Asymmetric Fixed Effects Models for Panel Data

Paul D. Allison
Statistical Horizons LLC
November 2018

Abstract

Standard fixed effects methods presume that effects of variables are symmetric: the effect of increasing a variable is the same as the effect of decreasing that variable but in the opposite direction. This is implausible for many social phenomena. York and Light (2017) showed how to estimate asymmetric models by estimating first-difference regressions in which the difference scores for the predictors are decomposed into positive and negative changes. In this paper, I propose some ways to improve their method for multi-period data. I also develop a data generating model that justifies the first-difference method but can be applied in more general settings. In particular, it can be used to construct asymmetric logistic regression models.

Asymmetric Fixed Effects Models for Panel Data

1. Introduction

In his 1985 book *Making It Count*, Stanley Lieberson devoted a whole chapter to arguing against the nearly universal presumption that causal effects are symmetric. What he meant is that, for both theorists and data analysts, there is usually an implicit assumption that if a one-unit increase in variable *X* produces a change of *B* units in variable *Y*, then a one-unit *decrease* in *X* will result in a change of *–B* units in *Y*. Essentially, this is an assumption that causal effects are exactly reversible.

Lieberson persuasively argued that there are strong reasons to suspect that many social and psychological phenomena do not work this way. Is the increase in happiness when a person gets married exactly matched by the decrease in happiness when that same person gets divorced? Is the effect of imprisonment on physical health cancelled out by the effect of release from prison? Does a 10% increase in income have the same effect on savings as a 10% decrease in income (in the opposite direction)? On the face of it, most people would probably answer no. But these are empirical questions and, unfortunately, good methodological tools for answering such questions have generally been lacking.

There have been several efforts to develop methods for testing asymmetrical hypotheses, but most of those efforts have focused on the distinction between necessary and sufficient conditions and have relied primarily on cross-sectional data. Rosenberg et al. (2006) provide a fairly comprehensive review of this literature. In contrast, I take it as axiomatic that longitudinal data are essential to capture Lieberson's notion of asymmetry. We need to see what happens when *X* goes up and compare that with what happens when *X* goes down. Chamlin and Cochran (1998) did address asymmetry with an autoregressive integrated moving average model applied

to bivariate time series data on oil prices and burglaries. However, their approach was simply to separate the time series into one long period in which the price of oil was increasing followed by another long period in which the price was decreasing. That method might be adequate for some applications, but it has little generality.

A more general solution was proposed by York and Light (2017) who showed how to estimate asymmetric models for panel data by

(a) calculating first differences for all time varying variables, both dependent and independent;

(b) decomposing each time-varying difference score into positive and negative components;

(c) estimating a first-difference regression model using the decomposed variables rather than the original ones.

I believe that York and Light have hit on a method that is essentially correct as a way to implement Lieberson's vision. However, their proposal is marred by a few problems:

1. They don't explain how their method is related to standard methods for analyzing panel data.

2. Their estimation method has some unnecessary peculiarities whose implications are not fully understood or explained;

3. Their method does not satisfactorily address the correlation structure that is typical of multi-period data that are first-differenced;

4. They do not present a data generating model with parameters that are estimable by their method.

The first goal of this paper is to fix these four problems. The second goal is to extend the method to logistic regression for categorical dependent variables. Computations for the examples were done using both SAS and Stata. Appendix 2 contains all the code used to produce the tables. Data sets are available at https://statisticalhorizons.com/resources/data-sets.

2. Symmetric Fixed Effects and First Differencing for Two-Period Data

Let's first consider traditional symmetric models for panel data. I begin with the case in which individuals are observed at only two points in time, $t=1$ and $t=2$. Things are much clearer and less complicated in this situation. For each individual i at each time t, we observe only two variables, Y_{it} and X_{it}. The extension to multiple X's is straightforward.

I'll begin with a linear model that is a common starting point for the analysis of panel data:

$$Y_{i1} = \mu_1 + \beta X_{i1} + \alpha_i + \varepsilon_{i1}$$
$$Y_{i2} = \mu_2 + \beta X_{i2} + \alpha_i + \varepsilon_{i2}$$

(1)

In this model, μ_1 and μ_2 are different intercepts that allow for a change over time that is unrelated to X. The effect of X on Y is β, which is assumed to be the same at both time points. The ε_{it}'s are random error terms that are specific to each time point and assumed to be independent of anything else on the right-hand side. Finally, α_i is an unobserved variable that represents the combined effects on Y of all variables that are specific to the individual but which do not change over time.

The α_i term can be treated as a set of random variables (a random effects or random intercepts model) or as a set of fixed constants (a fixed effects model). I'll treat them here as constants because then there is no need to make any assumptions about them. In particular, the

α's may be correlated with the X's, which implies that estimates of β control, in a fundamental way, for all time-invariant predictors.

There are several equivalent ways to estimate β, all producing identical estimates (Allison 2009). Here are the three most common methods:

1. <u>Least squares dummy variables</u> (LSDV). In this method, the data are arranged in "long form" with two records per person, one for each time point. Dummy (indicator) variables are created for all individuals (less one). Then Y is regressed on X plus the dummy variables for individuals and a dummy for time. LSDV produces "best linear unbiased" estimates, but it can be computationally intensive when the number of individuals is large.

2. <u>Mean deviation</u>. This is the classic fixed effects method. Again the data are in long form. But, for each individual, the mean of the two Y values is calculated, as well as the mean of the two X values. These individual-specific means are subtracted from the original values of X and Y for each record, to create the mean deviation variables Y^* and X^*. Then Y^* is regressed on X^*, along with a dummy variable for time. This produces exactly the same estimate of β as LSDV, but the computations are much faster. However, it does require an adjustment of the standard errors and degrees of freedom. Many statistical packages can automate all these steps (e.g., the **xtreg** command in Stata or the GLM procedure in SAS).

3. <u>First difference</u>. This is the simplest method in the two-period case. For eq. (1), subtract the time 1 equation from the time 2 equation, to produce

$$Y_{i2} - Y_{i1} = (\mu_2 - \mu_1) + \beta(X_{i2} - X_{i1}) + (\varepsilon_{i2} - \varepsilon_{i1}) \tag{2}$$

Notice that the α_i's have now been "differenced out" of the equation. To estimate β, simply create the two difference scores (using data in the "wide form") and then regress the Y difference on the X difference. No standard error adjustment is needed.

Example: 581 children were studied in 1990 and 1992 as part of the National

Longitudinal Survey of Youth (NLSY) (Center for Human Resource Research 2002). I'll use

three variables that were measured at each of the time points:

ANTI   antisocial behavior, measured one a scale from 0 to 6.

SELF   self-esteem, measured on a scale from 6 to 24.

POV   poverty status of family, coded 1 for family in poverty, otherwise 0.

The goal is to estimate the regression of ANTI on SELF and POV, corresponding to eq.

(1). Although all three methods described above produce exactly the same results, I'll focus on

the first difference method. Thus, using data in the wide form, I created new variables:

```
ANTIDIFF=ANTI2-ANTI1
SELFDIFF=SELF2-SELF1
POVDIFF=POV2-POV1
```

Let's take a look at the distributions of these difference scores. Figures 1 and 2 display

histograms for ANTIDIFF and SELFDIFF. Both graphs roughly resemble normal distributions

centered around 0. This is nice, but a little surprising given that the original variables are highly

skewed—ANTI to the right (most children are low on antisocial behavior) and SELF to the left

(most children are high on self-esteem).

Because POV is a dummy variable, POVDIFF has only three values, with the distribution

shown in Table 1. Over the two-year period, 45 children moved into poverty, 49 children moved

out of poverty, and the remainder (83%) stayed the same.

