INSTITUTE ON EDUCATION AND THE ECONOMY
Teachers College, Columbia University
439 Thorndike Hall
New York, NY 10027

Latent Curve Covariance Models for Longitudinal Data

Marc Scott	Institute on Education and the Economy
Marc Handcock	Pennsylvania State University

IEE Working Paper 13
August 1999

 Please direct all correspondence to Marc Scott, Institute on Education and the Economy, Box 174, Teachers College, Columbia University, 525 West 120th Street, New York, NY 10027, ms1098@columbia.edu.

Latent Curve Covariance Models for Longitudinal Data

Marc A. SCOTT and Mark S. HANDCOCK

We develop a class of covariance models for longitudinal data which capture underlying stochastic variation in an interpretable manner via latent curves. The approach decomposes the variation into a small number of data-adaptive components, the latent curves, which we call proto-splines. These components establish the shape of the stochastic variation and allow individual differences to be expressed via random coefficients, which adjust the scale of the components. This model class extends the scope of traditional random coeffcient models to effectively allow for an unknown design, while retaining their emphasis on the estimation of components of variance, which often have substantive meaning. We prove that the parameter estimates used to construct the latent curves and those associated with the random coefficients are consistent, asymptotically efficient, and Gaussian. An application to the analysis of wage inequality based on the National Longitudinal Survey of Young Men is presented.

KEY WORDS: Covariance Structure; Random Coefficient Model; Variance Components; Functional Data Analysis; Principal Components; National Longitudinal Survey.

1. INTRODUCTION

An increasing number of social, behavioral and biological science studies collect information from subjects at several points in time. These longitudinal studies enable researchers to study changes in the phenomena of interest over the life-course of the subjects. Explanatory covariates collected contemporaneously are often linked to the response in regression-like models known as mixed effects models. Longitudinal data require this more sophisticated model because the responses are correlated within individuals. This correlation structure can be quite complicated, and many parametric forms for covariance have been developed over the years. The most familiar of these are error-in-variables models such as compound symmetry, and time series models such as ARIMA models (see Diggle et. al. 1994 for others). One of the limitations of the approaches developed so far is that they require the researcher to have a substantial amount of knowledge (or make some strong assumptions) about the form of the covariance.

Marc A. Scott is Senior Research Associate, Institute on Education and the Economy, Teachers College, Columbia University, New York, NY 10027 (E-mail: ms1098@columbia.edu); and Mark S. Handcock is Associate Professor, Department of Statistics, Pennsylvania State University, University Park, PA 16802 (E-mail: handcock@stat.psu.edu). This research was partially supported by the Russell Sage Foundation and the Rockefeller Foundation. The authors thank Annette D. Bernhardt whose discussions and insights over the years are deeply embedded in this paper.

Journal of the American Statistical Association
Theory and Methods

In this paper we describe a more flexible model for covariance structure that allows important features of the covariance to be data-driven, rather than externally imposed. Strictly speaking, our model class is not a mixed effects model. Our models have strong ties to those developed by Rice and Silverman (1991), the common foundation being the decomposition of covariance into eigenfunctions using the Karhunen-Loève expansion. Under this formulation, the response is a stochastically weighted sum of eigenfunctions. We use a limited number of these eigenfunctions as a first approximation to the covariance, forming a non-traditional random coefficient model. What makes our model different from traditional mixed effects and random coefficient models is that we estimate the eigenfunctions directly as part of our model–this allows for uncertainty in the random effects design, which is non-standard. We add to the work of Rice and Silverman (1991) by establishing the asymptotic efficiency of the estimates (thus confidence bounds can be constructed for the curves themselves). We also provide a way to include orthogonality constraints on the eigenfunctions directly in our model, as an alternative to Silverman's (1996) approach.

As a motivating example, we look at inequality of wage outcomes for young workers in the United States. Cross-sectional analyses of wages show an increased polarization in the 1980s, despite a period of economic boom (Karoly 1993, Levy and Murnane 1992, Danziger and Gottschalk 1993, 1996). One explanation for the observed cross-sectional trends is that wage outcomes have become more volatile. That is, they vary greatly from one year to the next. Longitudinal wage data addresses this hypothesis directly. If individual wage trajectories are relatively smooth, then this indicates stable long-term trends, while bumpy trajectories are evidence of wage volatility. However, in many such studies, there are relatively few observations per individual, making these trends difficult to assess. A variance components model (Searle et al. 1992) for wages may be used to partition the variance into long-term, or permanent variation and short-term, transitory variation. Since long- and short-term trends in wages have substantive meaning, one can compare the resulting partitions from different economic periods.

But how does one go about partitioning longitudinal data into such trends? If one is willing to assume that individuals show quadratic wage growth as they age, and that the distribution of the coefficients forming such random quadratic curves is multivariate Gaussian, then traditional mixed effects models can be used (the random effects design matrix is taken to be a basis for the polynomials of degree two). Bernhardt et al. (1998) apply a mixed effects variance components model to wage data from two young adult cohorts in the National Longitudinal Survey (NLS) and show that the growth in overall inequality can be attributed to long-term trends and not to volatility (see also Gottshalk and Moffitt 1994, Haider 1996 and Baker 1997). While this is a reasonable decomposition, the proper specification of the underlying mechanism is crucial in matters with such strong policy implications. The methods we develop allow for more data-driven decompositions, and as such are a valuable check on more traditional model assumptions.

Continuing with our example, we shall refer to the vector of wages for each individual recorded over a set of time periods in his/her career as the *wage trajectory.* The trajectory is augmented by information on demographics, job acquisition, turnover, completion of educational milestones, enrollment status, and moves in and out of different industries and occupations to form the individual's wage profile. Typically (real) wages tend to grow as the individual ages, but there may be substantial heterogeneity in the timing and magnitude of growth. We will use age as the underlying time index as we are interested in comparative models of worker experience. A collection of age-indexed wage trajectories typically "fan out" over time, with some individuals experiencing high or low growth at both younger and older ages. This can be seen in the sample of wage trajectories from the National Longitudinal Survey of Youth (NLSY) in Figure 1. A description and analysis of this data is given in Section 4..

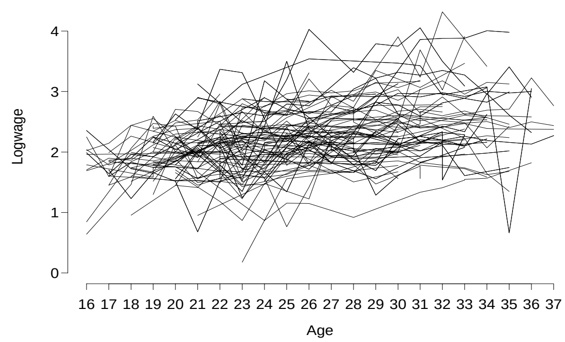

Figure 1. A sample of 1,000 wage trajectories from respondents in the NLSY. Note that in addition to the fanning out, there is a great deal of variability within and between trajectories–this is the heterogeneity we wish to describe.

To illustrate the use of new model class, we will partition the observed wage trajectories into short- and long-term components, yielding an interpretable statistical model.

To date, a number of different approaches have been developed that are appropriate for this task. Random coefficient models, as described in Longford (1993), model heterogeneity via random perturbations to a known structure captured in the design matrix (see also Vonesh and Chinchilli 1997). Their strengths include a well-established set of available inferential techniques, but these rest on the knowing a priori the correct form of the model, which effectively means knowing the form of the covariance. Xu, Hedeker and Ramakrishnan (1996) and Verbeke and Lesaffre (1996) relax the distributional assumptions in these models somewhat, by allowing the random coefficients to come from a Gaussian mixture. Nonlinear extensions of random coefficient models as well as nonparametric estimation of the coefficient densities are described in Davidian and Giltinan (1995). All of these methods either assume a highly restrictive form for the covariance or leave it largely unspecified. The former is often unrealistic, but more to the point, will rarely provide insight into long-term trends in variation unless they happen to take one of those simpler forms. The latter approach is often infeasible, requiring very large samples and producing highly variable estimates.

A different approach to covariance modeling comes from the functional data analysis literature, in which the data are sampled so frequently that they can be taken to be functions. The framework for this approach is given in Ramsey and Silverman (1997), and key models were developed by Rice and Silverman (1991), Kneip (1994), and Besse, Cardot and Ferraty (1997). Lindstrom (1995) combines ideas from both mixed effects and functional approaches, while Barry (1995) provides a Bayesian model for the covariance function. Brumback and Rice (1998) and Wang (1998) show that certain forms of covariance yield smoothing spline trajectory *predictions*, drawing on the work of Wahba (1990) and Speed (1991). Strictly speaking, the nonparametric approaches of Brumback and Rice (1998) and Wang (1998) do not estimate eigenfunctions–they predict individual-specific curves. This work differs in orientation from Rice and Silverman (1991) and our work, which focus directly on the covariance structure rather than individual curves.

