# Issues in Aggregating Multiple Time Series: Illustration through an AR(1) Model

Zhenqiu (Laura) Lu and Zhiyong Zhang

**Abstract** Intra-individual variation is time dependent variation within a single participant's time series. When data are collected from more than one subject, methods developed for single subject intra-individual relationship may not fully work and laws governing the inter-individual relationship may not apply to intra-individual relationship. There are relatively few methods to pool multiple time series for statistical data analysis. This article aims to investigate empirically several methods for pooling time series data and to address related issues through an AR(1) model. Specifically, multiple time series are formulated, pooling estimation methods are derived and compared, simulation studies results are summarized, and related practical issues are addressed.

**Key words:** Time Series Analysis, First-order Autoregressive Model, Pooling Multiple Subjects, Longitudinal Analysis, Maximum Likelihood Estimation.

## 1 Introduction

The variation analysis in psychological, social, and behavioral researches has many ramifications. Among them two main branches are inter-individual variation and intra-individual variation. The inter-individual variation is the variation between individuals, and also widely known as the analysis of cross-sectional data in many researches. The intra-individual variation is time dependent variation within a single participant's time series. It is also known as the analysis of time series data or P-technique in Cattell's (1952) data-box (Cattell, 1952). In this type of study,

---

Zhenqiu (Laura) Lu
University of Georgia, Athens, GA 30602, e-mail: zlu@uga.edu

Zhiyong Zhang
University of Notre Dame, Notre Dame, IN 46556, e-mail: zhangzhiyong@nd.edu

usually one subject is measured repeatedly and data on the variables of interest are collected for a large number of occasions. Data collected in this way do not have inter-individual differences since there is only one subject involved, but they can reflect the change across occasions. Intra-individual analysis has become popular advanced by Nesselroade, Molenaar, and colleagues and many methods are available for single time series analysis (e.g., Cattell, Cattell, & Rhymer, 1947; Molenaar, 1985; Nesselroade & Molenaar, 2003). (**add Zhang & Nesselroade, 2007: Zhang, Z., & Nesselroade J. R. (2007). Bayesian estimation of categorical dynamic factor models. Multivariate Behavioral Research, 42(4), 729-756.**)

In this article, attention will be drawn to multiple-subject intra-individual variation analysis. In many researches intra-individual relationship data are collected from more than one subject. When multiple subjects are involved, methods developed for single-subject intra-individual relationship may not fully work. In addition, laws governing the inter-individual relationship may not apply to intra-individual relationship (e.g., Molenaar, 2004; Nesselroade & Ram, 2004). **Check the references for Molenaar, they should be the same** Currently, there are relatively few methods in the literature on pooling multiple time series for data analysis (e.g., Cattell & Scheier, 1961; Daly, Bath, & Nesselroade, 1974; P. C. M. Molenaar, Huizenga, & Nesselroade, 2003; Nesselroade & Molenaar, 1999). Therefore, this study aims to evaluate methods for pooling time series and to address related issues through a first-order autoregressive (AR(1)) model. Specifically, we focus on five estimation methods for combining multiple time series: pooling conditional likelihood estimation, pooling exact likelihood estimation, connecting data conditional likelihood, connecting data exact likelihood, and multivariate analysis.

This rest of the article is organized as follows. In Section 2, we propose several methods for pooling time series data. Single time series and multiple time series are first described and formulated in terms of the AR(1) model. Then different estimation methods for multiple time series are introduced and derived. In Section 3, we evaluate the performance of different methods through a simulation study. We first present the simulation design and implementation. Then, we provide the simulation results. Section 4 discusses the implication and future direction of the study.

## 2 Models

In this section, we introduce the time series models and the corresponding estimation methods. We first focus on the single time series analysis, and then extend to multiple time series analysis.

## 2.1 Single Time Series AR(1) Model and Estimation Method

The simplest and the most popular single time series model to describe the intra-individual relationship is the first-order autoregressive model, also known as AR(1). It can be expressed as follows.

$$\begin{aligned} &y_1 : \text{the initial value} \\ &y_t = \mu + \alpha y_{t-1} + z_t \quad (t > 1) \\ &z_t \sim i.i.d.\ N(0, \phi) \end{aligned} \tag{1}$$

where $y_t$ is the observed value at time point $t$, $\alpha$ is the autoregressive coefficient at lag 1, $\mu$ is an unknown parameter related to the mean of $y$, and $z$ is a random shock or a white noise, which is assumed to follow a normal distribution with mean 0 and unknown variance $\phi$. The joint density function of $y_t$ $(t = 1, ..., T)$ given the AR(1) model in Eq (2) is

$$p(\mathbf{y}|\mu, \alpha, \phi) = p(y_1, y_2, ..., y_T|\mu, \alpha, \phi) = p(y_1) \prod_{t=2}^{T} p(y_t|y_{t-1}).$$

The path diagram of the AR(1) model can be portrayed as in Fig. (1).

