A Guide to Analyzing Large N, Large T Panel Data

Ryan P. Thombs

Abstract
Fixed effects estimation of a static model with robust or panel corrected standard errors is commonly used to model large N, large T panel data. However, this approach is biased and inconsistent in the presence of dynamic misspecification, slope heterogeneity, and cross-sectional dependence. Common correlated effects estimation of a dynamic model has been advanced to address these issues but is rarely used in sociology. Here, I provide an overview of the large N, large T panel data literature, and I conduct an array of Monte Carlo experiments to compare the fixed effects estimator to the common correlated effects estimator regarding the aforementioned issues. I show that fixed effects estimation with robust or panel corrected standard errors do not address these problems, which is most evident with high levels of slope heterogeneity and lag misspecification, and its performance worsens as the time dimension expands. In contrast, the common correlated effects estimator produces superior estimates as T increases and is robust to slope heterogeneity and cross-sectional dependence. Following the experiments, I present an example by examining the drivers of fossil fuel consumption at the U.S. state level from 1960 to 2018, and I conclude by presenting a decision-making framework for researchers to use to make informed decisions when modeling large N, large T panel data.

Keywords
panel data, quantitative methods

Introduction
Large N (cross-sections), large T (time periods) data is an increasingly common data structure analyzed by sociologists. In statistical theory, large N, large T panel data are a data structure where N and T grow at the same rate to infinity (Ditzen 2018; Pesaran 2006; Pesaran and Smith 1995). In practice, this means that T is large enough so that a time-series regression can be estimated for each unit in the analysis and N is large enough so that averaging across units produces consistent results and central limit theorems hold (Pesaran and Smith 1995). In this article, an N and T of at least 30 is considered large.

This type of data structure is common in macro-level research that analyzes cross-national or meso-level (states and provinces) data. For instance, the World Bank’s (2021) world development indicators database extends back to 1960 in many cases, as does state-level energy and income data in the United States (Bureau of Economic Analysis 2021; U.S. Energy Information Administration 2021). Although these data are typically annual, more granular temporal data are available on a monthly, weekly, or daily basis in some cases (e.g., financial, employment, and industry-level data).

As interest in big data increases and the temporal depth of widely used data sets expands, large N, large T data are becoming more ubiquitous, making it paramount that researchers have the appropriate tools to analyze these data. To date, most statistical work in this area comes from econometrics (Chudik and Pesaran 2015a; Pesaran 2006; 1This lower bound should not be followed too strictly. It is unlikely that an N or T less than 30 would be considered large, but it is possible if the model is rather parsimonious. If the model is fairly complex and includes a large number of regressors, then particularly for T, a number larger than 30 may be required so that there are appropriate degrees of freedom available.

1Boston College, Chestnut Hill, MA, USA

Corresponding Author:
Ryan P. Thombs, Department of Sociology, Boston College, 426 McGuinn Hall, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA.
Email: thombs@bc.edu

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Socius: Sociological Research for a Dynamic World

Pesaran and Smith 1995) and has yet to be widely used in sociology. Part of this is due to these advances not being readily available in popular software programs, but recently, particularly in Stata, a plethora of user-written commands emerged that eases this burden (Bersvendsen and Ditzen 2021; Ditzen 2018, 2021). But that is not the only issue because there are limited resources available to bring these methodological techniques to applied researchers in an approachable way.

To fill this gap, this article coalesces the literature on large \( N \) and large \( T \) data to guide researchers through analyzing these data. Most importantly, I build on the prior literature and show using Monte Carlo experiments that researchers must correctly model dynamics (including the correct number of lags), slope heterogeneity (whether effects vary across units), and cross-sectional dependence (units are correlated with each other because they share a common factor like technology or are commonly impacted by an event like a pandemic or recession) and that failing to address these issues can lead to biased and inconsistent estimates. The experiments indicate that using robust or panel corrected standard errors with fixed effects estimation do not mitigate these issues, and the performance of these approaches depend on the degree of lag misspecification, the presence of slope heterogeneity and cross-sectional dependence, and the number of time periods in the analysis. In contrast, the common correlated effects estimator produces superior estimates and is robust to slope heterogeneity and cross-sectional dependence, and it performs just as well as the fixed effects estimator without these issues present. When \( T = 30 \), the two-way fixed effects and common correlated effects estimators perform similarly, but the performance of the common correlated effects estimator is superior as \( T \) expands. Based on these findings, I conclude with a decision-making framework and a set of recommendations for researchers to consider when modeling large \( N \), large \( T \) data.

The article is separated into four sections. The first section outlines the modeling issues to consider with large \( N \), large \( T \) data, along with a discussion of the fixed effects and common correlated effects estimators. The second section reports a series of Monte Carlo experiments to compare several versions of the fixed effects estimator to the common correlated effects estimator. The third section provides an example by examining the drivers of fossil fuel consumption in U.S. states from 1960 to 2018. I then conclude with the decision-making framework and a set of recommendations for researchers.

**Modeling Large \( N \), Large \( T \) Panel Data**

In the following section, I first outline fixed effects estimation of a static model, which serves as the workhorse modeling approach in sociology. Following this discussion, I review dynamic models and the implications of dynamic misspecification. I then outline the issues associated with the fixed effects estimator in the cases of slope heterogeneity and strong cross-sectional dependence, followed by introducing the common correlated effects estimator which can address these concerns.

**The Workhorse: Fixed Effects Estimation of a Static Model**

Fixed Effects estimation of a static model (Equation 1) is the most commonly used approach to analyzing large \( N \), large \( T \) data in sociology:

\[
y_{it} = \alpha_i + \beta x_{it} + u_i + e_{it},
\]

where \( y_{it} \) is the dependent variable, \( x_{it} \) is a time-varying independent variable, \( \alpha_i \) are the unit-specific fixed effects, \( u_i \) are the time-specific fixed effects, and \( e_{it} \) is random variation for each unit at each point in time. This model is referred to as static because it models the contemporaneous effect of the independent variable on the dependent variable and assumes that the effect does not persist over time (De Boef and Keele 2008).

Fixed effects estimation can occur in two ways (Allison 2009). The first way is through the mean deviation approach, which works by constructing case-specific mean deviations for the dependent and independent variables and subtracting the individual-specific mean from each observed value:

\[
y_{it}^* = y_{it} - \bar{y}_i,
\]

\[
x_{it}^* = x_{it} - \bar{x}_i,
\]

and regressing \( y_{it}^* \) on \( x_{it}^* \) (Allison 2009).

The second way fixed effects can be estimated is by including a set of dummy variables for each case, which is also known as the least square dummy variable (LSDV) estimator (Allison 2009). This method allows a fixed effect coefficient to be estimated for each case (the mean-deviation approach does not allow for this), but it does constrain one of the units to zero, which drops it from the analysis. The LSDV estimator and the mean deviation approach produce equivalent estimates.

Fixed effects estimation of a static model has several limitations that can make it perform poorly. First, researchers tend to give little consideration to whether the static model is appropriate—often using it by default. Thus, many models are potentially misspecified due to incorrectly modeling the lag structure (i.e., not including the appropriate number of lags of the dependent and independent variables; De Boef and Keele 2008; Plümper and Troeger 2019). Second, even if the dynamics are properly specified, fixed effects estimation of a dynamic model in the presence of slope heterogeneity produces inconsistent estimates because the slope heterogeneity gets captured in the error term (Pesaran and Smith 2022; Vaisey and Miles 2017).
Dynamic Models and Dynamic Misspecification

Although the static model is commonly used, it is often statistically and theoretically inappropriate (Thombs, Huang, and Fitzgerald 2022). This model assumes that any potential autocorrelation is handled as a nuisance (e.g., measurement error)—rather than as part of the data generating process (i.e., that the estimated model is incorrect). When autocorrelation (and other issues like heteroskedasticity and cross-sectional dependence) arise, a robust variance estimator like clustered robust standard errors, panel corrected standard errors, and/or a generalized least squares transformation like Prais-Winsten are often used (Beck and Katz 1995). However, if the aforementioned issues are actually part of the data generating process, such corrections do little to mitigate their impact on estimation because the estimated model is not changed. Panel data are very likely to be autoregressive given that we expect social processes to be a function of the past. It would be unusual for any social theory to argue that history does not matter, which is what is assumed in a static model.

To see why applying a standard error correction to a static model likely fails to appropriately model the true data generating process, let us first estimate the autocorrelation in the residuals of a static model:

\[ y_{it} = \beta_0 x_{it} + e_{it}, \]

\[ e_{it} = \rho e_{it-1} + \xi_{it}, \]

then applying the Cochrane-Orcutt transformation to observations for \( T > 1 \) with \( \rho \):

\[ y_{it} - \hat{\rho} y_{i,t-1} = \beta (x_{it} - \hat{\rho} x_{it}) + v_{it}, \]

Then, to preserve the first time period, the following transformation is done to \( T = 1 \):

\[ \sqrt{1 - \rho^2} y_{1i} = \beta(\sqrt{1 - \rho^2} x_{1i}) + \sqrt{1 - \rho^2} v_{1i}. \]

The Prais-Winsten transformation is commonly used in tandem with panel corrected standard errors because they do not correct for autocorrelation (e.g., Jorgenson and Clark 2012; Kelly 2020; Kollmeyer and Peters 2019).

