A Multilevel Multiset Time-Series Model for Describing Complex Developmental Processes

Applied Psychological Measurement
2017, Vol. 41(4) 294–310
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146621616686058
journals.sagepub.com/home/apm

Xin Ma¹ and Jianping Shen²

Abstract

The authors sought to develop an analytical platform where multiple sets of time series can be examined simultaneously. This multivariate platform capable of testing interaction effects among multiple sets of time series can be very useful in empirical research. The authors demonstrated that the multilevel framework can readily accommodate this analytical capacity. Given their intention to use the multilevel multiset time-series model to pursue complicated research purposes, their resulting model is relatively simple to specify, to run, and to interpret. These advantages make the adoption of their model relatively effortless as long as researchers have the basic knowledge and skills in working with multilevel growth modeling. With multiple potential extensions of their model, the establishment of this analytical platform for analysis of multiple sets of time series can inspire researchers to pursue far more advanced research designs to address complex developmental processes in reality.

Keywords

multilevel models, assessment, achievement testing

Introduction

Measurement of Time Series

A time series is a collection of observations of a well-defined variable obtained through repeated measurements at predetermined equal intervals over time (i.e., time points). Data that are irregularly obtained are usually not a time series. A time-series model aims to discover the underlying mechanism that produces the data dynamics over time for monitoring and forecasting purposes. Both persons and groups can provide time-series data. Children's early vocabulary development is measured every month during the toddler year. This time series is about persons with 12 time points. Schools' (average) reading performance is measured every year over a decade. This time series is about groups with 10 time points. A time series of persons usually takes the form of panel data (i.e., the same persons are repeatedly measured). Although one collects data from

¹University of Kentucky, Lexington, KY, USA

²Western Michigan University, Kalamazoo, MI, USA

Corresponding Author:

Xin Ma, University of Kentucky, 347 Dickey Hall, Lexington, KY 40506, USA.

Email: xin.ma@uky.edu

persons for a time series of groups, different groups may provide data to form the time series. The average reading score of the first graders in a school can be measured annually, but each year there is a different group of first graders.

Researchers sometimes deal with more than one time series. There are, in general, two ways that multiple time series can occur. One way is to obtain repeated measures from multiple variables at one set of time points (i.e., one time series for each variable). The other way is more complicated that obtains repeated measurements from one variable at multiple sets of time points (i.e., time points within time points). The authors work with the second way in the present study.

In the case of persons, persons can be measured repeatedly within a day across multiple days, resulting in both a “within” time series and a “between” time series (within and between are relative).¹ The authors refer to this case that involves the same persons in both time series as the true multiset time series. In the case of groups, the percentage of students who reach the level of proficiency in reading (within a school) can be measured at multiple grade levels (e.g., Grades 1-5) within a calendar year across multiple calendar years, resulting in one time series along the dimension of grade level and one time series along the dimension of calendar year. This case involves different groups in both time series. There are different groups (i.e., Grades 1-5) along the dimension of grade level, and each grade level has a different group in each calendar year. The authors refer to this case as the pseudo-multiset time series. The statistical model (framework) the authors aim to develop works with both cases.

Measurement Properties of Multiset Time Series

The need to measure multiset time series occurs in the effort of monitoring persons and groups progressing through complex developmental processes. Although the term, multiset time series, sounds general, a measurement property of time points within time points is present in the data. The authors have within-day time points within between-day time points and grade levels within calendar years. Another measurement property concerns the concept of time. In the true multiset time series (of persons), time is naturally defined (e.g., 1 day). For the (pseudo) multiset time series for groups, it may need to go beyond natural time. Although grade level (e.g., Grades 1-5) does indicate a time-related progression, it is not a form of natural time. The use of grade level as a time dimension is not a rare practice in the literature. Muthén (1997) stated that

longitudinal educational data is frequently collected at fixed time points corresponding to grades. The relevant time dimension for achievement is grade because this reflects the amount of learning that has taken place. (p. 150)

A final (desirable) measurement property of the authors’ statistical model is the allowance of both irregular (unequal) time intervals and missing measurements along any time dimension.

The overarching measurement issue is that repeated measures are not necessarily for persons only but may include multiple measures at the group level. Students at multiple grade levels within a school are measured in a calendar year, and then the same multiple grade-level measurement procedure is carried out within the same school across multiple calendar years. School (group) not student (person) becomes the unit measured repeatedly. This measurement property is unique and important, in that it extends the concept of the measurement unit and allows the flexibility in data collection in which, say, only group statistics are available.

Analysis of Time Series

Traditionally, the moving average approach creates a series of averages from different subsets (of time points) in a time series to derive an overall understanding of the trend in the data (Chou, 1975), and the autoregression approach regresses a value from a time series (as a dependent variable) on one or more previous values from the same time series (as predictors or independent variables; Box & Jenkins, 1970). One popular contemporary approach is random coefficient growth modeling often in the form of a multilevel model with data hierarchy of time points, persons, and groups (Raudenbush & Bryk, 2002). If the time dimension is transformed into a multivariate vector (thus reducing the data hierarchy from three levels to two), latent variable regression approaches can be carried out (e.g., in Mplus; Muthén, 1997). The majority of the empirical analyses on time series in the literature deal with a single outcome measure.

With multiple outcome measures, statistical approaches are needed to analyze multiple time series. As discussed earlier, one way for multiple time series to occur is to obtain repeated measures from multiple variables at one set of time points. These variables are often correlated, resulting in what is referred to as multivariate time series. The vector autoregression approach, a direct extension of the (univariate) autoregression approach, reveals the linear correlations among multiple time series (see Enders, 2010). Rare are the contemporary approaches to the analysis of multiple outcome time series. Ma and Ma (2005) studied the correlation between the rates of growth in mathematics and science achievement across grade levels, using the random coefficient growth modeling approach in the form of a multivariate multilevel model.