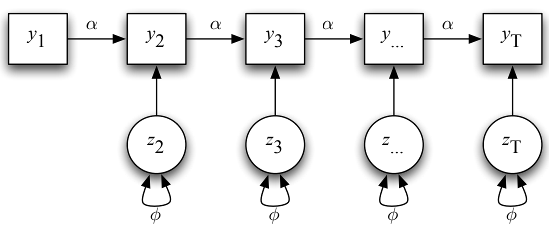


Fig. 1: The AR(1) Model

There are two maximum likelihood estimation (MLE) methods for AR(1): the conditional MLE and the exact MLE. The former treats the initial value $y_1$ as deterministic and focuses only on the conditional distribution in Eq. (2). By maximizing the conditional likelihood function without $y_1$, this estimation method makes the analysis relatively easy. Parameters are obtained by maximizing the conditional likelihood function as follows:

$$L_c(\alpha, \mu, \phi|\mathbf{y}) = \prod_{t=2}^{T} p(y_t|y_{t-1}, \alpha, \mu, \phi) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\phi}} \exp\left[-\frac{(y_t - \mu - \alpha y_{t-1})^2}{2\phi}\right].$$

Instead of treating $y_1$ as deterministic, the latter (exact MLE) estimation treats $y_1$ as random. It maximizes the exact likelihood function which includes the distribution of $y_1$. When exact MLE is adopted, a stationarity procedure is required. By assuming $|\alpha| < 1$, the covariance of $y_t$ in AR(1) is shown stationary, and we have

$$y_1 \sim N(\frac{\mu}{1-\alpha}, \frac{\phi}{1-\alpha^2}),$$
$$y_t|y_{t-1} \sim N(\mu + \alpha y_{t-1}, \phi), (t > 1).$$

With this assumption, the exact likelihood function of $\mathbf{y}$ is

$$L_e(\alpha, \mu, \phi|\mathbf{y}) = p(y_1|\alpha, \mu, \phi)L_c(\alpha, \mu, \phi|\mathbf{y})$$
$$= \frac{1}{\sqrt{2\pi(\frac{\phi}{1-\alpha^2})}} \exp\left[-\frac{(y_1 - \frac{\mu}{1-\alpha})^2}{2(\frac{\phi}{1-\alpha^2})}\right] \left\{\prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\phi}} \exp\left[-\frac{(y_t - \mu - \alpha y_{t-1})^2}{2\phi}\right]\right\}.$$

## 2.2 Multiple Time Series and Estimation

The AR(1) model introduced above is for single time series analysis. In reality, however, multiple time series are prevalent. There are many sources of multiple time series such as multiple-subject time series, in which data are collected from multiple similar subjects, multivariate time series, in which multiple dependent variables are collected from the same subjects and multiple-session time series, in which data are collected at different sessions from the same subjects. Suppose there are $N$ individuals (or sessions or other forms of series). Without loss of generality, we assume that each individual has $T$ observations collected from different time points, so totally there are $NT$ observations. Note that individuals can have different time series lengths. The model for individual $i$ at time point $t$ is expressed as follows.

$$y_{it} = \mu + \alpha y_{i(t-1)} + z_{it}, \ \ (i = 1, ..., N; \ t = 2, ..., T)$$

where $z_{it} \sim i.i.d. N(0, \phi)$. We focus on three strategies to analysis such multiple time series: pooling likelihoods together, connecting data directly, and using multivariate data analysis.

### 2.2.1 Pooling Likelihoods

There are various methods to estimate parameters $\mu$, $\alpha$ and $\phi$ in multiple time series analysis. They can be estimated by pooling the likelihood functions across all individuals. Based on different forms of the likelihood function, there are pooled conditional likelihood MLE and pooled exact likelihood MLE.

In the following analysis, we assume the $N$ individuals are from one population and have the same parameters $\mu$, $\alpha$ and $\phi$. But these assumptions can be relaxed. Pooled likelihood methods allow $\mu$, $\alpha$, and $\phi$ to vary and estimate.

The pooled conditional likelihood of an AR(1) model for $N$ individuals is

$$
\begin{aligned}
L_c(\alpha,\mu,\phi|\mathbf{y}) &= \prod_{i=1}^{N}\prod_{t=2}^{T} p(y_{it}|y_{i(t-1)},\alpha,\mu,\phi) \\
&= \prod_{i=1}^{N}\prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\phi}} \exp\left[-\frac{(y_{it}-\mu-\alpha\, y_{i(t-1)})^2}{2\phi}\right]. \qquad (2)
\end{aligned}
$$

To obtain the MLE of $\mu$, $\alpha$ and $\phi$, we take the first derivative with respect to each parameter and set it to 0 and make their corresponding second derivatives negative at $\hat{\theta} = (\hat{\mu},\hat{\alpha},\hat{\phi})$. For pooling conditional likelihoods, we have the closed form estimates for the parameters such that,

$$
\begin{aligned}
\hat{\mu} &= \frac{(\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i(t-1)}^2)(\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it}) - (\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i(t-1)})(\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i(t-1)}y_{it})}{N(T-1)\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i(t-1)}^2 - (\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i(t-1)})^2}, \\
\hat{\alpha} &= \frac{N(T-1)(\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i(t-1)}y_{it}) - (\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i(t-1)})(\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it})}{N(T-1)\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i(t-1)}^2 - (\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i(t-1)})^2}, \\
\hat{\phi} &= \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}(y_{it}-\hat{\mu}-\hat{\alpha}\, y_{i(t-1)})^2. \qquad (3)
\end{aligned}
$$

