A Method of Correcting Estimation Failure in Latent Differential Equations with Comparisons to Kalman Filtering

Kevin L. McKee[*],
Virginia Commonwealth University

Michael D. Hunter,
Georgia Institute of Technology

Michael C. Neale
Virginia Commonwealth University

Abstract

Studies have used the Latent Differential Equation (LDE, Boker et al. 2004) model to estimate the parameters of damped oscillation in various phenomena, but it has been shown that correct, non-zero parameter estimates are only obtained when the latent series exhibits little or no process noise (Deboeck and Boker, 2010). Consequently, LDEs are limited to modeling deterministic processes with measurement error rather than those with random behavior in the true latent state. The reasons for these limitations are considered, and a piecewise deterministic approximation (PDA) algorithm is proposed to treat process noise outliers as functional discontinuities and obtain correct estimates of the damping parameter. Comprehensive, random-effects simulations were used to compare results with those obtained using a State-Space Model (SSM) based on the Kalman filter. The LDE with the PDA algorithm (LDEPDA) successfully recovered the simulated damping parameter under a variety of conditions when process noise was present in the latent state. The LDEPDA had greater precision and accuracy than the SSM when estimating parameters from data with sparse jump discontinuities, but worse performance for diffusion processes overall. All three methods were applied to a sample of postural sway data. The basic LDE estimated zero damping, while the LDEPDA and SSM estimated moderate to high damping. The SSM estimated the smallest standard errors for both frequency and damping parameter estimates.

In this study, we examined systematic errors in estimating the parameters of the Latent Differential Equation (LDE) model of time series data (Boker et al., 2004), then used simulations to calibrate and test a novel method of correcting them. We then applied each of three methods to a sample of lateral, postural sway data and compared the results. We begin with a brief overview of common terms and methods, then review the advantages and limitations of the LDE in more detail.

A stochastic process is a series that changes according to a combination of deterministic and random components. Stochastic processes can be modeled in the frequency domain with spectral methods such as Fourier transforms, in the discrete time domain with

---

[*]Corresponding author: mckeek@vcu.edu.

Autoregressive Moving Average (ARMA) models (Box and Jenkins, 1976), or in continuous time as stochastic differential equations (SDEs; e.g., Wei 2006; Tuma and Hannan 1984). Psychological researchers have increasingly taken interest in continuous-time modeling because it offers a convenient method of handling measurement interval irregularity and estimates parameters that can be easily compared over different interval schemes. For these reasons, differential equation models have also been sought to estimate the dynamics of change as inherent, quantitative traits of a system, a common idea in econometrics (i.e., "deep structural parameters"), and more recently in psychological research (Boker, 2002).

When time series of data are obtained from sensors or questionnaires, variation may be due to multiple sources of randomness. *Measurement error* is defined as variation superimposed upon the true trajectory as a result of imperfect measurement, without influencing the process that is measured. If measurement error is present, then the true state of a system at any time can only be estimated from the data. If the trajectory includes *process noise*, then the true state varies randomly to some extent, often due to unmeasured, exogenous disturbances. In continuous time, the solution of a stochastic differential equation with process noise is called a diffusion process. Modeling diffusion processes requires distinguishing measurement noise from process noise while estimating the parameters of systematic variation. State-Space Models (SSM, e.g., Hunter, 2018) achieve this using iterative algorithms that numerically solve the SDE, conditioning each random distribution at time t on the estimated mean and covariance of the state at time $t-1$. One of the oldest and most well-known algorithms for this recursive conditioning is the Kalman filter (Kalman, 1960b; Kalman and Bucy, 1961), which solves the SDE by sequentially predicting each state from its previous estimate, then making corrections as a function of current observed values and the estimated variances of process and measurement noise. For linear differential equation models, the Kalman filter produces minimum-variance unbiased estimates (MVUE) of the latent states from one or more indicators and can be re-run over iterations of optimization to estimate the structural and measurement parameters of an SDE, including random noise distributions. Nonlinear models may differ in these properties, depending on the functional form of the nonlinearities and methods used to accomodate them. Estimation for nonlinear models can be accomplished using the Extended Kalman Filter (EKF, Kalman, 1960a), or the sigma point filter (or Unscented Kalman Filter, UKF, Wan and Van Der Merwe, 2000).

Techniques such as the Kalman filter often fit differential equations to data using the form of an analytic solution to predict each subsequent state. While analytically solving linear differential equations is straight-forward, some nonlinear differential equations do not have analytic solutions and are not approximated well by the linearized prediction methods most commonly used. One alternative method involves fitting polynomial splines to short, overlapping segments of the series and taking the estimated coefficients to be estimates of derivatives at the times around which each segment is centered (GLLA, Boker et al., 2010). The segments are obtained by "time-delay embedding" the data into shifted, duplicate columns:

$$Y^{(D)} = \begin{bmatrix} y_0 & y_\tau & y_{2\tau} & \cdots & y_{\tau(D-1)} \\ y_1 & y_{1+\tau} & y_{1+2\tau} & \cdots & y_{1+\tau(D-1)} \\ y_2 & y_{2+\tau} & y_{2+2\tau} & \cdots & y_{2+\tau(D-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{N-\tau(D-1)} & \cdots & y_{N-2\tau} & y_{N-\tau} & y_N \end{bmatrix} \tag{1}$$

The embedding dimension, D, or number of columns to embed, and the spacing parameter, τ, are decided *a priori* to capture dynamic behavior at a particular time scale and to adjust the degree of spline smoothing. The embedded data can then be projected onto a matrix of discrete polynomial approximations to produce the estimated latent states and their derivatives, which can subsequently be analyzed for linear or nonlinear structural relations.

Recently, a branch of differential equation modeling has appeared that uses this method, called the Latent Differential Equation (LDE, Boker et al., 2004; Boker, 2012). Rather than estimating derivatives and performing analyses in multiple steps, this approach simultaneously estimates the variances of each order of derivative as well as their linear structural relations as latent variables in a structural equation model (see: Bollen, 1989). The model is specified similarly to the latent growth curve (Preacher, 2008), but with rows representing time-embedded segments of the series rather than independent individuals.

Although the above derivative approximation method may confer advantages for analyzing certain nonlinear, possibly chaotic systems, the LDE has thus far only been used to model linear systems with simple solutions. The most commonly used specification is the Damped Linear Oscillator (Chow et al., 2005), in which variation of the latent state (x) over time (t) is governed by the two coefficients of a 2nd-order, linear differential equation: frequency of oscillation (η) and damping or change in amplitude (ζ):

$$\ddot{x}_t = \eta x_t + \zeta \dot{x}_t + \beta u_t + w_t \qquad\qquad w_t \sim N(0, \sigma_W) \tag{2}$$

$$y_t = cx_t + \epsilon_t \qquad\qquad \epsilon \sim N(0, \sigma_\epsilon) \tag{3}$$

The observed variable (y) is then a linear combination of the latent state and normally distributed measurement error (ϵ). Together, the coefficients η and ζ describe the stability of the state about some point of equilibrium over time and the rate of recovery from random movement due to process noise (w) and known, exogenous forces (u), hence they provide one possible model of stress and emotional resilience. Figure 1a shows a deterministic trajectory following from initial conditions described by the differential equation $\ddot{x}_t = -.62x_t - .17\dot{x}_t$ with initial conditions $x_0 = 1, \dot{x}_0 = 0$. A Monte-Carlo simulated data set generated from the same equation with the added process noise term w_t is shown in Figure 1b.

LDEs have been used to model affect (Steele and Ferrer, 2011; Chow et al., 2005), well-being following loss of a spouse (Bisconti et al., 2004), depressive symptomology

(Nicholson et al., 2011), feelings of intimacy (Boker and Laurenceau, 2006), working memory (Gasimova et al., 2014), and physiological synchrony in romantic (Helm et al., 2012) and maternal relationships (Zentall et al., 2006).

Despite the growing number of applications, it has been shown that the LDE does not estimate the damping parameter correctly when the latent state is disturbed by process noise (Deboeck and Boker, 2010). Many of the studies that did not begin measurement at a point of intervention while controlling for process noise after t_0 reported damping estimates $(\hat{\zeta})$ statistically indistinguishable from zero. We gathered several reported parameter estimates into Table 1. A raw value of ζ is not sufficient to judge its effect size because it depends on the corresponding value of the frequency, η. To standardize and interpret all of the ζ effect sizes across studies, we converted them to a dimensionless form as the the ratio (r) of amplitudes at consecutive oscillation peaks x_0 and x_1, using $r = x_1/x_0 = e^{\frac{\pi\zeta}{\sqrt{-\eta}}}$ (see Strogatz, 2000, p. 64–66 for more on dimensionless coefficients). An undamped system thus has $r \approx 1$, and a critically damped system has $r \approx 0$. The studies by Helm et al. (2012), Chow et al. (2005), and Gasimova et al. (2014) used the univariate LDE and reported no damping (i.e., an amplitude ratio of 1; See Table 1). Conversely, Bisconti et al. (2004) reported a 60% reduction in amplitude over the series of well-being measures in bereaved widows. The key difference is that in the study of bereaved widows, the series was initiated following a major, focal disturbance (the loss of a spouse), and the subject matter was such that subsequent disturbances of comparable magnitude were unlikely. The results are interpreted correctly as a decay in amplitude over the whole series, rather than in a recurrent impulse-response manner. Other studies estimated slight individual damping (Steele and Ferrer, 2011; Zentall et al., 2006) when using the coupled, multivariate version of the damped oscillator. We expect that the mechanism of bias is not fundamentally different for the coupled models, but the case is complicated by the allowance of forcing behavior between separate oscillators.

Many previous simulation studies have not shown the model's basic limitation because they excluded process noise entirely, focusing instead on measurement error over deterministic trajectories (Boker et al., 2004; Chow et al., 2005; Hu et al., 2014; McKee et al., 2018). The severe parameter estimate bias due to process noise limits the LDE to modeling deterministic solutions with measurement error but makes it inappropriate for diffusion processes. While the study by Deboeck and Boker (2010) first demonstrated this problem, subsequent LDE literature has not yet addressed it.

The inability of the LDE to estimate the damping parameter likely results from the combination of its polynomial approximation scheme and its method of decomposing the observed covariance structure. While the Kalman filter conditions each state estimate on prior estimates, the LDE model algebra does not include any expectation that higher order derivatives at time $t - $ determine lower-order derivatives at time t, though by definition, $x_t = x_{t-\Delta} + \Delta\dot{x}_{t-\Delta}$. Instead, derivatives are treated as independent and identically distributed over time, and only their contemporaneous covariance, $\Sigma_{X,\dot{X},\ddot{X}}$, is modeled:

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \eta & \zeta & 0 \end{bmatrix}, \qquad S = \begin{bmatrix} V_X & C_{X,\dot{X}} & 0 \\ C_{X,\dot{X}} & V_{\dot{X}} & 0 \\ 0 & 0 & V_W \end{bmatrix} \tag{4}$$

$$\widehat{\Sigma}_{X,\dot{X},\ddot{X}} = (I - A)^{-1} S (I - A)^{-\top} \tag{5}$$

$$= \begin{bmatrix} V_X & & \\ C_{X,\dot{X}} & V_{\dot{X}} & \\ \eta V_X + \zeta C_{X,\dot{X}} & \zeta V_{\dot{X}} + \eta C_{X,\dot{X}} & \eta^2 V_X + \zeta^2 V_{\dot{X}} + 2\eta\zeta C_{X,\dot{X}} + V_W \end{bmatrix} \tag{6}$$

Furthermore, the LDE fits polynomial splines over time-embedded intervals and treats the coefficients as estimated derivatives, though they are contemporaneously dependent for reasons not described by the model. To illustrate this, suppose that the time-delay embedding interval iterates forward by one occasion, introducing a large disturbance in its last column. The best-fitting polynomial over the new interval will include changes to the level, slope, and curvature, even though its expected values at the time points shared with the prior interval should remain nearly unchanged. The only paths describing dependence between derivatives in the LDE are the system dynamics, η and ζ, and the covariance $C_{X,\dot{X}}$. With no mechanism to contrast the influence of disturbances against expected changes due to prior states, contemporaneous dependence introduced by disturbances will appear as additional bias to the estimates of system dynamics. The LDE only decomposes the contemporaneous covariance matrix of the lagged data columns, which provides no information to distinguish covariance due to system dynamics from the inherent covariance of polynomial coefficients fit to the overlapping intervals of data. This is demonstrated in Figure 2. It can be seen that the overlapping polynomials create a smooth approximation that resembles an undamped system when the true series is actually a damped system with disturbances. The bottom scatter plot reflects the positive correlation of slope and curvature due to the intervals in which the disturbances accelerate the state away from equilibrium (shown in red). Excluding these intervals from the time-delay embedding structure results in two, independent segments (blue) for which the damping estimate, the slope of the dashed, blue line, matches the regression slope of the undisturbed, damped oscillator.

