Regularized Estimation of Multivariate Latent Change Score Models

Ross Jacobucci
University of Notre Dame

Kevin J. Grimm
Arizona State University

Regularized Estimation of Multivariate Latent Change Score Models

With longitudinal data, although a number of statistical frameworks are available for analysis, the use of structural equation modeling has become increasingly popular. Hypotheses about change trajectories, determinants of change, along with other components can be formulated as a sequence of models, each representing a specific theoretical formulation. For instance, it is common to first start with a No Change model, typically formulated as a model with only an intercept term. Next, a Linear Change model can be tested, generally resulting in a better fit to the data if respondents do in fact exhibit change across the time span. Finally, various nonlinear formulations can be tested, each representing a quantitative and qualitative difference with previously tested models. Each model in the sequence tests a specific hypothesis about the functional form of change.

Although testing a sequence of models, varying in their degree of complexity, is most commonly done using the latent growth curve modeling framework (McArdle & Epstein, 1987; Meredith & Tisak, 1990), the latent change score (LCS) framework (McArdle, 2001; McArdle & Hamagami, 2001) has been used in an increasing amount of research over the last decade. The LCS model can be used to test a rich array of linear and nonlinear models, allowing researchers to examine different influences of change across time. Additionally, the LCS framework can be extended to the bivariate space (and beyond) to assess the reciprocal relationship between changes in constructs over time. It is because of these rich extensions beyond the latent growth curve modeling framework that we focus this chapter on exploring and examining additional extensions of the LCS framework. Before our discussion of various extensions that incorporate regularization, we provide a brief overview of both univariate and bivariate latent change score models.
Latent Change Score Framework

The LCS framework for studying longitudinal change shares similarities with both the autoregressive cross-lagged model (Jöreskog, 1970, 1974; Usami, Hayes, & McArdle, 2016) and the latent curve model (McArdle & Epstein, 1987; Meredith & Tisak, 1990). The LCS framework allows for the examination of both linear and nonlinear change across time (age) in the univariate case, with extensions to examining change both within and between constructs in the case of multivariate LCS models (McArdle, 2001).

Traditional testing with the LCS framework includes examining models that include No Change, Constant Change, Proportional Change, and finally, a combination of both proportional and constant effects in the Dual Change Model (McArdle & Hamagami, 2001). Beyond these four univariate models, few extensions have been proposed. Proposed extensions include the Changes to Changes models (Grimm, An, McArdle, Zonderman, & Resnick, 2012), Latent Acceleration models (Hamagami & McArdle, 2007), and Bayesian estimation for both the univariate and bivariate LCS models (Hamagami, Zhang, & McArdle, 2009).

To describe the model, we begin with true score theory where an observed score at age \( t \) \((Y[age]_n)\) for respondent \( n \) is composed of a theoretical true score \((y[age]_n)\) and a unique score \((u[age]_n)\). This can be written as
\[
Y[age]_n = y[age]_n + u[age]_n. \tag{1}
\]

In this model, change over age (time) can be seen as a difference between the true score at the current age minus true score at the age in the prior year \((age - 1)\), such that
\[
\Delta y[age]_n = y[age]_n - y[age - 1]_n. \tag{2}
\]

Across the entirety of the age span, total change manifests itself as the initial true level plus the summation of all previous true changes, such that
where $y[0]_n$ is the initial true level and $\sum_{r=1}^{age} (\Delta y[r]_n)$ is the summation of changes until the current $age$.

**Univariate Latent Change Score Model.** Although we omit detail regarding all of the possible specification of the univariate LCS model, we follow the specification guidelines from Grimm, An, McArdle, Zonderman, and Resnick (2012). This involves specifying four different models: (a) the proportional change model where $\Delta y[age]_n = \pi \cdot y[age - 1]_n$, where $\pi$ is an estimated autoproportion parameter; (b) constant change where $\Delta y[age]_n = \alpha_{2n}$ where $\alpha_{2n}$ is the constant change component; (c) the dual change model which is a combination of both (a) and (b) where $\Delta y[age]_n = \alpha_{2n} + \pi \cdot y[age - 1]_n$; and (d) the changes to changes model where prior changes influence subsequent changes is added to model (c) $\Delta y[age]_n = \alpha_{2n} + \pi \cdot y[age - 1]_n + \phi \cdot \Delta y[age - 1]_n$. Each of the four models allows for different influences of change and different degrees of nonlinearity.

