Idiographic Personality Gaussian Process for Psychological Assessment

Yehu Chen, Muchen Xi, Jacob Montgomery
Joshua Jackson, Roman Garnett
Washington University in St Louis
chenyehu,m.xi,j.jackson,jacob.montgomery,garnett@wustl.edu

Abstract

Developing taxonomies for psychological assessment is crucial for understanding long-term human behaviors. However, existing psychometric methods tend to be nomothetic, lacking individualization, or rely on static cross-sectional data that overlooks the dynamic nature of psychological processes. We introduce an idiographic personality Gaussian process (IPGP) framework for time-series survey data, by leveraging Gaussian process coregionalization to conceptualize individualized taxonomies and stochastic variational inference for computational scalability. Through an extensive simulation study against benchmark methods and an exploratory factor analysis study of life outcomes of personality replication, we demonstrate that IPGP can simultaneously improve estimation of idiographic taxonomies and prediction of missing responses. We also assess IPGP using our IRB-approved data with a forecasting and a leave-one-trait-out prediction task, illustrating how IPGP identifies unique taxonomies of personality that display potential in advancing individualized approaches to psychological diagnosis.

1 Introduction

Building standard taxonomies for psychological assessment is crucial to understand long-term behaviors through repeated quantitative surveys, for instance, emotional stability after medical treatment or development of academic ability during secondary education [Molenaar, 2004, Wang et al., 2013, Dumas et al., 2020]. However, existing taxonomies face several limitations. First of all, a common group of latent concepts is usually constructed for describing a single psychological trait yet with different causes. For instance, depression is constructed to explaining behavior of people suffering from numerous syndromes that differ in etiology, symptoms, and biological processes but get grouped together and called depression [Borsboom et al., 2003, Molenaar, 2004]. Second, conceptualization of taxonomies tend not to be individualized by assuming (1) the correlation between latent traits of interest and survey responses are invariant across individuals and (2) the actual structure of the underlying latent dimensions is fixed across individuals. Lastly, current established taxonomies are usually developed from cross-sectional data that are collected only once from respondents, and might overlook the dynamic nature of psychological processes.

To address these limitations, recent work have proposed an *idiographic* approach that builds completely distinct taxonomy for everyone [Borkenau and Ostendorf, 1998, Beck and Jackson, 2020, 2021], compared to the *nomothetic* approach where everyone is described by the same taxonomy. However, complete personalization may sacrifice general interpretability to clinicians as any possible population commonality is completely ignored. Another line of research focus on building dynamic psychometric models with time-series data via item response theory [Rijmen et al., 2003, Reise and Waller, 2009, Dumas et al., 2020], longitudinal structural equation [Little, 2013, Kim and Willson, 2014, Asparouhov et al., 2018], vectorized autoregression [Lu et al., 2018, Haslbeck et al., 2021] and

Preprint. Under review.

Gaussian process (GP) [Wang et al., 2005, Damianou et al., 2011, Dürichen et al., 2014, Duck-Mayr et al., 2020]. Yet all these models adopt the nomothetic approach by fixing the same taxonomic structure across individuals, and hence fail to identify any deviation from subgroups. Furthermore, there has been attempt to combine approaches for creating individualization while maintaining group commonality [Beltz et al., 2016], but prioritizes predicting responses to clinical survey batteries at the expense of studying the latent structures that are the focus of domain researchers.

In this work, we propose an idiographic personality Gaussian process (IPGP) framework for assessing dynamic psychological taxonomies from time-series survey data, and combine the nomethetic and idiographic approaches by deploying a common structure for explaining the typical circumstance and individual structures for permitting deviations into distinct forms. We leverage the Gaussian process coregionalization model to conceptualize responses of grouped survey batteries, adjusted to non-Gaussian ordinal data, and utilize IPGP for hypothesis testing of domain theories. Computationally, our framework also exploits the stochastic variational inference for latent factor estimation, contrasting with other GP measurement models relying on Gibbs sampling that may not scale efficiently to intensive longitudinal setups [Dürichen et al., 2014, Duck-Mayr et al., 2020].

To our knowledge, our work is the first multi-task Gaussian process latent variable model for dynamic idiographic assessment, compared to previous literature focusing on either static setup that ignores dynamics in latent processes [Borkenau and Ostendorf, 1998, Bonilla et al., 2007, Beck and Jackson, 2021] or single-task approach that disregards inter-battery correlation [Snelson and Ghahramani, 2005, Hensman et al., 2015]. Methodologically, our approach intersects Gaussian process latent variable model (GPLVM) [Lawrence, 2003], Gaussian process dynamic system (GPDM) [Damianou et al., 2011, Dürichen et al., 2014] and GP ordinal regression for likert-type survey data [Croasmun and Ostrom, 2011, Chu and Ghahramani, 2005]. Through an extensive simulation study against benchmark methods and an exploratory factor analysis study of life outcomes of personality replication, we demonstrate that IPGP can simultaneously improve estimation of idiographic taxonomies and prediction of missing responses. We also assess IPGP using our IRB-approved data with a forecasting and a leave-one-trait-out prediction task, illustrating how IPGP identifies unique taxonomies of personality that display potential in advancing individualized approaches to psychological diagnosis.

2 Backgrounds

We start by laying out the ordinal factor model for building standard taxonomy from survey experiments [Digman, 1997, Baglin, 2014]. We then briefly discuss several existing idiographic longitudinal models in psychological assessment, and review the model of Gaussian process.

Ordinal factor analysis. Consider the following scenario in survery experiments of $i \in \{1, \dots, N\}$ units repeatedly answering the same set of $j \in \{i, \dots, J\}$ batteries over $t \in \{1, \dots, T\}$ periods with ordinal observations $y_{ijt} \in \{1, \dots, C\}$ up to C levels. For example, the responses could be Likert-typed, ranging from "strongly disagree" to "strongly agree". The latent factor model posits that the jth underlying latent variable $f_j^{(i)}(t)$ for unit i at time t are factored as $\mathbf{w}_j^T \mathbf{x}_i(t)$, where $\mathbf{x}_i(t) \in \mathbf{R}^K$ are unit-level latent factors and $\mathbf{w}_j \in \mathbf{R}^K$ are factor loadings. The $f_j^{(i)}(t)$s are then mapped to ordinal responses via an ordered logit model: $p\big(y_{ijt} = c \mid f_j^{(i)}(t) = f\big) = \Phi(b_c - f) - \Phi(b_{c-1} - f)$ with threshold parameters $b_0 < \cdots < b_C$. Usually b_0 and b_C are fixed to $-\infty$ and $+\infty$ such that the resulted categorical probability vector sums to 1, while $b_1, \dots, b_{C-1}$ are allowed to move freely. Stacking $\mathbf{x}_i(t)$s, $\mathbf{w}_j$s and y_{ijt}'s into matrices $\mathbf{x}$, $\mathbf{w}$ and tensor $\mathbf{y}$, the joint likelihood can be written as $\mathcal{L}(\mathbf{y} \mid \mathbf{x}, \mathbf{w}) = \prod_i \prod_j \prod_t p\big(y_{ijt} \mid \mathbf{x}_i(t), \mathbf{w}_j\big)$, while model identification is guaranteed by the general rule of factor models with additional orthogonality and normalization constraints [Bollen, 1989]. This factor model is also known as item response model [Samejima, 1969, Van der Linden and Hambleton, 1997], which estimates parameters via maximum likelihood, weighted least squares or EM algorithm [Bock and Aitkin, 1981, Forero et al., 2009, Li, 2016].