The conceptual framework for our models is known as latent curve analysis, beginning with Rao (1958) and formalized in Meredith and Tisak (1990). Much of our development is conceptually rooted in this work and the functional data analysis perspective offered by Rice and Silverman (1991) and Ramsay and Dalzell (1991).

In Section 2. we explore the foundations of latent curve models and formulate the proto-spline class. We discuss the estimation and asymptotic properties of proto-splines in Section 3.. In Section 4. we analyze wage inequality using proto-splines based on information in the NLSY dataset. In Section 5. we consider some extensions of the proto-spline model.

2. MODEL FORMULATION

2.1 Stochastic Process Formulation

Since change is observed over time, we can model longitudinal data as a stochastic process, $Y(t)$, observed at time points $t \in \mathcal{T}$. The set $\mathcal{T}$ may be continuous in an idealized case, but in practice it will be discrete. Ignoring any explanatory covariates other than time, the first and second order properties of $Y(t)$ are its mean and covariance functions $\mu(t)$ and $R(s,t) = E[(Y(s)-\mu(s))(Y(t)-\mu(t))]$. If we believe that the process is Gaussian, then the entire process is determined by these two functions.

Whether the underlying process is Gaussian or not, estimating its covariance function is a challenge. We can assume that the function is continuous, but beyond this we only know that it is positive definite. If $\mathcal{T}$ is discrete and finite, with size $|\mathcal{T}| = T$, then $R(s,t)$ is defined by $T(T+1)/2$ unique parameters. A natural approach is to assume that the covariance has a form that can be described using a smaller number of parameters. It is unlikely, however, that simple covariance model classes, such as compound symmetry and others described in (Diggle et. al. 1994, Chapter 5), can capture the very general structures that might exist. What is required is a framework within which we can closely *approximate* the covariance function using only a small number of parameters.

A natural approach is based on the Karhunen-Loève decomposition of the covariance function (Ash and Gardner 1975). This approach has the advantage that it has an interpretation in terms of a small number of latent underlying curves. Let $\mathcal{T}$ be the closed interval $[a, b]$, and let $Y(t)$ be an L^2 mean zero stochastic process with continuous covariance $R(s,t)$. Let $\{\phi_\nu(t)\}$ form an orthonormal basis for the space spanned by the eigenfunctions of the nonzero eigenvalues of $R(s,t)$ satisfying

$$\lambda_\nu \phi_\nu(s) = \int_a^b R(s,t)\phi_\nu(t)dt. \tag{1}$$

The collection of eigenvalues $\{\lambda_\nu\}$ and eigenfunctions $\{\phi_\nu\}$ form a decomposition that is a generalization of the eigenvalue and eigenvector decomposition of a matrix.

The Karhunen-Loève expansion states that

$$Y(t) = \sum_\nu \xi_\nu \phi_\nu(t), \quad t \in \mathcal{T}, \tag{2}$$

where $\xi_\nu = \int_a^b Y(t)\phi_\nu(t)dt$, and the ξ_ν are uncorrelated random variables with $E(\xi_\nu) = 0, E(|\xi_\nu|^2) = \lambda_\nu$. The series converges in L^2 to $Y(t)$ uniformly in t (Grenander 1981).

The expansion (2) suggests that a zero mean stochastic process, $Y(t)$, has a representation as a stochastically weighted sum of latent functions $\{\phi_\nu\}$. The stochastic weights are analogous to the coefficients in a random coefficient model (see Longford 1993 for a discussion). Again, such models do not allow the design to be unspecified. If the covariance structure $R(s,t)$ can be approximated by a *finite* sum of latent functions, then

$$\tilde{Y}(t) = \sum_{\nu=1}^{K} \xi_\nu \phi_\nu(t) \tag{3}$$

describes an estimable model for $Y(t)$ if the functions $\{\phi_\nu\}$ either are known, or come from a known class of functions, such as the cubic splines. Of course, determining the functions is an issue in itself. Note that the distributions of the $\{\xi_\nu\}$ are not specified; in most practical settings, some additional assumptions must be made for these variables.

2.2 A Class of Latent Curve Models: Proto-splines

Suppose that an individual i is observed at specific design points, $\vec{t_i}$ and let $Y_i(t) \equiv Y_{it}$ be the observation at time t. A class of latent curve models for this data is

$$Y_i(t) = \sum_{\nu=1}^{K} \omega_{i\nu} \phi_\nu(t) + \epsilon_i(t), \tag{4}$$

where $\phi_\nu(t)$ is a basis curve, $\omega_{i\nu} \sim f(\cdot)$ are individual specific coefficients for the ν^{th} curve drawn from some distribution f to be specified, and $\epsilon_i(t)$ is an error function. This class is quite similar to the finite approximation to the Karhunen-Loève expansion for stochastic processes (3), with a change of variables and the addition of an error term. Meredith and Tisak (1990) describe two ways to generate basis curves: explicit specification of the functions, ϕ_ν; and via factor analysis, in which the ϕ_ν are the underlying factors. The first approach forces us to specify most of the underlying structure of the covariance explicitly–that is, provide the eigenfunctions in the Karhunen-Loève expansion. We would rather let the data drive this level of specification. As for the second approach, a factor analysis requires either a complete balanced design or the use of a technique such as the EM algorithm to handle missing data, and estimation is known to produce highly variable and not necessarily interpretable results.

We define a new hybrid class by specifying the ϕ_ν in a data-adaptive manner. It is hybrid in the sense that we partially specify the ϕ_ν. Let $\{\psi_j\}$ be an orthonormal basis to a smooth function space $\mathcal{H}$ of dimension $S \leq T$. There is no restriction on these basis functions and in particular they need not be a spline basis. The main idea is to use only a subset of the $\{\psi_j\}$ to construct each latent curve ϕ_ν. Let $\mathcal{I}(\nu)$ be an indexing function defined on the integers $\{1, \ldots, S\}$ which selects the basis functions used to construct the ν^{th} latent curve. We construct the general latent curve ϕ_ν as a deterministically weighted sum of the basis functions specified by the indexing function $\mathcal{I}$, so

$$\phi_\nu(t) = \sum_{j \in \mathcal{I}(\nu)} \eta_j \psi_j(t), \tag{5}$$

and then (4) is fully specified. In order to insure orthogonality of the $\{\phi_\nu\}$, the index sets given by $\mathcal{I}(\nu)$ must be disjoint. This restriction implies that once we decide to estimate more than one principal function, each is in a proper subspace of $\mathcal{H}$. We call the resulting $\{\phi_\nu\}$ *proto-splines*, since they are partial or restricted versions of functions from a smooth space $\mathcal{H}$.

The method requires $S \leq T$ parameters to build all K proto-splines–a substantial reduction compared to principal components analysis, which effectively requires $T(T+1)/2$ parameters to define its principle functions. For identifiability, the random coefficients $\{\omega_\nu\}$ are all defined to have variance one, and we allow the ϕ_ν to have norm other than one. Note that other approaches to smooth estimates of covariance must enforce an orthogonality constraint between fitted principal functions, with the notable exception of Silverman (1996). In Silverman's formulation, the norm of the function space insures the orthogonality. The potential effect of imposing this specific structure has not been investigated. A lack of external constraints is one of the advantages of the proto-spline approach; flexibility and parsimony are among the others.

To place this model in the context of those that have been developed previously, we examine it for two extreme cases. First, if $K = S$, then each proto-spline is just a rescaled version of a basis function, and these are effectively treated as known. If we choose a particular set of basis functions, then this becomes equivalent to the stochastic portion of the model proposed in Brumback (1996) and Brumback and Rice (1998). We note, however, that their choice of basis functions were designed to produce cubic spline *predictions* for individual curves, and as such do not focus on covariance modeling itself. At the other extreme, when $K = 1$, we are estimating a single smooth principal function. much as in the smooth principal components analysis of Rice and Silverman (1991) and Ramsay and Silverman (1997).

For proto-splines, we choose K to be small in relation to S, so that for equal-sized index sets, S/K bases are available for each principal function, and the "proto-spline" nature of the principal function estimates becomes more apparent. This intermediate case is compatible with a principal functions analysis, in which we expect that most of the variation in the process is captured in a few of the largest principal functions. We enforce a small number of these by our choice of K and maintain the orthogonality requirement by design. How restrictive is this designed orthogonality? It is clear from the nonparametric approaches of Rice and Silverman (1991) and Silverman (1996) that two principal functions can share a basis function ψ_j and still be orthogonal, thus our approach is sufficient but not necessary to ensure orthogonality.