The literature is “thin” in the analysis of repeated measurements obtained from one variable at multiple sets of time points. The contemporary approaches may offer some extensions for this purpose. With this motivation and inspired by Ma and Ma (2005), the authors chose the random coefficient approach (over the latent variable approach) mainly for its relative simplicity. As Ma and Ma (2005) extended the random coefficient approach to the analysis of repeated measures obtained from multiple variables at one set of time points, the authors of this article aim to extend the same approach to the analysis of repeated measurements obtained from one variable at multiple sets of time points.

Multilevel Framework

A common statistical practice is to perform data analysis of time series in the analytical framework of multilevel modeling. Goldstein, Healy, and Rasbash (1994) presented the basic model of repeated measures as

$$y_{ij} = \sum_{k=0}^p \gamma_{jk} t_{ij}^k + \sum_{l=1}^q \alpha_l z_{ijl} + e_{ij}, \quad (1)$$

where y_{ij} represents the i th measurement for the j th person, t represents time, and z represents covariate. This model clearly separates within-subjects variation (Level 1) and between-subjects variation (Level 2) with (Level 1) error term (e_{ij}) assumed to be independently distributed with mean of 0 and constant variance. In the multilevel framework, the polynomial function descriptive of the within-subjects model generates coefficients to be taken as random to the between-subjects model (i.e., allowing variances and covariances among persons); covariates z_{ijl} can be applied as control variables at Level 1 (to adjust over the polynomial function) and/or predictor variables at Level 2 (to explain statistically significant variances and covariances of the polynomial function among persons).

If a linear function is taken as an example (see Raudenbush & Bryk, 2002), the multilevel model described above can be written as

$$y_{ij} = \gamma_{j0} + \gamma_{j1}t_{ij} + e_{ij}, \quad (2)$$

$$\gamma_{j0} = \alpha_{00} + \sum_{l=1}^q \alpha_{l0}z_{jl} + u_{j0}, \quad (3)$$

$$\gamma_{j1} = \alpha_{01} + \sum_{l=1}^q \alpha_{l1}z_{jl} + u_{j1}, \quad (4)$$

where γ_{j0} represents initial status and γ_{j1} represents rate of change (growth). They are taken from Level 1 (repeated measures) to Level 2 (persons) as random variables to be modeled by covariates z_{jl} . One of the major advantages of analyzing time series in the multilevel framework is the better handling of modeling error, with the addition of u_{j0} and u_{j1} as error terms at Level 2 (among persons). Extensions from this basic model are straightforward with the complication of the polynomial function (e.g., a quadratic function), the inclusion of (time-varying) covariates at Level 1 often as control variables to adjust the coefficients from the polynomial function, and the development of a higher level (e.g., group) to model the coefficients from Level 2 (e.g., individual differences in rate of growth) with group-level variables (at Level 3).

Multilevel Multiset Time Series

Multiset time series offers excellent opportunities to examine both “main effects” of different sets of time series and interaction effects among multiple sets of time series. Conceptually, if each set of time series can be considered as one “factor” in essence similar to that in ANOVA, then the ANOVA notions of main effects and interaction effects can be easily carried over. In the earlier example of reading proficiency measured from two sets of time series of calendar year and grade level, each time series can be examined as main effect, and the two time series can be tested for an interactive relationship as interaction effect. When this interaction effect is statistically significant, change in reading proficiency across grade levels is dependent on calendar year, or change in reading proficiency across calendar years is dependent on grade level. Both main effects and interaction effects of multiset time series can be studied in the multilevel framework as presented in Equations 2 to 4.

The authors aim to develop an analytical framework suitable for working with multiset time series that can be established using multilevel modeling techniques for some analytical advantages. They intend to describe their multilevel multiset time-series model, followed by an application of the model to examine changes (growth) in terms of the percentage of students who reach the level of proficiency in reading within a school (i.e., school as the unit of repeated measurement) across calendar years and grade levels under an intervention program (i.e., with treatment effects).

Model

Level 1 Model

The authors intend to use the case of two time series, the simplest of all multilevel multiset time-series models, to demonstrate the essential structure of their model. In this multilevel double time-series model, they incorporate the two time series into one two-level growth model.

With the outcome measure (dependent variable) and the two time series (time factors or time variables) plus the covariates that can be time varying, the Level 1 model can be expressed as

$$y_{ti} = \pi_{0i} + \pi_{1i}T_{1ti} + \pi_{2i}T_{2ti} + \pi_{3i}(T_1 \times T_2)_{ti} + \sum_{n=4}^m \pi_{ni}X_{(n-3)ti} + \varepsilon_{ti}, \quad (5)$$

where π_{0i} represents the initial status, π_{1i} represents the rate of change associated with the first time variable (T_1) when the second time variable (T_2) is 0, π_{2i} represents the rate of change associated with the second time variable (T_2) when the first time variable (T_1) is 0, and π_{3i} represents the interaction effect between the two time series (T_1 and T_2). One advantage of this multilevel double time-series model is its ability to model the interaction effect between changes. Time-varying covariates $X_{(n-3)ti}$ are collected within the sigma to adjust the outcome measure (dependent variable) for these covariates. Finally, the error term at Level 1, ε_{ti} , is independently distributed (i.e., residuals are uncorrelated).²

The interaction term in the model entails certain assumptions. The linear specification of both time series dictates that the change along the first time dimension is either strictly increasing or strictly decreasing for each time point of the second time series (and vice versa). The critical “turning point” if it exists cannot be identified with any linear specification of change. To be discussed later on, the model does have the potential to allow the nonlinear specification of change along any time dimension without substantially complicating the model.