If the model cannot distinguish process noise from system dynamics, then the non-random relationships estimated (i.e. η and ζ) will be biased by stochastic behavior, as we have seen. However, by estimating the influence of individual observations on the covariance structure of the derivatives, we can manipulate the time-delay embedding pattern to minimize derivative covariance due to process noise. This approach is proposed to work for cases in which the LDE's contemporaneous latent structure fits well to at least a plurality of the time-embedded data rows, which can thus be used as a reference distribution for identifying stochastic outliers. Outliers are therefore classified relative to the model's expectation. We discuss this requirement in more detail later.

In this study, we attempted to augment the LDE with an algorithm that automatically locates such outliers and treats them instead as functional discontinuities in the latent state. While it is unlikely that the LDE can be made to model diffusion processes analogous to iterative filtering methods without itself being reformulated as an iterative method, this approach allows the LDE to model a series as piecewise deterministic trajectories with measurement error. We therefore refer to this algorithm as Piecewise Deterministic Approximation (PDA), or LDEPDA when used specifically with the LDE.

Other problems arise from the formulation of the LDE that we did not aim to address with this study. Other authors have pointed out that modeling the data as independent when they are not violates the assumptions of the likelihood function and is expected to produce incorrect standard errors and likelihood ratio tests (Oud and Singer, 2008). This has usually been addressed in LDE applications by bootstrapping estimates of standard errors. Second, the best-fitting polynomial to approximate any given interval of a sine wave will systematically exhibit less curvature than the exact Taylor approximation of the sine wave at the center point of that interval (See Appendix). The error in curvature approaches zero as the interval width approaches zero, or as the order of the approximating polynomial approaches infinity. Resulting bias to parameters can be mitigated by using a sufficiently high-order polynomial approximation to the sinusoidal trajectory, or if possible, by using smaller sampling intervals and a smaller embedding dimension. We used an embedding dimension of half the true oscillation period, and for model simplicity and consistency with previous applications, only used derivatives up to the 2nd order. As a consequence, some minor bias to the parameters is expected.

Methods

Piecewise Deterministic Approximation (PDA) algorithm

When modeling time-delay embedded data, an outlier to the multivariate distribution of lags is not strictly an occasion of measurement, but a segment of consecutive occasions that does not exhibit the same dynamics as the best-fitting plurality of other segments. We assume that such a plurality exists due to true dynamic behaviors that are stable over time and accurately characterized by the model. If no such plurality exists, then the dynamics estimated are simply the average behavior, which may not necessarily be representative of any particular interval of the series. If multiple dynamics underlie the series, then common methods of mixture modeling and moderation variables can be applied equally well to the LDE as with any linear structural equation model. We limited our scope to the simplest case, wherein the data are best described by one underlying set of stationary dynamics, obfuscated by additive noise. As a result of time-delay embedding, individual measurements are repeated across multiple rows and will only be excluded if they represent extreme stochastic behavior fewer than D occasions apart. Otherwise, the following algorithm will only introduce discontinuities in the overlap of the time-delay embedded segments.

The ideal circumstances for our solution can be characterized as a "shot noise" process, in which disturbances are infrequent, of large magnitude, and interspersed by periods of deterministic behavior. The process noise may be in such cases defined as a Bernoulli-Gaussian process, with each disturbance following a probability of occurrence and a

Gaussian-distributed magnitude. If the probability of occurrence is small, few rows of data need to be excluded to estimate the true dynamic behavior. If process noise is simply Gaussian, then disturbances occur at every occasion and the series is better characterized as a diffusion process.

To detect outliers, we fit the LDE to the embedded data, then compute a vector of squared Mahalanobis distances (M) of each row of data ($\overrightarrow{x}$), given by:

$$m_t(\overrightarrow{x}) = \left(\overrightarrow{x} - \overrightarrow{\mu}_X\right)^T \widehat{\Sigma}^{-1}\left(\overrightarrow{x} - \overrightarrow{\mu}_X\right) \in M(X) \qquad (7)$$

Where $\overrightarrow{\mu}_X$ is the vector of expected column means, and $\widehat{\Sigma}$ is the expected covariance matrix estimated by the model (Mahalanobis, 1936). M will then be greatest for rows of data that are least well described by the model. Because M is χ^2-distributed with D degrees of freedom, one possible threshold for outlier classification would be a χ^2 critical value. However, we chose an alternative strategy for a few reasons: First, the unembedded data are presumed to have some sequential dependence due to the intrinsic dynamics, implying a non-diagonal covariance structure for each embedded row $\overrightarrow{x}$. This could be readily accounted for by the covariance $\widehat{\Sigma}$, but any mis-specification of the mean or covariance structure would leave M as a noncentral χ^2 distribution with the noncentrality parameter being a function of the kind of mis-specification. Second, recall that the data are embedded into overlapping segments of length D. If an outlier exists in the last element of a row, then it will exist in the next $D-1$ rows as well. Third, the sequence of Ms are themselves dependent due to the embedding: neighboring Ms rely on almost all of the same underlying observations. Thus although any particular M may be χ^2-distributed, the set of all Ms will not be. Therefore a simple, constant threshold based on the χ^2 distribution may not always suffice to classify outliers with sufficient sensitivity and specificity when M exhibits the non-stationary and autocorrelated behavior we expect, shown in Figure 3.

By subtracting out any autocorrelation in $M(\overrightarrow{x})$ we make the series of distances stationary and level, making it easier to distinguish singular outliers from neighboring points:

$$m_t'(\overrightarrow{x}) = m_t(\overrightarrow{x}) - \sum_{i=1}^{q} a_i m_{t-i}(\overrightarrow{x}) \in M'(X) \qquad (8)$$

where each a_i is an estimated autoregressive coefficient, making $m_t'(\overrightarrow{x})$ the residuals of an AR(q) model. To decide on the order (i.e., maximum lag) of the autoregression, we must further consider the structure of time-delay embedded, identical columns of data. The mean and variances of the column will be nearly identical, differing only by the exclusive of observations 1 through $D-1$ in columns 2 through D as those columns are time-shifted. Similarly, each off-diagonal of the covariance matrix will contain nearly identical values representing the covariances of observations at t and $t-n$. As equation 8 shows, $m_t(\overrightarrow{x})$ is function of this covariance structure as expected by the model and the row vector of data. Assuming the existence of a best-fitting plurality, if row vector $\overrightarrow{x}$ of length D contains an outlier in column D, covariances $C_{i,j}$ for columns i and j, where $i \neq j$ and $i,j \leq D$, will all be reduced. If an outlier exists in any column k, for $1 < k < D$, then only $C_{i,j}$, where $i \neq j$ and i,j

k, will be reduced. As $m_t(\vec{x})$ sums over the relative influence of each row upon each $C_{i,j}$, rows with outliers in element D will have the largest values of m_t. Assuming no additional outliers are introduced in the $D-1$ rows to follow, those rows will exhibit diminishing values of m_t, and no additional deviation when $k=1$. Figure 3a demonstrates this using sparse outliers, showing the highest values for the first rows in which each occurs. The simple decay pattern of m_t between sparse outliers can be described by a low-order autoregressive process. If an additional outlier is introduced in element l, for $1 < k < l \quad D$, then that row's deviation from the model-defined expectation once again increases, resulting in both patterns of accumulation and oscillation in m_t. Figure 3b shows how the signature increase in m_t accumulates with frequent outliers. This pattern is better described by a higher-order autoregressive process. The chosen autoregressive order q should therefore be appropriate to both D and the expected frequency of process noise outliers in the data. We chose to automate specification of the autoregression with $q = \left\lfloor Dp^{\frac{1}{2}} \right\rfloor$ where p is the outlier probability. Our formula is only a heuristic, and the selection of q can be made more or less conservative by modifying the exponent of p. Figure 3 illustrates the series of squared Mahalanobis distances for series with sparse process noise outliers (left) and frequent outliers (right).

After leveling the series, the next step is to ensure that all rows guaranteed to contain the outlier are classified. To do this, we "smear" M' forward by taking the maximum of the preceding $(D-1)$ values.

$$s(m_t') = max(m_t', \ldots, m_{t-D+1}') \in S(M') \tag{9}$$

Thus, an observation at column j that induces unusual dynamics over row i is removed from all rows in which it occurs except when $j = 1$, in which case the dynamic behavior of row i is initiated by the observation and not necessarily deviant. Using equation 9, we can define a threshold ϕ and classify all outliers as $S(M') > \mu_{M'} + \phi\sigma_{M'}$, that is, all points in M' greater than ϕ standard deviations away from its mean. The transformation from $M(X)$ to $S(M')$ is illustrated in Figure 4.

To summarize:

1. Fit the LDE to the whole embedded data matrix.

2. Compute $S(M)$, the $\vec{a}$-differenced and $(D-1)$-smeared Mahalanobis distances.

3. Re-fit the LDE model excluding rows where $S(M') > \mu_{M'} + \phi\sigma_{M'}$

4. Recompute $S(M')$ for all data using the expected values given by step 3.

5. Repeat from step 3 until convergence is reached as $range(M') \quad 0$

To specify the model as deterministic, we constrained $V_{\ddot{X}}$ to 0 so that the estimated curvature of each row of data is modeled only as a function of the corresponding slope and intercept, assuming no process noise. This increases the Mahalanobis distance for rows that depart from a deterministic expectation.

As the density of outliers is increased, the algorithm will result in a diminishing overlap of segments, and hence, reduced repeated measurement occasions across columns of the time-delay embedded data matrix. At a high enough density, some measurement occasions will be excluded altogether. The LDE currently overstates the precision of its parameter estimates due to time-delay embedding, but if this algorithm is used, estimates of precision are expected to become more accurate as redundant measurement occasions are excluded. If the series is accurately characterized as a system of stationary dynamics with discrete, stochastic outliers, then we expect an improvement in statistical power with the reduced residual variance. However, because the outlier classification is formulated in terms of the mean and standard deviation of M, its sensitivity scales inversely with the magnitude of outliers.

Consider two extreme cases: In the first, a stationary series exhibits no stochastic outliers beyond t_0, but is subject to Gaussian measurement error. Given an accurately specified model of the dynamics, every row of data will fit well in absolute terms, but nonetheless vary. If a high value of ϕ is used and the sample size is small, there may be no sufficiently deviant rows of data to exclude. If a low value of ϕ is used and the sample size is large, several may be excluded purely on the basis of measurement error. In this case, the primary concern is an unnecessary loss of power. In the second scenario, the series is independent over time with no stationary dynamics at all. The algorithm may result in spurious estimates of dynamics, as the data are reduced to any remaining, cohesive behaviors that coincide by chance. This error can be avoided by evaluating the autocorrelation function (ACF) of the series before fitting the model. An independent series will have no significant autocorrelation at any time lag. The algorithm can be extended to include stopping criteria to address some of these concerns, for instance, by defining a limit on the number or percentage of rows that can be excluded and by prerequisite tests of temporal dependence via the ACF. If the LDE is used with this algorithm, rates of row exclusion should be disclosed along with the standard errors of the estimates. In the next section, we outline a procedure for determining a value of ϕ as a function of D and series length.

Calibration of LDEPDA hyperparameters

We aimed to configure the PDA algorithm's hyperparameter, ϕ, to minimize bias to the estimated amplitude ratio. To do this, we constructed a calibration procedure that used simulated data and optimization to determine an appropriate value of ϕ. In this section, we describe our performance objective for LDEPDA, the procedure for reaching it, and the hyperparameter values obtained.

Following from our prior analysis of the classification threshold ϕ, we defined it as a linear combination of embedding dimension D and series length N with a vector of weights, β.

$$\phi = \beta_0 + \beta_1 D + \beta_2 N \tag{10}$$

To calibrate ϕ, LDEPDA was iteratively fit to a single set of simulated shot noise processes, optimizing the following error function f with respect to β:

$$r = e^{\frac{\pi\zeta}{\sqrt{-\eta}}} \qquad (11)$$

$$\epsilon = \hat{r} - r \qquad (12)$$

$$f(\beta) = \sigma_\epsilon + \mathbb{E}[\epsilon \,|\, \epsilon^2 < 1] \qquad (13)$$

here r is the amplitude peak ratio converted from ζ and η, and $\hat{r}$ is its estimated value from $\hat{\zeta}$ and $\hat{\eta}$. This dimensionless conversion ensures that the scale of estimate error remains constant over iterations between which the true, raw effect sizes differed by an order of magnitude. We used the sum of the estimate errors of the amplitude ratio and their outlier-trimmed mean (Equation 13) to jointly minimize both variance and bias in the estimates. We attempted multiple strategies of calibration and determined that this function results in a central mass of estimates with less bias than achieved by least-squares, due to a heavy tail of unidirectional outliers in the results from LDEPDA.

A representative set of shot noise series was generated with the parameter values and properties given in Table 2, with a total of 24 combinations over 8 conditions. Three series were generated under each condition, making 72 series in total. Dimensionless forms of the frequency and damping parameters were used to generate the data, with frequency specified by the number of data occasions per oscillation period as $\frac{2\pi}{\sqrt{-\eta}}$, and damping determined by the amplitude ratio.