3. INFERENCE FOR PROTOSPLINES

3.1 Preliminaries

We start with S basis functions, $\{\psi_1, \ldots, \psi_S\}$, spanning some space $\mathcal{H}$ to construct K latent curves, $\phi_1, \ldots, \phi_K$, where ϕ_ν is defined in (5). We add a mean component to the model, yielding

$$Y(t) = X(t)\beta(t) + \sum_{\nu=1}^{K} \omega_\nu \phi_\nu(t) + \epsilon(t), \tag{6}$$

in which $\epsilon(t)$ are i.i.d. $N(0, \sigma^2)$ random variables, and $X(t)$ is a design matrix capturing mean (fixed) effects. We assume that our observations are sampled at the same design points $\vec{t} = (t_1, \ldots, t_T)^T$, and drop the reference to $\vec{t}$ wherever possible. Thus the function $\beta(t)$ will become the familiar vector of fixed effects parameters, β, and $X(t)$ will become similarly discrete. To complete the model we specify that the ω_ν are i.i.d. standard Gaussian, so that the total variation is *implicitly* contained in the η_j. Alternatively we could parameterize it in Karhunen-Loève form (2) by leaving the

ω_ν unspecified and constraining the η_j to form ϕ_ν with norm one. After the parameters have been estimated, we can create a normalized fitted curve $\hat{\phi}_\nu^* = \hat{\phi}_\nu/(\hat{\phi}_\nu^T\hat{\phi}_\nu)^{\frac{1}{2}}$. The variance of the random effect associated with this curve is $s_\nu = \sum_{j\in\mathcal{I}(\nu)}\hat{\eta}_j^2$ mimicking λ_ν in (2).

If all of the ϕ_ν were known, then (6) would be a standard mixed effects model. What makes our latent curve model new is that we are able to leave the ϕ_ν only partially specified and proceed with model estimation–we let their exact form be driven by the data. A direct proof of the asymptotic properties of the maximum likelihood parameter estimates would require verification of regularity conditions. While our model is strictly speaking not a mixed effects model, we use asymptotics for mixed effects models as a starting point. Many mixed effects models take a relatively straightforward form, such as ARIMA and compound symmetry models. These and others are implemented in programs such as SAS PROC MIXED. For more complex mixed effects models, verification of regularity conditions is a non-trivial problem, which at a minimum involves a proof that the limiting value of the inverse of the information matrix is positive definite (Vonesh and Chinchilli, Section 6.2.3). Laird and Ware (1982) provided an estimation technique for traditional mixed effects models and developed the covariance properties of the fixed and random effects; no treatment was given to the parameters of the covariance structure itself. Miller (1977) provided the necessary conditions that establish the asymptotic distribution of all of the parameter estimates, including those of the covariance. Pinheiro (1994) built on this work to provide a simpler set of regularity conditions for mixed effects models, but these remain potentially non-trivial to show for new mixed effects models (cf. Jiang 1996, for conditions in an asymptotic analysis of REML estimates and a discussion of the dearth of asymptotic analyses for mixed models).

The above discussion sets the stage for our proof of the asymptotic properties of our latent curve model. Through a transformation of variables, we can construct a mixed effects model that has the same likelihood as the latent curve model. By the Likelihood Principle, estimation results and inferential procedures for the parameters of the mixed effects model can be applied to those of the latent curve model. We stress that while the new model we are about to describe is a mixed effects model, it is sufficiently complex that it is not embedded in a common covariance structure, such as those implemented in SAS PROC MIXED, and the asymptotics must be worked out–they do not simply follow from work already established for mixed effects models. Further, once the asymptotic properties have been established for our new model class, we must employ them under the original latent curve model formulation.

Let $Z_1 = [\psi_1 : \psi_2 : \cdots : \psi_{h(1)} : \psi_{h(1)+1} : \cdots : \psi_S]$ be the matrix formed using the selected basis vectors, ordered so that each group is in a contiguous block of the matrix. In the above, $h(\nu) = |\mathcal{I}(\nu)|$ is the number of basis functions in the latent curve ϕ_ν, and the subscript of "1" in Z_1 indicates that it is an individual-level matrix. Let $\gamma_\nu = (\eta_{\nu 1}, \ldots, \eta_{\nu h(\nu)})^T$

be the vector of parameters associated with ϕ_ν, and let $\Gamma = \bigoplus_{\nu=1}^{K} \gamma_\nu$. Then

$$\sum_{\nu=1}^{K} \omega_\nu \phi_\nu = Z_1 \Gamma \delta^*, \tag{7}$$

where $\delta^* \sim N(0, I_K)$. If we let $\delta_1 = \Gamma \delta^*$, then $\delta_1 \sim N(0, \Gamma \Gamma^T)$, and the random effects portion of the model has the standard form $Z_1\delta_1$, with all of the uncertainty relegated to the parameters in the covariance structure $\Gamma \Gamma^T$. Note that this structure is unique in the mixed effects model literature, and as such, asymptotics must be established for it. There are S random effects (instead of the original K) in this new model. This does not affect estimation of the variance components, but the individual-specific random effects should be estimated under the original model for interpretability. This underscores that fact that these are two very different models that happen to have the same likelihood.

In our asymptotic analysis for the likelihood-equivalent model, each individual is assumed to have the same mean effects design matrix X_1, with p_0 columns. The inclusion of explanatory covariates, which necessarily vary from individual to individual, would require a different analysis and is not explored here. The resulting model for one individual is then

$$Y_1 = X_1\beta + Z_1\delta_1 + \varepsilon_1, \tag{8}$$

in which Y_1 is a $T \times 1$ response vector, X_1 is a $T \times p_0$ design matrix, β is an $S \times 1$ set of parameters, Z_1 is the $T \times S$ matrix defined above, $\delta_1 \sim N(0, \Gamma \Gamma^T)$ is an $S \times 1$ set of individual random effects, and $\varepsilon_1 \sim N(0, \sigma^2 I)$ are the errors.

The model is reformulated for the entire set of m individuals by stacking the response vector and fixed effects design matrices for each individual into a common vector or matrix. The total number of observations is $n = mT$. This yields an $n \times 1$ response and error vector and an $n \times p_0$ design matrix X. The random effects are independent from individual to individual, so we redefine $Z = I_m \otimes Z_1$, and δ_1 is extended to an $mS \times 1$ vector δ.

For use in the proof to follow, we reformulate the model to conform to the structure established by Miller (1977). Let

$$Y = X\beta + \sum_{j=1}^{S} U^j a^j + \varepsilon, \tag{9}$$

where the U^j are the subset of columns in Z corresponding to ψ_j, and a^j is a vector of random effects associated with that basis vector. In other words, the i^{th} column of U^j is the column in Z corresponding to the j^{th} basis function for individual i. The a^j are the corresponding random effects for these "reordered" columns of Z. We let $\Gamma_m = I_m \otimes \Gamma$ be the transform extended to all individuals. The formulation (9) will be used with Pinheiro's (1994) regularity conditions for the asymptotics.

3.2 Likelihood

To simplify the analysis of the covariance of Y, define

$$\sigma_{ij} = \begin{cases} \eta_i \eta_j & \text{if } i \leq j \\ \sigma^2 & \text{if } i = j = 0. \end{cases} \tag{10}$$

where the indices are in the set $\Lambda = \{ij : \sigma_{ij} \neq 0; i, j = 1, \ldots, S\} \cup \{00\}$ Note that $\sigma_{ij} = 0$ whenever ψ_i and ψ_j correspond to different latent curves–these are ignorable terms. Define $G_{00} = I_n$, $a_{00} = 1$ and

$$G_{ij} = a_{ij}(U^i(U^j)^T + U^j(U^i)^T), \tag{11}$$

where

$$a_{ij} = \begin{cases} \frac{1}{2} & \text{if } i = j \\ 1 & \text{if } i \neq j. \end{cases} \tag{12}$$

for $ij \in \Lambda$ otherwise. The covariance of Y can then be expressed as:

$$\Sigma \stackrel{def}{=} \sum_{ij \in \Lambda} \sigma_{ij} G_{ij}. \tag{13}$$

Let $\alpha = (\sigma^2, \eta_1, \ldots, \eta_S)$ be the covariance parameters. The parameter space Θ for the model is

$$\Theta = \left\{ \theta \in \Re^{p_0+S+1} | \theta = (\beta^T, \alpha^T)^T,\ \beta \in \Re^{p_0}; \alpha \in \Re^{S+1} \text{ such that } \sigma^2 > 0, \eta_{\nu 1} > 0 \text{ for } \nu = 1, \ldots, K \right\}, \tag{14}$$

where $\eta_{\nu 1}$ is the "first" parameter (under the ordering imposed by the basis functions) in the ν^{th} group.

The log-likelihood for Y given θ is

$$\ell(\theta|Y) = \frac{1}{2} \left\{ n \log(2\pi) + \log(\det \Sigma) + e^T \Sigma^{-1} e \right\}, \tag{15}$$

where $e = Y - X\beta$ is a vector of deviations from the expected response. Now that we know the form of the likelihood, we can estimate the parameters using numerical optimization algorithms such as `nlminb` in Splus. `nlminb` is a very basic program that minimizes the objective function provided by the user. Even though the likelihood is that of a (Gaussian) mixed effects model, it is sufficiently complex so that it has not been implemented in the standard packages such as SAS PROC MIXED. We developed a set of Splus functions to compute the likelihood and gradient of any model in the proposed latent curve class. We determined the specific form of the gradient of the log-likelihood; using it in the optimization dramatically improves the rate of convergence. Details of an application are given in Section 4..