Idiographic longitudinal assessment. In psychological assessment, the idiographic approach emphasizes *intrapersonal* variation by requiring distinct loadings $\mathbf{w}_j^{(i)}$, while the nomothetic approach identifies general *interpersonal* variation assuming shared factor loadings $\mathbf{w}_j$s [Salvatore and Valsiner, 2010]. In terms of data collection, the idiographic approach usually surveys each individual multiple times ($n = 1$ and large T) for learning personalized taxonomy rather than many individuals at a

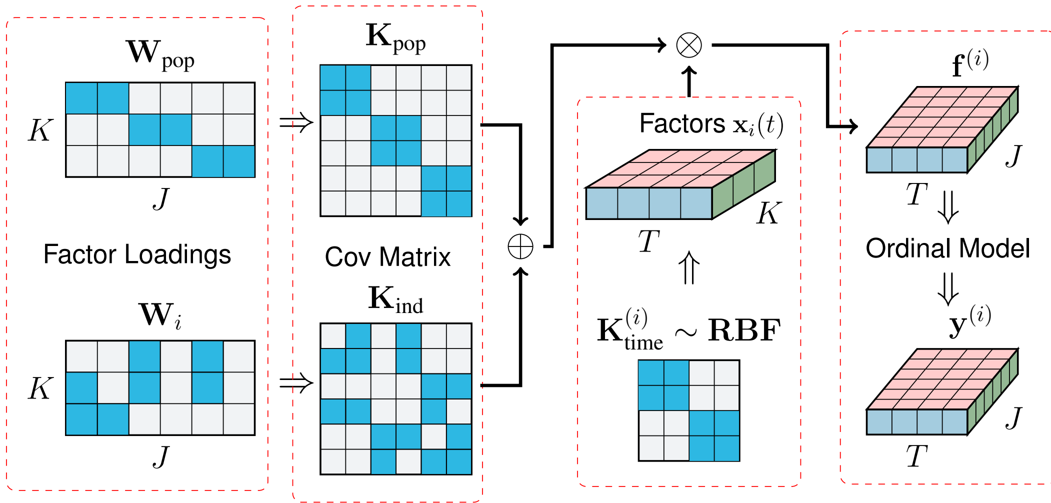

Figure 1: Proposed IPGP model for inferring latent factors and factor loadings from dynamic ordinal data. Input ordinal observations across channels are modeled as ordinal transformations of latent dynamic Gaussian processes with individualized RBF kernels and loading matrices.

single shot (large n and $T = 1$). To extract individualized dynamics from time-series data, recent psychometrics models have utilized longitudinal structural equations by explicitly specifying any intrapersonal and temporal relation yet sensitive to model mis-specification from domain theory [Little, 2013, Asparouhov et al., 2018]. Meanwhile, variants of hierarchical vector autoregression may automatically learn individual diffusion, but usually lack the ordered logit component as built in response space rather than latent space [Lu et al., 2018, Haslbeck et al., 2021].

Gaussian process. A Gaussian process (GP) can be used to define a distribution over f such that the evaluation of f at arbitrary subset of $\mathcal{X}$ is a joint multivariate Gaussian [Rasmussen and Williams, 2005]. To determine its mean and covariance, a $\mathcal{GP}(\mu, K)$ is specified with a mean function $\mu : \mathcal{X} \to \mathbf{R}$ and a positive-definite kernel function $K : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$. The most common kernel is the squared exponential (RBF) kernel $K(\mathbf{x_1}, \mathbf{x_2}) = \exp\left(-\frac{1}{2}\mathbf{x_1}^T\mathbf{P}\mathbf{x_2}\right)$ with precision matrix $\mathbf{P} = \mathbf{diag}(1/\ell_1^2, \dots, 1/\ell_d^2)$ and $d = \mathbf{card}(\mathcal{X})$. Posterior of a GP is usually analytical for Gaussian likelihood, but needs to be approximated in modeling latent variables. We discuss the variational approximation in Sec. (3).

3 Methodology

We propose an idiographic personality Gaussian process (IPGP) framework for assessing individualized dynamic psychological taxonomies from time-series survey data. Instead of joint estimation of latent factors and their loadings that cannot guarantee rotational and scaling invariance, we marginalize out the latent variables and focus on learning taxonomies of loadings. The overall architecture of IPGP is illustrated in Figure (1), where input ordinal observations across channels are modeled as ordinal transformations of latent dynamic GP with individualized RBF kernels and loading matrices.

3.1 Multi-task learning

Typically in psychological assessment, survey questions are meticulously grouped such that each group gauges a particular facet of personality. Hence, we conceptualize the assessment of psychological traits as a multi-task learning problem, where each question represents a distinct task but can be correlated with other tasks. A multi-task GP is an extension of the single-task GP but for vector-valued functions [Bonilla et al., 2007]. To motivate the multi-task framework, first consider the two-task scenario with two $T \times 1$ vector $\mathbf{f_1}^{(i)}$ and $\mathbf{f_2}^{(i)}$ denoting the latent temporal processes of unit i for question $j = 1, 2$. To fix the scale of latent factors, a time-level Gaussian process prior is placed on $\mathbf{x}_i(t) \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}^{(i)}_{\text{time}})$. Hence, by exploiting affine property of Gaussian, the induced joint distribution of vectorized $[\mathbf{f_1}^{(i)}, \mathbf{f_2}^{(i)}]^T$ can be written as:

$$\begin{bmatrix}\mathbf{f_1}^{(i)} \\ \mathbf{f_2}^{(i)}\end{bmatrix} \sim \mathcal{GP}\Big(\begin{bmatrix}\mathbf{0} \\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{w}_1^T\mathbf{w}_1\mathbf{K}_{\text{time}}^{(i)} & \mathbf{w}_1^T\mathbf{w}_2\mathbf{K}_{\text{time}}^{(i)} \\ \mathbf{w}_2^T\mathbf{w}_1\mathbf{K}_{\text{time}}^{(i)} & \mathbf{w}_2^T\mathbf{w}_2\mathbf{K}_{\text{time}}^{(i)}\end{bmatrix} \Big) \tag{1}$$

whose covariance of shape $2T \times 2T$ contains four block matrices $\mathbf{K}_{\text{time}}^{(i)}$ scaled by different $\mathbf{w}_j^T\mathbf{w}_{j'}$ $(j, j' \in \{1, 2\})$. Specifically, $\mathbf{w}_1^T\mathbf{w}_2$ controls the inter-task covariance between these two tasks and $\mathbf{w}_j^T\mathbf{w}_j$s $(j \in \{1, 2\})$ control their intra-task variance. This multi-task structure is also known as the linear model of coregionalization (LMC) [Alvarez et al., 2012], where the factor structure can be recovered from the relations $\mathbf{f_j}^{(i)} = \mathbf{w}_j^T\mathbf{x}_i(t)$ of linear combinations. Compared to vector autoregression, our GP dynamic system approach can better handle intrapersonal temporal structures. Now let $\mathbf{f}^{(i)} = [\mathbf{f_1}^{(i)}, \ldots, \mathbf{f_J}^{(i)}]^T$ represents the flattened $JT \times 1$ vector consisting of all J tasks. We write $\mathbf{f}^{(i)}$ in a formal multi-task GP notation using Kronecker product $\otimes$:

$$p(\mathbf{f}^{(i)}) \sim \mathcal{GP}\big(\mathbf{0}, \mathbf{K}_{\text{task}}^{(i)} \otimes \mathbf{K}_{\text{time}}^{(i)}\big) \tag{2}$$

$$\mathbf{K}_{\text{task}}^{(i)} = \mathbf{W}_{\text{pop}}^T\mathbf{W}_{\text{pop}} + \mathbf{w}_i^T\mathbf{w}_i + \text{diag}(\mathbf{v}) \tag{3}$$

where $\mathbf{K}_{\text{task}}^{(i)}$ denotes the unit-individualized task kernel, consisting of the self inner products of population loading $\mathbf{W}_{\text{pop}} = [\mathbf{w}_1, \ldots, \mathbf{w}_J]$ for explaining the interpersonal commonality and $J \times 1$ idiographic loading $\mathbf{w}_i$ for intrapersonal deviations, as well as a task-dependent noise component $\text{diag}(\mathbf{v}) = \text{diag}([\sigma_1^2, \ldots, \sigma_J^2])$. The Kronecker product $\otimes$ then multiplies each entry in the $J \times J$ task covariance with $\mathbf{K}_{\text{time}}^{(i)}$, and returns the stacked $JT \times JT$ covariance for $\mathbf{f}^{(i)}$. Here we use the common RBF kernel $\mathbf{K}_{\text{time}}^{(i)}(t, t') = \exp\big(-(t-t')^2/\ell_i^2\big)$ to account for dynamic changes in the latent attributes, whose bandwidth is determined by the unit-specific length scale ℓ_i, but any other kernel can substitute RBF as practitioners see fit.

3.2 Variational learning

Due to the non-Gaussian essence of ordinal likelihood, we adopt the stochastic variational inference technique (SVI) with inducing points introduced in [Hensman et al., 2015]. Dropping superscript for demonstration, SVI utilizes a variational distribution $q(\mathbf{u}) = \mathcal{N}(\mu_{\mathbf{u}}, \mathbf{\Sigma_u})$ on $m \ll n$ inducing variables $\mathbf{u}$ to approximate $p(\mathbf{f} \mid \mathbf{y})$ using the conditional $p(\mathbf{f} \mid \mathbf{u})$. Hence, the conditional log likelihood $\log p(\mathbf{y} \mid \mathbf{u})$ can be lower bounded by the expected log likelihood w.r.t. $p(\mathbf{f} \mid \mathbf{u})$, after exploiting the non-negativity of Kullback–Leibler (KL) divergence between $p(\mathbf{f} \mid \mathbf{u})$ and $p(\mathbf{f} \mid \mathbf{y})$:

$$\log p(\mathbf{y} \mid \mathbf{u}) \geq \mathbb{E}_{p(\mathbf{f}|\mathbf{u})} \log p(\mathbf{y} \mid \mathbf{f}) \tag{4}$$

Furthermore, a lower bound on model evidence (ELBO) can be obtained by combining Eq. (4) and an inequality derived by another KL divergence $\text{KL}[q(\mathbf{u}) \parallel p(\mathbf{u} \mid \mathbf{y})] \geq 0$ (see Appendix A for details):

$$\log p(\mathbf{y}) \geq \mathbb{E}_{q(\mathbf{u})}\big[\log p(\mathbf{y} \mid \mathbf{u})\big] - \text{KL}[q(\mathbf{u}) \parallel p(\mathbf{u})] \tag{5}$$

$$\geq \mathbb{E}_{q(\mathbf{f})}\big[\log p(\mathbf{y} \mid \mathbf{f})\big] - \text{KL}[q(\mathbf{u}) \parallel p(\mathbf{u})] \tag{6}$$

where the KL divergence $\text{KL}[q(\mathbf{u}) \parallel p(\mathbf{u})]$ between the variational $q(\mathbf{u})$ and prior $p(\mathbf{u})$ can be computed in closed form as both distributions are Gaussians. The expectation of log likelihood $\log p(\mathbf{y} \mid \mathbf{f})$ under the marginal distribution $q(\mathbf{f}) = \int p(\mathbf{f} \mid \mathbf{u})q(\mathbf{u})d\mathbf{u}$ is intractable but can be numerically approximated using Gauss-Hermite quadrature method. The variational parameters $\mu_{\mathbf{u}}$ and $\mathbf{\Sigma_u}$, individualized loadings $\mathbf{w}_i$ and $\text{diag}(\mathbf{v})$ as well as likelihood parameters $\{b_c\}$s are then optimized to maximize this lower bound. Finally, the predictive likelihood of new $p(\mathbf{y}^*) = \int p(\mathbf{y}^* \mid \mathbf{f}^*)p(\mathbf{f}^* \mid \mathbf{u})q^*(\mathbf{u})d\mathbf{u}$ is obtained by marginalizing out the optimized $q^*(\mathbf{u})$.

3.3 Theory testing

Our IPGP framework also naturally facilitates downstream tasks such as domain theory testing between models with and without shared or idiographic components. We adopt Bayes factor, the posterior $p(\mathcal{M}_i \mid \mathbf{y}) = \frac{p(\mathbf{y}|\mathcal{M}_i)p(\mathcal{M}_i)}{\sum_i p(\mathbf{y}|\mathcal{M}_i)p(\mathcal{M}_i)}$ over a pool of models $\{\mathcal{M}_i\}$ conditioning on observation $\mathbf{y}$ with prior weights $p(\mathcal{M}_i)$, as the hypothesis test on whether the latent structures for each individual are indeed distinct or are simply explainable by interpersonal commonality. Specifically, we refer the multi-task model in Eq. (3) as the *idiographic* model, and compare it with an *nomothetic* model without unit-specific components: $\mathbf{K}_{\text{task}}^{\text{pop}} = \mathbf{W}_{\text{pop}}\mathbf{W}_{\text{pop}}^{T} + \text{diag}(\mathbf{v})$.

Note that compared to this baseline nomothetic model, our proposed idiographic model in Eq. (3) introduces extra unit-level Jn loading parameters that enlarges the optimization space of hyperparameter. Hence, we propose to first learn the interpersonal loading matrix $\mathbf{W}_{\text{pop}}$ using the standard cross-sectional data from a nomothetic model that focuses on learning of population taxonomy, and then use the estimated $\mathbf{W}_{\text{pop}}$ as informative prior in the full model. We will show empirically in Sec. (4) that with this stronger prior IPGP achieves more precise estimation of individual taxonomies.

4 Experiments

We now evaluate IPGP in learning idiographic latent taxonomies and predicting actual responses against baseline methods from both psychometrics and Gaussian process literature. We then conduct an exploratory factor analysis study of life outcomes of personality replication, validating the popular Big Five personality theory standard using cross-sectional data [McCrae and John, 1992]. The inferred shared taxonomy structure is further incorporated as the informative prior for the population kernel in our case study, where we collected IRB-approved longitudinal survey data. We also highlight the predictive ability of IPGP through a forecasting and a leave-one-trait-out cross validation task, and illustrate how IPGP identifies unique taxonomies of personality that might advance individualized approaches to psychological diagnosis and inspire new theory.