**Bivariate Latent Change Score Model.** A fully specified bivariate model, for variables $X$ and $Y$, can be written as

$$
\Delta y[t]_n = \alpha_y \cdot s_{yn} + \pi_y \cdot y[t - 1]_n + \phi_y \cdot \Delta y[t - 1]_n + y_{yx} \cdot x[t - 1]_n + \xi_{yx} \\
\Delta x[t]_n = \alpha_x \cdot s_{xn} + \pi_x \cdot x[t - 1]_n + \phi_x \cdot \Delta x[t - 1]_n + y_{xy} \cdot y[t - 1]_n + \xi_{xy}
$$

where the univariate changes to changes model is specified in addition to coupling parameters, $\gamma$ and $\xi$, where effects from the other variable are included. The two coupling parameters add influence from the previous time and the prior latent change of the other variable, respectively.
This results in fitting seven models: (a) no coupling; (b) only $\gamma_{yx}$; (c) only $\gamma_{xy}$; (d) both $\gamma_{yx}$ and $\gamma_{xy}$; (e) model d + $\xi_{yx}$; (f) model d + $\xi_{xy}$; (g) model d + $\xi_{yx}$ and $\xi_{xy}$. A path diagram of the bivariate changes to changes model is depicted in Figure 1. Parameter labels in Figure 1 correspond to the specification in Equation 4, with latent variables depicted as circles and observed variables as squares. Following Grimm and colleagues (2012), the best fitting univariate model from each scale is then used in the bivariate LCS models, which tests the coupling between constructs $X$ and $Y$.

Choosing a final model is typically done using one of multiple information criteria, such as the Akaike Information Criteria (AIC; Akaike, 1973), the Bayesian Information Criteria (BIC; Schwarz, 1978), and the sample size-adjusted Bayesian information criterion (aBIC; Sclove, 1987). There are deficiencies in using information criteria to choose among LCS models (Usami, Hayes, & McArdle, 2016); however, using absolute fit indices can be difficult because in many cases the covariance coverage is low, preventing the calculation of such fit indices based on a saturated model.

The choice of a final bivariate (and univariate) model tells a very distinct story regarding the dynamic relationship between the variables of interest. For example, if model $b$ fits best, this leads us to believe that the previous level of $X$ exerts a significant influence on subsequent change in $Y$, whereas not including the same relationship from $Y$ to changes in $X$ does not significantly impact model fit, leading us to believe this relationship may not hold in the population. Additionally, the same interpretation holds if either models $e$ or $f$ fit best, but instead pointing to the significant influence that previous change in one variable has on subsequent change on the other variable. Despite the complexity of each bivariate model, each subsequent
model only examines the influence of one additional parameter, allowing us to examine nuanced relationships between both X and Y.

Exploratory Model Specification and Testing

With longitudinal SEM, most work has been focused on confirmatory modeling – the sequence of models is specified \textit{a priori}, each corresponding to a competing theoretical formulation of change. Less work has focused on exploratory modeling. The work that has focused on exploratory modeling has been focused in the context of latent growth curve models (McArdle & Epstein, 1987; Meredith & Tisak, 1990). In this, models can be tested with anywhere from a linear model with fixed factor loadings from the slope, to the latent basis growth model, which can be seen as an exploratory approach to determining the optimal shape of development (Grimm, Steele, Ram, & Nesselroade, 2013).

Despite work on exploratory modeling of change, increasing the flexibility \textit{across} models, not just \textit{within}, has received far less research. For example, testing the no change, linear change, and nonlinear change models represents a \textit{discrete} form of model testing, with no flexibility in testing models that lie in between each of these formulations. This form of testing can be altered to introduce flexibility into the model testing sequence by constraining individual parameters in each model. For instance, maximum likelihood estimation (MLE) of the linear growth curve model adds freely estimated mean and variance parameters to the latent variable slope factor. The addition of constraints on the mean and variance of the slope factor results in a model that exists somewhere in the continuum between a \textit{no change} and \textit{linear change} models. Where in this continuum depends on the level of constraint, a smaller constraint results in parameter estimates closer to \textit{linear change} model, whereas a greater constraint is closer to the \textit{no change} model.
One particular scenario that proves problematic when using discrete model selection is when one of the models fails to converge. For instance, when the amount of incompleteness is high, the LCS model has increasing amounts of nonconvergence (O'Rourke, Grimm, & MacKinnon, 2017). The question that remains is how to interpret a model that hasn’t converged. Particularly in the case when the most complex model (nonlinear change) does not converge, does this mean that linear change is the best fitting model? This dilemma is purely a product of discrete model fitting, where there is a large jump in complexity from one model to the next in the sequence. By increasing the flexibility in model testing, allowing for a continuum of models to be tested, it provides the opportunity to glean the maximal amount of information from a dataset.

