Biostatistics (2020) **21**, 4, *pp.* 676–691
doi:10.1093/biostatistics/kxy082
Advance Access publication on January 29, 2019

A two-stage estimation procedure for non-linear structural equation models

KLAUS KÄHLER HOLST*,†

Department of Biostatistics, University of Copenhagen, Øster Farimagsgade 5, entr. B, P.O.Box 2099, DK-1014 Copenhagen K, Denmark, Neurobiology Research Unit, Rigshospitalet, Copenhagen University Hospital, Juliane Maries Vej 28, building 6931, 3rd floor, DK-2100 Copenhagen, Denmark, and Maersk, Esplanaden 50, DK-1098 Copenhagen K, Denmark

klaus.holst@maersk.com

ESBEN BUDTZ-JØRGENSEN†

Department of Biostatistics, University of Copenhagen. Øster Farimagsgade 5, entr. B, P.O.Box 2099, DK-1014 Copenhagen K, Denmark

Summary

Applications of structural equation models (SEMs) are often restricted to linear associations between variables. Maximum likelihood (ML) estimation in non-linear models may be complex and require numerical integration. Furthermore, ML inference is sensitive to distributional assumptions. In this article, we introduce a simple two-stage estimation technique for estimation of non-linear associations between latent variables. Here both steps are based on fitting *linear* SEMs: first a linear model is fitted to data on the latent predictor and terms describing the non-linear effect are predicted by their conditional means. In the second step, the predictions are included in a linear model for the latent outcome variable. We show that this procedure is consistent and identifies its asymptotic distribution. We also illustrate how this framework easily allows the association between latent variables to be modeled using restricted cubic splines, and we develop a modified estimator which is robust to non-normality of the latent predictor. In a simulation study, we compare the proposed method to MLE and alternative two-stage estimation techniques.

Keywords: Latent variable; Neuroimaging; Non-linear estimation; Two-stage estimator.

1. Introduction

Over the last decades linear structural equation models have been useful in many fields of research. These models typically consists of two parts, a measurement part where observed outcomes are assumed to be reflections of underlying latent variables and a structural part relating the latent variables to each other. Important extensions of this framework have used more flexible measurement models to allow inclusion of binary, ordinal, and censored outcomes (Muthén, 1984; Skrondal and Rabe-Hesketh, 2004). In this

*To whom correspondence should be addressed.
†These authors contributed equally to this work.

article, we focus on the structural part and consider models which allow for non-linear relations between the latent variables. Until now most research in this topic has been on the interaction model assuming a linear effect of the product of two latent variables (i.e., $\eta = \beta_1\xi_1 + \beta_2\xi_2 + \beta_3\xi_1\xi_2 + \zeta$) or more general polynomial models including terms of higher order (e.g., $\eta = \beta_1\xi_1 + \beta_2\xi_1^2 + \zeta$), but a more general framework is of obvious interest.

Maximum likelihood inference in linear SEMs is facilitated by the fact that a closed form expression for the likelihood function is obtained when integrating out the latent variables. Non-linear models do not have this property and numerical methods are needed. Today, the so-called LMS algorithm (Klein and Moosbrugger, 2000) is perhaps the most widely used method. It approximates the likelihood function using a mixture of multivariate normal distributions and then this function is maximized with the EM-algorithm. For simple non-linear models (product-interaction model and second degree polynomial), this algorithm has been implemented in the widely used software package *Mplus* (Muthén and Muthén, 2012). An alternative EM-algorithm was proposed by Lee and Zhu (2002), while Wall (2009) used Adaptive Gaussian Quadrature (AGQ) and Rizopoulos and Moustaki (2008) suggested a hybrid EM-algorithm based on AGQ. Bayesian techniques have also been considered for non-linear models (Arminger and Muthén, 1998; Lee and Zhu, 2000), and these have been extended to more flexible semi-parametric models (Yang and Dunson, 2010; Song *and others*, 2010; Kelava and Brandt, 2014).

Due to the complexity of the ML-procedure for estimation of non-linear SEMs, a number of simpler so-called limited information methods have been developed. Kenny and Judd (1984) developed the first estimator for the product-interaction model by fitting a modified linear model including an additional latent variable with indicators given by products of indicators of the two interacting latent variables. Since then, this technique has been refined by several researchers (see Marsh *and others*, 2004 for an overview), but it remains rather *ad hoc* and cannot be used for more general non-linear models. For the polynomial model, Wall and Amemiya (2000) proposed a two-step method (2SMM), where the latent variables were first predicted using the so-called Bartlett score, while the second step estimated the parameters of the non-linear relations using a method-of-moments procedure allowing for uncertainty in the predicted variables. Despite some nice statistical properties, this method has not been used much in practice. Mooijaart and Bentler (2010) developed a method-of-moments procedure by including third-degree moments for estimation of non-linear effects in the polynomial model. A computationally very simple two-stage least squares (2SLS) method was developed by Bollen (1995). Here instrumental variables must be identified also for non-linear terms. It is not clear whether this method can be used in general models, but the method is non-iterative and easy to implement in standard statistical software. However, simulation studies have indicated that this method has a rather low efficiency compared to ML-estimation (Schermelleh-Engel *and others*, 1998).

In this article, we present a new two-stage method for estimation in non-linear SEMs. As in 2SMM, we first have a prediction step, but instead of the Bartlett score we use the Empirical Bayes method and instead of predicting the latent variables we predict the latent non-linear effect terms. Therefore, in the second step, it is sufficient to fit a linear SEM with the predicted variables included as covariates. We show that the method yields consistent estimation and derive expressions for asymptotic standard errors. We illustrate how splines can be included and by using mixture models in the first step, we extent the method so that it becomes robust to non-normality of latent predictor variables. In simulation studies the method is compared to ML, 2SMM, 2SLS, and an alternative two-stage estimator of semi-parametric associations between latent variables (Kelava *and others*, 2017). Finally, we illustrate the usefulness of the method analyzing data from neuroscience on the regional binding potential of different serotonergic markers in the human brain.

2. A NON-LINEAR STRUCTURAL EQUATION MODEL

We consider a model where a latent response variable $\eta_i = (\eta_{i1}, ..., \eta_{ip})^t$ of subject i ($i = 1, \ldots, n$) is assumed to be non-linearly related to a latent predictor $\xi_i = (\xi_{i1}, ..., \xi_{iq})^t$ after adjustment for covariates $Z_i = (Z_{i1}, ..., Z_{ir})$

$$\eta_i = \alpha + B\varphi(\xi_i) + \Gamma Z_i + \zeta_i, \tag{2.1}$$

such that $\varphi(\xi_i) = (\varphi_1(\xi_i), ..., \varphi_l(\xi_i))^t$ has finite variance. The main parameter $B(p \times l)$ describes the non-linear relation between ξ_i and η_i. Note, that φ may also depend on some of the covariates thereby allowing the introduction of interaction terms, but we here omit this to simplify notation.

The latent predictors (ξ_i) are related to each other and the covariates in a linear structural equation

$$\xi_i = \tilde{\alpha} + \tilde{B}\xi_i + \tilde{\Gamma} Z_i + \tilde{\zeta}_i, \tag{2.2}$$

where diagonal elements in $\tilde{B}$ are zero and the residual terms ζ_i and $\tilde{\zeta}_i$ are assumed to be independent with mean 0 and covariance matrices of Ψ and $\tilde{\Psi}$, respectively.

The observed variables $X_i = (X_{i1}, ..., X_{ih})^t$ and $Y_i = (Y_{i1}, ..., Y_{im})^t$ are linked to the latent variable in the two measurement models

$$Y_i = \nu + \Lambda \eta_i + KZ_i + \epsilon_i \tag{2.3}$$

$$X_i = \tilde{\nu} + \tilde{\Lambda}\xi_i + \tilde{K}Z_i + \tilde{\epsilon}_i, \tag{2.4}$$

where the error terms ϵ_i and $\tilde{\epsilon}_i$ are assumed to be independent with mean 0 and covariance matrices of Ω and $\tilde{\Omega}$, respectively. The parameters are collected into $\theta = (\theta_1, \theta_2)$, where $\theta_1 = (\tilde{\alpha}, \tilde{B}, \tilde{\gamma}, \tilde{\nu}, \tilde{\Lambda}, \tilde{K}, \tilde{\Omega}, \tilde{\Psi})$ are the parameters of the linear SEM describing the conditional distribution of X_i given Z_i. The rest of the parameters are collected into θ_2.