Parameters can also be estimated by maximizing the pooled exact likelihood function including the distribution of the initial value.

$$
L_e(\alpha,\mu,\phi|\mathbf{y}) = \prod_{i=1}^{N}\left[p(y_{i1}|\alpha,\mu,\phi)\prod_{t=2}^{T} p(y_{it}|y_{i(t-1)},\alpha,\mu,\phi)\right]
$$

Unfortunately, there is no analytic solution for $\hat{\theta} = (\hat{\mu},\hat{\alpha},\hat{\phi})$ in terms of $\{y_{it}\}, (1 \le i \le N, 1 \le t \le T)$. Instead, we have to adopt iterative algorithms to obtain numerical solutions. This method is recommended when participants are almost identical, or data are from multiple sessions of the same participant.

### 2.2.2 Connecting Data

Intuitively, one can also analyze multiple series data data by connecting all time series from multiple subjects as from a single subject. Fig. 2 shows the connected series from four individuals. This method assumes that the connecting points do not matter and, therefore, $y_{iT}$ and $y_{(i+1)1}$ can be connected as from a sequence. Also, it assumes all series share the same $\mu$, $\alpha$ and $\phi$. The assumption of equal $\mu$ can be relaxed, e.g., through centering means.

Based on different likelihood functions, there are conditional MLE and exact MLE for connected data. Let $j$ $(j = 1,2,\ldots,NT)$ be the new subscript for the connected series. The conditional likelihood function is

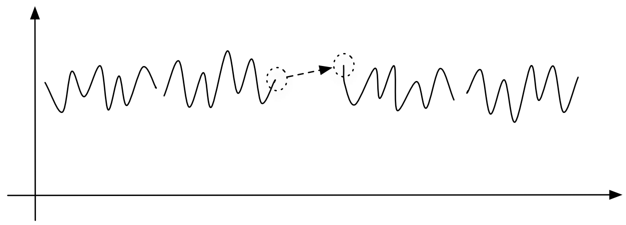

Fig. 2: Connecting data from multiple subjects

$$L(\alpha,\mu,\phi|\mathbf{y}) = \prod_{j=2}^{NT} p(y_j|\alpha,\mu,\phi) = \prod_{j=2}^{NT} \frac{1}{\sqrt{2\pi\phi}} \exp\left[-\frac{(y_j-\mu-\alpha y_{j-1})^2}{2\phi}\right].$$

The closed form parameter estimates can be easily obtained as

$$\begin{aligned}
\hat{\mu} &= \frac{(\sum_{j=2}^{NT} y_{j-1}^2)(\sum_{t=2}^{NT} y_j) - (\sum_{j=2}^{NT} y_{j-1})(\sum_{j=2}^{NT} y_{j-1}y_j)}{(NT-1)\sum_{j=2}^{NT} y_{j-1}^2 - (\sum_{j=2}^{NT} y_{j-1})^2},\\
\hat{\alpha} &= \frac{(NT-1)(\sum_{j=2}^{NT} y_{j-1}y_j) - (\sum_{j=2}^{NT} y_{j-1})(\sum_{j=2}^{NT} y_j)}{(NT-1)\sum_{j=2}^{NT} y_{j-1}^2 - (\sum_{j=2}^{NT} y_{j-1})^2},\\
\hat{\phi} &= \frac{1}{NT-1}\sum_{j=2}^{NT}(y_j-\hat{\mu}-\hat{\alpha}y_{j-1})^2.
\end{aligned}$$

For exact MLE, it requires a stationary AR(1) model. The exact likelihood functions of the connected data is

$$\begin{aligned}
&L(\alpha,\mu,\phi|\mathbf{y}) = p(y_1|\alpha,\mu,\phi)\prod_{j=2}^{NT} p(y_j|\alpha,\mu,\phi)\\
&= \frac{1}{\sqrt{2\pi(\frac{\phi}{1-\alpha^2})}} \exp\left[-\frac{(y_1-\frac{\mu}{1-\alpha})^2}{2\,(\frac{\phi}{1-\alpha^2})}\right]\left\{\prod_{j=2}^{NT}\frac{1}{\sqrt{2\pi\phi}}\exp\left[-\frac{(y_j-\mu-\alpha y_{j-1})^2}{2\phi}\right]\right\}.
\end{aligned}$$

Again, there is no analytical solutions for $\hat{\theta} = (\hat{\mu},\hat{\alpha},\hat{\phi})$.

### 2.2.3 Multivariate Time Series Analysis

Multivariate analysis is another alternative approach to multiple time series analysis. It views each time series as a $N$-dimensional multivariate variable. It also allows subject dependence. This method is relatively difficult to use compared the other two methods. It requires that the data are collected at the same time points, and certainly, all individuals have the same time series length.