**Rationale**

To overcome a number of these limitations in LCS model testing, we propose the incorporation of regularization (described below) for constraining parameters within the LCS model to allow for simpler, more flexible model testing. Particularly using the dual change score model, this allows us to test models ranging from highly nonlinear, to linear, to the intercept only models. Ultimately, this simplifies the model testing procedure. Furthermore, this approach is particularly useful in models where parameters are typically constrained to be equal over time, which is often the case in LCS models. For example, the autoproportion parameter is often constrained to be equal over time (in the dual change model) and it is difficult to detect efficiently where invariance may or may not hold. The same is true for bivariate LCS models as the autoproportion and coupling parameters are often constrained to be equal over time.

In this chapter, we introduce multiple ways for regularization to be used with the LCS model. One of these is a proposal of a penalized difference approach, where penalties are applied
to the difference between the constrained estimates (parameters estimates are the same across time) and freely estimated effects. We also demonstrate the use of penalizing parameters in the bivariate LCS models. In this, we use a hierarchical approach to test the range of models, ranging from *no* to *full* coupling, with one, instead of four models (or seven if including changes to changes coupling). To understand this form of model testing, we first present an overview of regularization and how this can be incorporated into both frequentist and Bayesian SEM estimation.

Regularization

The question that remains in testing models that lie on the continuum between traditionally specified models is how to systematically estimate models with constraints on specific parameter estimates. In the past, users of LISREL were able to place interval restrictions on parameter estimates (Rindskopf, 2012). However, this did not represent a systematic method that relied on a fit criterion to choose among many possible restrictions that could be placed on a parameter. In contrast, the method of regularization tests a sequence of models, each placing a varying degree of constraint on individual parameter estimates, while typically choosing among the sequence of constraints by examining performance on a holdout sample. The two most common forms of regularization are *Ridge Regression* (Hoerl & Kennard, 1970) and the *Least Absolute Shrinkage and Selection Operator* (Lasso; Tibshirani, 1996). In the case of using ordinary least squares regression, Lasso estimates can be found by optimizing the expression

$$\sum_{i=1}^{n} \left( Y_i - \beta_0 - \sum_{p=1}^{P} \beta_p X_{ip} \right)^2 + \lambda \sum_{p=1}^{P} |\beta_p| \tag{5}$$

In Equation 5, the left side of the expression is the traditional residual sum of squares from OLS regression. $\lambda$ is a constant, such that $\lambda \geq 0$. In Lasso regression, the $l_1$-norm (absolute value) is
used, instead of $l_2$-norm (squared) as in Ridge. In Lasso, the $\beta$ parameters shrink and, in many cases, $\beta$ parameters are shrunken all the way to zero; thus performing a form of subset selection (for more detail, see Hastie, Tibshirani, & Friedman, 2009; Tibshirani, 1996). Although there are a number of sparser penalties available (see Jacobucci, 2017), one worth detailing is the adaptive lasso (Zou, 2006). In this, each penalized parameter is scaled by its maximum likelihood estimate, which overcomes the propensity for the lasso to over-constrain large parameters (Fan & Li, 2001).

In the context of frequentist regularization, both Lasso and Ridge regression has been extended for use with SEMs, termed regularized structural equation modeling (RegSEM; Jacobucci, Grimm, & McArdle, 2016). RegSEM builds upon the traditional maximum likelihood (ML) fit function for SEM models,

$$F_{ML} = \log(|\Sigma|) + tr(C \cdot \Sigma^{-1}) - \log(|C|).$$  \hspace{1cm} (6)

and adds an element to the cost function

$$F_{regsem} = F_{ML} + \lambda P(\cdot).$$  \hspace{1cm} (7)

where $\lambda$ is the regularization parameter, taking on a value between zero and infinity, and $P(\cdot)$ is a general function for summing parameters. When $\lambda$ is zero, ML estimation is performed.

Particularly in more complex SEM models, due to the highly constrained nature of regularization, RegSEM convergence can be problematic. It is in this domain that Bayesian forms of regularization may be most useful. In the same way that frequentist regularization has been extended to SEM from regression, Bayesian regularization has been detailed in the context of regression (e.g. Kyung, Gill, Ghosh, & Casella, 2010; Park & Casella, 2008) as well as in SEM (Guo, Zhu, Chow, & Ibrahim, 2012; Feng, Wu, & Song, 2017; Jacobucci & Grimm, 2017; Lu, Chow, & Loken, 2016; Muthén & Asparouhov, 2012).
For Bayesian estimation, specific to our purpose is the recent implementation of Mplus' (Muthén & Muthén, 1998-2015) Bayesian SEM (Muthén & Asparouhov, 2012) small variance priors. Similar in a sense to using previous research to inform the choice of the prior distribution, the use of small variance priors that are centered around zero restricts the influence of non-target (e.g. cross-loadings) parameters. Small variance priors can lead to less biased estimates compared to when small parameter estimates, not of empirical interest, are inappropriately constrained to zero. This allows for more flexibility in model creation, where, through the use of restrictive priors, unimportant or uninformative parameters can be pared away.

