Integrated PLS-SEM-Latent Growth Curve Model: A New Conditional Time Invariant Method for Analysing Panel Survey Data
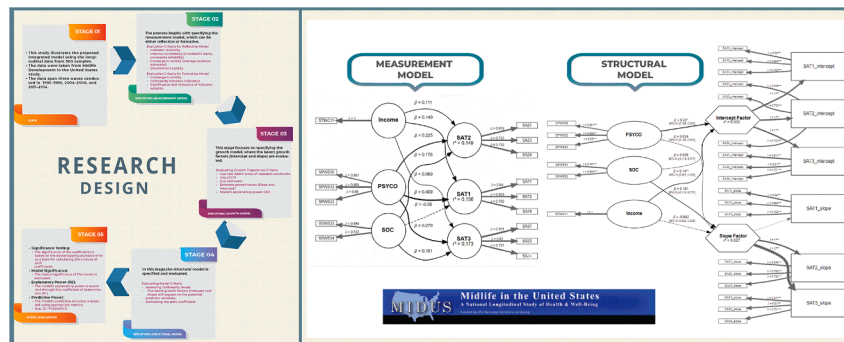
Zulkifli Mohd Ghazali [a], Wan Yaacob [b,c,*]

[a] Mathematical Sciences Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Perak, Kampus Tapah, Tapah Road 35400, Perak, Malaysia
[b] Mathematical Sciences Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam 40450, Selangor, Malaysia
[c] Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Kompleks Al-Khawarizmi, Universiti Teknologi MARA, Shah Alam 40450, Selangor, Malaysia

GRAPHICAL ABSTRACT

ARTICLE INFO

ABSTRACT

This study introduces an integrated methodological framework that combines Partial Least Squares Structural Equation Modelling (PLS-SEM) with a conditional time-invariant Latent Growth Curve Model (LGCM) to analyse panel survey data. The integration addresses limitations of existing PLS-SEM approaches in longitudinal research by enabling the simultaneous evaluation of measurement validity, growth trajectories, and predictors of change over time. A disjoint two-

---

* Corresponding author.
E-mail address: wnfairos@uitm.edu.my (W. Yaacob).

stage approach is applied to estimate measurement models and obtain latent scores, which are subsequently used to model growth factors (intercepts and slopes). The proposed method is illustrated using three waves of the Midlife in the United States (MIDUS) study, focusing on life satisfaction and its predictors. Results indicate that psychological well-being and social well-being significantly predict both baseline levels and growth in life satisfaction, while income influences only baseline levels. These findings demonstrate the capacity of the integrated model to disentangle inter-individual differences and developmental patterns in longitudinal data. The findings indicate that:

The integrated PLS-SEM–LGCM framework supports simultaneous analysis of constructs, trajectories, and predictors.

Empirical validation with MIDUS data demonstrates its ability to capture variability in life satisfaction.

Implementation in R through the custom *pls_growth* function enhances reproducibility and accessibility.

Specifications table

Subject area	Mathematics and Statistics
More specific subject area	Panel Survey Data
Name of your method	PLS-SEM Based Model with Conditional Time-Invariant of Latent Growth Curve for Panel Survey Data
Name and reference of original method	Original Method: PLS-SEM Higher-order Construct: Disjoint two-stage
	Reference: Becker, J. M., Klein, K., & Wetzels, M. (2012). Hierarchical Latent Variable Models in PLS-SEM: Guidelines for Using Reflective-Formative Type Models. Long Range Planning, 45(5–6), 359–394. https://doi.org/10.1016/j.lrp.2012.10.001
Resource availability	This study used a dataset from samples collected across three waves of the Midlife Development in the United States study (1995–1996, 2004–2006, and 2011–2014) https://www.icpsr.umich.edu/

Background

Partial Least Squares Structural Equation Modelling (PLS-SEM) is an alternative to the widely used covariance-based SEM (CB-SEM) method for analysing survey data in structural equation modelling techniques. The PLS-SEM approach was initially introduced by [1], and subsequent extensions were suggested by [2]. PLS-SEM has gained popularity due to its ability to estimate complex models, allowing researchers to explore intricate relationships among variables. Recent advancements in model specification and higher-order constructs in PLS-SEM are well-documented in [3], which provides updated guidelines for complex modelling contexts. This method offers flexibility in data requirements and enables the specification of relationships between constructs and indicator variables. Additionally, PLS-SEM is particularly effective in handling non-normal data and small sample sizes, making it suitable for studies with limited data or non-normal distributions [4–6]. Although PLS-SEM has gained popularity, especially in fields like education, marketing, and psychology, its application is limited to cross-sectional survey data ([7–10]). The utilisation of PLS-SEM in panel survey studies, which involves repeated measures over time, is relatively limited. This discrepancy becomes more apparent as there has been an increasing shift from cross-sectional to panel survey studies in recent years, particularly in the context of longitudinal research [11–13]. Several factors contributed to the underutilisation of PLS-SEM in panel survey data, despite the availability of five different approaches in PLS-SEM for analysing panel survey data. Among the approaches are; (i) Pre and Post approach with different construct (model 1), (ii) Path Comparison approach (model 2), (iii) Cross-lagged approach (model 3), (iv) Pre and Post approach with the same construct (model 4), and (v) Evaluation approach (model 5) ([14]).

Although there are five different approaches of PLS-SEM in analysing the panel survey data, these approaches still have limitations and spaces for improvement. The obvious limitation for models 1, 2, 3, and 4 is related to the number of waves for the study, since these approaches are only suitable for two periods of time. In addition, these four models focus on the pre- and post-evaluations and do not measure the evaluations of effects over time. Besides that, model 5 has a limitation in evaluating the growth of trajectory, even though this method is capable of handling studies with more than two periods of time. This model does not have a single structural model to measure the whole changes in the repeated measures simultaneously. In addition, the existing approach or model cannot capture an individual trajectory, the average of the trajectory of the sample or entire group, the evaluation of individual differences in trajectories, and assess the potential incorporation of predictors of individual differences in trajectories. Furthermore, it is also not flexible to simultaneously handle the independent and dependent latent constructs of the same model, allowing for complex representations of growth and correlations of change. These limitations hinder the effective application of PLS-SEM in this specific research context.

Therefore, this study introduces a novel integrated model by combining PLS-SEM with the Latent Growth Curve Model (LGCM) for analysing panel survey data. The integrated approach enhances the capabilities of PLS-SEM to better evaluate growth trajectories, inter-individual differences, and the predictors influencing these changes over time. This study also discusses how to evaluate the measurement model and structural model in PLS-SEM for panel survey data. In the measurement model, the adoption of the standard

PLS algorithms to the panel survey data is discussed. Meanwhile, this study also discusses how to develop the structural panel survey equation in the structural model, then evaluates the common latent trajectories, which are the characteristics of the mean trajectory of the entire group, the evaluation of individual differences in trajectories, and the potential incorporation of predictors of individual differences in trajectories.

The remainder of this paper is structured as follows. The Method Details section explains the theoretical framework and the three-stage procedure of the proposed integrated model, including measurement model evaluation, growth modelling, and structural analysis. The Method Validation section presents an empirical illustration using MIDUS panel data and the custom-built *pls_growth* function in R. Finally, the Conclusion section summarizes the key findings, contributions, and limitations of the study, and outlines directions for future research.

Method details

This section presents an exploration of the foundational concepts and methodologies that underpin this research. The proposed integrated model for panel survey data incorporates elements of both Partial Least Squares Structural Equation Modeling (PLS-SEM) and the Latent Growth Curve Model (LGCM). We begin by explaining the characteristics of panel survey data, the fundamentals of LGCM, and key aspects of PLS-SEM relevant to this study, such as its principles, algorithms, and higher-order models. Finally, we introduce a novel theoretical and conceptual framework specifically designed to apply PLS-SEM in the context of panel survey data.

Description of panel survey data

The panel survey data is a combination of cross-sectional and time series data types. The panel survey data consists of several units and many points in time. Using the panel survey data, we can observe the trajectories and the factors that influence these changes. The standard notation used for panel survey data is Y_{it}, where it is observed for all individuals $i = 1, 2, 3, ..., N$ across all time periods $t = 1, 2, 3, ..., T$. Panel survey data is a form of longitudinal study that focuses on capturing changes over time. This type of research often employs the growth equation (Eq. (1) to Eq. (6)) to analyse and understand the patterns of change observed in the data. By tracking the same individuals or groups over multiple points in time, panel surveys provide valuable insights into how various factors influence and shape these changes.