The underlying assumption of clustered robust standard errors (or panel corrected standard errors) and the Prais-Winsten transformation is that the estimated model is specified correctly and any heteroskedasticity or correlation in the errors is nuisances to correct for. However, as Plümper and Troeger (2019:19) lucidly argue, “This error structure exists not because nature invented a complex error process that ought to be controlled away, but because of a dynamic misspecification in the underlying data-generating process.” In other words, rather than treating autocorrelation (or other correlations in the errors) as a nuisance to get rid of, the presence of these issues is likely telling us something about our model. Thus, in the case of autocorrelation, researchers can explicitly model it by using a dynamic model. Dynamic models model the memory of a time series, where effects are distributed over time. In contrast to static models, dynamic models estimate both short-run (immediate) effects and long-run (cumulative) effects.

The defining feature of a dynamic model is that it includes the lag of the dependent variable. The most basic dynamic model is the lagged dependent variable (LDV) model, which is also referred to as the partial adjustment model or the Koyck model (De Boef and Keele 2008):

\[ y_{it} = \alpha_i + \phi_1 y_{i,t-1} + \beta_1 x_{it} + e_{it}. \]

In addition to the LDV, lags of the independent variables can be included, which is referred to as an autoregressive distributed lag (ARDL) model (see Table 1 for a summary of the

\[ (\hat{\Sigma}_n)_{OLS} = \hat{\sigma}^2 (XX)^{-1}. \]

OLS assumes that the variance (\( \hat{\sigma}^2 \)) is constant across units and uncorrelated (no autocorrelation, homoskedastic, and cross-sectionally independent). However, in the panel data context, it is likely these assumptions are violated because observations tend to be correlated over time and across units. To correct for autocorrelation and heteroskedasticity, researchers often turn to robust standard errors.

The clustered robust variance-covariance matrix is as follows:

\[ (\hat{\Sigma}_{Clustered}) = (XX)^{-1} \sum_{i=1}^{N} X_i^\prime \hat{u}_i \hat{u}_i^\prime X_i (XX)^{-1}. \]

where \( n \) represents a specific cluster, \( u_i \) is a vector of residuals in each cluster, and \( X_i \) is a matrix of the independent variables for the observations in each cluster. The idea here is that the OLS estimates of the coefficients are unbiased and consistent but are inefficient, so a robust variance estimator is used instead of the OLS variance estimator (Cameron and Trivedi 2010). This results in the standard errors being corrected, but the coefficients remain unchanged.

Using a Prais-Winsten transformation is another common way to account for autocorrelation. This approach works by first estimating the autocorrelation in the residuals of a static model:

\[ y_{it} = \beta_0 x_{it} + e_{it}, \]

\[ e_{it} = \rho e_{it-1} + \xi_{it}, \]

Then, to preserve the first time period, the following transformation is done to \( T = 1 \):

\[ \sqrt{1 - \rho^2} y_{1i} = \beta(\sqrt{1 - \rho^2} x_{1i}) + \sqrt{1 - \rho^2} v_{1i}. \]

The Prais-Winsten transformation is commonly used in tandem with panel corrected standard errors because they do not correct for autocorrelation (e.g., Jorgenson and Clark 2012; Kelly 2020; Kollmeyer and Peters 2019).

The underlying assumption of clustered robust standard errors (or panel corrected standard errors) and the Prais-Winsten transformation is that the estimated model is specified correctly and any heteroskedasticity or correlation in the errors is nuisances to correct for. However, as Plümper and Troeger (2019:19) lucidly argue, “This error structure exists not because nature invented a complex error process that ought to be controlled away, but because of a dynamic misspecification in the underlying data-generating process.” In other words, rather than treating autocorrelation (or other correlations in the errors) as a nuisance to get rid of, the presence of these issues is likely telling us something about our model. Thus, in the case of autocorrelation, researchers can explicitly model it by using a dynamic model. Dynamic models model the memory of a time series, where effects are distributed over time. In contrast to static models, dynamic models estimate both short-run (immediate) effects and long-run (cumulative) effects.

The defining feature of a dynamic model is that it includes the lag of the dependent variable. The most basic dynamic model is the lagged dependent variable (LDV) model, which is also referred to as the partial adjustment model or the Koyck model (De Boef and Keele 2008):

\[ y_{it} = \alpha_i + \phi_1 y_{i,t-1} + \beta_1 x_{it} + e_{it}. \]

In addition to the LDV, lags of the independent variables can be included, which is referred to as an autoregressive distributed lag (ARDL) model (see Table 1 for a summary of the
different dynamic models). The number of lagged dependent variables in the model is denoted as $p$ and the number of lags of the independent variable as $q$: ARDL($p,q$). A common ARDL model is the ARDL(1,1):

$$y_{it} = \alpha + \phi y_{it-1} + \beta_1 x_{it} + \beta_2 x_{it-1} + e_{it}. \quad (9)$$

The ARDL model can also be written as the algebraically equivalent error-correction model (ECM), where the contemporaneous values are modeled in first differences and the lags are in levels:

$$\Delta y_{it} = \alpha - \hat{\phi}_1 \Delta y_{it-1} + \beta_1 \Delta x_{it} + \beta_2 \Delta x_{it-1} + e_{it}. \quad (10)$$

$\Delta y_{it-1} - \lambda \Delta x_{it-1}$ is the cointegrating relationship between $x$ and $y$. Cointegration is when a linear combination of two nonstationary series make a stationary process (see Engle and Granger 1987). $\hat{\phi}_1$ is the error-correction rate, indicating how quickly the model returns to equilibrium following a shock to an independent variable, and $\lambda$ is the long-run effect of $x_{it}$ (Webb, Linn, and Lebo 2020). The ECM can be rewritten as a generalized ECM (GECM) by multiplying through $-\hat{\phi}_1$ in Equation 10 to get Equation 11:

$$\Delta y_{it} = \alpha + \hat{\theta}_1 \Delta y_{it-1} + \beta_1 \Delta x_{it} + \beta_2 \Delta x_{it-1} + e_{it}. \quad (11)$$

Equations 9 through 11 are algebraically equivalent to one another. $\phi_1 = 1 - \hat{\phi}_1$, and $\beta_1 + \beta_2$ in Equation 9 is equivalent to $\hat{\theta}_1 \lambda$ in Equation 10 and $\beta_1 + \beta_2$ in Equation 11. The short-run effect of $x_{it}$ is $\beta_1 = \hat{\beta}_1 = \beta_1$ in Equations 9 through 11, respectively. The long-run effect of $x_{it}$ is $\frac{\beta_1 + \beta_2}{1 - \hat{\phi}_1}$ in Equation 11.

Equation 9, which is equivalent to $\lambda$ in Equation 10 and $\frac{\beta_1}{\hat{\theta}_1}$ in Equation 11.

**Estimating the Distribution of the Long-Run Effect**

Researchers can calculate the distribution of the long-run effect over time by using the model estimates. In the calculation of the long-run effects, the denominator tells us how much the outcome variable changes over time given a shock to an independent variable (known as the speed of adjustment). In the LDV and ARDL models, an autoregressive coefficient near unity indicates that the dependent variable is highly persistent (and near zero for the ECM), and therefore, the long-run effect is distributed over a longer period of time. In contrast, an autoregressive coefficient near zero indicates that most of the effect occurs immediately.

We can manipulate the autoregressive coefficient in Equation 9 to illustrate this point. In a case where the dependent variable is highly persistent, $\phi_1 = 0.9$ and $\beta_1 = 0.4$ and $\beta_2 = 0.2$, then by construction, the long-run effect is $0.4 + 0.2 - 0.9 = 6.3$. We can then calculate the proportion of the long-run effect occurring in each time period. Assuming a one-unit shock to $x_{it}$, the proportion of the effect in the contemporaneous period is calculated by dividing the short-run effect by the long-run effect ($0.4 / 0.6 = 0.67$). In other words, 6.7 percent of the long-run effect occurs in the first period. Then in the second time period, we calculate the proportion of the long-run as $y_{it+1} = 0.9(0.4) + 0.2 = 0.56$. This suggests that $0.56 / 0.6 = 0.593$, or 9.3 percent of the long-run effect occurs in the second period. Thus, after two periods, 6.7 percent + 9.3 percent = 16 percent of the long-run effect occurred. The proportion of the effect in the third period (and for the remaining time periods) is given by $y_{it+2} = 0.9(0.56) = 0.50$, indicating that $0.50 / 0.6 = 0.83$, or 8.3 percent occurs in the third period. Thus, the long-run effect transpires relatively slowly over time.

