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

Longitudinal Data Models When the Covariance is the Feature of Interest

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

IEE Working Paper 14
August 1999

IMPORTANT: This is a draft copy of an article submitted to the *Sociological Methodology*. Do not circulate, cite or quote this document without the author's permission. Please direct all the 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.

Longitudinal Data Models When Covariance Structure is the Feature of Interest

Marc A. Scott
and
Mark S. Handcock

May 25, 1999

Abstract

We present several approaches to modeling the covariance in longitudinal studies when the covariance itself is the primary focus of the analysis. This is a departure from much of the work on longitudinal data analysis, in which attention is focused solely on the cross-sectional mean and the influence of covariates on the mean. Such analyses are particularly important in policy-related studies, in which the heterogeneity of the population is of interest. We describe a flexible, parsimonious class of models for covariance structure appropriate to such analyses, which we call proto-splines. This class extends the set of available random coefficient models by allowing a degree of uncertainty in the design matrix associated with the random coefficients. One important feature of this class is that it provides a decomposition of the variance into interpretable components. We compare several implementations of this class to a more commonly employed mixed effects model to describe the strengths and limitations of each approach. These alternatives are compared in an application to long-term trends in wage inequality for young workers. The findings provide additional guidance in the process of selecting appropriate models.

1 Introduction

1.1 Motivation

An increasing number of social and behavioral 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 via regression forms of individual-specific or population average models. Longitudinal data require more sophisticated models such as these to account for the dependence of the responses within individuals (Diggle, et al. 1994). In some studies, attention is focused solely on the cross-sectional mean, and the influence of covariates on the mean. As a result the motivation for the treatment of the covariance structure is to improve the model for the mean structure rather than forming realistic representations of the covariance itself. However, the covariance structure is of central importance in many policy-related studies. In these studies, one is trying to understand the structure of the heterogeneity in a population, and in so doing yield insight into social phenomena.

For example, in the field of stratification and mobility, the rising inequality of wage outcomes that began in the 1970s and has persisted into the 1990s has prompted labor market analysts to look at long-term trends in these outcomes. One explanation for the observed 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, permanent variation and short-term, transitory variation. We note that the distinction between long- and short-term trends is about fundamental economic mobility. The component of the observed pattern due to permanent variation is a measure of mobility relative to one's peers, while the transitory variation is a measure of individual economic instability. Since long- and short-term trends in wages have substantive meaning, one can compare the resulting partitions from different economic periods. Bernhardt, et al. (1997) apply a mixed effects variance components model to wage data from two young adult cohorts in the National Longitudinal Survey (NLS) and conclude 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).

In matters with such strong policy implications, proper specification of the underlying mechanism is crucial. Unfortunately, correct specification of covariance structure is hampered by a lack of exploratory methods for longitudinal data. All of this would be less important if the variance components were not sensitive to model misspecification. However, when a model is misspecified, the associated variance components are necessarily incorrect, and any subsequent "partitioning" may be suspect. A by-product of these models are estimates of the individual-specific effects defining the trajectories, and these are highly sensitive to model misspecification as well (Verbeke and Lesaffre 1996).[1]

1.2 Scope of analyses

We anchor our presentation by using an example from labor market economics, where these issues are paramount. We will be investigating two datasets from the NLS. The first, or original cohort, is a representative sample of young men aged 14-21 first interviewed in 1966 and interviewed annually for the next fifteen years (with the exception of 1972, 1974, 1977 and 1979). The second dataset began with a comparable sample of young men in 1979 who have been interviewed yearly since then for fifteen additional years. For comparability between cohorts, we selected only non-Hispanic whites, with resulting sample sizes of 2,614 and 2,373 respectively for the original and recent cohorts. For a detailed description of these datasets and their comparability, see Bernhardt, et al. (1997). According to Topel and Ward (1992), "the first 10 years of a career will account for 66 percent of lifetime wage growth for male high school graduates and almost exactly the same fraction of lifetime job changes," so it is important to understand trends manifesting themselves in this period.