Table 1. Frequency Distribution for POVDIFF

```
  povdiff |      Freq.       Percent          Cum.
----------+----------------------------------------
       -1 |         49          8.43          8.43
        0 |        487         83.82         92.25
        1 |         45          7.75        100.00
----------+----------------------------------------
    Total |        581        100.00
```
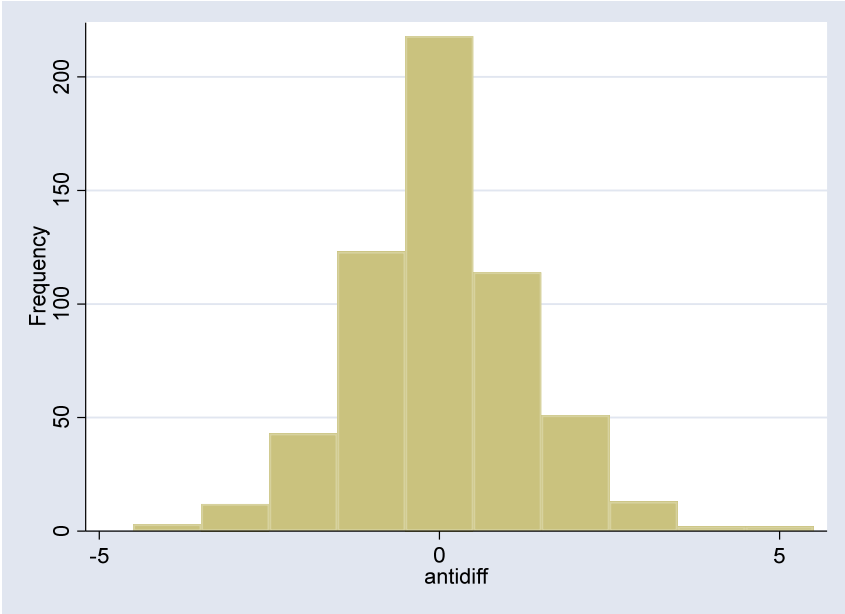

Figure 1. Histogram for ANTIDIFF
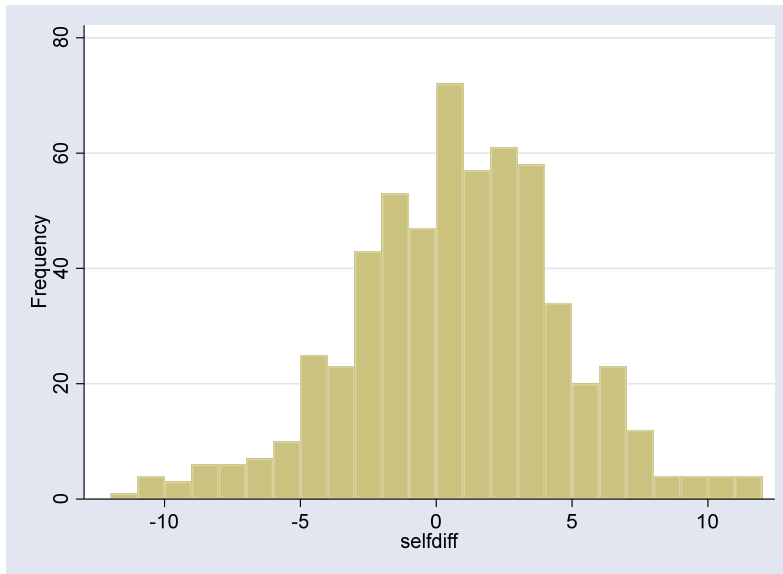
Figure 2. Histogram for SELFDIFF.

Table 2 shows the results of a least squares regression of ANTIDIFF on SELFDIFF and POVDIFF. Although not shown here, exactly the same results can be obtained with the LSDV method or the mean deviation method. The POVDIFF coefficient of .197 is not statistically significant. The SELFDIFF coefficient of -.039 is highly significant. For both coefficients, the interpretation is symmetric. Thus, a 1-unit increase in self-esteem is associated with a .039 decrease in antisocial behavior, while a 1-unit decrease in self-esteem is associated with a .039 increase in antisocial behavior. For poverty, moving into poverty increases antisocial behavior by .197 units, while moving out of poverty decreases antisocial behavior by .197 units.

Table 2. Symmetric Regression of Antisocial Behavior on Self-Esteem and Poverty.

```
-------------------------------------------------------------------------
 antidiff |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------+--------------------------------------------------------------
 selfdiff | -.0391292   .0136396   -2.87   0.004    -.0659185     -.01234
  povdiff |  .1969039   .1326352    1.48   0.138    -.0636018    .4574096
    _cons |  .0403031   .0533833    0.75   0.451    -.0645458    .1451521
-------------------------------------------------------------------------
```

7

3. An Asymmetric Model for Two-Period Data

Now let's modify the first-difference model to allow effects to be asymmetric. The key

insight of York and Light (2017) was that difference scores for predictor variables can be

decomposed into positive and negative components. Specifically, I define

$$X_i^+ = X_{i2} - X_{i1} \text{ if } (X_{i2} - X_{i1}) > 0, \text{ otherwise } 0,$$
$$X_i^- = -(X_{i2} - X_{i1}) \text{ if } (X_{i2} - X_{i1}) < 0, \text{ otherwise } 0.$$

Both of these variables are never negative, but the first represents an increase and the second

represents a decrease. It can easily be shown that

$$X_i^+ - X_i^- = X_{i2} - X_{i1}, \tag{3}$$

a result that will be needed later. The working model then becomes

$$Y_{i2} - Y_{i1} = (\mu_2 - \mu_1) + \beta^+ X_i^+ + \beta^- X_i^- + (\varepsilon_{i2} - \varepsilon_{i1}) \tag{4}$$

Thus, β^+ represents the change in Y for a one-unit increase in X and β^- represents the change in Y

for a one-unit decrease in X. Typically, β^+ and β^- will have opposite signs, although that's not

guaranteed. When $\beta^+ = -\beta^-$, then we are back to the original symmetric model.

Note that when X is a dummy variable, like POV, then the decomposed model has two

dummy variables:

$$X^+ = 1 \text{ if } X \text{ changed from 0 to 1, otherwise 0,}$$
$$X^- = 1 \text{ if } X \text{ changed from 1 to 0, otherwise 0.}$$

The implicit reference category is no change from time 1 to time 2.

Now let's estimate this model for the NLSY data, decomposing both the change in self-

esteem and the change in poverty. Creating the necessary variables is reasonably straightforward

in most software. Here, for example, is the Stata code to read in the data and generate the needed

variables:

```
use c:\data\nlsy.dta, clear
generate antidiff=anti92-anti90
generate selfdiff=self92-self90
generate povdiff=pov92-pov90
generate selfpos=selfdiff*(selfdiff>0)
generate selfneg=-selfdiff*(selfdiff<0)
generate povpos=povdiff*(povdiff>0)
generate povneg=-povdiff*(povdiff<0)
```

Table 3 displays the results for regressing the difference score for antisocial behavior on the

decomposed difference scores for self-esteem and poverty. The coefficient for SELFPOS is

small and far from statistically significant, indicating that increases in self-esteem have no

detectable effect on antisocial behavior. On the other hand, a decrease in self-esteem produces

an increase in antisocial behavior that is much larger in magnitude (.074) and highly significant.

Neither of the poverty coefficients is statistically significant, which is consistent with the overall

lack of effect in the symmetric model. Nevertheless, the coefficients are in the expected

direction: moving into poverty increases antisocial behavior while moving out of poverty reduces

antisocial behavior.

Table 3. Asymmetric Regression of Antisocial Behavior on Self-Esteem and Poverty.

antidiff	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
selfpos	-.0048386	.0251504	-0.19	0.848	-.0542362	.0445589
selfneg	.0743077	.025658	2.90	0.004	.023913	.1247024
povpos	.2502064	.2003789	1.25	0.212	-.143356	.6437688
povneg	-.126328	.1923669	-0.66	0.512	-.5041541	.2514981
_cons	-.0749517	.086383	-0.87	0.386	-.2446157	.0947123

The evidence in Table 3 suggests that the effect of a decrease in self-esteem is much larger than

the effect of an increase in that variable. But is the difference statistically significant? To

answer that question, one must test whether $\beta^+ = -\beta^-$. Most statistical packages have simple

commands for doing this. For example, in Stata, one can simply enter

```
test selfpos=-selfneg
```

That produced an *F*-statistic with a *p*-value of .106, not quite statistically significant. So there's

no compelling evidence that the effects of self-esteem are asymmetric.

4. Symmetric Estimation for Multi-Period Data

As we shall see in the next section, the decomposition of first-difference predictors into

positive and negative components can be easily extended to panel data with more than two

periods. Before going there, however, it's essential to address some critical issues regarding the

estimation of *symmetric* models with the first-difference approach. This section is a rather

extended digression from the topic of asymmetry but, unfortunately, a necessary one.

