The Stata Journal (2022)
22, Number 3, pp. 713–724 DOI: 10.1177/1536867X221124567

xtusreg: Software for dynamic panel regression under irregular time spacing

Yuya Sasaki
Vanderbilt University
Nashville, TN
yuya.sasaki@vanderbilt.edu

Yi Xin
California Institute of Technology
Pasadena, CA
yixin@caltech.edu

Abstract. We introduce a new command, `xtusreg`, that estimates parameters of fixed-effects dynamic panel regression models under unequal time spacing. After reviewing the method, we examine the finite-sample performance of the command using simulated data. We also illustrate the command with the National Longitudinal Survey Original Cohorts: Older Men, whose personal interviews took place in the unequally spaced years of 1966, 1967, 1969, 1971, 1976, 1981, and 1990. The methods underlying `xtusreg` are those discussed by Sasaki and Xin (2017, *Journal of Econometrics* 196: 320–330).

Keywords: st0690, xtusreg, dynamic panel regression, fixed effect, unequal time spacing

1 Introduction

Conventional methods to fit fixed-effects dynamic panel regression models require that a researcher observe three consecutive time periods or two pairs of two consecutive time periods. On the other hand, there are some panel datasets with irregularly spaced time intervals that do not satisfy these conventional time-spacing requirements for identification and estimation. For example, personal interviews were conducted in 1966, 1967, 1969, 1971, 1976, 1981, and 1990 for the National Longitudinal Survey (NLS) Original Cohorts: Older Men, and there are neither three consecutive time periods nor two pairs of two consecutive time periods in this list of years. Other examples with irregularly spaced panel data from the United States include Current Population Survey, Early Childhood Longitudinal Survey-K, NLS of Youth 1979, and Panel Survey of Income Dynamics—see section 2 for more details about specific survey periods of these datasets.

Even if panel data exhibit such irregular and unequal time spacing, Sasaki and Xin (2017) show that the fixed-effects dynamic panel regression parameters can be still identified as long as "two pairs of two consecutive time gaps" are available, which generalizes the conventional requirement of "two pairs of two consecutive time periods". In the NLS Original Cohorts: Older Men, for instance, there is a time gap of 0 between 1966 and 1966, a time gap of 1 between 1966 and 1967, a time gap of 2 between 1967 and 1969, and a time gap of 3 between 1966 and 1969. Thus, there are two pairs, $(0, 1)$ and $(2, 3)$, of consecutive time gaps in this panel dataset. This requirement is also satisfied by the list of the aforementioned panel datasets: Current Population Survey, Early

Childhood Longitudinal Survey-K, NLS of Youth 1979, and Panel Survey of Income Dynamics.

In this article, we introduce the `xtusreg` command, which executes estimation and inference for fixed-effects dynamic panel regression under irregular time spacing based on the method proposed in Sasaki and Xin (2017). After reviewing the method in section 2, we formally introduce the command in section 3. This is followed by simulation studies in section 4 and a real data illustration in section 5.

2 Review of the method

This section reviews a part of the method of identification and estimation of fixed-effects dynamic panel regression models under unequal time spacing proposed by Sasaki and Xin (2017). Consider the model

$$y_{it} = \gamma y_{i,t-1} + \beta x_{it} + \alpha_i + \varepsilon_{it} \tag{1}$$

where y_{it} denotes an observed state variable, x_{it} denotes an observed covariate, α_i denotes an unobserved individual fixed effect, and ε_{it} denotes an unobserved idiosyncratic shock. A researcher is often interested in the autoregressive parameter γ and the regression parameter β in such a model. The following two examples illustrate concrete regressions in which economists would be interested.

Example 1. The seminal article by Ashenfelter (1978) proposes models of the form (1) as "earnings generating functions" to analyze the effect β of job training x_{it} on the subsequent earnings y_{it}, accounting for the dynamics of y_{it}. With equally spaced panel datasets, such as the one used in Ashenfelter (1978), standard methods such as `xtabond` may be used to estimate such an effect of interest. However, with unequally spaced panel datasets, such as the NLS Original Cohorts: Older Men (which also contains information about earnings and exposure to job training), those existing commands incur biased estimates in general.

