.056	5.999
0.600	-0.554	1.961	2.945	5.775	1.924	2.966	5.706	1.890	3.031	5.729	1.903	2.989	5.688
0.400	-0.672	1.964	2.952	5.797	1.820	2.842	5.173	1.802	2.925	5.272	1.811	2.878	5.211
0.200	-0.819	1.967	2.959	5.819	1.507	2.377	3.582	1.574	2.581	4.062	1.569	2.522	3.958
0.000	-1.000	1.989	2.938	5.845	0.996	0.999	0.994	0.995	1.000	0.995	0.999	1.001	1.000

		SN-No averaging			SN-Centered average			SN-Partial average		
		V_0	V_1	V_2	V_0	V_1	V_2	V_0	V_1	V_2
σ_u	ϑ	Local level model								
1.000	-0.382	25.993	50.961	150.985	23.624	48.586	148.613	8.876	17.179	50.463
0.500	-0.610	7.249	13.491	38.497	5.990	12.231	37.238	2.249	4.310	12.620
0.200	-0.819	2.001	2.999	7.000	1.052	2.051	6.052	0.393	0.706	2.025
0.100	-0.905	1.252	1.500	2.500	0.346	0.596	1.597	0.128	0.192	0.511
0.000	-1.000	1.002	0.999	1.000	0.111	0.111	0.111	0.040	0.020	0.007
σ_λ	ϑ	Factor model								
1.000	-0.382	25.454	49.951	152.639	23.129	47.638	149.235	8.480	16.601	48.462
0.600	-0.554	9.803	18.620	55.592	8.397	17.221	53.797	3.079	5.989	17.450
0.400	-0.672	4.912	8.829	25.264	3.794	7.715	23.972	1.390	2.673	7.759
0.200	-0.819	1.978	2.956	7.066	1.032	2.012	6.077	0.378	0.683	1.945
0.000	-1.000	1.001	1.000	1.001	0.111	0.111	0.111	0.040	0.020	0.007

variances N, S and SN are closely in line with the results in Lemma 1 when σ_u, and hence the signal-to-noise ratio, is relatively high. This reflect the fanning out of the cross-sectional variance. When the dynamics of x is determined entirely of the dynamics of the stationary component w (i.e., $\sigma_u = 0$), the cross-sectional variance does not increase with T. Results for the factor model are given in the lower panel. Similarly to the local level model, the cross-sectional variances based on cases N, S and SN mimic these true ratios closely but in this common factor setup, the fanning out is driven by the heterogeneity of the factor loadings. When the factor loadings are homogeneous ($\sigma_\lambda = 0$), while the time-series variance (T case) is still expanding due to the non-stationarity of the common factor, the cross-sectional variance is no longer fanning out because this non-stationarity is common for all cross-sectional units.

The fanning out effect extends to $S \times N$ panels after the time dimension is collapsed. The

bottom panel of Table 1 shows the average of the variances over simulations. Comparing the variance of the averaged data and that of the raw data, the ratio is roughly 1/3. It would then seem that averaging is undesirable if the goal is to obtain data with maximum variance. The next section shows that averaging the data is desirable for estimation.

4 Implications for Cross-Section Regressions

In this section, we analyze the implications of using time averaged data for regressions. The first subsection analyzes the case when x_{it} is I(1). The second subsection studies the case when x_{it} is I(0).

4.1 Non-Stationary x

We consider a setting in which x_{it} is non-stationary, and y_{it} depends on x_{it} and possibly lags of y_{it} and x_{it}. We start with the static case:

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

$$y_{it} = \beta x_{it} + e_{it}, \tag{1b}$$

It is assumed that $u_{it} \sim (0, \sigma_i^2)$ and $e_{it} \sim (0, \xi_i^2)$, where u_{it} is independent of e_{it}. We know that a regression of using $T \times N$ panel data of y and x, or a time series regression using T observations of y and x of a particular unit will converge at the fast rate of T. While this super consistent rate gives precise estimates, the asymptotic distribution of $\hat{\beta}$ is, in general, non-standard. Instead, we consider a 'timeless' cross-section regression.

To motivate, consider a projection of $\bar{Y}_{iK}$ on $\bar{X}_{iK}$. The 'bias' from the projection is

$$\text{bias} = \frac{\text{cov}_i(\bar{X}_{iK}, \bar{e}_{iK})}{\text{var}_i(\bar{X}_{iK})}.$$

From the preceding analysis, $\bar{X}_{iK} = O(K^{1/2})$ because time averaging allows $\bar{X}_{iT}$ to be increasingly dominated by accumulated permanent shocks. If $e_{it} \sim (0, \xi^2)$, then $\bar{e}_{iK} = O(K^{-1/2})$. Because the signal in $\bar{X}_i$ grows while the noise $\bar{e}_{iK}$ decreases with K, the bias vanish with K.

We now analyze the (cross-section) between-group regression

$$\bar{Y}_{iK} = c + \beta \bar{X}_{iK} + \bar{e}_{ik} \tag{S}$$

and study the sampling properties of the least squares estimator

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

Proposition 1 *(SS) For $i = 1, \dots N$ and $t = 1, \dots T$, let the data be generated by (1a) and (1b). For each i, let $(\bar{X}_{iK}, \bar{Y}_{iK}, \bar{e}_{iK})$ be the average of (y_{it}, x_{it}, e_{it}) between $t = T_0 + 1$ and T_1. Then,*

- *As $N \to \infty$ with K fixed, $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0,\ Q_{XX}^{-1} V_K Q_{XX})$, where $c_K = \frac{(K+1)(2K+1)}{6K}$, $V_K = \mathbb{E}_i[(\bar{X}_{iK} - \mu_X)^2 \bar{e}_{iK}]$ and $Q_{XX} = var_i(\bar{X}_{iK}) = (T_0 + c_K)\mathbb{E}_i[\sigma_i^2]$.*
- *For large K, $\sqrt{N}K(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \mathcal{Q}_{XX}^{-1} \mathcal{V} \mathcal{Q}_{XX}^{-1})$ where $\mathcal{Q}_{XX} = \lim_{K\to\infty} K^{-1} Q_{XX}$, and $\mathcal{V} = \lim_{K\to\infty} V_K$.*

The proof is given in the Appendix. In the fixed K case, the variability in initial condition $\bar{x}_{i,T0}$ matters. The variance of $\sqrt{N}(\hat{\beta} - \beta)$ thus tends to zero at rate K^2. This is because the numerator of $\hat{\beta} - \beta$ is a sum of $O(1)$ terms, while the denominator is a sum of $O(K^2)$ terms. For this model, $\mathcal{V} = \mathbb{E}_i[\sigma_i^2 \xi_i^2]/3$ and $\mathcal{Q}_{XX} = \mathbb{E}[\sigma_i^2]/3$. Thus, for $K > 1$,

$$\sqrt{N}K(\hat{\beta} - \beta) \xrightarrow{d} N(0, \frac{3\mathbb{E}_i[\sigma_i^2 \xi_i^2]}{(\mathbb{E}_i[\sigma_i^2])^2}).$$

If $K = 1$, we obtain the standard cross-section result that $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \frac{\mathbb{E}_i[x_{i,T_0} \xi_i^2]}{(\mathbb{E}_i[x_{i,T_0+1}^2]^2)})$.

For large K, terms related to x_{iT_0} vanish. The cross-section 'between' estimator is super-consistent in the sense that the rate is faster than the usual rate of $\sqrt{N}$. Intuitively, the variation in $\bar{X}_{iK}$ accumulates as K grows while that of $\bar{e}_{iK}$ shrinks, so that $\text{var}_i(\bar{X}_{iK}) = O(K)$ but $\text{var}_i(\bar{e}_{iK}) = O(K^{-1})$. If u_{it} and e_{it} have homogeneous variance, $\sqrt{N}K(\hat{\beta} - \beta) \sim (0, 3\xi^2/\sigma_u^2)$. Of note is that the estimator is asymptotically normal. Proposition 1 implies that the most efficient cross-section regression is to perform full sample averaging. With $K = T$, $\hat{\beta}$ is $\sqrt{NT}$ consistent and asymptotically normal despite the fact that x_{it} is I(1) and we are performing a cross-section regression.

Proposition 1 assumes that the static regression (S) correctly specifies the simple data generating process. Now suppose that x_{it} continues to be a random walk as given in (1a), but now let

$$y_{it} = \alpha y_{it-1} + \beta_0 x_{it} + \beta_1 x_{it-1} + e_{it}, \qquad e_{it} \sim (0, \xi_i^2). \tag{2b}$$

Following Bewley (1979), (2b) can be rearranged as

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

where $\alpha^* = \alpha/(1-\alpha), \beta_1^* = \beta_1/(1-\alpha)$ and $e_{it}^* = e_{it}/(1-\alpha)$. The quantity

$$\theta^\dagger = \frac{\beta_0+\beta_1}{1-\alpha}$$

is often referred to as the long-run multiplier (LRM). A regression of y on x omits $\Delta y_{it}, \Delta x_{it-1}$ and $\mathbb{E}_i[\eta_{it}x_{it}] \neq 0$. Thus if x_{it} were stationary, a time series or a panel regression would yield inconsistent estimates. If x_{it} is I(1), a time series regression would yield super consistent estimates which converge at rate T because the omitted variable bias is of second order but the distribution of $\hat{\beta}$ is non-standard. Now in terms of time averaged data,

$$\bar{Y}_{iK} = \theta^\dagger \bar{X}_{iK} - \eta_{iK} + \frac{\bar{e}_{iK}}{1-\alpha}.$$

What would a least squares regression of $\bar{Y}_{iK}$ on $\bar{X}_{iK}$ yield?

