PSYCHOMETRIKA—VOL. 75, NO. 1, 158–175
MARCH 2010
DOI: 10.1007/S11336-009-9137-9

TIME DELAY EMBEDDING INCREASES ESTIMATION PRECISION OF MODELS OF INTRAINDIVIDUAL VARIABILITY

TIMO VON OERTZEN

MAX PLANCK INSTITUTE FOR HUMAN DEVELOPMENT

STEVEN M. BOKER

UNIVERSITY OF VIRGINIA

This paper investigates the precision of parameters estimated from local samples of time dependent functions. We find that *time delay embedding*, i.e., structuring data prior to analysis by constructing a data matrix of overlapping samples, increases the precision of parameter estimates and in turn statistical power compared to standard independent rows of panel data. We show that the reason for this effect is that the sign of estimation bias depends on the position of a misplaced data point if there is no a priori knowledge about initial conditions of the time dependent function. Hence, we reason that the advantage of time delayed embedding is likely to hold true for a wide variety of functions. We support these conclusions both by mathematical analysis and two simulations.

Key words: time delay embedding, dynamical systems models, intraindividual variability, oscillatory data.

1. Introduction

Many phenomena of interest in psychology exhibit intraindividual variability (Baltes, Reese, & Nesselroade, 1977; Nesselroade & Ram, 2004). That is to say, when variables that indicate a phenomenon are observed on multiple occasions, their values are likely to exhibit fluctuations—at some occasions, the variables values will be higher and at other occasions lower. If one wishes to model this variability, it is necessary to measure each individual sufficiently often to capture the fluctuations and for a sufficient number of occasions in order to be able to estimate interindividual differences in this intraindividual variability. The required density of measurement in time is dependent on how rapidly the phenomenon is fluctuating (Shannon & Weaver, 1949). The required number of occasions per individual is a function of the statistical power required to detect the effects of interest (Cohen, 1988). In addition, one may wish to predict individual differences in coefficients of models of intraindividual variability.

Experimental designs for longitudinal studies range from very short time series (e.g., 3 or 4 observations per individual) with many individuals to long time series (e.g., hundreds of observations per person) and fewer individuals. Such designs are frequently employed to understand how a trait-like characteristic of a person (e.g., sex) predicts a within-person coefficient representing the time-dependent behavior of the individual. Such models are sometimes expressed as multilevel models where the repeated observations are grouped by individual (Bisconti, Bergeman, & Boker, 2006; Pinheiro & Bates, 2000).

When only a few observations per individual (e.g., 5) are collected on many individuals (e.g., 200), there is reduced power to estimate the within-person time series coefficients. However, this

Requests for reprints should be sent to Timo von Oertzen, Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany. E-mail: vonoertzen@mpib-berlin.mpg.de

© 2009 The Psychometric Society. This article is published with open access at Springerlink.com

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

design has the advantage of a large sample of individuals, and thus the standard error of the mean within-person time series coefficient may be small. In this example design, 1,000 observations were made: 5 observations on each of 200 individuals.

When many observations per individual are collected, the estimates of within-person time series coefficients are more precise. But if a study has a fixed budget and a fixed cost of obtaining an observation, this means that fewer individuals can be recruited if a greater number of observations per person are obtained. If the budget allows 1,000 observations, another plausible design might be 50 observations on 20 individuals. In this case, improved within-person precision comes at the cost of decreased between-persons sample size. But a more precise within-person measure requires fewer individuals in order to estimate its mean and variance. Thus, there is a trade-off between within-person precision and between-persons precision for a fixed budget.

Others have explored trade-offs between number of individuals and number of observations per individual in multilevel designs (see, e.g., Hox, 2002; Snijders & Bosker, 1999), finding that the variance and sample size at each level of a multilevel model determines power, and thus impacts optimal design. The current article presents a method for improving power at the within-person level when the objective is to fit models to account for intraindividual fluctuations. This improvement can, for instance, substantially change the optimal balance between the number of observations per individual and number of individuals in any given experimental design.

Time delay embedding (sometimes called *state space embedding*) has been used for the estimation of coefficients in models of intraindividual fluctuations (e.g., Bisconti et al., 2006; Butner, Amazeen, & Mulvey, 2005; Chow, Ram, Boker, Fujita, & Clore, 2005; Deboeck, Boker, & Bergeman, 2009). Time delay embedding has also been widely used for the estimation of invariants of nonlinear time series in physics (see, e.g., Abarbanel, 1996; Grassberger & Procaccia, 1983; Kantz & Schreiber, 1997; Sauer, Yorke, & Casdagli, 1991). The interested reader is referred to these examples for illustrations of time delay embedding in realistic situations. Due to theorems of Whitney (1936) and Takens (1985), the dynamics of a time series can be estimated from a time delay embedding as long as the embedding delay is not poorly chosen. However, others state that the standard errors for model estimated from overlapping samples time delay embedding may be underestimated, and estimates are commonly adjusted to account for redundant degrees of freedom. The current article explores how overlapping samples time delay embedding affects the precision of estimates of within-person fluctuations relative to an independent samples design.

Both sampling methods, time delay embedding and independent bursts, focus on dynamic parameters (e.g., frequency of a sine or slope of a linear trend) rather than initial conditions (e.g., phase or intercept). In fact, we show that if initial conditions are known, the advantage of time delay embedding is reduced. In this article, we exemplify the gain on two-change processes, linear oscillators, and a simple linear trend. We assume that the model parameters are invariant over time (e.g., the frequency resp. slope is constant over time), and measurement error is homogenous and independent. The latter assumption may not be necessary, as the foundational mechanisms described in this article are independent of this assumption. However, since these assumptions are fairly mild, we concentrate on this situation for the scope of this article.

We will begin by describing what we mean by overlapping samples time delay embedding. We will next derive a reason why overlapping samples time delay embedding method can improve precision of coefficient estimates of intraindividual fluctuations. We will then use two simulations to compare the precision of coefficients obtained when a model is estimated using either (a) many short independent longitudinal *bursts* (essentially a panel design with 5 waves) or (b) an overlapping samples time delay embedding of a longer longitudinal sample.

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

2. Overlapping Samples Time Delay Embedding

Time delay embedding is a method for transforming a time series into a matrix of time-dependent chunks of data. By transforming a long sequence of data into a set of short time-dependent chunks, the time-dependence itself becomes the focus of the analysis rather than the prediction of any particular value at any particular moment in time. That is to say, each row in a time delay embedded data matrix encapsulates the time-dependence of a short chunk of the time series.

We will adopt the terminology "burst" (after Nesselroade, 1991) to describe short, independent, time-sequential samples from a longer time series. We will use the term "overlapping samples" to describe overlapping time-dependent samples created using time delay embedding. Finally, we will use the term "samples" when we are referring to both "bursts" and "overlapping samples."

In this article, we are concerned with the effect of time delay embedding on within-person precision. In order to simplify the exposition, we will assume that there are no between-persons differences and then compare two extremes of experimental design. Of course, a real world design would also be interested in between-person differences and thus an optimum design would include estimates of between-persons variance and fall somewhere between the two extremes examined here. But, by setting between-persons differences to zero, we can isolate the within-person portion of the problem and focus our discussion only on issues that have to do with overlapping samples. Although our exposition makes the assumption of no between-persons differences, the logic we use also applies to a multilevel case where coefficients can vary by individual.