4.1 Simulation and ablation

Setup. Our simulation considers longitudinal data of $n = 10$ units over $T = 30$ periods. We assume latent traits of each unit i has dimension $K = 5$, and each dimension latent vector is generated independently from a GP $\mathbf{x}_i^{(k)}(t) \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_{\text{time}}^{(i)})$ with unit-specific length scale uniformly randomly picked from $\ell_{\text{time}}^{(i)} \in [10, 20, 30]$. We split $m = 20$ batteries into K subsets of size $m/K = 4$, such that each subset dominates one dimensional in the latent traits. Specifically, we set high value of 3 in the population factor loading matrix $\mathbf{W}_{\text{pop}}$ for entries corresponding to the kth subset for dimension k, and low values drawn from Unif$[-1, 1]$ otherwise. We also set each unit-specific loading $\mathbf{w}_i$ from Unif$[-1, 1]$. To allow sparsity and reverse coding, we randomly drop half of loadings and flip signs of the remaining half. Finally, we generate the y_{ijt}s according to the ordered logit model with $C = 5$ levels, and apply $80\%/20\%$ splitting for training and testing.

Table 1: Comparison of averaged accuracy, log lik and correlation matrix distance between IPGP and baselines and ablated models in the simulated study. The full IPGP model (indicated in bold) significantly outperforms all ablated and baseline methods in both estimated correlation matrix and either in-sample or out-of-sample prediction in paired-t tests. Results from ablations imply that IPGP succeeds in predicting the correct labels due to its idiographic components and proper likelihood, and a well-informed population kernel is crucial in recovering the factor loadings. "—" indicates baseline software that cannot handle missing values.

	TRAIN ACC ↑	TRAIN LL ↑	TEST ACC ↑	TEST LL ↑	CMD ↓
GRM	0.261 ± 0.005	-3.556 ± 0.092	0.261 ± 0.006	-3.578 ± 0.098	0.657 ± 0.021
GPCM	0.562 ± 0.017	-2.067 ± 0.182	0.495 ± 0.012	-2.409 ± 0.143	0.545 ± 0.016
SRM	0.286 ± 0.006	-7.408 ± 0.063	0.289 ± 0.008	-7.341 ± 0.084	0.300 ± 0.024
GPDM	0.687 ± 0.010	-4.358 ± 0.028	0.667 ± 0.010	-4.377 ± 0.029	0.262 ± 0.016
DSEM	0.539 ± 0.021	-0.961 ± 0.015	—	—	0.256 ± 0.011
TVAR	0.554 ± 0.018	-1.168 ± 0.014	—	—	0.987 ± 0.013
IPGP-NOM	0.807 ± 0.007	-0.535 ± 0.015	0.790 ± 0.008	-0.555 ± 0.017	0.257 ± 0.009
IPGP-IND	0.932 ± 0.003	-0.243 ± 0.008	0.916 ± 0.004	-0.267 ± 0.009	0.530 ± 0.005
IPGP-LOW	0.897 ± 0.004	-0.313 ± 0.010	0.884 ± 0.005	-0.334 ± 0.011	0.397 ± 0.007
IPGP-NP	0.898 ± 0.003	-0.318 ± 0.009	0.883 ± 0.005	-0.342 ± 0.011	0.467 ± 0.010
IPGP	$\mathbf{0.957 \pm 0.002}$	$\mathbf{-0.159 \pm 0.005}$	$\mathbf{0.942 \pm 0.002}$	$\mathbf{-0.184 \pm 0.006}$	$\mathbf{0.128 \pm 0.006}$

Metrics and baselines. We consider two sets of metrics for evaluation: (1) the in-sample and out-of-sample predictive accuracy (ACC) and log likelihood (LL) of the actual responses, (2) the correlation matrix distance (CMD) between the estimated factor loading matrix and the true ones, which is defined for two covariance matrices $\mathbf{R}_1, \mathbf{R}_2$ as $\mathrm{d}(\mathbf{R}_1, \mathbf{R}_2) = 1 - \frac{\mathrm{tr}(\mathbf{R}_1\mathbf{R}_1)}{\|\mathbf{R}_1\|_f \|\mathbf{R}_1\|_f}$ [Herdin et al., 2005] with l_2 Frobenius norm. Note that CMD becomes zero if $\mathbf{R}_1, \mathbf{R}_2$ are equal up to a scaling factor, and one if they are orthogonal after flattening. We compare IPGP to (1) various latent variable models for ordinal responses, including the graded response model (GRM) [Samejima, 1969], the generalized partial credit model (GPCM) [Muraki, 1992] and the sequential response model (SRM) [Tutz, 1990], (2) Gaussian process dynamic model (GPDM) [Damianou et al., 2011, Dürichen et al., 2014] where the continuous predictions are rounded to the nearest ordinal level, (3) dynamic structural equation model (DSEM) [Asparouhov et al., 2018, McNeish et al., 2023] with trait-dependent latent variables and (4) time-varying vector autoregression (TVAR) with regularized kernel smoothing [Haslbeck et al., 2021]. We also compare IPGP with several ablated models: (1) IPGP-NOM without the idiographic kernel, (2) IPGP-IND without the population kernel, (3) IPGP-LOW with lower-rank factors of 2 than actual rank of 5 in the synthetic setup and (4) IPGP-NP where the population kernel is learned from scratch rather than fixed to the informative prior. Note that $\mathbf{W}_{\text{pop}}$ in the full IPGP model is fixed as learned from IPGP-NOM.

Results. We use 100 inducing points and ADAM optimizer of learning rate 0.05 to optimize ELBO for 10 epoches with batch size of 256. We repeat our simulation with 25 different random seeds using 300 Intel Xeon 268 CPUs. Table 1 shows comparison of averaged predictive accuracy, log likelihood and correlation matrix distance between IPGP and baselines and ablated models in the simulated study. Our IPGP model (indicated in bold) significantly outperforms all ablated models and baseline methods in estimated correlation matrix, predictive accuracy and log likelihood of both training and testing sets in paired-t tests. We found that IPGP succeeds in predicting the correct labels due to its idiographic components and proper likelihood, since IPGP-NOM and IPGP-GL are two of the worst ablations for all prediction metrics. In addition, IPGP-IND and IPGP-NP have the worst correlation matrix estimation, implying that a well-informed population kernel is crucial in recovering the underlying factor structures.

4.2 Exploratory factor analysis

We then validate the popular Big Five personality theory using standard cross-sectional data via an exploratory factor analysis, where a range of factors are tested and then determined according to model evidence rather than being fixed. We utilize the life outcomes of personality replication (LOOPR) data (see [Soto, 2019] for a full description of LOOPR), which is collected from $5,347$ unique participants on the Big Five Inventory [John et al., 1999] consisting of 60 battery questions. Our validation considers a range of latent trait dimension counts from $K = 1, \dots, 5$. For each dimension count, we first apply principal component analysis (PCA) directly on the correlation matrix of the cross-sectional observations to learn a vanilla population factor loading matrix. We then initialize $\mathbf{W}_{\text{pop}}$ in our model with this vanilla loading matrix, and optimize the loading matrix jointly with the variational parameters. Note that $T = 1$ in LOOPR, so we drop the idiographic components.