Latent growth curve model (LGCM) for panel survey data

The latent growth curve model (LGCM) is formed from the development of latent variable modelling and factor analysis [15–17]. Applying LGCM to panel survey data helps researchers understand dynamic processes. It reveals the underlying changes and transformations within a population. The LGCM has become widely used in behavioural sciences research. It is applied in various field, including social behavioural research, psychology (encompassing clinical, developmental, and educational research), studies on learning and memory, as well as research on personality [18–23]. This method has an advantage in the investigation of inter-individual differences in change over time, as it allows investigation the antecedents and consequences of change [24]. LGCM also provides group-level statistics, such as the mean growth rate and mean intercept, and can test hypotheses of specific trajectories. Additionally, it allows the incorporation of time-invariant covariates. The basic LGCM can be described as an unconditional growth model. The general unconditional latent growth curve model is as follows.

$$y_{it} = \alpha_i + \lambda_t \beta_i + \epsilon_{it} \tag{1}$$

where y_{it} is the value of the trajectory variable y for the ith case at time t, α_i is the random intercept for case i, and β_i is the random slope for case i. From Eq. (1), the random intercepts and slopes equations are as follows.

$$\alpha_i = \mu_\alpha + \zeta_{\alpha i} \tag{2}$$

$$\beta_i = \mu_\beta + \zeta_{\beta i} \tag{3}$$

where μ_α and μ_β are the mean intercepts and mean slopes.

Conditional model

To incorporate the potential independent variables that predict the growth factor, the conditional LGCM should be used. The conditional model starts with the same equation as the unconditional model, Eq. (1).

$$y_{it} = \alpha_i + \lambda_t \beta_i + \epsilon_{it} \tag{4}$$

However, in the level 2 equations, the equations for random intercepts and slopes (Eq. (2) and Eq. (3)) are modified to incorporate the potential predictor variables. The conditional random intercepts and slopes equations are as follows.

$$\alpha_i = \mu_\alpha + \gamma_{\alpha 1} x_{1i} + \zeta_{\alpha i} \tag{5}$$

$$\beta_i = \mu_\beta + \gamma_{\beta 1} x_{1i} + \zeta_{\beta i} \tag{6}$$

where μ_α and μ_β are the mean intercepts and mean slopes when x_{1i} is zero. The x_{1i} is covariate or predictor of the random intercepts and slopes, while $\gamma_{\alpha 1}$ and $\gamma_{\beta 1}$ are the covariate coefficients for x_{1i} in the random intercept and slope equations. This predictor is known as time-invariant covariate which does not vary over time.

Factor loadings

The factor loading (λ_t) is the time factor that can allow the incorporation of linear and curvilinear trajectories. There are two types of loading, which are fixed loading (represents a linear trajectory) and free loading (represents a curvilinear trajectory). In the linear trajectory, the model uses a fixed loading where λ_t is equal to $T-1$ for all T. In contrast, to estimate the curvilinear trajectory, [25] suggested the first loading $\lambda_1 = 0$ and the last loading $\lambda_T = 1$. Then, all other loadings between the first and last are freely estimated. However, in this study, we focus on the linear trajectory only.

Principles of pls-sem

Partial Least Squares Structural Equation Modelling (PLS-SEM) is a widely utilized statistical technique in survey studies. It offers the advantage of examining both observed and unobserved variables within a research framework. PLS-SEM consists of two main components which are measurement model and structural model. The measurement model establishes the relationship between constructs and their indicators, enabling researchers to assess the extent to which the indicators accurately measure the underlying constructs. On the other hand, a structural model represents the relationships between each of the constructs, enabling the exploration of direct and indirect effects among the variables of interest. The measurement and structural models are illustrated in Fig. 1.

PLS-SEM algorithm

The PLS-SEM algorithm involves a three-stage approach that belongs to the family of least squares algorithms as illustrated in Fig. 2. In the initialization stage, preliminary latent variable scores are calculated using unit weights for all indicators. These scores provide a starting estimate that is refined in later stages of the algorithm. To compute these scores, the algorithm typically uses unit weights (starting with 1) for all indicators in the measurement models. The initialization stage is important because it can affect the convergence and stability of the algorithm. Therefore, it is recommended to use a reliable initialization method, such as the centroid method, the factor-based method or path weighting scheme, to obtain more accurate estimates of the latent variables. The second stage is the outer model estimation stage, where the measurement model is estimated by regressing the observed variables on their respective latent variables. The third stage is the inner model estimation stage, where the structural model is estimated by regressing the endogenous latent variables on the exogenous latent variables.

Higher order construct

In the field of survey studies, evaluating higher-order constructs poses a methodological challenge, as these latent variables cannot be directly observed. Various methods are employed to address this challenge and to gain insight into the underlying relationships within the data. Within the realm of Partial Least Squares Structural Equation Modelling (PLS-SEM), two principal approaches are commonly utilized. The repeated indicators approach uses multiple indicators or observable variables to measure the higher-order
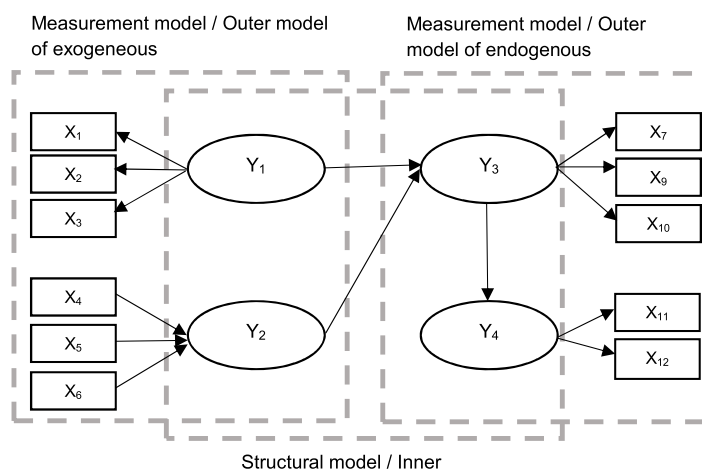

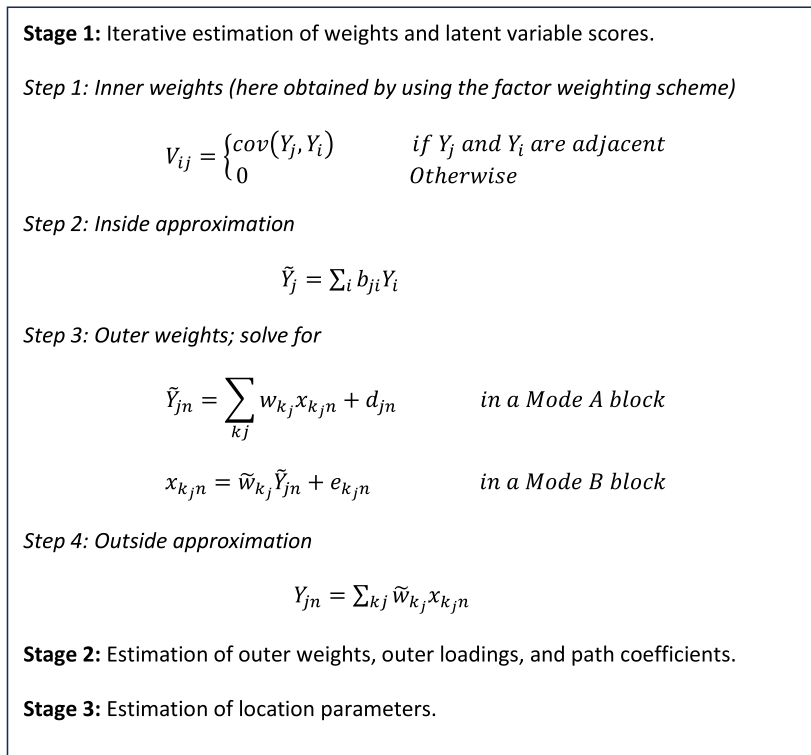**Fig. 1.** The measurement and structural models of the PLS-SEM.