At one end of the spectrum of design, one might observe many individuals a few times. For instance, suppose we observe 200 individuals each on 5 occasions separated by equal intervals of time, Δt. In such a design, each row in a data matrix corresponds to one individual's data and can be considered to be independent. If between-person differences are zero, each row is an independent sample in which the covariances between the observations and their time derivatives within each row have the same true scores. Any observed between-row differences in covariances are thus assumed to be due to error of measurement.

While this experimental design may seem to be assuming a great deal, analyses such as these are remarkably common in the literature. Consider a panel design with 5 waves that assumes stationarity and only analyzes mean point estimates and standard errors for coefficients of time dependence. We argue that this type of design makes the same assumptions that we are making in the following simplified examples.

Consider how the data might be arranged to analyze this experimental design. A 200×5 data matrix, $\mathbf{X}$, can be expressed as

$$\mathbf{X} = \begin{bmatrix} x_{(1,1)} & x_{(1,2)} & x_{(1,3)} & x_{(1,4)} & x_{(1,5)} \\ x_{(2,1)} & x_{(2,2)} & x_{(2,3)} & x_{(2,4)} & x_{(2,5)} \\ x_{(3,1)} & x_{(3,2)} & x_{(3,3)} & x_{(3,4)} & x_{(3,5)} \\ x_{(4,1)} & x_{(4,2)} & x_{(4,3)} & x_{(4,4)} & x_{(4,5)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{(200,1)} & x_{(200,2)} & x_{(200,3)} & x_{(200,4)} & x_{(200,5)} \end{bmatrix}, \tag{1}$$

where $x_{(i,j)}$ is the value of the variable x for the ith individual on the jth occasion of measurement. Each of 200 individuals was measured 5 times, so the total number of measurements was 1,000.

At the other extreme, suppose that we observe a single individual on 1,000 occasions separated by the same interval of time, Δt, as in the previous design. According to our assumptions

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

of stationarity and no between-persons differences, the expectation of the point estimates for covariances between observations and their derivatives is the same as in the previous example if we arrange the 1,000 observations into a 200 × 5 data matrix as

$$\mathbf{X} = \begin{bmatrix} x_{(1,1)} & x_{(1,2)} & x_{(1,3)} & x_{(1,4)} & x_{(1,5)} \\ x_{(1,6)} & x_{(1,7)} & x_{(1,8)} & x_{(1,9)} & x_{(1,10)} \\ x_{(1,11)} & x_{(1,12)} & x_{(1,13)} & x_{(1,14)} & x_{(1,15)} \\ x_{(1,16)} & x_{(1,17)} & x_{(1,18)} & x_{(1,19)} & x_{(1,20)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{(1,996)} & x_{(1,997)} & x_{(1,998)} & x_{(1,999)} & x_{(1,1000)} \end{bmatrix}, \tag{2}$$

where $x_{(i,j)}$ is the value of the variable x for the ith individual on the jth occasion of measurement. Since there is only one individual in this design, $i = 1$ for all observations.

The data matrices in Equations (1) and (2) are of the same order, 200 × 5. But note that there is time dependence between the last column in each row, $x_{(1,p)}$, and the first column in the next row, $x_{(1,p+1)}$, that is being ignored. Since estimating within-person time dependence is the objective of our analysis, we are losing information that could be used to help improve precision of parameter estimates.

Suppose now that we change how the data matrix is constructed and use overlapping samples time-delay embedding to create a 200 × 5 data matrix $\mathbf{X}^{(5)}$,

$$\mathbf{X}^{(5)} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} \\ x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} & x_{1,6} \\ x_{1,3} & x_{1,4} & x_{1,5} & x_{1,6} & x_{1,7} \\ x_{1,4} & x_{1,5} & x_{1,6} & x_{1,7} & x_{1,8} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1,200} & x_{1,201} & x_{1,202} & x_{1,203} & x_{1,204} \end{bmatrix}, \tag{3}$$

where $x_{(i,j)}$ is the value of the variable x for the ith individual on the jth occasion of measurement. Again, there is only one individual in this design, so $i = 1$ for all observations.[1] Note that in contrast to the 1,000 observations required in the first two designs, we now only require 204 observations to construct the 200 × 5 data matrix.

The two data matrices $\mathbf{X}$ and $\mathbf{X}^{(5)}$ in Equations (1) and (3) are of the same order, 200 × 5. But the overlapping samples time delay embedded matrix, $\mathbf{X}^{(5)}$, has two characteristics that appear to be problematic. First, there are only 204 measurements in $\mathbf{X}^{(5)}$ and there are 1,000 measurements in $\mathbf{X}$. So, we would expect that the time-delay embedded matrix, $\mathbf{X}^{(5)}$, would provide model parameter point estimates that would tend to vary more from sample to sample than would estimates using $\mathbf{X}$. Second, the rows of $\mathbf{X}^{(5)}$ are not independent of one another; in fact, the data is being reused from one row to the next. Thus, the assumption of independence of rows is not met.

Intuition would suggest that the lower number of observations and violation of standard assumptions in the time delay embedded matrix would lead to parameter point estimates for models of intraindividual variability that are less precise than those from the data matrix constructed from independent rows and to standard errors estimates that are smaller than the empirical variability in the parameter point estimates. In fact, we find that each of these intuitions are incorrect

[1]The matrix is a Toeplitz-matrix if the rows are reorganized in an inverted order. Such matrices have been used before to investigate the dynamics of time series (cf. Molenaar, 1985; Nesselroade et al., 2002), using a Toeplitz-matrix for the covariance matrix of a complete data series. Here, this matrix structure is used for a data matrix of small bursts taken from a longer time series.

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

both in the case of linear change and for the case of a linear oscillator. The remainder of this article demonstrates why this counter-intuitive result holds, first by deriving sources of imprecision in the case of linear change and the case of a linear oscillator, and then by presenting two simulations that show that the advantage derived algebraically does hold in practice.

3. Effects of Overlapping Time Delayed Samples

Assume a long data series that follows a known model with one parameter θ that describes the behavior of the model. Many local samples of this series are given to estimate θ. If a normally distributed unbiased measurement error is introduced on any data point x_t, then θ for all samples that include x_t and subsequently the total estimator $\hat{\theta}$ will decrease in precision.

In a specific situation, the error constitutes a displacement of x_t from its true value, and $\hat{\theta}$ will consequently be displaced, also. If x_t is included in two samples, though, the displacement of $\hat{\theta}$ will not necessarily be twice as high, as the error in x_t may have less influence in the second sample. Even more, if the displacement of $\hat{\theta}$ caused by a erroneous x_t has different signs in two different samples, the overall error on $\hat{\theta}$ may be smaller than if x_t would only be included in one sample, or even cancel out completely.

In the following, we show a situation where such a reduction of the displacement of $\hat{\theta}$ holds true for every fixed displacement of x_t. As these displacements are caused by unbiased measurement error, the effect will manifest in a higher precision (and not lower bias) of $\hat{\theta}$. So, assume in the following that all samples from the data series are accurate, with exception of a single data point x_t, which is displaced by a small error.

3.1. Linear Functions

Consider for example that the original data series is a line, and three local overlapping samples are taken from it as shown in the three panels of Figure 1. Assume the data point given are perfect, but the data point marked with an arrow (which is a member of all three samples) is a little bit too high. In the first panel of Figure 1, where x_t is at the end of the sample, the line that fits the data best has a positively biased slope (dotted line). In the middle panel, x_t is at the center of the panel, and the best fitting line has a positively biased intercept, but no bias in the slope. In the third panel, where x_t is at the beginning of the sample, the situation is inverted: Here, the slope is underestimated.