If $\phi_1 = 0.1$, we can see that the long-run effect is now $0.4 + 0.2 - 0.9 = 0.67$, and 59.7 percent $(0.4 / 0.67 = 0.597)$ of the long-run effect occurs in the first period. For the second period, the effect is $y_{it+1} = 0.1(0.4) + 0.2 = 0.24$. This suggests that

---

*Note: LDV = lagged dependent variable; ARDL = autoregressive distributed lag; ECM = error-correction model; GECM = generalized error-correction model.*

---

**Table 1. Dynamic Models.**

| Model Name | Model | Long-Run Effect |
|------------|-------|-----------------|
| LDV        | $y_{it} = \alpha_i + \phi y_{it-1} + \beta_1 x_{it} + \beta_2 x_{it-1} + e_{it}$ | \( \frac{\beta_1}{1-\phi_1} \) |
| ARDL       | $y_{it} = \alpha_i + \phi y_{it-1} + \beta_1 x_{it} + \beta_2 x_{it-1} + e_{it}$ | \( \frac{\beta_1 + \beta_2}{1-\hat{\phi}_1} \) |
| ECM        | $\Delta y_{it} = \alpha_i - \hat{\phi}_1 (y_{it-1} - \lambda x_{it-1}) + \kappa_i \Delta x_{it} + e_{it}$ | $\hat{\lambda}$ |
| GECM       | $\Delta y_{it} = \alpha_i + \hat{\theta}_1 \Delta y_{it-1} + \beta_1 \Delta x_{it} + \beta_2 \Delta x_{it-1} + e_{it}$ | \( \frac{-\beta_1}{\hat{\theta}_1} \) |

---

*Note: LDV = lagged dependent variable; ARDL = autoregressive distributed lag; ECM = error-correction model; GECM = generalized error-correction model.*

---

*The error correction rate should fall between $-1$ and 0. It is equivalent to the autoregressive coefficient in the ARDL model. Thus, a more persistent dependent variable will have an error correction rate near zero, and a less persistent dependent variable will have an error correction rate near $-1$. This discussion draws from De Boef and Keele (2008).*
0.24 / 0.67 = 0.358, or 35.8 percent of the long-run effect occurs in the second period, and $y_{i,t+2} = 0.1(0.24) = 0.024$ or $0.036 / 0.67 = 0.036$, or 3.6 percent in the third period. In other words, the long-run effect transpires quickly relative to when $\phi = 0.9$. Figure 1 illustrates this by plotting the distribution of the long-run effect over 10 periods of time for each of these examples. As the figure illustrates, nearly all of the long-run effect occurs within the first three time periods when $\phi = 0.1$ but only 63.8 percent of the effect by $T = 10$ when $\phi = 0.9$.

**Slope Heterogeneity**

Slope heterogeneity, which is where the effect of a variable differs across units (e.g., the effect of a variable in a cross-national study is different in the United States than in Nigeria), is theoretically and statistically relevant for sociologists. In terms of theory testing, many macro-level sociological theories make arguments regarding heterogeneity—making it important to model to appropriately test different hypotheses. Statistically, slope heterogeneity makes the fixed effects estimator (and other pooled estimators) problematic to use.

We can see this by considering the following dynamic model with heterogeneous slopes that is estimated by pooled, fixed effects regression:

$$y_{i,t} = \lambda_{i,t} y_{i,t-1} + \beta_{i,t} x_{i,t} + \alpha_t + \epsilon_{i,t}$$

$$\lambda_{i,t} = \lambda + \eta_{i,t}$$

$$\beta_{i,t} = \beta + \eta_{i,t}$$

$$\epsilon_{i,t} = \eta_{i,t} y_{i,t-1} + \eta_{i,t-1} x_{i,t} + \xi_{i,t}.$$ (12)

Regardless of whether the lags are properly specified, fixed effects estimation of a dynamic model in the presence of slope heterogeneity produces inconsistent estimates because the heterogeneity gets captured in the error term (Pesaran and Smith 1995). The fixed effects estimator will produce unbiased and consistent (but inefficient) estimates in the presence of slope heterogeneity (Pesaran and Smith 1995). Even if $\lambda_{i,t}$ is truly homogeneous, slope heterogeneity in $x_{i,t}$ will lead $\lambda_{i,t}$ to tend toward unity and $\beta_{i,t}$ to 0 (Pesaran and Smith 1995). When lags of the regressors are included in the model, $\beta_{i,t}$ and the coefficient of $x_{i,t-1}$ tend to take on the equal and opposite sign, and $\lambda_{i,t}$ will still tend toward 1 (Pesaran and Smith 1995). Thus, the long-run effect tends toward 0 because, based on the equation $(\hat{\beta}_{i,t} + \hat{\lambda}_{i,t})$, the numerator and the denominator tend toward 0 (Pesaran and Smith 1995:100).

To model slope heterogeneity, Pesaran and Smith (1995) propose the mean group (MG) estimator. The MG estimator runs a time-series regression on each unit and then takes the mean of all the estimates. MG estimation of the common ARDL(1,1) model is as follows:

6. This is not the case with a static model when the independent variables are strictly exogenous and the coefficients are uncorrelated with the observations.

7. Pesaran, Shin, and Smith (1999) note that the pooled mean group estimator can also be used, which restricts the long-run coefficients to equality but allows for short-run slope heterogeneity.
\[ y_{it} = \phi_{1it} + \beta_{1i} x_{it} + \beta_{2i} x_{it} + \alpha_{it} + \epsilon_{it}, \quad (13) \]

where the mean group coefficients are the average of the unit-specific coefficients (Ditzen 2018; Thombs et al. 2022):

\[ \delta_{MG} = \left( \phi_{1MG}, \beta_{1MG}, \beta_{2MG} \right) = \frac{1}{N} \sum_{i=1}^{N} \delta_{i}, \quad (14) \]

**Cross-Sectional Dependence**

Addressing cross-sectional dependence (cross-sectional units are correlated with one another) with large \( N \), large \( T \) data is of growing interest in panel data econometrics (Bailey, Kapetanios, and Pesaran 2016; Chudik and Pesaran 2015a; Everaert and De Groote 2016; Pesaran 2015; Thombs et al. 2022). This expanding body of literature shows that failing to correct for it leads to inconsistent and biased estimates (Ditzen 2018; Pesaran 2006). Cross-sectional dependence can generally be characterized as weak or strong (see Chudik and Pesaran 2015b). Weak cross-sectional dependence is when the correlation between cross-sectional units converges to zero as \( N \) increases to infinity, whereas the correlation converges to a nonzero constant as \( N \) increases to infinity with strong cross-sectional dependence (Chudik, Pesaran, and Tosetti 2011; Ditzen 2018, 2021). Examples of weak cross-sectional dependence are when the units are correlated due to spatial distance, and spatial regression can be used to model it. On the other hand, strong cross-sectional dependence stems from unobserved common factors that tie units together. Common examples are pandemics, recessions, or technological change. In the large \( N \), large \( T \) context, only weak cross-sectional dependence is needed to obtain consistent estimates, rather than the more restrictive assumption of cross-sectional independence, because correlation between a small number of units in the analysis does not impact estimates (Pesaran 2015).

Traditionally, cross-sectional dependence is dealt with using time fixed effects. Time fixed effects model the cross-sectional dependence by including \( T - 1 \) dummy variables of the time dimension. However, this approach makes strong assumptions that are often unrealistic, such as assuming common factors affect each unit in the analysis homogeneously. Such an assumption contrasts with an array of macro-comparative work that shows that structural factors have heterogeneous effects across nations and regions (Beckfield 2003; Chase-Dunn 1998; Givens, Huang, and Jorgenson 2019; Jorgenson and Clark 2012; VanHeuvelen and Summers 2019). Thus, time fixed effects may not adequately model the cross-sectional dependence in many cases.

Another approach is to use panel corrected standard errors, which adjust the standard errors but not the estimated coefficients (like robust standard errors). A fundamental difference between the two is that clustered robust standard errors assume that errors are not correlated across units. Beck and Katz (1995) first developed panel corrected standard errors and showed that they performed favorably to feasible generalized least squares, which were commonly used at the time. Assuming there are balanced data, panel corrected standard errors are estimated as follows:

\[ (\hat{\Omega})_{PCSE} = (X'X)^{-1} (X'\Omega X) (X'X)^{-1}, \quad (15) \]

where \( \Omega \) is a \( NT \times NT \) block matrix with the contemporaneous covariances on the diagonal denoted by \( \Sigma \), which is defined as:

\[ \hat{\Sigma}_{ij} = \frac{\hat{\epsilon_i} \hat{\epsilon_j}}{T}, \quad (16) \]

where \( \epsilon_i \) and \( \epsilon_j \) are the residuals for unit \( i \) and unit \( j \) and \( T \) is the number of time periods.