To explain these issues, a new example is needed. I'll use a subsample of the data

described in Firebaugh et al. (2013). For 1,134 respondents, five waves of survey data were

collected annually from 1983 to 1987 when the respondents were ages 18–22. Here are the

variables measured in each of the five years:

LWAGE	logged hourly wage for current or most recent job
MAR	1 if married, otherwise 0
EDU	years of education
URB	1 if living in an urban area, otherwise 0

Logged hourly wage rate will be the dependent variable for the example. For simplicity,

observations with any missing data on these four variables have been excluded.

For the two-period case, we saw in Section 2 that there are three methods for estimating a

symmetric fixed effects model that produce identical results: LSDV, mean deviation, and first

differencing. These three methods can also be used when there are more than two time points, but now there is a potential divergence. LSDV and mean deviation still produce identical results. With first differencing, on the other hand, there are many potential estimation methods. Those that are most commonly used will not produce the same estimates as LSDV and mean deviation. However, I will describe and demonstrate a first-difference estimation method that does produce estimates that are identical to those produced by mean deviation and LSDV.

The basic model for multi-period data is a straightforward extension of the model for two-period data.

$$Y_{it} = \mu_t + \beta X_{it} + \alpha_i + \varepsilon_{it} \qquad\qquad i = 1,\ldots,n;\ t = 1,\ldots,T \qquad\qquad (5)$$

The intercept, μ_t, is allowed to differ at each point in time. X_{it} can be regarded either as a single variable or as a vector of variables. As before, α_i represents the combined effect of all variables that do not vary with time, regardless of whether they are observed or unobserved. To make this a fixed effects model, α_i is treated as a set of constants. Finally, the ε_{it} are random errors that are assumed to be independent of X and α, independent of each other, and have a common variance.

Under these assumptions, the mean deviation method (or the equivalent LSDV method) yields estimates that are best linear unbiased. As a standard for comparison, let's apply the mean deviation method to the wage example. The first step is to put the data into long form, yielding 5,670 records, 5 per person. As in the two-period case, the mean is calculated for each variable for each person. These person-specific means are subtracted from the original values of each variable to create deviation scores. The final step is to do ordinary least squares regression with the deviation scores. These steps are automated in many software packages, including the **xtreg** command in Stata and PROC GLM in SAS. Results are shown in Table 4, with robust standard errors to compensate for any specification errors.

11

Table 4. Symmetric Fixed Effects Model with Five-Period Data Using Mean Deviation

```
                                (Std. Err. adjusted for 1,134 clusters in id)
       -----------------------------------------------------------------------
                    |               Robust
         lwage |      Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval]
       -------------+---------------------------------------------------------
           mar |   .0417479   .0228613     1.83   0.068    -.0031074    .0866031
           edu |   .0566356   .0156926     3.61   0.000     .0258458    .0874253
           urb |   .0812962    .029639     2.74   0.006     .0231426    .1394497
          year |
             2 |   .0310038   .0137951     2.25   0.025      .003937    .0580707
             3 |   .1034818   .0143946     7.19   0.000     .0752387     .131725
             4 |   .1657912   .0162096    10.23   0.000     .1339869    .1975955
             5 |   .2310036    .018366    12.58   0.000     .1949684    .2670388
         _cons |   .8036756   .1976148     4.07   0.000     .4159435    1.191408
       -------------+---------------------------------------------------------
```

Because the dependent variable is logged and the coefficients are all less than .10 in magnitude, they can be interpreted in terms of percentage changes. For example, the .04 coefficient for marital status says that the effect of getting married is a 4% increase in wages. And since the model is symmetric, the effect of getting divorced is a 4% *decrease* in wages. Each additional year of schooling produces a 5.7% increase in wages. The effect of moving from a non-urban area to an urban area is an 8% increase in wages, while moving from a non-urban to an urban area yields an 8% decrease.

Now let's estimate the same symmetric model using first differences. Equation (5) implies the working model,

$$Y_{it} - Y_{it-1} = (\mu_t - \mu_{t-1}) + \beta(X_{i1} - X_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1}) \qquad t = 2,\ldots,T \qquad (6)$$

Estimation of this equation is a fairly natural extension of the difference score method for two-period data, except that there are now multiple records for each person. With five-period data, there are four records for each person: The first record contains difference scores for periods 1 and 2. The second record contains difference scores for periods 2 and 3, and so on. Pooling all

these records into a single data set, we then estimate the regression of the *Y* difference score on the *X* difference scores.

Unfortunately, there are differing opinions on exactly how to estimate the difference score regression. Wooldridge (2010) proposed using ordinary least squares, with robust standard errors to correct for heteroscedasticity and for correlations among the errors. Under the assumptions stated for eq. (5), he showed that OLS will produce unbiased estimators of the regression coefficients. However, OLS estimators are less efficient than the ones produced by LSDV or mean deviation and, therefore, have larger true standard errors.

York and Light (2017), on the other hand, applied fixed effects regression to the difference scores (using the mean deviation method), along with robust standard errors to correct for heteroscedasticity and correlated errors. Their use of a fixed effects estimator is peculiar because the effects of any time-invariant variables are already removed by first differencing. A fixed effects model for the difference scores is equivalent to a model which says that the effect of time is linear with a slope that is unique to each individual. Although there is nothing intrinsically wrong with such a model, it goes well beyond what most people want to achieve when they do fixed effects. Moreover, the robustness that may be gained by this approach comes at a cost of larger standard errors. Finally, any attractions of this method are completely unrelated to the estimation of asymmetric effects.

As a possible alternative, York and Light also suggested the use of a random effects estimator. But a random effects model would only allow for positive correlations among the multiple records for each individual. In actuality, the most salient feature of first-difference data is a *negative* correlation between the error terms of adjacent difference scores.

Why is there a negative correlation? Consider the error terms in the difference regressions, $(\varepsilon_3 - \varepsilon_2)$ and $(\varepsilon_2 - \varepsilon_1)$. Clearly, whenever ε_2 is high, the first error term will tend to be low while the second will tend to be high. More specifically, it can be shown that, under the assumption that the ε_{it} are uncorrelated with each other and are homoscedastic, the correlation between any two adjacent error terms is -.50. Although robust standard errors adjust the standard errors for such correlations, they do nothing to address statistical inefficiency of the coefficient estimates themselves.

To accommodate this, I propose the use of generalized least squares (GLS) to estimate equation (6). Unlike the other difference score methods, GLS can produce estimates that are unbiased and efficient. As we will see, the choice among these methods can make a big difference.

I now apply these three first-difference methods to estimate a symmetric model for the wage example. Table 5 displays results from an OLS regression with robust standard errors (Wooldridge's method). Ideally, we would expect these results to be similar to those obtained with the classic fixed effects results in Table 4.

Table 5. Symmetric First-Difference Estimated with OLS and Robust Standard Errors.

```
                              (Std. Err. adjusted for 1,134 clusters in id)
-----------------------------------------------------------------------------
             |               Robust
wagediff |      Coef.    Std. Err.      t     P>|t|     [95% Conf. Interval]
---------+-------------------------------------------------------------------
 mardiff |    .0191042    .0256535     0.74    0.457    -.0312295      .069438
 edudiff |   -.0205905    .0248746    -0.83    0.408    -.0693959      .028215
 urbdiff |    .0764678    .0322208     2.37    0.018     .0132487     .1396869
    year |
       3 |    .0374366    .0228207     1.64    0.101     -.007339     .0822122
       4 |    .0247034    .0188542     1.31    0.190    -.0122897     .0616964
       5 |    .0259658    .0188678     1.38    0.169     -.011054     .0629857
   _cons |     .045713    .0136904     3.34    0.001     .0188517     .0725744
-----------------------------------------------------------------------------
```

Although the coefficient for urban residence is pretty close to the one in Table 4, marital status and education have coefficients that are much smaller than those in Table 4 and *p*-values that are much larger. (There's no reason to expect the year coefficients to be similar because the Table 4 coefficients estimate differences between years while the Table 5 coefficients estimate differences in differences between years).

Table 6 shows the results from applying the fixed effects method used by York and Light. Now the results are even more dramatically different from those produced by classic fixed effects in Table 4. Although the urban coefficient is still about the same (but with a substantially larger standard error), the education coefficient is now -.126, which says that each additional year of schooling yields a 12.6% *decrease* in wages. That's a hard estimate to swallow.