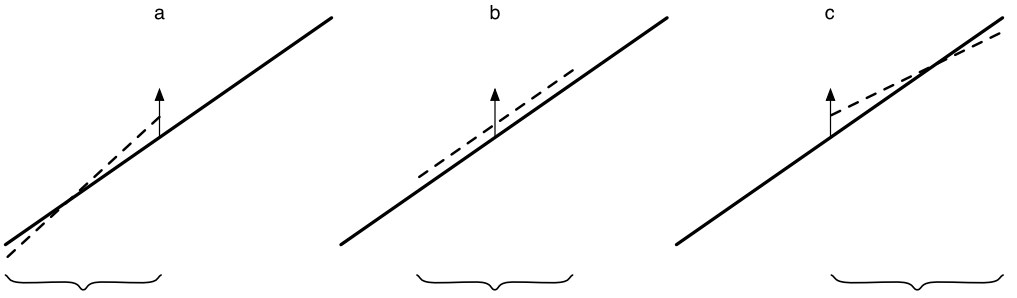


FIGURE 1.
Change of estimated slope with unknown intercept. Curly brackets indicate the observed time span. When in a perfect data set a positive error at the end of a sample is introduced, the slope estimate increases. If the positive error is introduced in the middle of the sample, the estimate of slope does not change, and if the error is introduced at the beginning of the sample, the slope estimate decreases. Thus, in a time delay embedding, a small error introduced in a perfect data set cancels out; indeed, if the estimate is made for each sample separately and the average is taken as overall estimator, the canceling is perfect.

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

This situation is typical for time delay embedding, where each data point (except the first and last few data points) will be at each possible position in one sample. So, we can expect that errors in the data set are partially canceling out.

Suppose that intercept and mean are separately estimated by a maximum likelihood estimation and then averaged over the overlapping samples. In practice, the slope would likely be fixed over the samples, and all samples estimated simultaneously, but arguably the behavior of the estimator would be very similar. Assume the data set is based on a model given by a linear curve with added normally distributed error

$$x_t = I + St + e_t \tag{4}$$

where I and S are normally distributed intercept and linear slope and e_t is normally distributed with zero mean, no covariances and variance σ_E^2. To support the intuition evoked by Figure 1, it will be shown in the following that the biases in a time delay embedding in fact cancel out perfectly for a small error.

For a given sample $x_0, \ldots, x_{T-1}$ of length T from the data set, the likelihood of the sample is

$$L(I, S, x) = \frac{1}{\sqrt{2\pi\sigma_E^{2T}}} \exp\left(-\frac{\sum_{t=0}^{T-1}(x_t - I - St)^2}{2\sigma_E^2}\right) \tag{5}$$

and hence the minus two log likelihood, denoted by $\mathcal{L}(I, S, x)$, is

$$\mathcal{L}(I, S, x) = \log(2\pi) + T\log\left(\sigma_E^2\right) + \frac{1}{\sigma_E^2}\sum_{t=0}^{T-1}(x_t - I - St)^2 \tag{6}$$

the gradient of $\mathcal{L}$ with respect to S and I is given by

$$\frac{\partial \mathcal{L}}{\partial S, I}(I, S, x) = \frac{1}{\sigma_E^2}\sum_{t=0}^{T-1}(x_t - I - St)\begin{pmatrix} t \\ 1 \end{pmatrix}. \tag{7}$$

To find the maximum likelihood estimator of I and S in the sample, the gradient is set equal to zero. The constant factor $\frac{1}{\sigma_E^2}$ does hence not influence the estimator and will be omitted from here on, i.e., we define $g(I, S, x) \propto \frac{\partial \mathcal{L}}{\partial S, I}(I, S, x)$ as

$$g(I, S, x) := \sum_{t=0}^{T-1}(x_t - I - St)\begin{pmatrix} t \\ 1 \end{pmatrix}. \tag{8}$$

Analogously, we define H as the Hessian of $\mathcal{L}$ divided by the same constant factor, i.e., $H(I, S, x) := \frac{\partial g}{\partial S, I}(I, S, x)$. H is then

$$H(I, S, x) = -\sum_{t=0}^{T-1}\begin{pmatrix} t^2 & t \\ t & 1 \end{pmatrix}. \tag{9}$$

We want to investigate the change of the maximum likelihood estimator $\hat{S}$ and $\hat{I}$ on an infinitesimally small increase of one of the data points x_τ. The change of the gradient on this small increase will be the derivative of g w.r.t. x_τ, which is

$$\frac{\partial g}{\partial x_\tau}(I, S, x) = \begin{pmatrix} \tau \\ 1 \end{pmatrix}. \tag{10}$$

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

To have the gradient return to zero, a change m in S and I has to be made such that the change of g on this is the negative of the above value. The change of g with respect to I and S is described by H, i.e., we have to choose m such that

$$Hm + \frac{\partial g}{\partial x_\tau} = 0 \tag{11}$$

$$\iff \quad m = -H^{-1}\begin{pmatrix}\tau\\1\end{pmatrix}. \tag{12}$$

The inverse of H is $H^{-1} = \frac{\operatorname{adj}(H)}{|H|}$, so we have

$$m = \frac{-1}{|H|}\operatorname{adj}(H)\frac{\partial g}{\partial x_\tau} \tag{13}$$

$$= \frac{1}{|H|}\sum_{t=0}^{T-1}\begin{pmatrix}1 & -t\\ -t & t^2\end{pmatrix}\begin{pmatrix}\tau\\1\end{pmatrix} \tag{14}$$

$$= \frac{1}{|H|}\sum_{t=0}^{T-1}\begin{pmatrix}\tau - t\\ t(t-\tau)\end{pmatrix}. \tag{15}$$

This vector describes the change of the estimators $\hat{S}$ and $\hat{I}$ on a small positive error on x_τ.

Now assume we have picked some overlapping samples from the model in a time delay embedding, i.e., in a way that the first sample contains the time points $0, \ldots, T-1$, the second $1, \ldots, T$, and so forth. Assume that the $(T-1)$th data point is biased; then for the ith sequence, the derivative of the estimators is

$$\frac{1}{|H|}\sum_{t=0}^{T-1}\begin{pmatrix}\tau_i - t\\ t(t-\tau_i)\end{pmatrix} \tag{16}$$

where $\tau_i = T - 1 - i$, hence

$$\frac{1}{|H|}\sum_{t=0}^{T-1}\begin{pmatrix}T-1-(i+t)\\ t(t-T+i+1)\end{pmatrix}. \tag{17}$$

Taking the average of the first column (the one corresponding to S) for all T overlapping samples, we get

$$\frac{1}{T|H|}\sum_{i=0}^{T-1}\sum_{t=0}^{T-1}\big(T-1-(i+t)\big). \tag{18}$$

We express the double sum as a single sum that runs over all values j of $(T-1-(i+t))$, i.e., j runs from $-(T-1)$ to $T-1$; each j appears $(T-|j|)$ in the sum, i.e., the above term is equal to

$$\frac{1}{T|H|}\sum_{j=-(T-1)}^{T-1}\big(T-|j|\big)j, \tag{19}$$

$$\frac{1}{T|H|}\sum_{j=1}^{T-1} j\big((T-j)-(T-j)\big) = 0. \tag{20}$$

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

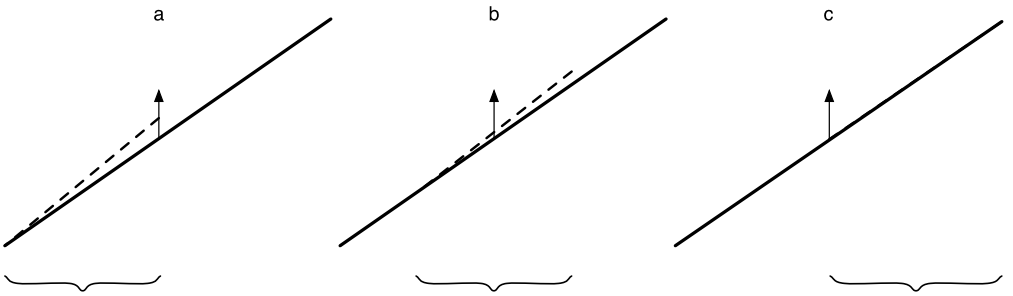


FIGURE 2.
Change of estimated slope with known first point. Curly brackets indicate the observed time span. When in a perfect data set a positive error is introduced while assuming a known first point, the slope estimate will increase, no matter where the error is introduced in the sample, to a higher extent if the error is far from the initial point, and decreasingly less the closer the error is to the initial point. In panel (**c**), the slope is not effected.