In this paper, we will present several competing models for covariance structure in longitudinal data and describe their distinct features. One of these, proto-splines, is novel in the

[1]Verbeke and Lesaffre (1997) note that the fixed effects are less sensitive to misspecification, but these effects are of less interest in the approach taken here.

literature, so it will be developed in some depth. We begin by describing available techniques for teasing out covariance structure in Section 2. We then discuss a general framework for describing covariance known as latent curve modeling and present the new proto-spline model class in Section 3. Our ultimate goal in the application is to compare the degree of permanent wage trajectory differences from two periods, the 1960s and 1980s. To keep the discussion manageable, however, we focus on the recent cohort in the next two sections. In Section 4, we make comparisons between the competing models and between the two NLS cohorts. We end with a discussion of the substantive meaning of the models, and extensions to other forms.

2 Current Approaches

2.1 Exploratory methods

Exploratory methods should provide a way to visualize a complex set of curves in space and inform model building. Longitudinal studies often involve many subjects, so that a traditional plot superimposing the curves will be hard to evaluate. We illustrate the problem by plotting wage profiles for about 800 young adults from the NLS recent cohort in Figure 1. It may appear that there are fewer than 800 trajectories displayed, but remember that these vary in length from two to sixteen years, so no single individual spans the entire age range. While we can gather some sense of where most of the curves lie and how they fan out over time, it is hard to describe their variation succinctly. For example, the curves do not simply start out higher or lower—the rate of growth appears to vary as well.

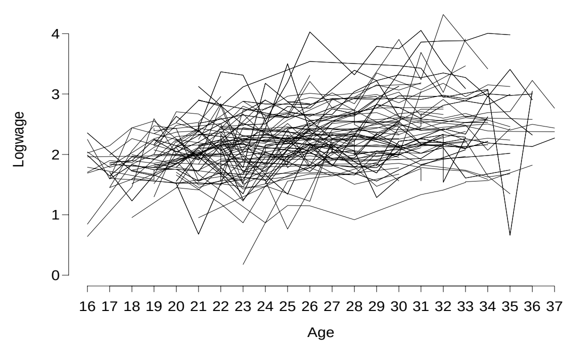

Figure 1: Wage trajectories for approximately 800 young adult workers in the NLS, 1979-1994 (recent) cohort: plot of logwage by age

Some techniques are designed to locate "representative curves" (Diggle et al. 1994) from a set of curves. These include principal components and principal points methods developed by Jones and Rice (1992) and Flury and Tarpey (1993). With these methods, the analyst attempts to identify one or more curves that capture the main patterns of variation in the full set. Representative curves for the above data could include one near the center of the mass, one exhibiting strong growth, and another flatter curve showing little to no growth. The proportion of curves near each of these exemplary ones would give some indication of the overall distribution.[2] Such techniques are an important step in understanding a general aspect of the covariance, such as where most curves lie, but they offer little information about how curves within each cluster vary with respect to the representatives. While the approximate regions of high, moderate and low growth may be identified, how often and by how much individuals exceed or fare worse within these clusters is not captured.

At the exploratory stage, our goal is to describe the variation between the curves. We can glean some global information about that variation without looking at every curve using two tools, which we now describe.

Stacked boxplots, in which the grouping variable is age in the time sequence, can yield some insight, as Figure 2 shows. Again, we witness a fanning out of wages as individuals age and can notice some reduction in variation at age 37. However, this graphic fails to incorporate any longitudinal features of the data, because it is a cross-sectional summary of age-specific characteristics. It essentially points to the next stage of an analysis, which would be to determine how often subjects with higher wages at young ages maintain these over the life course. In other words, if one were to superimpose the individual curves on the stacked boxplot, would they tend to fall strictly above or below the median, and does this shift over time?