3.3 Asymptotic Framework

Again, the framework we are employing was developed for mixed effects models. These models typically have many replications over a small number of design points. For our model, the design involves points in time, as in an annual survey. We expect to have a reasonably large number of subjects, so the asymptotics reflect what happens as the number of individuals goes to infinity, and we pursue this form of the asymptotics in the remainder of this section. Alternative forms, such as increasing the sampling frequency, will lead to alternate approximations and will not be pursued here.

3.4 Discussion

We are about to prove that for a balanced design, the maximum likelihood estimates for proto-splines are consistent and asymptotically Gaussian. We will do this by first establishing similar properties for the likelihood-equivalent mixed effects model. The results from that proof then apply to our new latent curve models. The asymptotic covariances are shown to be the inverse of the information matrix, so the estimates are asymptotically efficient. These asymptotic results may be used as approximations for unbalanced designs.

The likelihood-equivalent mixed effects model was developed because it falls within the general framework presented in Pinheiro (1994). The next step is to verify the abstract conditions he formulated. His work provides a convenient formulation of conditions that may be obviously satisfied for some (simpler) mixed effects models, but involved proofs are required for the model that we have developed.

3.5 Regularity Conditions

Let $\theta_0 = (\beta_0^T, \alpha_0^T)^T$ denote the true parameter vector and $\Sigma_0 = I_m \otimes \Psi_0$ the associated covariance matrix of the response vector Y.

Assumption 1 Each individual has the same fixed effects design matrix X_1 and it is of full rank.

Remark 3.1. Let $C_\beta = X_1^T \Psi_0^{-1} X_1$. This assumption can be replaced by the assumption that the fixed effects matrices behave as draws from a distribution and $E[C_\beta]$ is positive definite. The mean effect design matrix is under our control, and we will usually choose X to consist of the same basis functions as Z, so it is certainly of full rank. For saturated mean effects, X can also be chosen to be of full rank. In our application each individual has the same fixed effects design matrix.

Assumption 2 $n \geq p_0 + p_1 + 1$ and $T > S$, where $p_1 = \sum_{\nu=1}^{K} h(\nu)(h(\nu)+1)/2$.

Remark 3.2. Note that $p_0 \leq T$ and $p_1 \leq T(T+1)/2$, so the total number of observations, $n = mT$ is larger than $T + T(T+1)/2 + 1$ for $m > 2T$, with T fixed, which can be insured by observing a large enough sample. The second condition ensures that the number of observations per individual exceeds the number of random effects.

Assumption 3 Each latent curve ϕ_ν is comprised of at most five basis functions.

Remark 3.3. In Appendix A.3., we argue that the result is true for arbitrary numbers of basis functions. This assumption is true for most proto-spline models of interest.

Let $\upsilon_{ij} = \text{rank}(G_{ij})$, for $ij \in \Lambda \setminus 00$. In our model, $\upsilon_{ij} = 2ma_{ij}$. The final element is $\upsilon_{00} = n - \text{rank}[U^1 : \cdots : U^S] = (T-S)m$. Let C_1 be the $(p_1+1)\times(p_1+1)$ matrix defined by $[C_1]_{ij,kl} = (1/2)\lim_{n\to\infty}\text{tr}(\Sigma_0^{-1}G_{ij}\Sigma_0^{-1}G_{kl})/(\upsilon_{ij}\upsilon_{kl})^{\frac{1}{2}}$, $ij, kl \in \Lambda$. C_1 is the asymptotic covariance matrix of the variance components. It is proven in Appendix A.2. that the limits exist and an explicit expression is given for C_1. The positive definiteness of C_1 is proven in Appendix A.3..

The covariance matrix Σ is parameterized in terms of $[\sigma_{ij}]$ while the likelihood is parameterized in terms of α. However the mapping from the former to the latter is smooth and one-to-one. To translate our results from the variance components parameterization we form D, the $(p_1+1)\times(S+1)$ gradient matrix of $[\sigma_{ij}]$:

$$[D]_{ij,k} = \frac{\partial\sigma_{ij}}{\partial\alpha_k},\ ij\in\Lambda,\ k = 0,1,\ldots,S,$$

The specific entries in the gradient can easily be written down and are given in Appendix A.4..

Let $s = \text{diag}(1, s_{11}, \ldots, s_{(S-1)S})$ and define $C_\alpha \stackrel{def}{=} D^T s^{\frac{1}{2}} C_1 s^{\frac{1}{2}} D$, where s_{ij} is defined in Condition A.10 of the Appendix.

3.6 Main Asymptotic Result

We now give the main technical result: the consistency and asymptotic distribution of the parameter estimates.

Theorem Under Assumptions 1 through 3, and letting θ_0 be an interior point of Θ representing the true parameter vector and $J = \begin{bmatrix} C_\beta & 0 \\ 0 & C_\alpha \end{bmatrix}$, there exists a sequence of estimates $\hat\theta_n = (\hat\beta_n^T, \hat\alpha_n^T)^T$ with the following properties.

1. Given $\epsilon > 0, \exists\delta = \delta(\epsilon), 0 < \delta < \infty$ and $n_0 = n_0(\epsilon)$ such that $\forall n > n_0$

$$P_{\theta_0}\left(\left.\frac{\partial\ell(\theta)}{\partial\theta}\right|_{\theta=\hat\theta_n} = 0; \|\hat\beta_n - \beta_0\| < \frac{\delta}{m^{\frac{1}{2}}} \text{ and } |\hat\alpha_{ni} - \alpha_{0i}| < \frac{\delta}{m^{\frac{1}{2}}},\ i = 0,\ldots,S\right) \geq 1-\epsilon.$$

2. The (p_0+p_1+1)-dimensional vector with the first p_0 components given by $m^{\frac{1}{2}}(\hat\beta_n - \hat\beta_0)$ and the last p_1+1 components given by $m^{\frac{1}{2}}(\hat\alpha_{ni} - \alpha_{0i}),\ i = 0,\ldots,p_1$ converges in distribution to a $N(0, J^{-1})$.

The proof of this result is given in the Appendix.

The information matrix, J, indicates that the fixed and random effects are asymptotically uncorrelated, which is desirable. These results are for maximum likelihood estimation, and can be extended to restricted maximum likelihood estimation (REML) by altering certain regularity assumptions. Pinheiro (1994) describes such an extension in his

Section 3.2. Additional work is required to show that the asymptotics hold for unbalanced designs, for which the inclusion of explanatory covariates other than the time scale is an example.

3.7 Confidence intervals for Proto-splines

Based on the theorem, the parameter estimates (multiplied by $m^{\frac{1}{2}}$) are asymptotically Gaussian, with covariances given by the inverse of the information matrix, J. The components of J, C_β and C_α, are completely known and described in Appendix A.2..

We construct confidence intervals using the asymptotic covariance in the usual manner. Let $\hat{\alpha} = (\hat{\sigma}^2, \hat{\eta}_1, \ldots, \hat{\eta}_S)^T$ be the parameter estimates associated with the residual variance and random effects parameters, and let their variance-covariance matrix be denoted by $[mC_\alpha]^{-1}$, where m is the number of subjects. Since this vector is (asymptotically) Gaussian, we have a complete description of its distribution. Confidence intervals for these estimates are based on the variances located on the diagonal of $[mC_\alpha]^{-1}$. A similar result for the fixed effects parameter estimates follows using C_β.

We can construct confidence intervals for the latent curves ϕ_ν, since these are just (time-dependent) linear combinations of the parameter estimates. The vector given by $Z_1\hat{\alpha}$ for a Z_1 restricted to a specific group ν has the following distribution (asymptotically):

$$Z_1\hat{\alpha} \sim N(Z_1\alpha_0, Z_1[mC_\alpha]^{-1}Z_1^T),$$

where α_0 is the true parameter vector. The confidence intervals for these points are again a function of their variances, which are located on the diagonal of $Z[mC_\alpha]^{-1}Z^T$. Since the entire distribution and covariance structure of the parameter estimates are known, other inferential procedures, including hypothesis testing, multiple comparisons, and confidence bands, are possible.

4. APPLICATION TO THE ANALYSIS OF WAGE INEQUALITY

In this section we will use proto-spline models to describe the variation in wage mobility for young workers. Topel and Ward (1992) find that "the first 10 years of a career will account for 66 percent of lifetime wage growth for male high school graduates," so the early career of young workers is pivotal to their development.

4.1 Data Description

We will be investigating a dataset from the National Longitudinal Survey of Youth: a representative sample of young men aged 14-21 was interviewed in 1979 and has been interviewed yearly since then, with 1994 the most recently available year included in our data. For comparability with other cohorts, we also selected only non-Hispanic whites aged 14-21 in the first year of the surveys, with a resulting sample size of 2,427. A detailed description of the construction of the dataset is given in Bernhardt et. al. (1997).