Proposition 2 *Suppose that x_{it} is I(1), y_{it} is generated by the dynamic model given in (2b) with $|\alpha| < 1$, and $\mathbb{E}[u_{it}e_{is}] = 0$ for all (t,s). Consider least squares estimation of the static regression (S) using $\bar{Y}_{iK}$ and $\bar{X}_{iK}$.*

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

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

The omitted lags collected into η_{iK} in the time averaged regression are of the same order as the I(0) regressors. The omitted variable bias is $\frac{\mathrm{cov}_i(\bar{X}_{iK},\bar{e}_{iK}^*-\eta_{iK})}{\mathrm{var}_i(\bar{X}_{iK})}$ for fixed K. Since $\bar{X}_{iK}$ only depends on $\{u_{it}\}$ and $u_{it} \perp e_{is}$ by assumption, we only need to consider $\mathrm{cov}_i(\bar{X}_{iK}, \eta_{iK})$. As $\bar{X}_{iK} = O_p(K^{1/2})$ and $\eta_{iK} = O_p(K^{-1/2})$, their product is $O(1)$ and upon scaling by $\mathrm{var}(\bar{X}_{iK})$, we have a bias that is $O_p(K^{-1})$ that vanishes with K. We thus have a $\sqrt{N}K$ consistent estimate of the long-run multiplier as K increases. Though $\bar{e}_{iK}^*$ does not contribute to the bias, it affects the variance as is indicated in W_K. Because $e_{it}^* = \frac{e_{it}}{1-\alpha}$ and $\alpha < 1$, $\mathrm{var}(\bar{e}_{iK}) > \mathrm{var}(\bar{e}_{iK})$.

The preceding analysis assumes that $T \times N$ panels of y and x are available. Suppose we have data $(x_{is,t}, y_{is,t})$ for $i = 1, \dots N$ units in each location $s = 1, \dots S$ for $t = 1, \dots T$, a panel fixed effect regression

$$\bar{Y}_{is,K} = \beta X_{is,K} + \delta_i + \mu_s + \bar{\mathcal{E}}_{iK}$$

will result in $\sqrt{NS}K(\hat{\beta} - \beta) \xrightarrow{d} N(0, \mathcal{V})$ as $N, S \to \infty$. Effectively, we have a regression in NS observations of centered data $\tilde{X}_{is} = \bar{X}_{is} - \bar{X}_{i.} - \bar{X}_{.s} + \bar{X}_{..}$, and $\tilde{Y}_{is}$ is similarly defined.

4.2 Stationary x

The above subsection shows that time averaging data over K-periods yield $\sqrt{N}K$ consistent estimates of the estimand in the static regression, which can be β or θ depending on the model for y_{it}. Can we also consistently estimate β and θ when x_{it} is stationary? In this sub-section, we assume that

$$x_{it} = \rho x_{it-1} + u_{it}, \quad u_{it} \sim (0, \sigma_i^2), \tag{1b}$$

where $|\rho_i| < 1$, u_{it} has long-run variance

$$\omega_i^2 = \frac{\sigma_i^2}{(1-\rho_i)^2} > 0.$$

Recall that in the I(1) case, $\bar{X}_{iK} = O_p(K^{1/2})$. In the present I(0) case, $\bar{X}_{iK} = O_p(K^{-1/2})$. We start with estimation of the static model (S).

Proposition 3 *Suppose that x_{it} is generated as in (1b) and $|\rho| < 1$, with $\mathbb{E}[u_{it}e_{is}] = 0$ for all (t, s). Assume that $\mathbb{E}_i[\omega_i^2] > 0$. Consider least squares estimation of the static model (S).*

- *If y_t is generated by (1b) with $|\alpha| < 1$, then for all K, $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(0, Q_{XX}^{-1} V_K Q_{XX}^{-1})$ where $V_K = var_i((\bar{X}_{iK} - \mu_X)\bar{e}_{iK})$ and with $Q_{XX} = var_i(\bar{X}_{iK})$.*
- *If y_{it} is generated by (2b), then $\sqrt{N}\left(\hat{\beta} - \theta^\dagger + \frac{cov_i(\bar{X}_{iK}, \eta_{iK})}{var_i(\bar{X}_{iK})}\right) \xrightarrow{d} N(0, Q_{XX}^{-1} W_K Q_{XX}^{-1})$ with K fixed, where $W_K = var_i((\bar{X}_{iK} - \mu_X)(\bar{e}_{iK} - \eta_{iK}))$. For large K, $\sqrt{N}(\hat{\beta} - \theta^\dagger) \xrightarrow{d} N(0, \mathcal{Q}_{XX}^{-1} \mathcal{W} \mathcal{Q}_{XX}^{-1})$.*

Proposition 3 has two parts, both requiring $\mathbb{E}_i[\omega_i^2] > 0$ for identification because $Q_{XX} = \mathbb{E}_i(\omega_i^2)/K$. Variability can come from heterogeneity in ρ_i or σ_i^2, or both. With homoskedastic errors, there is no cross-sectional variation from heterogeneous means, and the sandwich degenerates. When this

is ruled out in both cases, $Q_{XX} = O(K^{-1})$ while V_K and W_K are $O(K^{-2})$. The variance is finite for every K and the asymptotic distributions are valid for all K.

In part (i), the model for y_{it} coincides with the regression model. Then, the estimator is unbiased at any K and is $\sqrt{N}$ consistent. Averaging now reduces both the signal $\text{var}(\bar{X}_{iK})$ and the noise V_K at the same rate of $1/K$. As a consequence, we no longer achieve the super-consistent rate that is obtained when x_{it} is I(1). Since the asymptotic distribution also holds when $K = 1$, this means that a single cross-section regression using the data sliced at $T_0 + 1$ will have the same precision as averaging over many periods. It should be noted, however, that $\bar{X}_{iK} \approx \omega_i^2/K$ and hence the persistence of x_{it} also affects the precision of $\hat{\beta}$.

In part (ii) when the static regression is misspecified, the estimator is biased for the long-run multiplier when K is fixed. Now $\eta_{iK} = O_p(K^{-1})$ which tends to zero faster than in the I(1) case when $\eta_{iK} = O_p(K^{-1/2})$. Nonetheless, the bias shrinks slower than in the I(1) case. This is because $\text{var}(\bar{X}_{iK})$ is $O(K^{-1/2})$ when x is I(0) compared to $O(K^{-1})$ when x is the I(1). Since $\text{cov}_i(\bar{X}_{iK}, \eta_{iK}) = O(K^{-1/2})O(K^{-1})$, the bias of $O_p(K^{-1/2})$ is larger than in the I(1) case, but nonetheless tends to zero as K increases.

It has been suggested in the literature that averaging the data over many periods allows us to estimate the medium and longer term effects. Our analysis shows that K affects the asymptotic properties of the estimator, but the estimand is always the long-run multiplier as defined in the textbook; it has no 'medium' or 'longer term' interpretation. In summary, when y_{it} is generated by (2b), whether x is I(1) or I(0), the cross-section estimator of $\theta^\dagger$ is biased when K is fixed, but is consistent and asymptotically normal as K increases. However, $\hat{\theta}^\dagger$ is super-consistent only when x is I(1). Quantities other than the long-run multiplier can be used to characterize the dynamic effects of changes in x, or of shocks to x, but for that, we will need to estimate a dynamic model.

4.3 Cross-Section Regressions of Dynamic Models

We now suppose that y_{it} is generated as in (2b) and incorporate the lagged variables which we will reparameterize as

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + (\beta_0 + \beta_1)x_{it} - \beta_1 \Delta x_{it-1} + e_{it} \\ &= \alpha y_{it-1} + \delta_0 x_{it} + \delta_1 \Delta x_{it-1} + e_{it}. \end{aligned}$$

But x_{it} is generated by u_{it} which is uncorrelated with e_{it} by assumption, so bias can only arise from correlation between $y_{i,t-1}$ and e_{it}. In standard dynamic panel regressions, $\text{cov}(y_{i,t-1} - \bar{y}_{i,-1}, e_{it} -$

$\bar{e}_i) = O(T^{-1})$. As $\text{var}(y_{i,t-1} - \bar{y}_{i,-1}) = O(1)$, Nickell bias is $O(T^{-1})$.

Cross-section regressions involve data without a time dimension, but it can still be used to estimate dynamic models. In particular, we want to perform a between-group dynamic regression in time-aggregated data

$$\begin{aligned}
\bar{Y}_{iK} &= c + \alpha \bar{Y}_{iK,-1} + \beta_0 \bar{X}_{iK} + \beta_1 \bar{X}_{iK} + \bar{e}_{iK} && \text{(2a)} \\
&= c + \alpha \bar{Y}_{iK,-1} + \delta_0 \bar{X}_{iK} + \delta_1 \Delta \bar{X}_{i,K} + \bar{e}_{iK} && \text{(2b)} \\
&= c + \alpha \bar{Y}_{iK,-1} + \delta_0 \bar{X}_{iK} + \delta_1 \frac{(x_{i,T_1} - x_{i_{T_0}})}{K} + \bar{e}_{iK}.
\end{aligned}$$

The relevant term for Nickell bias is

$$NB = \frac{\text{cov}(\bar{Y}_{iK,-1} e_{ik})}{\text{var}_i(\bar{Y}_{iK,-1})}.$$

Since the numerator is $O(K^{-1})$, regardless of whether x is I(1) or I(0), the difference is in the denominator. In particular, $\text{var}_i(\bar{Y}_{iK,-1}) = O(K)$ in the I(1) case and $\text{var}(\bar{Y}_{iK,-1}) = O(K^{-1})$ in the I(0) case. We need to consider them separately.