Nesting of Time Points

One of the measurement properties of multiset time series (i.e., time points within time points) was briefly discussed. The authors now have the context to discuss it in detail as a condition on time series. Time progresses and any two time series based on a one-on-one correspondence of time will demonstrate a perfect correlation. A child takes part in an experiment that collects data monthly on an outcome for 3 months following an intervention for a problem (i.e., 1, 2, 3 as data for one time series). This time series is used to examine the treatment effect. Meanwhile, the child is 3 months old when entering the experiment and grows over 3 months during the experiment (i.e., 3, 4, 5 as data for the other time series). This time series aims to examine whether maturity can help the child “grow” naturally out of the problem. However, because of the one-on-one correspondence between the two time series, 1 is their correlation coefficient.

Some sort of nesting of time points must exist as a condition or requirement to establish data for analysis of multiple sets of time series. In the above example, if data collection is bimonthly, then one time series is .5, 1, 1.5, 2, 2.5, 3 and the other is 3, 4, 5 (i.e., 3, 3, 4, 4, 5, 5 when data are established). These two time series are not perfectly correlated anymore (.96) because of the nesting (e.g., .5 and 1 from the first time series are nested within 3 from the second time series). The more complex the nesting between two time series, the weaker the correlation between them. This requirement is not unique to the authors' model. Any time series design involving more than one time series needs to deal with this requirement. Nonetheless, this nesting requirement has its advantage, in that two time series can be very different in length (i.e., in the number of time points). This allows for data analysis of complex and irregular sets of time series. The authors may change data collection in the above experiment to .9 nested within 3 (collecting data after .9 months of intervention when a child is 3 months old); 1.1, 1.3, 1.5, 1.7, and 1.9 nested within 4; and 2.1, 2.3, 2.5, 2.7, and 2.9 nested within 5 (i.e., .9, 1.1, 1.3, 1.5, 1.7, 1.9, 2.1, 2.3, 2.5, 2.7, 2.9 and 3, 4, 4, 4, 4, 4, 5, 5, 5, 5). In fact, complex and irregular nesting of time points reduces correlation substantially between two time series (.89 in this case).

Centering Time Points

In time-series analysis, it is typical to center time points (i.e., decide a time point as Time 0). Centering affects the intercept (status) not the slope (rate of change). The time series of months (3, 4, 5) may be centered around 3, turning the time series into 0, 1, 2, so that the intercept becomes the initial status. If those time points are not centered, Time 0 still exists. It is the time of birth, a hypothetical time point outside the time series of interest. Centering is important to the interpretation of the rate of change, particularly in the authors' multilevel multiset time-series model. Earlier, π_{1i} represents the rate of change associated with the first time variable (T_1) when the second time variable (T_2) equals 0 (i.e., Time 0). Stated differently, π_{1i} is the rate of change that the authors observe when they "stand" at Time 0 of T_2 . Without centering, they stand outside the time points of T_2 . Although π_{1i} is the same, its interpretation may not be relevant. Centering becomes an important issue again when they discuss the extension of their multilevel multiset time-series model to include some nonlinear (e.g., quadratic) relationships between two time series.

Level 1 Covariates

Although the authors highlighted earlier time-varying variables at Level 1, both time-varying and time-invariant covariates can be used within the sigma as adjustment over the outcome measure. When time-varying variables are applied, data can be collected according to either time series, even though data collection based on the longer time series is preferable for its potentially more heterogeneous distribution of data, which reduces standard error in estimation. Often, covariates at Level 1 function as control for a "purer" estimation of the rate of change, and their coefficients are usually not modeled at Level 2. In any case where the modeling of those coefficients makes sense, they can be taken with the coefficients of the polynomial function (descriptive of the nature of change) to the next level as dependent variables for modeling purposes.

Conception of Added Value

The concept of added value can be put into statistical practice at Level 1. As discussed earlier, working with the notion of Time 0 (e.g., the first time point or initial status that can be defined as Time 0), when T_2 is at its initial status, Equation 5 becomes

$$y_{ti} = \pi_{0i} + \pi_{1i}T_{1ti} + \sum_{n=4}^m \pi_{ni}X_{(n-3)ti} + \varepsilon_{ti}, \quad (6)$$

where π_{1i} represents the rate of change associated with T_1 . When T_2 takes on a special value of s (e.g., at the middle point of the second time series) in Equation 5, we have

$$y_{ti} = (\pi_{0i} + s\pi_{2i}) + (\pi_{1i} + s\pi_{3i})T_{1ti} + \sum_{n=4}^m \pi_{ni}X_{(n-3)ti} + \varepsilon_{ti}, \quad (7)$$

where the rate of change associated with T_1 is now $\pi_{1i} + s\pi_{3i}$. A comparison between Equations 6 and 7 indicates that $s\pi_{3i}$ can be defined as the value added to the rate of change associated with T_1 by T_2 when T_2 progresses to a special time point (s). This conception can bear important implications. In the example of intervention (time series) and age (time series), the added value

of maturity to the rate of change associated with intervention is often informative or even critical.

Level 2 Model

At the second level, the authors aim to model the initial status, the rate of change for each time series, and the interaction between the two time series with Level 2 variables. The need for the modeling of the polynomial function descriptive of change is obvious, but coefficients of covariates at Level 1, as they discussed, can also be modeled at this level (as random slopes) if necessary. Their multilevel double time-series model can be expressed at Level 2 as

$$\pi_{0i} = \theta_{00} + \sum_{p=1}^q \theta_{0p} Z_{pi} + u_{0i}, \quad (8)$$

$$\pi_{1i} = \theta_{10} + \sum_{p=1}^q \theta_{1p} Z_{pi} + u_{1i}, \quad (9)$$

$$\pi_{2i} = \theta_{20} + \sum_{p=1}^q \theta_{2p} Z_{pi} + u_{2i}, \quad (10)$$

$$\pi_{3i} = \theta_{30} + \sum_{p=1}^q \theta_{3p} Z_{pi} + u_{3i}, \quad (11)$$

where Z_p ($p = 1, 2, 3, \dots, q$) are Level 2 covariates, and their coefficients measure relationships of Level 2 covariates to the initial status (π_{0i}), both rates of change (π_{1i} and π_{2i}), and the interaction between rates of change (π_{3i}). Finally, u_{0i} to u_{3i} are Level 2 error terms for respective parameters.