Stage 1: Iterative estimation of weights and latent variable scores.

Step 1: Inner weights (here obtained by using the factor weighting scheme)

$$V_{ij} = \begin{cases} cov(Y_j, Y_i) & \text{if } Y_j \text{ and } Y_i \text{ are adjacent} \\ 0 & \text{Otherwise} \end{cases}$$

Step 2: Inside approximation

$$\tilde{Y}_j = \sum_i b_{ji} Y_i$$

Step 3: Outer weights; solve for

$$\tilde{Y}_{jn} = \sum_{kj} w_{k_j} x_{k_j n} + d_{jn} \qquad \text{in a Mode A block}$$

$$x_{k_j n} = \widetilde{w}_{k_j} \tilde{Y}_{jn} + e_{k_j n} \qquad \text{in a Mode B block}$$

Step 4: Outside approximation

$$Y_{jn} = \sum_{kj} \widetilde{w}_{k_j} x_{k_j n}$$

Stage 2: Estimation of outer weights, outer loadings, and path coefficients.

Stage 3: Estimation of location parameters.

Fig. 2. Procedure in PLS-SEM Algorithm for Measurement Model.

construct. This provides a direct and observable representation of the latent variable ([2,26]).

Conversely, the two-stage approach takes a more indirect route, introducing an additional layer to the analysis. Within the two-stage approach, researchers have proposed specific methods to navigate the intricacies of evaluating higher-order constructs. Two noteworthy methods are the embedded two-stage approach [27] and the disjoint two-stage approach [27–29]. In the context of our study, we adapted the disjoint two-stage approach. The disjoint two-stage approach involves initially assessing the lower-order component of the higher-order construct. In the first stage, the lower-order component is evaluated, and its latent score is calculated and saved. Then the saved score is used to assess the higher-order construct. Table 1 shows the summary of measurement specifications of the lower-order component and higher-order construct for the reflective-reflective and formative-reflective models as illustrated by Sarstedt et al. [30].
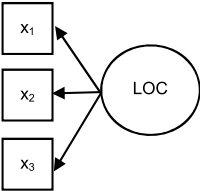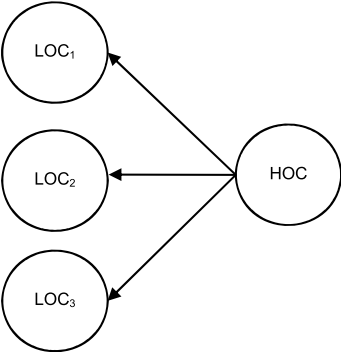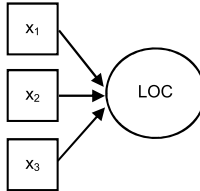
Proposed theoretical and conceptual framework of pls-sem panel survey data

Panel survey data offers a dynamic perspective on social phenomena by capturing both the breadth of cross-sectional data and the depth of time series data. In this study, the theoretical framework (Fig. 3) centres on employing Partial Least Squares Structural Equation Modelling (PLS-SEM) with a higher–order model, demonstrating its aptness for panel survey data. The comprehensive framework includes reflective–reflective and formative–reflective models, capturing the complexities of relationships among latent constructs. The estimation of this intricate model is facilitated by the disjoint two-stage approach, a methodical strategy that enhances precision in the analysis. This approach allows separate estimation of the lower-order components and the higher-order construct. It ensures a careful examination of individual constructs and their interrelationships within the specified models.

PLS-SEM is a powerful statistical method that excels in analysing complex relationships among latent variables. The adaptability of PLS-SEM is a cornerstone of its effectiveness in panel survey data analysis. PLS-SEM not only facilitates the evaluation of indicators for constructs but also flexibly empowers researchers to model higher-order constructs. This unique flexibility, making it well-suited for panel survey data analysis, allows for a holistic examination of the indicators for each repeated construct. In conjunction with this analytical framework, we incorporate Latent Growth Curve Models (LGCM), a class of statistical models designed for analysing the trajectories (individual and overall) of latent variables over time. Strengthening this approach, the conditional model in LGCM goes beyond merely describing the trajectories as it allows researchers to assess potential predictor variables and examine the factors influencing those trajectories.

The integration of PLS-SEM with conditional LGCM presents a powerful strategy for advancing panel survey data analysis, offering a range of benefits that collectively enhance the depth and precision of longitudinal research. PLS-SEM algorithm evaluates the reliability and validity of the indicators for each repeated construct, ensuring a robust assessment of the measurement model.

Table 1
Measurement Specification of the Higher-order Construct.

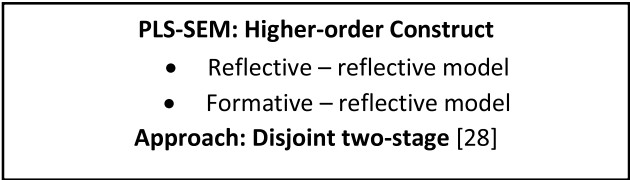| Measurement specification of the lower-order component	Measurement specification of the higher-order component
Reflective • Internal consistency (Cronbach's alpha, composite reliability) • Convergent validity (indicator reliability, average variance extracted) • Discriminant validity	Reflective • Internal consistency (Cronbach's alpha, composite reliability) • Convergent validity (indicator reliability, average variance extracted) • Discriminant validity

Fig. 3. Theoretical framework of PLS-SEM with higher-order constructs.

Additionally, LGCM contributes to the analysis by evaluating not only the trajectories but also examining the potential predictor variables that may influence the observed trajectories. This new approach establishes a comprehensive framework (Fig. 4) for understanding the complexities of panel survey data, encompassing measurement reliability and validity, trajectories, and exploring influential factors within the longitudinal context.

Integrated model specification

This integrated model has adapted a disjoint two-stage approach strategy for estimating the model, consisting of three sequential stages (Fig. 5). In stage 1, the process begins with specifying the measurement model, which can be either reflective or formative. For reflective models, evaluation criteria include indicator reliability, internal consistency (assessed using Cronbach's alpha and composite reliability), convergent validity (measured through average variance extracted), and discriminant validity (ensuring constructs that are supposed to be unrelated are distinct from each other). For formative models, the criteria include convergent validity (assessing the relationship between formative indicators and their latent construct), collinearity between indicators (ensuring that indicators are not
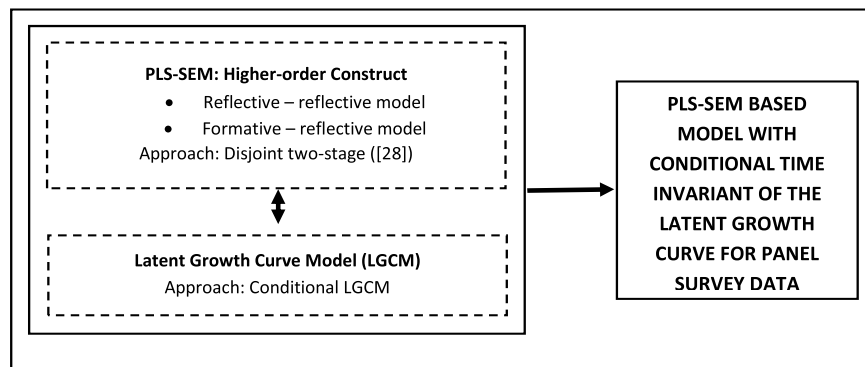
Fig. 4. Conceptual framework integrating PLS-SEM and LGCM for analysing panel survey data.
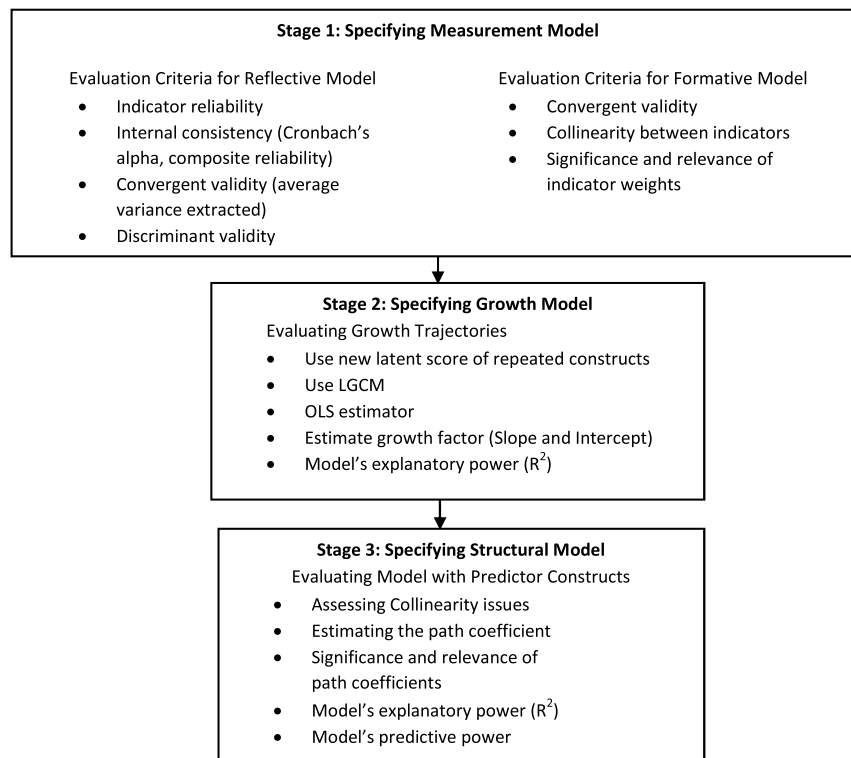