The use of small variance priors in Bayesian estimation has been formulated equivalently with regularization in frequentist regression (Kyung et al., 2010; Park & Casella, 2008, Tibshirani, 1996). That is, decreasing the prior variance in Bayesian estimation is equivalent to increasing the level of penalization in frequentist regularization. In the case of Ridge regression, a small variance prior with a normal distribution is used to constrain estimates and the Ridge estimate is the mean of the posterior distribution (Park & Casella, 2008; Tibshirani, 1996). In the case of Lasso regression, a small variance prior with a Laplace distribution is used to constrain estimates and the Lasso estimate is the mode of the posterior distribution (Park & Casella, 2008; Tibshirani, 1996).

Although there are a number of different formulations of Bayesian regularization, for the purposes of the applications in this chapter, we will focus on the hierarchical adaptive Lasso prior, in line with Feng, Wu, and Song (2017), of which we will term the Bayesian adaptive Lasso (BaLasso). A hierarchical model is one in which a prior on a specific parameter is further given a probabilistic specification, known as a hyperparameter (see Chapter 5 in Gelman, Carlin, Stern, & Rubin, 2014). In contrast to testing multiple different prior variances (e.g. 30 models) to
find a best-fitting final model, the use of a hierarchical prior allows for testing only one model, using the data to explicitly estimate the optimal prior on the parameters of interest.

**Purpose**

Our goal in pairing the LCS model with regularization is to demonstrate a more flexible approach to model testing. This will be demonstrated through the use of three examples, each demonstrating different types of penalties. The first adds penalties to the proportional change parameter in the dual change model, testing varying degrees of nonlinearity to the expected trajectory. The second example proposes a two-step procedure to capture the difference between allowing the proportional change parameter to be freely estimated across time and constraining the proportional change parameter to be invariant across time. This procedure allows us to test whether the assumption of constraining this parameter over time holds. In the third example, we add penalties to the coupling parameters in the bivariate LCS model, necessitating the testing of only one, instead of multiple models. Programming scripts for each example are available from both author’s websites.

**Univariate LCS Regularization**

To provide a more concrete example of the use of both types of regularization, we examine both BaLasso and RegSEM methods with the dual change model and longitudinal data from the Wechsler Intelligence Scale for Children dataset \((N = 204;\) Osborne & Suddick, 1972), which contains four measurement occasions (ages 6, 7, 9, and 11 years). For this analysis, we used the verbal scale. This examination can be conceptualized as testing whether the addition of the proportional change parameter, which increases the nonlinearity in the individual trajectory, is necessary and allows us to test models ranging from the constant change model to the dual change model. All parameters in the dual change model are either freely estimated or given
diffuse priors\textsuperscript{1} (see Hamagami, Zhang, & McArdle, 2009), outside of the penalties (RegSEM) and small variance priors (BaLasso) added to the proportional change. In RegSEM, the fit function becomes

\[ F_{\text{regsem}} = F_{ML} + \lambda |\pi|. \] (8)

Starting with an unpenalized model ($\lambda = 0$), we tested 40 penalty values, ranging from $\lambda = 0$ to $\lambda = 3.9$ in increments of .1. For the BaLasso, the prior on $\pi$ was

\[ \pi \sim N(0, \psi_j \tau_j^2) \]

\[ \tau_j^2 \sim \text{Gamma}(1, \gamma_j^2) \] (9)

\[ \psi_j^{-1} \sim \text{Gamma}(\alpha, \beta) \]

\[ \gamma_j^2 \sim \text{Gamma}(\alpha, \beta) \]

With $\alpha$ set to 1 and $\beta$ as .05 following the recommendations by Feng, Wu, and Song (2017). This prior on a prior, resulting in a gamma mixture of normals, capitalizes on the fact that the Laplace distribution can be expressed as a scale mixture of normal distributions with independent exponentially distribution variances (i.e. Gamma with $\alpha = 1$; Andrews & Mallows, 1974). In comparison to the Mplus small variance priors, which put a normal distribution prior on each parameter, the formulation in Equation 9 results in mean posterior distribution estimates that are closer to zero, while also allowing some estimates to be closer to what they would be with diffuse priors. One can think of this as a diffuse prior for significant parameters that influence the fit of the model, and highly constrained priors for those parameters that are not influential. Models were estimated in the R statistical environment (R Core Team, 2017) using the \texttt{regsem} package (Jacobucci, 2017) for the RegSEM Lasso model, and the \texttt{rjags} package (Plummer, 2016) to interface \texttt{JAGS} (Plummer, 2003) for the BaLasso model.
The maximum likelihood and the Bayesian estimation with diffuse prior results are displayed in Table 1. For the RegSEM results, the BIC only got worse (higher) as the value of the penalty increased and the model failed to converge when the penalty reached 2.2. Thus, using RegSEM, we concluded that the unpenalized dual change model was the best fitting model. For BaLasso, the model converged after 100,000 samples with thinning equal to 3. The mean estimates from the posterior distributions are displayed in Table 1.