Proposition 4 *Suppose y_{it} is generated as in (2b), with $u_{it} \perp e_{is}$ for all t, s and for all i.*

i If x_{it} is I(1), $var(\bar{Y}_{iK}) = O(K)$ so that NB=$O(K^{-2})$. Let $Z_{iK} = (\bar{Y}_{iK,-1}, \bar{X}_{iK}, \Delta x_i)'$, $B_1 = (\alpha, \delta_0, \delta_1)'$ and $D_K = diag(K, K, K^{-1})$. Then, estimation of (2b) yields

$$\sqrt{N} D_K (\hat{B}_1 - B_1) \xrightarrow{d} N(0, \mathcal{Q}_{ZZ}^{-1} \mathcal{V}_1 \mathcal{Q}_{ZZ}^{-1})),$$

where $\mathcal{Q}_{ZZ} = \lim_{K\to\infty} D_K^{-1} Q_{ZZ} D_K^{-1}$ with $Q_{ZZ} = var_i(\bar{Z}_{iK})$ and $\mathcal{V}_1 = \lim_{K\to\infty} D_K^{-1} V_{1K} D_K^{-1}$, $V_{1K} = var_i(\bar{Z}_{iK} \bar{e}_{iK})$.

ii. If x_{it} is generated by (1b) and $\omega_i^2 = \sigma_i^2/(1-\rho_i)^2 > 0$, then $var_i(\bar{Y}_{iK}) = O(K^{-1})$ and NB=$O(1)$. Let $Z_{iK} = (\bar{X}_{iK}, X_{ik,-1})'$ and $B_0 = (\beta_0, \beta_1)'$. Least squares estimation of (2a) yields

$$\sqrt{N}(\hat{B}_0 - B_0) \xrightarrow{d} N(0, \tilde{\mathcal{Q}}_{ZZ}^{-1} \tilde{\mathcal{V}}_0 \tilde{\mathcal{Q}}_{ZZ}^{-1}),$$

where $\tilde{Z}_{iK}$ projects out Y_{iK} from Z_{iK},$\mathcal{Q}_{\tilde{Z}\tilde{Z}} = \lim_{K\to\infty} Q_{\tilde{Z}\tilde{Z}}/K$ with $Q_{\tilde{Z}\tilde{Z}} = var_i(\tilde{Z}_{iK})$ and $\mathcal{V} = \lim_{K\to\infty} \tilde{V}_{0K}/K^2$, $\tilde{V}_{0K} = var_i(\tilde{Z}_{iK} \tilde{e}_{iK})$.

Proposition 4 states that the bias in $\hat{\alpha}$ from least squares estimation vanishes as K increases in the I(1) case but not in the I(0) case. Although the numerator is the same in both cases, the signal

$\text{var}(\bar{Y}_{iK})$ grows with K and overwhelms the noise in the I(1) case but not in the I(0) case. Though the estimates in the I(1) case are consistent, we have two non-stationary regressors $(\bar{Y}_{iK}, \bar{X}_{iK})$ and one stationary regressor $\Delta x_i = \frac{x_{i,T_1} - x_{i,T_0}}{K}$. The three estimates in $\hat{B}$ converge at different rates, with $\hat{\alpha}$ and $\hat{\delta}_0$ being $\sqrt{N}K$ consistent and $\hat{\delta}_1$ being $\sqrt{N}$ consistent.

If x_{it} is $I(0)$, y_{it} is also stationary, so that the Nickell bias is now $O(1)$ and $\hat{\alpha}$ is not consistent, nor can a bias-correction be made. Recall that the standard within-estimator has $\text{cov}(\ddot{y}_{it}, \dot{e}_{it})$ is $O(T^{-1})$. In the present setting,

$$\text{cov}(\bar{Y}_{iK,-1}, \bar{e}_{iK}) = \frac{\tau_i^2}{K^2(1-\alpha)}\left[K - \frac{1-\alpha^K}{1-\alpha}\right].$$

Averaging dilutes the covariance from K to T but does not eliminate it. If $K = T$ as in the case of full sample averaging, the covariance would be the same order as the within estimator. But while the denominator $var(\ddot{y}_{i,t-1}) = O(1)$ in within estimation to give a Nickell bias is $O(T^{-1})$, the denominator is also $O(K^{-1})$ giving a bias of $O(1)$ in the present setting, It is nonetheless useful to remark that $\text{cov}(\bar{Y}_{i,K-1}, \bar{e}_{iK})$ is zero when $K = 1$ and increases with K. The Nickell bias might not be severe in the I(0) case.

In spite of the bias in $\hat{\alpha}$, we can consistently estimate β_0 and β_1 with a rate of $\sqrt{N}$ when x is I(0). It is interesting that in this stationary case, estimation of the static model, though misspecified, would consistently estimate the long-run multiplier, θ. Yet, when we correctly specify the model, we can only estimate two of the three parameters. This is because in the static regression, $\bar{X}_{iK}$ is the only regressor and it is exogenous, but in the dynamic model, $\bar{Y}_{iK}$ is an endogeneous regressor correlated with the error. While simple averaging cannot eliminate Nickel bias, observe that it is zero when x is I(1). This implies that if we also specify a process x_{it}, computing impulse responses from the parameter estimates will not be an issue when x is I(1), but can be an issue when x is I(0) because coefficients on the lagged dependent variable will be biased. We will explore this in simulations.

Putting the results of the I(1) and I(0) case together, cross-section estimation of a static or a dynamic model can be consistent for the corresponding the long- and short-run effects of x_t without knowing a priori whether x is I(1) or I(0), and the estimates will be asymptotically normal.

5 A Class of Time Compression Functions

The above analysis uses time averaging to collapse the time variation, but other time compression schemes are possible. We now let $\tilde{X}_{iK} = \sum_{t=T_0}^{T_1} w(t)x_{it}$ for some time compression function $w(t)$. K period averaging puts $w(t) = 1/K$ for $t \in [T_0 + 1, T_1]$ while point sampling at $t = T_1$ puts $w(t) = 1$ at $t = T_1$ and zero others. Other weighting schemes are possible.

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

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

where ϕ_s is the sum of the total weight placed on periods at or after s. The sequence $\{\phi_s\}$ is thus the survival function of the weight distribution w, starting at $\phi_{T_0+1} = 1$ to $\phi_{T_1} = \tilde{w}(T_1)$. The time compressed variable $\tilde{X}_{iK}$ depends on the initial condition and the shocks, its variability $\text{var}(\tilde{X}_{iK}) = \sigma_i^2 \Phi_K$ is dominated by Φ_K, shown in the Appendix to be $O(K)$. We thus have

$$\Phi_K = \sum_{s=T_0+1}^{T_1} \phi_s^2 = \tilde{c}_K K,$$

where $\tilde{c}_K$ depends on the time compression scheme. For K period averaging, $\tilde{c}_K = 1/3$, and for point sampling, $\tilde{c}_K = 1$.

In the stationary case, we have, by definition,

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

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

$$\text{var}(\tilde{X}_{iK}) = \frac{\sigma_i^2}{1-\rho_i^2} \sum_{h=-(K-1)}^{(K-1)} \sum_{t=-s=h} w(t)w(s)\rho_i^{|t-s|} \equiv \sigma_i^2 \Psi_K(\rho).$$

The function $\psi_s(\rho_i)$ plays the role of ϕ_s in the unit root case, and the Appendix shows that

$$\Psi_K = O(K^{-1}).$$

The result is immediate for K period averaging because the variance of the mean of a covariance stationary process is $O(K^{-1})$. Interestingly, $\psi_s \to \phi_s$ as $\rho_i \to 1$. More precisely, in the local-to-unity case when $\rho_i = 1 - \kappa_i/K$, $\frac{1}{1-\rho_i^2} = \frac{1}{(1-\rho_i)(1+\rho_i)} \propto \frac{K}{2\kappa_i}$. Thus, $\Psi_K = O(K)$ and behaves like Φ_K. This allows $\text{var}(\bar{X}_i)$ to transition smoothly from $O(K^{-1})$ (stationary) to $O(1)$ (unit root).

Lemma 2 *Let $w(t)$ be a time compression function that satisfies (i) $w(t) \geq 0$ for all t and (ii) $\sum_{t=T_0+1}^{T_1} w(t) = 1$. Let also $\bar{X}_{iK} = \sum_{t=T_0+1}^{T_1} w(t) x_{it}$. Then, $\psi_s(\rho_i) \xrightarrow{\rho_i \to 0} w(s)$ and $\psi_s(\rho_i) \xrightarrow{\rho_i \to 1} \phi_s$.*

The two conditions on $w(t)$ ensure that it is a proper probability distribution over the interval $[T_0, T_1]$. Part (i) says that the tail sum ϕ_s interpolates continuously between the two extremes, with the implication that $\sum \psi(\rho_i)^2 \to \sum_t w(t)^2 = O(1/K)$ when $\rho_i = 0$, and $\sum \phi_s^2 = O(K)$ when $\rho_i = 1$.