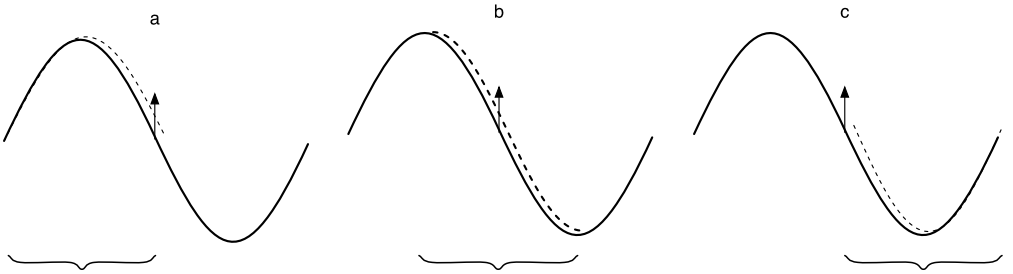


FIGURE 3.
Change of estimated frequency with unknown phase. When in a perfect data set a positive error at the end of a sample (with a half cycle and phase zero) is introduced, the frequency estimate decreases. If the positive error is introduced in the middle of a sample with phases $\frac{\pi}{2}$, the estimate of frequency does not change, and if the error is introduced at the beginning of a sample with phases π, the frequency estimate increases. Thus, in a time delay embedding, a small error introduced in a perfect data set cancels out. If the error is introduced at a phase which is a integer multiple $\frac{\pi}{2}$ and the estimate is made for each sample separately and then averaged, the canceling is perfect.

This means that the averaged estimator of S does not change on a small error introduced in one data point of a linear curve if time delay embedding is applied. The biases of the overlapping samples cancel out each other perfectly.

If part of the global information is given, though, the effect vanishes; consider, for example, that we know the position of the first data point in each sample, and the estimate the slope only. Figure 2 illustrates this situation; as can be seen, the estimated slope now would change to higher values, and only the absolute value of the slope bias is dependent on the position of the error in the sample.

3.2. Oscillatory Functions

The reasons for a local sampling are usually even more pronounced in oscillatory functions like sine curves. So, a natural question is whether the canceling effect of bias for time delay embedding as can be found for linear functions can be found here, also. Figure 3 illustrates the situation of overlapping samples for a sine curve; again, we assume that perfect data is given, and a single measurement x_t is misplaced by a small positive error, indicated by the arrow. In the first Panel of Figure 3, the error is at the end of the sample (which in this example is the end of the first half cycle). The dotted line gives the best fitting sine approximation to the curve, with freely estimated phase and frequency. We can see that the frequency is decreased.

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

In the middle panel, the sample is taken from a quarter of the cycle to three quarters of the cycle, and the same error is introduced at the middle of the sample. Here, the best fitting sine curve is the same curve shifted slightly to the right; so the frequency estimate in this sample does not change.

In the last panel, finally, the frequency is increased to counter the effect of the error. So, it is reasonable to assume that on average, the biases cancel out, at least partly. To show this, the following is an analogue development of the one given for linear curves. Here, we will concentrate on the change of the average frequency estimator for a small deviation of an otherwise perfect sample of the sine curve.

So assume the model is now given by a sine curve with added normally distributed error

$$x_t = \sin(\theta + ft) + e_t \tag{21}$$

where f is a parameter representing the frequency and e_t is normally distributed with zero mean, no covariances and variance σ_E^2. Again, the likelihood for a given sample $x_0, \ldots, x_{T-1}$ is

$$L(\theta, f, x) = \frac{1}{\sqrt{2\pi\sigma_E^{2T}}} \exp\left(-\frac{\sum_{t=0}^{T-1}(x_t - \sin(\theta + ft))^2}{2\sigma_E^2}\right) \tag{22}$$

and hence the minus two log likelihood $\mathcal{L}(\theta, f, x)$

$$\mathcal{L}(\theta, f, x) = \log(2\pi) + T\log\left(\sigma_E^2\right) + \frac{1}{\sigma_E^2}\sum_{t=0}^{T-1}\left(x_t - \sin(\theta + ft)\right)^2. \tag{23}$$

We again consider the gradient with respect to the two model parameters f and θ

$$\frac{\partial \mathcal{L}}{\partial f, \theta}(\theta, f, x) = -\frac{1}{\sigma_E^2}\sum_{t=0}^{T-1}\cos(\theta + ft)\left(x_t - \sin(\theta + ft)\right)\begin{pmatrix} t \\ 1 \end{pmatrix} \tag{24}$$

again, since the gradient is set to zero to find the maximum likelihood estimators of f and θ, we omit the constant factor and define $g(\theta, f, x)$ with $g \propto \frac{\partial \mathcal{L}}{\partial f, \theta}$ as

$$g(\theta, f, x) := \sum_{t=0}^{T-1}\cos(\theta + ft)\left(x_t - \sin(\theta + ft)\right)\begin{pmatrix} t \\ 1 \end{pmatrix}. \tag{25}$$

Let H again be the Hessian of $\mathcal{L}$ divided by the same constant, i.e., $H(\theta, f, x) := \frac{\partial g}{\partial \theta, f}(\theta, f, x)$, which is

$$H(\theta, f, x) = \sum_{t=0}^{T-1}\left(-\sin(\theta + ft)\left(x_t - \sin(\theta + ft)\right) - \cos^2(\theta + ft)\right)\begin{pmatrix} t^2 & t \\ t & 1 \end{pmatrix}. \tag{26}$$

Assume a situation where the data set x is perfect. To investigate the change of the maximum likelihood estimators $\hat{f}$ and $\hat{\theta}$ on an infinitesimally small increase of one of the data points x_τ, consider again the change of g with respect to x_τ

$$\frac{\partial g}{\partial x_\tau} = \cos(\theta + f\tau)\begin{pmatrix} \tau \\ 1 \end{pmatrix}. \tag{27}$$

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

Let m again be the vector of changes in f and θ such that the gradient returns to zero after a change in x_τ. Since $Hm + \frac{\partial g}{\partial x_\tau} = 0$, we get $m = -H^{-1}\frac{\partial g}{\partial x_\tau}$. Since we assumed a perfect data set, we have $x_t = \sin(\theta + ft)$, which simplifies the expansion of this term to

$$m = \frac{-1}{|H|}\operatorname{adj}(H)\frac{\partial g}{\partial x_\tau} \tag{28}$$

$$= \frac{\cos(\theta + f\tau)}{|H|}\sum_{t=0}^{T-1}\cos^2(\theta + ft)\begin{pmatrix} 1 & -t \\ -t & t^2 \end{pmatrix}\begin{pmatrix} \tau \\ 1 \end{pmatrix} \tag{29}$$

$$= \frac{\cos(\theta + f\tau)}{|H|}\sum_{t=0}^{T-1}\cos^2(\theta + ft)\begin{pmatrix} \tau - t \\ t(t-\tau) \end{pmatrix}. \tag{30}$$

As before in the linear case, m describes the change of the estimator $\hat{f}$ and $\hat{\theta}$ on a small positive error in x_τ. If we compare to the linear case, we have two differences: The cosine in front of the sum, which is merely a factor out of the sum and not important, and the squared cosine in the sum.