Table 1
Parameter Estimates for the Dual Change Score Model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>MLE</th>
<th>Diffuse Bayes</th>
<th>BaLasso</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>19.17</td>
<td>19.17</td>
<td>19.22</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-1.84</td>
<td>-1.88</td>
<td>-1.34</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.43</td>
<td>0.43</td>
<td>0.40</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>18.35</td>
<td>18.23</td>
<td>18.40</td>
</tr>
<tr>
<td>$\sigma_2^2$</td>
<td>1.14</td>
<td>1.16</td>
<td>1.06</td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>-2.91</td>
<td>-2.84</td>
<td>-2.37</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>11.19</td>
<td>11.41</td>
<td>11.46</td>
</tr>
<tr>
<td>Fit</td>
<td>BIC = 5028</td>
<td>DIC=4586</td>
<td>DIC = 4587</td>
</tr>
</tbody>
</table>

Note: BIC = the Bayesian Information Criterion; DIC = the Deviance Information Criterion; $\alpha$ is the mean of the intercept ($\alpha_1$) and the slope ($\alpha_2$); $\sigma$ is the variance of the intercept ($\sigma_1^2$), the slope ($\sigma_2^2$), and the covariance between slope and intercept ($\sigma_{12}$); $\sigma^2$ is the residual variance at each time point, constrained to be equal across time.

The interesting thing to note is how the mean of the constant change component ($\alpha_2$) was shrunken towards the estimate from the constant change model, 4.67 (other parameters not displayed). This occurs due to the dependence between the proportional change parameter and the mean of the slope. As the proportional change parameter was shrunken towards zero, the mean of the slope correspondingly moved towards the parameter estimate in the constant change model. As a result, the BaLasso model only resulted in a small amount of shrinkage towards the constant change model. Given the results from both RegSEM and BaLasso, we concluded that
the dual change model was appropriate for this data, as the BaLasso model led to a small amount of shrinkage and RegSEM chose a model equivalent to the unpenalized maximum likelihood estimation. Given this, there is an important proportional effect from the previous time point leading to the subsequent change in addition to the additive effect on change from the constant change component. This results in a nonlinear model, comprising both linear and proportional change.

In a second step, we tested whether the nonlinear change was influenced from the previous time point (proportional change), the previous change (changes to changes), or both (model $d$ in the univariate formulation). This involves testing a fully specified univariate model, and penalizing both the autoproportion and changes to changes parameters. However, similar to penalizing the autoproportion parameter, the best fitting model, across regularization methods, kept both parameters as non-zero. As a result, we do not present these results further.

**Time-Varying Effects Regularization**

The problems with non-convergence in the LCS model creates problems not only in model selection, but also in parameter estimation. Furthermore, the proportional change parameters are often held invariant across time, but this assumption is rarely challenged or tested. This constraint is imposed for theoretical interpretation, but also to limit the number of freely estimated parameters as a way to prevent problems with estimation. To overcome these limitations when fitting LCS models, we propose a two-step procedure that we term the *penalized difference* approach. In this, we can use either Bayesian or frequentist forms of regularization.

This two-step process can be accomplished in the following manner. In the first step, the autoproportion parameters are constrained to be equal across time, resulting in a single estimate.
In the second step, a deviation parameter is created, capturing the difference between the time-varying parameter estimate and the time-constrained estimate at each time point. Thus, any parameter estimate that differs meaningfully from the invariant estimate should have a deviation parameter that is non-zero. This can be used as a final, more flexible model, or to inform subsequent analyses that incorporate adjustments to the initial model to take into account the time-varying effects discovered in the second step. Note that one alternative procedure is to use the estimate from step one as the prior mean for each parameter in the second step. We prefer to use the penalized difference approach as it captures the amount that each autoproportion parameter deviates from the invariance assumption.