$f(\beta)$ was non-differentiable with respect to β due to the role of dichotomous outlier classifications in the resulting parameter estimate error. Therefore, derivative-free optimization via the Differential Evolution algorithm (Mullen et al., 2011) was used (maximum of 5000 generations, population size=20, F=0.5, Crossover probability=.9, strategy: DE / rand / 1 / bin with per-vector-dither) to search for a global minimum of f with respect to β. Optimization converged by 240 generations ($f(\beta) = 0.317456$) to the following values:

$$\phi = -0.152344 + 0.004885D + 0.0022N, \qquad (14)$$

These values were used in each subsequent application of LDEPDA.

Validation of the LDEPDA Procedure

We aimed to validate the LDEPDA procedure by determining whether it can correct the estimation errors in the LDE and successfully recover the amplitude ratio when the data-generating values are known. Both the SSM and LDE were used as benchmarks for comparisons of performance. The simulated data were generated from a wide variety of parameters to obtain a comprehensive comparison of the robustness of each model and to map scenarios in which each would be preferable.

Design.—We used random-effects simulations to evaluate the precision and accuracy of parameter estimation with the algorithm for a comprehensive range of parameter values, data properties, and process noise distributions, and to compare estimates to those obtained using both the basic LDE and the Kalman-filtered SSM. In the first simulation, each series was generated according to the 2nd-order SDE given in equation 2 (e.g., Figure 1b), where $w \sim N(0, \sigma_w)$. For simulated diffusion processes, the AR order for LDEPDA was set equivalent to D (i.e., $p = 1$). In the second simulation, each series was generated according to a shot noise process where outliers were additive discontinuities in the level (x) that occurred randomly, with normally-distributed magnitude (i.e., a Bernoulli-Gaussian process). For shot noise processes, the AR order of LDEPDA was set at each iteration to $\left\lfloor Dp^{\frac{1}{2}} \right\rfloor$, using the data-generating value of p (though this must be guessed in real applications). All models were fit to the same data at each iteration, allowing direct comparisons of the parameter estimates' precision. Random samples were generated indefinitely for each simulation until sufficient statistical power was achieved to model the estimate error in the results. 40,000 diffusion processes and 80,000 shot noise processes were generated and modeled. Trials were only excluded on account of OpenMx output status codes indicating optimization failure.

Parameters and conditions were generated from uniform distributions with ranges given in Table 3. Data were generated using the dimensionless forms of the parameters to ensure that their possible topological effects were evenly represented over the given ranges. Frequency was represented by the number of occasions per oscillation with a minimum value of 8, thus excluding the special case in which $D = 3$, i.e., polynomial interpolation with no smoothing and hypersensitivity to all noise. Amplitude ratios were limited to values representing damped systems ($r < 1$) to exclude unstable series with continually growing variance.

Software.—All simulations and analyses were conducted in R version 3.4.2 (R Core Team, 2017). Model specification and data generation used the R package OpenMx (Neale et al., 2016), a free, open-source platform for Structural Equation, State-Space (see: Hunter, 2018), and other statistical modeling. Diffusion processes were generated using mxGenerateData() from a continuous-time state space model. Shot noise processes were generated using the R package deSolve with function lsoda and a list of randomly generated, additive events. Autoregression for the PDA algorithm was fit using the R function ar() using Ordinary Least Squares.

Models.—Three versions of the 2nd-order damped linear oscillator model were specified: one LDE tested without PDA but with process noise variance freely estimated, one LDE, identical to the first, but with process noise variance constrained to 0 and used with PDA, and one SSM that included estimation of process noise variance. Both LDE models used an embedding dimension set at each iteration to half of the expected period to best capture the oscillation frequency.

The study by Deboeck and Boker (2010) that first demonstrated the LDE's estimation problems only fit models using observed and expected covariance matrices. We used Full Information Maximum Likelihood (FIML), fitting the model to each row of data for both basic LDE and for use with PDA.

Optimization.—All models used the default OpenMx optimization algorithm, but with the gradient descent step replaced with Nelder-Mead (Nelder and Mead, 1965) for robustness to sub-optimal, local solutions. The OpenMx default configuration of the Nelder-Mead algorithm was used[1] with a maximum of 5×10^5 iterations and the command mxTryHard(), with argument extraTries=15 for quasi-global search of the likelihood space using stochastically varied start values. Box constraints were used to ensure $\hat{\eta} < 0$ and all estimated variances > 0.

Analyses.—Data were only excluded from analysis on the basis of a non-zero OpenMx status code, representing model non-convergence. Raw parameter estimates $\hat{\eta}$ and $\hat{\zeta}$ were converted back to their dimensionless forms, as measures per period $(\hat{\lambda})^2$ and the ratio of amplitude over consecutive peaks $(\hat{r})$, respectively, for direct comparison with the corresponding data-generating values λ and r. Parameter estimate error magnitude was calculated as $(\hat{\lambda} - \lambda)^2$ and $(\hat{r} - r)^2$ and regressed upon the corresponding values for each simulated condition. Data were subset for the multiple regressions according to error outlier thresholds for both amplitude ratio $((\hat{r} - r)^2 < 2)$ and period $((\hat{\lambda} - \lambda)^2 < 20)$ to only model the central mass of estimates and avoid obfuscation due to extreme and infinite values. Pearson correlation was used to determine whether estimate errors followed a similar pattern of occurence between different models applied to the same data.

Results

Model non-convergence rates are given in Table 4, grouped by OpenMx output status code. Only a very small proportion of iterations included convergence failure in at least one of the models, and these iterations were excluded from further analyses. The highest rates occurred for the basic LDE when modeling shot noise processes, with 3.92% of the iterations in total failing to converge.

Coefficients of the multiple regression of squared estimate error onto each simulated condition are given in Table 5 and may be used to directly estimate the empirical error variance of each parameter from values of each condition. Because of the high dimensionality of the results, it is not possible to give a complete visualization of the relationships between the conditions and the estimation error. However, Figures 5, 6, 7, and 8 are vignettes of the parameter space comparing scenarios in which the LDEPDA or SSM either failed differentially or achieved comparable results.

The first simulation examined typical diffusion processes with normally distributed, continuous process noise. When modeling diffusion processes, the basic LDE had the greatest bias in both period and damping. The LDEPDA was robust to diffusion processes as long as the true period was low and a suitably low embedding dimension was chosen. The SSM estimates of the diffusion process were in all cases unbiased and with smaller variance than the LDEPDA. Estimates from the LDEPDA came closer to their true values than those

---

[1]The OpenMx default control options for Nelder-Mead include $\alpha = 1$, $beta_O = 0.5$, $\beta_i = 0.5$, $\gamma = 2$, $\sigma = 0.5$.
[2]Equivalent to 2π times the natural frequency, as data were generated using the natural frequency.

from the SSM in 21.8% of the diffusion process iterations, across all conditions. LDEPDA estimates of damping were more accurate when period was small (less than 14 occasions), and became highly unreliable for larger periods.

The second simulation examined shot noise processes in which additive, functional discontinuities are randomly interspersed between periods of deterministic behavior. In these simulations, the LDEPDA outperformed the SSM in both the bias and variance of the damping estimates under specific conditions. These conditions included a large period length, low outlier probablity, low measurement error, and small to moderately high damping. The SSM tended to underestimate the period and overestimate the effect of damping. When discontinuities occurr at high frequency, the shot noise series somewhat better resembles a diffusion process, but does not become identical to one. Figure 8 shows the recovered performance of the SSM under such conditions, and the appearance of a curvilinear bias to the results from the LDEPDA, dependent on the true amplitude ratio. Both LDE models showed less bias to period than the SSM overall. Estimates of the damping parameter from the LDEPDA were closer to the true values than those from the SSM in 48.5% of shot noise process iterations across all conditions. Figure 9 shows a map of all conditions in which the LDEPDA tended to outperform or parallel the SSM for modeling shot noise processes. Specifically, the LDEPDA damping estimates were better when period was greater than 20 to 30 measurement occasions, scaling with measurement error, when damping effects were small to moderately large, and when outlier probability was below 20%. Outlier probability had a non-linear relationship with the performance of each model, with the SSM excelling under higher outlier probabilities. The SSM damping estimates were also more robust to measurement error.

The basic LDE invariably estimated all varieties of data to be undamped series, regardless of other conditions. This bias is apparent both visually in Figures 5–8 and by its large, negative coefficient on the true damping value in Table 5. The LDE did obtain a small number of correct estimates of damping by virtue of a low outlier probability resulting in series with no discontinuities at all. As we have shown, that is the only scenario in which it can correctly estimate the damping. Estimates of the period were relatively unbiased in the case of shot noise processes.

Figure 10 shows the multivariate distributions of amplitude ratio and period estimates over all conditions for the SSM and LDEPDA. Estimate errors for both parameters, over both types of process were uncorrelated between models overall. The variances and distribution tails for the LDEPDA amplitude ratio estimates were much larger than the SSM for both types of process. LDEPDA estimates of period were generally biased upward for diffusion processes, while estimates of period from the SSM were biased downward for shot noise processes.

Application: Postural Sway

To briefly demonstrate an application of all three models to real data and the possible variation of their results, we used a public postural balance data set published by Santos and Duarte (2016). We extracted a single series of lateral center-of-pressure from one healthy

individual, sampled at 100 Hz over 60 seconds. The series was then truncated to a 40-second run of stationary oscillations and downsampled to 10 Hz, making 401 observations total.

Many methods may be used to decide on an embedding dimension. Because the model supposes damped, sinusoidal oscillation, spectral analysis provides a natural description of oscillatory patterns in the series. Spectral data analyses use the Discrete Fourier Transform (DFT) to decompose the series into the sum of sinusoids over a range of harmonic frequencies, with amplitudes and phase shifts unique to each frequency (Wei, 2006). Thus, the amplitude spectrum can be used to determine the relative contributions of each oscillation to the variance of the signal, typically with noise contributing to the high frequencies. We used the simple periodogram of the data to detect the most prominent oscillation frequency and decide an embedding dimension for both LDE models. If data are abundant, alternative spectral approaches such as Bartlett's method or windowing techniques may be preferable by canceling out corruption of the spectrum due to noise.

The raw series and its spectral density are shown in Figure 11. Major frequency components ranged from 0.1–0.5 Hz, or 2 to 10 seconds per oscillation period. To model the shortest oscillations and avoid excessive bias to parameters, we chose the embedding dimension $D =$ 6 for both models, a 0.6-second interval, or one-third of the expected period. The classification threshold parameters were set to the values given in equation 14, and the PDA AR order was set to 2, expecting a frequency of outliers (p) less than 15%.

Table 6 contains point estimates $\hat{\eta}$ and $\hat{\zeta}$ with bootstrapped 95% quantile confidence intervals. Boostrapping used 500 iterations of resampling for each model. Resampled data was sorted by time index for the LDEPDA to avoid complications of time-dependence with LDEPDA. Each iteration re-ran the entire algorithm to account for variance in the classification results. The SSM used block resampling with randomly located blocks of length 80 to avoid introducing an abundance of discontinuities that may bias the bootstrapped standard error distribution. Conversions to damped frequency (in Hz) and amplitude ratio are given in the second and fourth row, also with bootstrapped quantile confidence intervals. The basic LDE produced the expected zero estimate for the damping parameter, which translates to an amplitude decay rate of less than 1%. The LDEPDA and SSM estimated much larger change in amplitude due to damping. Both estimating an amplitude decay rate around 30–60%, though the LDEPDA's estimate of amplitude ratio was much more variable, with a confidence interval that included one (and equivalently, $\hat{\zeta}$ included zero). $\hat{\eta}$ was similar between LDE and LDEPDA, but the conversion to damped frequency in Hz took into account the larger damping effects in the LDEPDA. Thus, the LDE describes a somewhat larger oscillation period as a result of $\hat{\zeta}$. The raw $\hat{\eta}$ and damped frequency estimated by the SSM fell entirely outside of the confidence intervals of the LDE. The confidence intervals of those parameters in the LDEPDA overlapped with both those of the LDE and the SSM.

Table 6 also contains the number of rows of data, total observations, and free parameters for each model. The LDEPDA removed 153 rows of time-delay embedding, leaving a total of 1488 observations over 248 rows. Because the sample structure differs between all three

models, we cannot compare them on the basis of common model fit indices such as the −2Log-Likelihood, AIC, or BIC, which are all determined by sample size.

Discussion

LDEs have been used in many settings of behavioral research but display some important weaknesses. This paper presented a method for correcting one such weakness: the consistently incorrect estimation of the damping parameter in the presence of process noise. We first showed that when using local polynomial approximation methods to estimate the derivatives of a function, stochastic disturbances result in false dynamics that bias the covariance of the estimated derivatives and consequently result in bias to the linear effects estimated with an LDE. Our correction method, Piecewise Deterministic Approximation (PDA), detects and removes rows of data that are particularly laden with noise, treating them instead as functional discontinuities in the latent state. We conducted two simulation studies to evaluate the performance of PDA for LDEs and compared the results to those obtained by SSM with Kalman filtering. Last, we applied all three methods to a sample of lateral postural sway data and compared the resulting estimates of oscillation frequency and damping.