For identification of the model, we need to impose some parameter constraints (Bollen, 1989). Generally, measurement models can be made identifiable by selecting a reference indicator for each latent variable. For this variable, we fix the regression coefficient of the latent variable (element of Λ or $\widetilde{\Lambda}$) to 1 and the intercept (element of ν or $\widetilde{\nu}$) to 0. Alternatively, the variance of latent variables can be fixed to 1, and their intercepts (element of α or $\widetilde{\alpha}$) set to 0. In the estimation procedure described below, it turns out to be crucial to use the reference indicator restriction in the measurement model for Y_i . Also, we model the covariance Ψ of the latent outcomes (η_i) using an unrestricted covariance matrix.

The likelihood function is $L(Y, X, Z, \theta) = \prod_{i=1}^{n} \int p_\theta(Y_i, X_i|\eta_i, \xi_i, Z_i) p_{\theta_2}(\eta_i|\xi_i, Z_i) p_{\theta_1}(\xi_i|Z_i) \mathrm{d}\xi_i \mathrm{d}\eta_i$. Assuming joint normality of $(\zeta_i, \tilde{\zeta}_i, \epsilon_i, \tilde{\epsilon}_i)'$, a closed form expression for L, is available only if φ is linear. In the general case numerical integration techniques are necessary for ML-estimation, which in practice for even moderately sized problems (number of latent variables) is computationally intractable. Instead, we exploit that the structural equation (2.1) is linear in the parameters to get

$$\mathbb{E}(\eta_i|X_i, Z_i) = \alpha + B\mathbb{E}_{\theta_1}[\varphi(\xi_i)|X_i, Z_i] + \Gamma Z_i, \tag{2.5}$$

noting that the conditional expectation on the right-hand side depends only on the distribution parametrized by θ_1. Equation (2.5) suggests that parameters can be estimated in two steps. First, the linear SEM given by equations (2.2) and (2.4) is fitted to $(X_i, Z_i, i = 1, \ldots, n)$ and the latent covariate $\varphi(\xi_i)$ is predicted by the conditional mean $E_{\widehat{\theta}_1}[\varphi(\xi_i)|X_i, Z_i]$. Step 2 then estimates θ_2 in a linear SEM, where the measurement model is given by equation (2.3) and the structural model is equation (2.1) with the latent predictor replaced by $\mathbb{E}_{\widehat{\theta}_1}[\varphi(\xi_i)|X_i, Z_i]$. Thus, the key idea is to replace $\varphi(\xi_i)$ by (an estimate of) the conditional mean of $\varphi(\xi_i)$

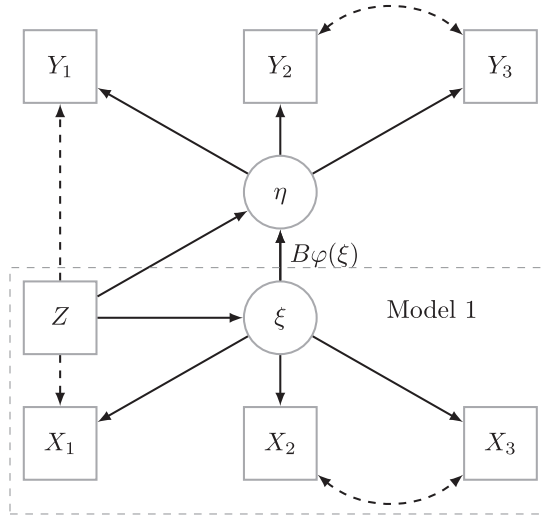

Fig. 1. Path diagram showing an example of the non-linear structural equation models considered. The two-stage estimator is constructed by separately estimating parameters of Model 1, and (non-linear) associations between the two models are then estimated in a subsequent step based on predictions from the Model 1 analysis.

given X and Z. Previous methods have used φ of the conditional mean $E_{\widehat{\theta}_1}[\xi_i|X_i, Z_i]$ (Jöreskog and Yang, 2000; Schumacker, 2002).

In the following, we will use the notation $\varphi^*(\xi_i) = \mathbb{E}[\varphi(\xi_i) \mid X_i, Z_i]$ and $\varphi_n^*(\xi_i) = \mathbb{E}_{\widehat{\theta}_1}[\varphi(\xi_i) \mid X_i, Z_i]$ to distinguish between the conditional expectation and the plug-in estimator, where the expectation is taken with respect to the distribution indexed by the estimated parameter values from the stage one model. To summarize, we define the two-stage estimator in the following way

REMARK 2.1 (Two-Stage Structural Equation Model estimator (2SSEM))

1. The linear SEM given by (2.2) and (2.4) is fitted to $(X_i, Z_i, i = 1, \ldots, n)$ to achieve an estimate of the parameter θ_1.
2. The parameter θ_2 is estimated via a linear SEM with measurement model given by (2.3) and structural model given by (2.1), where the latent predictor, $\varphi(\xi_i)$, is replaced by $\varphi_n^*(\xi_i)$.

We can now formulate the consistency properties of the proposed estimator (the asymptotic distribution of 2SSEM is derived in Supplementary Material D available at *Biostatistics* online).

THEOREM 1 Under a correctly specified non-linear SEM (2.1)–(2.4) including correctly specified distribution of the residual terms, 2SSEM will yield consistent estimates of all parameters (θ) except for the residual covariance, Ψ, of the latent variables in step two.

Proof. Since the exposure part of the model is correctly specified, θ_1 is estimated using ML-estimation and therefore $\widehat{\theta}_1$ is consistent under mild regularity conditions (Anderson and Amemiya, 1988). In step two, the model is misspecified as we fit a linear model to a non-linear association. We prove the theorem by showing that even though the model is misspecified it includes the true mean and covariance of the data. When fitting a linear SEM with correctly specified mean and variance, the estimator $\widehat{\theta}_2$ will converge to the

parameter value which induces the true mean and variance (Arminger and Schoenberg, 1989). Finally, we characterize this parameter value and see that it is identical to the truth (θ_2) except for elements describing the residual covariance of the latent variables in Step 2 (Ψ). We make these arguments assuming that the predicted values φ^* were available. The result then follows from the Continuous Mapping Theorem and by noting that $\varphi_n^* \longrightarrow \varphi^*$ *a.s.* as $n \to \infty$. The proof is illustrated in the situation where $K = 0$, as this simplifies matrix expressions without affecting theoretical insights (see the Appendix of the supplementary material available at *Biostatistics* online for a more general formulation).

The next step is to show that the model fitted in step-two induces the correct mean and variance structure for the data $V_i = (Z_i^t, \varphi^*(\xi_i)^t, Y_i^t)^t$. To save space, we illustrate this only for the variance (full proof is given in Appendix of the supplementary material available at *Biostatistics* online). In step two, we fit a linear model with correct measurement part ($Y_i = \nu_2 + \Lambda_2\eta_i + \epsilon_i$) and structural part $\eta_i = \alpha_2 + B_2\varphi^*(\xi_i) + \Gamma_2 Z_i + \zeta_i$. Here the subscript 2 is used to distinguish the parameters of step-two model from the true parameter value (no subscript). Using standard results from linear models (Bollen, 1989), and the notation $\Sigma_{z,\varphi} = \Sigma_{\varphi,z}^t$, the modeled variance is

$$\Sigma_{V_i} = \begin{pmatrix} \Sigma_Z & \Sigma_{z,\varphi} & [\Sigma_{z,\varphi}B_2^t + \Sigma_z\Gamma_2^t]\Lambda_2^t \\ . & \Sigma_\varphi & [\Sigma_\varphi B_2^t + \Sigma_{\varphi,z}\Gamma_2^t]\Lambda_2^t \\ . & . & \Lambda_2\Sigma_\eta\Lambda_2^t + \Omega_2 \end{pmatrix}, \tag{2.6}$$