To examine this two-step procedure, we conducted a small scale simulation study. We simulated data adhering to the dual change model with six time points and a sample of 500. The proportional change parameter was simulated to have an estimate of 0 for the first two change factors, and 0.2 for the last three change factors. Using 100 replications, we compared two models using both frequentist and Bayesian regularization. In the first step, the invariance model constrained the autoproportion parameter to be equal over time. In the second model, the parameter estimate for the autoproportion parameter in the first model was used to create a deviation parameter. In this, each proportional change parameter was

\[ \pi_j = \pi_1 + dev_j. \]

\[ dev_j \sim N(0,.01) \]

(10)

where \( \pi_j \) is the autoproportion parameter at time \( j \), \( \pi_1 \) is the autoproportion parameter from model 1, and \( dev_j \) is the deviation parameter at time \( j \). For the prior on \( dev_j \) we chose a small variance normal distribution (Ridge) with a variance of .01. Our main concern was not deriving
posterior mean estimates at zero, but instead to capture an interpretable deviation parameter, while also removing bias from other model parameters.

For frequentist regularization, estimation was formulated as

\[ F_{\text{diffLasso}} = F_{\text{ML}} + \lambda |dev_j| \]  

where \( dev_j \) is the vector of deviations for each freely estimated \( \pi_j \) from the estimate of \( \pi \) in step 1 using ML. Although we could have used one of multiple types of penalty, we chose the Lasso for its ability to set the deviations to 0. In this, although the unconstrained estimates start at points away from \( \pi \) and their simulated estimates, as the penalty increases, these parameters become closer and closer to \( \pi \). Because of difficulties in deriving the effective degrees of freedom in this approach, we used five-fold cross-validation and the chi-square fit statistic to choose a final model.

The percent bias for each method is displayed in Table 2. For the autoproportion parameters, both the invariant model and the model that freely estimated each parameter had the largest amounts of bias. Difficulties in the estimation of these parameters resulted in biased estimates for the slope mean for each of the four models. For the entire model, both the Bayesian and diffLasso models has lower amounts of bias in comparison to both ML models. In comparing both of these models that used deviation parameters, Bayesian estimation resulted in higher bias for the slope variance, while the diffLasso model had high bias for the slope mean and for the first autoproportion parameter. However, differences in performance were almost negligible. Therefore, preference for one form of difference estimation can be given to software familiarity, as each require completely dissimilar frameworks for testing.

*Table 2*

Percent Bias for Each Method
REGULARIZED LATENT CHANGE SCORE MODELS

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Simulated</th>
<th>Inv.</th>
<th>Free</th>
<th>Bayes</th>
<th>diffLasso</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>1.00</td>
<td>-0.56</td>
<td>-0.39</td>
<td>-0.73</td>
<td>-0.03</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>1.00</td>
<td>-32.24</td>
<td>9.19</td>
<td>3.99</td>
<td>-5.10</td>
</tr>
<tr>
<td>$\pi_1$</td>
<td>0</td>
<td>26.07</td>
<td>-9.17</td>
<td>3.15</td>
<td>5.77</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>0</td>
<td>26.07</td>
<td>-4.62</td>
<td>4.42</td>
<td>2.82</td>
</tr>
<tr>
<td>$\pi_3$</td>
<td>0.20</td>
<td>6.07</td>
<td>-2.93</td>
<td>0.19</td>
<td>1.67</td>
</tr>
<tr>
<td>$\pi_4$</td>
<td>0.20</td>
<td>6.07</td>
<td>-2.10</td>
<td>0.61</td>
<td>0.98</td>
</tr>
<tr>
<td>$\pi_5$</td>
<td>0.20</td>
<td>6.07</td>
<td>-1.08</td>
<td>1.02</td>
<td>1.07</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>0.50</td>
<td>-7.87</td>
<td>0.32</td>
<td>-0.98</td>
<td>-1.72</td>
</tr>
<tr>
<td>$\sigma_2^2$</td>
<td>0.50</td>
<td>-17.41</td>
<td>12.14</td>
<td>6.47</td>
<td>-0.46</td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>0.10</td>
<td>-5.54</td>
<td>2.40</td>
<td>0.74</td>
<td>-1.35</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>1.00</td>
<td>0.75</td>
<td>0.32</td>
<td>-0.29</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Note: $\alpha$ is the mean of the intercept ($\alpha_1$) and the slope ($\alpha_2$); $\sigma$ is the variance of the intercept ($\sigma_1^2$), the slope ($\sigma_2^2$), and the covariance between slope and intercept ($\sigma_{12}$); $\sigma^2$ is the residual variance at each time point, constrained to be equal across time. Inv. refers to the model with invariant autoproportion parameters, Free to the model with freely estimated autoproportion parameters, Bayes and diffLasso refer to both Bayesian and RegSEM estimation, respectively, of the penalized difference method.