The first task of the study was to calibrate the classification threshold φ according to a range of possible scenarios that can arise in the data. We accomplished this using a small, representative sample of simulated series with variation over conditions expected to impact estimation. Three parameter values were discovered which jointly minimize the bias and error variance of estimates as a function of the embedding dimension and series length. The results of the calibration depended chiefly on two decisions: the simulated sample and the objective function. Our calibration of the algorithm may be improved with the use of a larger, more diverse simulated sample that expresses a wider range of data conditions. Increasing the number of series per condition would also improve its representativeness and convergence properties. However, with increased conditions and overall sample size, the runtime of optimization also increases multiplicatively. For that reason, we chose a somewhat limited sample that we intend to nonetheless capture common properties and limitations of psychological data, in total observations per series, oscillation period, and measurement error variances. Second, the objective function was specifically chosen to find a solution which balances minimal bias with variance, though the two statistics are in many cases inversely related. If bias does not obscure the relative, between-persons comparability of the estimates and such comparisons are the primary goal of the analysis, then an objective function may be chosen which instead minimizes the variance only. In this case, our primary aim concerned the elimination of bias. Another way to improve the generality of the algorithm's performance would be to include more terms and parameters in the objective function, especially if aspects like measurement error and outlier probability can be determined beforehand.

The random-effects simulation studies mapped the performance of both the SSM and the LDEPDA over a comprehensive space of data-generating properties and model parameters. It is clear from the results that the LDEPDA can be used to model both diffusion and shot noise processes, but is only preferable for the latter under the specific conditions shown in

Figure 9. The LDEPDA's performance was worst when applied to series with high-frequency oscillations and a high frequency of disturbances. Similarly, it did not outperform the SSM under any conditions for which the series was a diffusion process. As evidenced by both Figure 10 and by comparing the multiple regression coefficients in Table 5, LDEPDA error was far more sensitive to each condition than the SSM and for some, liable to consistently produce extreme errors.

Figure 7 shows that for shot noise processes, the SSM consistently deviates from the true amplitude ratio for shot noise processes with an oscillation period longer than 14 occasions and the outlier probability is low. It is unsurprising that the LDEPDA better reflects shot noise processes, given that it fits a deterministic model between sparse, dichotomously classified disturbances. Despite violation of the distributional assumptions of the SSM, it nonetheless produced reliable and unbiased estimates of damping when period length was short. SSMs can be correctly specified for non-normal stochastic distributions, or be fit using objective functions such as Ordinary Least Squares that relax the assumptions of normality invoked here by maximum likelihood. It is also possible to use row-wise model fit indices to segment data for an SSM in a similar manner as the PDA algorithm. A truly fair comparison of models for shot noise processes should include possible modifications to the SSM that we specified and is beyond the scope of this study.

The results of our brief application to postural control data accorded with our expectations, given the results of the simulations. The basic LDE estimated that posture control is an undamped oscillator, but both the LDEPDA and SSM estimated that it is moderate to highly damped, with each oscillation tending toward a 30–60% reduction in amplitude per successive peak. Besides the evidence we have provided that the basic LDE's zero estimate of damping is spurious, it is likely that natural, organic, and highly complex processes such as human postural regulation will include components of random disturbance and compensate with damping-like behavior. One drawback of the LDEPDA is apparent in the large confidence intervals for damping, which encompassed zero despite a moderate effect size. The large confidence intervals are due to uncertainty in the classification step, which was re-run at each iteration of bootstrapping. It is difficult to guarantee statistical power by these methods because the algorithm reduces the number of usable data automatically. It may be possible to improve that aspect by including parameteric standard errors or the number of rows removed in the calibration objective function. More importantly however, the high variability in outlier classification likely indicates that disturbances were too small and too frequent to be easily classified. Visual inspection of the data reveal almost no large jumps that can obviously qualify as discontinuities. It is more likely that the standing posture was only disturbed by a multitude of small, frequent, internal forces such as blood circulation and inconsistent muscle tension. The power of the analyses benefits from correct choice in the underlying noise distributions and type of process, and such a pattern of disturbances is consistent with a Gaussian diffusion process. We therefore have reason to think that the results of the SSM are more credible in this case.

It is also notable that the raw frequency estimate from the SSM fell outside of the basic LDE's confidence intervals, while the LDEPDA frequency estimate was much closer. The

oscillation period estimated by the LDEPDA was, as a function of the much higher damping value, substantially lower than that estimated by the basic LDE.

Even with our outlier classification algorithm, using LDEs to model stochastic processes may be inadvisable for several reasons. The true generality of our results and the threshold values used to obtain them cannot be concluded from necessarily limited simulations and the lack of a comprehensive, analytic exposition. It would be prudent to precede any application to real data with simulations and a calibration process that reflects the nature of that data as best as possible. Results can vary widely depending upon several technical decisions without obvious answers, including the autoregressive model for the squared Mahalanobis distances, time-delay embedding dimension, and classification threshold, though we have provided basic principles and simulation results to assist with each of these. Theoretical considerations are equally important. The best choice in model depends on *a priori* evidence for the kind of forcing functions and noise that affect the dynamical system in question. We were unable to identify any certain advantages of the LDEPDA over the SSM with Kalman filtering for estimating the parameters of linear diffusion processes, and indeed the Kalman filter has been proven for that task analytically. The best scenarios in which to use the LDEPDA are those with sparse stochasticity that cannot or need not be described parametrically, such as unknown jump discontinuities.

The generality of the algorithm with respect to model specification is also not certain. LDEs have not been widely generalized beyond simple first and second-order linear, ordinary differential equations, so we have only sought to correct errors present in that context. Our review of the literature suggests a similar pattern of bias in the multivariate, coupled oscillator LDE, but we constrained our study and development to the univariate case only. Those interested in further developing or applying LDE-based methods should consider the limitations present in its simplest manifestations before attempting more complex analyses such as nonlinear models and whether the same model is possible through alternative methods. For example, we previously demonstrated a method of modeling compound, multi-timescale oscillators as an LDE (McKee et al., 2018). If our outlier detection algorithm is used with this method, the multiplicatively larger embedding dimension (D) implies that much larger stretches of data must be removed to completely eliminate stochastic outliers and obtain unbiased estimates. For a much simpler solution, the same model can be specified as an SSM with a few changes to the transition matrix (A), the state loading matrix (C), and the process noise variance-covariance matrix (Q):

$$\dot{X}_t = AX_t + W_t$$

$$Y = CX_t + \epsilon_t$$

$$X_t = \begin{bmatrix} x_{1,t} \\ \dot{x}_{1,t} \\ x_{2,t} \\ \dot{x}_{2,t} \end{bmatrix} \quad \dot{X}_t = \begin{bmatrix} \dot{x}_{1,t} \\ \ddot{x}_{1,t} \\ \dot{x}_{2,t} \\ \ddot{x}_{2,t} \end{bmatrix} \quad A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \eta_1 & \zeta_1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \eta_2 & \zeta_2 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \quad Q = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & Var_{W_1} & 0 & Cov_{W_1,W_2} \\ 0 & 0 & 0 & 0 \\ 0 & Cov_{W_1,W_2} & 0 & Var_{W_2} \end{bmatrix}$$

As mentioned earlier, the Kalman filter estimates a predicted and corrected value for each state in the series, so disturbances and measurement error scores can be computed as well. For a model such as this, stochastic outliers may be unique to each timescale within a univariate series and can be extracted from the model as the difference of the predicted and corrected state estimate: $w_t = x_{t/t} - x_{t/t-1}$, while measurement error is the difference of the observed value from the state estimate: $\epsilon_t = y_t - C_{t/t}$. Post-hoc analyses of estimated process noise can help determine whether the model is correctly specified, as autocorrelation in the scores should be minimized. Estimation of stochastic outliers may also be of substantive interest, for instance, by revealing unmeasured, stressful events in data on health and well-being, likely instances of drug use, or changes in dynamic behavior.

The methods examined in this study represent only two approaches to continuous-time differential equation modeling. For an alternative SEM approach, ctsem (Driver et al., 2017) provides continuous-time panel modeling for time-series data (the Exact Discrete Method, or EDM), with advantages for analyzing large, multi-person ensembles (Oud and Singer, 2008). A comparison of SEM and Kalman filtering methods is also given by Chow et al. (2010), and Oud (2017) compared estimation of damped linear oscillator parameters using the multivariate LDE, EDM, and an Approximate Discrete Method (ADM). Functional Data Analysis (Ramsay, 2005) includes a variety of techniques for non-parametric curve-fitting and differential equation modeling, including techniques for piecewise deterministic processes with measurement error. For the multi-level modeling of linear dynamics, Bayesian estimation methods have been used with state-space models (Oravecz et al., 2011).

It is possible to model stochastic oscillators with piecewise deterministic trajectories using the LDE with our algorithm to detect stochastic outliers and treat them as functional discontinuities. This approach appears to reliably estimate damping and frequency parameters under specific conditions and is even preferable to the SSM when the data behave according to a shot noise processes with sparse disturbances and long oscillation periods. There are, however, several disadvantages to consider: The augmented LDE involves many more heuristically-determined technical steps over the SSM with room for error, inaccurate standard error estimates due to unmodeled autocorrelation between rows of data and duplicated measures, parameter bias due to local polynomial approximation, inability to model continuous process noise with the system dynamics, and uncertain generalizability of both the model and the algorithm as a consequence. It is also difficult to determine the requirements for statistical power by this method, as the end results are governed by automated transformations of the data structure that depend highly on the

underlying behavior of the data. The limited scope of our study did not include alternations to the SSM that would correct its specification for Bernoulli-Gaussian process noise. Despite this, the SSM was still more accurate than the LDEPDA for a majority of the shot noise processes generated. We nonetheless demonstrated that when the data act according to deterministic behavior with infrequent jump discontinuities or process noise outliers, estimates of the dynamics by the basic SSM can be unpredictable and biased, while the LDEPDA is a workable alternative.

Acknowledgments

Appendix: Derivative bias from quadratic approximation

In this appendix we show that the best quadratic approximation over any non-zero interval of a sinusoid underestimates the true derivatives at the interval's center point. Take a quadratic function $\phi(x, a, b, c)$ that minimizes the $\mathcal{L}^2$-norm with $\cos(x + x_0)$ over the interval $[-r, r]$:

$$\phi(x) = ax^2 + bx + c, \tag{15}$$

$$f(\phi(x), r, x_0) = \int_{-r}^{r} (\cos(x + x_0) - \phi(x))^2 dx, \tag{16}$$

We show that the derivatives of $\phi(x)$ with respect to x are less than the derivatives of $\cos(x + x_0)$ for positive interval r, and approach equality as r goes to zero.

If we minimize f with respect to constants a, b, and c, we obtain the best-fitting quadratic approximation of $\cos(x + x_0)$ over the interval $[-r, r]$. To do this, set $\nabla f(r, x_0) = 0$ and solve for a, b, and c to get:

$$a(r, x_0) = \frac{15(r^2\cos(x_0)\sin(r) + 3r\cos(x_0)\cos(r) - 3\cos(x_0)\sin(r))}{2r^5}, \tag{17}$$

$$b(r, x_0) = \frac{3(r\cos(r)\sin(x_0) - \sin(x_0)\sin(r))}{r^3}, \tag{18}$$

$$c(r, x_0) = \frac{-ar^3 + 3\cos(x_0)\sin(r)}{3r}, \tag{19}$$

Each coefficient of $\phi(x)$ approaches 0 as r is increased. There is a singularity at $r = 0$. Using L'Hôpital's rule, we find that as $r \to 0$, each coefficient approaches the coefficient of its corresponding term in the 2nd-order Taylor approximation of $\cos(x + x_0)$:

$$\lim_{r \to \infty} \begin{cases} a(r, x_0) = 0, \\ b(r, x_0) = 0,, \\ c(r, x_0) = 0, \end{cases} \quad \lim_{r \to 0} \begin{cases} a(r, x_0) = \dfrac{-\cos(x_0)}{2}, \\ b(r, x_0) = -\sin(x_0), \\ c(r, x_0) = \cos(x_0), \end{cases} \tag{20}$$

To see that the limit at $r = 0$ is the global maximum of a, and maxima and minima of a with respect to x_0 are simply $n\pi$ where $\cos(x_0) = \pm 1$:

$$\underset{x_0}{\mathrm{argmax}} \left| a(r, x0) \right| = n\pi, \text{ for any } n \in \mathbb{Z}, \tag{21}$$

$$\underset{x_0}{\max} a(r) = \pm \frac{15\left(r^2\sin(r) + 3r\cos(r) - 3\sin(r)\right)}{2r^5} \tag{22}$$

Because the denominator contains the highest power of r, the function oscillates with monotonically decreasing amplitude and its global maximum is approached as $r \to 0$. Therefore, the 2nd derivative of quadratic approximation $\phi(x)$ over any interval greater than zero will be of a smaller magnitude than the true derivatives of $\cos(x + x_0)$. If such estimates are used in place of exact derivatives in a differential equation, then estimated linear, proportional relations will also be biased toward zero.