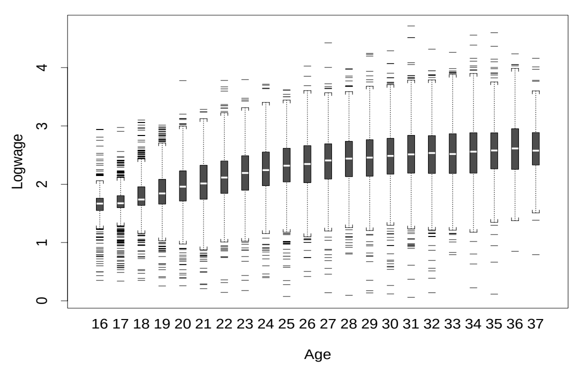

Figure 2: Boxplots summarizing logwage by age for young adult workers in the recent cohort

The variance of logwages increases as an individual ages and so the underlying process is non-stationary. Suppose that a response $Y(t)$ at age t is non-stationary with $\sigma^2(t) = \mathrm{Var}(Y(t))$. If the increments of $Y(t)$ in t form a stationary process then $Y(t)$ is called intrinsically stationary. Under this weaker assumption, the variation is often summarized by the variogram (Diggle, et al. 1994):

$$\gamma(u) = \tfrac{1}{2}\mathrm{Var}\left[Y(t) - Y(t+u)\right],\ u \geq 0, \tag{1}$$

where u measures the distance (in our case, time) between observations. This model specifies that the covariance of $Y(t)$ and $Y(t+u)$ is

$$\mathrm{Cov}(Y(t), Y(t+u)) = \tfrac{1}{2}\sigma^2(t) + \tfrac{1}{2}\sigma^2(t+u) - \gamma(u). \tag{2}$$

A nonparametric estimate of the variogram for the NLS data is displayed in Figure 3.

[2]This partitioning can be viewed as a clustering problem. Clustering longitudinal data, which is often unbalanced, may require some preliminary adjustment to the data and is discussed in Scott (1998).

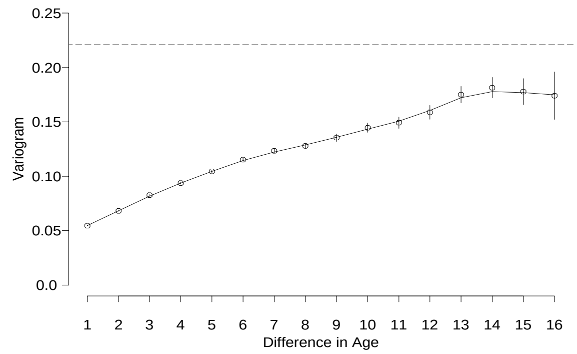

Figure 3: Variogram of mean-reduced logwage for young adult workers in the recent cohort

We placed a horizontal line at the level of the weighted average of the age-specific process variances, so that the scale relevant to a stationary process is apparent. We then computed confidence intervals for the variogram point estimates following Cressie (1991),[3] and we fit a smooth curve to those points.

The shape of the variogram suggests that a fair amount of structure remains in this data after removing a mean process. Had this not been the case, the variogram would have resembled a horizontal line. As the gap between observations widens, the variogram increases, indicating a decrease in correlation between observed wages. The fact that even at the widest gaps there is still some positive correlation may indicate that long-term stratification in wages does persist. In other words, for a single individual initial and final wages are often both larger or both smaller, describing a persistent stratification between individuals over the life course. We will explore aspects of this hypothesis in later sections. Finally, we note that the variogram is often used to construct covariance models directly. That is, parametric models for the variogram are often assumed and estimated from the data. This is appropriate in other contexts, but when population heterogeneity is of key interest, these models are less substantively interpretable than the models presented below. We do not pursue them here—see Jones 1990 for a detailed discussion.