Now assume T samples taken in a time delay embedding from $2T-1$ data points, again with the $(T-1)$th data point biased. This bias will change the estimates in the ith sample to

$$\frac{\cos(\theta + if + f\tau_i)}{|H|}\sum_{t=0}^{T-1}\cos^2(\theta + if + ft)\begin{pmatrix} \tau_i - t \\ t(t-\tau_i) \end{pmatrix}, \tag{31}$$

where $\tau_i = T - 1 - i$, hence

$$\frac{\cos(\theta + (T-1)f)}{|H|}\sum_{t=0}^{T-1}\cos^2\big(\theta + (i+t)f\big)\begin{pmatrix} T-1-(i+t) \\ t(t-T+i+1) \end{pmatrix}. \tag{32}$$

Taking the average estimate of f for all T samples, we obtain

$$\frac{\cos(\theta + (T-1)f)}{T|H|}\sum_{i=0}^{T-1}\sum_{t=0}^{T-1}\cos^2\big(\theta + (i+t)f\big)\big(T-1-(i+t)\big). \tag{33}$$

Again, we express the double sum as a single one running over all values j of $T-1-(i+t)$. The jth summand appears $T-|j|$ times, so the above term becomes

$$\frac{\cos(\theta + (T-1)f)}{T|H|}\sum_{j=-(T-1)}^{T-1} j\big(T-|j|\big)\cos^2\big(\theta + (T-1)f + jf\big) \tag{34}$$

$$= \frac{\cos(\theta + (T-1)f)}{T|H|}\sum_{j=1}^{T-1} j\big(T-|j|\big)\big[\cos^2\big(\theta + (T-1)f + jf\big) - \cos^2\big(\theta + (T-1)f - jf\big)\big]. \tag{35}$$

As the cosine square is in $[0, 1]$, we see that the effect of the error at least partially cancels out. An easy bound from above shows that the average bias of the error in this way is at most half as much as the bias would be if just one sample was taken with the error in it. Perfect cancellation occurs if the cosine square is symmetric around $\theta + (T-1)f$, i.e., symmetric around the position at which the error was introduced. That is the case if $\theta + (T-1)f$ is an integer multiple of $\frac{\pi}{2}$.

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

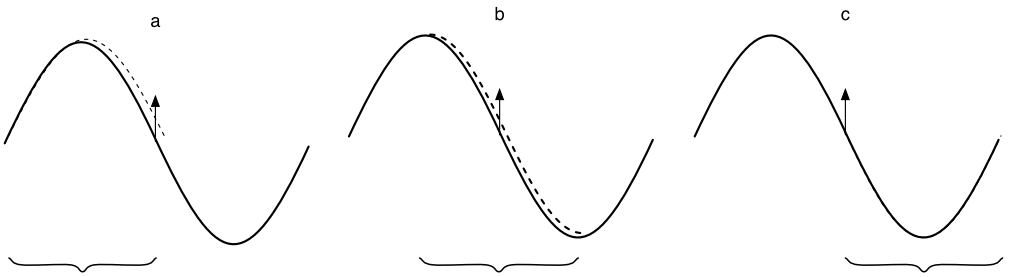


FIGURE 4.
Change of estimated slope with known first point. When in a perfect data set a positive error is introduced while assuming a known phase (as is approximately the case if the phases of the samples are linked in a known way), the frequency estimate will change in the same direction if the samples are made in a time delay embedding.

So, we see that for the sine curve the same effect as for the linear case occurs when a small error is introduced and samples are taken in a time delay embedding. Cancellation for the sine is not necessarily perfect, but as the cosine square has even more symmetry points than the sine curve, it still happens at many positions.

As in the linear case, the effect vanishes when a global embedding is known, for example, if the phase is known (which fixes the first data point). Figure 4 illustrates this situation: In the first two panels, the estimated frequency decreases, and stays constant in the last (as the error is introduced at a data point which was assumed to be known).

4. Simulation Fitting Sine Curves

To illustrate the effect of overlapping samples time delay embedding, we simulated data to conform to (a) the independent samples case and (b) the overlapping samples time delay embedding case. We simulated the data as an undamped linear oscillator (i.e., a pure sine curve) with amplitude equal to one and wavelength equal to 100 observations, i.e., we simulated 100 data points per cycle. In these units, this corresponds to a frequency of 0.01. To this latent score, we added a homogeneous fixed measurement error for each single data point. The measurement error was normally distributed with variance 0.01 or 0.05. From this data, we sampled 40, 200, or 1,000 samples of length five each. For the independent sample case, we simulated each burst separately. Starting with a uniformly distributed phase between 0 and 2π, we simulated five consecutive measurements with frequency and amplitude described above. For the overlapping time delay sampling, we simulated $N+4$ (i.e., 44, 204, or 1004, respectively) consecutive points with one uniformly distributed starting phase, and then sampled N overlapping samples of length five.

In both the independent and the overlapping cases, the same method was used to estimate the frequency. A sine curve was fit to all samples simultaneously, each with an individual starting phase but the same frequency parameters (so $N+1$ parameters in total), using maximum likelihood estimation. For each cell, the simulation was repeated until the standard deviation of the estimated frequency was no longer changing significantly, but using at least 200 repetitions per cell.

For a frequency of 0.01, the error variance of 0.01 corresponds to a signal to noise ratio in the range of 0.000905 (for samples close to the inflexion point of the sine) to 0.72 (for samples close to the intercept of the sine with the x axis). The average signal to noise ratio is 0.35. Expressed in reliabilities (i.e. latent variance over total variance), these values correspond to a range from 0.000904 to 0.42, with a mean of 0.26. For an error variance of 0.05, the signal to noise range

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

TABLE 1.
Standard deviation of estimated frequency, $\hat{f}$, for different homogenous error variance and sample size. "Rows" refers to the number of rows in the data matrix as described in Equations (1) and (3) (i.e., number of samples) and "Obs" refers to the total number of observations required to build a data matrix of the given type (overlapping time delay embedding or independent bursts). In no cell was the proportion of bias of the mean estimated frequency more than 0.005 of the true value.

Condition		Time delay embedding			Independent bursts		
Error	Rows	Obs	mean ($\hat{f}$)	SD	Obs	mean ($\hat{f}$)	Stdv
0.01	40	44	0.00999	0.000053	200	0.01001	0.000126
0.01	200	204	0.01000	0.000019	1000	0.01000	0.000052
0.01	1000	1004	0.01000	0.000009	5000	0.01000	0.000024
0.05	40	44	0.01003	0.000296	200	0.00998	0.000553
0.05	200	204	0.01003	0.000105	1000	0.01004	0.000254
0.05	1000	1004	0.01003	0.000037	5000	0.01001	0.000106

is 0.00018 to 0.144, with a mean of 0.07. Thus, the simulated time series were of markedly low reliability, i.e., were contaminated with a great deal of noise. We set the frequency parameter to be equal in all samples, while estimating a separate phase parameter for each sample.