Let $\mathbf{Y}_t$, $\mathbf{Y}_{t-1}$ and $\mathbf{z}_t$ be three $N$-dimensional column vectors, $\mathbf{Y}_t' = (y_{1t}, y_{2t}, ..., y_{Nt})$, $\mathbf{Y}_{t-1}' = (y_{1(t-1)}, y_{2(t-1)}, ..., y_{N(t-1)})$, and $\mathbf{z}_t' = (z_{1t}, z_{2t}, ..., z_{Nt})$, and $\beta$ be a $(2 \times 1)$ vector including parameters $\mu$ and $\alpha$. At time point $t$, the multiple time series can be expressed as $\mathbf{Y}_t = (\mathbf{1}, \mathbf{Y}_{t-1})\beta + \mathbf{z}_t$, which is

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{pmatrix} = \begin{pmatrix} 1 & y_{1(t-1)} \\ 1 & y_{2(t-1)} \\ \vdots & \vdots \\ 1 & y_{N(t-1)} \end{pmatrix} \begin{pmatrix} \mu \\ \alpha \end{pmatrix} + \begin{pmatrix} z_{1t} \\ z_{2t} \\ \vdots \\ z_{Nt} \end{pmatrix}.$$

If we combine all time points $t$ from 2 to $T$, then we have the least-squares (LS) estimate of $\beta$,

$$\hat{\beta} = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} N(T-1) & \sum_{t=2}^{T}\sum_{i=1}^{N} y_{i(t-1)} \\ \sum_{t=2}^{T}\sum_{i=1}^{N} y_{i(t-1)} & \sum_{t=2}^{T}\sum_{i=1}^{N} y_{i(t-1)}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=2}^{T}\sum_{i=1}^{N} y_{it} \\ \sum_{t=2}^{T}\sum_{i=1}^{N} y_{i(t-1)} y_{it} \end{bmatrix}.$$

Note that with the normality assumption, maximizing (2) with respect to $\mu$ and $\alpha$ is equivalent to minimizing

$$\sum_{i=1}^{N}\sum_{t=2}^{T} (y_{it} - \mu - \alpha y_{i(t-1)})^2$$

with respect to $\mu$ and $\alpha$. Therefore, the LS solution for $\theta = (\mu, \alpha, \phi)$ is exactly the same as the pooled conditional likelihood MLE solution as shown in (3).

## 3 Simulation Study

To investigate the performance of different pooling methods on estimating multiple times series, we conducted a simulation study.

### *3.1 Design and Implementation*

First, multiple time series under various conditions were generated from an AR(1) model. The true population parameter values were $\mu = 0$, $\alpha = 0.5$, and $\phi = 0.25$. As a main difference among various methods on parameter estimation is the treatment of the initial value $y_1$, we generated data under three conditions: (i) with a fixed $y_1$ at 0, (ii) with a random $y_1$ drew from $N(0, \phi)$, and (iii) with a random $y_1$ drew from $N(\frac{\mu}{1-\alpha}, \frac{\phi}{1-\alpha^2})$. Other conditions included the number of time points, $T = (5, 10, 15, 20, 30)$, and the number of subjects, $N = (10, 20, 30, 40, 50)$. In total, there are $3 \times 5 \times 5 = 75$ different conditions in the simulation. For each condition, 1,000 replications of data were generated and analyzed using the different pooling methods.

Second, model parameters ($\mu$, $\alpha$, and $\phi$) were estimated by using different multiple time series estimation methods. As the multivariate LS estimation method yields the same results as the pooled likelihood conditional MLE when data are normally distributed, we adopted four estimation methods in this study: pooled likelihood conditional MLE, pooled likelihood exact MLE, pooled data conditional MLE, and pooled data exact MLE. As there is no analytical solutions for exact MLE, iterative algorithms were employed to obtain numerical solutions.

Finally, results were summarized across all simulation replications. For each parameter estimate, $Est.$ is the average estimate across 1,000 replications; the absolute bias ($Bias.abs$) was calculated as the absolute value of the difference between the estimated value and its true value; the relative bias ($Bias.rel$) of the estimate was the ratio of the absolute bias to the true value; the empirical s.e. ($SE.emp$) was calculated as the standard deviation of the parameter estimates across 1,000 replications; the average s.e. ($SE.avg$) was the average standard error across 1,000 replications; the mean square error ($MSE$) of each parameter estimate was calculated as $MSE = Bias.abs^2 + SE.emp^2$; and the $Cover$ is the coverage rate.

In the simulation study, we used the R language to generate data, estimate parameters, and summarize results. Complete R code is available from the first author of the article.

### 3.2 Results

Totally, there were $3 \times 5 \times 5 = 75$ simulation result tables. Due to the limited space, we only showed part of the results here. Tables 1 to 6 summarized the estimates from different estimation methods: pooling likelihood (P.L.) conditional MLE, pooling likelihood (P.L.) exact MLE, connecting data (C.D.) conditional MLE, and connecting data (C.D.) exact MLE.