Table 2: In-sample accuracy and averaged log lik of our method and baselines for various K in LOOPR. Best model for each K is indicated in bold and the best model across different Ks is further indicated in italic.

	ACC ↑					LL / N ↑				
MODEL	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$
PCA	0.106	0.099	0.123	0.217	0.192	−1.957	−1.990	−2.009	−2.036	−2.051
GRM	0.238	0.107	0.178	0.113	0.146	−1.838	−1.832	−1.814	−1.838	−1.841
GPCM	0.213	0.156	0.186	0.159	0.163	−1.754	−1.761	−1.764	−1.750	−1.756
SRM	0.243	0.134	0.179	0.125	0.155	−1.784	−1.784	−1.783	−1.780	−1.767
GPDM	0.268	0.272	0.266	0.268	0.263	−2.155	−2.158	−2.158	−2.159	−2.158
DSEM	0.188	0.114	0.110	0.105	0.104	−1.997	−1.960	−1.908	−1.845	−1.775
IPGP	**0.322**	**0.319**	***0.323***	**0.318**	**0.318**	**−1.478**	**−1.477**	**−1.477**	**−1.477**	***−1.476***

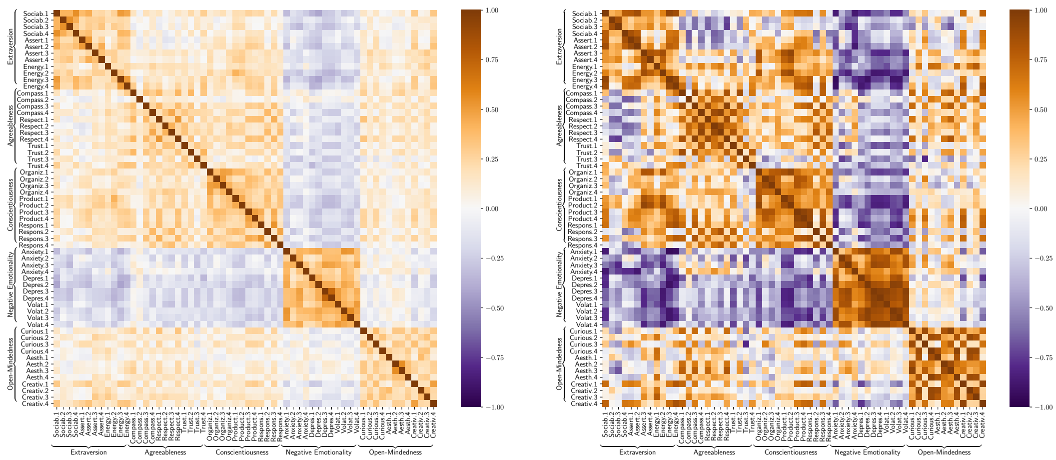

Figure 2: Illustration of raw correlation matrix (left) and our estimated Big Five loading matrix (right). Both correlation matrices displace a *block* pattern, where estimated interpersonal variation show strong correlation between questions within the same factor of the Big Five personalities and weak correlation across different factors. Besides, questions corresponding negative emotionality show minor negative correlation with those corresponding to extraversion and conscientiousness, suggesting trait-by-trait interaction effects.

Validation of Big Five. Table 2 shows the predictive accuracy and averaged log likelihood of our method and baseline methods (excluding TVAR for lacking low-rank assumption) for various K in LOOPR. Best model for each K is indicated in bold numbers and the best model across different Ks is further indicated in italic numbers. Despite having slightly worse in-sample predictive accuracy than factor 3 model, IPGP with factor 5 has significant higher model evidence than all the other models, with the second best model is $\exp(-79)$ more unlikely indicated by Bayes factors. Therefore, our results indicate that when psychological measurements are estimated from standard cross-sectional data, Big Five personality is necessary for learning interpersonal variation.

Estimated interpersonal variation. We also show the raw correlation and our estimated Big Five correlation in Figure 2. Both correlation matrices display a *block* pattern, where estimated interpersonal variation show strong correlation between questions within the same factor of the Big Five and weak correlation across different factors. Besides, questions corresponding negative emotionality show minor negative correlation with those corresponding to extraversion and conscientiousness, suggesting appropriate trait-by-trait interaction effects.

4.3 Case study

To further demonstrate IPGP in longitudinal setting for learning idiographic psychological taxonomies, we collected an intensive longitudinal data using experience sampling measures (ESM). We also highlight the predictive ability of IPGP through a forecasting and a leave-one-trait-out cross validation task, and illustrate how IPGP identifies unique taxonomies of personality that might advance individualized approaches to psychological diagnosis and inspire new theory.

Data collection. In ESM design, each participant was asked to complete personality assessments six times per day for three weeks, resulting maximum 126 assessments per person. With 93 valid student participants, we acquired 8,770 assessments in total with an average of 94 assessments per person. The personality assessment is derived from the BFI-2 [Soto and John, 2017] to ensure identification of latent factors and ample coverage of the late trait space. The BFI-2 includes 60 items with four unique items assessing each of the three different sub-factors for each Big-Five domains. We removed one item for each sub-factors that are not appropriate for contextualized assessments of ESM design. To mitigate the fatigue and learning effect from repeated measures, we employed a planned missing design where participants were randomly tested on only two out of three items assessing the same sub-factors, resulting only 30 items for each assessment.

Comparison between nomothetic and idiographic models. We run the full IPGP model with idiographic component and unit-specific time kernel on the collected longitudinal data. Again we set the ranks of the population and individual loading matrices to 5 and 1 respectively, and incorporate the prior knowledge of the cross-sectional data by fixing the population loadings as the Big Five loadings estimated in Sec. (4.2) and optimizing the individual loadings. We contrast our proposed idiographic model (IPGP) and baselines in Table 3, which shows the in-sample prediction, averaged log likelihood and Bayes factors. We found that IPGP outperforms IPGP-NOM with higher predictive accuracy and log likelihood, and is favored decisively by a Bayes factor of $\exp(1.06 \times 10^4)$.

Table 3: In-sample prediction and averaged log likelihood of our proposed model (IPGP) and baselines for the longitudinal data, as well as Bayes factors to IPGP. "—" indicates self comparison of Bayes factors.

	ACC	LL/N	log(**BF**)
GRM	0.210	-2.266	-2.32×10^5
GPCM	0.288	-1.516	-3.80×10^4
SRM	0.260	-1.927	-1.44×10^5
GPDM	0.382	-3.865	-7.80×10^5
DSEM	0.226	-1.399	-7.72×10^3
TVAR	0.382	-1.546	-4.47×10^4
IPGP-NOM	0.403	-1.410	-1.06×10^4
IPGP	**0.417**	$\mathbf{-1.369}$	—

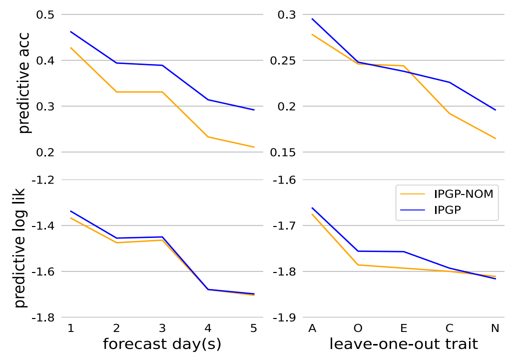

Figure 3: Predictive accuracy and log lik of IPGP and IPGP-NOM for the forecasting task and leave-one-trait-out cross-validation task.