2.2 Covariance models

In order to go beyond the above descriptions of the covariation, we introduce models that incorporate covariance structure directly in their formulation. Broadly speaking, there have been two approaches to modeling the heterogeneity in longitudinal data. The first allows the covariance structure to be specified nonparametrically. The methods of Davidan and Gallant (1993), Diggle and Verbyla (1998) and Gasser, et al. (1998) describe the multivariate density of the response process. As exploratory methods, they should expose interesting structure, but their nonparametric nature limits one's ability to interpret the results, especially in the context of variance components. The second approach is to propose a parametric model for the process,

[3]We note that simultaneous confidence bands would be more informative, since we are effectively estimating a whole curve, but we do not explore these in this paper.

and then use the parameter estimates to describe the variation. The simplest model upon which many others are built is the random coefficient (or mixed effects) model (Longford 1993). These assume there is a common mean structure, called the fixed effects, and then provide latitude for the analyst to specify the systematic variation about the mean (these are called the random effects).

For example, a random coefficient model may postulate that wage growth is reasonably approximated by a random quadratic curve in time, as follows.

$$Y(t) = X(t)\beta + Z(t)\delta + \epsilon(t) \tag{3}$$

In this model, $Y(t)$ is the response, the vector $X(t)$ captures the structure of the fixed effects and the vector $Z(t)$ does the same for the random effects, while $\epsilon(t)$ is a residual process. For quadratic growth, $X(t)$ and $Z(t)$ would have a component for constant, linear and quadratic polynomials in time. The vector β consists of coefficients corresponding to the columns of $X(t)$ and δ is a set of random coefficients, one for each column of $Z(t)$. Note that the above formulation for the response $Y(t)$ is that of a stochastic process. The δ are often assumed to be multivariate Gaussian, as are the $\epsilon(t)$. Continuing with our quadratic growth example, δ may be modeled as trivariate Gaussian, with the six parameters of the covariance matrix left unrestricted.[4] Any systematic variation is captured in these parameters; for example, the constant and linear terms may be negatively correlated, so that high initial responses show less growth over time, while smaller initial responses grow more rapidly. We will fit such a model in Section 4.1.

Our goal is to describe interesting trends in covariance, but these models typically summarize this structure in general terms. With random quadratics, for example, the extent of divergence from the mean structure is of interest, substantively, but no single covariance parameter summarizes this extent—all the variance components must be interpreted simultaneously. Thus, we may be able to describe the density of random quadratics, but the result may be difficult to interpret.

3 Latent curve models

3.1 General description

Meredith and Tisak (1990) propose that covariance structure can be decomposed into meaningful components known as latent curves. In this formulation,

$$Y_i(t) = \mu(t) + \sum_k \omega_{ik}\phi_k(t) + \epsilon_i(t), \tag{4}$$

where $Y_i(t)$ represents the response for the i^{th} individual, $\mu(t)$ is a mean function, ω_{ik} is the individual-specific coefficient associated with the k^{th} latent curve $\phi_k(t)$ and $\epsilon_i(t)$ is the residual process. If the ϕ_k are known, then the above is similar to ordinary regression, with ϕ_k functioning as the columns of the design matrix $X(t)$ ($\mu(t)$ can be incorporated into the fixed design as well). However, subjects have their own regression coefficients; removing the subscript i from all terms

[4] This random quadratic model, with six parameters, is relatively complex. Simpler models, such as an AR(1) process may fit the data, but do not have a straightforward variance components interpretation, so they are not as useful in this context (see Jones 1990 for a discussion).

and evaluating the expression at finitely many time points will reduce the above to ordinary regression. We suppress explicit reference to the cross-sectional mean $\mu(t)$ in most of what follows.

When this model was originally presented, it was assumed that the latent curves would be estimated directly from the data. With a few additional assumptions, this is a factor analysis, which is a particular decomposition of the covariance structure. The large variability inherent in such covariance estimates prompted researchers to impose smoothness constraints on the curves (Rice and Silverman 1991). The factor analytic decomposition can be justified on theoretical grounds using the Karhunen-Loève expansion of an L_2 stochastic process.