Table 6. Symmetric First-Difference with Fixed Effects and Robust Standard Errors.

```
                        (Std. Err. adjusted for 1,134 clusters in id)
-----------------------------------------------------------------------------
            |               Robust
wagediff |     Coef.    Std. Err.       t     P>|t|      [95% Conf. Interval]
---------+-------------------------------------------------------------------
 mardiff |   .0131029   .0287179     0.46   0.648    -.0432433      .0694491
 edudiff |    -.12558    .043132    -2.91   0.004    -.2102075     -.0409524
 urbdiff |   .0748975   .0354895     2.11   0.035     .0052649       .14453
    year |
       3 |   .0321924   .0226377     1.42   0.155    -.0122241      .0766089
       4 |   .0156451    .019018     0.82   0.411    -.0216694      .0529595
       5 |   .0141105   .0189854     0.74   0.457      -.02314      .0513609
   _cons |   .0646659   .0136071     4.75   0.000     .0379679      .0913638
---------+-------------------------------------------------------------------
```

I believe that GLS is a better way to go than either of these two methods. With GLS, it's necessary to specify or estimate the variances of the error terms and the correlations among them. With four records per person, there are six correlations and four variances. We can estimate an unstructured model that estimates all of these additional parameters—and there's

15

something to be said for that. Or we can impose some appropriate structure on the variances and

correlations.

Let's first try an unstructured model, estimated by maximum likelihood with robust

standard errors (using Stata's **mixed** command). Results in Table 7 are reasonably close to those

for classic fixed effects in Table 4, although the coefficients are a little smaller and the *p*-values

are a little larger.

In the lower panel of Table 7 we see estimates for the standard deviations of the four

error terms (which are all quite similar), along with estimates of the six correlations. The three

correlations between error terms that are adjacent in time are negative and quite large: -.50, -.33,

and -.38. On the other hand, the correlations between error terms that are not adjacent in time

(and therefore do not share a common component) are quite small, and their 95% confidence

intervals all include 0.

Table 7. Symmetric First-Difference Model with Unstructured GLS.

```
                              (Std. Err. adjusted for 1,134 clusters in id)
    ---------------------------------------------------------------------------
             |               Robust
    wagediff |     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
    ---------+-----------------------------------------------------------------
     mardiff |   .0349179    .022229    1.57   0.116    -.0086502    .0784859
     edudiff |   .0368013   .0158223    2.33   0.020     .0057902    .0678124
     urbdiff |   .0753664    .028231    2.67   0.008     .0200347     .130698
             |
        year |
           3 |   .0404232   .0230346    1.75   0.079    -.0047238    .0855702
           4 |     .02955   .0189069    1.56   0.118    -.0075069    .0666069
           5 |   .0324069   .0189648    1.71   0.087    -.0047635    .0695773
             |
       _cons |   .0348007   .0137085    2.54   0.011     .0079325    .0616688
```

```
----------------------------------------------------------------
                          |              Robust
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
--------------------------+-------------------------------------------------
Residual: Unstructured    |
               sd(e2) |   .4565056    .0225466     .4143867    .5029056
               sd(e3) |   .4414226    .0230753     .3984356    .4890475
               sd(e4) |   .4235591    .0246425     .3779126     .474719
               sd(e5) |   .4560596    .0290936      .402458    .5168001
          corr(e2,e3) |  -.4959771    .0460547    -.5808013   -.4004982
          corr(e2,e4) |  -.0427685    .0352129    -.1114718    .0263417
          corr(e2,e5) |    .021735    .0361802    -.0491673    .0924193
          corr(e3,e4) |  -.3258869    .0521469     -.424009   -.2202028
          corr(e3,e5) |  -.0596874    .0385681    -.1347953    .0161024
          corr(e4,e5) |  -.3764606    .0567421    -.4819381    -.260228
----------------------------------------------------------------
```

The good thing about the unstructured model is that it allows for departures from the standard assumptions. In particular, it allows error variances that change over time, as well as for serial correlations among the ε_{it} in the original model (5).   On the other hand, if we want a method that's equivalent to the mean deviation method, we should impose the following constraints: (a) all error standard deviations are the same, (b) all correlations among adjacent error terms are -.50, and (c) all correlations among non-adjacent error terms are 0. These constraints can be imposed with PROC MIXED in SAS (as shown in Appendix 2). Results are identical to those produced by the mean deviation method in Table 4.

Stata can't fully achieve these constraints. Its Toeplitz 1 model constrains the error standard deviations to be the same and the correlations between non-adjacent error terms to be 0. It also constrains all the correlations between adjacent error terms to be the same, but it can't constrain them to be exactly -.50. Table 8 displays results from fitting this model to the wage data using the **mixed** command in Stata. These estimates are quite similar to those in Table 7 for the unstructured model. The estimated correlation between adjacent error terms is -.42, not quite the -.50 that was specified in PROC MIXED.

Table 8. Symmetric First-Difference Model Estimated with Toeplitz 1 GLS

```
                                (Std. Err. adjusted for 1,134 clusters in id)
-------------------------------------------------------------------------------
            |               Robust
 wagediff   |     Coef.   Std. Err.      z     P>|z|      [95% Conf. Interval]
------------+------------------------------------------------------------------
   mardiff  |    .036767   .0224893    1.63    0.102    -.0073113     .0808452
   edudiff  |   .0377946   .0165048    2.29    0.022     .0054458     .0701434
   urbdiff  |   .0816084   .0289154    2.82    0.005     .0249353     .1382814
            |
      year  |
         3  |   .0404957   .0230238    1.76    0.079    -.0046302     .0856216
         4  |   .0297125   .0189145    1.57    0.116    -.0073593     .0667843
         5  |   .0320945   .0189502    1.69    0.090    -.0050472     .0692362
            |
     _cons  |   .0345768   .0136988    2.52    0.012     .0077277      .061426
-------------------------------------------------------------------------------
            |               Robust
 Random-effects Parameters   |   Estimate   Std. Err.     [95% Conf. Interval]
----------------------------+--------------------------------------------------
id:                 (empty) |
----------------------------+--------------------------------------------------
Residual: Toeplitz(1)       |
                    rho1    |  -.4185508    .0151597    -.4478095    -.3883967
                    sd(e)   |   .4463315    .0157754     .4164589     .4783468
```

Keep in mind that all three of these GLS models should produce estimates that are consistent (and, hence, approximately unbiased). The attraction of the more parsimonious models is that they should have smaller true standard errors.

The decision to use an unstructured or structured model will depend in part on T, the number of time points. In an unstructured model, the number of error variances and correlations that must be estimated is $T(T+1)/2$. If there are 20 time points, that's 210 extra parameters! On the other hand, the Toeplitz 1 structure only has one error variance and one correlation, and the fully constrained model estimates only the single error variance. So when the number of time points is small, the unstructured model will be fine, especially when the number of individuals is

18

large. On the other hand, when there are many time points, the Toeplitz 1 model or the fully constrained model will be strongly preferred.

5. Asymmetric Estimation for Multi-Period Data

We are finally ready to estimate an asymmetric first-difference model for multi-period data using GLS. Once the data are in the long form, the decomposed difference scores for each data record can be defined in exactly the same way it was done for two-period data. That is, for $t=2,\ldots,T$.

$$X_{it}^+ = X_{it} - X_{it-1} \text{ if } (X_{it} - X_{it-1}) > 0, \text{ otherwise } 0,$$
$$X_{it}^- = -(X_{it} - X_{it-1}) \text{ if } (X_{it} - X_{it-1}) < 0, \text{ otherwise } 0.$$

The working model then becomes

$$Y_{it} - Y_{it-1} = (\mu_t - \mu_{t-1}) + \beta^+ X_{it}^+ + \beta^- X_{it}^- + (\varepsilon_{it} - \varepsilon_{it-1}) \tag{7}$$

which can be estimated by GLS. One important issue is that it's not obvious what sort of model for Y_{it} would lead to this model for the difference scores. I will address that issue in the next section.

For the wage data, there's no reason to decompose the difference scores for education because that variable only increases—you can't experience a decrease in years of schooling. We'll allow for asymmetry in the effect of marital status because there are definitely changes in both directions, as shown in Table 9.