where $\Sigma_\eta = B_2\Sigma_\varphi B_2^t + B_2\Sigma_{\varphi,z}\Gamma_2^t + \Gamma_2\Sigma_{z,\varphi}B_2^t + \Gamma_2\Sigma_Z\Gamma_2^t + \Psi_2$ is the modeled variance of η and Σ_Z, $\Sigma_{z,\varphi}$, Σ_φ are completely unstructured parameters modeling the variance of the step-two covariates $(Z_i, \varphi^*(\xi_i))$. To calculate the true variance, we write the structural model as

$$\eta_i = \alpha + B\varphi(\xi_i) + \Gamma Z_i + \zeta_i = \alpha + B\varphi^*(\xi_i) + B[\varphi(\xi_i) - \varphi^*(\xi_i)] + \Gamma Z_i + \zeta_i, \tag{2.7}$$

and from this we derive the variance of the latent variable η_i

$$\begin{aligned} \mathbb{V}\text{ar}(\eta_i) = {} & B\mathbb{V}\text{ar}[\varphi^*(\xi_i)]B^t + B\mathbb{C}\text{ov}[\varphi^*(\xi_i), Z_i]\Gamma^t + \Gamma\mathbb{C}\text{ov}[Z_i, \varphi^*(\xi_i)]B^t \\ & + B\mathbb{V}\text{ar}[\varphi(\xi_i) - \varphi^*(\xi_i)]B^t + \Gamma\mathbb{V}\text{ar}(Z)\Gamma^t + \Psi, \end{aligned} \tag{2.8}$$

because $\mathbb{C}\text{ov}[\varphi^*(\xi_i), \varphi(\xi_i) - \varphi^*(\xi_i)] = 0$ and $\mathbb{C}\text{ov}[Z_i, \varphi(\xi_i) - \varphi^*(\xi_i)] = 0$, which follows from the law of iterated expectations. We can now determine the covariances $\mathbb{C}\text{ov}[\varphi^*(\xi_i), Y_i] = \mathbb{C}\text{ov}[\varphi^*(\xi_i), \Lambda\eta_i] = \mathbb{C}\text{ov}[\varphi^*(\xi_i), B\varphi^*(\xi_i) + \Gamma Z_i]\Lambda^t$ as $\mathbb{C}\text{ov}[\varphi^*(\xi_i), \varphi(\xi_i) - \varphi^*(\xi_i)] = 0$ and $\mathbb{C}\text{ov}(Z_i, Y_i) = \mathbb{C}\text{ov}(Z_i\Lambda\eta_i) = \mathbb{C}\text{ov}[Z_i, B\varphi^*(\xi_i) + \Gamma Z_i]\Lambda^t$ as $\mathbb{C}\text{ov}[Z_i, \varphi(\xi_i) - \varphi^*(\xi_i)] = 0$. Finally, the variance is

$$\mathbb{V}\text{ar}(V_i) = \begin{pmatrix} \mathbb{V}\text{ar}(Z_i) & \mathbb{C}\text{ov}[Z_i, \varphi^*(\xi_i)] & [\mathbb{C}\text{ov}\{Z_i, \varphi^*(\xi_i)\}B^t + \mathbb{V}\text{ar}(Z_i)\Gamma^t]\Lambda^t \\ . & \mathbb{V}\text{ar}[\varphi^*(\xi_i)] & [\mathbb{V}\text{ar}\{\varphi^*(\xi_i)\}B^t + \mathbb{C}\text{ov}\{\varphi^*(\xi_i), Z_i\}\Gamma^t]\Lambda^t \\ . & . & \Lambda\mathbb{V}\text{ar}(\eta_i)\Lambda^t + \Omega \end{pmatrix}. \tag{2.9}$$

It can now be seen that the modeled variance is equal to the true variance [(2.6) = (2.9)] and the modeled mean is equal to the true mean (equations not shown) if $B_2 = B, \Lambda_2 = \Lambda, \Gamma_2 = \Gamma, \Omega_2 = \Omega, \alpha_2 = \alpha, \nu_2 = \nu, \mu_\varphi = \mathbb{E}[\varphi(\xi_i)], \mu_Z = \mathbb{E}(Z_i), \Sigma_{z,\varphi} = \mathbb{C}\text{ov}[Z_i, \varphi^*(\xi_i)], \Sigma_\varphi = \mathbb{V}\text{ar}[\varphi^*(\xi_i)], \Sigma_Z = \mathbb{V}\text{ar}(Z_i)$, and $\Psi_2 = \Psi + B\mathbb{V}\text{ar}[\varphi(\xi_i) - \varphi^*(\xi_i)]B^t$. Note, that although the expression for the modeled variance of the latent variable (Σ_η) is missing the term $B\mathbb{V}\text{ar}[\varphi(\xi_i) - \varphi^*(\xi_i)]B^t$, the model can achieve the correct variance for Y_i by adding the missing term to the residual variance Ψ_2. Since the model includes the true mean and variance, $\widehat{\theta_2}$ will converge to the specific parameter value inducing this mean and variance. Therefore, all parameters are consistently estimated except for the variance of the latent variable, which will be overestimated. □

An important advantage of the proposed method is that closed form expressions for $\varphi^*(\xi_i)$ are available for large classes of functions including polynomials and splines basis functions (see Supplementary Material B available at *Biostatistics* online) making the implementation of the estimator straightforward and computationally fast. Also, note that the formulation of Theorem 1 is overly restrictive, as it states that consistency relies not only on a correctly specified model structure, but also on correctly specified distributions of residuals. However, linear SEMs only require a correctly specified mean and variance (conditional on covariates) to yield consistent estimation (Arminger and Schoenberg, 1989) and therefore 2SSEM will be robust to distributional misspecifications. Thus, non-normality of the residuals of the measurement models in step two will not affect consistency. However, for $\widehat{\theta}_2$ to be consistent, we note that the following conditions must hold: (a) $\mathbb{E}[\varphi(\xi_i) - \varphi^*(\xi_i)] = 0$, (b) $\mathbb{C}\text{ov}[\varphi^*(\xi_i), \varphi(\xi_i) - \varphi^*(\xi_i)] = 0$, (c) $\mathbb{C}\text{ov}[Z_i, \varphi(\xi_i) - \varphi^*(\xi_i)] = 0$. These are obviously fulfilled when the conditional mean $\mathbb{E}[\varphi(\xi_i) \mid X_i, Z_i]$ is correctly modeled. To address this, both the stage 1 and stage 2 model fits should be assessed using standard model checking tools for linear SEMs (Sánchez *and others*, 2009). Here the critical distribution is that for the residual $\widetilde{\zeta}_i$ of the latent exposure, whereas correct specification of error terms in the measurement model of the stage one model is less important as shown in simulations of Section 3. In Section 2.1, we describe an extended method which allows flexible models for the distribution of the latent exposure variable.

A few consequences of the calculations in the proof of Theorem 1 are important to note. Firstly, it is important that the step-two model uses an unstructured covariance matrix (Ψ_2) for the variance of the latent variable. If that is not the case, it may be impossible for the model to account for the misspecification of Σ_η ($\Sigma_\eta \neq \mathbb{V}\text{ar}(\eta_i)$), therefore the modeled variance of $V_i = (Z_i^t, \varphi^*(\xi_i)^t, Y_i^t)^t$ may be wrong and estimation will likely become inconsistent. In particular, the model should not be made identifiable by fixing the variance of the latent variable to one. Secondly, covariates that are used in the step-two model, must also be present in the first step. If that is not the case, $\mathbb{C}\text{ov}[Z_i, \varphi(\xi_i) - \varphi^*(\xi_i)]$ may not be zero and the step-two model will have an incorrect variance leading to inconsistent estimation. Thirdly, in the step-two model, the covariance terms $\mathbb{C}\text{ov}[Z, \varphi^*(\xi_i)]$ and $\mathbb{V}\text{ar}[\varphi^*(\xi_i)]$ are modeled using unstructured matrices so that any information these terms might have had about the parameters is disregarded in this approach. Finally, note that in Theorem 1 it is assumed that variables in the stage one model affect the variables in the stage two model only indirectly through the latent variable. In the presence of a direct effect from one of the indicators, X_i, either on the latent variable or the indicators of the stage two model, the proposed method can easily be extended by simply including the relevant indicators as covariates in the stage two model. The consistency of this approach is proven in Supplementary Material A available at *Biostatistics* online.