Predictive performance of IPGP. We also evaluate the out-of-sample performance of the idiographic and nomothetic models using two prediction tasks: forecasting future responses and leave-one-trait-out cross validation. For the forecasting task, we train both models with data from the first 40 days and predict future responses for the last 5 days. For the cross validation task, we predict responses of each trait by training on data belonging to the other four traits. Figure (3) shows the predictive accuracy and log likelihood of IPGP and IPGP-NOM for the forecasting task over varying horizons and the leave-one-trait-out cross-validation task. IPGP has consistently better performance than IPGP-NOM in both tasks except for being slightly less accurate in predicting extraversion. Overall, IPGP is favored than IPGP-NOM by Bayes factors of $\exp(43)$ and $\exp(716)$ in these tasks.

Figure 4: Four residual correlations as identified by our k-mean clustering. Each heatmap displays the trait-level residual correlation averaged across corresponding batteries for one cluster, with darker red and blue indicating larger positive and negative deviations. For instance, agreeableness (A) is more correlated to extraversion (E) than the population profile in the first profile, but less correlated to openness (O) in the second profile. Moreover, these two directions of deviations are even exacerbated in the third and fourth profiles.

Discovery of unique taxonomies. Despite our small cohort size (93 respondents), we also manage to identify distinct profiles of personality that substantially differ from the interpersonal commonality. Specifically, we first perform a k-mean clustering using all 93 estimated individual correlation matrix with CMD as the distance metric, and then compute the residual correlation between each estimated clustering centroid and the population correlation. Figure (4) illustrates four residual correlations as identified by our k-mean clustering. Each heatmap displays the trait-level residual correlation averaged across corresponding batteries for one cluster, with darker red and blue indicating larger positive and negative deviations. For instance, agreeableness (A) is more correlated to extraversion (E) than population profile in the first profile, but less correlated to openness (O) in the second profile. Moreover, these two directions of deviations are even exacerbated in the third and fourth profiles.

5 Related work

Idiographic assessment emphasizes the important aspects of individuals otherwise missing in oversimplified taxonomies of psychological behaviors [Hamaker and Dolan, 2009]. Empirical evidence across many psychometrics fields has shown the lack of generalizability of the nomothetic models only focusing on interpersonal variation [Molenaar, 2004]. Hence, Song and Ferrer [2012] incorporated simple random effect method with dynamic factor models for analyzing psychological processes. Jongerling et al. [2015] proposed a multilevel first-order autoregressive model with random intercepts to measure daily positive effects over several weeks. Beltz et al. [2016] combined the nomothetic and idiographic approaches in analyzing clinical data by adding individual components to the group iterative multiple model (GIMME). However, all of these methods focus on modeling in response space rather than latent space for ordinal survey data.

Gaussian process latent variable model (GPLVM) is a dimensional reduction method for Gaussian data, where the latent variables are optimized after integrating out the function mappings [Lawrence, 2003, Lalchand et al., 2022]. Our proposed framework differs from GPLVM as we optimize the factor loading matrix while marginalizing the latent variables. In addition, our model contrasts GPLVM and (variational) Gaussian Process dynamical model (GPDM) [Wang et al., 2005, Damianou et al., 2011] in our non-Gaussian ordered logistic observation model. Finally, our longitudinal framework with stochastic variational inference learning differs from the static GP item response theory (GPIRT) [Duck-Mayr et al., 2020] with more computationally demanding Gibbs sampling.

Longitudinal measurement models integrate temporal dynamics into psychological theories with growing popularity of longitudinal design in survey methods [Jebb et al., 2015, Ariens et al., 2020]. For instance, families of longitudinal structural equation models (SEM) such as multiple-group longitudinal SEM and longitudinal growth curve model were developed for repeated measurement studies [Little, 2013], where M*plus* software was developed later for dynamic SEM with Bayesian Gibbs sampling [Asparouhov et al., 2018, McNeish et al., 2023]. Dynamic item response models [Rijmen et al., 2003, Reise and Waller, 2009, Dumas et al., 2020] and time-varying vector autoregressive model [Lu et al., 2018, Haslbeck et al., 2021] were also proposed to estimate the trajectories of latent traits. Despite previous work in behavioral literature focusing on Gaussian observations [Dürichen et al., 2014], multi-task Gaussian process time series has not yet been exploited for survey experiments with non-Gaussian likelihood when exact inference is not plausible.

6 Conclusion

We propose a novel idiographic personality Gaussian process (IPGP) model for personalized psychological assessment and learning of intrapersonal taxonomy from longitudinal ordinal survey data, an under-explored setup in Gaussian process dynamic system literature. We exploit Gaussian process coregionalization for capturing between-battery structure and stochastic variational inference for scalable inference. Future directions include adaptation of IPGP to other psychological studies such as emotion, and incorporation of contextual information such as behaviors or demographics.

Our proposed IPGP framework also provides insights to domain theory testing, in this work, addressing the substantive debate in psychometrics surrounding the shared versus unique structures of psychological features. Besides predicting the actual responses, we also include learning of the true underlying taxonomy structures as evaluation metrics to minimize risks of inductive bias. Our experimental results show that IPGP is decisively favored than the nomothetic baseline, and substantive deviations from the common trend persist in considerable individuals. Hence, our framework has a great potential in advancing individualized approaches to psychological diagnosis and treatment.

Acknowledgments and Disclosure of Funding

This work was supported by the 2023 Seed Grant of Transdisciplinary Institute in Applied Data Sciences at Washington University.

References

Peter CM Molenaar. A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement*, 2(4):201–218, 2004.

Xiaojing Wang, James O Berger, and Donald S Burdick. Bayesian analysis of dynamic item response models in educational testing. *The Annals of Applied Statistics*, 7(1):126–153, 2013.

Denis Dumas, Daniel McNeish, and Jeffrey A Greene. Dynamic measurement: A theoretical–psychometric paradigm for modern educational psychology. *Educational Psychologist*, 55(2): 88–105, 2020.

Denny Borsboom, Gideon J Mellenbergh, and Jaap Van Heerden. The theoretical status of latent variables. *Psychological Review*, 110(2):203, 2003.

Peter Borkenau and Fritz Ostendorf. The Big Five as States: How Useful Is the Five-Factor Model to Describe Intraindividual Variations over Time? *Journal of Research in Personality*, 32(2):202–221, 1998.

Emorie D Beck and Joshua J Jackson. Consistency and Change in Idiographic Personality: A Longitudinal ESM Network Study. *Journal of Personality and Social Psychology*, 118(5):1080, 2020.

Emorie D Beck and Joshua J Jackson. Within-person variability. In *The Handbook of Personality Dynamics and Processes*, pages 75–100. Elsevier, 2021.

Frank Rijmen, Francis Tuerlinckx, Paul De Boeck, and Peter Kuppens. A nonlinear mixed model framework for item response theory. *Psychological Methods*, 8(2):185, 2003.