3.2 A stochastic process representation

The Karhunen-Loève expansion represents a mean-zero stochastic process $Y(t)$ as a stochastically weighted (potentially infinite) sum of basis functions as follows.

$$Y(t) = \sum_k \omega_k \phi_k(t) \tag{5}$$

The $\phi_k(t)$ are taken to be orthonormal, and the ω_k are uncorrelated random variables. If we require the ω_{ik} be stochastic in the latent class model (4), then it is a finite approximation of (5) with additional mean and error terms. The important aspect of model building for stochastic processes is thus the specification of the latent curves ϕ_k and the distribution of the stochastic weights ω_{ik}.

The analyst must consider these specifications simultaneously, as the following illustrates. The Karhunen-Loève expansion requires that the stochastic weights be uncorrelated, but not necessarily independent. If the weights are independent, then the data generating mechanism specifies that subjects receive a series of random shocks, each one taking the form of a latent curve. These shocks are independent, so receiving a large shock from the first curve does not mean one will receive a large shock from the next. This may or may not be appropriate for the substantive problem, depending on the shape of the shocks. In the social and behavioral sciences, it is often assumed that various influences emanate from common but unobservable factors, sometimes referred to as latent traits. In this case, dependent shocks seem more reasonable, since we would expect certain perturbations to occur together. If the shocks are constrained to be orthogonal to each other, independent weights are more tenable.

3.3 The proto-spline model class

Scott (1998) and Scott and Handcock (1998) introduced the proto- spline class of heterogeneity models to address these and other concerns. Motivated by a longitudinal study of wage growth, the authors formulated a class of models that capture long- and short-term features of the covariance structure. These components are of substantive interest in labor market economics. The models use a latent curve formulation to identify long-term patterns of variation, which yields a meaningful variance components decomposition. The proto-spline class is distinguished by the data-adaptive manner in which the curves are estimated. We will now describe this class in detail.

Proto-splines incorporate some of the smoothness features of nonparametric models but remain estimable as a parametric form of covariance. The approach also manages to satisfy a between-curve orthogonality condition without imposing external constraints on the estimation. The way we incorporate these features keeps the model parsimonious, interpretable, and easy to estimate.

The proto-splines class is derived from the model class of Meredith and Tisak (1990),

$$Y(t) = \mu(t) + \sum_k \omega_k \phi_k(t) + \epsilon(t). \tag{6}$$

We have dropped the individual subscript i to emphasize that the ω_k are viewed as random coefficients. This model class is still too general to be estimated. We need to assume a distributional form for the random coefficients and we need to specify the form of the functions $\phi_k(t)$ somehow. For several reasons, including interpretability, we currently restrict these functions to be orthonormal. Using the Karhunen-Loève expansion, our model is a finite approximation of an L_2 stochastic process with the addition of mean and error terms.

We create a link between members of this class and nonparametric models in the way that we construct the latent curves. Suppose for the moment that we believe that most of the stochastic variation can be described by one curve, ϕ_1 (this is a single latent curve model). What remains is the choice of a class of functions $\mathcal{H}$ in which our estimate $\hat{\phi}_1$ will reside. One choice for $\mathcal{H}$ is the smooth set of functions known as the natural cubic splines. Splines are defined in terms of T ordered knots $t_1, \ldots, t_T$ on an interval $\mathcal{T} = [a, b]$, where $a < t_1 < \ldots < t_T < b$. A *cubic spline* is a cubic polynomial on all intervals $(a, t_1), (t_1, t_2), \ldots, (t_T, b)$, it is everywhere continuous, and it has continuous first and second derivatives on the entire interval $\mathcal{T}$. *Natural cubic* splines have second and third derivatives equal to zero at endpoints a and b (Green and Silverman 1994). The natural cubic splines are of dimension T, which means that specifying the value of such a spline at its knots describes it completely. Note that $\phi_k(t)$, while specified and estimated at T time points, has a functional interpretation in that there is a corresponding curve defined at all points in the interval $\mathcal{T}$. Features of this function, such as its derivative, are thus available and could yield insight into the observed process.