References

Bisconti TL, Bergeman CS, and Boker SM (2004). Emotional well-being in recently bereaved widows: a dynamical systems approach. The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 59(4). doi:10.1093/geronb/59.4.P158.

Boker S, Deboeck PR, Schiller CE, and Keel PK (2010). Generalized local linear approximation of derivatives from time series In Statistical methods for modeling human dynamics : an interdisciplinary dialogue, Notre Dame series on quantitative methodologies. Routledge, New York. doi:10.4324/9780203864746.

Boker S, Neale M, and Rausch J. (2004). Latent differential equation modeling with multivariate multi-occasion indicators In Recent developments on structural equation models, pages 151–174. Springer. doi:10.1007/978-1-4020-1958-6_9.

Boker SM (2002). Consequences of continuity: The hunt for intrinsic properties within parameters of dynamics in psychological processes. Multivariate Behavioral Research, 37(3):405–422. doi:10.1207/S15327906MBR3703_5. [PubMed: 26751295]

Boker SM (2012). Dynamical systems and differential equation models of change In Cooper H, Panter A, Camic P, Gonzalez R, Long D, and Sher K, editors, APA Handbook of Research Methods in Psychology, pages 323–333. American Psychological Association, Washington, DC. doi:10.1037/13621-016.

Boker SM and Laurenceau J-P (2006). Dynamical systems modeling: An application to the regulation of intimacy and disclosure in marriage In Models for Intensive Longitudinal Data. Oxford University Press. doi:10.1093/acprof:oso/9780195173444.003.0009.

Bollen KA (1989). Structural equations with latent variables. Wiley series in probability and mathematical statistics Applied probability and statistics. Wiley, New York. doi:10.1002/9781118619179.

Box GEP and Jenkins GM (1976). *Time Series* Analysis: Forecasting and Control (Rev. ed.). Holden-Day, San Francisco.

Chow S-M, Ho M-HR, Hamaker EL, and Dolan CV (2010). Equivalence and differences between structural equation modeling and state-space modeling techniques. Structural Equation Modeling: A Multidisciplinary Journal, 17(2):303–332. doi:10.1080/10705511003661553.

Chow S-M, Ram N, Boker SM, Fujita F, and Clore G. (2005). Emotion as a thermostat: representing emotion regulation using a damped oscillator model. Emotion, 5(2):208. doi:10.1037/1528-3542.5.2.208. [PubMed: 15982086]

Deboeck P and Boker S. (2010). Modeling noisy data with differential equations using observed and expected matrices. Psychometrika, 75(3):420–437. doi:10.1007/s11336-010-9168-2.

Driver CC, Oud JHL, and Voelkle MC (2017). Continuous time structural equation modeling with r package ctsem. Journal of Statistical Software, 77(1):1–35. doi:10.18637/jss.v077.i05.

Gasimova F, Robitzsch A, Wilhelm O, Boker SM, Hu Y, and Hülür G. (2014). Dynamical systems analysis applied to working memory data. Frontiers in psychology, 5:687–687. doi:10.3389/fpsyg.2014.00687. [PubMed: 25071657]

Helm JL, Sbarra D, and Ferrer E. (2012). Assessing cross-partner associations in physiological responses via coupled oscillator models. Emotion, 12(4):748–762. doi:10.1037/a0025036. [PubMed: 21910541]

Hu Y, Boker S, Neale M, and Klump KL (2014). Coupled latent differential equation with moderators: Simulation and application. Psychological methods, 19(1):56. doi:10.1037/a0032476. [PubMed: 23646992]

Hunter MD (2018). State space modeling in an open source, modular, structural equation modeling environment. Structural Equation Modeling: A Multidisciplinary *Journal*, 25(2):307–324. doi:10.1080/10705511.2017.1369354.

Kalman RE (1960a). Contributions to the theory of optimal control. Boletín de la Sociedad Matemática Mexicana, 5(2):102–119.

Kalman RE (1960b). A new approach to linear filtering and prediction problems. Basic Engineering, 82:35–45. doi:10.1115/1.3662552.

Kalman RE and Bucy RS (1961). New results in linear filtering and prediction theory. Transactions of the ASME, Series D, Journal of Basic Engineering, 83:95–108. doi:10.1115/1.3658902.

Mahalanobis PC (1936). On the generalized distance in statistics. National Institute of Science of India.

McKee KL, Rappaport LM, Boker SM, Moskowitz DS, and Neale MC (2018). Adaptive equilibrium regulation: Modeling individual dynamics on multiple timescales. Structural Equation Modeling: A Multidisciplinary Journal, pages 1–18. doi:10.1080/10705511.2018.1442224.

Mullen KM, Ardia D, Gil DL, Windover D, and Cline J. (2011). Deoptim: An r package for global optimization by differential evolution. Journal of Statistical Software, 40(6).

Neale MC, Hunter MD, Pritikin JN, Zahery M, Brick TR, Kirkpatrick RM, Estabrook R, Bates TC, Maes HH, and Boker SM (2016). OpenMx 2.0: Extended structural equation and statistical modeling. Psychometrika, 81(2):535–549. doi:10.1007/s11336-014-9435-8. [PubMed: 25622929]

Nelder JA and Mead R. (1965). A simplex method for function minimization. The Computer Journal, 7(4):308–313. doi:10.1093/comjnl/7.4.308.

Nicholson JS, Deboeck PR, Farris JR, Boker SM, and Borkowski JG (2011). Maternal depressive symptomatology and child behavior: Transactional relationship with simultaneous bidirectional coupling. Developmental Psychology, 47(5):1312–1323. doi:10.1037/a0023912. [PubMed: 21639624]

Oravecz Z, Tuerlinckx F, and Vandekerckhove J. (2011). A hierarchical latent stochastic differential equation model for affective dynamics. Psychological Methods, 16(4):468–490. doi:10.1037/a0024375. [PubMed: 21823796]

Oud J. (2017). Comparison of four procedures to estimate the damped linear differential oscillator for panel data In Longitudinal models in the behavioral and related sciences, pages 19–39. Routledge. doi:10.4324/9781315091655-2.

Oud JHL and Singer H. (2008). Continuous time modeling of panel data: Sem versus filter techniques. Statistica Neerlandica, 62(1):4–28. doi:10.1111/j.1467-9574.2007.00376.x.

Preacher KJ (2008). Latent growth curve modeling Quantitative Applications in the Social Sciences. SAGE, Los Angeles, [Calif.]; London. doi:10.4135/9781412984737.

R Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay J. O. a. (2005). Functional Data Analysis. Springer Series in Statistics Second edition.. edition. doi:10.1007/978-1-4757-7107-7.

Santos DA and Duarte M. (2016). A public data set of human balance evaluations. PeerJ, 4. doi:10.7717/peerj.2648.

Steele JS and Ferrer E. (2011). Latent differential equation modeling of self-regulatory and coregulatory affective processes. Multivariate Behavioral Research, 46(6):956–984. doi:10.1080/00273171.2011.625305. [PubMed: 26736119]

Strogatz SH (2000). Nonlinear Dynamics and Chaos. Perseus Books, Cambridge, MA.

Tuma NB and Hannan MT (1984). Social Dynamics Models and Methods. Elsevier.

Wan E and Van Der Merwe R. (2000). The unscented kalman filter for nonlinear estimation. pages 153–158. IEEE Publishing. doi:10.1109/ASSPCC.2000.882463.

Wei WWS (2006). Time series analysis : univariate and multivariate methods. Pearson Addison Wesley, Boston, 2nd ed. edition. doi:10.1093/oxfordhb/9780199934898.013.0022.

Zentall SR, Boker SM, and Braungart-Rieker JM (2006). Mother-infant synchrony: A dynamical systems approach. In Proceedings of the Fifth International Conference on Development and Learning.
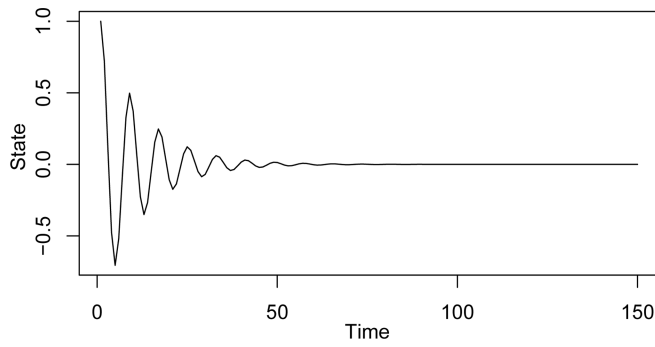
(a) $\ddot{x}_t = -.62x_t - .17\dot{x}_t$
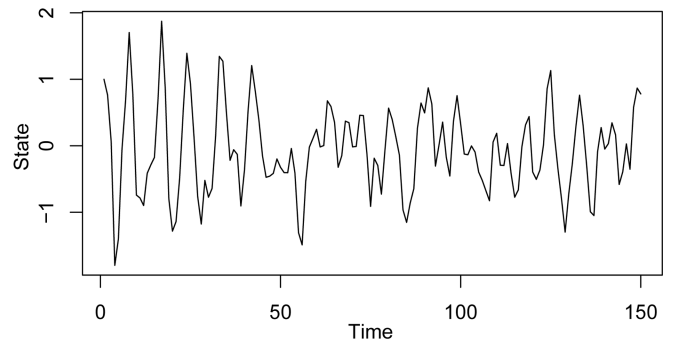
(b) $\ddot{x}_t = -.62x_t - .17\dot{x}_t + w_t, \; w_t \sim N(0, .3)$

Figure 1.
Solutions of ordinary (left) and stochastic (right) 2nd-order differential equations with initial values $x_0 = 1$ and $\dot{x}_0 = 0$
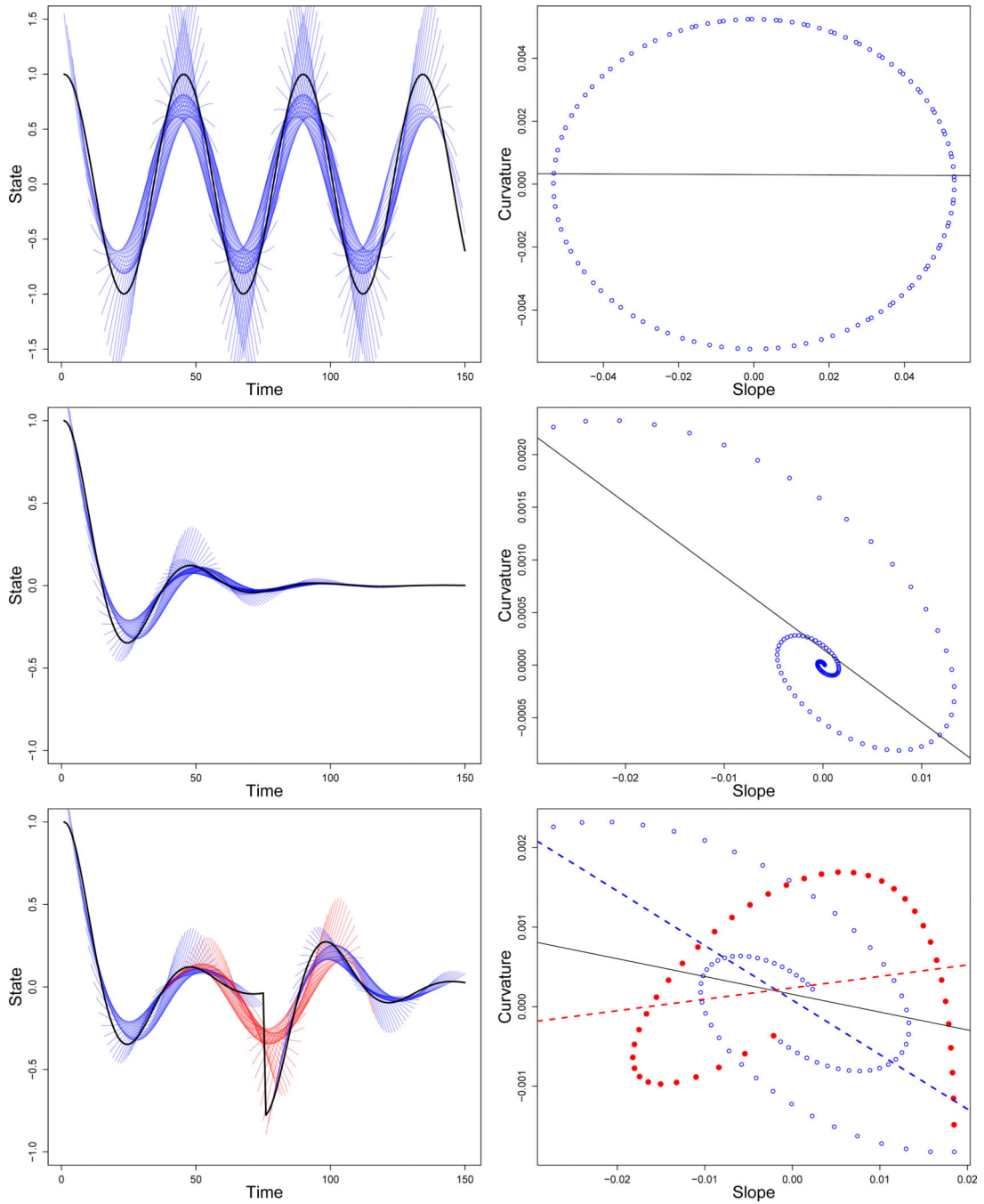
Figure 2.
Polynomial approximations fit to overlapping intervals of an undamped oscillator (top), damped oscillator (middle), and an equivalently damped oscillator with one exogenous disturbance (bottom). The colored curves are quadratics fit to overlapping intervals of the state function. Scatter plots of the slope and curvature coefficients are shown to the right with regression lines showing the direction of covariance. Intervals that include the disturbance are shown in red. False dynamics (red) result from spline-smoothing over disturbances and will bias the slope of the regression line, i.e. the estimated damping (ζ),