It is interesting to compare our method to the two-stage method-of-moments (2SMM) of Wall and Amemiya (2000). Here, predictions $\widehat{\xi}$ and $\widehat{\eta}$ of the latent variables are calculated from a confirmatory factor analysis model and then the second step fits a non-linear regression model $\widehat{\eta} = \beta\varphi(\widehat{\xi}) + \zeta$ allowing for uncertainty in $\widehat{\xi}$. For the latter task, a method moments estimator is used, but it works only for polynomial models. Our method is different from 2SMM in two important ways. Rather than predicting ξ in step one, we predict the non-linear terms $\varphi(\xi)$ and therefore we are left with a linear model in step two. Of course the predicted terms are different from the latent true terms, so, as in 2SMM, we have measurement error in covariates in the second step. Here the choice of prediction method in step one becomes important. We use the conditional mean $\varphi^*(\xi_i) = \mathbb{E}[\varphi(\xi)|X, Z]$ which has Berkson errors, that is the prediction in uncorrelated with the prediction error $\mathbb{C}\text{ov}[\varphi^*(\xi_i), \varphi(\xi_i) - \varphi^*(\xi_i)] = 0$. In linear regression models, it is well-known that Berkson errors will not bias regression coefficients (Carroll *and others*, 2006). In step two, we estimate parameters using a linear model and therefore the Berkson errors do not lead to inconsistency as we show in Theorem 1. In contrast, Bartlett predictions (used in 2SMM) have classic error and therefore adjustments are needed in order to achieve unbiased estimation.

2.1. *Extension to mixtures of structural equation models*

In this section, we extend the structural part of the model to allow for non-normal latent predictor variables. This is done through a mixture model. Thus, let $G_i \sim \text{multinom}(\pi)$ be the class indicator $G_i \in \{1, \ldots, K\}$, and ξ_i the q-dimensional latent predictor $\xi_i = \sum_{k=1}^{K} I(G_i = k)\xi_{ki}$, where each component ξ_{ki} follows a linear structural equation

$$\xi_{ki} \quad = \quad \tilde{\alpha}_k + \tilde{B}\xi_{ki} + \tilde{\Gamma} Z_i + \tilde{\zeta}_{ki}, \tag{2.10}$$

with $\tilde{\zeta}_{ki} \sim N(0, \tilde{\Psi}_k)$. We assume G_i to be independent of $(\tilde{\zeta}_{1i}, ..., \tilde{\zeta}_{Ki})$ and $(\epsilon_i, \tilde{\epsilon}_i)$. Results can be extended also to the case where π depends on covariates. Note, that the extension concerns only the conditional distribution of the latent predictor given the covariates. Thus, the only parameters that depend on k is the intercept and the variance in the structural model for ξ, whereas the measurement models and the structural model for η_i remain as shown in equations (2.1, 2.3, 2.4) independent of k.

In the extended model, the two-step procedure is modified by the fact that ML-estimation in step 1 will be more complex (typically done via the EM-algorithm) and the predicted latent variables now become

$$\begin{aligned} \mathbb{E}[\varphi(\xi_i)|X_i, Z_i] &= \mathbb{E}\{\mathbb{E}[\varphi(\xi_i)|X_i, Z_i, G_i]|X_i, Z_i\} \\ &= \textstyle\sum_{k=1}^{K} P(G_i = k|X_i, Z_i)\, \mathbb{E}[\varphi(\xi_{ki})|X_i, Z_i, G_i = k] \\ &= \textstyle\sum_{k=1}^{K} P(G_i = k|X_i, Z_i)\, \mathbb{E}[\varphi(\xi_{ki})|X_i, Z_i], \end{aligned} \tag{2.11}$$

where the last step uses the fact that G_i is independent of $(\zeta_{ki}, \tilde{\epsilon}_i)$. So the predictions of step 1 are the sum of the product of the posterior probabilities (class probabilities which are by-products of the EM-algorithm) and predictions of the type described in Supplementary Material B available at *Biostatistics* online. Of course, the second step in the estimation procedure is unchanged: a linear SEM is fitted with the predictions included as covariates.

3. SIMULATION STUDY

In this section we explore the finite sample properties of 2SSEM. We first considered scenarios where all distributional assumptions of the model were fulfilled. Next, we explored the robustness of the proposed methods in misspecified models and finally we explored 2SSEM in a non-parametric setting using a flexible spline model. The Monte Carlo simulations were based on the model illustrated in Figure 1 with the stage 1 model defined as $\xi_i = \gamma_1 Z_i + \widetilde{\zeta}_i$, and $X_{ij} = \xi_i + \widetilde{\epsilon}_{ij}$ and the stage 2 model given by

$$\eta_i = \beta^T \varphi(\xi_i) + \gamma_2 Z_i + \zeta_i, \quad Y_{ij} = \eta_i + \epsilon_{ij},$$

with mutually independent residual terms $\widetilde{\zeta}_i, \zeta_i, \widetilde{\epsilon}_{ij}, \epsilon_{ij}, j = 1, 2, 3$. The distribution of these terms was varied throughout the simulations. The parameters of primary interest were the structural parameters, β, defining the association between the two latent variables ξ_i and η_i.

The source code for the simulations are available in the `lava` git repository (commit hash: `53e2e18`): https://github.com/kkholst/lava/tree/twostage/inst/simulations.

3.1. *Simulations I: correctly specified model*

Data were generated from a quadratic structural model $\eta_i = \beta_0 + \beta_1 \xi_i + \beta_2 \xi_i^2 + \gamma_2 Z_i + \zeta_i$, with all residuals $\widetilde{\zeta}_i, \zeta_i, \widetilde{\epsilon}_{ij}, \epsilon_{ij}, j = 1, 2, 3$ being standard normal and $\beta_0 = 1, \beta_1 = 1, \beta_2 = 0.5$. The simulation was run with sample sizes of $n = 1000$ and $n = 500$ (see Supplementary Material G available at

Table 1. *Performance of the estimators: Gaussian 2SSEM, 2SEMM with two-component mixture (2SSEM mixture), 2SLS, 2SLS with heteroskedasticity standard errors (2SLS robust), methods of moments estimator (2SMM and 2SMM robust, with the former deriving moments from a Gaussian distribution. Standard error are omitted here but are derived in Wall and Amemiya (2000)), and approximate ML (Laplace, AGQ9) in a simulation study from a quadratic model* $\mathbb{E}(\eta \mid \xi) = \beta_0 + \beta_1\xi + \beta_2\xi^2$ *(true parameters* $\beta_1 = 1$ *and* $\beta_2 = 0.5$*) where all assumptions hold*

		Mean	SD	SE	$\frac{\text{SE}}{\text{SD}}$	Cov.	RMSE
$\boldsymbol{\beta_1 = 1}$	2SSEM	0.996	0.371	0.349	0.942	0.940	0.371
	2SSEM mixture	1.031	0.393	0.424	1.079	0.941	0.394
	2SLS	0.975	0.647	0.606	0.938	0.938	0.647
	2SLS robust	0.975	0.647	0.631	0.975	0.948	0.647
	2SMM	1.031	0.431				0.432
	2SMM robust	1.042	0.466				0.468
	Laplace	1.127	0.306	0.292	0.957	0.939	0.331
	AGQ9	1.002	0.277	0.272	0.984	0.947	0.277
$\boldsymbol{\beta_2 = 0.5}$	2SSEM	0.499	0.072	0.068	0.944	0.933	0.072
	2SSEM mixture	0.507	0.077	0.084	1.095	0.934	0.077
	2SLS	0.496	0.115	0.100	0.864	0.919	0.115
	2SLS robust	0.496	0.115	0.112	0.973	0.942	0.115
	2SMM	0.506	0.081				0.081
	2SMM robust	0.508	0.088				0.088
	Laplace	0.526	0.060	0.057	0.958	0.939	0.065
	AGQ9	0.501	0.054	0.053	0.986	0.943	0.054

Biostatistics online), without any covariate in the model ($\gamma_1 = 0, \gamma_2 = 0$) and with a covariate ($\gamma_1 = 1, \gamma_2 = 1$, see Supplementary Material available at *Biostatistics* online). The simulation study was based on 1000 replications. The 2SSEM methods were compared to Bollen's 2SLS estimator (Bollen, 1995) (see Supplementary Material E available at *Biostatistics* online), 2SMM of Wall and Amemiya (2000), and approximate ML based on a Laplace approximation as well as AGQ with nine quadrature points.