Table 9. Frequency Distribution for MARDIFF

```
 mardiff |     Freq.     Percent        Cum.
---------+-----------------------------------
     -1 |       117        2.58        2.58
      0 |     4,095       90.28       92.86
      1 |       324        7.14      100.00
---------+-----------------------------------
   Total |     4,536      100.00
```

19

Thus, there were 324 entries into marriage and 117 exits from marriage. Similarly, Table 10 shows the frequency distribution for URBDIFF.

Table 10. Frequency Distribution for URBDIFF

```
  urbdiff |      Freq.     Percent        Cum.
----------+---------------------------------
       -1 |        158        3.48        3.48
        0 |      4,188       92.33       95.81
        1 |        190        4.19      100.00
----------+---------------------------------
    Total |      4,536      100.00
```

There were 190 moves from non-urban areas to urban areas and 158 moves from urban areas to non-urban areas.

Table 11 displays results for a fully constrained GLS model (estimated with PROC MIXED). The coefficients are all in the expected direction. Getting married increases wages by 2.5%. Leaving a marriage decreases wages by 6.7%. But neither of these coefficients is significantly different from 0, and they are not significantly different from each other. Moving to an urban area increases wages by 9%. Leaving an urban area reduces wages by 7.4%. Both of these effects are significantly different from 0, but they are not significantly different from each other (test not shown). Thus, there is no evidence for asymmetry in this example.

Table 11. Asymmetric First-Difference Model Estimated with Constrained GLS

Effect	Estimate	Error	t Value	Pr > |t|	Lower	Upper
Intercept	0.06686	0.01403	4.77	<.0001	0.03933	0.09438
marpos	0.03206	0.02290	1.40	0.1616	-0.01284	0.07697
marneg	-0.06732	0.04782	-1.41	0.1593	-0.1611	0.02644
edudiff	0.05606	0.01571	3.57	0.0004	0.02526	0.08685
urbpos	0.08155	0.03447	2.37	0.0180	0.01397	0.1491
urbneg	-0.08002	0.03616	-2.21	0.0270	-0.1509	-0.00913
t2	-0.03456	0.01905	-1.81	0.0697	-0.07192	0.002791
t3	0.006903	0.01984	0.35	0.7280	-0.03200	0.04581
t4	-0.00332	0.02182	-0.15	0.8792	-0.04611	0.03947
t5	0	.	.	.	.	.

6. An Asymmetric Linear Model for Multi-Period Data

We now have a pretty good method for estimating asymmetric fixed effects models using first differences. This method is reasonably easy to implement, intuitively appealing, and seems to produce plausible results. What we don't have (yet) is a model for the data generating process that would imply the first-difference equations whose parameters we have been estimating. Specifically, we need a set of equations with individual Y_{it}'s on the left-hand side instead of difference scores.

In this section I propose a simple model that leads to the difference equations used in previous sections. This model has several benefits:

1. It enriches our understanding of what sort of process would lead to asymmetric effects.

2. It provides a platform for generating simulated data that can be used evaluate the performance of alternative estimators.

3. It makes it possible to estimate an asymmetric model using the LSDV or mean deviation methods rather than the first-difference method.

4. Most importantly, it makes it possible to generalize the model to handle categorical dependent variables.

Here is the model. We observe Y_{it} and X_{it} for $t=1,\ldots,T$. Although X_{it} is a single variable, the extension to multiple variables should be straightforward. As in earlier sections, for $t=2,\ldots,$ T, define

$$X_{it}^+ = X_{it} - X_{it-1} \text{ if } (X_{it} - X_{it-1}) > 0, \text{ otherwise } 0,$$
$$X_{it}^- = -(X_{it} - X_{it-1}) \text{ if } (X_{it} - X_{it-1}) < 0, \text{ otherwise } 0.$$

For $t=1$, in which case X_{it-1} is not observed, both X_{it}^+ and X_{it}^- are set to 0. Now define

$$Z_{it}^+ = \sum_{s=1}^{t} X_{is}^+$$
$$Z_{it}^- = \sum_{s=1}^{t} X_{it}^-$$

(8)

Thus, Z^+ is the accumulation up to time t of all previous positive changes in X, and Z^- is the accumulation of all previous negative changes in X. When X is a dummy variable, Z^+ is just the number of previous changes from 0 to 1, and Z^- is the number of previous changes from 1 to 0. For example, Z^+ might be the number of previous marriages and Z^- might be the number of previous divorces.

The model is simply

$$Y_{it} = \mu_t + \beta^+ Z_{it}^+ + \beta^- Z_{it}^- + \alpha_i + \varepsilon_{it}, \qquad t=1,\ldots,T \tag{9}$$

where α and ε satisfy the usual assumptions for a fixed effects regression model.

I now show that this model implies the difference score model that we used in previous sections. If we construct difference scores based on this model, we have

$$Y_{it} - Y_{it-1} = (\mu_t - \mu_{t-1}) + \beta^+ (Z_{it}^+ - Z_{it-1}^+) + \beta^- (Z_{it}^- - Z_{it-1}^-) + (\varepsilon_{it} - \varepsilon_{it-1}). \tag{10}$$

But, but by definition,

$$Z_{it}^+ - Z_{it-1}^+ = \sum_{s=1}^{t} X_{is}^+ - \sum_{s=1}^{t-1} X_{is}^+$$
$$= X_{it}^+$$

(11)

Similarly

$$Z_{it}^- - Z_{it-1}^- = X_{it}^-.$$

So we end up with

$$Y_{it} - Y_{it-1} = (\mu_t - \mu_{t-1}) + \beta^+ X_{it}^+ + \beta^- X_{it}^- + (\varepsilon_{it} - \varepsilon_{it-1})$$

which is just the working first difference model in eq. (7). I can't claim that this is the *only* data generation model that would imply the difference score model. But I have been unable to find any other.

One surprising implication of this model is that Y_{it} depends on the entire previous history of changes in *X*. That raises two issues:

1. We don't know the history of *X* prior to time 1. That's not a problem, however, because that history does not vary over the observed time periods. Therefore, it gets absorbed into α_i, which is removed by first differencing or otherwise adjusted for by standard fixed-effects methods.

2. What happens when the effects of *X* are symmetric, i.e., $\beta^+ = - \beta^-$? In that case, it can be shown (see Appendix 1) that

$$\beta^+ Z_{it}^+ + \beta^- Z_{it}^- = \beta^+ X_{it}$$

and we are back to a model in which *Y* depends only on the current value of *X*.

We are now in a position to estimate an asymmetric model using conventional fixed effects rather than first differences. The only tricky part is the programming needed to compute the *Z* variables, that is, the accumulated positive and negative changes for each predictor

variable. See Appendix 2 for the details. After constructing appropriate Z variables, I applied the mean deviation method to the five-period example of Section 4. As anticipated, results were *identical* to those in Table 11 that that were obtained by applying the fully constrained GLS method to the first-difference data.

Is there any advantage to doing it this way rather than using the first-difference method? Well, the mean deviation method is more familiar than first differencing to most researchers who use fixed effects. However, in my experience, the programming for asymmetric models is actually a little easier with the first-difference method.

One benefit is that the mean deviation method may be better for certain patterns of missing data, specifically, if data are always observed for predictor variables but missing for some values of the dependent variable. For example, suppose that in the three-period case, some people have missing data on Y at time 2. With the first difference method, those people would be lost entirely because the middle value of Y is needed to calculate both of the difference scores. With the mean deviation method, however, they could still contribute observations for times 1 and 3.

Another big payoff from this alternative approach is that it is easily extended to categorical data. I'll show how to do that in the next section. One final attraction is that this approach can be incorporated into the SEM panel methods proposed by Allison et al. (2017), which make it possible to relax many of the hidden constraints of conventional fixed effects models. For example, instead of assuming that the unobserved individual component α_i has the same effect at all points in time, one can allow it to have different effects at different time points.

7. An Asymmetric Logistic Model for Multi-Period Dichotomous Data

Suppose the dependent variable Y_{it} is dichotomous with values of 1 and 0. For many well-known reasons (Allison 2012), a linear model for Y is not very attractive. Instead, let's consider an asymmetric logistic regression model. Letting p_{it} be the probability that $Y_{it} = 1$, the model is

$$\log\left(\frac{p_{it}}{1-p_{it}}\right) = \mu_t + \beta^+ Z_{it}^+ + \beta^- Z_{it}^- + \alpha_i, \qquad t=1,\ldots,T,$$

where the Z's are the accumulated positive and negative difference scores as defined in Section 6. As in previous sections, the α_i's are treated as constants that represent the combined effects on Y of all time-invariant variables.