Fig. 5. Flow in the Integrated Model.

excessively correlated), and the significance and relevance of indicator weights (evaluating the importance of each indicator in forming the construct). These criteria align with recent guidelines recommending Confirmatory Composite Analysis (CCA) for assessing reflective measurement models in PLS-SEM [31].

Moving to the second stage, this integrated model focuses on specifying the growth model, where the latent growth factors (intercept and slope) are evaluated. The Ordinary Least Squares (OLS) estimator is used to estimate the model parameters, and the model's explanatory power (R^2) is assessed to determine how well the model explains the variability in the outcome.

In stage 3, the structural model is specified and evaluated. This includes assessing collinearity issues to ensure that predictor variables are not highly collinear, estimating path coefficients to calculate the strength and direction of relationships between constructs (predictors and latent growth factors), and determining the significance and relevance of these path coefficients. The model's explanatory power (R^2) is re-evaluated to see how well it explains the variance in the latent growth factor, and the model's predictive power is assessed to determine its ability to predict future observations. Finally, bootstrapping is employed to evaluate the statistical significance of the model parameters. This resampling method involves repeatedly sampling from the data with replacement to create a distribution of the parameter estimates, which helps in evaluating the precision of the estimates and testing the stability and reliability of the model.

Integrated model evaluation

This section discusses the criteria and the rules of thumb for each stage as illustrated in Fig. 5.

Stage 1: Specifying Measurement Model

Reflective measurement model

Fig. 6 shows the reflective measurement model evaluation procedure. Reflective measurement models are evaluated by looking at measures' internal consistency dependability and their level of reliability (indicator reliability). The average variance extracted (AVE) is used in the validity assessment process to target each measure's convergent validity. Furthermore, the discriminant validity of a reflectively measured construct is assessed relative to other construct measures in the same model using the heterotrait-monotrait (HTMT) ratio of correlations.

In step 1, reflective indicator loadings are used to assess the relationship between latent constructs and their observed indicators. The threshold for reflective indicator loadings is set at ≥ 0.708. This criterion ensures that each indicator sufficiently explains the variance of the underlying construct. A loading of 0.708 implies that approximately 50 % of the variance in the indicator is explained by the construct (since $0.708^2 \approx 0.50$). Indicators with loadings below this threshold may need to be reevaluated or potentially excluded to maintain the construct's validity.

Moving to step 2, internal consistency reliability measures the extent to which indicators of a construct are correlated, providing a gauge of the construct's overall reliability. Cronbach's alpha and composite reliability (rho$_c$) are the primary metrics for this criterion. Cronbach's alpha is considered the lower bound, while rho$_c$ serves as the upper bound for internal consistency reliability. The reliability coefficient rho$_A$, typically lying between these bounds, offers a balanced representation of internal consistency. The thresholds for internal consistency reliability are a minimum of 0.70 (or 0.60 in exploratory research), a maximum of 0.95 to avoid indicator redundancy and compromised content validity, and a recommended range of 0.80 to 0.90. These thresholds ensure a balance between sufficient internal consistency by avoiding excessive redundancy among indicators.

Step 3 involves convergent validity assessing whether indicators of a construct that are theoretically related are indeed related in practice. The average variance extracted (AVE) is the metric used to evaluate convergent validity, with a threshold of AVE ≥ 0.50. An AVE of 0.50 or higher indicates that at least 50 % of the variance in the indicators is explained by the construct, signifying adequate convergent validity.

Finally, discriminant validity examines whether constructs that are supposed to be unrelated are indeed distinct from one another. The heterotrait-monotrait ratio of correlations (HTMT) is employed to evaluate discriminant validity. The thresholds for HTMT are less than 0.90 for conceptually similar constructs and less than 0.85 for conceptually different constructs. Additionally, it is important to test if the HTMT value is significantly lower than the specified threshold, ensuring that constructs are not only theoretically but also empirically distinct.

Formative measurement model

Fig. 7 shows the formative measurement model evaluation procedure. Formative measurement models are evaluated using several criteria, including convergent validity, collinearity, the statistical significance of indicator weights, and the relevance of indicators.

Firstly, the convergent validity in formative measurement models is assessed through redundancy analysis. This involves examining the correlation between the formative construct and a reflective (or single item) measurement of the same concept. The threshold for adequate convergent validity is a correlation of ≥ 0.708. This high correlation ensures that the formative construct adequately captures the same concept as the reflective measure, indicating that the indicators collectively represent the construct well.

Moving to the second step, the collinearity refers to the degree to which indicators are correlated with each other. High collinearity can pose significant issues in formative measurement models, potentially distorting the results and interpretations. The variance inflation factor (VIF) is used to assess collinearity, with critical collinearity issues likely occurring if VIF is 5 or greater, collinearity issues usually being uncritical if VIF is between 3 and 5, and collinearity not being problematic if VIF is less than 3. Maintaining VIF values below these thresholds is crucial to ensure that collinearity does not adversely affect the model's validity and reliability.

Finally, the statistical significance of indicator weights is a crucial aspect of formative measurement models. The t-values of indicator weights should be greater than $2.576(\alpha = 0.01)$, $1.960(\alpha = 0.05)$, *or* $1.645(\alpha = 0.10)$ respectively, for two-tailed tests. Furthermore, the 95 % percentile confidence interval ($\alpha = 0.05$) should not include zero. These criteria ensure that the weights of the indicators are statistically significant, confirming their contribution to the construct. Indicators with significant weights contribute meaningfully to the construct, and the larger the significant indicator weight, the higher the relative contribution of that indicator to the construct. This relevance indicates that certain indicators play a more crucial role in defining the construct, and their significance

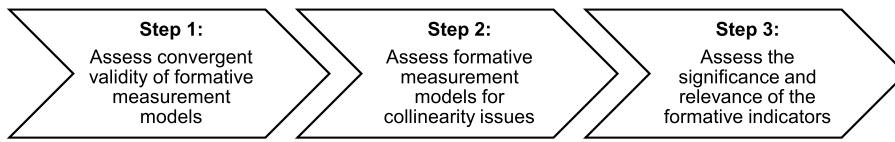**Fig. 6.** Reflective measurement model evaluation procedure.

Fig. 7. Formative measurement model evaluation procedure.

should be carefully considered in the model assessment. Even if an indicator's weight is not statistically significant, it can still be relevant if it has a loading of 0.50 or greater and is statistically significant. Such indicators are considered important for the construct as they represent substantial aspects of the concept being measured. This criterion ensures that indicators with moderate but significant contributions are not overlooked in the assessment process.

Besides evaluating the measurement model, the latent scores for each construct are also estimated and saved for subsequent analyses.