Steven P Reise and Niels G Waller. Item Response Theory and Clinical Measurement. *Annual Review of Clinical Psychology*, 5:27–48, 2009.

Todd D Little. *Longitudinal Structural Equation Modeling*. Guilford Press, 2013.

Eun Sook Kim and Victor L Willson. Testing Measurement Invariance Across Groups in Longitudinal Data: Multigroup Second-Order Latent Growth Model. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4):566–576, 2014.

Tihomir Asparouhov, Ellen L Hamaker, and Bengt Muthén. Dynamic Structural Equation Models. *Structural Equation Modeling: a Multidisciplinary Journal*, 25(3):359–388, 2018.

Feihan Lu, Yao Zheng, Harrington Cleveland, Chris Burton, and David Madigan. Bayesian hierarchical vector autoregressive models for patient-level predictive modeling. *PloS One*, 13(12): e0208082, 2018.

Jonas MB Haslbeck, Laura F Bringmann, and Lourens J Waldorp. A Tutorial on Estimating Time-Varying Vector Autoregressive Models. *Multivariate Behavioral Research*, 56(1):120–149, 2021.

Jack Wang, Aaron Hertzmann, and David J Fleet. Gaussian Process Dynamical Models. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems*, volume 18. MIT Press, 2005.

Andreas Damianou, Michalis Titsias, and Neil Lawrence. Variational Gaussian Process Dynamical Systems. *Advances in Neural Information Processing systems*, 24, 2011.

Robert Dürichen, Marco AF Pimentel, Lei Clifton, Achim Schweikard, and David A Clifton. Multitask Gaussian Processes for Multivariate Physiological Time-Series Analysis. *IEEE Transactions on Biomedical Engineering*, 62(1):314–322, 2014.

JBrandon Duck-Mayr, Roman Garnett, and Jacob Montgomery. GPIRT: A Gaussian Process Model for Item Response Theory. In *Conference on Uncertainty in Artificial Intelligence*, pages 520–529. PMLR, 2020.

Adriene M. Beltz, Aidan G. C. Wright, Briana N. Sprague, and Peter C. M. Molenaar. Bridging the Nomothetic and Idiographic Approaches to the Analysis of Clinical Data. *Assessment*, 23(4): 447–458, 2016.

Edwin V Bonilla, Kian Chai, and Christopher Williams. Multi-task Gaussian Process Prediction. *Advances in Neural Information Processing Systems*, 20, 2007.

Edward Snelson and Zoubin Ghahramani. Sparse Gaussian Processes using Pseudo-inputs. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems*, volume 18. MIT Press, 2005.

James Hensman, Alexander Matthews, and Zoubin Ghahramani. Scalable Variational Gaussian Process Classification. In *Artificial Intelligence and Statistics*, pages 351–360. PMLR, 2015.

Neil Lawrence. Gaussian Process Latent Variable Models for Visualisation of High Dimensional Data. *Advances in Neural Information Processing Systems*, 16, 2003.

James T Croasmun and Lee Ostrom. Using Likert-Type Scales in the Social Sciences. *Journal of Adult Education*, 40(1):19–22, 2011.

Wei Chu and Zoubin Ghahramani. Gaussian Processes for Ordinal Regression. *Journal of Machine Learning Research*, 6(35):1019–1041, 2005.

John M Digman. Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, 73(6):1246, 1997.

James Baglin. Improving Your Exploratory Factor Analysis for Ordinal Data: A Demonstration Using FACTOR. *Practical Assessment, Research, and Evaluation*, 19(1):5, 2014.

Kenneth A Bollen. *Structural Equations with Latent Variables*, volume 210. John Wiley & Sons, 1989.

Fumiko Samejima. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 1969.

Wim J Van der Linden and RK Hambleton. Handbook of Item Response Theory. *Taylor & Francis Group. Citado na pág*, 1(7):8, 1997.

R Darrell Bock and Murray Aitkin. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4):443–459, 1981.

Carlos G Forero, Alberto Maydeu-Olivares, and David Gallardo-Pujol. Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4):625–641, 2009. doi: 10.1080/10705510903203573.

Cheng-Hsien Li. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48:936–949, 2016.

Sergio Salvatore and Jaan Valsiner. Between the General and the Unique: Overcoming the Nomothetic versus Idiographic Opposition. *Theory & Psychology*, 20(6):817–833, 2010.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 11 2005.

Mauricio A Alvarez, Lorenzo Rosasco, Neil D Lawrence, et al. Kernels for Vector-Valued Functions: A Review. *Foundations and Trends® in Machine Learning*, 4(3):195–266, 2012.

Robert R McCrae and Oliver P John. An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, 60(2):175–215, 1992.

Markus Herdin, Nicolai Czink, Hüseyin Ozcelik, and Ernst Bonek. Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels. In *2005 IEEE 61st Vehicular Technology Conference*, volume 1, pages 136–140. IEEE, 2005.

Eiji Muraki. A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, 16(2):159–176, 1992.

Gerhard Tutz. Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, 43(1):39–55, 1990.

Daniel McNeish, Jennifer A Somers, and Andrea Savord. Dynamic structural equation models with binary and ordinal outcomes in M*plus*. *Behavior Research Methods*, pages 1–27, 2023.

Christopher J Soto. How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project. *Psychological Science*, 30(5): 711–727, 2019.

Oliver P John, Sanjay Srivastava, et al. The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. *Handbook of personality: Theory and research*, 2:102—–138, 1999.

Christopher J Soto and Oliver P John. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1):117, 2017.

Ellen L Hamaker and Conor V Dolan. Idiographic Data Analysis: Quantitative Methods—From Simple to Advanced. In *Dynamic Process Methodology in the Social and Developmental Sciences*, pages 191–216. Springer, 2009.

Hairong Song and Emilio Ferrer. Bayesian estimation of random coefficient dynamic factor models. *Multivariate Behavioral Research*, 47(1):26–60, 2012.

Joran Jongerling, Jean-Philippe Laurenceau, and Ellen L Hamaker. A Multilevel AR(1) Model: Allowing for Inter-Individual Differences in Trait-Scores, Inertia, and Innovation Variance. *Multivariate Behavioral Research*, 50(3):334–349, 2015.

Vidhi Lalchand, Aditya Ravuri, and Neil D. Lawrence. Generalised GPLVM with Stochastic Variational Inference. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 7841–7864. PMLR, 28–30 Mar 2022.

Andrew T Jebb, Louis Tay, Wei Wang, and Qiming Huang. Time series analysis for psychological research: examining and forecasting change. *Frontiers in Psychology*, 6:727, 2015.

Sigert Ariens, Eva Ceulemans, and Janne K Adolf. Time series analysis of intensive longitudinal data in psychosomatic research: A methodological overview. *Journal of Psychosomatic Research*, 137:110191, 2020.