Simulations showed that the 2SSEM estimator had good properties in finite samples (Table 1). The method seemed approximately unbiased and confidence intervals had coverage probabilities close to the nominal level. It is interesting to see that in this case, where residual terms were normal, nothing seemed to have been lost by applying the robust mixture model extension. In addition to providing effectively unbiased inference, this method yielded standard errors that were very close to those obtained assuming normality. As expected, ML analysis was more efficient, but the loss of the 2SSEM procedure was modest. The Laplace approximation showed some bias compared to the rest of the methods, but preformed well as measured by the RMSE and was almost as efficient as the more sophisticated AGQ-approximation. Bollen's 2SLS provided unbiased estimation, but it was clearly less efficient than both ML and 2SSEM. Also, the non-robust standard errors as suggested by Bollen (1995) underestimated the uncertainty in the second order term. The methods of moments estimator 2SMM was less efficient than the 2SSEM estimator but performed clearly better than 2SLS. The conclusions were consistent across all scenarios we examined (see Supplementary Material G available at *Biostatistics* online).

In contrast to 2SLS and 2SMM, our method is not restricted to polynomial structures, and we also examined the performance with an exponential effect, $\eta_i = \beta_1\xi_i + \beta_2 \exp(\xi_i) + \zeta_i$. The 2SSEM estimator also performed well in this setting being effectively unbiased with correct coverage of the confidence limits (see Supplementary Material G available at *Biostatistics* online for details).

3.2. *Simulations II: robustness*

Here we explored the properties of the estimators in misspecified models. First, we examined data generating mechanisms identical to the previous model except of the conditional distribution of the latent variable ξ_i which was not Gaussian but followed a mixture distribution, i.e., $\widetilde{\zeta}_i \sim 0.25\mathcal{N}(0,1) + 0.75\mathcal{N}(3,1)$. The

Table 2. *Performance of the two-stage estimator assuming a Gaussian distribution (2SSEM) and with two-component mixture (2SSEM mixture) in a simulation study from a quadratic model in three scenarios, where modeling assumptions were not all satisfied. First the latent predictor followed a two-component mixture distribution, then the latent predictor followed a uniform distribution and finally the residuals in the measurement model of stage one followed a uniform distribution*

		Mean	SD	SE	$\frac{\text{SE}}{\text{SD}}$	Cov.	RMSE
$\widetilde{\zeta} \sim$ GMM							
	2SSEM	1.349	0.088	0.086	0.976	0.021	0.360
	2SSEM mixture	1.000	0.104	0.103	0.988	0.948	0.104
$\boldsymbol{\beta_1 = 1}$	2SLS robust	1.012	0.169	0.162	0.956	0.936	0.170
	2SMM	1.000	0.112				0.112
	2SMM robust	0.999	0.114				0.114
	2SSEM	0.380	0.030	0.029	0.976	0.025	0.123
	2SSEM mixture	0.499	0.038	0.038	0.990	0.947	0.038
$\boldsymbol{\beta_2 = 0.5}$	2SLS robust	0.496	0.056	0.055	0.976	0.929	0.056
	2SMM	0.500	0.040				0.040
	2SMM robust	0.500	0.041				0.041
$\widetilde{\zeta} \sim$ Unif							
	2SSEM	0.998	0.069	0.069	0.998	0.948	0.069
	2SSEM mixture	0.998	0.075	0.075	1.007	0.957	0.075
$\boldsymbol{\beta_1 = 1}$	2SLS robust	1.003	0.085	0.087	1.015	0.959	0.086
	2SMM	1.001	0.078				0.078
	2SMM robust	1.001	0.077				0.077
	2SSEM	0.310	0.054	0.056	1.030	0.102	0.198
	2SSEM mixture	0.485	0.101	0.104	1.027	0.944	0.102
$\boldsymbol{\beta_1 = 0.5}$	2SLS robust	0.489	0.160	0.159	0.994	0.934	0.161
	2SMM	0.544	0.124				0.131
	2SMM robust	0.520	0.126				0.128
$\widetilde{\epsilon}_j \sim$ Unif							
	2SSEM	0.997	0.075	0.075	1.002	0.945	0.075
	2SSEM mixture	0.998	0.082	0.082	1.000	0.935	0.082
$\boldsymbol{\beta_1 = 1}$	2SLS robust	0.997	0.090	0.091	1.014	0.951	0.090
	2SMM	1.000	0.079				0.079
	2SMM robust	1.000	0.079				0.079
	2SSEM	0.506	0.063	0.063	0.996	0.951	0.063
	2SSEM mixture	0.521	0.072	0.076	1.055	0.956	0.075
$\boldsymbol{\beta_1 = 0.5}$	2SLS robust	0.499	0.088	0.088	0.996	0.937	0.088
	2SMM	0.521	0.070				0.073
	2SMM robust	0.512	0.074				0.074

results are summarized in Table 2, where we see some bias in the 2SSEM estimator using a Gaussian distribution for the stage 1 model. The mixture 2SSEM estimator is unbiased with correct coverage. As an observation, we noted that 2SSEM was generally much faster, more computational stable and less dependent on starting values than the Laplace and AGQ approximations. This was especially the case in the mixture setting where the ML-methods had convergence problems and need for fine-tuning across different implementations (results not shown).

The above simulation setup corresponds exactly to the assumptions of our mixture model extension, so to test the robustness of the extension we also included a study where $\widetilde{\zeta}_i$ followed a uniform distribution with mean zero and variance one, and a simulation where the residuals of the indicators, $\widetilde{\epsilon}_{ij}$ followed a uniform distribution. In both cases, the 2SSEM mixture estimator was effectively unbiased with coverage close to the nominal level. Interestingly, the Gaussian 2SSEM was robust to the misspecification of the indicator distribution where it performed slightly better than the mixture model. Both 2SLS and 2SMM appeared to be robust to the considered misspecifications. In most cases both estimators were less efficient than 2SSEM (see Supplementary Material G available at *Biostatistics* online); however, as the sample size increased we observed that 2SMM seemed to catch up with 2SSEM. We note however, that a severe limitation of the 2SMM approach is the lack of generalizations (and implementations) allowing for example relaxation of conditional independence assumptions, inclusion of covariates, and most importantly specifications of functional forms beyond the polynomial structure.

3.3. *Simulations III: non-parametric estimation*

To study 2SSEM in a non-parametric setting, we also simulated data from the measurement models defined in the previous sections but with unknown functional relationship between the latent variables given by $\varphi(\xi;\boldsymbol{\beta}) = \beta_1\xi + \beta_2\xi^2 + \sin(\beta_3\xi)$. A natural cubic spline with k knots $t_1 < t_2 < \cdots < t_k$ is given by $\mathbb{E}(\eta_i \mid \xi_i) = \gamma_0 + \gamma_1\xi_i + \sum_{j=1}^{k-2}\gamma_{j+1}f_j(\xi_i)$, with $f_j(\xi_i) = g_j(\xi_i) - \frac{t_k - t_j}{t_k - t_{k-1}}g_{k-1}(\xi_i) + \frac{t_{k-1} - t_j}{t_k - t_{k-1}}g_k(\xi_i)$, $j = 1,\ldots,k-2$. Here $g_j(\xi_i) = (\xi_i - t_j)^3 1_{\{\xi_i > t_j\}}$ so to apply 2SSEM we calculated $\mathbb{E}[g_j(\xi_i)|X_i, Z_i]$ (see Supplementary Material B available at *Biostatistics* online). As a benchmark, we compared 2SSEM with the estimator proposed by Kelava *and others* (2017) and the corresponding Matlab implementation (https://github.com/tifasch/nonparametric/tree/ead709097d6). Here, the number of equidistant knots were chosen by dividing the simulated data into a single test and training data of equal size and choosing the spline basis (degrees of freedom varying from 1 to 11) as the one that minimized the RMSE evaluated in the test data. We noted that slightly better results were obtained for 2SSEM when the hyper-parameters (spline knots) were chosen using 5-fold cross validation. To make the results more comparable we, however, adopted the same method for choosing the degrees of freedom for the spline using the exact same split of the testing and training data.