Let $\psi_1(t), \ldots, \psi_T(t)$ be an orthogonal basis for the cubic splines observed at the T knots. Then $\phi_1 \in \mathcal{H}$ has representation

$$\phi_1(t) = \sum_{j=1}^{T} \eta_j \psi_j(t), \tag{7}$$

where the η_j are *non-random* parameters that define the curve with respect to the basis. Extending this to the response variable, for the full model, we have

$$\begin{aligned} Y(t) &= \mu(t) + \omega_1 \phi_1(t) + \epsilon(t) & (8) \\ &= \mu(t) + \omega_1 \sum_{j=1}^{T} \eta_j \psi_j(t) + \epsilon(t), & (9) \end{aligned}$$

where ω_1 is a mean zero random coefficient with variance one and $\epsilon(t)$ are i.i.d. Gaussian random variables with variance σ_ϵ^2.[5]

We require T parameters to specify *one* smooth curve. Our model has two variance components, the variance of ω_1 and the residual variance, σ_ϵ^2, and it is a *parametric* covariance model

[5]Note that the magnitude of the variation associated with ϕ_1 is contained in its norm, and the error structure $\epsilon(t)$ can be made more complex.

with uncertainty in the basis function $\phi_1(t)$. This uncertainty is a distinguishing feature of proto-spline models. They are effectively random coefficient models of the linear mixed effects variety, in which the random effects design matrix is constrained to lie in a function space, but the choice of functional element is estimated from the data.[6] In Scott and Handcock (1998) we show that a proto-spline model has the equivalent likelihood of a specific mixed effects model, and we use this fact to do efficient estimation and inference.

3.4 A comparison of proto-spline and mixed effects models

The distinction between proto-spline models and other mixed effects models can be made at this point using the single proto-spline model. The standard mixed effects model has the following form:

$$Y(t) = X(t)\beta + Z(t)\delta + \epsilon(t). \tag{10}$$

A key feature of this model is that X(t) and Z(t) are *known* designs. The single latent curve proto-spline model is precisely the above model, with $Z(t) = \phi_1(t)$ and $\delta = \omega_1$, only $Z(t)$ is *unknown*. To illustrate the conceptual difference, we will consider three random quadratic models. For Model I, we assume that we know the exact quadratic curve that describes the structured covariation about the mean. Let $Z(t) = t + \frac{1}{2}t^2$, so $Z(t)$ is a scalar-valued function describing a *particular* growth structure. Further, let the random effects, δ, be a Gaussian random variable with unknown variance (the variance is one of the model's variance components). For a specific individual,

$$Y_i(t) = X(t)\beta + (t + \frac{1}{2}t^2)\delta_i + \epsilon_i(t). \tag{11}$$

Every subject gets some random multiple of the fixed curve $t + \frac{1}{2}t^2$.

For Model II we consider a mixed effects model in which each individual has their own quadratic perturbation as follows. Let the three columns of $Z(t)$ be given by the vector $(1, t, t^2)$, and let $\delta = (\delta_1, \delta_2, \delta_3)$ be a vector of random coefficients, with a multivariate Gaussian distribution. Then for an individual specific curve,

$$Y_i(t) = X(t)\beta + \delta_{1i} + \delta_{2i}t + \delta_{3i}t^2 + \epsilon_i(t). \tag{12}$$

While this is quite flexible, the variance components analysis requires a full description of the estimated covariance structure of the random effects, which are contained in a 3×3 matrix that includes important covariance as well as variance components. We must use all of this information when describing any variance partitioning.