Example 2. In economics of education, dynamic models of the form (1) are often used as "value-added models" to analyze persistent impact β of intervention x_{it} on the human capital y_{it} of children. The Early Childhood Longitudinal Survey-K is one of the most relevant datasets to study the value-added models, but this dataset is unequally spaced as $T = 3$, 4, 8, 12, and 18. With the unequal spacing, standard commands such as `xtabond` incur biased estimates of the value-added models in general.

For convenience of writing, we first define a few shorthand and auxiliary notations. Write $E_i(\cdot) := E(\cdot|\alpha_i)$ for the expectation conditional on individual i's specific fixed effect α_i. With this notation, we in turn define the auxiliary random variables

$$Z_{i\tau} := E_i(y_{it} y_{i,t+\tau}) \qquad z_{i\tau} := E_i(x_{it} x_{i,t+\tau})$$
$$\zeta_{i\tau} := E_i(y_{it} x_{i,t+\tau}) \qquad \zeta_{i,-\tau} := E_i(x_{it} y_{i,t+\tau})$$

for each time t and time gap τ. Furthermore, we denote their cross-sectional means by $Z_\tau := E(Z_{i\tau})$, $z_\tau := E(z_{i\tau})$, $\zeta_\tau := E(\zeta_{i\tau})$, and $\zeta_{-\tau} := E(\zeta_{i,-\tau})$.

Let T be the set of unequally spaced time periods t for which a researcher observes the panel data $\{(y_{it}, x_{it})\}_{i=1}^{N}$. We define the set of survey gaps by

$$\mathcal{T} = \{|t_1 - t_2| : t_1, t_2 \in T\}$$

We also define the set of gap-associated survey years by

$$T(\tau) = \{t \in T : t + \tau \in T\}$$

for each time gap $\tau \in \mathcal{T}$ and let $T(\tau) = \emptyset$ if $\tau \notin \mathcal{T}$. For the NLS Original Cohorts Older Men, for example, personal interviews were conducted in 1966, 1967, 1969, 1971, 1976, 1981, and 1990. In this case, we can write $T = \{1966, 1967, 1969, 1971, 1976, 1981, 1990\}$, $\mathcal{T} = \{0, 1, 2, 3, 4, 5, 7, 9, 10, 12, 14, 15, 19, 21, 23, 24\}$, $T(0) = T$, $T(1) = \{1966\}$, $T(2) = \{1967, 1969\}$, $T(3) = \{1966\}$, $T(4) = \{1967\}$, $T(5) = \{1966, 1971, 1976\}$, and so on.

If panel data with unequal spacing have $T(1) \neq \emptyset$, $T(\Delta t) \neq \emptyset$, and $T(\Delta t + 1) \neq \emptyset$ for some gap Δt in the set of natural numbers, then we call its spacing structure the "U.S. spacing" (compare Sasaki and Xin [2017, def. 2]). For instance, one can verify that each of the following datasets has this U.S. spacing structure:

NLS Original Cohorts: Older Men	$T = \{1966, 1967, 1969, 1971, 1976, \ldots\}$
Current Population Survey	$T = \{\ldots, 3, 4, 13, 14, \ldots\}$
Early Childhood Longitudinal Survey-K	$T = \{3, 4, 8, 12, 18\}$
NLS of Youth 1979	$T = \{\ldots, 1993, 1994, 1996, 1998, \ldots\}$
Panel Study of Income Dynamics	$T = \{\ldots, 1996, 1997, 1999, 2001, \ldots\}$

See example 2 in Sasaki and Xin (2017) for detailed discussions.