The assumption of homogenous error can make the data difficult to fit, since some samples will always be close to the inflexion point and suffer from an extremely bad signal to noise ratio. The problem becomes more emphasized the lower the frequency is because with low frequencies, each single data point in a sample close to the inflexion will have the low signal to noise ratio. We chose the low frequency and the homogeneous error to show that nevertheless the frequency parameter can be estimated, and the estimation can be improved drastically by constructing an overlapping samples time delay embedding.

Table 1 gives the results of our simulation. The first two columns give the error variance and the number of bursts simulated in each simulation cell. Columns three to five report the results from the overlapping samples time delay embedding, while columns six to eight report the results of the independent sampling.

In columns three and six, the overall number of simulated data points (respectively, the total number of observations) are calculated for both sampling methods. For the time delay embedding, this is the number of overlapping samples plus four. For the independent bursts, this is the number of bursts multiplied by five. In all cells with both sampling methods, frequency was estimated within half a percent of the true value (see columns four and seven, respectively).

The most important information of Table 1 can be found in columns five and eight, which report the standard deviations of the estimated frequency for both methods of sampling. Note that estimation precision was relatively high despite the high error. Nevertheless, precision was always higher for a time delay embedding, usually more than twice as precise. Note, in addition, that roughly only a fifth of the data points were needed to create the same number of samples in the time delay embedding, since four of the five data points were reused from the sample before. To compare the methods assuming equally many data points, time delay embedding with 200 samples (line two and five, respectively) must be compared to random sampling with 40 samples (line one and four, respectively), and 1,000 samples to 200 samples. The advantage of the time delay embedding then becomes even more pronounced.

5. Simulation Using Latent Differential Equations

Time delay embedding is simply a method for setting up a data matrix prior to analysis. To illustrate that this method can be used with other estimation techniques, we next fit a latent differential equations (LDE) model (see Boker, Neale, & Rausch, 2004 for a more complete

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

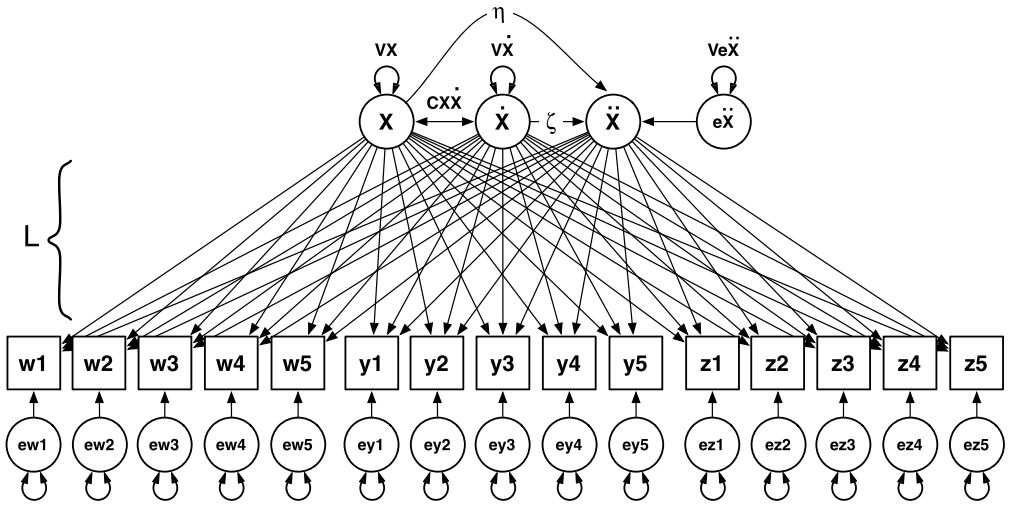


FIGURE 5.
Multivariate Latent Differential Equation structural model with three indicators for a damped linear oscillator latent process.

description of this model) to simulated data and show that the advantage holds for estimation of frequency and damping parameters for a damped linear oscillator. Each row in the data matrices in Equations (1) and (3) can be used to estimate the displacement and derivatives at a time t by filter basis functions similar to quadratic growth curve model. The covariances between these estimated displacements and the derivatives can then estimate the coefficients η and ζ in a second order differential equation model for the damped linear oscillator expressed as

$$\ddot{x}_t = \eta x_t + \zeta \dot{x}_t + e_t, \tag{36}$$

where x_t is the displacement of the variable from its equilibrium value at time t and $\dot{x}_t$ and $\ddot{x}_t$ are its first and second derivatives at time t.

The data for the simulation were generated to conform to a factor model as shown in Figure 5, but with four indicator variables rather than three. The indicator variables were generated for a damped linear oscillator latent process, and time-independent normally distributed error was added to each indicator variable. The latent process variable was created by numerically integrating Equation (36) using Mathematica. Values for each of the coefficients was chosen from the sets $\zeta = \{0.0, -0.02\}$ and $\eta = \{-0.6\}$. Thus, we generated two oscillators: one with no damping and one with mildly negative damping. All indicator variables were simulated with communalities of 0.8, and the four factor loadings were selected as 1.0, 0.9, 0.8, and 0.7, respectively. We drew samples either using the randomized initial condition method (bursts of 5) or using the time delayed embedding method so that the resulting data matrices either had 200 or 400 rows. Thus, we created a $2 \times 2 \times 2$ design with 2 levels of damping, 2 levels of matrix rows, and 2 types of sampling. For each cell in the simulation design, we simulated 1,000 data sets and fit an LDE model to each data set.

The LDE factor model was specified as follows. First a loading matrix, $\mathbf{L}$, was constructed to conform to the four indicators by 5 occasions of measurement per row in the data matrices

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

(Boker et al., 2004)

$$\mathbf{L} = \begin{bmatrix} 1 & -2s & (-2s)^2/2 \\ 1 & -1s & (-1s)^2/2 \\ 1 & 0 & 0 \\ 1 & 1s & (1s)^2/2 \\ 1 & 2s & (2s)^2/2 \\ a & -2as & a(-2s)^2/2 \\ a & -1as & a(-1s)^2/2 \\ a & 0 & 0 \\ a & 1as & a(1s)^2/2 \\ a & 2as & a(2s)^2/2 \\ b & -2bs & b(-2s)^2/2 \\ b & -1bs & b(-1s)^2/2 \\ b & 0 & 0 \\ b & 1bs & b(1s)^2/2 \\ b & 2bs & b(2s)^2/2 \\ c & -2cs & c(-2s)^2/2 \\ c & -1cs & c(-1s)^2/2 \\ c & 0 & 0 \\ c & 1cs & c(1s)^2/2 \\ c & 2cs & c(2s)^2/2 \end{bmatrix} \tag{37}$$

where s is the sampling interval between successive samples, and a, b, and c are free parameters accounting for the factor loadings of the indicators for the latent process. Next, the structural part of the model is specified according to the RAM notation (McArdle & McDonald, 1984)

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \eta & \zeta & 0 \end{bmatrix}, \tag{38}$$

$$\mathbf{S} = \begin{bmatrix} V_F & C_{FdF} & 0 \\ C_{FdF} & V_{dF} & 0 \\ 0 & 0 & V_{d2F} \end{bmatrix}. \tag{39}$$

Finally, a 20×20 diagonal matrix $\mathbf{U}$ is constructed for the residual variances and then the expected covariances of the model can be calculated as

$$\mathbf{R} = \mathbf{L}(\mathbf{I} - \mathbf{A})^{-1}\mathbf{S}(\mathbf{I} - \mathbf{A})^{-1'}\mathbf{L}' + \mathbf{U}. \tag{40}$$

The structural equation modeling software Mx (Neale, Boker, Xie, & Maes, 2003) was then used to fit the LDE model to all data sets using full information maximum likelihood as the objective function.