Stage 2: Specifying Growth Model

Building on the latent scores estimated in Stage 1, the second stage focuses on specifying the growth model, where the latent growth factors (intercept and slope) are evaluated. The unconditional latent growth curve model (Eq. (1) to Eq. (3)) is established, which involves only repeated constructs. To estimate this model, key parameters include the individual-specific intercepts $(\widehat{\alpha}_i)$ and slopes $(\widehat{\beta}_i)$, their population means (μ_α, μ_β), residual variance $(Var(\epsilon))$ and the variances of the random effects $(\psi_{\alpha\alpha}, \psi_{\beta\beta})$. This model assumes a linear growth trajectory, where change in the repeated constructs occurs at a constant rate over time, and equally spaced time scores are applied using fixed loadings $\lambda_t = t - 1$. These assumptions correspond to a conventional linear latent growth model in which a single intercept and slope represent each individual's trajectory.

In addition, the model assumes that residuals (ϵ_{it}) are independently and identically distributed with constant variance and zero mean $(\epsilon_{it} \sim N(0, \sigma^2))$, and that the random intercepts and slopes are drawn from a multivariate distribution. However, because the latent scores used in this stage are obtained from the PLS-SEM algorithm using Ordinary Least Squares (OLS), the estimation does not require multivariate normality. This feature aligns with the distribution-free nature of PLS-SEM and supports the use of non-parametric bootstrapping in the final structural model stage. Although the linear trajectory provides a parsimonious and interpretable framework, future studies should extend the model to accommodate non-linear change patterns. For example, quadratic growth models capture acceleration or deceleration trends, and piecewise models allow examination of segmented growth phases with varying slopes. These potential extensions enhance the model's flexibility and provide a more nuanced representation of individual developmental trajectories over time.

Following the measurement model evaluation in Stage 1, the standardized latent scores of the repeated constructs are extracted. These scores represent each individual's estimated position on the latent variable at each time point. To prepare these scores for trajectory estimation, reverse standardization is applied. The standardized scores are multiplied by the standard deviation and then added to the mean of the original observed construct values. This step restores the latent scores to the original scale of the construct.

Next, the latent growth factors (α_i and β_i) are estimated using ordinary least square estimator (OLS). This is done by treating the reverse-standardized latent scores as outcomes and using time scores λ_t as predictors. This can be expressed in the Eq. (7) and Eq. (8). This estimation produces the intercept and slope for individual i.

$$\widehat{\beta}_i = \frac{\sum_{t=1}^{T}(\lambda_t - \overline{\lambda})(y_{it} - \overline{y}_i)}{\sum_{t=1}^{T}(\lambda_t - \overline{\lambda})^2} \tag{7}$$

$$\widehat{\alpha}_i = \overline{y}_i - \widehat{\beta}_i\overline{\lambda} \tag{8}$$

The resulting coefficients α_i and β_i represent the individual's initial status and rate of change, respectively. The mean trajectory for the entire sample is estimated by averaging these parameters as shown in Eq. (9) and Eq. (10).

$$\widehat{\mu}_\alpha = \frac{\sum_{i=1}^{N}\widehat{\alpha}_i}{N} \tag{9}$$

$$\widehat{\mu}_\beta = \frac{\sum_{i=1}^{N}\widehat{\beta}_i}{N} \tag{10}$$

Next, to estimate the differences of trajectories between the individual, the variance of intercept and slope is assessed. The equation for assessing the variance is shown in Eq. (13) and Eq. (14).

$$var(\epsilon) = \frac{\sum_{i=1}^{N}var(\epsilon_i)}{N} \tag{11}$$

$$var(\epsilon_i) = \frac{\sum_{t=1}^{T} e_{it}^2}{T - 2} \tag{12}$$

$$\widehat{\psi}_{\alpha\alpha} = var(\widehat{\alpha}) - \frac{var(\epsilon)\sum_{t=1}^{T}\lambda_t^2}{\left[\sum_{t=1}^{T}(\lambda_t - \overline{\lambda})^2\right]T} \tag{13}$$

$$\widehat{\psi}_{\beta\beta} = var(\widehat{\beta}) - \frac{var(\epsilon)}{\sum_{t=1}^{T}(\lambda_t - \overline{\lambda})^2} \tag{14}$$

Then, the R^2 is used to evaluate the model fit. R^2 values for each regression provide information on the closeness of the linear trajectory to the data points observed. This second stage thus translates PLS-estimated latent scores into individual growth trajectories while retaining robustness to non-normal data. The estimated growth parameters are subsequently analyzed in Stage 3 to identify significant predictors within the structural model using the PLS-SEM framework.

Stage 3: Specifying Structural Model

In the third stage, the structural model is evaluated based on the PLS-SEM framework as illustrated in Fig. 8. This stage involves evaluating potential predictor variables towards the trajectories by employing the OLS estimator. The latent growth factors (intercept and slope) will regress on the potential predictor variables. The coefficients are estimated, and the structural model must assess the potential of collinearity issues before proceeding to the subsequent tests, which are to examine the significance of these potential predictor variables. The significance of the coefficients is based on the bootstrapping standard error as a basis for calculating the t-values of path coefficients. This stage plays a crucial role in understanding the factors influencing the trajectories. Finally, the overall significance of the model is evaluated using the R^2 value, which offers a comprehensive perspective on the explanatory power of the entire integrated model.

The collective findings from these three stages contribute to a nuanced understanding of the evaluation of the measurement model, the mean trajectory of the entire group, the evaluation of individual differences in trajectories, and the potential incorporation of predictors of individual differences in trajectories.

Analysis using R software

The proposed integrated model is implemented in R using a custom function, "***pls_growth***", developed specifically for this framework. The function enables modelling growth trajectories within the PLS-SEM context, supported by the SEMinR package. Table 2 shows how to install and begin using the PLSgrowth package. Additionally, various functions from the SEMinR package, such as *construct, relationships, and bootstrap_model,* were used to define the measurement model, estimate the structural model, and perform bootstrapping analysis, respectively. Table 3 illustrates the use of the pls_growth function and several functions in the SEMinR package.

Method validation

This section illustrates the proposed integrated model using data from 300 samples collected across three waves of the Midlife Development in the United States study (1995–1996, 2004–2006, and 2011–2014). In this empirical example, we employed a reflective-reflective model, which focuses on how indicators reflect the underlying constructs, as opposed to the formative-reflective model. The repeated construct in this model is life satisfaction (SAT1, SAT2 and SAT3), measured over time. Additionally, the model includes three predictor constructs that serve as time-invariant covariates, namely psychological well-being (PSYCO), social well-being (SOC), and household income (INCOME). This integrated approach aims to examine the trajectories or rate of change in life satisfaction and how these predictors influence life satisfaction across the different waves of the study.

In the first stage, we create the initial model to evaluate the measurement model by connecting all the latent exogenous variables to the lower-order component known as latent repeated constructs as displayed in Fig. 9. Then, the model needs to satisfy all the criteria mentioned earlier in Fig. 6, and the results are shown in Table 4 and Table 5. The latent score for the repeated constructs is also saved, which will be used in the next stage of the analysis to further explore their trajectories on life satisfaction over time.

In the second stage, latent growth estimates were derived by regressing repeated measures of life satisfaction over time. This approach enabled the estimation of individual trajectories using two growth parameters, namely the intercept, which represents the initial status, and the slope, which captures the rate of change over time. The distribution of intercepts, as visualised in Fig. 10, shows that individuals began with a high level of life satisfaction. The mean intercept in Table 6 was 7.5228 (SE = 0.0717, 95 % CI [7.3823,

Fig. 8. Structural model evaluation procedure.