The four equations examine and describe at least three things: First, they produce grand point estimates for the initial status (θ_{00}), the rate of change for both time series (θ_{10} and θ_{20}), and the interaction between rates of change (θ_{30}). Second, they generate a matrix of variances and covariances among Level 2 units for testing the statistical significance of the variance in the initial status (u_{0i}), both rates of change (u_{1i} and u_{2i}), and the interaction between rates of change (u_{3i}) across Level 2 units. Finally, for each coefficient of the polynomial function with statistically significant variance, Z_p ($p = 1, 2, 3, \dots, q$) can be applied to account for that variance. Each coefficient can be modeled with a different set of Level 2 covariates if necessary.

With Equations 8 to 11 specifying four random effects at Level 2, the assumed variance–covariance structure (matrix) takes the unstructured form with variances on the diagonal and covariances symmetrically off the diagonal. The covariances often bear important implications, particularly when a model aims to understand institution effects. The common practice is to turn the variance–covariance matrix into a correlation matrix. In this case, the correlation matrix concerns the relationships among the initial status, the rate of change for each time series, and the interaction between the two time series (e.g., whether the rate of change for one time series is positively and strongly correlated with the rate of change for the other time series). Whether correlations depend upon Level 2 variables can also be examined by comparing the models with and without Level 2 variables. If correlations change little between the two models, then relationships among the initial status, rates of change, and the interaction are independent of Level 2 variables (thus generating an opportunity to examine institution effects).

Model Flexibility

The multilevel multiset time-series model is a very flexible general analytical framework for time-series data. It is first noted that the denotation of i in Equation 5 can stand for both person (e.g., student) and group (e.g., school). In the case of i for person, the authors deal with what they referred to earlier as the true multiset time series such as persons measured 3 times a day (one time series) across multiple days (one time series). In the case of i for group, they deal with what they referred to earlier as the pseudo-multiset time series such as schools measured at multiple grade levels within a calendar year (one time series) across multiple calendar years (one time series). Such a flexibility allows either persons or groups to be situated in the model.

The multilevel multiset time-series model can handle any number of time series, so that single time-series techniques become special cases. In the case of one time series, the authors' model can work with both (multiple) persons and multiple groups (i.e., group time series). If interventions are manifested in groups, treatment effects can be conveniently examined with their model.

Model Extensions

The multilevel double time-series model contains two levels: Equation 5 as Level 1 and Equations 8 to 11 as Level 2. A typical application is repeated measures nested within persons. This model can be easily extended to include some higher levels of units. In the case of three levels with repeated measures nested within persons nested within groups, y_{tij} can be used to represent the t th measurement for the i th person in the j th group:

$$y_{tij} = \pi_{0ij} + \pi_{1ij}T_{1tij} + \pi_{2ij}T_{2tij} + \pi_{3ij}(T_1 \times T_2)_{tij} + \sum_{n=4}^m \pi_{nij}X_{(n-3)tij} + \varepsilon_{tij}, \quad (12)$$

$$\pi_{1ij} = \theta_{10j} + \sum_{p=1}^q \theta_{1pj}Z_{pij} + u_{1ij}, \quad (13)$$

$$\theta_{10j} = \gamma_{100} + \sum_{k=1}^l \gamma_{10k}W_{kj} + v_{10j}, \quad (14)$$

where θ_{10j} represents average rate of change (associated with the first time series) for the j th group taken to Level 3 to be modeled by group-level covariates W_{kj} ($k = 1, 2, 3, \dots, l$) and v_{10j} is Level 3 error term (associated with θ_{10j}). Now, γ_{100} is the grand rate of change for the first time series, v_{10j} measures the variance in average rate of change across groups, and γ_{10k} ($k = 1, 2, 3, \dots, l$) measures relationships of Level 3 covariates to average rate of change among groups.

Obviously, the authors used the first time series as an example to demonstrate the model structure at Levels 2 and 3, Equations 13 and 14. Other parts of the polynomial function (that describes the nature of change) can be treated in the same manner. The three-level double time-series model is capable of exploring person and group effects on the coefficients from the polynomial function.

Instead of working with the levels of data hierarchy, the other extension of the authors' multilevel multiset time-series model focuses on the polynomial function that describes the nature of change. Both two-level and three-level double time-series models deal with the linear change as in Equation 5 with the linear rate of change for both time series. From the linear change to the nonlinear change represents another model extension. Equation 5 can be rewritten as

$$y_{ti} = \pi_{0i} + \pi_{1i}T_{1ti} + \pi_{2i}T_{1ti}^2 + \pi_{3i}T_{2ti} + \pi_{4i}(T_1 \times T_2)_{ti} + \pi_{5i}(T_1^2 \times T_2)_{ti} + \sum_{n=6}^m \pi_{ni}X_{(n-5)ti} + \varepsilon_{ti}, \quad (15)$$

where the first time series takes on a quadratic function and, for simplicity of demonstration, the second time series remains in a linear form. The addition of the quadratic term (which represents the acceleration or deceleration of change) allows the authors to examine a new interaction effect between the quadratic term of the first time series and the second time series. This new addition can bear very informative insights on whether the quadratic change (i.e., acceleration or deceleration of change) associated with the first time series depends on the linear change associated with the second time series. In Equation 15, if π_{5i} is statistically significant, then the rate of change of the first time series is accelerated or decelerated (in a quadratic manner) depending on a faster or slower linear rate of change of the second time series.