4.2 Model Description

In this section we describe two proto-spline models. They are chosen to give two disparate descriptions of the variation in trajectories. Many alternate models can be developed to reflect different labor economic theories about the process.

Model I consists of a cubic spline spanning all ages, with a knot at age 23. This curve represents the strongest single systematic variation in the stochastic process, much as the first principal component of a covariance matrix does. By including every age in each basis function, the model should flexibly identify the long-range dependencies in the variation, if these dominate the overall variation.

Model II involves a partitioning of the ages into different (and thus independent) groups and choosing a different basis for each. This allows local features of variation to emerge. By choosing non-overlapping intervals, we have much more freedom in setting the basis for each proto-spline, since they are necessarily orthogonal to each other. The age groupings are 16-17, 18-22, 23-25, and 26-37, so $K = 4$. The bases are respectively a constant, a cubic spline (one centered knot), a quadratic, and another cubic spline. The results of this model could yield insight into any medium-range dependencies in the process. By creating partitions, we have excluded the possibility that an individual response is the sum of overlapping latent curves. We do this for illustrative purposes only. While overlapping curves could yield additional insight, the orthogonality requirement often places a restriction on the set of available bases for each curve.

4.3 Model Fit

The proto-spline models are fit by maximizing the log–likelihood (15). The MLEs were determined numerically using the Splus function `nlminb` with gradient information. The latent curves ϕ_ν are of primary interest, and their MLEs are $\hat{\phi}_\nu(t) = \sum_{j \in \mathcal{I}(\nu)} \hat{\eta}_j \psi_j(t)$, based on (5). These unnormalized latent curves represent variation beyond the mean response. They are mutually orthogonal with the random coefficient ω_ν for each curve drawn from i.i.d. $N(0, 1)$ distributions.

The mean effects for these models were fit using a saturated design. The fits resembled quadratic curves and their shapes were quite similar, so we do not display them here. The squared values of the fitted latent curve is an estimate of the structured variation contributed by that curve at a given age. The total variation is simply the sum of these structured parts and the residual variation.

Model I is quite revealing. The single latent curve depicted in Figure 2 captures most of the structured variation, starting out near zero at age 16, with fairly a consistent slope until age 25, where the rate of growth begins to diminish.

Table 1. Analysis of Variance

Source	*Model I*	*% of Total Variance (0.273) Explained* *Model II*
Fixed Effects	22.8	22.8
Proto-splines	40.4	49.3
Residual	36.8	27.9
Total	100.0	100.0

One interpretation for this curve is that wage gains (or losses–the random coefficient is negative half of the time, and the overall mean effect is positive quadratic growth) in the early ages establish the later magnitude of growth (loss). This implies that wages remain consistently high or low over time, resulting in long-term wage stratification. However, this interpretation rests on the assumption that nearly all wage growth is captured in this one curve. In fact, the residual variation here is the largest of our two models, so there may be quite a bit of structured variation left in the trajectories. Proto-spline models are designed to capture the longitudinal features of the data. Note, however, the model's prediction of the cross-sectional variances still capture the general shape of their empirical counterparts (Figure 3).

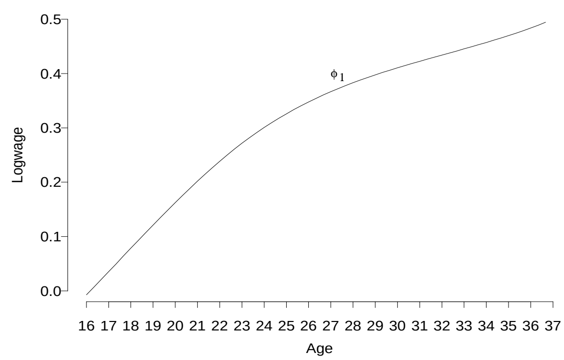

Figure 2. The fitted latent curves for a single curve proto-spline model. This single curve accounts for 40.4% of the variation between the trajectories. See Table 1 for detailed variance components analysis.

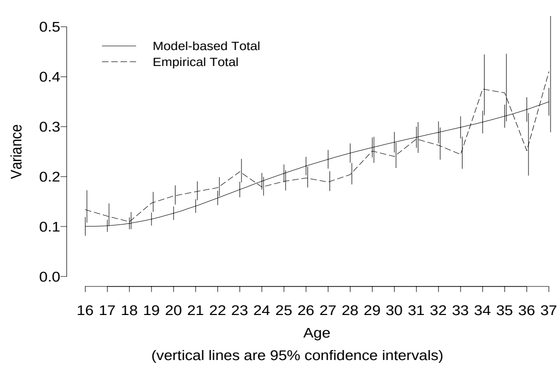

Figure 3. Model-based and empirical variances for the single curve proto-spline model. The vertical lines are pointwise 95% confidence intervals for each value. While the model matches the overall shape of the empirical variances, there are some systematic deviations.

Model I captures the strong dependence of early and later wages while, in contrast, Model II posits that wage variation for age groups 16-17, 18-22, 23-25, and 26-37, are independent. We can view Model II as four separate time windows from which we are viewing wage growth. Ignoring the earliest age group, we can label these groupings as *college* (for some), *matching* of employees to employers, and a subsequent *stable* period in which many career choices have been made. These groupings are all approximate, since each period involves a mixture of many different activities. The fitted latent curves are presented in Figure 4. At the college ages, ϕ_2 is a dramatically increasing curve representing deviations from the mean response. Translated into total wage, the steep curve implies that some people are making consistent gains in wages, while others are staying relatively flat or decreasing with respect to the mean wage (again, the random coefficient is negative half of the time). This latent curve captures the sharp growth in wages that immediately follows college for some individuals and the relative stagnation in wages for others. The period of employer matching represented by ϕ_3 shows a slight increase or decrease in that age range, with a change in the opposite direction in between. We interpret this as some indication of the "job churning" process prevalent at these ages. We fit a quadratic to these points, which *allows* for such fluctuation, but does not require it. The strength of the proto-spline model is that it flexibly captures the dominant variation on a time domain. The curve for the last age group shows moderate wage growth/loss for most of this "stable" period, with a slight change of direction in the oldest ages. It is striking that these "stylized facts" of labor market economics are easily extracted by proto-spline models.

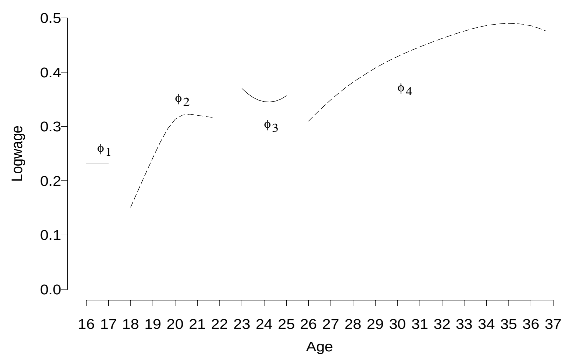

Figure 4. The four fitted latent curves for the multiple regime proto-spline model. Taken together, these curves encapsulate many stylized facts about early-career wage trajectories.

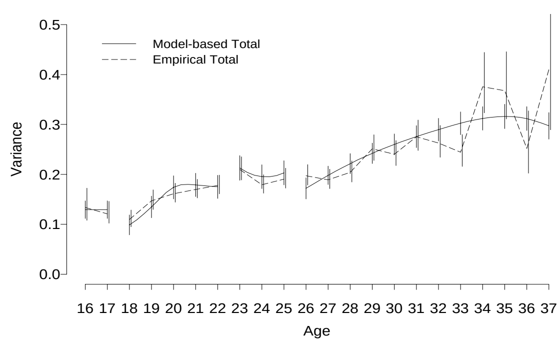

Figure 5. Model-based and empirical variances for multiple regime proto-spline model. The vertical lines are pointwise 95% confidence intervals for each value. The model-based variances are more reflective of the empirical variances than those of the match between the model variances using a single curve proto-spline model.

The comparison to the empirical variance in Figure 5 is encouraging. The model closely tracks the cross-sectional variation at all but the later ages. Thus this age-partitioned analysis of the data captures cross-sectional and medium-range dependence quite well.

Model I can be extended to include additional basis vectors and hence orthogonal latent curves. Note, however, that the orthogonality requirement severely restricts the form of these latent curves.