Suppose now $y_{it} = \beta x_{it} + e_{it}$, where $u_{is} \perp e_{it}$ for all i, s, t, $\tilde{(}X_{iK}, \tilde{e}_{iK})$ are iid over i with finite fourth moments. Let $\tilde{\beta}$ be obtained from the cross-section regression $\tilde{Y}_{iK} = c + \beta \tilde{X}_{iK} + \tilde{e}_{iK}$ using time compressed data $(\tilde{Y}_{iK}, \tilde{X}_{iK})$. By strict exogeneity,

$$\text{cov}_i(\tilde{X}_{iK}, \tilde{e}_{iK}) \quad = \quad \sum_s \phi_s \sum_t w(t) \mathbb{E}_i[u_{it} e_{is}] = 0.$$

We can thus have consistent estimates for any K, but for large K, the convergence rate is $\sqrt{NK}$ when x is I(1), and the usual rate of $\sqrt{N}$ when it is I(0). Then, when x is I(1), $\tilde{\beta}$ has the properties given in Proposition 1 with c_K replaced by $\tilde{c}_K$, and when x is I(0), $\tilde{\beta}$ has the same properties as given in part (i) of Proposition 3. Crucially, the asymptotic distribution is continuous in the persistence parameter ρ. The fact that $\text{var}_i(\tilde{X}_{iK}) = O(K)$ when ρ_i is close to one is important because $\text{var}_i(\tilde{X}_{iK})$ is Q_{XX}. Hence, Q_{XX} transitions smoothly from $O(1/K)$ when $\rho_i = 0$ to to $O(K)$ when $\rho_i = 1$ and in the neighborhood of 1. This is unlike time series regression where the asymptotic distribution is not continuous in ρ.

The above fixed K results assume estimation of a correctly specified model, by which we mean that error term of the regression is uncorrelated with the regressor. Though the results do not apply to estimation of a static model when the DGP is dynamic because of correlation between η_{iK} and $\bar{X}_{iK}$, favorable large K results continue to hold. The long-run multiplier is $\sqrt{NK}$ consistent and asymptotically normal as in Proposition 2 for the I(1) case, and $\sqrt{N}$ consistent and asymptotically normal as in Proposition 3 for the I(0) case, and the asymptotic distribution moves smoothly between the two cases. Though the results do not apply to estimation of ADL models because of

non-vanishing Nickel bias in the I(0) case even when K is large, the results carry over to estimation of a distributed lag model without the autoregressive part.

6 Long Differencing

The class of functions satisfying Lemma 2 is large, but there are also time compression functions that do not fit into that class. Simple long differencing and long differencing of averaged data (LD) are two examples. LD is based on a difference of the local averaging of $M = 2m + 1$ observations centered at T_0 and T_1, with $K = T_1 - T_0$. Unlike K period averaging, it requires an additional parameter, m. The popularity of LD could be due to the fact that it gives the regression a diff-in-diff flavor. In spite of its popularity, the properties of these regressions are not well understood. Burke and Emerick (2016) simulate x_{it} as having a deterministic trend and show that if $K = T_1 - T_0$ is sufficiently large, the LD estimates from a cross-section regression are more precise than the ones obtained from a $T \times N$ panel of raw data.[6] We now analyze the LD assuming that x_{it} has a stochastic trend which seems to be a better characterization of the data.

We will now argue that Lemma 2 does not apply to LD because the condition $\sum_t w(t) = 1$ is not satisfied. An example makes this clear. Suppose that $T_0 = 4, T_1 = 10$, $m = 1$ and thus $M = 3$.

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

Several features are of note: (i) $\tilde{w}(t) < 0$ for some t, (ii) $\sum_t w(t) = 0$. Hence both conditions required for Lemma 2 to hold are violated. Furthermore, (iii) the initial condition x_{i,T_0} is not used; (iv) $\tilde{\phi}_s$ jumps from 0 to $1/M$ at $T_0 - m + 1$, and from $1/M$ to 0 at $T_1 + m + 1$.

A LD transformation is actually the same filter considered in Kutznets (1958). Howrey (1968) shows that the 20-year cycle documented in Kutznets (1958) could be an artifact of long-differencing time-averaged data, even if the raw data were a white noise.[7] For stationary data, it is known that averaging attenuates the high frequency variations and differencing attenuates the low frequency variations. For regressions, Engle and Liu (1972) show that a dynamic regression using averaged

[6] Burke and Emerick (2016) simulated $x_{it} = \theta t + w_{it}$ 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. Arguably, it is the mean of $X^2_{is,t}$ that should be computed, not the square of the mean.

[7] Amemiya and Wu (1972) shows that if x_t is a stationary AR(p), time averaging overlapping data $\bar{X}_t$ is ARMA(p,q), where q is at most p.

data will yield biased estimates because of correlation between the error term and the regressors. What makes the LD work is that transformation is applied only at $t = T_0$ and T_1 with the effect of collapsing the time variations into $\tilde{X}_{iK} = \bar{X}_{iM}(T_1) - \bar{X}_{iM}(T_0)$, and then using the cross-section variation in estimation. This is unlike in time series regression based on dynamic correlations that are distorted by the transformation.

The above example shows that for I(1) variables, LD puts a weight of 1 on the $K - M + 1$ innovations that are in the middle, while the observations at the two end points have smaller weights. Generalizing the pattern,[8] we obtain

$$\text{var}(\tilde{X}_{iM}) = \omega_i^2 \bigg((K - M + 1) + 2 \sum_{j=1}^{2m} (j/M)^2 \bigg) = \omega_i^2 \bigg(K - \frac{M^2 - 1)}{3M} \bigg).$$

If we let $K \to \infty$ keeping M and T_0 in LD fixed, then $\text{var}(\tilde{X}_{iK}) = \omega_i^2 K$, which is three times larger than $\text{var}(\bar{X}_{iK}) = \omega_i^2 K/3$ from simple averaging. This is a consequence of the fact that when $K \to \infty$ with M fixed, $K - M + 1$ innovations will receive a full weight of 1.

It is useful to compare LD with simple averaging when both use the maximum of the sample size, being T. With simple averaging, this requires setting $T_0 = 1$ and $T_1 = T$. With $K = T$, we obtain $\text{var}(\bar{X}_{iK}) = \omega_i^2 c_T$. With LD, the full use of the sample is a span of $T - 2m$, obtained by setting $T_0 = m$ and $T_1 = T - m$, and $\text{var}(\tilde{X}_{iK}) = \omega_i^2 (T - M + 1 - \frac{M^2-1}{3M}$. The LD variance is smaller than K averaging if and only if $M < 3T/4$, or $m < T/3$. Once the window width of $M = 2m + 1$ is larger than $3T/4$, there will not be enough observations that receive unit weighting to realize the variance advantage over simple averaging. The choice of m is thus important. In this regard, we note that Dell et al. (2012) report 'medium-term' effects using two 15-year averages positioned 15 years apart (i.e., $m = 7$ and $K = 15$), which implies $K - M$ is small. In contrast, Burke and Emerick (2016) report 'longer-term effects' using $m = 2$ and $K = 20$, and thus $K - M$ is much larger.

Though the signal in $\text{var}_i(\tilde{X}_{iK})$ is larger, $\tilde{e}_{iK} = \bar{e}_{iM}(T_1) - \bar{e}_{iM}(T_0)$, $\text{var}_i(\tilde{e}_{iK}) = 2\xi_i^2/M$ and $\text{var}_i(\bar{e}_{iK}) = \xi_i^2/K$, the LD has more noise. Nonetheless, the LD does seem to have efficiency advantage with a suitable choice of M. If x is I(1), a static LD regression will give $\sqrt{NK}$ consistent estimates of β if the DGP is a static model and $\sqrt{NK}$ consistent estimates of the long-run multiplier if the true model is a dynamic model. If we perform a dynamic regression and the DGP is dynamic, Nickell bias still precludes consistent estimation of α, but we will get $\sqrt{NK}$ consistent estimates of $\delta_0 = \beta_0 + \beta_1$ and of $\delta_1 = -\beta_0$. The convergence rates are identical to simple averaging.

[8] We have used $2 \sum_{j=1}^{2m} (j/M)^2 = \frac{(M-1)(2M-1)}{3M}$ and $M = 2m + 1$ to simplify.