Under the U.S. spacing, Sasaki and Xin [2017, corollary 1 (ii)] show that $(\gamma, \beta)'$ can be identified by

$$\begin{pmatrix} \gamma \\ \beta \end{pmatrix} = \frac{1}{|\Delta|} \begin{pmatrix} (z_{\Delta t+1} - z_1)(Z_{\Delta t+1} - Z_1) + (\zeta_1 - \zeta_{\Delta t+1})(\zeta_{-(\Delta t+1)} - \zeta_{-1}) \\ (\zeta_0 - \zeta_{-\Delta t})(Z_{\Delta t+1} - Z_1) + (Z_{\Delta t} - Z_0)(\zeta_{-(\Delta t+1)} - \zeta_{-1}) \end{pmatrix}$$

with

$$|\Delta| := (Z_{\Delta t} - Z_0)(z_{\Delta t+1} - z_1) - (\zeta_{-\Delta t} - \zeta_0)(\zeta_{\Delta t+1} - \zeta_1)$$

provided that $|\Delta| \neq 0$ holds.

Given the identification result, one can now take the sample counterpart of it to obtain an estimator. The auxiliary random variables, Z_τ, z_τ, ζ_τ, and $\zeta_{-\tau}$, may be estimated, respectively, by

$$\widehat{Z}_\tau = \frac{1}{N}\sum_{i=1}^{N} \overline{Z}_{i\tau} \qquad \widehat{z}_\tau = \frac{1}{N}\sum_{i=1}^{N} \overline{z}_{i\tau} \qquad \widehat{\zeta}_\tau = \frac{1}{N}\sum_{i=1}^{N} \overline{\zeta}_{i\tau} \qquad \widehat{\zeta}_{-\tau} = \frac{1}{N}\sum_{i=1}^{N} \overline{\zeta}_{i,-\tau}$$

for any $\overline{Z}_{i\tau}$, $\overline{z}_{i\tau}$, $\overline{\zeta}_{i\tau}$, and $\overline{\zeta}_{i,-\tau}$ such that $Z_\tau = E(\overline{Z}_{i\tau})$, $z_\tau = E(\overline{z}_{i\tau})$, $\zeta_\tau = E(\overline{\zeta}_{i\tau})$, and $\zeta_{-\tau} = E(\overline{\zeta}_{i,-\tau})$. Under a time-invariant moment assumption (assumption 2 of Sasaki and Xin [2017]), $\overline{Z}_{i\tau}$, $\overline{z}_{i\tau}$, $\overline{\zeta}_{i\tau}$, and $\overline{\zeta}_{i,-\tau}$ can be constructed by any linear combination of the form

$$\overline{Z}_{i\tau} = \sum_{t \in T(\tau)} a_t^\tau y_{it} y_{i,t+\tau} \qquad \overline{z}_{i\tau} = \sum_{t \in T(\tau)} b_t^\tau x_{it} x_{i,t+\tau}$$

$$\overline{\zeta}_{i\tau} = \sum_{t \in T(\tau)} c_t^\tau y_{it} x_{i,t+\tau} \qquad \overline{\zeta}_{i,-\tau} = \sum_{t \in T(\tau)} d_t^\tau x_{it} y_{i,t+\tau}$$

where $a^\tau = (a_t^\tau)_{t \in T(\tau)}$, $b^\tau = (b_t^\tau)_{t \in T(\tau)}$, $c^\tau = (c_t^\tau)_{t \in T(\tau)}$, and $d^\tau = (d_t^\tau)_{t \in T(\tau)}$ satisfy, respectively, $\sum_{t \in T(\tau)} a_t^\tau = 1$, $\sum_{t \in T(\tau)} b_t^\tau = 1$, $\sum_{t \in T(\tau)} c_t^\tau = 1$, and $\sum_{t \in T(\tau)} d_t^\tau = 1$. The `xtusreg` command uses the simple arithmetic means. The sample counterpart of the identifying formula yields the explicit estimator