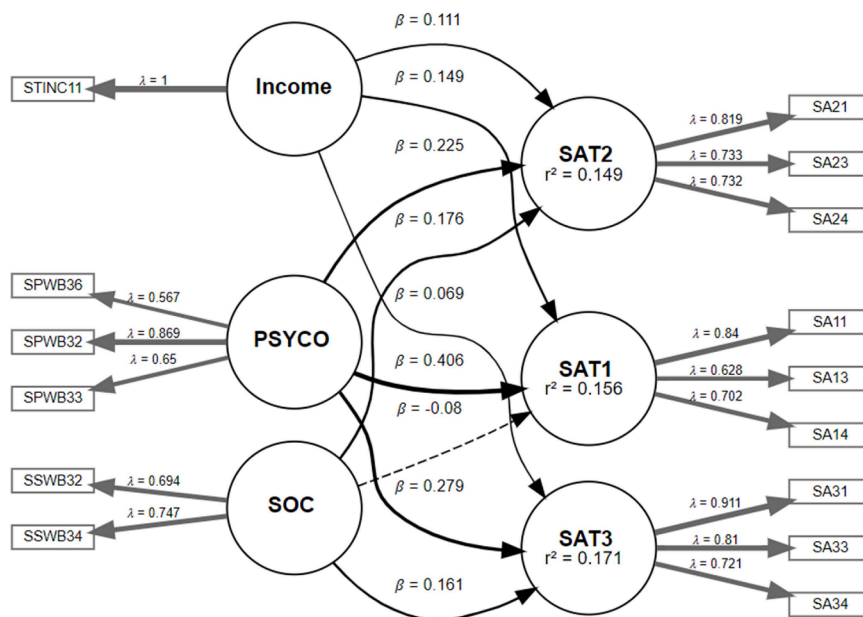**Table 2**

Installation of the PLSgrowth package.

Install the devtools package if not already installed.	*install.packages("devtools")*
Install PLSgrowth from GitHub.	*devtools::install_github("MrZMG/PLSgrowth")*
Load the package into your R session.	*library(PLSgrowth)*

Table 3

Summary of R code for the proposed integrated model.

Define the measurement and structural model using constructs and relationships function from SEMinR package.	sat_mm <- constructs(reflective("SAT1", multi_items("SA1", c(1, 3, 4))), reflective("SAT2", multi_items("SA2", c(1, 3, 4))), reflective("SAT3", multi_items("SA3", c(1, 3, 4))), reflective("PSYCO", multi_items("SPWB3", c(2, 3, 6))), reflective("SOC", multi_items("SSWB3", c(2, 4))), reflective("Income", single_item("STINC11")), higher_composite("GROWTH", c("SAT1", "SAT2", "SAT3"))) sat_sm <- relationships(paths(from = "PSYCO", to = "GROWTH"), paths(from = "Income", to = "GROWTH"), paths(from = "SOC", to = "GROWTH"))
Estimate the measurement, growth and structural model using '*pls_growth*' function.	sat_pls <- pls_growth(data = data, measurement_model = sat_mm, structural_model = sat_sm, missing = mean_replacement, include_tvcs = FALSE, missing_value = FALSE)
Bootstrap for significance and relevance of path coefficients.	boot <- bootstrap_model(seminr_model = sat_pls, seed = 123)

Fig. 9. Evaluation of the measurement model in the first stage.

7.6632]), indicating that the typical participant perceived themselves as moderately to highly satisfied with life at the initial measurement occasion. However, the variance in intercepts, recorded at 0.8642, shows considerable interindividual variability in baseline levels of life satisfaction. This means that while some participants reported relatively low initial satisfaction, others reported much higher values.

Additionally, the slope distribution, presented in Fig. 11, reveals more pronounced heterogeneity in the direction and rate of change. The mean slope in Table 6 was estimated at −0.2061 (SE = 0.0472, 95 % CI [−0.2987, −0.1135]), indicating that participants

Table 4
Reliability and validity statistics of constructs.

	Cronbach's α	ρ_A	Composite reliability ρ_c	Average variance extracted (AVE)
PSYCO	0.743	0.776	0.743	0.500
SOC	0.682	0.685	0.683	0.519
SAT1	0.766	0.785	0.770	0.531
SAT2	0.804	0.809	0.806	0.581
SAT3	0.854	0.867	0.857	0.669

Table 5
Discriminant validity assessment using the HTMT criterion.

	PSYCO	Income	SOC	SAT1	SAT2	SAT3
PSYCO						
Income	0.051					
SOC	0.704	0.018				
SAT1	0.355	0.17	0.191			
SAT2	0.342	0.12	0.324	0.661		
SAT3	0.395	0.077	0.346	0.478	0.638	

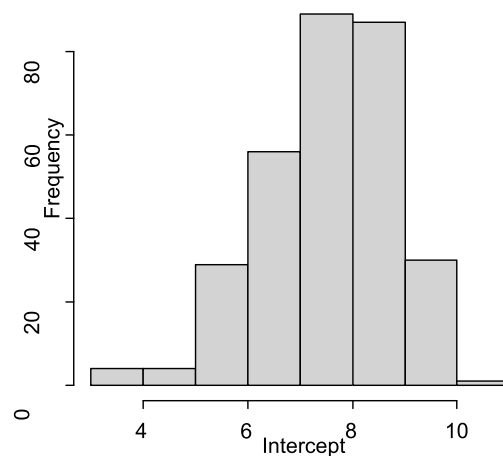**Fig. 10.** Histogram of intercept of life satisfaction.

Table 6
Parameter estimate for the growth model.

Parameter	Estimate
$var(\epsilon)$	0.8327
μ_α	7.5228 (0.0717)
μ_β	$-0.2061(0.0472)$
$\psi_{\alpha\alpha}$	0.8462
$\psi_{\beta\beta}$	0.2530
CI_{μ_α}	(7.3823, 7.6632)
CI_{μ_β}	$(-0.2987, -0.1135)$

experienced a decline in life satisfaction throughout the study period. The variance in slope values, which reached 0.2530, indicates that while some participants exhibited a downward trend, others either maintained stability or showed improvement. This is evident from the presence of both positive and negative slopes in the histogram.

Such interindividual variability is a key advantage of the latent growth modelling approach, as it emphasises that not all participants follow the same developmental pattern. The combination of declining, stable, and improving trajectories indicates that a variety of personal, social, and environmental factors influence life satisfaction during midlife. These findings demonstrate diverse developmental experiences across individuals.

Regarding theoretical relevance, the overall decline in life satisfaction can be understood through lifespan development theories, which suggest that midlife is often marked by increasing life demands, transitions, and role strain, all of which could negatively impact
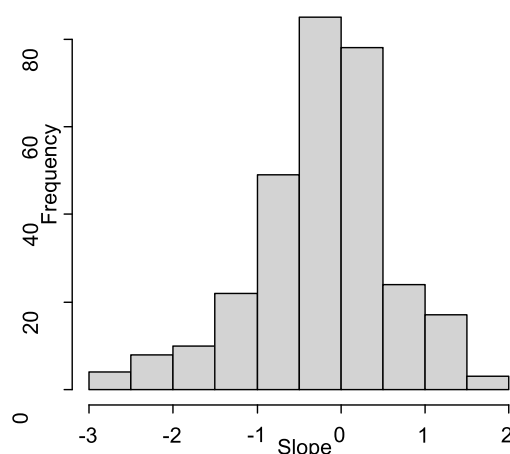
Fig. 11. . Histogram of slope of life satisfaction.

subjective well-being. Simultaneously, the significant variability in growth slopes indicates the influence of buffering or enhancing factors, such as resilience, coping resources, or social support. This variability emphasises the practical importance of identifying conditions that promote more positive developmental pathways in life satisfaction.

Finally, the structural model is evaluated by addressing the potential predictor construct that influences the trajectories. This evaluation is made by regressing the slope and intercept over the potential predictor construct. Then, the bootstrap method is employed to assess the significance of the path coefficients based on t-values or confidence intervals, as shown in Fig. 12. Fig. 12 illustrates the final stage of the integrated model, where the structural relationships between the predictor constructs (psychological well-being, social well-being, and income) and the two latent growth parameters (intercept and slope of life satisfaction) are examined. The path diagram highlights the standardized path coefficients and their corresponding levels of significance obtained via bootstrapping. In this empirical application, psychological well-being (PSYCO) exhibited a substantial and statistically significant positive effect on the intercept, indicating that individuals with higher baseline psychological well-being also reported higher initial life satisfaction. Similarly, PSYCO demonstrated a small but significant negative effect on the slope, suggesting a buffering role in the decline of life satisfaction over time. Meanwhile, social well-being (SOC) showed a significant positive influence on both the intercept and slope, implying that higher levels of perceived social support are associated with initially higher life satisfaction and a more stable or less negative trajectory across time. Lastly, household income was found to be significantly associated only with the intercept, but not the slope, indicating its role in shaping the baseline level of life satisfaction rather than the rate of change. Overall, these findings validate the capacity of the proposed integrated model to disentangle the differential effects of predictors on baseline status and developmental trajectories, offering meaningful insights into the dynamic processes of well-being over time.