While the LD seems appealing when x is I(1), it has a major drawback when x is I(0). Consider the example above, but now $x_{it} = \rho x_{it-1} + u_{it}$ with $\rho < 1$ and thus $x_{it} = \sum_{r=0}^{\infty} \rho^r u_{i,t-r}$ with $\bar{X}_{iM}(T_j) = \frac{1}{3}(x_{i,T_j-1} + x_{i,T_j} + x_{i,T_j+1})$. The weight on innovation u_{is} in $\bar{X}_{iM}(T_j)$ for $s \leq T_j - 1$ is:

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

Denoting $\kappa = (1+\rho+\rho^2)/3$, the weights on u_{is} in $\tilde{X}_{iK} = \bar{X}_{iM}(T_1) - \bar{X}_{iM}(T_0)$ 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, $\psi_s(\rho) = \sum_{t \geq s} w(t)\rho^{t-s}$ $w(s) = w_1 - w_0$. For $\rho = 1$, this reduces to a weight of 1 for $s = 6, 7, 8, 9$ obtained earlier. For $\rho < 1$, innovations d periods from the T_1 window receive weight ρ_i^d in $\bar{X}_{iM}(T_1)$, and the weight becomes negligible when ρ_i^d is small. In the end, $\text{var}(\bar{X}_{iM}(T_j)) \approx \omega_i^2/M$ as predicted by the theory for covariance stationary processes, but the rate does not change with K. Accordingly, the Q_{XX} is $O(M)$ for any K. The variance of $\tilde{X}_{iK}$ 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, $\text{var}(\tilde{X}_{iK}) = 2\sigma_i^2/M$. In a sense, the effective sample size of an LD regression is M instead of K when x is stationary. As such, it does not enjoy the 'robustness' of simple averaging to whether x is I(1) or I(0). For example, the omitted variable $\tilde{\eta}_{iK}$ is now correlated with $\tilde{X}_{iK}$. The fix K bias from estimating the long run multiplier using $\bar{X}_{iK}$ is now $O(M^{-1})$ and does not vanish with K.

7 Simulations

This section evaluates estimates from time series, cross-section, and panel regressions using time averaged data. We first consider univariate and panel regressions with a time dimension. We then specify a bivariate model where x_{it} may not be stationary or exogenous. Finally, we entertain models that specify x_{it} as the sum of a stationary and a non-stationary component .

We consider the following regressions (which can be static or dynamic):

1. T: time series regressions with T observations at give $(i, s) = (i_*, s_*)$.

2. S or N: : cross section regressions each with N or S observations of averaged data.
3. TN: panel fixed effect regression using $T \times N$ data matrices at given $i = i^*$.
4. SN: panel effect regression using a $S \times N$ data matrices at given $t = t^*$.
5. LD_{01}: long-difference regression using the difference of averaged data centered at T_0 and T_1

For each $i = 1, \ldots, N$, $s = 1, \ldots, S$ and $t = 1, \ldots, T$, data are generated as

$$\begin{aligned} y_{is,t} &= \beta_\tau \tau_{is,t} + \beta_w w_{is,t} + e_{is,t} \\ x_{is,t} &= \tau_{is,t} + w_{is,t}, \\ w_{is,t} &= \rho w_{is,t-1} + \varepsilon_{is,t}, \end{aligned}$$

where $e_{is,t} \sim N(0,1)$, and $\varepsilon_{is,t} \sim N(0,1)$. We set $\rho = 0$ or 0.5, $\beta_\tau = -1$ and $\beta_w = -2.5$. The nonstationary component $\tau_{is,t}$ is either μ_{ist} in the local level model with innovation variance σ_u^2, or the common component where the random walk μ_t has unit innovation variance and loadings $\lambda_i \sim (0, \sigma_\lambda^2)$. As discussed above, the relative importance of the two components is driven by the signal-to-noise ratio. Thus, we consider different values of σ_u^2 and σ_λ^2 and the implied moving-average parameter ϑ.

Table 2 presents estimates and corresponding standard errors, obtained from 1,000 replications, from several cross-section regressions for different values of σ_u (or σ_λ), ρ and the implied ϑ. The true values of the regression coefficients are $(\beta_\tau, \beta_w) = (-1, -2.5)$ and the sample size is $(T, N, S) = (50, 50, 50)$. The estimates from the infeasible regression on $\tau_{is,t}$ and $w_{is,t}$, which are not observed by the econometrician, are labeled ORACLE. In practice, the econometrician observes only $x_{is,t}$ and not its individual components. The feasible regression of y on x yields the cointegrating vector. As $\text{cov}(x, y) = \beta_\tau \text{var}(\tau) + \beta_w \text{var}(w)$, we have

$$\beta_X = \beta_\tau \frac{\text{var}(\tau)}{\text{var}(x)} + \beta_w \frac{\text{var}(w)}{\text{var}(x)}.$$

'S' denotes a cross-section regression of S observations sampled at $T_1 = 20$, $T = 50$ and averaged from 1 to T. 'S-CA' denotes a cross-section regression of centered averaged $k = 9$ observations around $T_0 = 10$ and $T_2 = 40$ while 'S-PA' stands for cross-section regressions with partial averaging between T_0 and T_1, and T_1 and T_2. Finally, '$LD_{T_0,T_2}(k)$' is a long-differenced regression between T_2 and T_0, with averaging of m observations around these two time periods.

Even though the non-stationary variations should dominate when T is large, in finite samples the least squares estimate $\hat{\beta}_X$ will be a weighted sum of β_μ with weights that depend on the

signal-noise-ratio, with an indicator being ϑ. Regardless of data used, we see that $\hat{\beta}_\tau$ is close to β_τ when ϑ is away from -1, but is contaminated by β_w as ϑ moves closer to -1. Even though all data give somewhat similar estimates, the standard errors are largest when the raw data are used and smallest with partial averaging, especially if the averaging is performed over the whole time dimension $1:T$. Overall, averaging and using information later in the sample helps while long differencing adds little benefit. The finite-sample size and power properties of a t-test based on these methods are presented in Table 3. These results suggest that inference, based on the limiting distribution of these estimators and tests, appears to be quite reliable.

Table 4 reports estimates and standard errors from panel regressions (SN panel) with fixed effects. The methods are denoted similarly as in Table 2 with 'S' replaced by 'SN'. Again, the SN regression with averaged data over the whole time period $(1:T)$ appears to estimate most accurately the coefficient β_τ on the nonstationary component with the estimates being contaminated with β_w as the signal from the nonstationary component (σ_u and σ_λ) decreases.

Finally, Table 5 presents the results from a dynamic panel estimation of the bivariate dynamic model along with the implied estimates of $\theta^\dagger$ and θ^*. 'SN-ADL(1,1)' reports the estimates from the first (ADL(1,1)) equation and 'SN-System' also reports the estimates from the second equation. Overall, the model parameters and long-run effects are very precisely estimated and are largely robust to the value of ρ. It is also notable that there is practically no downward bias in $\hat{\rho}$ which, in turn results in very accurate estimates of θ^*.

8 Conclusions

Temperature data tend not to be stationary, and estimating the effect of temperature on growth at any horizon must confront this fact. This paper considers regressions using time-compressed data in this context. The main result is that if the regressor is nonstationary, time compression (removing the T dimension of the data), such as by time averaging, can yield (super-)consistent and asymptotically normal estimates which obviates the need for unit-root asymptotics. The key mechanism is that the paths of the I(1) variables across units 'fan out' with time. Time averaging amplifies the signal in the I(1) regressor and dampens the stationary regression noise. A result of general interest is a class of time compression functions that allow smooth transition of the asymptotic distribution from I(1) to I(0) so that standard normal inference can be used without a priori knowledge of whether the regressor is I(1) or I(0). We use simulations to verify the finite-sample properties of estimates in static and dynamic regressions using time compressed data.

Appendix

Suppose that the scalar series 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 $\frac{\bar{X}}{\sqrt{T}} = T^{-3/2} \sum_{t=1}^{T} x_t \Rightarrow \omega_v^2 \int_0^1 W(r)dr = A$. 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).

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.

Proof of Lemma 1

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

$$\begin{aligned}
\bar{X}_{iK} &= x_{i,T_0} + \frac{1}{K} \sum_{t=1}^{K} \sum_{j=1}^{t} u_{i,T_0+j} \\
&= x_{i,T_0} + \frac{1}{K} \sum_{j=1}^{K} (K - j + 1) u_{i,T_0+j} \\
&= x_{i,T_0} + \sum_{j=1}^{K} \phi_j u_{i,T_0+j}, \qquad \phi = \frac{K - j + 1}{K},
\end{aligned}$$

since each u_{i,T_0+j} appears in all sums with $t > j$. With $k = K - j + 1$, $\sum_{j=1}^{K} \phi_j^2 = \frac{1}{K^2} \sum_{k=1}^{K} k^2 = K(K+1)(2K+1)/6$. We thus have $\text{var}(\bar{X}_{iK}) = T_0 \omega_i^2 + c_K \omega_i^2$.

Note that $\mu_X \equiv \mathbb{E}_i[\bar{X}_{iK}] = \mathbb{E}_i[x_{i,T_0}]$, because u_{it} have mean zero. Thus, $\text{Var}_i(x_{i,T_0}) = \mathbb{E}_i[(x_{i,T_0} - \mu_X)^2]$ is the cross-section variance of initial conditions. Therefore, $\text{var}_i(\bar{X}_{iK}) = \text{var}_i(x_{i,T_0}) + c_K \mathbb{E}_i[\omega_i^2]$.

Proof of Proposition 1

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:

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

Since $E[u_{is}e_{it}] = 0$ for all t, s by assumption, a CLT applies to the numerator $\frac{1}{\sqrt{N}} \sum_i (\bar{X}_{iK} - \bar{\bar{X}})\bar{e}_{iK}$ and gives a mean zero process with variance:

$$V_K = \mathbb{E}_i\left[(\bar{X}_{iK} - \mu_X)^2 \, \bar{e}_{iK}^2\right] = \mathbb{E}_i\left[(\bar{X}_{iK} - \mu_X)^2 \, \frac{\tau_i^2}{K}\right] = \frac{c_K \, \mathbb{E}_i[\omega_i^2 \tau_i^2] + \mathbb{E}_i[(x_{i,T_0} - \mu_X)^2 \tau_i^2]}{K}$$

since $u_{it} \perp e_{is}$ by assumption. Combining the results for the numerator and denominator gives the variance stated in the Proposition.