**Bivariate LCS Regularization**

To demonstrate the utility of conducting regularized estimation with the bivariate LCS model, we used both the verbal and performance scales from the WISC dataset. Traditionally, in the bivariate LCS model, the following sequence of models are tested: no coupling, X to Y coupling, Y to X coupling, and dual coupling. Omitting the $\xi$ parameters, the dual coupling model is depicted in Figure 1. Once these four models are tested, one or multiple of various fit indices are used to choose a best fitting model. However, as mentioned before, this model testing procedure can be hampered by non-convergence of specific solutions, leaving researchers with uncertainty regarding what model is best, and what this says about the underlying relationship between processes.

The model testing procedure that we propose for bivariate LCS models is similar to what was used with the dual change model. Instead of penalizing the proportional change parameter,
we add BaLasso\textsuperscript{2} priors to both coupling parameters. This can be seen as a hierarchical form of testing, one that encompasses the four models traditionally used in the model testing sequence. If either or both of the coupling parameters are estimated as non-significant, with mean posterior estimates near zero, this leads us to believe that this directional parameter is unnecessary to understand the dynamic relationship between the two constructs. More so, this one step procedure simplifies the model testing process for researchers, while allowing a greater degree of flexibility in constraining the model to adhere to previous knowledge in the subject area.

Additionally, we included the RegSEM adaptive lasso. The sequence of penalties was added to both coupling parameters, using the BIC to choose among this sequence of models.

The same prior specification as depicted in Equation 9 was used for both coupling parameters, while all other parameters were estimated using diffuse priors as specified in Hamagami, Zhang, and McArdle (2009). The results are displayed in Table 3.

Table 3
Parameter Estimates for the Bivariate LCS Model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>\textit{Balasso}</th>
<th>\textit{RegSEM ALasso}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>95% Int</td>
</tr>
<tr>
<td>( \alpha_{v_1} )</td>
<td>19.29</td>
<td>18.59, 20.00</td>
</tr>
<tr>
<td>( \alpha_{v_2} )</td>
<td>-2.27</td>
<td>-4.48, -.02</td>
</tr>
<tr>
<td>( \alpha_{p_1} )</td>
<td>18.37</td>
<td>17.15, 19.61</td>
</tr>
<tr>
<td>( \alpha_{p_2} )</td>
<td>1.27</td>
<td>-1.19, 3.51</td>
</tr>
<tr>
<td>( \pi_v )</td>
<td>.42</td>
<td>.22, .62</td>
</tr>
<tr>
<td>( \pi_p )</td>
<td>-.37</td>
<td>-.54, -.19</td>
</tr>
<tr>
<td>( \sigma_{v_1}^2 )</td>
<td>17.43</td>
<td>13.02, 22.06</td>
</tr>
<tr>
<td>( \sigma_{v_2}^2 )</td>
<td>.99</td>
<td>.33, 1.75</td>
</tr>
<tr>
<td>( \sigma_{v_1,v_2} )</td>
<td>-2.75</td>
<td>-5.01, -.56</td>
</tr>
<tr>
<td>( \sigma_{p_1}^2 )</td>
<td>56.51</td>
<td>42.47, 70.82</td>
</tr>
<tr>
<td>( \sigma_{p_2}^2 )</td>
<td>4.93</td>
<td>2.52, 7.37</td>
</tr>
<tr>
<td>( \sigma_{p_1,p_2} )</td>
<td>2.76</td>
<td>-2.07, 7.90</td>
</tr>
<tr>
<td>( \sigma_{v_1,p_1} )</td>
<td>26.36</td>
<td>19.99, 33.35</td>
</tr>
<tr>
<td>( \sigma_{v_2,p_2} )</td>
<td>.83</td>
<td>.07, 1.63</td>
</tr>
</tbody>
</table>
\[
\begin{align*}
\sigma_{V_1,P_2} & \quad -3.39 \quad -6.17, -.79 \quad -4.04 \\
\sigma_{V_2,P_1} & \quad -3.51 \quad -7.24, -.24 \quad -3.48 \\
\sigma^2_V & \quad 12.18 \quad 10.47, 13.87 \quad 11.68 \\
\sigma^2_P & \quad 22.13 \quad 19.33, 24.95 \quad 21.50 \\
\gamma_{V,P} & \quad .74 \quad .51, .97 \quad 0.80 \\
\gamma_{P,V} & \quad .03 \quad -.01, .18 \quad 0
\end{align*}
\]

Note: \( \alpha \) is the mean of the intercept (\( \alpha_1 \)) and the slope (\( \alpha_2 \)); \( \gamma_{V,P} \) is the coupling parameter from Verbal to Performance; \( \sigma \) is the variance of the intercept (\( \sigma^2_1 \)), the slope (\( \sigma^2_2 \)), and the covariance between slope and intercept (\( \sigma_{12} \)); \( \sigma^2 \) is the residual variance at each time point, constrained to be equal across time. For RegSEM ALasso, the BIC was used to choose a final model.