In each simulation, $r = 1,\ldots,100$, we generated $n = 200$ observations, and for each estimator we calculated $\text{RMSE}_r = \left(\sum_{i=1}^{n}\left[\varphi(\xi_i;\boldsymbol{\beta}) - \tilde{\varphi}(\xi_i;\widehat{\boldsymbol{\gamma}}_r)\right]^2\right)^{\frac{1}{2}}$, where $\boldsymbol{\beta}$ denotes the true parameter and $\widehat{\boldsymbol{\gamma}}_r$ is the estimated parameters of the spline model, i.e., $\widehat{\eta} = \tilde{\varphi}(\xi;\widehat{\boldsymbol{\gamma}}_r) = \boldsymbol{B}(\xi)\widehat{\boldsymbol{\gamma}}_r$, where $\boldsymbol{B}(\xi)$ is the spline basis design matrix. With $\beta_1 = 1, \beta_2 = 0, \beta_3 = 1$ and with $\widetilde{\zeta}, \zeta \sim \mathcal{N}(0,1)$ the average RMSE over all replications were 0.314 for Kelava's estimator and 0.112 for 2SSEM, and similarly when $\widetilde{\zeta}, \zeta \sim U(-6,6)$ the average RMSE were 0.933 and 0.608 in favor of 2SSEM. Similar conclusions were drawn when using a stronger non-linear functional form given by $\beta_1 = 1, \beta_2 = 0, \beta_3 = 1$. In the Gaussian case the average RMSE was 1.177 and 0.827, and in the uniform case 3.613 and 1.988, all in favor of 2SSEM. In addition, we note that another advantage of the 2SSEM estimator is the immediately available expressions of the asymptotic variance through the estimated influence functions. See Supplementary Material G.4 available at *Biostatistics* online for more details.

4. Application: modeling *in vivo* brain serotonin measurements

Serotonin (5-HT, 5-hydrotryptamine) is known to play an important role in the regulation of appetite, sleep, mood, sex, and memory function. Variation in cerebral 5-HT levels is also recognized as being influential on addiction and development of psychiatric disorders such as schizophrenia and depression. With positron emission tomography (PET) techniques it is possible to quantify post and pre-synaptic markers of the serotonergic system in the living human brain, such as the serotonin 2A receptor (5-HT_{2A}) and the serotonin transporter (SERT). These markers have been intensively studied and associations to eating, sleeping, and mood disorders (Meyer, 2007; Meyer *and others*, 1999) have been identified.

Animal studies examining the consequence of manipulation of central 5-HT levels indicate an approximate (negative) linear relationship between normal synaptic 5-HT levels and 5-HT_{2A} receptor binding (Licht *and others*, 2009). It has been suggested that (Meyer, 2007) 5-HT_{2A} receptor binding may act as an indicator of the cerebral 5-HT levels. Similarly, experimental studies have shown that manipulation of synaptic 5-HT levels causes change in the SERT binding (Pineyro *and others*, 1994) with a suggested non-linear functional form (inverted u-shape where low and high serotonin levels both are associated with low levels of SERT). This association was studied in Erritzoe *and others* (2010) hypothesizing that underlying low 5-HT levels could lead to a compensatory up-regulation of 5-HT_{2A} receptor binding and down-regulation of SERT.

We will here present an analysis of the same sample as in the original paper, while taking into account the measurement error in both the 5-HT_{2A} and SERT measurements by using a non-linear SEM. The 5-HT_{2A} receptor binding potential (BP_p) and the SERT binding potential (BP_{ND}) was measured in 56 normal subjects. For each subject, the measurements were summarized in a number of regions of interest. We refer to the original paper for details on the method used in acquiring the data.

For the serotonergic markers the concept of a measurement model seems to be ideal in capturing the idea of an underlying common regulator of the two types of measurements. For the 5-HT_{2A} receptor outcome in a given region, we will assume a model

$$\text{BP}_{p,\text{ROI}} = \mu_{\text{ROI}} + \lambda_{\text{ROI}} \cdot \xi + \epsilon_{\text{ROI}}, \tag{4.12}$$

with a single latent variable ξ. The flexibility in letting each region have its own intercept, μ_{ROI}, loading parameter, λ_{ROI}, and residual term $\epsilon_{\text{ROI}} \sim \mathcal{N}(0, \sigma^2_{\text{ROI}})$ allows us to model data, where different regions have different degrees of binding potential, variation, and correlation with other regions. We will assume independence between residuals though this is not a necessary assumption. We propose a similar model for SERT binding with a measurement model described by a latent variable η. For both markers we chose four high binding regions of interest which previously have been demonstrated to be reliable measurements of 5-HT_{2A} and SERT binding, respectively (see Figure 2). To describe the association between 5-HT_{2A} receptor binding and SERT binding we added a simple structural model $\eta = \mu_{\text{sex}} + \beta\xi + \zeta$, to see how well a linear approximation would describe the relationship. In this simple model, ξ takes the role of the common regulator, i.e. the central 5-HT level, as measured directly by the 5-HT_{2A} receptor binding, and the common regulator predicts the levels of the global SERT variable η.

We estimated the parameters of the model using ML. The estimate of the primary parameter of interest, β, was 0.046 $\text{BP}_{ND}/\text{BP}_p$ (with parietal cortex and thalamus as reference regions, allowing us to interpret the effect as change in SERT BP_p in thalamus per unit change in parietal cortical BP_{ND}) and 95% CI $[-0.057; 0.149]$. The lack of statistical significance may be explained by the lack of non-linear effects in our model specification. A χ^2 omnibus-test of goodness-of-fit (a likelihood ratio test against an unstructured 8D normal distribution) yielded a p-value of 0.23. Thus, based solely on this test there was no evidence against the model. Clearly, this goodness-of-fit test is, however, not adequate for detecting non-linearities (Mooijaart and Satorra, 2009).

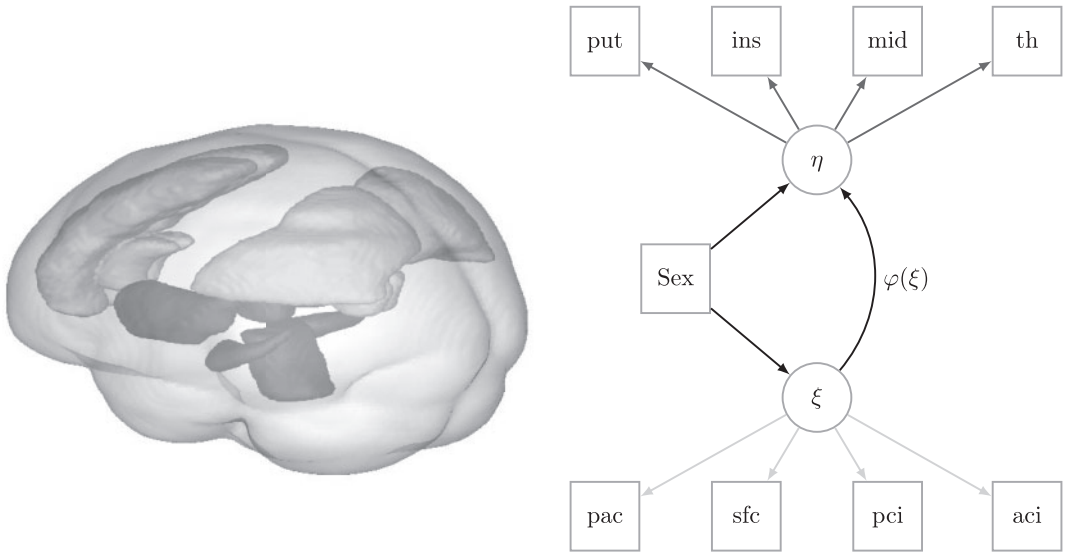

Fig. 2. Path diagram for a structural equation model describing the relationship between 5-HT_{2A} receptor binding (light gray) and SERT binding (dark gray). Each of the two types of markers are described by a single latent variable. For the 5-HT_{2A} receptor the following regions (light gray regions in the glass brain of the right figure) was chosen as measurements: *Parietal cortex* (pac), *Superior frontal cortex* (sfc), *Posterior cingulate gyrus* (pci), and *Anterior cingulate gyrus* (aci). For the serotonin transporter we chose the regions: *Putamen* (put), *Insula* (ins), *Midbrain* (mid), and *Thalamus* (th).