5.1. LDE Simulation Results

The distributions of η and ζ for the conditions $\zeta = -0.02$ and $N = 500$ are plotted against the normal distribution as a QQ plot in Figures 6a and 6b, respectively. Since each of these plots approximates a straight line, the distributions of estimates of η and ζ are approximately normal. The greater the slope of the line, the wider the distribution of the parameters—thus it is evident that the distributions of η and ζ for time delay embedding in the top two plots are much more precise than those of the independent rows sampled cases in the bottom two plots. Note also that

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

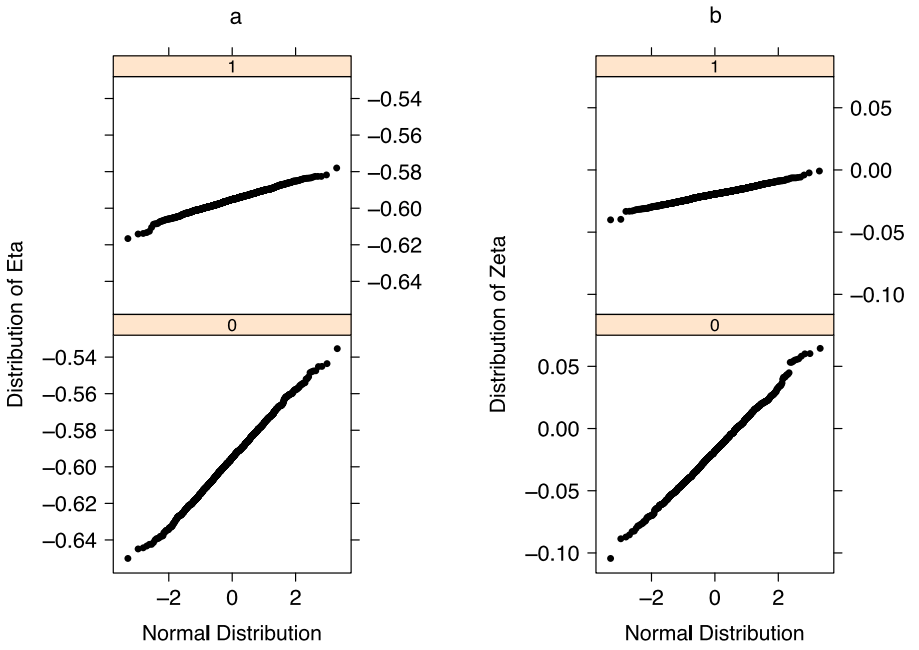


FIGURE 6.
QQ plots of coefficient distributions of (**a**) η and (**b**) ζ versus normal distribution for $\zeta = -0.02$ and $N = 500$. The *top row* of plots in (**a**) and (**b**) are the overlapping time delay embedded data and the *bottom row* of plots are the independent rows sampled data.

the mean of the distributions for all four plots are approximately the true values of $\eta = -0.6$ and $\zeta = -0.02$. The pattern of results in these plots are representative of all of the other cells in the simulation. In all cases, the time delay embedding estimates were considerably more precise than the independent rows sampled estimates, there was no appreciable bias, and the parameter estimates approximated a normal distribution.

Table 2 displays the results of predicting the standard deviations of the parameter values from two of the simulation conditions: number of rows in the data matrix (coded in hundreds) and whether or not the matrix was constructed as a time delayed embedding (coded as 0 = burst sampling and 1 = overlapping samples). Thus, for each 100 observations, the simulation predicts a standard deviation decrease of 0.0032 for η and 0.0047 for ζ. There are also large independent effects of constructing an overlapping samples time delay data matrix: a decrease of 0.0181 for the standard deviation of η and 0.0263 for ζ. Note that the effect on the standard deviation of the parameters of adding one hundred rows to the data matrix is about one fifth the effect of using time delayed embedding.

The bias of η and ζ was also modeled and the results are shown in Table 3. Overall, there was a small, but statistically significant bias for η (bias = 0.005, $p < 0.001$) and the bias for ζ was not significantly different from zero. However, there was not a significant effect on the bias of η or ζ from whether the matrix was constructed as a time delay embedding.

The gain in precision can be rephrased as a gain in power. For illustrative purposes, assume a researcher wants to reject the hypotheses that the oscillator in the example is undamped, i.e., $\zeta = 0$. Under the assumption that the variance of the parameter estimate is independent from the true parameter value, a researcher may apply a Wald test (cf. Bollen, 1989). Using a bootstrap method allows one to get an estimate of the standard deviation of ζ.

In the situation of the simulation above, ζ was -0.02. When using 500 random bursts of length 5, the standard deviation of ζ is 0.0273. Using a significance level of 5%, a one-sided

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

TABLE 2.
Standard deviations of η and ζ predicted as a multiple regression from simulation conditions. N is the number of rows in the matrix in hundreds (i.e., 2 or 5). Embedded is coded 0 for the independent rows data and 1 for the overlapping samples time delay embedded data.

η	Value	SE	t	p
(Intercept)	0.0369	0.00255	14.480	<0.0001
N (in hundreds)	−0.0032	0.00062	−5.082	0.00383
Embedded	−0.0181	0.00187	−9.665	0.00020
Multiple R-squared: 0.960, Adjusted R-squared: 0.943				
ζ	Value	SE	t	p
(Intercept)	0.0520	0.00354	14.693	<0.0001
N (in hundreds)	−0.0047	0.00086	−5.421	0.00289
Embedded	−0.0263	0.00259	−10.121	0.00016
Multiple R-squared: 0.964, Adjusted R-squared: 0.949				

TABLE 3.
Bias for η and ζ predicted as a multiple regression from simulation conditions. N is the number of rows in the matrix in hundreds (i.e., 2 or 5). Embedded is coded 0 for the burst sample matrix and 1 for the embedded matrix.

η	Value	SE	t	p
Intercept	0.00560	0.00065	8.562	<0.001
N (in hundreds)	−0.00006	0.00016	−0.402	0.704
Embedded	−0.00083	0.00048	−1.737	0.143
Multiple R-squared: 0.389, Adjusted R-squared: 0.144				
ζ	Value	SE	t	p
Intercept	0.00057	0.00233	0.244	0.817
N (in hundreds)	0.00004	0.00057	0.077	0.941
Embedded	−0.00066	0.00171	−0.388	0.714
Multiple R-squared: 0.030, Adjusted R-squared: −0.358				

Wald test will reject the hypotheses of no damping if the estimated ζ is smaller than -0.0309. Considering the distribution described in Figure 6, this will be the case in 28.0% of the cases. In other words, the power to reject the hypothesis of no damping with a Wald test is only 28.0%, which is usually considered a rather low power.

When using a data matrix of 500 rows with a time delay embedding, the standard error of ζ is only 0.0092. In this case, the critical value for the Wald test would be -0.0151. Again considering the distribution from Figure 6, this corresponds to a power of 70.2%. We can see that the power is drastically increased even though only a fifth of the samples had been taken.

It is reasonable to assume that the increase in power also holds for other statistical tests than the Wald test. A more complete investigation of this topic needs to be developed in the future.