$$\begin{pmatrix} \widehat{\gamma} \\ \widehat{\beta} \end{pmatrix} = \frac{1}{|\widehat{\Delta}|} \begin{pmatrix} \left(\widehat{z}_{\Delta t+1} - \widehat{z}_1\right)\left(\widehat{Z}_{\Delta t+1} - \widehat{Z}_1\right) + \left(\widehat{\zeta}_1 - \widehat{\zeta}_{\Delta t+1}\right)\left(\widehat{\zeta}_{-(\Delta t+1)} - \widehat{\zeta}_{-1}\right) \\ \left(\widehat{\zeta}_0 - \widehat{\zeta}_{-\Delta t}\right)\left(\widehat{Z}_{\Delta t+1} - \widehat{Z}_1\right) + \left(\widehat{Z}_{\Delta t} - \widehat{Z}_0\right)\left(\widehat{\zeta}_{-(\Delta t+1)} - \widehat{\zeta}_{-1}\right) \end{pmatrix} \tag{2}$$

where

$$|\widehat{\Delta}| = \left(\widehat{Z}_{\Delta t} - \widehat{Z}_0\right)\left(\widehat{z}_{\Delta t+1} - \widehat{z}_1\right) - \left(\widehat{\zeta}_{-(\Delta t)} - \widehat{\zeta}_0\right)\left(\widehat{\zeta}_{\Delta t+1} - \widehat{\zeta}_1\right)$$

While the above procedure focuses on the just-identified case for the sake of clarity, the generic generalized method of moments (GMM) restriction is provided by

$$E\left\{\boldsymbol{g}\left(\boldsymbol{w}_i, \boldsymbol{\theta}_0\right)\right\} = \boldsymbol{0}$$

where $\boldsymbol{\theta} = (\gamma, \beta)$, $\boldsymbol{w}_i = (x_{it}, y_{it})_{t \in T}$, and the rows of the moment function $\boldsymbol{g}(\boldsymbol{w}_i, \boldsymbol{\theta})$ consist of

$$\begin{aligned} g_{1,tt't''t'''}(\boldsymbol{w}_i, \boldsymbol{\theta}) = &\left(y_{i,t'''} y_{i,t'''+\Delta t+1} - y_{i,t'} y_{i,t'+1}\right) \\ &- \gamma\left(y_{i,t''} y_{i,t''+\Delta t} - y_{it} y_{it}\right) - \beta\left(y_{i,t'''} x_{i,t'''+\Delta t+1} - y_{i,t'} x_{i,t'+1}\right) \\ g_{2,tt't''t'''}(\boldsymbol{w}_i, \boldsymbol{\theta}) = &\left(x_{i,t'''} y_{i,t'''+\Delta t+1} - x_{i,t'} y_{i,t'+1}\right) \\ &- \gamma\left(x_{i,t''} y_{i,t''+\Delta t} - x_{it} y_{it}\right) - \beta\left(x_{i,t'''} x_{i,t'''+\Delta t+1} - x_{i,t'} x_{i,t'+1}\right) \end{aligned}$$

for all $(t, t', t'', t''') \in T(0) \times T(1) \times T(\Delta t) \times T(\Delta t + 1)$. The `xtusreg` command implements the GMM based on these moment restrictions, so it can handle both the just-identified and overidentified cases. Specifically, the estimator is given by

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta} \left\{\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}(\boldsymbol{w}_i, \boldsymbol{\theta})\right\}' \boldsymbol{W}_N \left\{\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}(\boldsymbol{w}_i, \boldsymbol{\theta})\right\}$$

where $\boldsymbol{W}_N$ is a weighting matrix. The `xtusreg` command uses the identity matrix in the first step and uses the estimated optimal weighting matrix in the second step. In the just-identified case, the GMM estimator coincides with the closed-form estimator (2). Even with this GMM framework, we emphasize that it is still necessary for panel data to exhibit the U.S. spacing in order for the command to run.