Table 1 : Simulation results for the condition of fixed $y_1 = 0$, $N = 10$ individuals, $T = 5$ observations (1000 replications)

| | | | True[a] | Est.[b] | Bias.abs[c] | Bias.rel[d] | SE.emp[e] | SE.avg[f] | MSE[g] | Cover[h] |
|---|---|---|---|---|---|---|---|---|---|---|
| P.L.[i] | Exact[j] | $\mu$ | 0 | -0.0027 | 0.0027 | 0.0027 | 0.0640 | 0.0574 | 0.0041 | 0.9440 |
| | | $\alpha$ | 0.5 | 0.3567 | 0.1433 | 0.2867 | 0.1399 | 0.1318 | 0.0401 | 0.8250 |
| | | $\phi$ | 0.25 | 0.1942 | 0.0558 | 0.2234 | 0.0466 | 0.0392 | 0.0053 | 0.5990 |
| | Cond.[k] | $\mu$ | 0 | -0.0039 | 0.0039 | 0.0039 | 0.0884 | 0.0782 | 0.0078 | 0.9200 |
| | | $\alpha$ | 0.5 | 0.4534 | 0.0466 | 0.0932 | 0.1794 | 0.1693 | 0.0344 | 0.9330 |
| | | $\phi$ | 0.25 | 0.2372 | 0.0128 | 0.0510 | 0.0572 | 0.0531 | 0.0034 | 0.8750 |
| C.D.[l] | Exact | $\mu$ | 0 | -0.0032 | 0.0032 | 0.0032 | 0.0708 | 0.0641 | 0.0050 | 0.9370 |
| | | $\alpha$ | 0.5 | 0.3245 | 0.1755 | 0.3509 | 0.1325 | 0.1325 | 0.0484 | 0.7630 |
| | | $\phi$ | 0.25 | 0.2039 | 0.0461 | 0.1845 | 0.0501 | 0.0408 | 0.0046 | 0.6480 |
| | Cond. | $\mu$ | 0 | -0.0033 | 0.0033 | 0.0033 | 0.0729 | 0.0660 | 0.0053 | 0.9350 |
| | | $\alpha$ | 0.5 | 0.3310 | 0.1690 | 0.3381 | 0.1353 | 0.1355 | 0.0469 | 0.7790 |
| | | $\phi$ | 0.25 | 0.2078 | 0.0422 | 0.1686 | 0.0511 | 0.0420 | 0.0044 | 0.6830 |

[a] The true value of the corresponding parameter.

[b] The average of the estimate of the corresponding parameter across 1000 replications.

[c] The absolute bias of the estimate.

[d] The relative bias of the estimate.

[e] The empirical s.e. across 1000 replications.

[f] The average of the s.e. obtained from the model.

[g] The mean square error of the estimate, $MSE = Bias.abs^2 + SE.emp^2$.

[h] The coverage probability of the estimate.

[i] The method of pooling likelihood functions.

[j] Parameters are estimated by maximizing the exact likelihood function of the original data.

[k] Parameters are estimated by maximizing the conditional likelihood function of the original data.

[l] The method of connecting data.

Table 2: Simulation results for the condition of fixed $y_1 = 0$, $N = 50$ individuals, $T = 30$ observations (1000 replications)

| | | | True | Est. | Bias.abs | Bias.rel | SE.emp | SE.avg | MSE | Cover |
|---|---|---|---|---|---|---|---|---|---|---|
| P.L. | Exact | $\mu$ | 0 | 0.0002 | 0.0002 | 0.0002 | 0.0126 | 0.0123 | 0.0002 | 0.9450 |
| | | $\alpha$ | 0.5 | 0.4810 | 0.0190 | 0.0380 | 0.0225 | 0.0223 | 0.0009 | 0.8680 |
| | | $\phi$ | 0.25 | 0.2413 | 0.0087 | 0.0348 | 0.0087 | 0.0088 | 0.0002 | 0.8280 |
| | Cond. | $\mu$ | 0 | 0.0002 | 0.0002 | 0.0002 | 0.0134 | 0.0131 | 0.0002 | 0.9450 |
| | | $\alpha$ | 0.5 | 0.4974 | 0.0026 | 0.0052 | 0.0233 | 0.0233 | 0.0005 | 0.9590 |
| | | $\phi$ | 0.25 | 0.2495 | 0.0005 | 0.0019 | 0.0090 | 0.0093 | 0.0001 | 0.9560 |
| C.D. | Exact | $\mu$ | 0 | 0.0002 | 0.0002 | 0.0002 | 0.0131 | 0.0128 | 0.0002 | 0.9410 |
| | | $\alpha$ | 0.5 | 0.4802 | 0.0198 | 0.0397 | 0.0227 | 0.0226 | 0.0009 | 0.8580 |
| | | $\phi$ | 0.25 | 0.2438 | 0.0062 | 0.0248 | 0.0088 | 0.0089 | 0.0001 | 0.8720 |
| | Cond. | $\mu$ | 0 | 0.0002 | 0.0002 | 0.0002 | 0.0131 | 0.0128 | 0.0002 | 0.9410 |
| | | $\alpha$ | 0.5 | 0.4805 | 0.0195 | 0.0391 | 0.0227 | 0.0227 | 0.0009 | 0.8600 |
| | | $\phi$ | 0.25 | 0.2440 | 0.0060 | 0.0241 | 0.0088 | 0.0089 | 0.0001 | 0.8780 |

Note: With the same notations as in Table 1.