Part (ii): $K \to \infty$ With large K, $c_K = (K+1)(2K+1)/(6K) \approx K/3$ 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, $\mathcal{V} = \mathbb{E}_i[\omega_i^2 \tau_i^2]/3$ and $\mathcal{Q}_{XX} = \mathbb{E}_i[\omega_i^2]/3$, giving an asymptotic variance of $3\mathbb{E}_i[\omega_i^2 \tau_i^2]/(\mathbb{E}_i[\omega_i^2])^2$.

Proof of Proposition 2:

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

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

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

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

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,

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

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

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

and

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

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

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

Proof of Proposition 3

We now let $x_{it} = \rho_i x_{it-1} + u_{it}$, where $u_{it} \sim (0, \sigma_i^2)$ and $\omega_i^2 = \frac{\sigma_i^2}{(1-\rho_i^2}$ is the long-run variance of x_{it}. It follows from the theory of covariance stationary processes that $\text{var}(\bar{X}_{iK}) = \frac{\omega_i^2}{K}$, and $\bar{X}_{iK} = O_p(K^{-1/2})$.

For part (i) when the static model is correctly specified, it follows from $\text{cov}_i(\bar{X}_{iK}, \bar{e}_{iK}) = 0$ that $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(, Q_{XX}^{-1} V_K Q_{XX}^{-1})$, where $V_K = \text{var}_i((\bar{X}_{iK} - \mu_X)\bar{e}_{iK})$ for all K.

For part (ii), the omitted variables are collected in

$$\eta_{iK} = \frac{\alpha \Delta y_{iK} + \beta_1 \Delta x_{iK}}{1 - \alpha} = O_p(K^{-1})$$

since $\Delta x_{iK} = \frac{1}{K}(x_{i,T_1} - x_{i,T_0}) = O_p(K^{-1})$, and similarly, $\Delta y_{iK} = O_p(K^{-1})$. Though the bias is shrinking faster than in the I(1) case, $\bar{X}_{iK} = O_p(K^{-1/2})$, so the omitted variable bias is $OK(K^{-1/2})$, which is larger than in the I(1) case.

Proof of Proposition 4

We consider the reparameterized model

$$\bar{Y}_{iK} = \alpha \bar{Y}_{iK,-1} + \underbrace{(\beta_0 + \beta_1)}_{|delta_0} \bar{X}_{iK} \underbrace{-\beta_1}_{\delta_1} \Delta \bar{X}_{iK} + \bar{e}_{iK}$$

Since $Y_{iK} = O(K^{1/2})$ when x and thus y is non-stationary, Nickell bias in the I(1) case is

$$NB = \frac{\text{cov}_i(\bar{Y}_{iK,-1}, \bar{e}_{iK})}{\text{var}_i(\bar{Y}_{iK,-1})} = \frac{O(K^{-1})}{O(K)} = O(K^{-2}) \to 0.$$

However, in the I(0) case, x is stationary and thus $\text{var}_i(\bar{Y}_{iK,-1}) = O(K^{-1})$. Furthermore, analogous to dynamic panel regressions,

$$\text{cov}(\bar{Y}_{iK,-1}, \bar{e}_{iK}) = \frac{\tau_i^2}{K^2(1-\alpha)}\left[K - \frac{1-\alpha^K}{1-\alpha}\right]$$

which is $O(K^{-1})$ for large K. Thus, in the I(0) case, NB=O(1) and does not change with K.

Proof of Lemma 2

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:

$$\tilde{X}_i = x_{i,T_0} + \sum_{s=T_0+1}^{T_1} \phi_s \, u_{is}, \qquad \phi_s \equiv \sum_{t=s}^{T_1} w(t), \tag{3}$$

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,

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

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

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

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 $A_h > 0$ and $\sum_h A_h = (\sum_t w(t))^2 = 1$, so $\{A_h\}$ forms a probability distribution over the lags. Thus, $G(\rho_i) = \mathbb{E}_A[\rho_i^{|h|}$. 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. 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 $\rho_i = 1 - \kappa_i/K$, $\omega_i^2 = K^2\sigma^2/c_i^2$, 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.

Figures and Tables

Table 2: Cross-Section Regressions using Averaged Data: $(\beta_\tau, \beta_w) = (-1, -2.5)$, $(T, S, N) = (50, 50, 50)$

DGP: $y_{is,t} = \beta_\tau \tau_{is,t} + \beta_w w_{is,t} + e_{is,t}$,
cross-section regression: $y_s = \beta_x x_s + error$

Local-Level Model

σ_u	ρ	ϑ	Oracle		S			S-CA		S-PA		$LD_{T_0,T_2}(m)$		
			$\hat{\beta}^*_\tau$	$\hat{\beta}^*_w$	T_1	T	$1:T$	$T_0 \pm k$	$T_2 \pm k$	$T_0 : T_1$	$T_1 : T_2$	$m=0$	$m=4$	$m=9$
1.00	0.00	-0.38	-1.000	-2.501	-1.074	-1.029	-1.002	-1.020	-1.005	-1.012	-1.003	-1.093	-1.013	-1.007
0.50	0.00	-0.61	-1.000	-2.501	-1.253	-1.111	-1.008	-1.075	-1.018	-1.044	-1.012	-1.313	-1.049	-1.027
0.20	0.00	-0.82	-1.001	-2.501	-1.835	-1.500	-1.045	-1.367	-1.103	-1.233	-1.068	-1.929	-1.258	-1.151
0.10	0.00	-0.90	-1.002	-2.500	-2.251	-2.005	-1.161	-1.843	-1.339	-1.633	-1.236	-2.294	-1.679	-1.462
1.00	0.50	-0.67	-1.001	-2.503	-1.096	-1.039	-1.007	-1.064	-1.015	-1.038	-1.011	-1.122	-1.042	-1.025
0.50	0.50	-0.79	-1.002	-2.502	-1.320	-1.145	-1.028	-1.226	-1.058	-1.140	-1.041	-1.391	-1.153	-1.093
0.20	0.50	-0.91	-1.005	-2.501	-1.942	-1.602	-1.155	-1.785	-1.299	-1.581	-1.221	-2.027	-1.616	-1.435
0.10	0.50	-0.95	-1.009	-2.501	-2.309	-2.096	-1.469	-2.219	-1.746	-2.073	-1.611	-2.340	-2.102	-1.928
			S.E.											
1.00	0.00	-0.38	0.021	0.066	0.066	0.041	0.010	0.034	0.016	0.025	0.013	0.075	0.027	0.020
0.50	0.00	-0.61	0.041	0.041	0.118	0.078	0.021	0.067	0.032	0.050	0.025	0.129	0.053	0.039
0.20	0.00	-0.82	0.103	0.030	0.187	0.151	0.051	0.139	0.078	0.113	0.062	0.189	0.116	0.090
0.10	0.00	-0.90	0.207	0.029	0.206	0.190	0.097	0.185	0.136	0.173	0.115	0.200	0.171	0.148
1.00	0.50	-0.67	0.035	0.092	0.070	0.043	0.016	0.052	0.024	0.038	0.020	0.079	0.039	0.030
0.50	0.50	-0.79	0.070	0.051	0.121	0.082	0.033	0.094	0.048	0.072	0.039	0.131	0.072	0.057
0.20	0.50	-0.91	0.175	0.030	0.177	0.149	0.077	0.137	0.102	0.128	0.087	0.177	0.123	0.111
0.10	0.50	-0.95	0.349	0.026	0.185	0.175	0.120	0.130	0.137	0.137	0.130	0.178	0.130	0.136