Next, we applied the 2SSEM procedure to the SEM of Figure 2 with the association between ξ and η described by a second order polynomial: $\eta = \mu_{\text{sex}} + \beta_1\xi + \beta_2\xi^2 + \zeta$. The estimates were $\widehat{\beta}_1 = 0.676$ (95% CI $[0.321; 1.030]$, $p = 0.00018$) and $\widehat{\beta}_2 = -0.153$ (95% CI $[-0.233; -0.074]$, $p = 0.0002$), thus confirming our hypothesis of a non-linear association between 5-HT_{2A} receptor and SERT binding potential (Wald test for the hypothesis of no association: $p = 0.0008, df = 2$).

A more flexible natural cubic spline model was next applied. The predicted latent variable of the measurement error model for 5-HT_{2A} receptor binding potential was in the range 0.5–3.5 BP_p, and we choose 4 knot points equidistantly in the interval 1 to 3. The association between the two markers (and comparison with the linear and quadratic model) is shown in Figure 3. The natural cubic spline suggests a more flat association between 5-HT_{2A} BP_p and SERT BP_{ND} for high 5-HT_{2A} binding potential values, but otherwise there was a close agreement with the quadratic model. We conducted a more formal comparison of the two models using 5-fold cross validation with all indicator variables normalized. The RMSE was in slight favor of the quadratic model (0.94 vs 1.03). We also examined natural cubic splines with increased number of knots, but they all exhibited over-fitting with higher RMSE. Finally, we used a two-component mixture in the stage one model and got results that were almost identical to those of the Gaussian model. Also, as in the original paper we observed that the estimated non-linear association was not sensitive to removing the observations with the highest values of 5-HT_{2A} binding potential from the data (results not shown).

Thus, in agreement with animal and experimental studies, we were able to show a non-linear association between these two serotonergic markers. Our refined analysis also confirmed the findings of the original paper (Erritzoe *and others*, 2010), where the same data were analyzed using standard regression techniques and hence results are likely to be susceptible to bias due to measurement error in both variables.

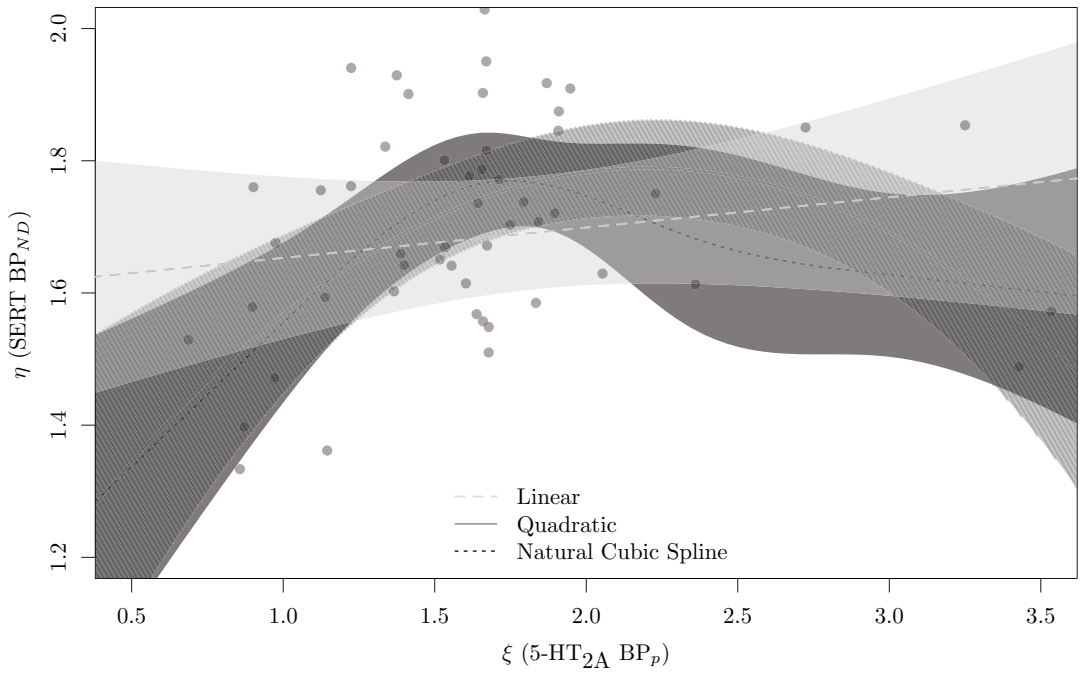

Fig. 3. Association between 5-HT_{2A} BP_p binding potential and SERT BP_{ND} binding potential as estimated using linear, quadratic, and natural cubic spline model. The points are the Empirical Bayes estimates from two separate linear SEMs.

5. DISCUSSION

ML-inference in non-linear SEMs is complex. Computational intensive methods based on numerical integration are needed and results are sensitive to distributional assumptions. This article presented the two-stage estimator 2SSEM as a computationally simple alternative to ML. Here, both steps are based on linear models: first we predict the non-linear terms and then these are related to latent outcomes in the second step. We identified the asymptotic distribution of 2SSEM, developed a robust extension based on mixture models and implemented the methods in a `R`-package (see Supplementary Material F available at *Biostatistics* online) available from the `twostage` branch of https://github.com/kkholst/lava.

Simulations indicated a modest loss of efficiency compared to ML-estimation and our method was shown to be more powerful than two computationally simple and robust alternatives, i.e., 2SLS and 2SMM. In addition, 2SSEM can be applied to a larger class of non-linear functions than 2SLS and 2SMM. In particular, the class of restricted cubic splines is an important example which have shown to be very useful in applications of regression models. The introduction of stable and fast estimation algorithms for spline functions in the structural equation framework is likely to lead to important improvements in applications which for too long have been restricted to linear relationships.

In linear models, 2SSEM is equivalent to regression calibration (Carroll *and others*, 2006). This method has been investigated in linear SEMs e.g., by Skrondal and Laake (2001) and Sánchez *and others* (2009). In non-linear SEMs, the idea of using mixture modeling to achieve more robust estimation has been exploited for ML-estimators e.g., by Kelava *and others* (2014). The handling of splines in 2SSEM is related to the non-parametric estimators suggested by Carroll *and others* (1999) for linear regression models with measurement error in covariates. Bayesian methods have been developed for semi-parametric estimation

in SEMs (Song *and others*, 2013; Guo *and others*, 2012; Kelava and Brandt, 2014), but frequentist methods are rare. Bauer (2005) and Kelava *and others* (2017) presented interesting methods but did not provide results on asymptotic standard errors. The latter procedure was included in our simulation study where it yielded larger prediction errors than 2SSEM.

The two-stage approach may be especially useful in data bases with many different research projects. Here, the SEM for the exposure may be fitted only once and the predicted non-linear terms can be stored along with the influence function. Then the predictions of exposure terms can be related to different outcomes by different research groups using linear structural models with corrected standard errors. Even in situations where ML-inference is the goal, 2SSEM will likely be very useful in providing good starting values.

When using 2SSEM for assessing associations between latent variables an obvious strategy would be to start the analysis with a rich parametric model (e.g., spline model) which may then be reduced by backward selection using Wald tests. An obvious extension would be to develop lasso-type regularization for the 2SSEM estimator. Similarly, different parametric forms may be compared using Wald tests in a nested model. An alternative is to base the model selection on the estimated out-of-sample predictive performance through cross-validation as demonstrated in the application. It may also be possible to develop fit criteria to detect general non-linear misfit. Recent theory in non-linear SEMs have focused on the development of such criteria and it may be possible to extend these so that they can be used together with 2SSEM. For example, Schermelleh-Engel *and others* (2014) developed a χ^2-test comparing the observed and expected covariance matrix of the observed variables appended with selected products of indicators. Another interesting possibility would be to consider cumulative residuals as described by Sánchez *and others* (2009) for linear SEMs.