The variance matrix of the GMM estimator is calculated based on the asymptotic normality

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right) \stackrel{d}{\to} N\left\{0,(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}\boldsymbol{G}\,(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}\right\}$$

where $\boldsymbol{S}$ is the variance matrix of $\boldsymbol{g}(\boldsymbol{w}_i,\boldsymbol{\theta}_0)$ and $\boldsymbol{G}$ is given by

$$\boldsymbol{G} = E\begin{bmatrix} \vdots & \vdots \\ -(y_{i,t''}y_{i,t''+\Delta t} - y_{it}y_{it}) & -(y_{i,t'''}x_{i,t'''+\Delta t+1} - y_{i,t'}x_{i,t'+1}) \\ \vdots & \vdots \\ -(x_{i,t''}y_{i,t''+\Delta t} - x_{it}y_{it}) & -(x_{i,t'''}x_{i,t'''+\Delta t+1} - x_{i,t'}x_{i,t'+1}) \\ \vdots & \vdots \end{bmatrix}$$

See theorem 2 of Sasaki and Xin (2017).

In addition to the U.S. spacing, there is another spacing pattern, called the "U.K. spacing" (Sasaki and Xin 2017, ex. 1). However, the `xtusreg` command introduced in this article does not handle the U.K. spacing, because it tends to incur a weak identification as reported in Sasaki and Xin (2017, sec. 5). A weak-identification–robust approach to inference is recommended under the U.K. spacing, and this suggests a direction for future research in development of a Stata program.

The `xtusreg` command introduced in this article does not use all the combinations of two pairs of two consecutive time gaps. Instead, it uses two pairs of the two smallest consecutive time gaps. This is because larger time gaps tend to incur larger finite-sample biases in practice.

Before forming the moment restrictions presented above, the `xtusreg` command introduced in this article automatically drops observations with missing values. Further, it uses the rectangle of a balanced portion of the panel data.

3 The xtusreg command

3.1 Syntax

The syntax of the `xtusreg` command is as follows:

`xtusreg` *depvar* [*indepvars*] [*if*] [*in*] [`, onestep stationary sweight(`*var*`)`
`gamma(`*real*`) beta(`*real*`)`]

Here *depvar* stands for the dependent variable y, and *indepvars* include independent variables x. Exactly one *depvar* variable should be included to run the command, while *indepvars* are optional and may include multiple variables.

3.2 Options

The `xtusreg` command has five options.

`onestep` sets an indicator for the one-step GMM estimation. By default, two-step efficient GMM estimation is set. This option will not make a difference in the results if parameters are just identified.

`stationary` sets an indicator for not executing the location-scale normalization of variables. By default, implementation of the location-scale normalization is set. Appendix C.1 of Sasaki and Xin (2017) recommends the implementation of the location-scale normalization under nonstationary distributions.

`sweight(`*var*`)` sets sampling weights. By default, uniform weights for all the observations are set.

`gamma(`*real*`)` sets the initial value of the autoregressive coefficient for a numerical optimization in the GMM estimation. The default is `gamma(0)`.

`beta(`*real*`)` sets the initial value of the regression coefficients for a numerical optimization in the GMM estimation. The default is `beta(0)`.

3.3 Stored results

`xtusreg` stores the following in `e()`:

Scalars	
`e(NT)`	number of observations
`e(N)`	number of cross-sectional units
`e(T)`	number of time periods
`e(objective)`	value of the GMM objective
Macros	
`e(cmd)`	`xtusreg`
`e(steps)`	number of GMM steps: `one` or `two`
`e(properties)`	`b V`
Matrices	
`e(b)`	coefficient vector
`e(V)`	variance–covariance matrix of the estimators
`e(tlist)`	list of time periods
`e(gap1)`	first pair of consecutive time gaps
`e(gap2)`	second pair of consecutive time gaps
Functions	
`e(sample)`	marks estimation sample

4 Simulation studies

In this section, we present the finite sample performance of the `xtusreg` command using simulated data. We provide do-files to implement all the simulation results presented below.