Clearly, computing difference scores for the dependent variable would not work for this model. The standard method for estimating fixed effects logistic regression models is conditional maximum likelihood, which removes the α_i's by conditioning each individual's likelihood on the total number of 1's and 0's observed for that individual. We'll apply this method to the following example.

The sample consists of 1,151 teenaged girls who were interviewed annually for five years, beginning in 1979. The dependent variable is POV which, for each year, equals 1 if the girl's household was in poverty otherwise 0. The following predictor variables are also measured in each of the five years:

MOTHER   1 if respondent currently had a least one child, otherwise 0

SPOUSE   1 if respondent currently lives with a husband, otherwise 0

INSCHOOL 1 if respondent was currently enrolled in school, otherwise 0

HOURS   Hours worked during the week of the survey

I'll begin with a symmetric model that includes year as a categorical variable. Table 12 displays the estimates, which were obtained with the **clogit** command in Stata. All the predictors have coefficients that are significantly different from 0. When girls are currently enrolled in school or if they have children, they are more likely to be in poverty. When they are living with a husband or work more hours, they are less likely to be in poverty. Keep in mind that these are within-person estimates, so they represent the effects of changing from one condition to another.

Table 12. Symmetric Logistic Regression Via Conditional Likelihood.

```
                                 (Std. Err. adjusted for clustering on id)
-----------------------------------------------------------------------------
             |               Robust
       pov |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+---------------------------------------------------------------
     mother |   .5824322   .1668935     3.49   0.000     .255327     .9095374
     spouse |  -.7477585   .1824145    -4.10   0.000   -1.105284    -.3902326
   inschool |   .2718653   .1203546     2.26   0.024    .0359746     .5077559
      hours |  -.0196461   .0033695    -5.83   0.000   -.0262501    -.0130421
             |
       year |
          2 |   .3317803   .0965172     3.44   0.001      .14261     .5209506
          3 |   .3349777    .107582     3.11   0.002    .1241209     .5458345
          4 |   .4327654   .1177925     3.67   0.000    .2018964     .6636344
          5 |   .4025012   .1365389     2.95   0.003    .1348898     .6701126
-----------------------------------------------------------------------------
```

Now we'll estimate an asymmetric model by doing conditional likelihood with the accumulated first-differences of the predictors. Note that we cannot decompose the effect of MOTHER because it only increases from 0 to 1, never from 1 to 0. Results in Table 13 show a slightly larger effect of MOTHER than in Table 12. For SPOUSE, we see a strong negative effect of getting married on the log-odds of being in poverty, but a rather small positive effect of becoming unmarried. Only 48 people got unmarried (the overwhelming majority presumably by divorce), so that might account for the high *p*-value for this coefficient. Nevertheless, the difference in the magnitudes of the positive and negative coefficients is statistically significant at

conventional levels (*p*=.04). This is our first example in which there is noteworthy evidence of asymmetry.

Table 13. Asymmetric Logistic Regression Via Conditional Likelihood.

```
                                     (Std. Err. adjusted for clustering on id)
-----------------------------------------------------------------------------
               |               Robust
         pov   |      Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval]
---------------+-------------------------------------------------------------
        mother |   .6314079   .1681145     3.76    0.000     .3019095    .9609062
  spousecumpos |  -.8300194   .1850695    -4.48    0.000    -1.192749   -.4672898
  spousecumneg |    .070639   .3748448     0.19    0.851    -.6640433    .8053214
inschoolcumpos |  -.0816474   .2402821    -0.34    0.734    -.5525917    .3892969
inschoolcumneg |  -.3085952   .1274934    -2.42    0.016    -.5584776   -.0587128
   hourscumpos |  -.0227499   .0039225    -5.80    0.000    -.0304378   -.0150621
   hourscumneg |   .0150824   .0044844     3.36    0.001     .0062931    .0238718
          year |
             2 |   .3694268   .0982753     3.76    0.000     .1768107    .5620429
             3 |   .4244777   .1138939     3.73    0.000     .2012498    .6477057
             4 |    .595898   .1362824     4.37    0.000     .3287895    .8630065
             5 |   .6333666   .1645651     3.85    0.000      .310825    .9559083
-----------------------------------------------------------------------------
```

The INSCHOOL coefficients are a little surprising. Entering school reduces the odds of poverty by about 27% (calculated as $100(e^{\beta}-1)$), but leaving school also reduces the odds, by about 8%. The latter figure is far from statistically significant, so the negative coefficient could just be random error. But it does serve to point out that there is no necessity that changes in opposite directions for the predictors also have opposite sign.

For hours worked, the coefficients do have the expected opposite signs and are both highly significant. An increase of one hour reduces the odds of poverty by 2.3% while a decrease of one hour increases the odds of poverty by about 1.5%. However, the difference between these two magnitudes is not quite statistically significant (*p*=.12).

There is one other unexpected finding. Under the asymmetric model, the YEAR effects show a clear ordering (with the odds of poverty steadily increasing) that was not nearly so apparent in the symmetric model.

Although I only examine the dichotomous case here, the method could easily be applied to estimate asymmetric fixed effects for ordered logistic regression, multinomial logistic regression, or negative binomial regression for count data. It's all a matter of appropriately coding the predictor variables.

8. Discussion

In this paper, I have proposed methods to improve and extend the method of York and Light (2017) for estimating asymmetric fixed effects models for panel data. Fixed effects models are the natural way to go for asymmetric causal effects because they focus on within-individual change rather than between-individual differences. That allows us to separate out the differential effects of increases and decreases of the predictor variables.

For linear models, first-differencing is the most intuitively appealing and computationally convenient method for estimating asymmetric fixed effects models. However, for best results, generalized least squares is necessary to handle the negative correlations that naturally arise between adjacent first-difference observations. GLS can produce efficient estimates under a variety of correlation structures that are appropriate both when there are and are not correlations among the original *undifferenced* error terms.

In Section 6, I proposed a data generating model that implies the working equations of the first-difference approach. In this model, Y_{it} is a function of the accumulated positive changes in X and the accumulated negative changes in X over the observed history of the individual. This implies that any given change has an effect that persists over time, unless cancelled out (fully or

28

partially) by a change in the opposite direction. To better understand this, suppose that X is a dummy variable for employed vs. unemployed. The model says that each entry into employment will raise the expected value of Y (say, income) by a fixed amount. And, holding other factors constant, that expected value will persist until employment is terminated. At that point, the expected value of Y will be reduced by a fixed but different amount.

Is this a believable model? Well, keep in mind that it's less restrictive than the standard fixed effects model which assumes that the effect of becoming unemployed is just the reverse of the effect of becoming employed. However, there are many applications where it might be more plausible to suppose that the effect of a given change may be temporary rather than persistent. Allison (1994) discusses some ways to model this kind of process, but both the formulation and estimation of such models can become substantially more complicated.

The new data generating model proposed here makes it possible to estimate asymmetric effects without using first differences of the dependent variable. That's only mildly attractive for linear models, but it's a great boon for estimating asymmetric models for categorical data, including logistic regression models of all kinds, as well as models for count data.

References

Allison, Paul D. "Using panel data to estimate the effects of events." *Sociological Methods & Research* 23.2 (1994): 174-199.

Allison, Paul D. *Fixed effects regression models*. Vol. 160. SAGE publications, 2009.

Allison, Paul D. *Logistic regression using SAS: Theory and application*. SAS Institute, 2012.

Allison, Paul D., Richard Williams, and Enrique Moral-Benito. "Maximum likelihood for cross-lagged panel models with fixed effects." *Socius* 3 (2017): 1-17.

Arellano, Manuel. *Panel Data Econometrics*. Oxford University Press, 2003.

Chamlin, Mitchell B., and John K. Cochran. "Causality, economic conditions, and burglary." *Criminology* 36.2 (1998): 425-440.

Center for Human Resource Research (2002), *NLSY97 User's Guide*. Washington, DC: U.S. Department of Labor.