Obviously, Equation 15 depicts the Level 1 model. The complication of the polynomial function that defines the nature of change provides more opportunities for the coefficients of the complicated polynomial function to be modeled at the person level and even at the group level in the same manner as demonstrated earlier. For example, the new interaction effect between the quadratic term of the first time series and the second time series can be examined in terms of person (individual) differences (at Level 2) and group effects (at Level 3).

In Equation 15, the second time series can also be defined as a quadratic function of change, which will allow the formation of even more interaction terms between elements of both time series. For example, a new interaction effect between the quadratic terms of both time series can be established. The further complication of the polynomial function that defines the nature of change then allows more modeling at different (higher) levels (e.g., person and group levels).

Model Limitations

Despite the multiple strengths of the multilevel multiset time-series model, limitations exist. Related to measurement, the (research) design based on which measurement is carried out needs to create a situation of time points within time points, so that the two time series are not perfectly correlated. As a result, some measurement concerns may occur (e.g., carry over effects if the same measure is used many times over a short period of time). Related to estimation, the (pseudo) multilevel multiset time-series model (with groups rather than individuals at Level 1) does not inform about individual growth in the way a traditional repeated measures model would and thus is somewhat limited in its potential to draw strong inferences about, say, treatment effects. Overall, the multilevel multiset time-series model does not eliminate the need to collect repeated measures data at the individual level. The authors' application to follow relies on the pseudo-multilevel multiset time-series model. The extent to which the composition of students changes across both grade levels and calendar years can also create a potential problem for the model. Contextual information about student composition needs to be collected and used as (time-varying) covariates at Level 1 to adjust for the outcome measure as shown in Equation 5. Although this control alleviates this problem, resources may be needed to collect additional contextual data on student composition across both grade levels and calendar years.

Application

The application of the authors' multilevel multiset time-series model deals with the evaluation of the Learning-Centered Leadership Development (LCLD) Program for Practicing and Aspiring Principals developed at Western Michigan University. Under this multiyear school

leadership intervention program, a team of principals and teacher leaders from each participating school were trained to develop and implement school renewal activities. The authors carried out an investigation that aimed to assess the effect of the LCLD program on school-level student achievement.

Data

The authors used school as their sampling unit, and their sample was comprised of 46 schools with grade levels between Grade 3 and Grade 8. Each school provided data that covered 4 years, between 2010 and 2013, of information on school background and school-level student achievement. Therefore, the time span was 4 years (from 2010-2013), while the grade span was six grade levels (from Grades 3-8). Among these 46 schools, 20 were assigned to the experimental group (with LCLD), while 26 were assigned to the control group (without LCLD).

Measures

Student achievement was adopted from the Michigan Educational Assessment Program (MEAP) for students in Grades 3 to 9. MEAP assessment is content based, is aligned directly to the statewide content standards for elementary and middle schools, and shows satisfactory internal consistency (Michigan Department of Education, 2006). The authors obtained (school-level) MEAP scores in English and calculated for each school a percentage of students who reached the proficient level in English. Those (school) percentages became the outcome measure for their evaluation. The authors also collected information on critical school contextual characteristics from schools' websites, including school (enrollment) size, proportion of male students (a measure of school gender composition), proportion of White students (a measure of school racial-ethnic composition), and proportion of students eligible for free lunch (a measure of school socioeconomic composition). Because these school contextual variables change from year to year, they used them as school-level time-varying variables to adjust for school-level student (English) achievement, which also changes from year to year.³

Analysis

Each school contained information on school-level student (English) achievement and school contextual characteristics across four calendar years. The authors could develop a multilevel growth model with calendar years nested within schools to compare the rate of growth in school-level student (English) achievement across four calendar years between the experimental and control groups. Meanwhile, each school contained a number of grade levels for each of the four calendar years. Within each year, the authors could develop a multilevel growth model with grade levels (at Level 1) nested within schools (at Level 2) to compare the rate of growth in school-level student (English) achievement across grade levels between the experimental and control groups. Obviously, a series of univariate analyses of these two time series (calendar years and grade levels) would be undesirable or even flawed.

The analytical situation in which each school contained information on school-level student (English) achievement and school contextual characteristics across a span of six grade levels (from Grades 3-8) over a period of four calendar years (from 2010-2013) fits well into the authors' multilevel multiset time-series model, with (school-level) repeated measures (at Level 1) nested within schools (at Level 2). The authors incorporated both time series into one multilevel growth model (at Level 1). The Level 1 model, Equation 16, is the same as Equation 5.

Table 1. Results of a Multilevel Double Time-Series Model Estimating Growth in English Performance Across Grade Levels (3-8) and Calendar Years (2010-2013).

Fixed effects			Effect	SE		
IS = initial status (Grade 3 in 2010)			44.87*	1.66		
G = rate of growth across grade levels (3-8)			-0.80	0.71		
Y = rate of growth across calendar years (2010-2013)			-1.46*	0.60		
G × Y			0.86*	0.30		
Random effects			IS	G	Y	G × Y
Within subjects (schools)		71.08				
IS		74.80*				
G		2.07	6.61*			
Y		5.91	0.71	2.81		
G × Y		-6.21	-2.36	-0.53		1.27*

Note. For χ^2 tests of variances, $df = 44$. Estimates are adjusted for time-varying (by year) school characteristics of school (enrollment) size, proportion of male students, proportion of White students, and proportion of students eligible for free or reduced-price lunch. Deviance = 3,852.29 (number of estimated parameters = 11).