toward zero. Excluding the disturbed intervals results in two independent trajectories with the true, damped behavior.
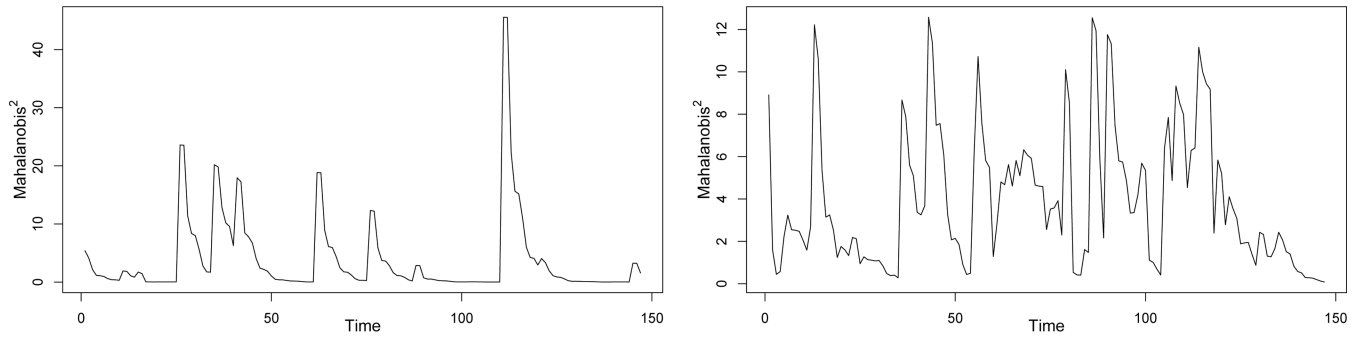
(a) Sparse outliers (6.25% probability)

(b) Frequent outliers (50% probability)

Figure 3.
Squared Mahalanobis distances (M) given two frequencies of process noise outliers. The time dynamics of M depend on the frequency of outliers.
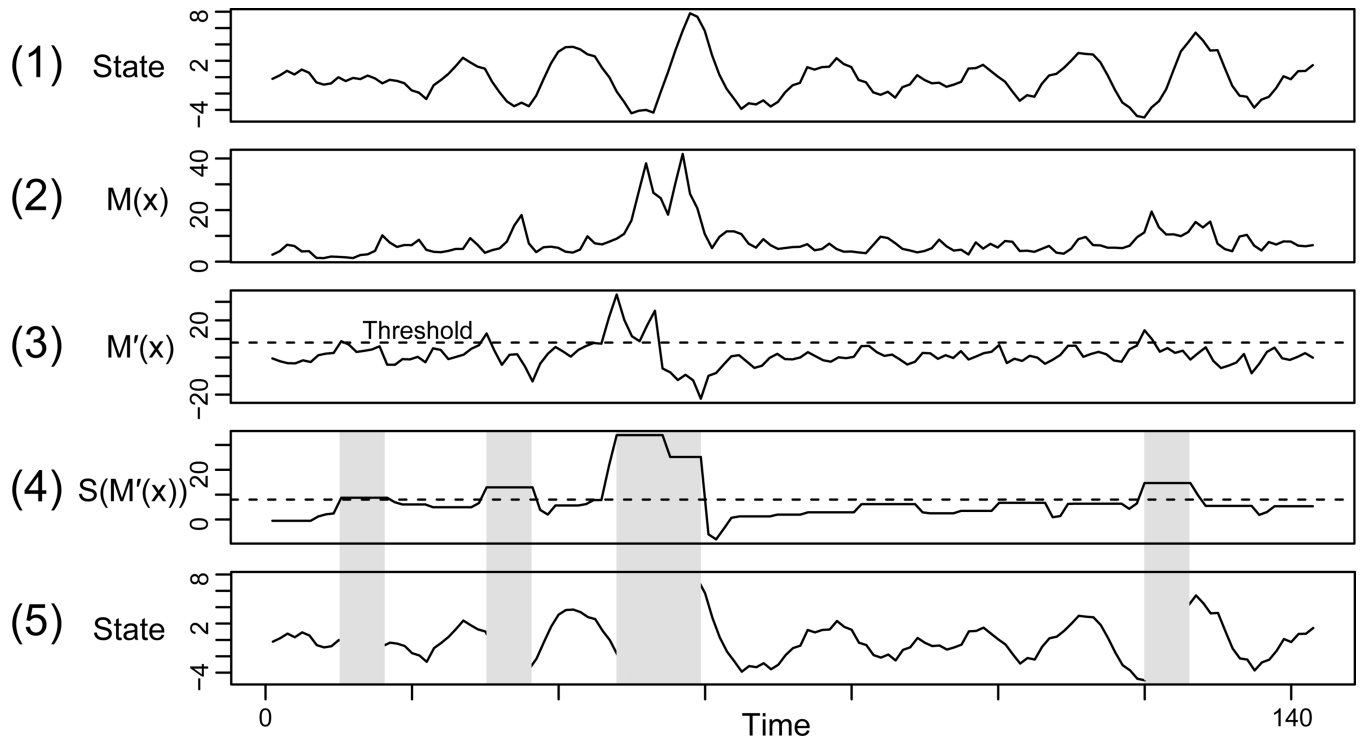
Figure 4.
Steps for outlier / discontinuity classification using the Mahalanobis distance
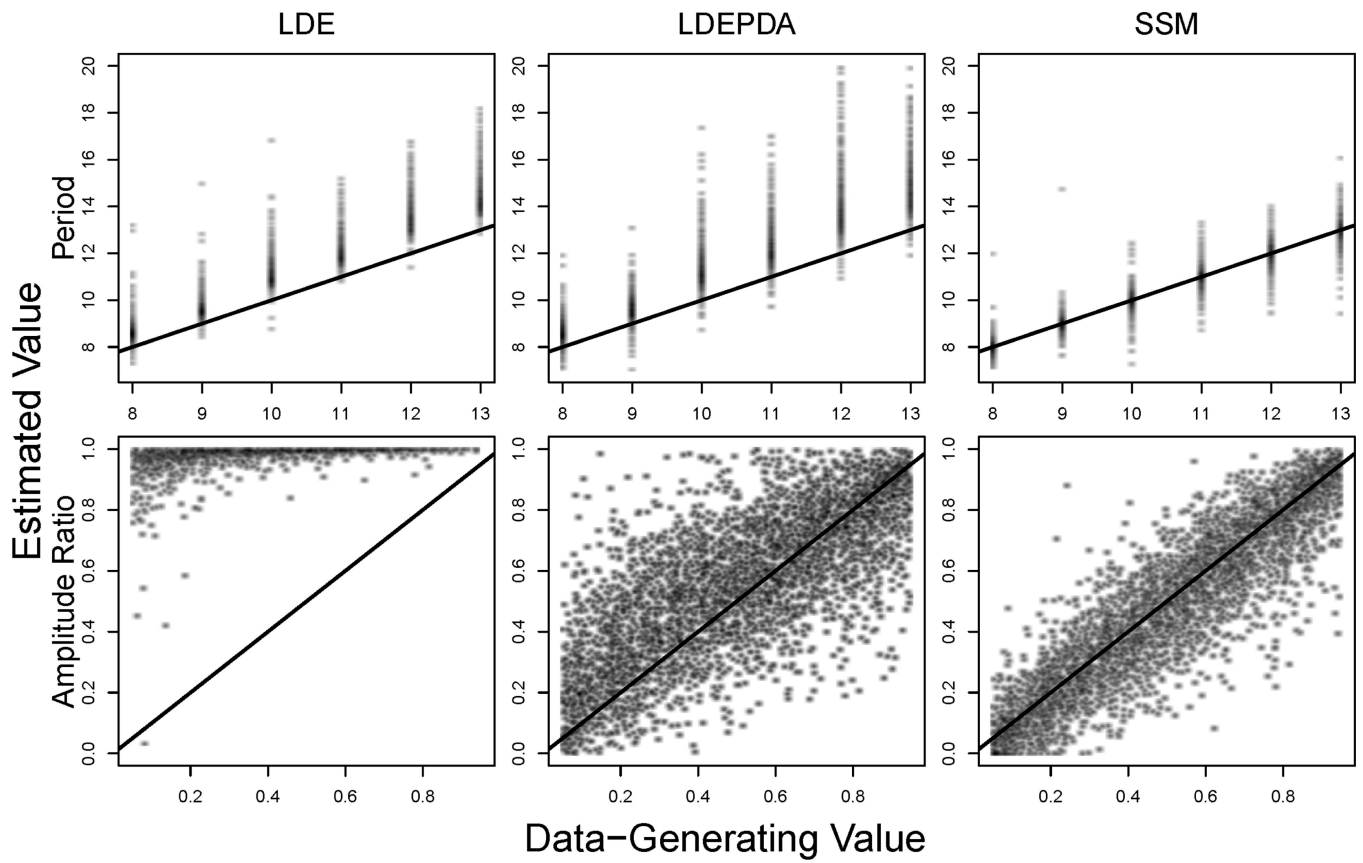
Figure 5.

Results vignette 1, diffusion process: Period 14, $\sigma_\epsilon \geq \frac{1}{4}$. LDEPDA is robust to diffusion

processes with high measurement error when the period and embedding dimension are small.
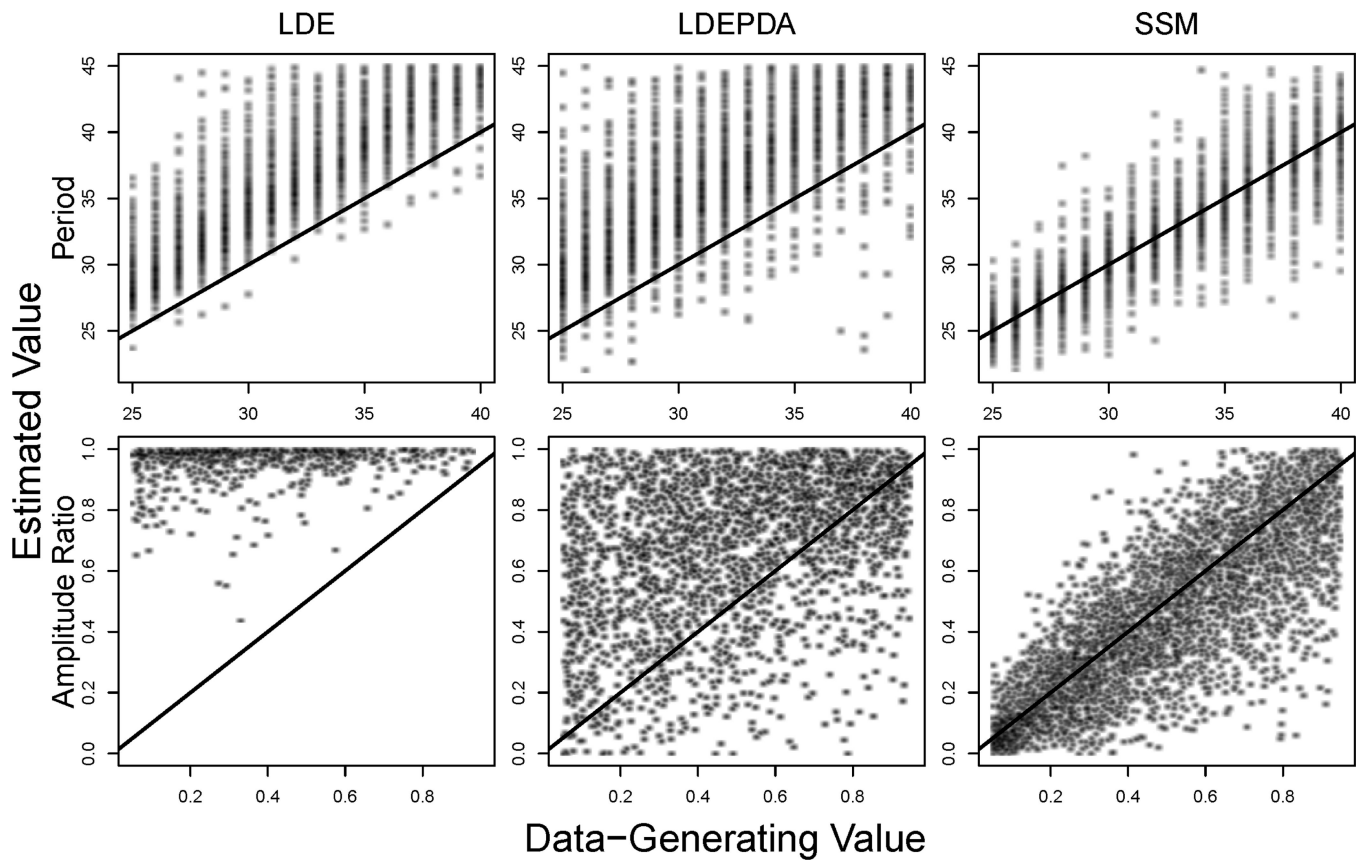
Figure 6.