However, treating the cross-sectional dependence as a nuisance in the error term suffers from the same problems discussed in the previous section regarding autocorrelation. These approaches are sufficient if the cross-sectional dependence is truly part of the error term, but they will not adequately model the dependence if it is truly part of the data generating process. Given these considerations, we can turn to the common correlated effects estimator that offers a way to directly model the cross-sectional dependence.

**Common Correlated Effects**

Pesaran (2006) advanced the common correlated effects estimator to eliminate the effects of the unobserved common factors by using cross-sectional averages with OLS. This method contrasts with other common approaches that rely on correcting standard errors rather than explicitly modeling the cross-sectional dependence. The common correlated effects estimator allows these factors to affect each unit differently (Pesaran 2006). Assuming a model with heterogeneous slopes with one unobserved common factor, the estimator is based on Equation 17:

\[ y_{it} = \alpha_i + \beta_{1i} x_{it} + u_{it}, \quad (17) \]

where \( y_{it} \) is the function of time-varying variable \( x_{it} \), \( \alpha_i \) is the unit-specific fixed effect, and \( u_{it} \) is the error term that consists of the unobserved common factor \( f_t \), with heterogeneous factor loadings \( \gamma_i \), and an independent and identically distributed error term, \( \epsilon_{it} \), that is unit-specific (Ditzen 2018; Pesaran 2006). The actual equation the common correlated effects estimator estimates is Equation 18:

\[ y_{it} = \alpha_i + \beta_{1i} x_{it} + \gamma_{it} \tilde{v}_t + \epsilon_{it}, \quad (18) \]

where \( \tilde{v}_t = \left( \tilde{x}_t, \tilde{y}_t \right) \) and \( \gamma_{it} \) are the unit-specific factor loadings. Pesaran (2006) demonstrates that the cross-sectional

---

8Driscoll and Kraay (1998) standard errors can also be used to correct for cross-sectional dependence in the error term.
averages will consistently estimate the common factors as long as \( x_{it} \) is strictly exogenous. Chudik and Pesaran (2015a) show that the common correlated effects estimator is consistent with a dynamic model by adding a minimum of \( \sqrt{T} \) lags of the cross-sectional averages. Thus, the equation for an ARDL(1,1) model estimated by the common correlated effects estimator with \( p \) lags of the cross-sectional averages is as follows:

\[
y_{it} = \lambda_{1i} y_{i,t-1} + \beta_{1i} x_{it} + \beta_{2i} x_{it-1} + \sum_{t=0}^{p} \gamma_{i} \overline{y}_{i,t-1} + \alpha_{i} + \epsilon_{it},
\]

(19)

Monte Carlo Experiments and Evidence

Monte Carlo Experiments

Five Monte Carlo experiments were conducted to compare the robustness of the fixed effects estimator and the common correlated effects estimator to autoregression (or persistence), slope heterogeneity, and cross-sectional dependence.9

The data generating process of the experiments are based on the following ARDL \((p,q)\) model:

\[
y_{it} = \lambda_{1i} y_{i,t-p} + \beta_{1i} x_{it} + \beta_{2i} x_{it-q} + \gamma_{i} f_{i} + c_{it} + u_{it},
\]

where \( y_{it} \) is the dependent variable, \( x_{it} \) is the independent variable, \( c_{it} \) is the unit-specific fixed effect, and \( \gamma_{i} \) are the heterogeneous factor loadings for the common factor \( f_{i} \). \( \lambda_{1i}, \beta_{1i}, \) and \( \beta_{2i} \) are the coefficients for the lagged dependent variable, the contemporaneous value of \( x_{it} \), and the lagged value of \( x_{it} \), respectively. \( x_{it} \) is autoregressive \((\alpha_{it} = 0.5)\) and a function of a unit-specific fixed effect \((c_{it})\) and a common factor \((f_{i})\):

\[
x_{it} = \alpha_{it} x_{i,t-1} + c_{it} + \gamma_{i} f_{i} + \epsilon_{it}.
\]

The unit-specific fixed effects are generated as:

\[
c_{it} \sim IIDN(1,1)
\]

\[
c_{it} = \sigma_{i} + \eta_{it} \sim IIDN(0,1).
\]

The common factor is autoregressive:

\[
f_{i} = \rho_{f} f_{i-1} + \xi_{i} \sim IIDN(0,1-\rho^{2})
\]

\[
\rho_{f} = 0.7,
\]

and the error terms are independent and identically distributed:

\[
\begin{align*}
    \epsilon_{it} & \sim IIDN(0,1) \\
    u_{i,t} & \sim IIDN(0,1).
\end{align*}
\]

I performed five experiments based on Equation 20, which I summarize in Table 2, Table 3, and Table 4. Experiments 1 through 3 follow an ARDL \((1,1)\) and include a common factor and differ according to the degree of slope heterogeneity. Experiment 1 allows for the most variation in the slope coefficients, whereas Experiment 3 allows for the least, and Experiment 2 falls in the middle. Experiment 4 also follows an ARDL \((1,1)\), but the slope coefficients are homogeneous. Experiment 5 restricts Equation 20 to an ARDL \((0,0)\), that is, a static model, where the slope coefficients are homogeneous and there is no cross-sectional dependence. Experiment 5 is a “best case” scenario for the one-way fixed effects estimator. One thousand simulations are performed for each combination of \( N \) and \( T \) of 30, 60, and 90. Robust standard errors and panel corrected standard errors (potentially combined with a Prais-Winsten transformation) are used with the fixed effects estimator to test the robustness of these approaches (see Table 4). One hundred initial observations are generated and then dropped for each panel to prevent issues with initial conditions.

Monte Carlo Results

The results of the Monte Carlo experiments are summarized for each experiment in the following. The results are reported in tabular format in the supplemental material. The bias, root mean square error (RMSE), and coverage are reported for each estimator. The bias reports the average difference between the estimator and the true parameter in percentage terms. The RMSE provides information on how much each estimator deviates from the true parameter, which is a function of how biased and inefficient an estimator is, and the coverage is the proportion of the confidence intervals that include the true parameter.

| Table 2. Summary of Monte Carlo Experiments. |
|-----------------|-----------------|-----------------|
| Experiment | Model | Slope Heterogeneity | Common Factors |
| 1 | Dynamic | Yes | 1 |
| 2 | Dynamic | Yes | 1 |
| 3 | Dynamic | Yes | 1 |
| 4 | Dynamic | No | 0 |
| 5 | Static | No | 0 |

As expected, the DCCE10 performs the best of the estimators in Experiment 1. The dynamic fixed effects estimators and the DCCE perform similarly when \( T = 30 \), but the DCCE is far superior as \( T \) increases. For the

9The analyses in this study are performed with Stata 17. The fixed effects estimator with robust standard errors is implemented with the xtabreg, fe command, the fixed effects estimator with panel corrected standard errors with the xtpcse command, and the common correlated effects estimator with the xtdcce2 command (Ditzen 2018, 2021).

10The abbreviations used in this section correspond to Table 4.
Table 3. Coefficients in Monte Carlo Experiments.

| Coefficients                  | 1    | 2    | 3    | 4    | 5    |
|------------------------------|------|------|------|------|------|
| $\lambda_{i,1}$ (autoregressive coefficient) | IIDU (0, 0.9) | IIDU (0.2, 0.7) | IIDU (0.4, 0.6) | 0.5   | —    |
| $\beta_{i,1}$ (short-run effect) | IIDU (0, 1) | IIDU (0.25, 0.75) | IIDU (0.4, 0.6) | 0.5   | 0.5  |
| $\beta_{2,i}$ (coefficient of $x_{i,t-1}$) | IIDU (0.2, 0.9) | IIDU (0.3, 0.7) | IIDU (0.4, 0.6) | 0.5   | —    |
| $\theta$ (long-run effect) | $\hat{\beta}_{i,1} + \hat{\beta}_{2,i}$ | $\hat{\beta}_{i,1} + \hat{\beta}_{2,i}$ | $\hat{\beta}_{i,1} + \hat{\beta}_{2,i}$ | $\hat{\beta}_{i,1} + \hat{\beta}_{2,i}$ | —    |
| $\gamma_1, \gamma_2$ (factor loadings) | IIDU (0.5, 1.5) | IIDU (0.5, 1.5) | IIDU (0.5, 1.5) | —    | —    |

Table 4. Estimators and Models used in the Monte Carlo Experiments.