Table 3: Simulation results for the condition of $y_1 \sim N(0,\phi)$, $N = 10$ individuals, $T = 5$ observations (1000 replications)

| | | | True | Est. | Bias.abs | Bias.rel | SE.emp | SE.avg | MSE | Cover |
|---|---|---|---|---|---|---|---|---|---|---|
| P.L. | Exact | $\mu$ | 0 | -0.0007 | 0.0007 | 0.0007 | 0.0657 | 0.0604 | 0.0043 | 0.9470 |
| | | $\alpha$ | 0.5 | 0.4387 | 0.0613 | 0.1225 | 0.1345 | 0.1318 | 0.0218 | 0.9370 |
| | | $\phi$ | 0.25 | 0.2301 | 0.0199 | 0.0795 | 0.0481 | 0.0466 | 0.0027 | 0.8550 |
| | Cond. | $\mu$ | 0 | -0.0008 | 0.0008 | 0.0008 | 0.0874 | 0.0786 | 0.0076 | 0.9230 |
| | | $\alpha$ | 0.5 | 0.4638 | 0.0362 | 0.0724 | 0.1491 | 0.1443 | 0.0235 | 0.9310 |
| | | $\phi$ | 0.25 | 0.2382 | 0.0118 | 0.0473 | 0.0550 | 0.0533 | 0.0032 | 0.8890 |
| C.D. | Exact | $\mu$ | 0 | -0.0009 | 0.0009 | 0.0009 | 0.0753 | 0.0713 | 0.0057 | 0.9450 |
| | | $\alpha$ | 0.5 | 0.3613 | 0.1387 | 0.2775 | 0.1322 | 0.1312 | 0.0367 | 0.8450 |
| | | $\phi$ | 0.25 | 0.2512 | 0.0012 | 0.0048 | 0.0535 | 0.0503 | 0.0029 | 0.9140 |
| | Cond. | $\mu$ | 0 | -0.0004 | 0.0004 | 0.0004 | 0.0772 | 0.0729 | 0.0060 | 0.9370 |
| | | $\alpha$ | 0.5 | 0.3621 | 0.1379 | 0.2757 | 0.1331 | 0.1319 | 0.0367 | 0.8470 |
| | | $\phi$ | 0.25 | 0.2517 | 0.0017 | 0.0069 | 0.0537 | 0.0509 | 0.0029 | 0.9170 |

Note: With the same notations as in Table 1.

Table 4: Simulation results for the condition of $y_1 \sim N(0,\phi)$, $N = 50$ individuals, $T = 30$ observations (1000 replications)

| | | | True | Est. | Bias.abs | Bias.rel | SE.emp | SE.avg | MSE | Cover |
|---|---|---|---|---|---|---|---|---|---|---|
| P.L. | Exact | $\mu$ | 0 | 0.0007 | 0.0007 | 0.0007 | 0.0125 | 0.0125 | 0.0002 | 0.9490 |
| | | $\alpha$ | 0.5 | 0.4949 | 0.0051 | 0.0103 | 0.0236 | 0.0225 | 0.0006 | 0.9320 |
| | | $\phi$ | 0.25 | 0.2474 | 0.0026 | 0.0103 | 0.0092 | 0.0090 | 0.0001 | 0.9280 |
| | Cond. | $\mu$ | 0 | 0.0009 | 0.0009 | 0.0009 | 0.0132 | 0.0131 | 0.0002 | 0.9480 |
| | | $\alpha$ | 0.5 | 0.4989 | 0.0011 | 0.0022 | 0.0241 | 0.0229 | 0.0006 | 0.9340 |
| | | $\phi$ | 0.25 | 0.2494 | 0.0006 | 0.0022 | 0.0094 | 0.0093 | 0.0001 | 0.9390 |
| C.D. | Exact | $\mu$ | 0 | 0.0007 | 0.0007 | 0.0007 | 0.0130 | 0.0130 | 0.0002 | 0.9490 |
| | | $\alpha$ | 0.5 | 0.4822 | 0.0178 | 0.0356 | 0.0238 | 0.0226 | 0.0009 | 0.8650 |
| | | $\phi$ | 0.25 | 0.2521 | 0.0021 | 0.0084 | 0.0094 | 0.0092 | 0.0001 | 0.9300 |
| | Cond. | $\mu$ | 0 | 0.0007 | 0.0007 | 0.0007 | 0.0131 | 0.0130 | 0.0002 | 0.9480 |
| | | $\alpha$ | 0.5 | 0.4823 | 0.0177 | 0.0355 | 0.0238 | 0.0226 | 0.0009 | 0.8680 |
| | | $\phi$ | 0.25 | 0.2521 | 0.0021 | 0.0086 | 0.0094 | 0.0092 | 0.0001 | 0.9280 |

Note: With the same notations as in Table 1.