In examining the coupling parameters in Table 3, it is clear that the path from verbal to performance is much stronger than the reciprocal path. The coupling parameter from performance to verbal was estimated as near zero with the BaLasso, and was zero with RegSEM ALasso, whereas the other coupling parameter had a much larger estimate, which was significant according to the credible intervals. Therefore, according to this model, we concluded that prior levels of verbal ability influenced future changes in performance ability. More specifically, those that were higher in prior verbal ability had greater subsequent increases in performance ability.

**Discussion**

The purpose of this chapter was to introduce a number of novel methods that increase the flexibility in assessing longitudinal change with the LCS framework. We proposed three different strategies: penalizing the proportional change parameter, assessing invariance of the autoproportion parameter by using a penalized difference approach, and finally a single hierarchical model to test the presence of coupling in the bivariate LCS model. Our focus is less on the three specific examples, and more on the introduction of a new way of thinking about model evaluation. The use of both frequentist and Bayesian regularization methods opens the door to many more types of application than just the three detailed here. Regularization allows
for more flexibility, in both simplifying the model testing process, while promoting the incorporation of more complex specifications.

With the three proposed methods that alter the way of testing LCS models, additional simulation work is required to understand the settings in which the methods work or don’t work. Particularly with the bivariate coupling, simulation studies are warranted to test whether using sparse hierarchical priors captures the true model better than discrete model testing. Specifically, given that the influence of a prior is determined by the sample size, this hierarchical approach is expected to demonstrate variability in performance across both small and large sample sizes.

Across the three examples, one limitation to RegSEM was problems with convergence, derived from pairing constrained estimation with a model that was highly constrained to begin with. As the RegSEM optimization methods (see Jacobucci, 2017) improve, we expect increased pairing of RegSEM with the various LCS models. In the second example, with penalizing the difference with the time-constrained parameter estimate, a frequentist penalization method has been worked out with RegSEM. This entails directly adding the difference between the time-constrained estimate and current parameter estimate to Equation 11. However, a number of difficulties remain in implementing this approach. In the future, particularly with large sample sizes, we expected RegSEM to hold an advantage in this form of testing given the ability of RegSEM Lasso to push parameter estimates all the way to zero, thus simplifying the process of variable/parameter selection.

The use of the hierarchical BaLasso is just one implementation of hierarchical Bayesian regularization methods. In comparison to the BaLasso, two specific methods that accomplish setting parameter estimates to zero are the spike-and-slab (Ishwaran & Rao, 2005; Lu, Chow, & Loken, 2016; O’Hara & Sillanpää, 2009) and Horseshoe (Carvalho, Polson, & Scott, 2009)
priors (see Jacobucci & Grimm, 2017 for a comparison in SEM). Particularly with large models, setting parameters directly to zero could be advantageous, due to simplifying inference. We expect an increased incorporation of Bayesian regularization into structural equation models in the future.

In conclusion, this chapter introduced a framework for incorporating regularization into the LCS framework. Across both univariate and bivariate LCS models, using both frequentist and Bayesian regularization allowed for a simplified process in choosing a best model, while also increasing the plasticity of the LCS model to incorporate additional parameterizations. We expect that researchers are able tailor their specific research questions to incorporate aspects of the three examples presented. The three novel ways of pairing regularization and the LCS model represent a small subset of possible applications and modifications. We look forward to seeing additional proposals in this arena going forward.
Footnote

1 Note that in this specification, outside the context of regularization, both the frequentist and Bayesian specifications are expected to give similar results (Hamagami, Zhang, & McArdle, 2009).

2 We solely use the BaLasso with the Bivariate model given problems with RegSEM and model convergence. The complex, highly constrained nature of the Bivariate LCS model make frequentist estimation particularly problematic.
References


Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible
representation of substantive theory. Psychological Methods, 17, 313-335.
doi: 10.1037/a0026802

Angeles, CA: Muthén & Muthén

What, how and which. Bayesian Analysis, 4, 85–11

change score models. Manuscript under review.


Association, 103(482), 681–686.

sampling. In Proceedings of the 3rd international workshop on distributed statistical
computing (Vol. 124, p. 125).

https://CRAN.R-project.org/package=rjags

R Core Team. (2017). R: A language and environment for statistical computing [Computer
software manual]. Vienna, Austria.

Rindskopf, D. (2012). Next steps in Bayesian structural equation models: Comments on,
variations of, and extensions to Muthe´n and Asparouhov (2012). Psychological
Methods, 17, 336–339. doi:10.1037/a0027130


Figure 1

Diagram of the Bivariate Latent Change Score Model