We generate data following the time-spacing pattern of the NLS Original Cohorts: Older Men. First, we consider a simple data-generating process with just a state variable y and without an observed predictor x. Independently generate

$$y_{i1} \sim \mathcal{N}(0, 1^2) \qquad \alpha_i \sim \mathcal{N}(0, 1^2)$$

for individuals $i = 1, \ldots, N$, where $N = 1000$. For the subsequent time periods $t = 2, \ldots, 70$, we iteratively and autoregressively generate

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it} \qquad \varepsilon_{it} \sim \mathcal{N}(0, 1^2)$$

where we set $\gamma = 0.5$ for the autoregressive coefficient. After we have accumulated the whole data $[(y_{it}) : i \in \{1, \ldots, N\},\ t \in \{1, \ldots, 70\}]$ for 70 periods, we drop all the time periods except for $t = 1966$, 1967, and 1969. This leaves us with $[(y_{it}) : i \in \{1, \ldots, N\},$ $t \in \{1966, 1967, 1969\}]$, similarly to the first three survey years for the NLS Original Cohorts: Older Men. The first row of table 1 shows simulation results based on 2,000 Monte Carlo iterations.

Table 1. Baseline simulation results

Unequally spaced	γ			
time periods	Bias	SD	RMSE	95%
$\{1966, 1967, 1969\}$	0.004	0.093	0.093	0.941
$\{1966, 1967, 1970\}$	0.000	0.045	0.045	0.937
$\{1966, 1967, 1971\}$	−0.000	0.043	0.043	0.946

The displayed statistics include the bias, standard deviation (SD), root mean squared error (RMSE), and 95% coverage frequency for the parameter, γ. The bias, SD, and RMSE are all small relative to the magnitude of the parameter value $\gamma = 0.5$, and the coverage frequencies by the 95% confidence interval are close to the nominal probability of 0.95. We repeat this simulation analysis using extended lists of unequally spaced time periods. The results based on $T = \{1966, 1967, 1970\}$ are displayed in the second row in table 1. The results based on $T = \{1966, 1967, 1971\}$ are displayed in the third row in table 1. The results are very similar to those found in the first row, and therefore the same conclusion applies.

Next, we consider a data-generating process with an observed predictor x as well as the state variable y. Independently generate

$$y_{i1} \sim \mathcal{N}(0, 1^2) \qquad x_{i1} \sim \mathcal{N}(0, 1^2) \qquad \alpha_i \sim \mathcal{N}(0, 0.5^2)$$

for individuals $i = 1, \ldots, N$, where $N = 1000$. For the subsequent time periods $t = 2, \ldots, 70$, we iteratively and autoregressively generate

$$y_{it} = \gamma y_{i,t-1} + \beta x_{it} + \alpha_i + \varepsilon_{it} \qquad \varepsilon_{it} \sim \mathcal{N}(0, 1^2)$$
$$x_{it} = \rho x_{i,t-1} + \alpha_i + \eta_{it} \qquad \eta_{it} \sim \mathcal{N}(0, 5^2)$$

where we set $\gamma = \beta = \rho = 0.5$. Again, we drop all the time periods except for $t = 1966$, 1967, and 1969, leaving us with $[(y_{it}, x_{it}) : i \in \{1, \ldots, N\},\ t \in \{1966, 1967, 1969\}]$.

The first row of table 2 shows simulation results based on 2,000 Monte Carlo iterations. We repeat this simulation analysis using extended lists of unequally spaced time periods. The results based on $T = \{1966, 1967, 1970\}$ are displayed in the second row in table 2. The results based on $T = \{1966, 1967, 1971\}$ are displayed in the third row in table 2.

Table 2. Simulation results for a model with a covariate

Unequally spaced	γ				β			
time periods	Bias	SD	RMSE	95%	Bias	SD	RMSE	95%
$\{1966, 1967, 1969\}$	−0.003	0.115	0.115	0.920	0.011	0.083	0.084	0.905
$\{1966, 1967, 1970\}$	−0.009	0.070	0.070	0.913	0.019	0.082	0.084	0.934
$\{1966, 1967, 1971\}$	−0.009	0.066	0.067	0.924	0.021	0.090	0.092	0.932

These unequally spaced time periods cannot be handled by conventional commands, such as `xtabond`, for fixed-effects dynamic panel regressions. On the other hand, the simulation results demonstrate that `xtusreg` can still produce estimates along with their valid standard errors.