Table 5: Simulation results for the condition of $y_1 \sim N(\frac{\mu}{1-\alpha}, \frac{\phi}{1-\alpha^2})$, $N = 10$ individuals, $T = 5$ observations (1000 replications)

| | | | True | Est. | Bias.abs | Bias.rel | SE.emp | SE.avg | MSE | Cover |
|---|---|---|---|---|---|---|---|---|---|---|
| P.L. | Exact | $\mu$ | 0 | -0.0006 | 0.0006 | 0.0006 | 0.0692 | 0.0616 | 0.0048 | 0.9400 |
| | | $\alpha$ | 0.5 | 0.4561 | 0.0439 | 0.0879 | 0.1384 | 0.1325 | 0.0211 | 0.9380 |
| | | $\phi$ | 0.25 | 0.2417 | 0.0083 | 0.0333 | 0.0501 | 0.0490 | 0.0026 | 0.9010 |
| | Cond. | $\mu$ | 0 | -0.0011 | 0.0011 | 0.0011 | 0.0896 | 0.0789 | 0.0080 | 0.9120 |
| | | $\alpha$ | 0.5 | 0.4603 | 0.0397 | 0.0794 | 0.1505 | 0.1392 | 0.0242 | 0.9250 |
| | | $\phi$ | 0.25 | 0.2382 | 0.0118 | 0.0470 | 0.0551 | 0.0533 | 0.0032 | 0.8920 |
| C.D. | Exact | $\mu$ | 0 | -0.0004 | 0.0004 | 0.0004 | 0.0800 | 0.0736 | 0.0064 | 0.9310 |
| | | $\alpha$ | 0.5 | 0.3666 | 0.1334 | 0.2669 | 0.1324 | 0.1311 | 0.0353 | 0.8500 |
| | | $\phi$ | 0.25 | 0.2670 | 0.0170 | 0.0680 | 0.0586 | 0.0534 | 0.0037 | 0.9230 |
| | Cond. | $\mu$ | 0 | -0.0011 | 0.0011 | 0.0011 | 0.0820 | 0.0752 | 0.0067 | 0.9290 |
| | | $\alpha$ | 0.5 | 0.3662 | 0.1338 | 0.2675 | 0.1327 | 0.1315 | 0.0355 | 0.8490 |
| | | $\phi$ | 0.25 | 0.2668 | 0.0168 | 0.0672 | 0.0593 | 0.0539 | 0.0038 | 0.9190 |

Note: With the same notations as in Table 1.

Table 6: Simulation results for the condition of $y_1 \sim N(\frac{\mu}{1-\alpha}, \frac{\phi}{1-\alpha^2})$, $N = 50$ individuals, $T = 30$ observations (1000 replications)

| | | | True | Est. | Bias.abs | Bias.rel | SE.emp | SE.avg | MSE | Cover |
|---|---|---|---|---|---|---|---|---|---|---|
| P.L. | Exact | $\mu$ | 0 | -0.0002 | 0.0002 | 0.0002 | 0.0127 | 0.0125 | 0.0002 | 0.9540 |
| | | $\alpha$ | 0.5 | 0.4998 | 0.0002 | 0.0003 | 0.0220 | 0.0225 | 0.0005 | 0.9470 |
| | | $\phi$ | 0.25 | 0.2495 | 0.0005 | 0.0021 | 0.0091 | 0.0091 | 0.0001 | 0.9570 |
| | Cond. | $\mu$ | 0 | -0.0001 | 0.0001 | 0.0001 | 0.0134 | 0.0131 | 0.0002 | 0.9540 |
| | | $\alpha$ | 0.5 | 0.4997 | 0.0003 | 0.0007 | 0.0222 | 0.0227 | 0.0005 | 0.9440 |
| | | $\phi$ | 0.25 | 0.2494 | 0.0006 | 0.0024 | 0.0093 | 0.0093 | 0.0001 | 0.9530 |
| C.D. | Exact | $\mu$ | 0 | -0.0002 | 0.0002 | 0.0002 | 0.0132 | 0.0130 | 0.0002 | 0.9560 |
| | | $\alpha$ | 0.5 | 0.4833 | 0.0167 | 0.0334 | 0.0222 | 0.0226 | 0.0008 | 0.8900 |
| | | $\phi$ | 0.25 | 0.2549 | 0.0049 | 0.0195 | 0.0095 | 0.0093 | 0.0001 | 0.9260 |
| | Cond. | $\mu$ | 0 | -0.0002 | 0.0002 | 0.0002 | 0.0133 | 0.0131 | 0.0002 | 0.9560 |
| | | $\alpha$ | 0.5 | 0.4833 | 0.0167 | 0.0334 | 0.0221 | 0.0226 | 0.0008 | 0.8900 |
| | | $\phi$ | 0.25 | 0.2549 | 0.0049 | 0.0195 | 0.0095 | 0.0093 | 0.0001 | 0.9260 |

Note: With the same notations as in Table 1.