| Estimator          | Model                                      |
|--------------------|--------------------------------------------|
| 1FE $^R$           | $y_{i,t} = \beta_1 x_{i,t} + \alpha_i + \epsilon_{i,t}$ |
| 2DFE $^R$          | $y_{i,t} = \lambda_i y_{i,t-1} + \beta_1 x_{i,t} + \beta_2 x_{i,t-1} + \alpha_i + \epsilon_{i,t}$ |
| 2FER $^R$          | $y_{i,t} = \beta_1 x_{i,t} + \alpha_i + \epsilon_{i,t}$ |
| 2DFE $^R$/PCSE     | $y_{i,t} = \lambda_i y_{i,t-1} + \beta_1 x_{i,t} + \beta_2 x_{i,t-1} + \alpha_i + \epsilon_{i,t}$ |
| 2DFE $^R$/PSAR     | $y_{i,t} = \lambda_i y_{i,t-1} + \beta_1 x_{i,t} + \beta_2 x_{i,t-1} + \alpha_i + \epsilon_{i,t}$ |
| DFE $^R$           | $y_{i,t} = \lambda_i y_{i,t-1} + \beta_1 x_{i,t} + \beta_2 x_{i,t-1} + \alpha_i + \epsilon_{i,t}$ |
| CCE $^R$           | $y_{i,t} = \beta_1 x_{i,t} + \gamma_i \tilde{V}_{i,t} + \alpha_i + \epsilon_{i,t}$ |
| DCCE $^R$          | $y_{i,t} = \lambda_i y_{i,t-1} + \beta_1 x_{i,t} + \beta_2 x_{i,t-1} + \sum_{i=0}^{p} \gamma_i \tilde{V}_{i,t-i} + \alpha_i + \epsilon_{i,t}$ |

Note: 1FE = one-way fixed effects; 2FE = two-way fixed effects; R = clustered robust standard errors; PCSE = panel corrected standard errors; PSAR = panel-specific autoregressive(1) correction; CCE = common correlated effects; D = dynamic.

The results for $\beta_{i,1}$, $\lambda_{i,1}$, and $\beta_{2,i}$ in Experiment 2 are similar to those of Experiment 1, but the biases of the dynamic fixed effects estimators are not as great, which is expected given that there is less slope heterogeneity. The 2DFE produces slightly less bias (in absolute terms) than the DCCE when $T = 50$ (6.5 percent compared to –9.8 percent), but the bias and RMSE of the 2DFE increase as $T$ increases. When $T = 90$, the bias is between 10.6 percent and 11.1 percent, and the RMSE is between 20.9 and 21.9. In contrast, the bias and RMSE for the DCCE tend toward zero as $T$ increases. When $T = 90$, the bias is between –2.7 percent and –2.4 percent, and the RMSE is between 5 and 6.5. Like in Experiment 1, the coverage is greater for the 2DFE than for the 2DFE $^R$/PCSE, with the coverage for the 2DFE $^R$ being between 45 and 61 percentage points higher.

The results for $\beta_{i,1}$, $\lambda_{i,1}$, and $\beta_{2,i}$ are also similar to Experiment 1. For $\beta_{i,1}$, the 2DFE and the DCCE have minimal bias (near zero), low RMSEs, and high coverage. For $\lambda_{i,1}$, the bias of the dynamic fixed effects estimator is positive, which increases as $T$ increases, but is smaller in magnitude than in Experiment 1. When $T = 90$, the bias of $\lambda_{i,1}$ is roughly 32 percent. The bias for $\beta_{2,i}$ in the 2DFE models is between –25.7 percent and –26.2 percent. In comparison, the bias for $\lambda_{i,1}$ is approximately –7 percent for the DCCE when $T = 90$, and the bias for $\beta_{2,i}$ is approximately 5 percent.
Like Experiment 1, the dynamic models are superior to the static models with regards to estimating $\hat{\beta}_{1,i}$. The 1FE$^R$ performs by far the worst of any of the estimators (the bias is greater than 196 percent for all combinations of $N$ and $T$), while the 2FE$^\text{PCSE/PSAR}$ performs the best of the static models. The bias for the 2FE$^\text{PCSE/PSAR}$ does not exceed $-7.67$ percent.

**Experiment 3: dynamic, low slope heterogeneity, and cross-sectional dependence.** The results for the long-run effect ($\theta$) in Experiment 3 follow a similar trend as Experiments 1 and 2, but the 2DFE exhibits minimal bias. The 2DFE performs well in estimating the long-run effect, with a bias less than 2 percent for all combinations of $N$ and $T$, and the coverage is greater than 91 percent for all combinations of $N$ and $T$. In comparison, the DCCE also performs well, and its bias and RMSE decrease as $T$ increases, while its coverage increases as $T$ increases. The 1DFE$^R$ performs poorly compared to the other dynamic estimators because it exhibits bias (>$23$ percent for all combinations of $N$ and $T$) and has very poor coverage (0 percent for all combinations of $N$ and $T$ >60). Overall, these results indicate that the 2DFE performs well with low levels of slope heterogeneity but is inconsistent as $T$ increases.

The results for $\hat{\beta}_{1,i}$, $\lambda_{1,i}$, and $\beta_{2,i}$ are also similar to Experiments 1 and 2, but the impact of slope heterogeneity is relatively small. For $\hat{\beta}_{1,i}$, the dynamic two-way fixed effects estimator and the DCCE have minimal bias (near zero), low RMSEs, and high coverage. For $\lambda_{1,i}$, the bias of the dynamic fixed effects estimator remains positive (tends toward 1) but is significantly smaller in magnitude compared to the high slope heterogeneity case. For example, when $N$ and $T = 90$, the bias of the 1DFE is 9.34 percent, and the bias for the 2DFE models is 5.35 percent. In comparison, the bias for $\lambda_{1,i}$ in the DCCE model is negative and tends toward 0 as $T$ increases.

For $\hat{\beta}_{2,i}$, the bias of the dynamic fixed effects estimator increases as $T$ increases but less so than in the first two experiments. The bias remains negative, indicating the coefficient tends toward the equal and opposite sign of $\hat{\beta}_{1,i}$. For example, when $N$ and $T = 90$, the bias for the 1DFE is $-14.69$ percent, and the bias for the 2DFE models is $-4.93$ percent. In contrast, the DCCE produces a positive bias, and it converges to the true parameter as $T$ increases, as evidenced by the bias decreasing from approximately 20 percent when $T = 30$ to approximately 5.9 percent when $T = 90$ for all combinations of $N$.

Like Experiments 1 and 2, the dynamic models outperform the static models in estimating $\hat{\beta}_{1,i}$. The 1FE$^R$ performs the worst of any of the estimators, with a bias greater than 209 percent for all combinations of $N$ and $T$. This suggests that the 1FE$^R$ tends to perform poorly regardless of the level of slope heterogeneity. The 2FE$^\text{PCSE/PSAR}$ performs the best of the static models, with a bias between $-2.16$ percent and $-3.68$ percent for all combinations of $N$ and $T$. Its RMSE and coverage also far outperform the 1FE$^R$ and 2FE$^R$.

**Experiment 4: dynamic, no slope heterogeneity, and cross-sectional dependence.** Experiment 4 differs from the first three experiments because the slope coefficients are homogeneous. The 2DFE performs the best in terms of estimating the long-run effect ($\theta$), and the bias ($<-3.15$ percent for all combinations of $N$ and $T$) and the RMSE tend to 0 as $T$ increases. The DCCE also performs well, with the bias ($<-8.76$ percent for all combinations of $N$ and $T$) and RMSE tending toward 0 as $T$ increases. The 1DFE$^R$ performs poorly, with the bias and RMSE increasing as $T$ increases, and it has very poor coverage. The bias (between 23.24 percent and 28.07 percent) is similar to its bias in the other experiments with slope heterogeneity.

Like Experiment 1, the dynamic models are superior to the static models, with biases near 0 for all combinations of $N$ and $T$, and it has a much lower RMSE and higher coverage compared to the other static models. The 1FE$^R$ performs the worst of any of the estimators, with a bias greater than 209 percent for all combinations of $N$ and $T$.

Moving to the static models, the 2FE$^\text{PCSE/PSAR}$ performs the best of the static models, with biases near 0 for all combinations of $N$ and $T$, and it has a much lower RMSE and higher coverage compared to the other static models. The 1FE$^R$ performs the worst of any of the estimators, with a bias greater than 209 percent for all combinations of $N$ and $T$.

**Experiment 5: static, no slope heterogeneity, and no cross-sectional dependence.** Experiment 5 is a “best case” scenario for the one-way fixed effects estimator because there is no slope heterogeneity or cross-sectional dependence. The estimator does perform well, with biases near 0, a low RMSE, and high coverage for all combinations of $N$ and $T$ for the static and dynamic models. However, two-way fixed effects and CCE estimation of dynamic and static models perform just as well. This indicates that the one-way fixed effects estimator does not offer superior performance compared to the other estimators even in a best case scenario. A summary table of the estimators/models across the experiments is found in Table 5.