5. DISCUSSION

We have presented a new model for the variation in longitudinal data that makes the original idea of latent curve analysis proposed by Meredith and Tisak (1990) feasible. It further extends that work by establishing the asymptotic properties of the associated parameter estimates. The approach is rooted in the theory of stochastic processes and the Karhunen-Loève expansion. We provide for extensive flexibilty in proto-spline models by giving the researcher control over the basis functions and groupings employed, and we suggest that they be chosen to reflect theoretical hypotheses about the structure of the phenomena. Describing complex covariance structures using a small number of stochastically weighted *latent curves* provides the researcher with a parsimonious and highly interpretable model. This will be useful when the covariance diverges from simpler structures. When long-term variation is of interest, we contend that traditional time series error models do not offer interpretable variance decompositions (see Jones 1990 for a discussion). Further, by using these parsimonious models, we avoid the potentially infeasible estimation of a fully general covariance matrix while still avoiding the assumption of an exact form for the basis functions, which is required in traditional models.

In future work, we would like to identify function spaces that handle overlapping multiple-curve models and still maintain the orthogonality property, and those that are specialized to handle jump processes at several time points. Mixtures of wavelet and spline bases are natural candidates. Some features may force us to reconsider the orthogonality requirement, if its imposition severely limits the set of available spaces. To identify and allow for differences in short-term behavior, we might incorporate more complex error structures into the modeling class as well. The basis functions (and knots, when appropriate) can be chosen adaptively using model selection criteria such as BIC – see Kooperberg and Stone (1992). The properties of such approaches, however, need to be explored for these models in realistic applications.

APPENDIX: PROOF OF THE MAIN RESULT

In this Appendix we prove the maximum likelihood estimates for the proto-spline class of models are consistent and asymptotically Gaussian. This work involves a detailed analysis of the model to insure that it conforms to a set of regularity conditions established by Pinheiro (1994). Once these conditions have been verified, the asymptotic results follow.

A.1 Proof of Regularity Conditions

We now formally verify the regularity conditions. The numbering is the same as that in Pinheiro (1994), section 3.1. The first two conditions mirror the assumptions in Section 3.5.

Condition A.1 *The matrix X is of full rank p_0.*

Condition A.2 $n \geq p_0 + p_1 + 1$.

Condition A.3 *The partitioned matrix $[X : U^j]$ has rank greater than p_0 for $j = 1, \ldots, S$.*

We first verify the condition when X is composed of the same basis functions as Z. Note that the row rank of X is $T > S = p_0$, so the rank of $[X : U^j]$ is determined by its column rank. Appending U^j does introduce some dependence as follows: Null Space$[X : U^j] = \{(-e_j^T, 1_m^T)^T\}$, where e_j is the standard basis vector of length S consisting of a 1 in the j^{th} row, and 1_m is a vector of m ones. Thus rank$[X : U^j] = p_0 + m - \text{Nullity}[X : U^j] = p_0 + m - 1 > p_0$. A similar argument applies for saturated mean effects, X, so the condition is satisfied. □

Condition A.4 *The matrices $G_{11}, G_{22}, \ldots, G_{(S-1)S}, G_{00}$ are linearly independent.*

We first show that $\{G_{11}, G_{22}, \ldots, G_{(S-1)S}\}$ are always independent. The basis functions ψ_j are orthonormal and form an independent set, so the corresponding matrices U^j are linearly independent. If

$$\sum_{ij \in \Lambda \setminus 00} \tau_{ij} G_{ij} = \sum_{ij \in \Lambda \setminus 00} \tau_{ij} \alpha_{ij} (U^i (U^j)^T + U^j (U^i)^T) = 0, \tag{A.1}$$

then a simply counting argument removes the coefficient α_{ij} and implies

$$\sum_{i=1}^{S} \sum_{j=1}^{S} \tau_{ij} U^i (U^j)^T = 0, \tag{A.2}$$

where $\tau_{ij} \stackrel{def}{=} \tau_{ji}$ if $i > j$. Then

$$\sum_{i=1}^{S} U^i \sum_{j=1}^{S} \tau_{ij} (U^j)^T = 0, \tag{A.3}$$

and the independence of U^i forces each of the terms in the second sum to be zero, so for all i,

$$\sum_{j=1}^{S} \tau_{ij} (U^j)^T = 0. \tag{A.4}$$

Again, the independence of U^j implies that $\tau_{ij} = 0$ for every $j = 1, \ldots, S$. Since the result is true for all i, $\tau_{ij} = 0$ for all $ij \in \Lambda \setminus 00$, and thus the G_{ij} are linearly independent. We can chose the basis functions ψ_j so that $G_{00} = I_n \notin \text{span}\{\psi_1, \ldots, \psi_S\}$, and then all G_{ij} are linearly independent. The requirement on G_{00} and the span is easy to insure in practice, but must not be overlooked: if the ψ_j are the T standard basis functions (which is more than we allow), then I_n is in the span and the random effects will be confounded by the residual effects. □

Condition A.5 *The number of observed levels (m) of the random effects goes to infinity.*

In the proto-spline model, the levels are the m individuals, which asymptotically go to infinity.□

Condition A.6 $\lim_{n\to\infty} \upsilon_{00}/n$ *exists and is positive.*

The rank of each U^j is m, they are of dimension $mT \times m$, and are linearly independent, so the partitioned matrix has full rank: $\text{rank}[U^1 : \cdots : U^S] = mS$ and thus $\upsilon_{00} = mT - mS$.

$$\lim_{n\to\infty} \upsilon_{00}/n = \lim_{m\to\infty} \frac{(mT - mS)}{mT} = \frac{T-S}{T} \tag{A.5}$$

As long as $S < T$, this condition is satisfied. □

Condition A.7 (i) *There exists a sequence of positive quantities υ_n depending on n and going to infinity such that $C_0 = \lim_{n\to\infty} X^T\Sigma_0^{-1}X/\upsilon_n$ exists and is positive definite.*

From (8),

$$\Psi_0 = Z_1 , \, , ^T Z_1^T + \sigma^2 I, \tag{A.6}$$

$\Sigma_0 = I_m \otimes \Psi_0$, and $\Sigma_0^{-1} = I_m \otimes \Psi_0^{-1}$. If Ψ_0 is positive definite, then Ψ_0^{-1} and Σ_0^{-1} will be as well.

The matrix $Z_1 , \, , ^T Z_1^T$ is positive semi-definite because it is the variance-covariance matrix of the random variable $Y = (Z_1 , \,)\delta^*$, where $\delta^* \sim N(0, I_K)$. Its eigenvalues are all non-negative, so the eigenvalues of $Z_1 , \, , ^T Z_1^T + \sigma^2 I$ are all greater than or equal to σ^2, which we assume is non-zero. So the eigenvalues of Ψ_0 are all positive, and Ψ_0 is positive definite.

Using Hadi (1996) Result 6.5, it can be deduced that

$$\Psi_0^{-1} = \frac{1}{\sigma^2}\left[I - Z_1 [\bigoplus_{\nu=1}^{K} \lambda_\nu \gamma_\nu \gamma_\nu^T] Z_1^T \right], \tag{A.7}$$

and the entire structure of Σ_0^{-1} is now known.

We assume that each individual has the same fixed effects design matrix X_1 and construct them to be of full rank p_0. By setting $\upsilon_n = n/T$,

$$\lim_{n\to\infty} \frac{X^T\Sigma_0^{-1}X}{n/T} = \lim_{m\to\infty} \frac{X^T(I_m \otimes \Psi_0^{-1})X}{m} = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} X_1^T\Psi_0^{-1}X_1 = X_1^T\Psi_0^{-1}X_1 \tag{A.8}$$

for an arbitrary i. Since this last expression involves a matrix that does not depend on n, the limit C_0 exists. Observation 7.1.6 in Horn and Johnson (1985) states that given Ψ_0^{-1} positive definite and X_1 of full rank $p_0 \leq T$, $X_1^T\Psi_0^{-1}X_1$ is positive definite, and thus so is C_0. □

A.2 Existence and Structure of C_1

We prove the existence of C_1, the asymptotic covariance matrix of the variance components, and give an explicit expression for it. Consider first the case where $ij, kl \in \Lambda \setminus 00$. Recalling that $\Sigma_0^{-1} = I_m \otimes \Psi_0^{-1}$, and $\upsilon_{ij} = m$ or $2m$ when $i \neq j$ or $i = j$

respectively, we can express

$$[C_1]_{ij,kl} = \frac{1}{2}\alpha^*_{ij}\alpha^*_{kl}\text{tr}\left[\Psi_0^{-1}(\psi_i\psi_j^T + \psi_j\psi_i^T)\Psi_0^{-1}(\psi_k\psi_l^T + \psi_l\psi_k^T)\right] \quad \text{(A.9)}$$

where

$$\alpha^*_{ij} = \begin{cases} a_{ij} & \text{if } i = j \\ a_{ij}/\sqrt{2} & \text{if } i \neq j. \end{cases} \quad \text{(A.10)}$$

As $[C_1]_{ij}$ does not depend on m, the limit exists.