5 Illustration of the command

In this section, we illustrate the `xtusreg` command with the real data of NLS Original Cohorts: Older Men.

In the software package, a subsample of this dataset can be loaded by the following command line:

```
. use nls_original_cohort
```

These data contain six variables: `id`, `year`, `logincome`, `age`, `edu`, and `white`. We first set the individual and time indices to refer to `id` and `year`, respectively, by the following command line:

```
. xtset id year
  (output omitted)
```

Having set the panel-data structure, we first run the simple autoregression of $y =$ `logincome` by the following line:

```
. xtusreg logincome
  (output omitted)
Number of observations:          8994
Number of cross-section units:   2998
Number of time periods:          3
List of time periods:            65, 66, 68
-----------------------------------------------------------------------------
L1 = Autoregressive Coefficient (gamma)
-----------------------------------------------------------------------------
             |  Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+---------------------------------------------------------------
          L1 |    .4785156   .1030199     4.64   0.000     .2766003     .680431
-----------------------------------------------------------------------------
Reference: Sasaki, Y. & Xin, Y. (2017) Unequal Spacing in Dynamic Panel Data:
Identification and Estimation. Journal of Econometrics 196 (2), pp. 320-330.
```

The estimate of the autoregressive parameter γ is a significant positive below one value, implying that the log income is positively autocorrelated and follows a stationary process. (Note that income was received in the years $\{1965, 1966, 1968\}$, while personal interviews were conducted in years $\{1966, 1967, 1969\}$.)

We next include `age` as a control x by the following line:

```
. xtusreg logincome age
  (output omitted)
Number of observations:          8994
Number of cross-section units:   2998
Number of time periods:          3
List of time periods:            65, 66, 68
-----------------------------------------------------------------------------
L1 = Autoregressive Coefficient (gamma)
-----------------------------------------------------------------------------
             |  Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+---------------------------------------------------------------
          L1 |    .1245117   .0198636     6.27   0.000     .0855799    .1634436
         age |   -.7397461   .2570027    -2.88   0.004    -1.243462     -.23603
-----------------------------------------------------------------------------
Reference: Sasaki, Y. & Xin, Y. (2017) Unequal Spacing in Dynamic Panel Data:
Identification and Estimation. Journal of Econometrics 196 (2), pp. 320-330.
```

The inclusion of the control affects the point estimate of the autoregressive parameter γ, but the level of statistical significance is larger than before. Furthermore, the coefficient β of `age` is significant.

The dataset has two remaining candidates, `white` and `edu`, of controls x. However, because both of these two variables are generally time invariant, the parameters of fixed-effects panel regression models may not be identified if these variables are included as controls. (This also applies to conventional commands to estimate fixed-effects panel regressions.) To account for the observed demographic heterogeneity, `white`, one can instead run a regression for each of the ethnic categories.

```
. xtusreg logincome if !white
```
(*output omitted*)
```
Number of observations:          2853
Number of cross-section units:   951
Number of time periods:          3
List of time periods:            65, 66, 68

L1 = Autoregressive Coefficient (gamma)
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
L1	.2578125	.1275462	2.02	0.043	.0078265	.5077985

```
Reference: Sasaki, Y. & Xin, Y. (2017) Unequal Spacing in Dynamic Panel Data:
Identification and Estimation. Journal of Econometrics 196 (2), pp. 320-330.
. xtusreg logincome if white
```
(*output omitted*)
```
Number of observations:          6141
Number of cross-section units:   2047
Number of time periods:          3
List of time periods:            65, 66, 68

L1 = Autoregressive Coefficient (gamma)
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
L1	.6328125	.144489	4.38	0.000	.3496193	.9160057

```
Reference: Sasaki, Y. & Xin, Y. (2017) Unequal Spacing in Dynamic Panel Data:
Identification and Estimation. Journal of Econometrics 196 (2), pp. 320-330.
```

The autoregressive parameter γ is larger for white individuals than for nonwhite individuals, implying that log income is more persistent for white individuals than for nonwhite individuals.