### 3.2.1 Findings from the Simulation

Based on the simulation results, we can draw the following conclusions.

(1) In general, the $Bias.abs$, $Bias.rel$, $SE.emp$, $SE.avg$, and $MSE$ values for large $T$ or large $N$ (e.g., see Tables 2, 4, 6) were smaller than those for small $T$ or small $N$ (e.g., see Tables 1, 3, 5). The coverage rates for large $T$ or large $N$ were more close to 0.95 than those for small $T$ or small $N$. Both indicated that for these four estimation methods, although the data sets had different initial values, (a) the longer the time series, the more accurate the estimate, and (b) the more participants, the more accurate the estimates.

(2) By comparing bias statistics (such as $Bias.abs$ and $Bias.rel$) and coverage rates (*Cover*) across all tables, in most cases the pooled likelihood methods outperformed the connecting data methods. For the situation of large $T$ and small $N$, connecting data methods also estimated well.

(3) The pooled likelihood conditional MLE performed best, especially on the recovery of the autoregressive coefficient $\alpha$ (see tables 1-4), except when initial value $y_1 \sim N(\frac{\mu}{1-\alpha}, \frac{\phi}{1-\alpha^2})$ both the pooled likelihood conditional MLE and the pooled likelihood exact MLE performed well (e.g., see tables 5 and 6).

(4) For large $T$ and small $N$, conditional MLE performed similarly as the exact MLE.

(5) For small $T$ and large $N$, exact MLE is more efficient but may have large bias depending on the state of $y_1$ and stationarity of time series.

(6) The parameter coverages for data with random initial values were closer to 0.95 than those for fixed initial value $y_1 = 0$. Among two types of random initial values, the one drew from the stationary distribution $y_1 \sim N(\frac{\mu}{1-\alpha}, \frac{\phi}{1-\alpha^2})$ performed better than the other distribution.

(7) For the data with fixed initial values $y_1 = 0$, the parameter coverage rates were low, especially for $\phi$ (see Table 1).

## 4 Discussion, and Future Directions

To compare the different methods, we contrast them in Table 4. **You might as well describe the difference in words here**.

Table 7: A comparison of multiple time series estimation methods

| Pooling Method | Comparision |
|---|---|
| Connecting data | Easy to use; |
| | Allow $\mu$ to vary through centering; |
| | Can be used for large T and small N situation. |
| Pooling likelihood | Easy to use; |
| | Allow $\mu$, $\alpha$, $\phi$ to vary and estimate; |
| | For large T and small N, Conditional MLE $\approx$ Exact MLE; |
| | For small T and large N, Exact MLE is more efficient but may have large bias depending on the state of $y_1$ and stationarity of time series. |
| Multivariate method | Relatively difficult to use; |
| | The same time series length; |
| | Data measured at the same time; |
| | Allow subject dependence. |

### 4.1 Future Directions

This study serves as an initial inquiry to pooling multiple time series. Future research can be conducted in the following areas. First, in the current study, we simply assume that $\mu$, $\alpha$, and $\phi$ are homogenous across participants. Such an assumption can be tested before pooling data together. Second, another extension is to fit a multilevel time series model to the data to account for possible heterogenous in the model parameters. Third, in read data collection, the Likert items are often used and therefore, data collected are often ordinal and not normally distributed. How to analyze multiple time series of ordinal data can be investigated in the future. Fourth, in collecting time series data, missing data can occur more often than cross-sectional data collection. Therefore, dealing with missing data is another potential topic.

## References

Cattell, R. B. (1952). The three basic factor–analytic research designs–their interrelations and derivatives. *Psychological Bulletin*, *49*, 499–520.

Cattell, R. B., Cattell, A. K. S., & Rhymer, R. M. (1947). P-technique demonstrated in determining psychophysical source traits in a normal individual. *Psychometrika*, *12*, 267–288.

Cattell, R. B., & Scheier, I. H. (1961). *The meaning and measurement of neuroticism and anxiety*. New York: Ronald Press.

Daly, D. L., Bath, K. E., & Nesselroade, J. R. (1974). On the confounding of inter– and intraindividual variability in examining change patterns. *Journal of Clinical Psychology*, *30*, 33–36.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181-202.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology – this time forever. *Measurement: Interdisciplinary Research and Perspectives*, *2*, 201–218.

Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (pp. 339–360). Norwell, Massachusetts: Kluwer Academic Publishers.

Nesselroade, J. R., & Molenaar, P. C. M. (2003). Quantitative models for developmental processes. In J. Valsiner & K. Connolly (Eds.), *Handbook of developmental psychology* (pp. 622–639). London: Sage Publications.

Nesselroade, J. R., & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time–series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 224–250). Newbury Park, CA: Sage Publications.

Nesselroade, J. R., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development*, *1*, 9–29.