* $p < .05$.

$$y_{ti} = \pi_{0i} + \pi_{1i}T_{1ti} + \pi_{2i}T_{2ti} + \pi_{3i}(T_1 \times T_2)_{ti} + \sum_{n=4}^7 \pi_{ni}X_{(n-3)ti} + \varepsilon_{ti}. \tag{16}$$

Specifically, i stands for school (group) not student (person), and the Level 1 model contains the outcome measure (school-level student achievement in English) and two time factors denoting calendar year and grade level. The authors' model is able to test the interaction effect between growth (i.e., between two time series). There are also time-varying school contextual variables functioning as control for school-level student achievement. In Equation 16, π_{0i} is the initial status (with Time 0 defined as Grade 3 and 2010, respectively), π_{1i} is the rate of growth associated with calendar years, π_{2i} is the rate of growth associated with grade levels, and π_{3i} is the interaction effect (between the two time series). Time-varying variables ($X_{(n-3)ti}$) are collected within the sigma to adjust for school-level student achievement. Finally, ε_{ti} is the error term.

In general, the second level of this multilevel multiset time-series model tests the initial status, both rates of growth, and the interaction between growth (all as school-level parameters) with school-level variables descriptive of educational interventions, program initiatives, and educational policies and practices. The authors' focus was the implementation of LCLD. The (school-level) dummy variable, E_i , was the only variable at the school level denoting participation of schools in LCLD (comparing the experimental group with the control group).

$$\pi_{si} = \theta_{s0} + \theta_{s1}E_i + u_{si} \ (s = 0, 1, 2, 3). \tag{17}$$

So θ_{s1} ($s = 0, 1, 2, 3$) represents a measure of the effectiveness of LCLD (i.e., treatment effects) on the initial status, particularly on two rates of growth and the interaction between growth. Finally, u_{0i} to u_{3i} are error terms for respective (school-level) parameters.

The authors used the HLM 7.01 software (Raudenbush, Bryk, & Congdon, 2013) to generate the estimates using the full-maximum-likelihood (FML) approach with the expectation-maximization (EM) algorithm.⁴ Table 1 presents the results from the authors' multilevel multiset time-series model on the growth patterns in (school-level) student English achievement across grade levels (Grades 3-8) and calendar years (2010-2013).

Results (Meanings) of Initial Status

The intercept (π_{0i}) indicates school-level initial status in student English achievement that represents the average school percentage of Grade 3 students who performed at the proficient level in English in 2010 in a (hypothetical) school where all covariates take on the value 0. Centering those covariates around their grand means as the authors did makes the initial status more interpretable. Then, the hypothetical school becomes one with average school contextual characteristics (e.g., with average school enrollment size). Overall, the intercept (π_{0i}) indicates some kind of status, depending on how centering of time points is done for each time series.

School-level initial status was statistically significant (i.e., significant fixed effect) and varied statistically significantly across schools (i.e., significant random effect). The fixed effect shows that 44.87% of students in Grade 3 performed at the proficient level in English in 2010 in the (hypothetical) school. This value should be understood as the average school percentage because there is a percentage for each school. The initial status is an interesting measure, but the rates of growth along time dimensions (i.e., slopes) are the focus of interpretation in most cases.

Results (Meanings) of Rates of Growth

The authors reiterate that π_{1i} and π_{2i} represent the rate of growth associated with one time variable (only) when the other time variable equals 0. They interpret π_{1i} as school-level rate of growth across grade levels (Grades 3-8) in 2010. Stated differently, in 2010, a certain amount of growth in percentage (π_{1i}) might be expected from one grade level to the next (e.g., from Grades 3-4). Nonetheless, π_{1i} was not statistically significant and did not vary statistically significantly across schools. Meanwhile, the authors interpret π_{2i} as school-level rate of growth across calendar years (2010-2013) for Grade 3. Stated differently, for Grade 3, a certain amount of growth in percentage (π_{2i}) might be expected from one calendar year to the next (e.g., from 2010-2011). In this case, π_{2i} was statistically significant but did not vary statistically significantly across schools. In fact, π_{2i} was negative, indicating that schools declined rather than grew in percentage across calendar years in the case of Grade 3. Nonetheless, these results cannot be taken as independent statements (findings) about the growth pattern of either time series, because of the statistically significant interaction effect between grade level and calendar year.

Regardless of the multilevel multiset time-series model, there are two ways to make sense of rate of growth. First, each (linear) rate of growth represents a constant change from one time point to the next (i.e., the same slope) along a certain time dimension (series). Second, to some researchers, because measuring multiple grade levels (in a calendar year) does not produce panel data, there is not really a rate of growth across grade levels. Instead, they prefer the term of an average difference in percentage between grade levels. Both interpretations speak to the same phenomenon because, even in the case of calendar year as time series for which the first way makes natural sense, the rate of growth can still be interpreted as an average difference in percentage between calendar years.

Results (Meanings) of Interaction Effect

The slope (π_{3i}) measures the interaction between time series. A fixed interaction effect examines whether the rate of growth across one time dimension differs between time points of the other time dimension, or the degree to which persons or groups grow along one time dimension varies between time points of the other time dimension. A random interaction effect examines

Table 2. Change in School-Level Percentage of Students at Proficient Level in English by Calendar Year and Grade Level.

	%
Calendar year	
2010	−0.80
2011	0.06
2012	0.92
2013	1.78
Grade level	
3	−1.46
4	−0.60
5	0.26
6	1.12
7	1.98
8	2.84

whether the pattern that the rate of growth across one time dimension differs between time points of the other time dimension is consistent across persons or groups.

The interaction effect between grade level and calendar year was statistically significant and varied statistically significantly across schools. It was positive, indicating that school-level rate of growth across grade levels (from Grades 3-8) increased across calendar years (from 2010-2013). The authors rely on some algebraic manipulations to bring numerical values into Equation 16 as a way to understand how the rate of growth of one time factor changes at each given point of the other time factor (see Table 2). In 2010, the (within) school percentage of students who performed at the proficient level in English decreased from Grade 3 to Grade 8 at 0.80% from one grade level to the next. Then, the (within) school percentage of students who performed at the proficient level in English increased from Grade 3 to Grade 8 between 2011 and 2013, and the degree of such an increase became stronger across calendar years. The magnitude stood at 0.06% in 2011, 0.92% in 2012, and 1.78% in 2013 (all from one grade level to the next).