### An Example: Drivers of U.S. State-Level Fossil Fuel Consumption

Large $N$, large $T$ data are commonly used in environmental sociology to study the drivers of environmental change (Dietz 2017; Jorgenson and Clark 2012; Jorgenson et al. 2019). I extend this literature by examining the drivers of fossil fuel consumption in the United States from 1960 to 2018. Fossil fuel consumption data are obtained from the
Socius: Sociological Research for a Dynamic World

U.S. Energy Information Agency (2021). Drawing from the Stochastic Impacts by Regression on Population, Affluence and Technology (STIRPAT) framework (York et al. 2003), I include real disposable personal income deflated by the CPI-U-US (Bureau of Economic Analysis 2021), total population (U.S. Energy Information Administration 2021), the industrial sector’s share of energy consumption (U.S. Energy Information Administration 2021), the share of income going to the top 10 percent of earners (Frank 2021), and the percentage energy consumption from renewables (U.S. Energy Information Administration 2021) as independent variables.11 In the following, I report the dynamic and static models for the one-way fixed effects estimator with robust standard errors (1FER), the two-way fixed effects estimator with robust standard errors (2FER), the two-way fixed effects estimator with panel-corrected standard errors (2FE PCSE),12 and the common correlated effects (CCE) estimator. Models that are dynamic start with the letter “D.” The descriptive statistics are reported in Table 6, and the slope heterogeneity tests are reported in Table 7. The variables exhibit strong

Table 5. Summary of the Monte Carlo Experiments by Estimator and Model.

| Estimator/Model | Summary |
|-----------------|---------|
| 1DFER<sup>11</sup> | Performs poorly in the presence of slope heterogeneity and cross-sectional dependence. It produces highly biased estimates, and the bias increases as \( T \) increases. |
| 2DFER<sup>11</sup> | Performs poorly with a high degree of slope heterogeneity but does pretty well with medium to low levels of heterogeneity. Tends to outperform the DCCE when \( T = 30 \), but its performance worsens as \( T \) increases. |
| 2DFE<sup>PCSE</sup> | Performs the same as the 2DFER<sup>11</sup> in terms of bias and RMSE because it produces the same estimated coefficient. Its coverage is worse than the 2DFER<sup>11</sup> because it tends to underestimate the standard errors. |
| DCCE | Performs the best in terms of bias, RMSE, and coverage when there is slope heterogeneity and strong cross-sectional dependence, particularly as \( T \) increases. |
| 1FE<sup>11</sup> | Strong assumptions are needed for it to perform well (no autoregression, no slope heterogeneity, and no strong cross-sectional dependence). Should be used with extreme caution. |
| 2FE<sup>11</sup> | Performs poorly (but better than the 1FE<sup>11</sup>) because it assumes no autoregression, no slope heterogeneity, and homogeneous factor loadings. |
| 2FE<sup>PCSE/PSAR</sup> | Performs the best of the static models and produces negatively biased coefficients, whereas the other static models produce positive bias. However, like the other static models, its performance worsens as \( T \) increases. |
| CCE | Although it models slope heterogeneity and cross-sectional dependence, it produces highly biased estimates (because it assumes no autoregression). Its performance worsens as \( T \) increases. |

Note: 1FE = one-way fixed effects; 2FE = two-way fixed effects; R = clustered robust standard errors; PCSE = panel corrected standard errors; PSAR = panel-specific autoregressive(1) correction; CCE = common correlated effects; D = dynamic; RMSE = root mean square error.

Table 6. Descriptive Statistics.

| Variable | Measurement Unit | Mean (SD) | Between SD | Within SD | CD |
|----------|------------------|-----------|------------|-----------|----|
| Fossil fuel consumption | British thermal units (millions) | 1.39 (1.62) | 1.57 | .43 | 174.17<sup>10</sup> |
| Personal income | Per capita | 30,882.33 (10,283.30) | 4,329.06 | 9,347.03 | 269.84<sup>10</sup> |
| Population | Total | 4,945.89 (5,577.46) | 5,380.77 | 1,647.32 | 231.75<sup>10</sup> |
| Industrial share | % of energy consumption | 35.08 (12.91) | 11.39 | 6.27 | 117.55<sup>10</sup> |
| Top 10% share | % of income | 28.42 (6.39) | 3.04 | 5.64 | 248.58<sup>10</sup> |
| Renewable energy (%) | % of energy consumption | 9.73 (11.36) | 10.89 | 3.59 | 88.03<sup>10</sup> |

Note: Data are perfectly balanced with 3,009 total observations (\( N = 51, T = 59 \)). Generated with the `xtsum` and `xtcd2` (Ditzen 2018) commands in Stata 17. CD = Statistic for Pesaran test for weak cross-sectional dependence (\( H_0 = \text{weak cross-sectional dependence} \)). \(* p < .05\).

Table 7. Slope Heterogeneity Tests.

| Test | Adjusted Delta Statistic |
|------|--------------------------|
| Blomquist and Westerlund HAC | 11.11<sup>10</sup> |
| Pesaran-Yamagata test with CSAs | 9.75<sup>10</sup> |
| Pesaran-Yamagata test AR adjusted | 15.02<sup>10</sup> |

Note: \( H_0 = \text{slope homogeneity} \). The Bartlett-Westerlund HAC test is estimated with bandwidth of 3. The Pesaran-Yamagata test with cross-sectional averages incorporates three lags of each contemporaneous variable. Generated with the `xthst` (Bersvendsen and Ditzen 2021) command in Stata 17. AR = autoregressive; CSA = cross-sectional average; HAC = heteroscedasticity and autocorrelation consistent. \(* p < .05\).

11The variables are in natural logarithms.

12For the static model, a panel-specific AR(1) correction is used with the PCSEs, so it is labeled 2FE<sup>PCSE/PSAR</sup>. 
cross-sectional dependence (see Table 6), and the slope coefficients are heterogeneous according to all three slope heterogeneity tests.

The results of the dynamic models are reported in Table 8, and the static models are reported in Table 9 for reference. For the dynamic models, the coefficients behave as expected. The autoregressive coefficient is near unity in the fixed effects models, while it is .504 for the DCCE model. The short-run coefficients and the coefficients on the lags of the independent variables tend to take on equal and opposite signs in the fixed effects models, which is also expected. Pesaran’s test for weak cross-sectional dependence (reported

---

Table 8. Dynamic Models of Fossil Fuel Consumption, 1960–2018.

|                      | 1DFE<sup>R</sup> | 2DFE<sup>R</sup> | 2DFE<sup>PCSE</sup> | DCCE  |
|----------------------|------------------|------------------|----------------------|-------|
| **Fossil fuel consumption** | .915* (.010) | .925* (.012) | .925* (.009) | .508* (0.23) |
| **Short-run effects** |                  |                  |                      |       |
| Personal income      | .301* (.078)     | .064 (.047)     | .064 (.048)          | -.024 (0.054) |
| Population           | .558* (.119)     | .697* (.139)    | .697* (.140)         | .343# (0.207) |
| Industrial share     | .133* (.051)     | .128* (.046)    | .128* (.016)         | .214* (0.042) |
| Top 10% share        | .032 (.028)      | .061* (.029)    | .061* (.029)         | .005 (0.026)  |
| Renewable energy (%) | -.085* (.010)    | -.067* (.009)   | -.067* (.007)        | -.089* (0.014) |
| **Long-run effects** |                  |                  |                      |       |
| Personal income      | -.274* (.074)    | -.029 (.046)    | -.029 (.047)         | .040 (0.047)  |
| Population           | -.519* (.116)    | -.653* (.135)   | -.653* (.139)        | -.155 (0.217) |
| Industrial share     | -.102* (.048)    | -.104* (.044)   | -.104* (.016)        | -.082# (0.032) |
| Top 10% share        | -.033 (.026)     | -.065* (.032)   | -.065* (.029)        | .025 (0.027)  |
| Renewable energy (%) | .074* (.010)     | .062* (.010)    | .062* (.007)         | .042# (0.010) |
| **CSA variables (lags)** |            |                  |                      |       |
| Personal income      | .311* (.071)     | .472* (.181)    | .472* (.249)         | .062 (1.12)   |
| Population           | .457* (.118)     | .579* (.137)    | .579* (.107)         | .486# (2.09)  |
| Industrial share     | .370* (.057)     | .326* (.045)    | .326* (.075)         | .310* (0.74)  |
| Top 10% share        | -.017 (.177)     | -.051 (.272)    | -.051 (.264)         | .063 (0.91)   |
| Renewable energy (%) | -.126* (.026)    | -.065* (.036)   | -.065* (.034)        | -.112* (0.30) |
| Observations (N/T)   | 2,958 (51/58)    | 2,958 (51/58)   | 2,958 (51/58)        | 2,805 (51/55) |
| CD statistic         | 70.650*          | -3.875*         | -3.875*              | -4.60  |
| CIPS                 | -6.164*          | -6.060*         | -6.075*              | -6.162* |
| CSA variables (lags) |                  |                  |                      |       |

Note: Standard errors are in parentheses. Generated with the `xtreg`, `xtpcse`, and `xtdcce2` (Ditzen 2018) commands in Stata 17. 1FE = one-way fixed effects; 2FE = two-way fixed effects; PCSE = panel corrected standard errors; CCE = common correlated effects; CD-Statistic = Test Statistic forWeak Cross-Sectional Dependence Test; CIPS = Pesaran Panel Unit Root Test in the Presence of Cross-Sectional Dependence; CSA = crosssectional average. *p < .05; #p < .10.