We applied non-linear models to PET measurements of the serotonergic system. Based on biological evidence, we proposed a statistical model describing the underlying cerebral 5-HT level by inclusion of latent components in a SEM. The underlying 5-HT level was here assumed to be measured indirectly by PET measurements of 5-HT_{2A} receptor binding and SERT binding. In agreement with animal and experimental studies, we were able to show a non-linear association between these two serotonergic markers. Our model represents a first step towards linking several measurements of the serotonergic system into a simultaneous description of central 5-HT levels. An interesting longer-term perspective of this model is the possibility to explore the association between latent 5-HT levels and the development of neuropsychiatric diseases such as major depressive episodes. The extension of 2SSEM to allow for binary and time-to-event endpoints will be a topic for future research.

Supplementary materials

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

The project is part of the Dynamical Systems Interdisciplinary Network, University of Copenhagen, and it has also received support from Innovation Fund Denmark. The authors wish to thank the staff at the Neurobiology Research Unit, Rigshospitalet, for providing the data example and help with the interpretation of the data. *Conflict of Interest:* None declared.

References

Anderson, T. W. and Amemiya, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics* **16**, 759–771.

ARMINGER, G. AND MUTHÉN, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika* **63**, 271–300.

ARMINGER, G. AND SCHOENBERG, R. J. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika* **54**, 409–425.

BAUER, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling* **12**, 513–535.

BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

BOLLEN, K. A. (1995). Structural equation models that are nonlinear in latent variables: a least-squares estimator. *Sociological Methodology* **25**, 223–252.

CARROLL, R. J., MACA, J. D. AND RUPPERT, D. (1999). Nonparametric regression in the presence of measurement error. *Biometrika* **86**, 541–554.

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. AND CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. CRC Press.

ERRITZOE, D., HOLST, K. K., FROKJAER, V. G., LICHT, C. L., KALBITZER, J., NIELSEN, F. A., SVARER, C., MADSEN, J. AND KNUDSEN, G. M. (2010). A nonlinear relationship between cerebral serotonin transporter and 5-HT2A receptor binding: an in vivo molecular imaging study in humans. *Journal of Neuroscience* **30**, 3391–3397.

GUO, R., ZHU, H., CHOW, S. M. AND IBRAHIM, J. G. (2012). Bayesian lasso for semiparametric structural equation models. *Biometrics* **68**, 567–577.

JÖRESKOG, K. G. AND YANG, F. (2000). Latent variable scores and their uses. On-line paper www.ssicentral.com/lisrel/techdocs/lvscores.pdf.

KELAVA, A. AND BRANDT, H. (2014). A general non-linear multilevel structural equation mixture model. *Frontiers in Psychology* **5**, 748.

KELAVA, A., KOHLER, M., KRZYŻAK, A. AND SCHAFFLAND, T. F. (2017). Nonparametric estimation of a latent variable model. *Journal of Multivariate Analysis* **154**, 112 – 134.

KELAVA, A., NAGENGAST, B. AND BRANDT, H. (2014). A nonlinear structural equation mixture modeling approach for nonnormally distributed latent predictor variables. *Structural Equation Modeling: A Multidisciplinary Journal* **21**, 468–481.

KENNY, D. A. AND JUDD, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin* **96**, 201.

KLEIN, A. AND MOOSBRUGGER, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika* **65**, 457–474.

LEE, S. Y. AND ZHU, H. T. (2000). Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology* **53**, 209–232.

LEE, S. Y. AND ZHU, H.-T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika* **67**, 189–210.

LICHT, C. L., MARCUSSEN, A. B., WEGENER, G., OVERSTREET, D. H., AZNAR, S. AND KNUDSEN, G. M. (2009). The brain 5-HT4 receptor binding is down-regulated in the Flinders Sensitive Line depression model and in response to paroxetine administration. *Journal of Neurochemistry* **109**, 1363–1374.

MARSH, H. W., WEN, Z. AND HAU, K. T. (2004). Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychological Methods* **9**, 275.

MEYER, J. H. (2007). Imaging the serotonin transporter during major depressive disorder and antidepressant treatment. *Journal of Psychiatry & Neuroscience* **32**, 86–102.

MEYER, J. H., KAPUR, S., HOULE, S., DASILVA, J., OWCZAREK, B., BROWN, G. M., WILSON, A. A. AND KENNEDY, S. H. (1999). Prefrontal cortex 5-HT2 receptors in depression: an [18F]setoperone PET imaging study. *American Journal of Psychiatry* **156**, 1029–34.

MOOIJAART, A. AND BENTLER, P. M. (2010). An alternative approach for nonlinear latent variable models. *Structural Equation Modeling* **17**, 357–373.

MOOIJAART, A. AND SATORRA, A. (2009). On insensitivity of the chi-square model test to nonlinear misspecification in structural equation models. *Psychometrika* **74**, 443–455.

MUTHÉN, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* **49**, 115–132.

MUTHÉN, L. K. AND MUTHÉN, B. O. (2012). *Mplus Version 7 User's Guide*. Los Angeles, CA: Muthén & Muthén.

PINEYRO, G., BLIER, P., DENNIS, T. AND DE MONTIGNY, C. (1994). Desensitization of the neuronal 5-HT carrier following its long-term blockade. *Journal of Neuroscience* **14**, 3036–3047.

RIZOPOULOS, D. AND MOUSTAKI, I. (2008). Generalized latent variable models with non-linear effects. *British Journal of Mathematical and Statistical Psychology* **61**, 415–438.

SÁNCHEZ, B. N., BUDTZ-JØRGENSEN, E. AND RYAN, L. M. (2009a). An estimating equations approach to fitting latent exposure models with longitudinal health outcomes. *The Annals of Applied Statistics* **3**, 830–856.

SÁNCHEZ, B. N., HOUSEMAN, E. A. AND RYAN, L. M. (2009b). Residual-based diagnostics for structural equation models. *Biometrics* **65**, 104–115.

SCHERMELLEH-ENGEL, K., KERWER, M. AND KLEIN, A. G. (2014). Evaluation of model fit in nonlinear multilevel structural equation modeling. *Frontiers in Psychology* **5**, 181.

SCHERMELLEH-ENGEL, K., KLEIN, A. AND MOOSBRUGGER, H. (1998). Estimating nonlinear effects using a latent moderated structural equations approach. In: SCHUMACKER, R. E. AND MARCOULIDES, G. A. (editors), *Interaction and Nonlinear Effects in Structural Equation Modeling*. Lawrence Erlbaum Associates Publishers, pp. 203–238.

SCHUMACKER, R. E. (2002). Latent variable interaction modeling. *Structural Equation Modeling* **9**, 40–54.

SKRONDAL, A. AND LAAKE, P. (2001). Regression among factor scores. *Psychometrika* **66**, 563–575.

SKRONDAL, A. AND RABE-HESKETH, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. CRC Press.

SONG, X. Y., LU, Z. H., CAI, J. H. AND IP, E. H. S. (2013). A Bayesian modeling approach for generalized semiparametric structural equation models. *Psychometrika* **78**, 624–647.

SONG, X. Y., PAN, J. H., KWOK, T., VANDENPUT, L., OHLSSON, C. AND LEUNG, P. C. (2010). A semiparametric Bayesian approach for structural equation models. *Biometrical Journal* **52**, 314–332.

WALL, M. M. (2009). Maximum likelihood and Bayesian estimation for nonlinear structural equation models. In: MILLSAP, R. E. AND Maydeu-OLIVARES, A. (editors), *The SAGE Handbook of Quantitative Methods in Psychology*. California: SAGE Publications, pp. 540–567.

WALL, M. M. AND AMEMIYA, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association* **95**, 929–940.

YANG, M. AND DUNSON, D. B. (2010). Bayesian semiparametric structural equation models with latent variables. *Psychometrika* **75**, 675–693.

[*Received December 8, 2017; revised October 21, 2018; accepted for publication October 28, 2018*]