6. Discussion

Overlapping samples time delay embedding of a time series improves precision of parameter estimates in comparison to sampling short independent rows when the assumptions made in this article are met. There is a surprisingly large benefit to using the time delay embedding method, but there is a mathematical reason why this benefit appears. If one is interested in modeling intraindividual variability as a time dependent process then it appears to be a much better strategy

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

to put effort into gathering a few more samples for a single longer time series than into gathering another independent row of data.

Note that in time delay embedding, fewer but consecutive observations are drawn from the time series, compared to random burst sampling, where more observations are drawn, but in small consecutive chunks. This seems to fit with the general observation that fewer observations over more time points can contain more information than many short observations.

We expect that one beneficiary of this method will be models of intraindividual processes that are formulated as differential equations. Differential equations have long been considered as appropriate for social science data (Hotelling, 1927), but only recently have sufficient data and methods for estimation been available. There are now several ways to estimate the parameters of differential equations such as Equation (36). We expect that the result presented in the current article may improve power in some of these methods. For instance, functional data analysis (Ramsay, Hooker, Campbell, & Cao, 2007) and the approximate discrete method (Oud & Jansen, 2000) are candidate estimation methods that might exhibit improved power when using overlapping samples time delay data matrices.

Note that time-varying covariates can be added to the data matrix in additional columns. The covariate should be aligned so that the time of occurrence of the covariate is at the time of the center of the window. For example, in the matrix $X^{(5)}$ of Equation (3) in the introduction, the value of the covariate at the third time point would be placed in a sixth column in the data matrix and then could be used in modeling. If such an approach is taken, a mixed-effects model grouped by person ID will likely be used.

We caution that overlapping samples time delay embedding is a tool that, like any statistical tool, could lead to poor results when assumptions are violated. In our simulations, we assumed that (a) we are interested in modeling fluctuations about an equilibrium and not growth from a known zero point, (b) the model parameters were invariant over time, (c) the error was homogeneous and uncorrelated over time, and (d) the models were linear change and linear oscillators. Our results show that the time delay embedding method has no advantage when the initial conditions are known and do not need to be estimated. However, more work needs to be done in order to find out whether the other three assumptions are actually necessary and how robust the advantage of overlapping samples time delay might be if the assumptions are violated.

Furthermore, we caution that one should not assume that it is always a better strategy to put one's money and effort into sampling a longer time series than to gather more independent rows of data. For instance, suppose one is interested in how short-term regulation changes over the long term, e.g., how emotional regulation changes as we age. In this case, one must carefully weigh the power requirements for detecting time-related change in the short-term regulatory coefficients. More reliable estimates of the coefficients will help, indicating more observations within a independent row; but longer intervals between independent rows will also improve ones chance of detecting time-related change in the regulatory coefficients. Additional work needs to be done in order to integrate multilevel power calculations (e.g., Snijders & Bosker, 1999) with power calculations for overlapping samples time delay embedding.

In conclusion, we found much to recommend overlapping samples time delay embedding as a method for improving precision of parameter estimates for models of intraindividual processes in relatively short and noisy time series such as are commonly seen in social science data. We have found that this advantage holds in the case of linear change and for a linear oscillator. Other models for intraindividual process may also benefit from this method, but we have not yet shown that the advantage holds in general. The advantage of time delay embedding does not appear to depend on the estimation method used, thus many methods for estimation may be able to take advantage of this data structuring technique. When applied in situations where the assumptions made in the present article are satisfied, we recommend considering time delay embedding prior to estimation of parameters.

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Abarbanel, H.D.I. (1996). *Analysis of observed chaotic data*. New York: Springer.

Baltes, P.B., Reese, H.W., & Nesselroade, J.R. (1977). *Life–span developmental psychology: introduction to research methods*. Pacific Grove: Brooks/Cole.

Bisconti, T.L., Bergeman, C.S., & Boker, S.M. (2006). Social support as a predictor of variability: an examination of the adjustment trajectories of recent widows. *Psychology and Aging*, *21*(3), 590–599.

Boker, S.M., Neale, M.C., & Rausch, J. (2004). Latent differential equation modeling with multivariate multi-occasion indicators. In K. van Montfort, H. Oud, & A. Satorra (Eds.), *Recent developments on structural equation models: theory and applications* (pp. 151–174). Dordrecht: Kluwer Academic.

Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.

Butner, J., Amazeen, P.G., & Mulvey, G.M. (2005). Multilevel modeling to two cyclical processes: extending differential structural equation modeling to nonlinear coupled systems. *Psychological Methods*, *10*(2), 159–177.

Chow, S.M., Ram, N., Boker, S.M., Fujita, F., & Clore, G. (2005). Capturing weekly fluctuation in emotion using a latent differential structural approach. *Emotion*, *5*(2), 208–225.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.

Deboeck, P.R., Boker, S.M., & Bergeman, C.S. (2009). Modeling individual damped linear oscillator processes with differential equations: using surrogate data analysis to estimate the smoothing parameter. *Multivariate Behavioral Research*, *43*(4), 497–523.

Grassberger, P., & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D*, *9*, 189–208.

Hotelling, H. (1927). Differential equations subject to error, and population estimates. *Journal of the American Statistical Association*, *22*(159), 283–314.

Hox, J.J. (2002). *Multilevel analysis: techniques and applications*. Hillsdale: Erlbaum.

Kantz, H., & Schreiber, T. (1997). *Nonlinear time series analysis*. Cambridge: Cambridge University Press.

McArdle, J.J., & McDonald, R.P. (1984). Some algebraic properties of the Reticular Action Model for moment structures. *British Journal of Mathematical and Statistical Psychology*, *87*, 234–251.

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

Neale, M.C., Boker, S.M., Xie, G., & Maes, H.H. (2003). *Mx: statistical modeling* (6th edn.). Richmond: VCU, Department of Psychiatry.

Nesselroade, J.R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D.S. Palermo (Eds.), *Visions of aesthetics, the environment, and development: the legacy of Joachim F. Wohlwill* (pp. 213–240). Hillsdale: Erlbaum.

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*(1–2), 9–29.

Nesselroade, J.R., McArdle, J.J., Aggen, S.H., & Meyers, J.M. (2002). Dynamic factor analysis models for representing process in multivariate time-series. In D. Moskowitz, & S. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: methods and applications* (pp. 235–265). Hillsdale: Lawrence Erlbaum Associates.

Oud, J.H.L., & Jansen, R.A.R.G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrica*, *65*(2), 199–215.

Pinheiro, J.C., & Bates, D.M. (2000). *Mixed-effects models in S and S-plus*. New York: Springer.

Ramsay, J.O., Hooker, G., Campbell, D., & Cao, J. (2007). Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society B*, *69*(5), 774–796.

Sauer, T., Yorke, J., & Casdagli, M. (1991). Embedology. *Journal of Statistical Physics*, *65*(3–4), 95–116.

Shannon, C.E., & Weaver, W. (1949). *The mathematical theory of communication*. Champaign: University of Illinois Press.

Snijders, T.A.B., & Bosker, R.J. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Newbury Park: Sage.

Takens, F. (1985). Detecting strange attractors in turbulence. In A. Dold & B. Eckman (Eds.), *Lecture notes in mathematics: Vol. 1125. Dynamical systems and bifurcations* (pp. 99–106). Berlin: Springer.

Whitney, H. (1936). Differentiable manifolds. *Annals of Mathematics*, *37*, 645–680.

Manuscript Received: 3 MAR 2009
Final Version Received: 9 JUN 2009
Published Online Date: 11 SEP 2009

https://doi.org/10.1007/s11336-009-9137-9 Published online by Cambridge University Press