By rearranging the terms inside of the trace (Result 5.1.d in Hadi 1996), so that the expression is a scalar, we can show that all but the very last row and column of C_1 is a block diagonal matrix, with each group in its own block:

$$[C_1]_{ij,kl} = \frac{\alpha^*_{ij}\alpha^*_{kl}}{\sigma^4}\left[(e_l^T e_i - \lambda_\nu\eta_l\eta_i)(e_j^T e_k - \lambda_\nu\eta_j\eta_k) + (e_l^T e_j - \lambda_\nu\eta_l\eta_j)(e_i^T e_k - \lambda_\nu\eta_i\eta_k)\right], \quad \text{(A.11)}$$

where all of $i, j, k, l \in \mathcal{I}(\nu)$ for some common group ν.

A similar reduction when taking limits occurs when one or both of $ij, kl \in 00$. One can show:

$$[C_1]_{00,kl} = \frac{2\alpha^*_{kl}}{\sigma^4\sqrt{T-S}}\left[e_l^T e_k + \lambda_\nu\eta_k\eta_l(\lambda_\nu s_\nu - 2)\right]. \quad \text{(A.12)}$$

where ν is the common group of index kl, and s_ν are the variance components defined in the text of Section 3.1.

$$[C_1]_{00,00} = \frac{1}{\sigma^4(T-S)}\left[T - 2\sum_{\nu=1}^{K}\lambda_\nu s_\nu + \sum_{\nu=1}^{K}\lambda_\nu^2 s_\nu^2\right], \quad \text{(A.13)}$$

where $T > S$ is the total number of design points in the model. Let A_ν be the portion of C_1 corresponding to the ν^{th} group. Recall that $p_1 = \sum_{\nu=1}^{K} h(\nu)(h(\nu)+1)/2$, and define the $p_1 \times p_1$ submatrix $C_1^* = \text{diag}(A_1, A_2, \ldots, A_K)$. Then C_1 has the form

$$C_1 = \begin{pmatrix} A_1 & 0 & 0 & 0 & b_1 \\ 0 & A_2 & 0 & 0 & b_2 \\ 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & 0 & A_K & b_K \\ b_1^T & b_2^T & \cdots & b_K^T & d \end{pmatrix}, \quad \text{(A.14)}$$

where b_ν are the elements $[C_1]_{00,kl}$ for all $kl \in \mathcal{I}_\nu$, and $d = [C_1]_{00,00}$.

A.3 Positive definiteness of C_1

Condition A.7 (ii) *C_1 is positive definite.*

Our proof that C_1 is positive definite involves an analysis of the structure induced not only by the trace operation, but by the elements of the γ_ν as well (see also Appendix A.1 of Scott 1998 where we discuss a more direct approach based on the positive definiteness of the original Σ_0^{-1}).

We divide the problem into two subproblems. We first show that the submatrix C_1^* (the first p_1 rows and columns of C_1) is positive definite. We then show that after the addition of the final row and column, this property is still maintained.

The structure of the model insures that C_1^* is block diagonal with blocks A_ν so if the A_ν are positive definite, then so is C_1^*.

Lemma *The eigenvalues for each block in C_1^* corresponding to a group ν, with $1 \leq h(\nu) \leq 5$ are as follows:*

- $1/\sigma^4$, *with multiplicity* $\binom{h(\nu)}{2}$
- λ_ν/σ^2, *with multiplicity* $h(\nu)-1$
- λ_ν^2, *with multiplicity* 1.

Proof of the Lemma

This has been verified for $h(\nu) = 2, 3, 4, 5$ using Mathematica software. As long as $\sigma^2 > 0$, all of the eigenvalues of the ν^{th} block are positive, and thus the matrix in block ν is positive definite. When $h(\nu) = 1$, the corresponding block is the scalar $(1/\sigma^4)(1-\lambda_\nu\eta_j^2)^2$ for some index $j \in \mathcal{I}(\nu)$, which can be simplified to $(\sigma^2\lambda_\nu)^2/\sigma^4 = \lambda_\nu^2$. Provided $\sigma^2 > 0$, this term is positive, so a block with $h(\nu) = 1$ is positive definite.

For most models, latent curves employing at most 5 basis functions should suffice. Directly computing the eigenvalues when $h(\nu) >> 5$ is infeasible using existing software. However, because $h(\nu) = 4$ contains all of the possible "types" of elements in the matrix C_1^*, we believe that our findings will hold true for larger $h(\nu)$. Approaches to extending the results to larger $h(\nu)$ are discussed in Appendix A.3 of Scott (1998).

We still must show that the inclusion of the last row and column results in C_1 being positive definite.

The matrix C_1 will be positive definite if

$$d > \sum_{\nu=1}^{K} b_\nu^T A_\nu^{-1} b_\nu. \tag{A.15}$$

The exact form of the inverses A_ν^{-1} are known and are discussed in Appendix A.2 of Scott (1998). Using these, we compute the terms in the sum explicitly, and find that

$$b_\nu^T A_\nu^{-1} b_\nu = \frac{1}{T-S}\left[\frac{h(\nu)-1}{\sigma^4} + \lambda_\nu^2\right],$$

so their sum is

$$\frac{1}{T-S}\left[\frac{\sum_\nu (h(\nu)-1)}{\sigma^4} + \sum_\nu \lambda_\nu^2\right].$$

Further,

$$\begin{aligned} d &= \frac{1}{\sigma^4(T-S)}\left[T - 2\sum_{\nu=1}^{K}\lambda_\nu s_\nu + \sum_{\nu=1}^{K}\lambda_\nu^2 s_\nu^2\right] \\ &= \frac{1}{\sigma^4(T-S)}\left[T - K + \sum_{\nu=1}^{K}(1 - 2\lambda_\nu s_\nu + \lambda_\nu^2 s_\nu^2)\right] \\ &= \frac{1}{\sigma^4(T-S)}\left[T - K + \sum_{\nu=1}^{K}(1 - \lambda_\nu s_\nu)^2\right] \\ &= \frac{1}{\sigma^4(T-S)}\left[T - K + \sum_{\nu=1}^{K}(\sigma^2\lambda_\nu)^2\right] \\ &= \frac{1}{T-S}\left[\frac{T-K}{\sigma^4} + \sum_{\nu=1}^{K}\lambda_\nu^2\right], \end{aligned}$$

which reduces requirement (A.15) to

$$\frac{T-K}{\sigma^4} + \sum_{\nu=1}^{K}\lambda_\nu^2 > \frac{\sum_{\nu=1}^{K}(h(\nu)-1)}{\sigma^4} + \sum_{\nu=1}^{K}\lambda_\nu^2.$$

Since $\sum_\nu h(\nu) = S$, this requirement is nothing more than $T - K > S - K$ or $T > S$, which we require for other reasons.

A.4 Conditions Related to the Parameterization

The covariance in the proto-spline model is parameterized or structured, so we must meet the following additional regularity conditions that correspond to 3.3.2 – 3.3.4 of Pinheiro (1994).

Condition A.8 *The function f that maps the parameters to the respective variance components is twice differentiable with continuous second derivatives.*

The parameters in the proto-spline model are α. Let $f_{ij}(\alpha) = \sigma_{ij}$, where σ_{ij} is the ij^{th} covariance component described in Section 3.2. So

$$f_{ij}(\alpha) = \eta_i\eta_j, \tag{A.16}$$

for $ij \neq 00$ and $f_{00}(\alpha) = \sigma^2$. We now state the first and second partial derivatives of f.

$$\frac{\partial f_{ij}}{\partial \eta_k} = \begin{cases} 2\eta_i & \text{if } i = j = k \\ \eta_j & \text{if } i = k \ \& \ i \neq j \\ \eta_i & \text{if } j = k \ \& \ i \neq j \\ 0 & \text{otherwise.} \end{cases} \tag{A.17}$$

$$\frac{\partial f_{ij}}{\partial \sigma^2} = 0, \quad \frac{\partial f_{00}}{\partial \sigma^2} = 1. \tag{A.18}$$

The second partial derivatives are all zero, with the following exceptions:

$$\frac{\partial^2 f_{ii}}{\partial \eta_i^2} = 2, \quad \frac{\partial^2 f_{ij}}{\partial \eta_i \partial \eta_j} = 1. \tag{A.19}$$

Since a constant function is continuous, the requirement is satisfied. □

Condition A.9 *f is one-to-one, i.e. $\alpha \neq \alpha' \Rightarrow f(\alpha) \neq f(\alpha')$.*

It is easier to check the contrapositive of this statement. Recall (A.16) to see that each component of f depends on at most two elements of α. Abusing notation somewhat, we must show

$$f(\eta_i, \eta_j) = f(\eta_i', \eta_j') \Rightarrow \eta_i = \eta_i' \ \& \ \eta_j = \eta_j' \tag{A.20}$$

for all $ij \in \Lambda \setminus 00$. The image $f(\alpha)$ involves η_i and η_j in the definitions of σ_{ii}, σ_{jj}, and σ_{ij}, so the above reduces to showing that

$$\eta_i^2 = (\eta_i')^2 \ \& \ \eta_j^2 = (\eta_j')^2 \ \& \ \eta_i \eta_j = \eta_i' \eta_j' \Rightarrow \eta_i = \eta_i' \ \& \ \eta_j = \eta_j'. \tag{A.21}$$

This holds except when one of $\eta_i' = -\eta_i$ or $\eta_j' = -\eta_j$. If we *fix the sign* of one of the parameters η_j in each group ν to be positive, we will be able to identify the parameters in the model. We arbitrarily set the sign of the "first" parameter in each group (under the ordering induced by the ordering of the basis functions) to be positive. When $ij = 00$, the parameter $f_{00}(\sigma^2) = \sigma^2$ is directly identifiable. □

Condition A.10 $\lim_{n \to \infty} v_{ij}/m = s_{ij}$ *exists and is positive.*

In the proto-spline model, $v_{ij} = m$ or $2m$ so the limit is respectively 1 or 2. □

Using the chain rule, Pinheiro (1994) lays out the relationship between the derivatives of the likelihood with respect to the original and induced parameters. The gradient matrix of f, D, was defined in Section 3.5.