Taking a different perspective, school-level rate of growth across calendar years (from 2010-2013) increased across grade levels (from Grades 3-8). Similar algebraic manipulations (see Table 2) show that, for both Grades 3 and 4, the (within) school percentage of students who performed at the proficient level in English decreased from 2010 to 2013, and Grade 4 indicated a smaller degree of decrease across calendar years than Grade 3. The magnitude of decrease from one calendar year to the next was 1.46% for Grade 3 and 0.60% for Grade 4. For each of Grades 5 to 8, the (within) school percentage of students who performed at the proficient level in English increased from 2010 to 2013, and the degree of such an increase became stronger from Grades 5 to 8. The magnitude stood at 0.26% for Grade 5, 1.12% for Grade 6, 1.98% for Grade 7, and 2.84% for Grade 8 (all from one calendar year to the next).

Finally, the (statistically significant) random interaction effect indicated that schools did not share the same pattern of school-level rate of growth across grade levels (from Grades 3-8) increasing across calendar years (from 2010-2013) or of school-level rate of growth across calendar years (from 2010-2013) increasing across grade levels (from Grades 3-8).

Results (Meanings) of Program Effects

The authors then added program effects to growth patterns (see Table 3). They used the program dummy to model the initial status, both rates of growth, and the interaction effect. They did not

Table 3. Results of a Multilevel Double Time-Series Model Estimating Program Effects on Growth in English Performance Across Grade Levels (3-8) and Calendar Years (2010-2013).

Fixed effects		Effect	SE		
IS = initial status (Grade 3 in 2010)					
Intercept		45.89*	2.39		
Program (1 = yes, 0 = no)		−2.04	3.37		
G = rate of growth across grade levels (3-8)					
Intercept		−0.54	0.80		
Program (1 = yes, 0 = no)		−0.75	1.46		
Y = rate of growth across calendar years (2010-2013)					
Intercept		−1.81 *	0.82		
Program (1 = yes, 0 = no)		0.58	1.17		
G × Y					
Intercept		0.63	0.39		
Program (1 = yes, 0 = no)		0.69	0.58		
Random effects		IS	G	Y	G × Y
Within subjects (schools)		70.72			
IS		77.25*			
G		1.50	6.51*		
Y		6.74	0.79	2.90	
G × Y		−6.28	−2.13	−0.61	1.18*

Note. For χ^2 tests of variances, $df = 43$. Estimates are adjusted for time-varying (by year) school characteristics of school (enrollment) size, proportion of male students, proportion of White students, and proportion of students eligible for free or reduced-price lunch. Deviance = 3,852.29 (number of estimated parameters = 11).

* $p < .05$.

find any statistically significant program effects. Statistically significant results indicate that 45.89% of Grade 3 students in the school with average school contextual characteristics that did not participate in LCLD performed at the proficient level in English in 2010 (i.e., when all predictors in the model take on the value 0 with school contextual variables centered). In addition, schools that did not participate in LCLD, on average, declined in terms of percentage of Grade 3 students who performed at the proficient level in English by 1.81 from one calendar year to the next between 2010 and 2013. For one example of statistically significant random effects, school-level initial status (Grade 3 in 2010) varied statistically significantly across schools (after controlling for participation in LCLD).

What if there are statistically significant program effects? The (four) intercepts in Table 3 can be considered respectively measures for schools not in LCLD (see the above interpretation), and program effects can then be considered respectively “added values” (either positive or negative) of schools in LCLD. If program effects along the time dimension of calendar years (0.58) were statistically significant, then the rate of growth across calendar years would be $-1.81 + 0.58 = -1.23$ for schools in LCLD. Thus, the notion of added values can generate measures for schools in LCLD. Especially, if interaction effects are statistically significant, tables similar to Table 2 can be generated for both experimental and control groups to show program effects.

Results (Meanings) of Covariances

Random effects are presented in the form of variance–covariance matrix in Tables 1 and 3. The authors have interpreted statistically significant variances (on the diagonal). Covariances (off

Table 4. Correlations Among Random Effects of Initial Status, Rates of Growth, and Interaction.

	IS	G	Y	G × Y
IS = initial status (Grade 3 in 2010)		.09	.41	−.64
G = rate of growth across grade levels (3-8)	.07		.17	−.81
Y = rate of growth across calendar years (2010-2013)	.45	.18		−.28
G × Y	−.66	−.77	−.33	

Note. The upper and lower triangles correspond to the variance–covariance matrices in Tables 1 and 3, respectively.

the diagonal) can also be important. Correlations are usually calculated from the matrix to show relationships among random effects (see Table 4). For example, the rate of growth across grade levels was negatively and strongly correlated to the interaction, indicating that schools with a stronger rate of growth across grade levels showed a weaker interaction. A comparison between the upper and lower triangles shows trivial changes, indicating that the relationships among initial status, rates of growth, and interaction were largely independent of Level 2 (school-level) variables (the program dummy was the only Level 2 variable in this case).

The correlation matrices in Table 4 also make the model useful in assessing the potential to deconfound grade-level and calendar year effects within a school (as both represent causal effects due to the school). High correlations indicate the model’s inability to deconfound grade-level versus calendar year effects, likely resulting in the lack of program effects along either time dimension. The authors encountered the other scenario in which correlations were low (.17 and .18) between grade-level and calendar year effects. Therefore, they were reasonably deconfounded, likely due to changes in educational policies and practices (e.g., standard setting) over time, given the presence of time-varying covariates descriptive of school composition.