Results vignette 2, diffusion process: Period > 24, $\sigma_\epsilon \le \frac{1}{8}$. LDEPDA fails when the period and embedding dimension are large, even if measurement error is low.
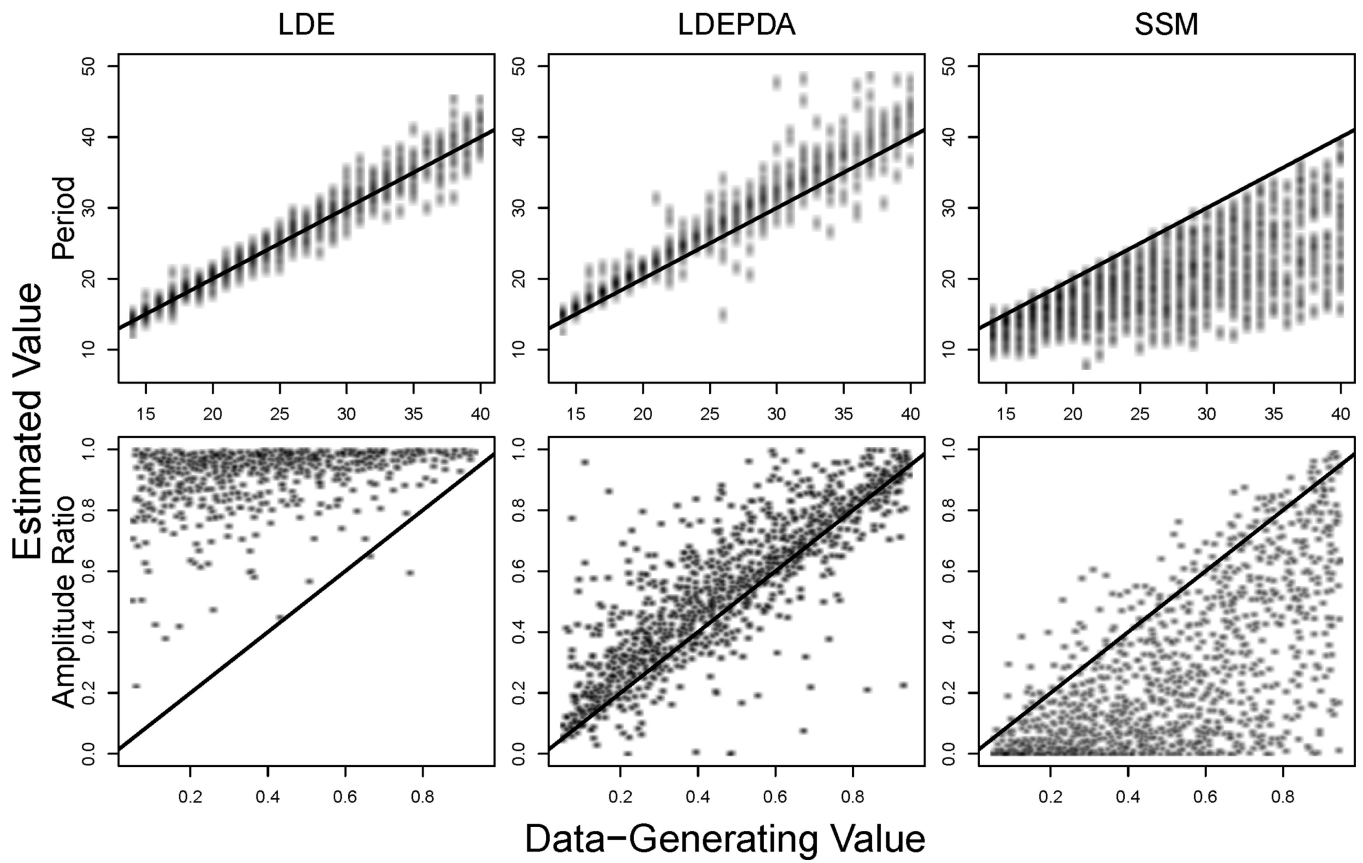
Figure 7.

Results vignette 3, shot noise process: Period > 14, $\sigma_\epsilon \leq \frac{1}{8}$, $p \leq .15$. SSM fails when process

noise behaves as infrequent jump discontinuities and the period is large. LDEPDA had low bias and variance.
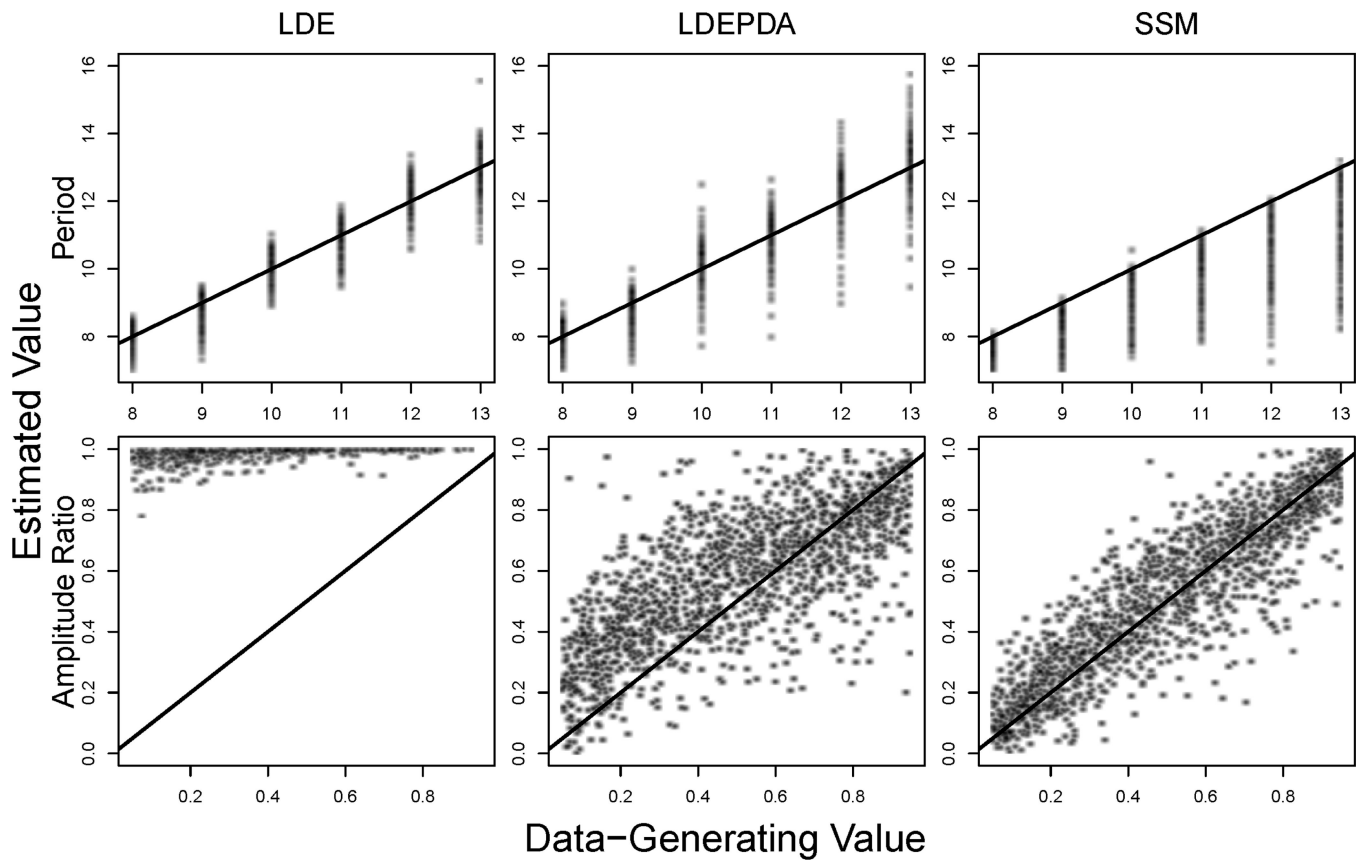
Figure 8.