We form H_f, the $(S+1) \times (S+1) \times (p_1+1)$ Hessian array of f

$$[H_f]_{ij,k,l} = \frac{\partial^2 f_{ij}(\alpha)}{\partial \alpha_k \partial \alpha_l}, \ ij \in \Lambda, \ k, l = 0, 1, \ldots, S.$$

This yields

$$\begin{aligned}
\frac{\partial \ell(\beta, \alpha|y)}{\partial \alpha} &= D^T \frac{\partial \ell(\beta, \sigma|y)}{\partial \sigma} \\
\frac{\partial^2 \ell(\beta, \alpha|y)}{\partial \alpha \partial \alpha^T} &= D^T \frac{\partial^2 \ell(\beta, \sigma|y)}{\partial \sigma \partial \sigma^T} D + H_f \frac{\partial \ell(\beta, \alpha|y)}{\partial \sigma} \\
\frac{\partial^2 \ell(\beta, \alpha|y)}{\partial \alpha \partial \beta^T} &= D^T \frac{\partial^2 \ell(\beta, \sigma|y)}{\partial \sigma \partial \beta^T}
\end{aligned}$$

The specific entries in the gradient and Hessian for f can be found in equations (A.17) through (A.19).

The theorem follows from the model formulation and the regularity conditions established in the previous section combined with Theorem 3.3.2 of Pinheiro (1994). □

[Received July 15, 1998. Revised .]

REFERENCES

Ash, Robert B., and Gardner, Melvin F. (1975), *Topics in Stochastic Processes,* New York: Academic Press.

Baker, Michael (1997), "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings," *Journal of Labor Economics,* 15, 338-375.

Barry, Daniel (1995), "A Bayesian Model for Growth Curve Analysis," *Biometrics,* 51, 639-655.

Bernhardt, Annette D., Morris, Martina, Handcock, Mark S., and Scott, Marc A. (1997), "Work and Opportunity in the Post-Industrial Labor Market," Report to the Russell Sage Foundation, New York, pp. 84.

Besse, Philippe, Cardot, Hervé, and Ferraty, Frédéric (1997), "Simultaneous Non-parametric Regressions of Unbalanced Longitudinal Data," *Computational Statistics and Data Analysis,* 24, 255-270.

Besse, Philippe, and Ramsay, J. O. (1986), "Principal Components Analysis of Sampled Functions," *Psychometrika,* 51, 285-311.

Brumback, Babette A. (1996), "Statistical Models for Hormone Data," Ph.D Thesis, University of California, Berkeley.

Brumback, Babette A., and Rice, John A. (1998), "Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves," *Journal of the American Statistical Association*, 93, 961-976.

Cappelli, Peter. (1995), "Rethinking Employment," *British Journal of Industrial Relations* 33, 563-602.

Danziger, Sheldon and Gottschalk, Peter (1993), *Uneven Tides: Rising Inequality in America,* New York: Russell Sage Foundation.

——— 1996. *America Unequal.* New York: Russell Sage Foundation.

Davidian, Marie, and Giltinan, David M. (1995), *Nonlinear Models for Repeated Measurement Data,* London: Chapman and Hall.

Diggle, Peter J., Liang, Kung-Yee, and Zeger, Scott L., (1994), *Analysis of Longitudinal Data,* Oxford: Oxford University Press.

Gottschalk, Peter and Moffit, Robert. (1994), "The Growth of Earnings Instability in the US Labor Market," *Brookings Papers on Economic Activity* 2, 217-72.

Grenander, Ulf (1981), *Abstract Inference,* New York: Wiley.

Hadi, Ali S. (1996), *Matrix Algebra as a Tool*, Belmont, CA: Wadsworth, ITP.

Haider, Steven (1996), "Earnings Instability and Earnings Inequality in the United States: 1967-1991," Manuscript, University of Michigan.

Harrison, Bennett. (1994), *Lean and Mean: The Changing Landscape of Corporate Power in the Age of Flexibility.* New York: Basic Books.

Horn, R. A., and Johnson, C. R. (1985), *Matrix Analysis,* New York : Cambridge University Press.

Jiang, Juming (1996), "REML Estimation: Asymptotic Behavior and Related Topics," *Annals of Statistics,* 24, 255-286.

Jones, Richard H. (1990), "Serial Correlation or Random Subject Effects?" *Communications in Statistics Series B*, 19(3), 1105-1123.

Karoly, Lynne A. (1993), "The Trend in Inequality Among Families, Individuals, and Workers in the United States: A Twenty-Five Year Perspective," In *Uneven Tides: Rising Inequality In America*, edited by S. Danziger and P. Gottschalk, 19-97. New York: Russell Sage.

Kneip, Alois (1994), "Nonparametric Estimation of Common Regressors of Similar Curve Data," *Annals of Statistics*, 22, 1386-1427.

Kooperberg, Charles and Stone, Charles J. "Logspline density estimation for censored data," *Journal of Computational and Graphical Statistics,* 1, 301-328.

Laird, Nan M. and Ware, James H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963-974.

Levy, Frank and Murnane, Robert (1992), "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature,* 30, 1333-1381.

Lindstrom, Mary J. (1995), "Self-modelling with Random Shift and Scale Parameters and a Free-knot Spline Shape Function," *Statistics in Medicine,* 14, 2009-2021.

Longford, Nicholas T. (1993), *Random Coefficient Models,* Oxford: Oxford University Press.

Meredith, W., and Tisak, J. (1990), "Latent Curve Analysis," *Psychometrika,* 55, 105-122.

Miller, J. J. (1977). "Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Model Analysis of Variance," *Annals of Statistics,* 5, 746-762.

Pinheiro, José C. (1994), "Topics in Mixed Effects Models," Ph.D Thesis, University of Wisconsin, Madison

Ramsay, J. O. and Dalzell, C. J. (1991), "Some Tools for Functional Data Analy sis," *Journal of the Royal Statistical Society Series B,* 53, 539-572.

Ramsay, J. O. and Silverman, B. W. (1997), *Functional Data Analysis.* New York: Springer.

Rao, C. R. (1958), "Some statistical methods for comparison of growth curves," *Biometrics,* 14, 1-17.

Rice, John A. and Silverman, B. W. (1991), "Estimating the Mean and Covariance Structure Nonparametrically when the Data are Curves," *Journal of the Royal Statistical Society Series B,* 53, 233-243.

Scott, Marc A. (1998), "Statistical Models for Heterogeneity in the Labor Market," Ph.D Thesis, New York University.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992), *Variance Components,* New York: Wiley.

Silverman, B. W. (1996), "Smoothed Functional Principal Components Analysis by Choice of Norm," *Annals of Statistics,* 24, 1-24.

Speed, T. P. (1991), "Comment on 'That BLUP is a Good Thing: The Estimation of Random Effects'," *Statistical Science*, 6, 42-44.

Topel, Robert H., and Ward, Michael P. (1992), "Job Mobility and the Careers of Young Men," *Quarterly Journal of Economics,* 107, 439-479.

Verbeke, Geert and Lesaffre, Emmanuel (1996), "A Linear Mixed-Effects Models with Heterogeneity in the Random-Effects Population," *Journal of American Statistical Association* 91, 217-221.

Vonesh, E. F. and Chinchilli, V. M. (1997), "Linear and Nonlinear Models for the Analysis of Repeated Measurements," New York: Marcel Dekker

Wahba, Grace (1990), *Spline Models for Observational Data,* Philadelphia: Society for Industrial and Applied Mathematics (CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59).

Wang, Yuedong (1996), "Smoothing Spline Models with Correlated Random Errors," *Journal of the American Statistical Association*, 93, 341-348.

Technical Report 966, University of Wisconsin, Madison, Dept. of Statistics.

Xu, Weichun, Hedeker, Donald, and Ramakrishnan, Viswanathan (1996), "Mixtures in Random-effects Regression Models," Manuscript, School of Public Health, University of Illinois at Chicago.