Finally, we also run the fixed-effects dynamic panel regression controlling for `edu`. Similarly to `white`, we may not include this time-invariant variable `edu` as a control x, and hence we run a regression for the following two education level categories:

```
. xtusreg logincome if edu < 12
```
(*output omitted*)
```
Number of observations:          5766
Number of cross-section units:   1922
Number of time periods:          3
List of time periods:            65, 66, 68

L1 = Autoregressive Coefficient (gamma)
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
L1	.4023438	.1042255	3.86	0.000	.1980655	.606622

```
Reference: Sasaki, Y. & Xin, Y. (2017) Unequal Spacing in Dynamic Panel Data:
Identification and Estimation. Journal of Econometrics 196 (2), pp. 320-330.
```

```
. xtusreg logincome if edu >= 12
  (output omitted)
Number of observations:          3228
Number of cross-section units:   1076
Number of time periods:          3
List of time periods:            65, 66, 68
-----------------------------------------------------------------------------
L1 = Autoregressive Coefficient (gamma)
-----------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+---------------------------------------------------------------
          L1 |   .7363281   .2291288     3.21   0.001     .2872439    1.185412
-----------------------------------------------------------------------------
Reference: Sasaki, Y. & Xin, Y. (2017) Unequal Spacing in Dynamic Panel Data:
Identification and Estimation. Journal of Econometrics 196 (2), pp. 320-330.
```

The autoregressive parameter γ is larger for those individuals with 12 years of education or higher, implying that log income is more persistent for this group of individuals.

Existing commands could not have produced these results, because of the irregular time spacing of the NLS Original Cohorts: Older Men. The new command, `xtusreg`, allows researchers to make new discoveries with these types of panel datasets, to which existing commands did not apply.

6 Conclusion

Conventional methods to fit fixed-effects dynamic panel regression models require an observation of three consecutive time periods or two pairs of two consecutive time periods. Sasaki and Xin (2017) show that the availability of "two pairs of two consecutive time gaps", which generalizes the conventional requirement of "two pairs of two consecutive time periods", suffices for the identification of the model parameters. In this article, we introduced the `xtusreg` command, which executes estimation and inference for fixed-effects dynamic panel regression under unequal time spacing based on the methods proposed in Sasaki and Xin (2017). The command was illustrated with an analysis of income dynamics using the NLS Original Cohorts: Older Men, for which personal interviews were conducted in 1966, 1967, 1969, and so on.

Finally, we discussed a limitation of the `xtusreg` command and a direction for future research. The current command accommodates only a certain pattern of unequal time spacing in panel data, namely, the U.S. spacing (Sasaki and Xin 2017, def. 2). Another important class of spacing pattern, called U.K. spacing (Sasaki and Xin 2017, ex. 1), tends to incur a weak identification as reported in Sasaki and Xin (2017, sec. 5). Hence, an implementation of the standard estimation and inference is not preferable under the U.K. spacing. An alternative is to conduct inference robust to weak identification. We leave development of a command to meet this goal for future research.

7 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 22-3
. net install st0690     (to install program files, if available)
. net get st0690         (to install ancillary files, if available)
```

`xtusreg` can also be downloaded from the Statistical Software Components Archive by typing

```
. ssc install xtusreg
```

8 References

Ashenfelter, O. 1978. Estimating the effect of training programs on earnings. *Review of Economics and Statistics* 60: 47–57. https://doi.org/10.2307/1924332.

Sasaki, Y., and Y. Xin. 2017. Unequal spacing in dynamic panel data: Identification and estimation. *Journal of Econometrics* 196: 320–330. https://doi.org/10.1016/j.jeconom.2016.10.002.

About the authors

Yuya Sasaki is an associate professor of economics at Vanderbilt University

Yi Xin is an assistant professor of economics at the California Institute of Technology.