Table 9. Static Models of Fossil Fuel Consumption, 1960–2018.

|                      | 1FER | 2FER | 2FEP<sup>C</sup>PSAR | CCE  |
|----------------------|------|------|-----------------------|------|
| Δ Personal income    | .384* (.092) | .060 (.046) | .069 (.045) | .024 (.042) |
| Δ Population         | .780* (.139) | .786* (.161) | .803* (.121) | .520* (.107) |
| Δ Industrial share   | .127* (.052) | .121* (.047) | .118* (.016) | .210* (.041) |
| Δ Top 10% share      | .035 (.028) | .070* (.030) | .071* (.028) | -.014 (.023) |
| Δ Renewable energy (%) | -.086* (.011) | -.066* (.011) | -.064* (.007) | -.088* (.013) |
| Observations (N/T)   | 2,958 (51/58) | 2,958 (51/58) | 2,958 (51/58) | 2,958 (51/58) |
| CD statistic         | 76.859* | -3.792* | 177.513* | -2.32 |
| CIPS                 | -6.151* | -6.161* | -2.565* | -6.132* |
| CSA variables        | Δ Fossil fuel consumption, renewable energy (%) |       |       |       |

Note: Δ = first difference. Standard errors are in parentheses. Generated with the `xtreg`, `xtpcse`, and `xtdcce2` (Ditzen 2018) commands in Stata 17. 1FE = one-way fixed effects; 2FE = two-way fixed effects; PCSE = panel corrected standard errors; PSAR(1) = panel-specific autoregressive(1) correction; CCE = common correlated effects; CD-Statistic = Test Statistic for Weak Cross-Sectional Dependence Test; CIPS = Pesaran Panel Unit Root Test in the Presence of Cross-Sectional Dependence; CSA = crosssectional average. *p < .05.

---

The autoregressive coefficient is near unity in the fixed effects models, while it is .504 for the DCCE model. The short-run coefficients and the coefficients on the lags of the independent variables tend to take on equal and opposite signs in the fixed effects models, which is also expected. Pesaran’s test for weak cross-sectional dependence (reported
as the CD-statistic in Table 8) indicates strong cross-sectional dependence in the 1DFE. The CD statistic is less in magnitude with the inclusion of time fixed effects, but the statistic is significant at the .05 level. The use of PCSEs does little to mitigate the strong cross-sectional dependence because the CD statistic remains the same value.

Turning to the short-run effects, industrial share and population are positively associated with fossil fuel consumption, while renewable energy is negatively associated with fossil fuel consumption across all four models. There is disagreement between the models regarding the short-run impacts of personal income and the top 10 percent share. The 1DFER model finds that personal income is positive and statistically significant, whereas the 2DFE models and the DCCE find no statistically significant relationship. The top 10 percent share coefficient is positive and statistically significant for both 2DFE models, but it is not statistically significant in the 1DFE or DCCE model.

The long-run effects are similar across the four models, but there are some notable differences between the estimators. The DCCE finds that personal income is not associated with fossil fuel consumption, but the 1DFE and 2DFE estimate a sizeable long-run association between income and fossil fuel consumption. The discrepancy between the 1/2DFE and DCCE arises primarily from the autoregressive coefficient tending toward unity, which creates highly misleading results, as theory predicts (Pesaran and Smith 1995) and the Monte Carlo experiments showed. Likewise, although population is statistically significant across all of the models, the 1DFE and 2DFE estimates are counterintuitive because they indicate that the short-run effect is greater than the long-run effect. For the 2DFE models, a 1 percent increase in population is associated with a .697 percent increase in fossil fuel consumption at $T = 1$, but the effect is .579 percent in the long run. Given that prior research consistently finds that population has at least a proportionate impact on environmental outcomes, it is unlikely that the effect would become smaller over time (Rosa and Dietz 2012).

In the static models (Table 9), the differences between the models are similar to the dynamic case. First differences are used because the 2FE models in levels are nonstationary according to the Pesaran panel unit-root test in the presence of cross-section dependence (available on request). Personal income is positive and statistically significant for the 1FE but not for the 2FE models or the CCE. The top 10 percent share coefficient is statistically significant in both 2FE models but not the other models. The other coefficients are similar across the models. Both 2FE models suffer from strong cross-sectional dependence, and the CD statistic for 2FEPCSE/PSAR is even larger in absolute magnitude than the 1FE model—suggesting the use of PCSEs does not mitigate strong cross-sectional dependence. A summary of the results across Table 8 and Table 9 is found in Table 10.

### Conclusion and Recommendations for Modeling Large N, Large T Data

There are three key takeaways from the Monte Carlo experiments and the example presented. First, failing to correctly model the lag structure, slope heterogeneity, and cross-sectional dependence produces biased and inconsistent results. Second, the performance of the dynamic two-way fixed effects estimator relative to the common correlated effects estimator depends on $T$. Third and finally, use of the one-way fixed effects estimator requires the researcher to make strong...
Thombs

assumptions that are often implausible. Even when it is theoretically appropriate to use, it performs similarly to other estimators. Its use is only appropriate when the researcher has strong evidence that there is no slope heterogeneity or cross-sectional dependence.

Based on the theoretical discussion earlier in the article and the takeaways from the experiments, I present a decision-making framework in flowchart form in Figure 2. As the framework illustrates, there are theoretical and statistical decisions the researcher must make. Researchers must first decide the appropriate ARDL\((p, q)\) or ECM\((p, q)\) model to estimate. In Step 2, they should test for slope heterogeneity and then, based on these results, estimate a model using the MG estimator (3a) or 1FE estimator (3b). Following estimation, researchers should test for strong cross-sectional dependence in the residuals. If there is no strong cross-sectional dependence, then researchers can report their estimates from 3a or 3b. If there is strong cross-sectional dependence, then they should use the CCE (5a) or the pooled CCE (assumes homogeneous slopes) or the 2FE estimator (5b) depending on whether it eliminates the strong cross-sectional dependence.

To conclude, I make several remarks and suggestions regarding the framework:

1. **Use theory to guide the modeling approach.** Rather than assuming fixed effects estimation of a static model is appropriate, researchers should explicitly connect their modeling strategy to the theories they are testing. Researchers should consider whether theories are static or dynamic in nature and whether effects are heterogeneous.

   If there is uncertainty regarding lag specification, the best way to tell whether a dynamic model is appropriate is to include lags of the dependent (and independent) variables in the model. Including them will at once model the dynamic process and correct the issue of autocorrelation (Pickup 2015). If a standard error correction is used instead, it will be misspecified if the data generating process is truly dynamic. Thus, researchers should always start with a dynamic model unless they have a strong theoretical reason to believe otherwise.

   Moreover, many sociological theories expect that social processes function differently across spatial and social contexts. Thus, the homogeneity assumptions of the fixed effects estimator often are too restrictive, making the flexibility of heterogeneous panel estimators like the CCE more apt to use in many cases.
2. **Estimate several models to consider the impact of lag misspecification.** Although using theory can give researchers a general idea whether processes are static or dynamic, they should always estimate several different models to derive robust estimates. Estimating multiple models can help researchers better understand their data and results, and it provides transparency to readers. Following a general-to-specific approach to modeling (Hendry 1995), researchers should start by specifying a general model that serves as a probable representation of the underlying data generating process and then restrict the model as necessary if lags are not statistically significant. Starting with an ARDL(1,1) model is common. However, more lags can be added, and researchers should examine how changing the lag length influences their results. Of course, this is not a flawless approach. Using too few lags can lead to biased estimates, and too many lags can reduce efficiency and affect power (Chudik et al. 2016).

3. **Testing for slope heterogeneity and strong cross-sectional dependence.** Testing for strong cross-sectional dependence and slope heterogeneity is easy to do in common software programs like Stata. Pesaran’s test for weak cross-sectional dependence (the CD test) can be implemented with the xtdc2 package (Ditzen 2018), and the exponent test for cross-sectional dependence developed by Bailey et al. (2016) can be implemented with the xtcse2 command (Ditzen 2021). The exponent test estimates the strength of the common factors responsible for cross-sectional dependence. For slope heterogeneity, Bersvendsen and Ditzen’s (2021) xthst command can be used. Finding that the fixed effects estimator produces a coefficient on $x_{i,t} \cdot 1$ and equal and opposite coefficients on $x_{i,t}$ and $x_{i,t-1}$ can also alert researchers that slope heterogeneity is likely an issue.