Results vignette 4, shot noise process: Period ≤ 14, $\sigma_\epsilon \leq \frac{1}{8}$, $p \geq .5$. At higher frequencies, jump discontinuities begin to behave similarly to a diffusion process. Variance increases and curvilinear bias appears in LDEPDA estimates of amplitude ratio. SSM estimates amplitude ratio accurately, but with bias to period.
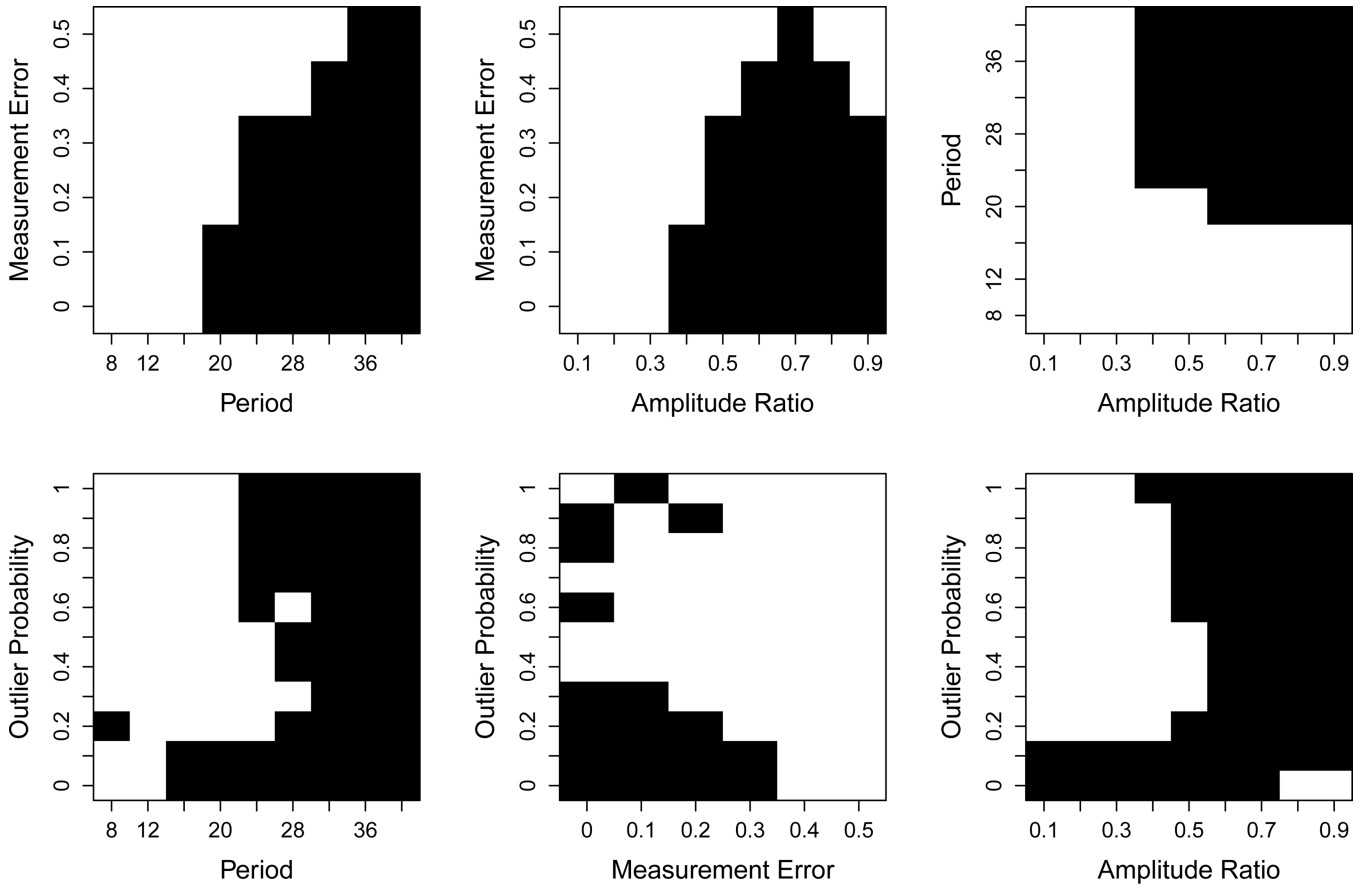
Figure 9.
Overall model performance comparisons of damping estimate accuracy for shot noise processes. Each condition was binned into 10 cells. Cells colored black represent parts of the condition and parameter domain where the LDEPDA damping estimates were more frequently closer to the data-generating values than the the SSM estimates. When the data represented a shot noise process, LDEPDA outperformed SSM for larger oscillation period, smaller damping effects, low outlier probability, and low measurement error. Grid comparisons of diffusion processes are not shown because LDEPDA did not outperform SSM under any conditions.
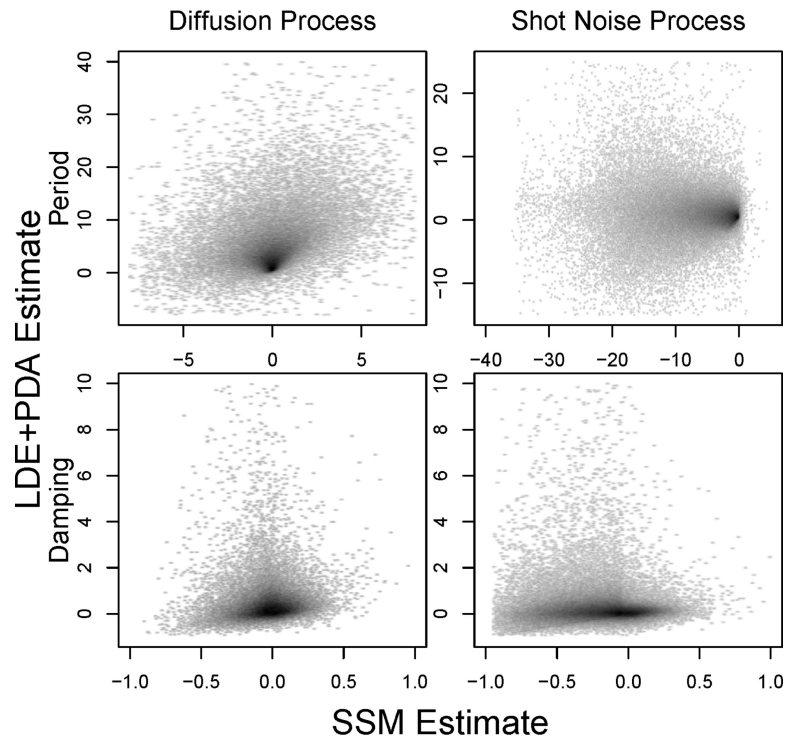
Figure 10.
Estimate error covariance plots comparing the estimate error distributions of the LDEPDA and the SSM.
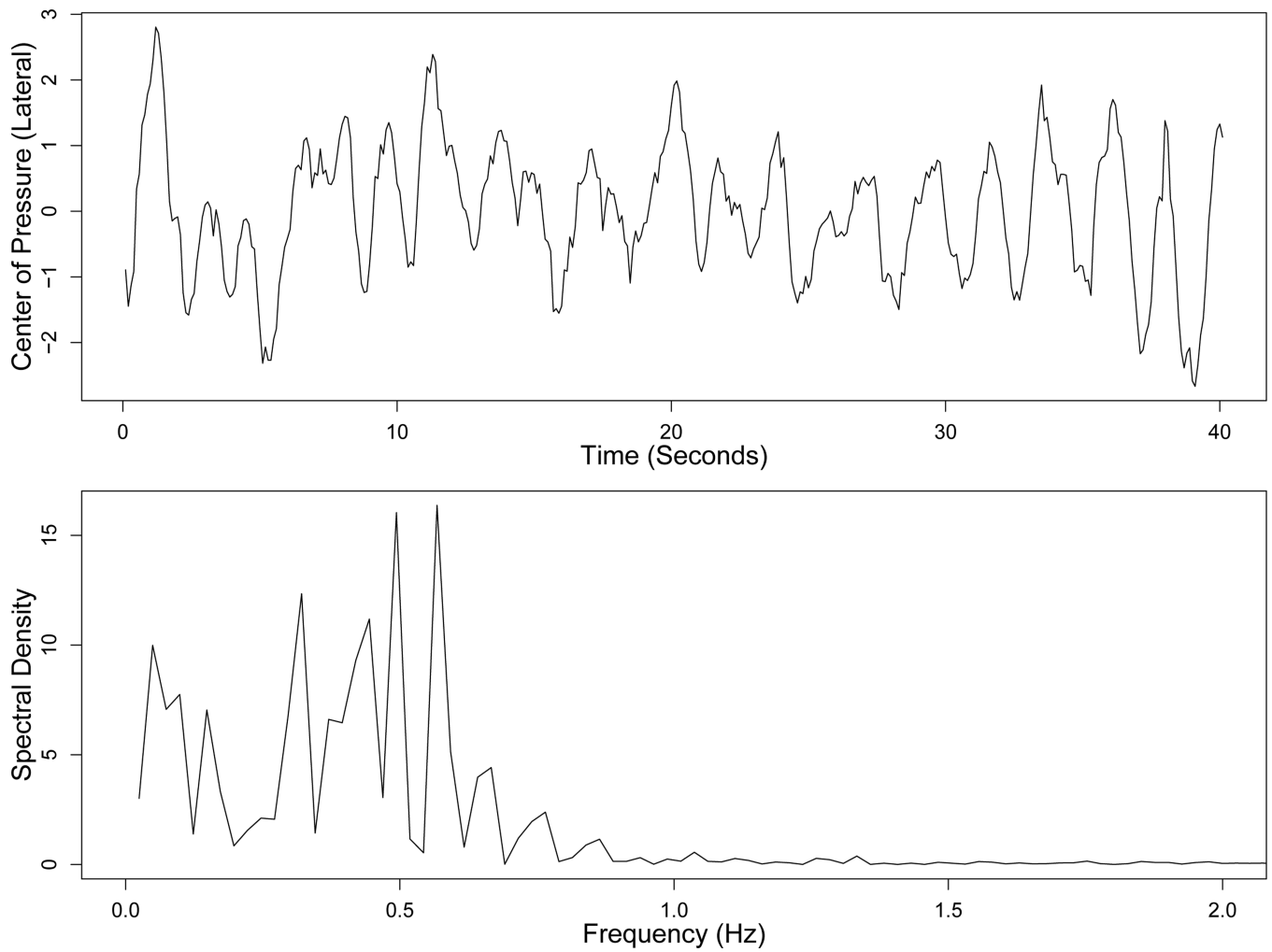
Figure 11.
Top: Lateral postural sway data sampled at 10Hz for 40 seconds. Bottom: Spectral Density plot of the same data showing peaks around 0.5Hz.

Table 1

Estimates of frequency and damping obtained from six past studies, with damping converted to ratio of amplitude over successive peaks. Studies without asterisks used the univariate LDE. The others used variants such as the coupled LDE or GOLD, which may have slightly different properties regarding the estimation of damping.

Study	$\hat{\eta}$	$\hat{\zeta}$	Amplitude Ratio
[**]Steele and Ferrer (2011)	−0.3700	−0.0420	0.8050
[**]	−0.5760	−0.0030	0.9877
[**]	−0.4100	0.0040	1.0198
[**]	−0.4480	−0.0340	0.8525
[**]	−0.5440	0.0120	1.0524
[**]	−0.6110	−0.0220	0.9154
Helm et al. (2012)	−0.0040	−0.0010	0.9515
	−0.0040	−0.0010	0.9515
	−0.0108	−0.0007	0.9791
	−0.0100	−0.0020	0.9391
	−0.0090	−0.0010	0.9674
	−0.0141	−0.0006	0.9843
	−0.0630	0.0010	1.0126
	−0.0490	0.0020	1.0288
	−0.0650	−0.0010	0.9878
	−0.0840	0.0010	1.0109
	−0.0470	0.0010	1.0146
	−0.0620	0.0010	1.0127
[**]Zentall et al. (2006)	−0.088	−0.053	0.573
[**]	−0.173	0.003	1.023
Chow et al. (2005)	−0.830	−0.010	0.966
	−0.830	−0.010	0.966
	−0.840	−0.020	0.934
	−0.840	−0.010	0.966
	−0.840	−0.010	0.966
	−0.940	−0.010	0.968
Gasimova et al. (2014)	−0.061	−0.004	0.950
[*]	−0.010	0.000	1.013
[*]	−0.025	0.000	1.004
Bisconti et al. (2004)	−0.015	−0.018	0.630

[*]
LDE Variants: Results from GOLD,

[**]
Results from coupled oscillator

Table 2

Parameters and settings for the simulated calibration data set

Parameter	Values
Series length N (Occasions)	100, 200
Measurement Error (σ_e)	.0625
Period λ (Occasions)	8, 20
Damping r (Amplitude ratio)	.25, .75
Outlier probability	.125, .25, .5

Table 3

Simulation parameters and settings

Parameter / Setting	Values
Series length N (Occasions)	$U(80, 500) \in \mathbb{Z}$
Measurement Error (σ_ϵ	$U(0.03125, 0.5) \in \mathbb{R}$
*Process noise (σ_w^2)	0.3
Period λ (Occasions)	$U(8, 40) \in \mathbb{Z}$
Damping r (Amplitude ratio)	$U(0.1, 0.9) \in \mathbb{R}$
**Outlier probability	$U(0, 1) \in \mathbb{R}$

*
Diffusion processes only

**
Shot noise processes only

Table 4

Rates of non-convergence for all three models. Status code 0 indicates successful optimization, 5 indicates a non-convex Hessian, and 6 indicates a non-zero gradient with no possible improvement.

Diffusion Process

Code	LDE	LDEPDA	SSM
0	42060 (98.66%)	42424 (99.51%)	42618 (99.96%)
5	321 (0.75%)	174 (0.41%)	15 (0.04%)
6	252 (0.59%)	35 (0.08%)	0 (0%)

Shot Noise Process

Code	LDE	LDEPDA	SSM
0	80869 (96.08%)	82739 (98.31%)	83362 (99.05%)
5	1742 (2.07%)	618 (0.73%)	522 (0.62%)
6	1554 (1.85%)	808 (0.96%)	281 (0.33%)

Table 5

Coefficients of multiple regression for Period error $(\hat{\lambda} - \lambda)^2$ *and amplitude ratio error* $(\hat{r} - r)^2$ *onto the simulation conditions.*

Diffusion Process

	Amplitude ratio error $(\hat{r} - r)^2$			Period error $(\hat{\lambda} - \lambda)^2$ (Occasions)		
	LDE	**LDEPDA**	**SSM**	**LDE**	**LDEPDA**	**SSM**
Intercept	0.93910	0.18202	0.01156	1.00278	1.68523	1.15073
σ_ϵ	0.00431	−0.05445	0.00465	−0.14330	−0.19550	0.19767
N	−0.00054	−0.00007	−0.00013	0.00610	0.01100	−0.00604
λ	0.00482	0.00981	0.00092	0.35211	0.25444	0.14105
r	−1.04321	−0.36127	0.04019	−6.05156	−5.10173	−2.68161

Shot Noise Process

	Amplitude ratio error $(\hat{r} - r)^2$			Period error $(\hat{\lambda} - \lambda)^2$ (Occasions)		
	LDE	**LDEPDA**	**SSM**	**LDE**	**LDEPDA**	**SSM**
Intercept	0.82099	0.07163	−0.14472	−0.55648	−0.05119	4.57989
σ_ϵ	0.00421	0.01940	−0.10322	0.23471	0.52872	−3.89788
N	−0.00035	−0.00041	−0.00014	−0.00361	−0.00325	−0.00142
λ	0.00345	0.00625	0.00700	0.14088	0.21189	0.44433
r	−0.95630	−0.05849	0.21161	−0.36047	−0.61647	−11.73631
p	0.09937	0.03854	0.03838	0.00283	−1.15405	1.72661

Table 6

Model information and point estimates of frequency ($\hat{\eta}$) and damping ($\hat{\zeta}$) with estimated 95% quantile confidence intervals and conversions to Hz and peak amplitude ratio.

	LDE	LDEPDA	SSM
$\hat{\eta}$	−7.45 (−8.38, −6.53)	−7.43 (−8.75, −5.96)	−4.55 (−5.57, −4.60)
$\hat{\zeta}$	−0.01 (−0.44, 0.33)	−0.49 (−0.96, 0.21)	−0.65 (−0.98, −0.58)
Frequency (Hz)	0.43 (0.37, 0.46)	0.36 (0.27, 0.46)	0.28 (0.24, 0.30)
Amplitude Ratio r	0.99 (0.59, 1.47)	0.57 (0.34, 1.27)	0.39 (0.25, 0.44)
Rows	396	248	401
Observed Statistics	2376	1488	401
Parameters	8	7	7