Firebaugh, Glenn, Cody Warner, and Michael Massoglia. "Fixed effects, random effects, and hybrid models for causal analysis." *Handbook of causal analysis for social research*. Springer, Dordrecht, 2013. 113-132.

Lieberson, Stanley. *Making it count: The improvement of social research and theory*. Univ of California Press, 1987.

Rosenberg, Andrew S., Austin J. Knuppe, and Bear F. Braumoeller. "Unifying the study of asymmetric hypotheses." *Political Analysis* 25.3 (2017): 381-401.

Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. MIT press, 2010.

York, Richard, and Ryan Light. "Directional asymmetry in sociological analyses." *Socius* 3 (2017): 1-13.

Appendix 1

Proof that if $\beta^+ = -\beta^-$, then $\beta^+ Z_{it}^+ + \beta^- Z_{it}^- = \beta^+ X_{it}$:

By definition and by assumption, we have

$$
\begin{aligned}
\beta^+ Z_{it}^+ + \beta^- Z_{it}^- &= \beta^+ Z_{it}^+ - \beta^+ Z_{it}^- \\
&= \beta^+ (Z_{it}^+ - Z_{it}^-) \\
&= \beta^+ \left(\sum_{s=2}^{t} X_{is}^+ - \sum_{s=2}^{t} X_{is}^- \right) \\
&= \beta^+ \sum_{s=2}^{t} (X_{is}^+ - X_{is}^-)
\end{aligned}
$$

The summations can start at time 2 because X_{is}^+ and X_{is}^- are defined to be 0 when $s=1$. Using the

result in eq. (3) in the main text, we can rewrite this as

$$
\begin{aligned}
\beta^+ Z_{it}^+ + \beta^- Z_{it}^- &= \beta^+ \sum_{s=2}^{t} (X_{is} - X_{is-1}) \\
&= \beta^+ \left(\sum_{s=2}^{t} X_{is} - \sum_{s=2}^{t} X_{is-1} \right) \\
&= \beta^+ \left(\sum_{s=2}^{t} X_{is} - \sum_{s=1}^{t-1} X_{is} \right) \\
&= \beta^+ X_{it}
\end{aligned}
$$

The final equality follows because the second summation exactly cancels all but the final term in

the first summation.