Conclusion

The integration of Partial Least Squares Structural Equation Modelling (PLS-SEM) with the Latent Growth Curve Model (LGCM) proposed in this study offers a novel and robust approach for analysing panel survey data. This integrated model overcomes several limitations of traditional PLS-SEM, particularly in longitudinal research, by enabling the evaluation of growth trajectories, individual differences, and potential predictors over time. The use of the disjoint two-stage approach for higher-order constructs, combined with the flexibility of PLS-SEM and the dynamic analysis capabilities of LGCM, provides a comprehensive framework for examining longitudinal data.

Compared to conventional latent growth curve modelling based on covariance-based SEM, the proposed model offers several methodological advantages. Traditional LGCM assumes multivariate normality and typically focuses on reflective constructs, which limit flexibility in handling more complex model structures. In contrast, this integrated approach capitalises on the distribution-free nature of PLS-SEM, allowing both reflective and formative measurement models, and is well-suited for small to moderate sample sizes and non-normal data. By estimating latent scores through PLS in Stage 1 and using these to model growth trajectories in Stage 2, followed by structural evaluation in Stage 3, the model supports a modular yet unified analysis of measurement, growth, and prediction components. This makes it a practical and flexible alternative for researchers dealing with panel survey data. Several key contributions that emphasise the novelty of this approach are shown in Table 7.

This new approach not only improves the reliability and validity of panel data analysis but also provides deeper insights into individual trajectories and the factors that influence changes over time. As a result, this study offers a valuable contribution to both the methodological literature and practical applications in the fields of behavioural sciences, education, and social research, where understanding dynamic processes over time is critical.
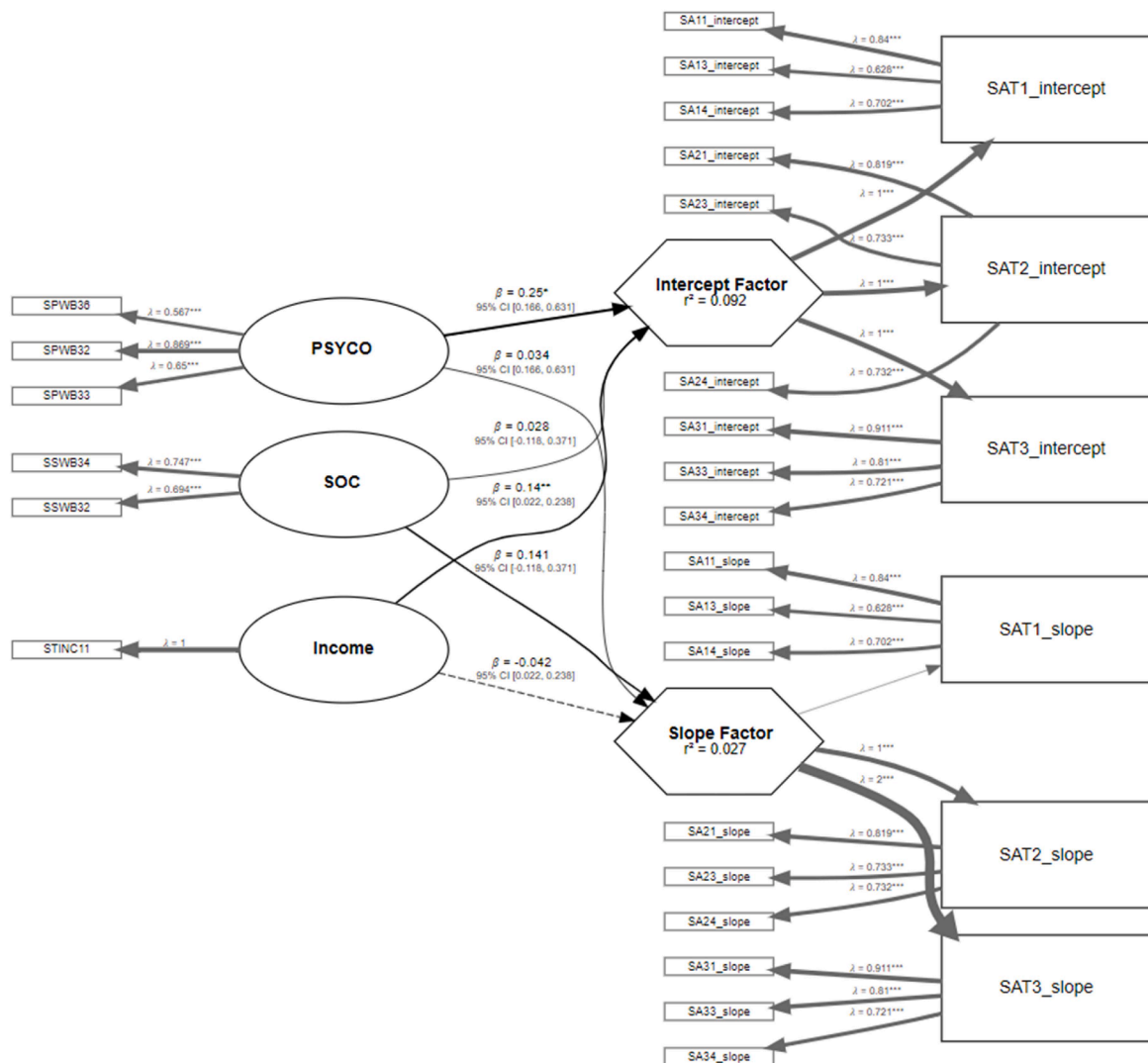
Fig. 12. Bootstrapping of Integrated PLS-SEM with LGCM results.

Table 7
Elaboration on the contributions to this study.

Methodological Innovation	The combination of PLS-SEM with LGCM is a novel advancement, allowing for a more thorough analysis of the reliability and validity of construct and growth trajectories within the same model while capturing complex relationships and changes over time.
Extension to Longitudinal Data	By adapting PLS-SEM for panel survey data, this study expands the method's applicability beyond cross-sectional analyses. This method enables longitudinal modelling that accommodates growth trajectories and latent constructs simultaneously and evaluates the factors that influence trajectories.
Enhanced Analytical Framework	The use of the disjoint two-stage approach for higher-order constructs within the integrated model adds a methodological contribution. The detailed discussion of measurement model specifications, growth model estimation, and structural modelling in a longitudinal context offers greater precision and flexibility in modelling complex relationships in panel data.
Practical Implementation in R	The development of the custom ***pls_growth*** function in R, alongside the SEMinR package, facilitates the application of this integrated model, making it accessible to researchers and practitioners who work with longitudinal data.

Limitations

- This study focuses on linear growth models to examine changes over time. It does not consider non-linear or polynomial models, which better capture complex developmental patterns such as acceleration or deceleration in certain variables.

- The proposed integrated model emphasises time-invariant covariates (TICs) to predict individual differences in growth trajectories. It does not incorporate time-varying covariates (TVCs), which may offer more detailed insights into how dynamic predictors influence changes over time.

Future work

Future research may expand upon the current framework by incorporating non-linear growth trajectories, including quadratic and piecewise models, to capture more intricate developmental patterns over time. The integration of time-varying covariates (TVCs) could also enhance the model's ability to examine within-person changes and dynamic influences across measurement occasions. Additionally, extending the framework to multilevel or hierarchical structures may allow for more nuanced analyses of nested data commonly found in education, health, and organisational studies. Further methodological advancements could explore moderation and mediation effects within the longitudinal context, providing a richer understanding of causal mechanisms and conditional relationships over time. Applying the proposed model in various empirical domains will also help evaluate its generalizability and practical utility across diverse fields.

Related research article

LGCM and PLS-SEM in Panel Survey Data: A Systematic Review and Bibliometric Analysis

For a published article

Mohd Ghazali, Z., Wan Yaacob, W. F., & Wan Omar, W. M. (2023). LGCM and PLS-SEM in Panel Survey Data: A Systematic Review and Bibliometric Analysis. Data, 8(2), 32. https://doi.org/10.3390/data8020032

Ethics statements

This study used an open-source dataset from samples collected across three waves of the Midlife Development in the United States study (1995–1996, 2004–2006, and 2011–2014)
https://www.icpsr.umich.edu/

Supplementary material *and/or* additional information [OPTIONAL]

To facilitate replication and understanding of the proposed model, a simulated dataset and annotated R script have been made publicly available on GitHub: https://github.com/MrZMG/PLSgrowth_time-invariant_supplement

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Data availability

Data will be made available on request.