Pseudo R²

Following the tradition of multilevel growth modeling, Table 1 is considered the baseline model that contains only dependent and time variables (at Level 1). The performance of a new model such as Table 3 built upon the baseline model is assessed using *Pseudo R²* (i.e., proportion of variance explained by the new model):

$$Pseudo\ R^2 = \frac{Var_{(baseline)} - Var_{(new)}}{Var_{(baseline)}}, \tag{18}$$

where *Var* indicates variance. *Pseudo R²* can be calculated for each random coefficient (intercept or slope) from Level 1. For example, variance associated with the rate of growth across grade levels is 6.61 from Table 1 (baseline) and 6.51 from Table 3 (new). *Pseudo R²* is .02, indicating that program effects were unable to explain the variance in the rate of growth across grade levels.

Occasionally, the addition of variables at Level 2 may produce larger variance in the new model, resulting in negative *Pseudo R²*. For example, variance associated with initial status is 74.80 from Table 1 (baseline) and 77.25 from Table 3 (new), so that *Pseudo R²* is negative. Holden, Kelley, and Agarwal (2008) stated that

it should be noted that in some circumstances, it is possible for *Pseudo R²* statistic to be negative because it is possible the inclusion of an additional predictor can increase the magnitude of the variance component (Hedeker & Gibbons, 2006; Singer & Willett, 2003). When a negative *Pseudo R²*

statistic occurs, the Pseudo R^2 is largely uninterpretable and should not generally be used. (Raudenbush & Bryk, 2002; Singer & Willett, 2003; pp. 798-799)

Summary

The authors have attempted to establish an analytical platform where multiple sets of time series can be examined simultaneously.⁵ They believe that this multivariate platform is capable of testing interaction effects among multiple sets of time series, and that any analytical platform with such a capacity can be very useful in empirical research. They have then demonstrated that the multilevel framework can readily accommodate this analytical capacity. Interestingly, given the authors' intention to use the multilevel multiset time-series model to pursue complicated research purposes, the resulting model is relatively simple to specify, to run, and to interpret.

The multilevel multiset time-series model is relatively simple to specify, in that a three-level model is sufficient to examine growth patterns and program effects up to the group level (e.g., repeated measures, students, and schools). Because interaction effects among multiple sets of time series are specified at Level 1, complex interactive relationships can be unfolded at both Levels 2 and 3 (e.g., students and schools). The authors showed that extensions of their model embrace many substantive research issues without the need to venture beyond a three-level model. The multilevel multiset time-series model is also relatively simple to run, in that there is no special requirement for data preparation. When multiple sets of time series have a natural structure of data hierarchy, the formation of data requires very little special manipulation. Finally, the multilevel multiset time-series model is relatively simple to interpret. The (qualitative) description of growth patterns and program effects is not strange to anyone who can apply (ANOVA) factorial design for data analysis. The quantitative interpretation can be aided by simple algebraic manipulations to unfold interactive growth patterns and to compare program effects on growth patterns. Overall, this analytical platform can inspire researchers to pursue far more advanced research designs to address complex developmental processes in reality.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study is supported by a grant from the U.S. Department of Education's school leadership program.

Supplemental Material

The online supplementary materials are available at <http://journals.sagepub.com/doi/suppl/10.1177/0146621616686058>.

Notes

1. Some researchers may prefer to ignore the within-day repeated measures and simply count time points continuously each time a measurement is made, resulting in only one time series. The choice of measurement options depends on the research questions.
2. Uncorrelated residuals are typically not assumed in time-series modeling. However, time-series analysis in any form of regression inherits the assumption of uncorrelated residuals. The benefit of this

assumption to the authors' specific model is that it allows them to "shift" more variance to the relationship between the two time series (i.e., how the pattern of one time series is related to that of the other time series) through the specification of uncorrelated residuals.

3. Descriptive statistics on school-level outcome measure and school contextual variables by calendar years and grade levels are available in the Online Appendix A.
4. HLM syntax is available in the Online Appendix B.
5. Apart from the number of outcome measures modeled in time-series analysis that separates Ma and Ma (2005) from the multilevel multiset time-series model of this study (see discussion earlier), the analytical "philosophy" differs. Ma and Ma (2005) take the traditional latent variable approach to examine multiple time series based on covariance between their random effects. Therefore, multiple time series are discussed in terms of association (e.g., persons or groups who grow faster along one time series also grow faster along the other time series). The multilevel multiset time-series model of this study takes a different approach that directly manipulates covariates. Therefore, multiple time series are discussed in terms of interaction (e.g., how fast persons or groups grow along one time series differs between time points of the other time series).

References

- Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.
- Chou, Y. (1975). *Statistical analysis*. Toronto, Ontario: Holt, Rinehart and Winston of Canada.
- Enders, W. (2010). *Applied econometric time series* (3rd ed.). New York, NY: John Wiley.
- Goldstein, H., Healy, M. J. R., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures. *Statistics in Medicine*, 13, 1643-1655.
- Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: John Wiley.
- Holden, J. E., Kelley, K., & Agarwal, R. (2008). Analyzing change: A primer on multilevel models with applications to nephrology. *American Journal of Nephrology*, 28, 792-801.
- Ma, L., & Ma, X. (2005). Estimating correlates of growth between mathematics and science achievement via a multivariate multilevel design with latent variables. *Studies in Educational Evaluation*, 31, 79-98.
- Michigan Department of Education. (2006). *Michigan educational assessment program technical report 2005-2006*. Lansing: Author.
- Muthén, B. (1997). Latent variable growth modeling with multilevel data. In M. Berkane (Ed.), *Latent variable modeling and applications to causality* (pp. 149-161). New York, NY: Springer-Verlag.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: SAGE.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2013). *HLM 7.01 for windows*. Skokie, IL: Scientific Software International.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York, NY: Oxford University Press.