31

Appendix 2. Computer Code for Stata and SAS

```
***Stata code for Asymmetric Fixed Effects Models for Panel Data

***Section 2
use c:\data\nlsy.dta, clear
generate antidiff=anti92-anti90
generate selfdiff=self92-self90
generate povdiff=pov92-pov90
***Table 1
table povdiff
***Figure 1
hist antidiff, freq discrete
***Figure 2
hist selfdiff, freq
***Table 2
regress antidiff selfdiff povdiff


***Section 3
generate selfpos=selfdiff*(selfdiff>0)
generate selfneg=-selfdiff*(selfdiff<0)
generate povpos=povdiff*(povdiff>0)
generate povneg=-povdiff*(povdiff<0)
***Table 3
regress antidiff selfpos selfneg povpos povneg
test selfpos=-selfneg
test povpos=-povneg


***Section 4
use "C:\data\wagerate.dta", clear
gen id = _n
reshape long lwage mar edu urb, i(id) j(year)
xtset id year
***Table 4
xtreg lwage mar edu urb i.year, fe robust
***Table 5
gen wagediff=d.lwage
gen mardiff=d.mar
gen edudiff=d.edu
gen urbdiff=d.urb
reg wagediff mardiff edudiff urbdiff i.year, cluster(id)
***Table 6
xtreg wagediff mardiff edudiff urbdiff i.year, fe robust
***Table 7
mixed wagediff mardiff edudiff urbdiff i.year || id:, ///
   nocon res(uns,t(year)) stddev robust
***Table 8
mixed wagediff mardiff edudiff urbdiff i.year || id:, ///
   nocon res(toeplitz1,t(year)) stddev robust
```

```
***Section 5
***Table 9
tab mardiff
***Table 10
tab urbdiff
***Table 11 Not available in Stata

***Section 6
use "C:\data\wagerate.dta", clear
gen id = _n
reshape long lwage mar edu urb, i(id) j(year)
xtset id year
gen mardiff=d.mar
gen urbdiff=d.urb
replace mardiff=0 if mardiff==.
replace urbdiff=0 if urbdiff==.
gen marpos=mardiff*(mardiff>0)
gen marneg=-mardiff*(mardiff<0)
gen urbpos=urbdiff*(urbdiff>0)
gen urbneg=-urbdiff*(urbdiff<0)
gen marcumpos=marpos+L1.marpos+L2.marpos +L3.marpos if year==5
replace marcumpos=marpos+L1.marpos+L2.marpos if year==4
replace marcumpos=marpos+L1.marpos if year==3
replace marcumpos=marpos if year==2
replace marcumpos=0 if year==1
gen marcumneg=marneg+L1.marneg+L2.marneg +L3.marneg if year==5
replace marcumneg=marneg+L1.marneg+L2.marneg if year==4
replace marcumneg=marneg+L1.marneg if year==3
replace marcumneg=marneg if year==2
replace marcumneg=0 if year==1
gen urbcumpos=urbpos+L1.urbpos+L2.urbpos +L3.urbpos if year==5
replace urbcumpos=urbpos+L1.urbpos+L2.urbpos if year==4
replace urbcumpos=urbpos+L1.urbpos if year==3
replace urbcumpos=urbpos if year==2
replace urbcumpos=0 if year==1
gen urbcumneg=urbneg+L1.urbneg+L2.urbneg +L3.urbneg if year==5
replace urbcumneg=urbneg+L1.urbneg+L2.urbneg if year==4
replace urbcumneg=urbneg+L1.urbneg if year==3
replace urbcumneg=urbneg if year==2
replace urbcumneg=0 if year==1
xtreg lwage marcumpos marcumneg urbcumpos urbcumneg edu i.year, fe robust

***Section 7
***Table 12
use c:\data\teenyrs5.dta, clear
xtset id year
clogit pov mother spouse inschool hours i.year, group(id) robust
***Table 13
gen spousediff=d.spouse
gen inschooldiff=d.inschool
gen hoursdiff=d.hours
```

```
gen spousepos=spousediff*(spousediff>0)
gen spouseneg=-spousediff*(spousediff<0)
gen inschoolpos=inschooldiff*(inschooldiff>0)
gen inschoolneg=-inschooldiff*(inschooldiff<0)
gen hourspos=hoursdiff*(hoursdiff>0)
gen hoursneg=-hoursdiff*(hoursdiff<0)
gen spousecumpos=spousepos+L1.spousepos+L2.spousepos +L3.spousepos if year==5
replace spousecumpos=spousepos+L1.spousepos+L2.spousepos if year==4
replace spousecumpos=spousepos+L1.spousepos if year==3
replace spousecumpos=spousepos if year==2
replace spousecumpos=0 if year==1
gen spousecumneg=spouseneg+L1.spouseneg+L2.spouseneg +L3.spouseneg if year==5
replace spousecumneg=spouseneg+L1.spouseneg+L2.spouseneg if year==4
replace spousecumneg=spouseneg+L1.spouseneg if year==3
replace spousecumneg=spouseneg if year==2
replace spousecumneg=0 if year==1
gen inschoolcumpos=inschoolpos+L1.inschoolpos+L2.inschoolpos +L3.inschoolpos if
year==5
replace inschoolcumpos=inschoolpos+L1.inschoolpos+L2.inschoolpos if year==4
replace inschoolcumpos=inschoolpos+L1.inschoolpos if year==3
replace inschoolcumpos=inschoolpos if year==2
replace inschoolcumpos=0 if year==1
gen inschoolcumneg=inschoolneg+L1.inschoolneg+L2.inschoolneg +L3.inschoolneg if
year==5
replace inschoolcumneg=inschoolneg+L1.inschoolneg+L2.inschoolneg if year==4
replace inschoolcumneg=inschoolneg+L1.inschoolneg if year==3
replace inschoolcumneg=inschoolneg if year==2
replace inschoolcumneg=0 if year==1
gen hourscumpos=hourspos+L1.hourspos+L2.hourspos +L3.hourspos if year==5
replace hourscumpos=hourspos+L1.hourspos+L2.hourspos if year==4
replace hourscumpos=hourspos+L1.hourspos if year==3
replace hourscumpos=hourspos if year==2
replace hourscumpos=0 if year==1
gen hourscumneg=hoursneg+L1.hoursneg+L2.hoursneg +L3.hoursneg if year==5
replace hourscumneg=hoursneg+L1.hoursneg+L2.hoursneg if year==4
replace hourscumneg=hoursneg+L1.hoursneg if year==3
replace hourscumneg=hoursneg if year==2
replace hourscumneg=0 if year==1
clogit pov mother spousecumpos spousecumneg inschoolcumpos ///
 inschoolcumneg hourscumpos hourscumneg i.year, group(id) robust
```

```
***SAS code for Asymmetric Fixed Effects Models for Panel Data
***Note: proc glm and proc logistic, used below, cannot produce robust standard
errors;

***Section 2;
data nlsydiff;
set my.nlsy;
antidiff=anti2-anti1;
selfdiff=self2-self1;
povdiff=pov2-pov1;
selfpos=selfdiff*(selfdiff>0);
selfneg=-selfdiff*(selfdiff<0);
povpos=povdiff*(povdiff>0);
povneg=-povdiff*(povdiff<0);
run;
***Table 1;
proc freq data=nlsydiff;
table povdiff;
run;
***Table 2;
proc reg data=nlsydiff;
model antidiff=selfdiff povdiff;
run;

***Section 3;
***Table 3;
proc reg data=nlsydiff;
model antidiff=selfpos selfneg povpos povneg;
test selfpos=-selfneg;
test povpos=-povneg;
run;

***Section 4;
data wagediff;
set my.wagerate;
id=_n_;
array lwagea (*) lwage1-lwage5;
array mara (*) mar1-mar5;
array urba (*) urb1-urb5;
array edua (*) edu1-edu5;
do t=1 to 5;
lwage=lwagea(t);
mar=mara(t);
urb=urba(t);
edu=edua(t);
if t>1 then do;
  wagediff=lwagea(t)-lwagea(t-1);
  mardiff=mara(t)-mara(t-1);
  urbdiff=urba(t)-urba(t-1);
  edudiff=edua(t)-edua(t-1);
  marpos=mardiff*(mardiff>0);
```

```
   marneg=-mardiff*(mardiff<0);
   urbpos=urbdiff*(urbdiff>0);
   urbneg=-urbdiff*(urbdiff<0);
   t3=(t=3);
   t4=(t=4);
   t5=(t=5);
   end;
output;
end;
run;
***Table 4;
proc glm data=wagediff;
absorb id;
class t;
model lwage=mar edu urb t /solution ;
run;
***Table 5;
proc reg data=wagediff;
model wagediff=mardiff edudiff urbdiff t3 t4 t5 / hcc ;
run;
***Table 6;
proc glm data=wagediff;
absorb id;
model wagediff=mardiff edudiff urbdiff t3 t4 t5  ;
run;
***Table 7;
proc mixed data=wagediff empirical;
class t;
model wagediff=mardiff edudiff urbdiff t / solution;
repeated t / subject=id type=unr;
run;
***Table 8;
proc mixed data=wagediff empirical;
class t;
model wagediff=mardiff edudiff urbdiff t / solution;
repeated t / subject=id type=toep(2);
parms (-.5) (.2)/ ratios hold=1;
/* The PARMS statement constrains the ratio of the covariance to the residual variance
to be -.5 */
run;

***Section 5;
***Tables 9 & 10;
proc freq data=wagediff;
table mardiff urbdiff;
run;
***Table 11;
proc mixed data=wagediff empirical;
class t;
model wagediff=marpos marneg edudiff urbpos urbneg t / solution cl;
repeated t / subject=id type=toep(2);
```

```
parms (-.5) (.2)/ ratios hold=1;
contrast 'martest' marpos 1 marneg 1;
contrast 'urbtest' urbpos 1 urbneg 1;
run;

***Section 6;
data wagecum;
set my.wagerate;
id=_n_;
array mar (*) mar1-mar5;
array urb (*) urb1-urb5;
array edu (*) edu1-edu5;
array lwage(*) lwage1-lwage5;
array mardiff (*) mardiff1-mardiff5;
array urbdiff (*) urbdiff1-urbdiff5;
array marpos (*) marpos1-marpos5;
array marneg (*) marneg1-marneg5;
array urbpos (*) urbpos1-urbpos5;
array urbneg (*) urbneg1-urbneg5;
t=1;
marposcum=0; marnegcum=0; urbposcum=0; urbnegcum=0;
mart=mar1; urbt=urb1; edut=edu1; lwaget=lwage1;
output;
do t=2 to 5;
mart=mar(t);
urbt=urb(t);
edut=edu(t);
lwaget=lwage(t);
mardiff(t)=mar(t)-mar(t-1);
urbdiff(t)=urb(t)-urb(t-1);
marpos(t)=mardiff(t)*(mardiff(t)>=0);
marneg(t)=-mardiff(t)*(mardiff(t)<0);
urbpos(t)=urbdiff(t)*(urbdiff(t)>=0);
urbneg(t)=-urbdiff(t)*(urbdiff(t)<0);
marposcum=marpos(t)+marposcum;
marnegcum=marneg(t)+marnegcum;
urbposcum=urbpos(t)+urbposcum;
urbnegcum=urbneg(t)+urbnegcum;
output;
end;
run;
proc glm data=wagecum;
absorb id;
class t;
model lwaget=marposcum marnegcum urbposcum urbnegcum edut t / solution;
contrast 'mar test' marposcum 1 marnegcum 1;
contrast 'urb test' urbposcum 1 urbnegcum 1;
run;
```

```
***Section 7;
***Table 12;
proc logistic data=my.teenyrs5 desc;
 class year / param=ref;
 model pov = mother spouse inschool hours year;
 strata id ;
run;
***Table 13;
data teencum;
set my.teenpov;
array mothera (*) mother1-mother5;
array spouse (*) spouse1-spouse5;
array inschool (*) inschool1-inschool5;
array hours (*) hours1-hours5;
array spousediff (*) spousediff1-spousediff5;
array hoursdiff (*) hoursdiff1-hoursdiff5;
array inschooldiff (*) inschooldiff1-inschooldiff5;
array pova (*) pov1-pov5;
array spousepos (*) spousepos1-spousepos5;
array spouseneg (*) spouseneg1-spouseneg5;
array inschoolpos (*) inschoolpos1-inschoolpos5;
array inschoolneg (*) inschoolneg1-inschoolneg5;
array hourspos (*) hourspos1-hourspos5;
array hoursneg (*) hoursneg1-hoursneg5;
t=1;
spouseposcum=0; spousenegcum=0; inschoolposcum=0; inschoolnegcum=0;
hoursposcum=0; hoursnegcum=0;
pov=pov1; mother=mother1;
output;
do t=2 to 5;
pov=pova(t);
mother=mothera(t);
spousediff(t)=spouse(t)-spouse(t-1);
inschooldiff(t)=inschool(t)-inschool(t-1);
hoursdiff(t)=hours(t)-hours(t-1);
spousepos(t)=spousediff(t)*(spousediff(t)>0);
spouseneg(t)=-spousediff(t)*(spousediff(t)<0);
inschoolpos(t)=inschooldiff(t)*(inschooldiff(t)>0);
inschoolneg(t)=-inschooldiff(t)*(inschooldiff(t)<0);
hourspos(t)=hoursdiff(t)*(hoursdiff(t)>0);
hoursneg(t)=-hoursdiff(t)*(hoursdiff(t)<0);
spouseposcum=spousepos(t)+spouseposcum;
spousenegcum=spouseneg(t)+spousenegcum;
inschoolposcum=inschoolpos(t)+inschoolposcum;
inschoolnegcum=inschoolneg(t)+inschoolnegcum;
hoursposcum=hourspos(t)+hoursposcum;
hoursnegcum=hoursneg(t)+hoursnegcum;
output;
end;
run;
```

```
proc logistic data=teencum desc;
 class t / param=ref;
 model pov = mother spouseposcum spousenegcum
   inschoolposcum inschoolnegcum hoursposcum hoursnegcum t;
 strata id ;
 test spouseposcum=-spousenegcum;
 test hoursposcum=-hoursnegcum;
 test inschoolposcum=-inschoolnegcum;
run;
```