References

[1] H. WOLD, Path Models with Latent Variables: the NIPALS Approach. Quantitative Sociology, Elsevier, 1975, pp. 307–357, https://doi.org/10.1016/b978-0-12-103950-9.50017-4.
[2] J.-B. Lohmöller, Latent Variable Path Modeling with Partial Least Squares, Physica-Verlag HD, Heidelberg, 1989, https://doi.org/10.1007/978-3-642-52512-4.
[3] J.F. Hair Jr., M. Sarstedt, C.M. Ringle, S.P. Gudergan, Advanced Issues in Partial Least Squares Structural Equation Modeling, 2nd Edition, SAGE Publications, Thousand Oaks, CA, 2024.
[4] C. Fornell, F.L. Bookstein, Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory, J. Market. Res. 19 (1982) 440, https://doi.org/10.2307/3151718.
[5] J.F. Hair Jr., L.M. Matthews, R.L. Matthews, M. Sarstedt, PLS-SEM or CB-SEM: updated guidelines on which method to use, Int. J. Multiv. Data Anal. 1 (2017) 107, https://doi.org/10.1504/ijmda.2017.10008574.

[6] H.W. Willaby, D.S.J. Costa, B.D. Burns, C. MacCann, R.D. Roberts, Testing complex models with small sample sizes: a historical overview and empirical demonstration of what Partial Least Squares (PLS) can offer differential psychology, Pers. Individ. Dif. 84 (2015) 73–78, https://doi.org/10.1016/j.paid.2014.09.008.

[7] E. Roemer, A tutorial on the use of PLS path modeling in longitudinal studies, Industr. Manag. Data Syst. 116 (2016) 1901–1921, https://doi.org/10.1108/IMDS-07-2015-0317.

[8] M. Haenlein, A.M. Kaplan, A Beginner's Guide to Partial Least Squares Analysis, Understand. Statist. 3 (2004) 283–297, https://doi.org/10.1207/s15328031us0304_4.

[9] J. Henseler, G. Hubona, P.A. Ray, Using PLS path modeling in new technology research: updated guidelines, Industr. Manag. Data Syst. 116 (2016) 2–20, https://doi.org/10.1108/IMDS-09-2015-0382.

[10] C. Nitzl, W.W. Chin, The case of partial least squares (PLS) path modeling in managerial accounting research, J. Manag. Control 28 (2017) 137–156, https://doi.org/10.1007/s00187-017-0249-6.

[11] J.S. Michel, M.A. Rotch, J.E. Carson, N.A. Bowling, N.V. Shifrin, Flattening the Latent Growth Curve? Explaining Within-Person Changes in Employee Well-Being during the COVID-19 Pandemic, Occup. Health. Sci. 5 (2021) 247–275, https://doi.org/10.1007/s41542-021-00087-4.

[12] J. Milicev, P. Qualter, C. Goodfellow, J. Inchley, S.A. Simpson, A.H. Leyland, K. Kharicha, E. Long, The prospective relationship between loneliness, life satisfaction and psychological distress before and during the COVID-19 pandemic in the UK, J. Public Health (Germany) (2022), https://doi.org/10.1007/s10389-022-01719-x.

[13] Á. Szabó, J. Sheridan, D. Newcombe, Ten-Year Trajectories of Alcohol Consumption in Older Adult New Zealanders, J. Gerontol. - Series B Psychol. Sci. Soc. Sci 76 (2021) 496–506, https://doi.org/10.1093/geronb/gbz143.

[14] Z. Mohd Ghazali, W.F. Wan Yaacob, W.M. Wan Omar, LGCM and PLS-SEM in Panel Survey Data: a Systematic Review and Bibliometric Analysis, Data (Basel) 8 (2023) 32, https://doi.org/10.3390/data8020032.

[15] W. Meredith, J. Tisak, Latent Curve Analysis, Psychometrika 55 (1990) 107–122, https://doi.org/10.1007/BF02294746.

[16] T.E. Duncan, S.C. Duncan, An introduction to latent growth curve modeling, Behav Ther 35 (2004) 333–363, https://doi.org/10.1016/S0005-7894(04)80042-X.

[17] K.A. Bollen, P.J. Curran, Latent Curve models : a Structural Equation Perspective, Wiley-Interscience, 2006.

[18] M.S. Birkeland, T. Holt, S.M. Ormhaug, T.K. Jensen, Perceived social support and posttraumatic stress symptoms in children and youth in therapy: a parallel process latent growth curve model, Behav Res Ther 132 (2020) 103655, https://doi.org/10.1016/j.brat.2020.103655.

[19] D. Zhang, E.S. Huebner, L. Tian, Neuroticism and cyberbullying among elementary school students: a latent growth curve modeling approach, Pers Individ Dif 171 (2021) 110472, https://doi.org/10.1016/j.paid.2020.110472.

[20] M. Isiordia, E. Ferrer, Curve of Factors Model: a Latent Growth Modeling Approach for Educational Research, Educ. Psychol. Meas. 78 (2018) 203–231, https://doi.org/10.1177/0013164416677143.

[21] M.J. Tomasik, L.A. Helbling, U. Moser, Educational gains of in-person vs. distance learning in primary and secondary schools: a natural experiment during the COVID-19 pandemic school closures in Switzerland, Int. J. Psychol. 56 (2021) 566–576, https://doi.org/10.1002/ijop.12728.

[22] T.K. Lee, K.(K A.S.). Wickrama, C.W. O'Neal, Application of Latent Growth Curve Analysis With Categorical Responses in Social Behavioral Research, Struct. Equat. Modeling 25 (2018) 294–306, https://doi.org/10.1080/10705511.2017.1375858.

[23] W. Wu, S.G. West, A.B. Taylor, Evaluating Model Fit for Growth Curve Models: integration of Fit Indices From SEM and MLM Frameworks, Psychol. Methods 14 (2009) 183–201, https://doi.org/10.1037/a0015858.

[24] K.J. Preacher, A.L. Wichman, R.C. Maccallum, N.E. Briggs, LATENT GROWTH CURVE MODELING, SAGE Publications Inc., 2008.

[25] J.J. McArdle, Dynamic but Structural Equation Modeling of Repeated Measures Data. Handbook of Multivariate Experimental Psychology, Springer US, Boston, MA, 1988, pp. 561–614, https://doi.org/10.1007/978-1-4613-0893-5_17.

[26] Richard J. Wold, in: H.O.A. Jöreskog, K.G. Wold (Eds.), Soft Modeling: the Basic Design and Some Extensions, 1982, pp. 1–54. North-Holland.

[27] C.M. Ringle, M. Sarstedt, D.W. Straub, A Critical Look at the Use of PLS-SEM in "MIS Quarterly, MIS Quart. 36 (2012). http://www.jstor.org/stable/41410402.

[28] J.M. Becker, K. Klein, M. Wetzels, Hierarchical Latent Variable Models in PLS-SEM: guidelines for Using Reflective-Formative Type Models, Long Range Plann. 45 (2012) 359–394, https://doi.org/10.1016/j.lrp.2012.10.001.

[29] R. Agarwal, E. Karahanna, Time Flies When You're Having Fun: cognitive Absorption and Beliefs about Information Technology Usage, MIS Quart. 24 (2000) 665, https://doi.org/10.2307/3250951.

[30] M. Sarstedt, J.F. Hair, J.-H. Cheah, J.-M. Becker, C.M. Ringle, How to Specify, Estimate, and Validate Higher-Order Constructs in PLS-SEM, Australas. Market. J. 27 (2019) 197–211, https://doi.org/10.1016/j.ausmj.2019.05.003.

[31] J.F. Hair, M.C. Howard, C. Nitzl, Assessing measurement model quality in PLS-SEM using confirmatory composite analysis, J Bus Res 109 (2020) 101–110, https://doi.org/10.1016/J.JBUSRES.2019.11.069.