Factor Model

σ_λ	ρ	ϑ	Oracle		S			S-CA		S-PA		$LD_{T_0,T_2}(m)$		
			$\hat{\beta}^*_\tau$	$\hat{\beta}^*_w$	T_1	T	$1:T$	$T_0 \pm k$	$T_2 \pm k$	$T_0 : T_1$	$T_1 : T_2$	$m=0$	$m=4$	$m=9$
1.00	0.00	-0.38	-1.000	-2.499	-1.383	-1.255	-1.060	-1.202	-1.095	-1.178	-1.093	-1.419	-1.159	-1.114
0.60	0.00	-0.55	-0.999	-2.501	-1.531	-1.363	-1.093	-1.303	-1.141	-1.262	-1.139	-1.587	-1.239	-1.173
0.40	0.00	-0.67	-0.999	-2.501	-1.687	-1.486	-1.134	-1.414	-1.198	-1.351	-1.189	-1.750	-1.335	-1.246
0.20	0.00	-0.82	-0.998	-2.500	-1.925	-1.697	-1.216	-1.610	-1.317	-1.512	-1.286	-1.990	-1.512	-1.389
0.00	0.00	-1.00	-0.003	-2.501	-2.496	-2.498	-2.496	-2.495	-2.499	-2.497	-2.495	-2.500	-2.499	-2.497
1.00	0.50	-0.38	-1.000	-2.502	-1.406	-1.272	-1.114	-1.341	-1.159	-1.295	-1.160	-1.453	-1.270	-1.206
0.60	0.50	-0.55	-1.000	-2.501	-1.593	-1.410	-1.182	-1.510	-1.253	-1.430	-1.242	-1.653	-1.417	-1.325
0.40	0.50	-0.67	-0.999	-2.501	-1.756	-1.543	-1.254	-1.663	-1.351	-1.558	-1.324	-1.822	-1.557	-1.444
0.20	0.50	-0.82	-0.999	-2.501	-1.996	-1.766	-1.393	-1.900	-1.537	-1.774	-1.477	-2.060	-1.786	-1.652
0.00	0.50	-1.00	-0.001	-2.501	-2.500	-2.501	-2.498	-2.497	-2.500	-2.499	-2.498	-2.500	-2.499	-2.499
			S.E.											
1.00	0.00	-0.38	0.028	0.058	0.447	0.391	0.203	0.350	0.268	0.343	0.270	0.444	0.315	0.269
0.60	0.00	-0.55	0.044	0.042	0.476	0.443	0.246	0.411	0.311	0.406	0.321	0.474	0.371	0.321
0.40	0.00	-0.67	0.063	0.033	0.480	0.472	0.291	0.454	0.354	0.452	0.366	0.469	0.419	0.369
0.20	0.00	-0.82	0.103	0.025	0.439	0.480	0.361	0.486	0.413	0.492	0.426	0.414	0.469	0.431
0.00	0.00	-1.00	0.070	0.020	0.146	0.152	0.143	0.142	0.140	0.144	0.141	0.150	0.142	0.146
1.00	0.50	-0.38	0.045	0.089	0.450	0.404	0.270	0.422	0.321	0.422	0.338	0.455	0.385	0.344
0.60	0.50	-0.55	0.073	0.057	0.478	0.455	0.334	0.468	0.380	0.474	0.397	0.475	0.442	0.407
0.40	0.50	-0.67	0.104	0.042	0.471	0.477	0.383	0.477	0.420	0.491	0.438	0.456	0.468	0.445
0.20	0.50	-0.82	0.171	0.029	0.414	0.469	0.438	0.439	0.457	0.474	0.476	0.383	0.460	0.468
0.00	0.50	-1.00	0.074	0.018	0.129	0.133	0.073	0.077	0.077	0.077	0.073	0.126	0.076	0.077

Table 3: Rejection Rates for t-test for $H_0 : \beta = -1$

β	ρ	ϑ	S			S-CA		S-PA		$LD_{T_0,T_2}(m)$		
			T_1	T	$1:T$	$T_0 \pm k$	$T_2 \pm k$	$T_0 : T_1$	$T_1 : T_2$	$m = 0$	$m = 4$	$m = 9$
-1.00	0.00	-0.00	0.071	0.075	0.072	0.059	0.068	0.089	0.079	0.080	0.086	0.090
-0.99	0.00	-0.00	0.067	0.083	0.354	0.095	0.186	0.121	0.228	0.073	0.119	0.162
-0.97	0.00	-0.00	0.101	0.163	0.937	0.232	0.651	0.359	0.819	0.114	0.320	0.500
-0.95	0.00	-0.00	0.231	0.462	1.000	0.631	0.989	0.824	1.000	0.193	0.782	0.947
-1.00	0.50	-0.50	0.071	0.074	0.073	0.058	0.068	0.090	0.079	0.078	0.086	0.089
-0.99	0.50	-0.50	0.067	0.083	0.353	0.096	0.185	0.121	0.229	0.073	0.118	0.163
-0.97	0.50	-0.50	0.101	0.163	0.937	0.232	0.651	0.356	0.819	0.115	0.319	0.500
-0.95	0.50	-0.50	0.230	0.465	1.000	0.631	0.989	0.824	1.000	0.193	0.781	0.947

Table 4: Panel Regressions Using Averaged Data: $(\beta_\tau, \beta_w) = (-1, -2.5)$, $(T, S, N) = (50, 50, 50)$

Local Level Model

σ_u	ρ	ϑ	Oracle		SN			SN-CA		SN-PA		$LD_{T_0,T_2}(m)$		
			$\hat{\beta}^*_\tau$	$\hat{\beta}^*_w$	T_1	T	$1:T$	$T_0 \pm k$	$T_2 \pm k$	$T_0 : T_1$	$T_1 : T_2$	$m=0$	$m=4$	$m=9$
1.00	0.00	-0.38	-1.001	-2.503	-1.072	-1.029	-1.002	-1.019	-1.004	-1.011	-1.003	-1.094	-1.012	-1.007
0.50	0.00	-0.61	-1.003	-2.501	-1.251	-1.111	-1.007	-1.074	-1.017	-1.042	-1.011	-1.316	-1.048	-1.027
0.20	0.00	-0.82	-1.007	-2.500	-1.834	-1.500	-1.042	-1.368	-1.101	-1.230	-1.066	-1.937	-1.256	-1.151
0.10	0.00	-0.90	-1.013	-2.500	-2.251	-1.999	-1.156	-1.848	-1.336	-1.629	-1.233	-2.304	-1.677	-1.463
1.00	0.50	-0.67	-1.002	-2.505	-1.094	-1.039	-1.007	-1.064	-1.015	-1.037	-1.010	-1.122	-1.041	-1.024
0.50	0.50	-0.79	-1.004	-2.502	-1.317	-1.144	-1.026	-1.226	-1.057	-1.137	-1.040	-1.393	-1.151	-1.093
0.20	0.50	-0.91	-1.011	-2.501	-1.938	-1.600	-1.152	-1.789	-1.296	-1.577	-1.220	-2.034	-1.618	-1.437
0.10	0.50	-0.95	-1.022	-2.501	-2.304	-2.091	-1.465	-2.224	-1.743	-2.071	-1.611	-2.348	-2.106	-1.932
			S.E.											
1.00	0.00	-0.38	0.020	0.068	0.009	0.006	0.001	0.005	0.002	0.004	0.002	0.010	0.004	0.003
0.50	0.00	-0.61	0.041	0.043	0.017	0.011	0.003	0.009	0.005	0.007	0.004	0.018	0.008	0.005
0.20	0.00	-0.82	0.102	0.033	0.027	0.022	0.007	0.019	0.011	0.016	0.009	0.027	0.017	0.013
0.10	0.00	-0.90	0.204	0.031	0.029	0.028	0.013	0.026	0.019	0.024	0.017	0.028	0.025	0.021
1.00	0.50	-0.67	0.034	0.092	0.010	0.006	0.002	0.007	0.003	0.005	0.003	0.011	0.006	0.004
0.50	0.50	-0.79	0.069	0.052	0.017	0.012	0.004	0.013	0.007	0.010	0.006	0.019	0.011	0.008
0.20	0.50	-0.91	0.172	0.032	0.026	0.022	0.010	0.019	0.014	0.018	0.013	0.025	0.018	0.016
0.10	0.50	-0.95	0.344	0.028	0.027	0.027	0.017	0.018	0.019	0.019	0.018	0.025	0.019	0.019