A Mathematical Details of Evidence Lower Bound

We provide the full mathematical details of the evidence lower bound defined in Eq. (6). As KL divergence is always non-negative, we first consider the KL divergence between $p(\mathbf{f} \mid \mathbf{u})$ and $p(\mathbf{f} \mid \mathbf{y})$:

$$
\begin{aligned}
\mathrm{KL}[p(\mathbf{f} \mid \mathbf{u}) \parallel p(\mathbf{f} \mid \mathbf{y})] &= \mathbb{E}_{p(\mathbf{f}|\mathbf{u})} \log \frac{p(\mathbf{f} \mid \mathbf{u})}{p(\mathbf{f} \mid \mathbf{y})} && (7) \\
&= \mathbb{E}_{p(\mathbf{f}|\mathbf{u})} \log \frac{p(\mathbf{f} \mid \mathbf{u}) p(\mathbf{y})}{p(\mathbf{y} \mid \mathbf{f}) p(\mathbf{f})} && (8) \\
&= \mathbb{E}_{p(\mathbf{f}|\mathbf{u})} \log \frac{p(\mathbf{f} \mid \mathbf{u}) p(\mathbf{y} \mid \mathbf{u}) p(\mathbf{u})}{p(\mathbf{y} \mid \mathbf{f}) p(\mathbf{f})} && (9) \\
&= \mathbb{E}_{p(\mathbf{f}|\mathbf{u})} \log \frac{p(\mathbf{y} \mid \mathbf{u})}{p(\mathbf{y} \mid \mathbf{f})} && (10) \\
&= \log p(\mathbf{y} \mid \mathbf{u}) - \mathbb{E}_{p(\mathbf{f}|\mathbf{u})} \log p(\mathbf{y} \mid \mathbf{f}) \geq 0 && (11)
\end{aligned}
$$

Moving $\mathbb{E}_{p(\mathbf{f}|\mathbf{u})} \log p(\mathbf{y} \mid \mathbf{f})$ to the R.H.S of the above inequality will lead to Eq. (4). We then exploit the inequality given by $\mathrm{KL}[q(\mathbf{u}) \parallel p(\mathbf{u} \mid \mathbf{y})] \geq 0$:

$$
\begin{aligned}
\mathrm{KL}[q(\mathbf{u}) \parallel p(\mathbf{u} \mid \mathbf{y})] &= \mathbb{E}_{q(\mathbf{u})} \log \frac{q(\mathbf{u})}{p(\mathbf{u} \mid \mathbf{y})} && (12) \\
&= \mathbb{E}_{q(\mathbf{u})} \log \frac{q(\mathbf{u}) p(\mathbf{y})}{p(\mathbf{y} \mid \mathbf{u}) p(\mathbf{u})} && (13) \\
&= -\mathbb{E}_{q(\mathbf{u})} \log p(\mathbf{y} \mid \mathbf{u}) + \mathrm{KL}[q(\mathbf{u}) \parallel p(\mathbf{u})] + \log p(\mathbf{y}) \geq 0 && (14)
\end{aligned}
$$

Rearranging the above inequality, applying Eq. (4) and exploiting notation $q(\mathbf{f}) = \int p(\mathbf{f} \mid \mathbf{u}) q(\mathbf{u}) d\mathbf{u}$ leads to the ELBO:

$$
\begin{aligned}
\log p(\mathbf{y}) &\geq \mathbb{E}_{q(\mathbf{u})} \log p(\mathbf{y} \mid \mathbf{u}) - \mathrm{KL}[q(\mathbf{u}) \parallel p(\mathbf{u})] && (15) \\
&= \mathbb{E}_{q(\mathbf{u})} \left[\mathbb{E}_{p(\mathbf{f}|\mathbf{u})} \log p(\mathbf{y} \mid \mathbf{f}) \right] - \mathrm{KL}[q(\mathbf{u}) \parallel p(\mathbf{u})] && (16) \\
&= \mathbb{E}_{q(\mathbf{f})} \log p(\mathbf{y} \mid \mathbf{f}) - \mathrm{KL}[q(\mathbf{u}) \parallel p(\mathbf{u})] && (17)
\end{aligned}
$$

B Estimated Correlations of Selective Individuals

Figure (5) shows the estimated correlations of selective individuals for the identified four profiles in the longitudinal study.

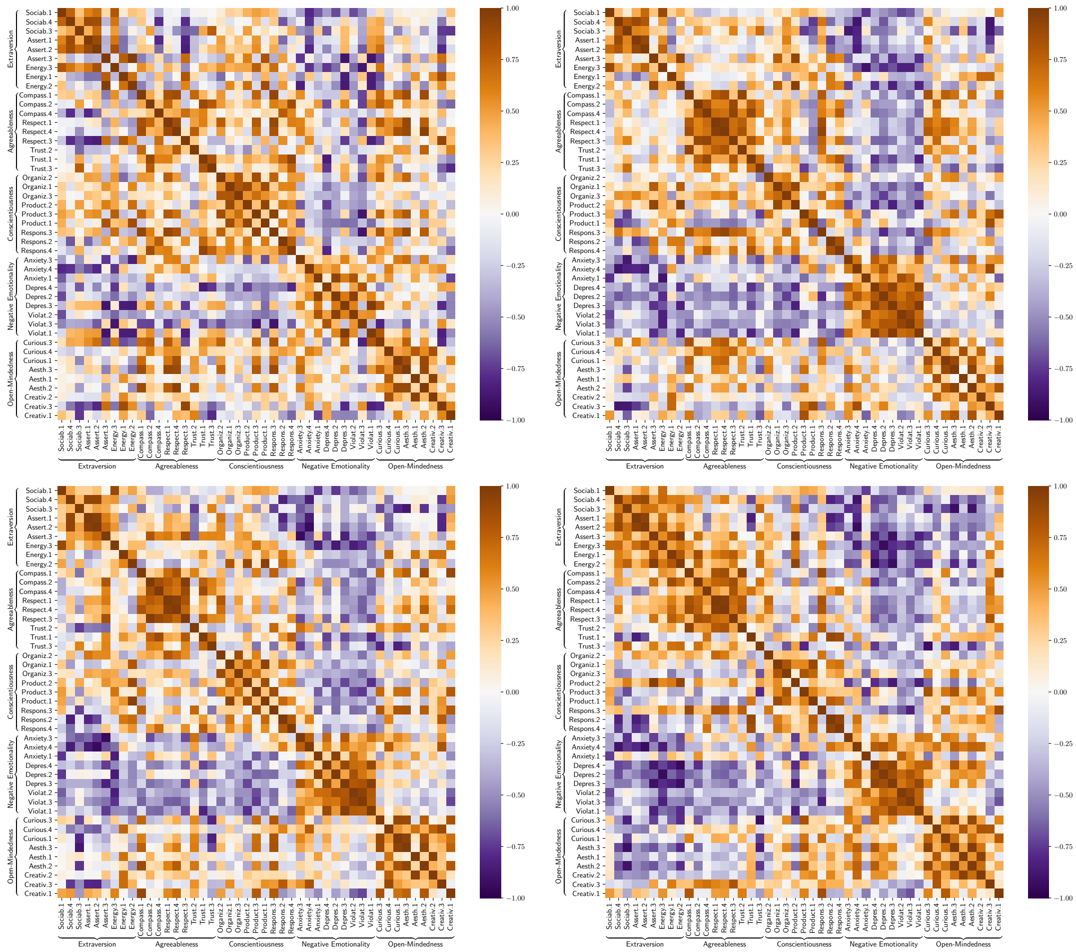

Figure 5: Estimated correlations of selective individuals for the identified four profiles in the longitudinal study.