4. **The number of time periods matters.** Taken together, these recommendations suggest that the common correlated effects estimator is more appropriate to use in high slope heterogeneity, so the fixed effects estimator may be more appropriate to use in this case.

**Acknowledgments**

I thank Andrew Jorgenson for helpful comments and feedback.

**ORCID iD**

Ryan P. Thombs https://orcid.org/0000-0002-8823-6143

**Supplemental Material**

Supplemental material for this article is available online.

**References**

Allison, Paul D. 2009. *Fixed Effects Regression Models.* Thousand Oaks, CA: SAGE.

Allison, Paul D., Richard Williams, and Enrique Moral-Benito. 2017. “Maximum Likelihood for Cross-Lagged Panel Models with Fixed Effects.” *Socius* 3. doi:10.1177/2378023117710578.

Bailey, Natalia, George Kapetanios, and M. Hashem Pesaran. 2016. “Exponent of Cross-Sectional Dependence: Estimation and Inference.” *Journal of Applied Econometrics* 31(6):929–60.

Beck, Nathaniel, and Jonathan N. Katz. 1995. “What to Do (and Not to Do) with Time-Series Cross-Section Data.” *The American Political Science Review* 89(3):634–47.

Beckfield, Jason. 2003. “Inequality in the World Polity: The Structure of International Organization.” *American Sociological Review* 68(3):401–24.

Bersvendsen, Tore, and Jan Ditzen. 2021. “Testing for Slope Heterogeneity in Stata.” *The Stata Journal* 21(1):51–80.

Bureau of Economic Analysis. 2021. “Personal Income by State.” https://www.bea.gov/data/income-saving/personal-income-by-state.

Cameron, A. Colin, and Pravin K. Trivedi. 2010. *Microeconometrics Using Stata.* College Station, TX: Stata Press.

Chase-Dunn, Christopher. 1998. *Global Formation: Structures of the World-Economy.* Lanham, MD: Rowman & Littlefield.

Chudik, Alexander, Kamiar Mohaddes, M. Hashem Pesaran, and Mehdi Raissi. 2016. “Long-Run Effects in Large Heterogeneous Panel Data Models with Cross-Sectionally Correlated Errors.” Pp. 85–135 in *Essays in Honor of Man Ullah. Advances in Econometrics,* vol. 36, edited by G. González-Rivera, R. Carter Hill, and T.-H. Lee. Bingley, UK: Emerald Group Publishing Limited.

Chudik, Alexander, and M. Hashem Pesaran. 2015a. “Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors.” *Journal of Econometrics* 188(2):393–420.

Chudik, Alexander, and M. Hashem Pesaran. 2015b. “Large Panel Data Models with Cross-Sectional Dependence: A Survey.” Pp. 2–45 in *The Oxford Handbook of Panel Data*, edited by B. H. Baltagi. Oxford, UK: Oxford University Press.

Chudik, Alexander, M. Hashem Pesaran, and Elisa Tosetti. 2011. “Weak and Strong Cross-Section Dependence and Estimation of Large Panels.” *The Econometrics Journal* 14(1):C45–90.
De Boef, Suzanna, and Luke Keele. 2008. “Taking Time Seriously.” *American Journal of Political Science* 52(1):184–200.

Dietz, Thomas. 2017. “Drivers of Human Stress on the Environment in the Twenty-First Century.” *Annual Review of Environment and Resources* 42(1):189–213.

Ditzen, Jan. 2018. “Estimating Dynamic Common-Correlated Effects in Stata.” *The Stata Journal* 18(3):585–617.

Ditzen, Jan. 2021. “Estimating Long-Run Effects and the Exponent of Cross-Sectional Dependence: An Update to xtdcce2.” *The Stata Journal* 21(2):687–707.

Driscol, John C., and Aart C. Kraay. 1998. “Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data.” *The Review of Economics and Statistics* 80(4):549–60.

Engle, Robert F., and C. W. J. Granger. 1987. “Co-Integration and Error Correction: Representation, Estimation, and Testing.” *Econometrica* 55:251–76.

Eveaert, Gerdie, and Tom De Groote. 2016. “Common Correlated Effects Estimation of Dynamic Panels with Cross-Sectional Dependence.” *Econometric Reviews* 35(3):428–63.

Frank, Mark W. 2021. “U.S. State Level Income Inequality Data.” https://www.shsu.edu/eecomf/inequality.html.

Givens, Jennifer E., Xiaorui Huang, and Andrew K. Jorgenson. 2019. “Ecologically Unequal Exchange: A Theory of Global Environmental Injustice.” *Sociology Compass* 13(5):e12693. doi:10.1111/soc4.12693.

Hendry, David F. 1995. *Dynamic Econometrics*. Oxford, UK: Oxford University Press.

Jorgenson, Andrew K., and Brett Clark. 2012. “Are the Economy and the Environment Decoupling? A Comparative International Study, 1960–2005.” *American Journal of Sociology* 118(1):1–44.

Jorgenson, Andrew K., Shirley Fiske, Klaus Hubacek, Jia Li, Tom McGovern, Torben Rick, Juliet B. Schor, William Solecki, Richard York, and Ariela Zycherman. 2019. “Social Science Perspectives on Drivers of and Responses to Global Climate Change.” *Wiley Interdisciplinary Reviews: Climate Change* 10(1):e554. doi:10.1002/wcc.554.

Kelly, Orla. 2020. “The Silver Bullet? Assessing the Role of Education for Sustainability.” *Social Forces* 99(1):178–204.

Kollmeyer, Christopher, and John Peters. 2019. “Financialization and the Decline of Organized Labor: A Study of 18 Advanced Capitalist Countries, 1970–2012.” *Social Forces* 98(1):1–30.

Leszczenksy, Lars, and Tobias Wolbring. 2022. “How to Deal with Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study” *Sociological Methods & Research* 51(2):837–65.

Pesaran, M. Hashem. 2006. “Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure.” *Econometrica* 74(4):967–1012.

Pesaran, M. Hashem. 2015. “Testing Weak Cross-Sectional Dependence in Large Panels.” *Econometric Reviews* 34:1089–117.

Pesaran, M. Hashem, Yongcheol Shin, and Ron P. Smith. 1999. “Pooled Mean Group Estimation of Dynamic Heterogeneous Panels.” *Journal of the American Statistical Association* 94(446):621–34.

Pesaran, M. Hashem, and Ron Smith. 1995. “Estimating Long-Run Relationships from Dynamic Heterogeneous Panels.” *Journal of Econometrics* 68(1):79–113.

Pickup, Mark. 2015. *Introduction to Time Series Analysis*. Thousand Oaks, CA: SAGE.

Plümper, Thomas, and Vera E. Troeger. 2019. “Not So Harmless After All: The Fixed-Effects Model.” *Political Analysis* 27(1):21–45.

Rosa, Eugene A., and Thomas Dietz. 2012. “Human Drivers of National Greenhouse-Gas Emissions.” *Nature Climate Change* 2:581–86.

Thombs, Ryan P., Xiaorui Huang, and Jared B. Fitzgerald. 2022. “What Goes Up Might Not Come Down: Modeling Directional Asymmetry with Large N, Large T Data.” *Sociological Methodology* 52(1):1–29.

U.S. Energy Information Administration. 2021. “State Energy Data System (SEDS).” https://www.eia.gov/state/seds/seds-data-complete.php?sid=US.

Vaisey, Stephen, and Andrew Miles. 2017. “What You Can—and Can’t—Do With Three-Wave Panel Data.” *Sociological Methods & Research* 46(1):44–67.

VanHeuvelen, Tom, and Nik Summers. 2019. “Divergent Roads: A Cross-National Intercohort Analysis of Affluence and Environmental Concern.” *Social Science Research* 82:72–91.

Webb, Clayton, Suzanna Linn, and Matthew J. Lebo. 2020. “Beyond the Unit Root Question: Uncertainty and Inference.” *American Journal of Political Science* 64(2):275–92.

World Bank. 2021. “World Development Indicators.” https://data.worldbank.org/source/world-development-indicators.

York, Richard, Eugene A. Rosa, and Thomas Dietz. 2003. “Human Drivers of Global Environmental Injustice.” *American Sociological Review* 68(2):279–300.

**Author Biography**

Ryan P. Thombs is a PhD candidate in the sociology department at Boston College. His research examines the drivers of global environmental change, the drivers of health disparities and population health outcomes, and quantitative methodology, particularly related to panel data and time-series modeling. His work appears in journals such as *Sociological Methodology*, the *Journal of Health and Social Behavior*, *Sociological Forum*, *The Sociological Quarterly*, *Sociology of Development*, *Climatic Change*, and *Energy Research & Social Science.*