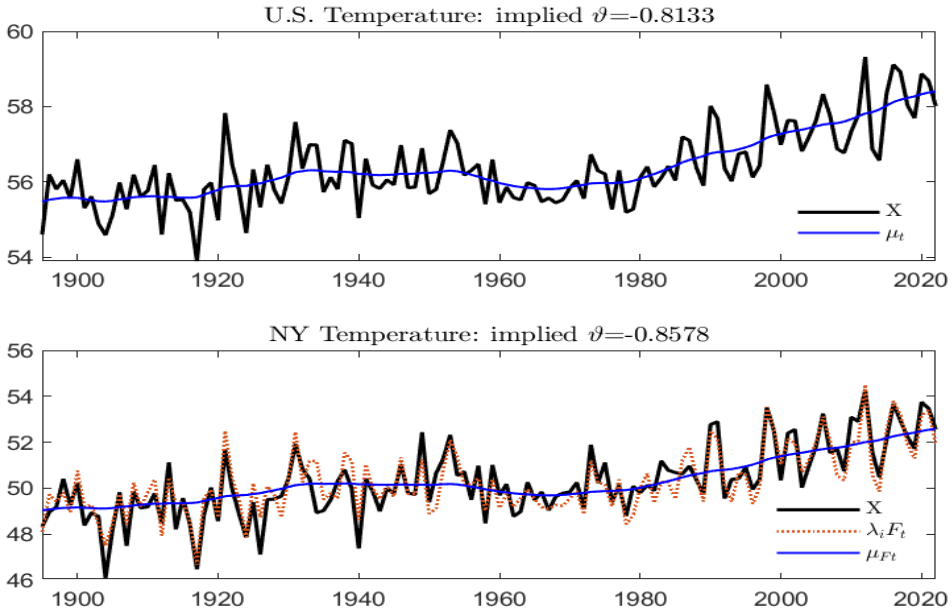
Factor Model

σ_λ	ρ	ϑ	Oracle		SN			SN-CA		SN-PA		$LD_{T_0,T_2}(m)$		
			$\hat{\beta}^*_\tau$	$\hat{\beta}^*_w$	T_1	T	$1:T$	$T_0 \pm k$	$T_2 \pm k$	$T_0 : T_1$	$T_1 : T_2$	$m=0$	$m=4$	$m=9$
1.00	0.00	-0.38	-1.001	-2.503	-1.072	-1.029	-1.002	-1.019	-1.004	-1.011	-1.003	-1.094	-1.012	-1.007
0.50	0.00	-0.61	-1.003	-2.501	-1.251	-1.111	-1.007	-1.074	-1.017	-1.042	-1.011	-1.316	-1.048	-1.027
0.20	0.00	-0.82	-1.007	-2.500	-1.834	-1.500	-1.042	-1.368	-1.101	-1.230	-1.066	-1.937	-1.256	-1.151
0.10	0.00	-0.90	-1.013	-2.500	-2.251	-1.999	-1.156	-1.848	-1.336	-1.629	-1.233	-2.304	-1.677	-1.463
1.00	0.50	-0.67	-1.002	-2.505	-1.094	-1.039	-1.007	-1.064	-1.015	-1.037	-1.010	-1.122	-1.041	-1.024
0.50	0.50	-0.79	-1.004	-2.502	-1.317	-1.144	-1.026	-1.226	-1.057	-1.137	-1.040	-1.393	-1.151	-1.093
0.20	0.50	-0.91	-1.011	-2.501	-1.938	-1.600	-1.152	-1.789	-1.296	-1.577	-1.220	-2.034	-1.618	-1.437
0.10	0.50	-0.95	-1.022	-2.501	-2.304	-2.091	-1.465	-2.224	-1.743	-2.071	-1.611	-2.348	-2.106	-1.932
			S.E.											
1.00	0.00	-0.38	0.020	0.068	0.009	0.006	0.001	0.005	0.002	0.004	0.002	0.010	0.004	0.003
0.50	0.00	-0.61	0.041	0.043	0.017	0.011	0.003	0.009	0.005	0.007	0.004	0.018	0.008	0.005
0.20	0.00	-0.82	0.102	0.033	0.027	0.022	0.007	0.019	0.011	0.016	0.009	0.027	0.017	0.013
0.10	0.00	-0.90	0.204	0.031	0.029	0.028	0.013	0.026	0.019	0.024	0.017	0.028	0.025	0.021
1.00	0.50	-0.67	0.034	0.092	0.010	0.006	0.002	0.007	0.003	0.005	0.003	0.011	0.006	0.004
0.50	0.50	-0.79	0.069	0.052	0.017	0.012	0.004	0.013	0.007	0.010	0.006	0.019	0.011	0.008
0.20	0.50	-0.91	0.172	0.032	0.026	0.022	0.010	0.019	0.014	0.018	0.013	0.025	0.018	0.016
0.10	0.50	-0.95	0.344	0.028	0.027	0.027	0.017	0.018	0.019	0.019	0.018	0.025	0.019	0.019

Table 5: Dynamic Panel SN Regressions in Averaged Data

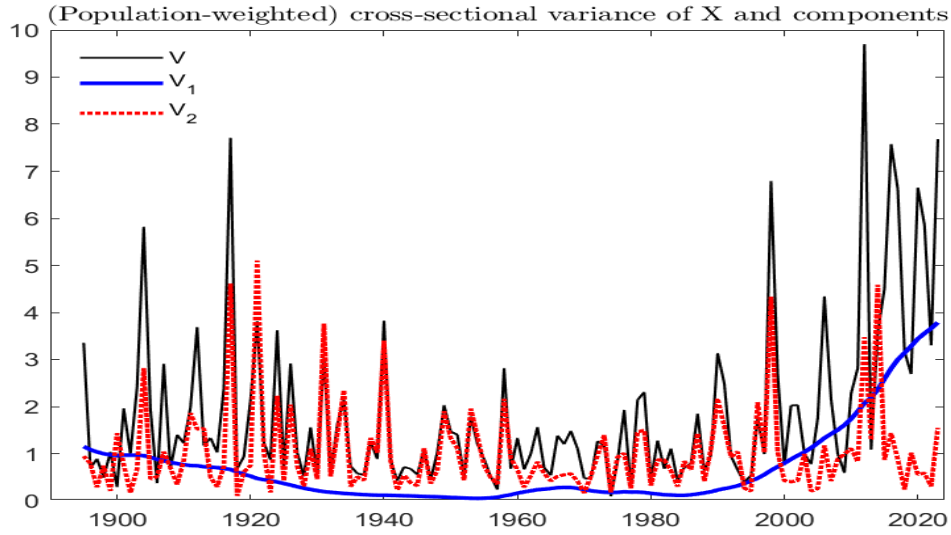
					SN-ADL(1,1)				SN-System		
ρ	α	γ	$\theta^{\dagger}$	θ^{*}	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}$	$\hat{\theta}^{\dagger}$	$\hat{\rho}$	$\hat{\gamma}$	$\hat{\theta}^{*}$
0.0	0.0	0.0	-1.5	-1.5	-0.999	-0.499	0.001	-1.500	-0.002	-0.000	-1.499
0.5	0.0	0.0	-1.5	-3.0	-0.999	-0.500	0.001	-1.500	0.498	-0.001	-3.000
0.8	0.0	0.0	-1.5	-7.5	-0.999	-0.501	0.001	-1.500	0.798	-0.001	-7.523
0.9	0.0	0.0	-1.5	-15.0	-0.999	-0.501	0.000	-1.500	0.898	-0.001	-15.124
1.0	0.0	0.0	-1.5	-Inf	-0.999	-0.501	0.000	-1.500	0.988	-0.001	-297.584
0.0	0.0	0.1	-1.5	-1.3	-0.999	-0.499	0.001	-1.500	-0.001	0.100	-1.303
0.5	0.0	0.1	-1.5	-2.3	-0.999	-0.500	0.001	-1.500	0.498	0.099	-2.307
0.8	0.0	0.1	-1.5	-4.3	-0.999	-0.501	0.001	-1.500	0.798	0.099	-4.289
0.9	0.0	0.1	-1.5	-6.0	-0.999	-0.501	0.000	-1.500	0.898	0.099	-6.011
1.0	0.0	0.1	-1.5	-10.0	-0.999	-0.501	0.000	-1.500	0.988	0.099	-9.451
0.0	0.5	0.0	-3.0	-3.0	-0.999	-0.500	0.501	-3.002	-0.001	-0.000	-3.002
0.5	0.5	0.0	-3.0	-6.0	-0.999	-0.500	0.500	-3.002	0.498	-0.001	-6.013
0.8	0.5	0.0	-3.0	-15.0	-0.999	-0.501	0.500	-3.001	0.798	-0.001	-15.086
0.9	0.5	0.0	-3.0	-30.0	-0.999	-0.501	0.500	-3.000	0.898	-0.001	-30.339
1.0	0.5	0.0	-3.0	-Inf	-0.999	-0.501	0.500	-3.000	0.990	-0.000	-473.554
0.0	0.5	0.1	-3.0	-2.3	-0.999	-0.500	0.500	-3.002	-0.001	0.100	-2.308
0.5	0.5	0.1	-3.0	-3.8	-0.999	-0.500	0.500	-3.002	0.498	0.100	-3.753
0.8	0.5	0.1	-3.0	-6.0	-0.999	-0.500	0.500	-3.002	0.798	0.099	-6.010
0.9	0.5	0.1	-3.0	-7.5	-0.999	-0.501	0.500	-3.002	0.898	0.099	-7.515
1.0	0.5	0.1	-3.0	-10.0	-0.999	-0.501	0.500	-3.001	0.990	0.099	-9.769
		med. se SN(T)			0.020	0.030	0.016	0.023	0.024	0.016	0.384
		med. se SN(5)			0.021	0.033	0.016	0.032	0.025	0.016	0.440

Figure 1: U.S. Annual Temperature Data



;

Figure 2: Fanning Out of Cross-Section Variance in x and its Components



References

Amemiya, T. and Wu, R. 1972, The effect of aggregation on prediction in autoregressive model, *Journal of American Statistical Association* **67**, 628–632.

Auffhammer, M. 2018, Quantifying Economic Damages from Climate Change, *Journal of Economic Perspectives* **32:4**, 33–52.

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

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

Carleton, T., Duflo, E., Jack, K. and Zappala, G. 2024, Adaptation to Climate Change, NBER Working paper 33264.

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

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.

Dell, M., Jones, B. and Olken, B. 2014, What do we Learn from the Weather? The New Climate-Economy Literature, *Journal of Economic Literature* **52:3**, 740–798.

Engle, R. F. and Liu, T. 1972, Effects of Aggregation over Time on Dynamic Characterization of an Econonometric Model, *in* B. G. Hickman (ed.), *Econometric Models of Cyclical Behavior*, UMI, pp. 673–739.

Estrada, F., perron, P. and Yamamoto, Y. 2023, On the Persistence of Near-Surface Temperature in a Warming World, *Annals of the New York Academy of Science* **1531**, 69–83.

Hendry, D. F. 1995, *Dynamic Econometrics*, Oxford University Press, Oxford.

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

Hsiang, S. 2016, Climate Econometrics, *Annual Review of Resource Economics* **8**, 43–75.

Kolstad, C. and Moore, F. 2020, Estimating the Economic Impacts of Climate Change Using Weather Observations, *Review of Environmental Economics and Policy* **14:1**, 1–24.

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

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

Nath, I., Ramey, V. and Klenow, P. 2024, How Much will Global Warming Cool Global Growth, NBER Workign Paper 32761.

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.

Reed, W. R. and Zhu, M. 2017, On Estimating Long-Run Effects in Models with Lagged Dependent Variables, **64**, 302–311.

Sahm, C. 2019, Direct Stimulus Payments to Individuals, *in* H. Shambaugh, R. Boushey and J. Nunn (eds), *Recession Ready: Fiscal Policies to Stabilize the American Economy*, Brookings Institute.