Model I is highly inflexible in that we must impose an exact form for growth beyond the mean. However, the variance component for δ_i is highly interpretable—it is the variance of the coefficient of precisely determined shocks to the system, so a larger variance means there is greater dispersion in the growth curves, and that they all follow that form. It would be difficult to make a similar statement about Model II.

For Model III we consider a single latent curve proto-spline model, which offers the interpretability of the simpler model (I), and the flexibility of the more complex model (II). Let $\phi_1(t) = \eta_1\psi_1(t) + \eta_2\psi_2(t) + \eta_3\psi_3(t)$, where $\psi_1(t) = 1$, $\psi_2(t) = t$ and $\psi_3(t) = t^2$. While this might

[6] Strictly speaking, this is not a mixed effects model, since the design is unknown.

resemble model II, the vector (η_1, η_2, η_3) is common to each individual and does not represent individual-specific random effects. Every individual curve has the following form:

$$Y_i(t) = X(t)\beta + \delta_i(\eta_1 + \eta_2 t + \eta_3 t^2) + \epsilon_i(t), \tag{13}$$

with the parameters (η_1, η_2, η_3) fixed and identical for each individual, a reparameterization (9) that keeps the notation consistent. Each of these models is different, and we claim that the proto-splines offer an effective compromise between the rigidity and flexibility of Models I and II, while remaining highly interpretable from a variance components perspective.

3.5 Estimation of multiple curve proto-splines

In this section we extend the development of the single proto-spline model in (7)-(9) to the general multiple proto-spline model. We note that the choice of our function space implies that ϕ_1 has a nonparametric interpretation, since the curve lies in a (smooth, continuous) function space. Note that the model is not restricted to the space of cubic spline functions. Any finite-dimensional function space (or vector space) could be employed, but the smoothness feature of this particular space and its links to nonparametric models was relevant to the development. An advantage to the proto-spline class of models is that the bases may be chosen to reflect the form expected in the substantive process without knowing which specific version of that form is present.[7] If we choose models that result in latent curve estimates with a functional interpretation, features such as the derivative become available.

The formulation of proto-spline models satisfies the orthogonality conditions without imposing external constraints on the estimation and limits the number of parameters required, keeping the model parsimonious. Proto-splines exist in the space between parametric and nonparametric models, sharing certain features of each. They also yield highly interpretable latent curves.

We define the full proto-spline class by extending the above example to more than one curve without introducing additional parameters. The main idea is to use only a subset of the T bases ψ_j to construct each curve ϕ_ν. Let $\mathcal{I}_\nu$ be an indexing function defined on the integers $1, \ldots T$, which selects the basis functions used to construct the ν^{th} curve. In our simple example, ϕ_1 uses all T basis functions, so $\mathcal{I}_1 = \{1, \ldots, T\}$. We construct latent curves as a deterministically weighted sum of the basis functions specified by the indexing function $\mathcal{I}$ so

$$\phi_\nu(t) = \sum_{j \in \mathcal{I}_\nu} \eta_j \psi_j(t) \tag{14}$$

and then

$$Y(t) = \mu(t) + \sum_{\nu=1}^{K} \omega_\nu \phi_\nu(t) + \epsilon(t) \tag{15}$$

as before. In order to insure orthogonality of the ϕ_ν, the index sets given by $\mathcal{I}_\nu$ must be disjoint. This restriction implies that once we decide to estimate more than one latent curve, the curves are no longer from the class $\mathcal{H}$. Since $\mathcal{H}$ was originally chosen to be the natural cubic splines, we named the resulting ϕ_ν *proto-splines*, because they are *partial* versions of a full spline fit. This

[7] An unknown quadratic curve was expected in Model III. More complex function spaces, such as those including jump processes and wavelet bases can be used to adapt the model to the substantive questions of interest.

method requires T parameters to build all K curves; if we do not normalize the curves, then for identifiability the random coefficients ω_ν are all presumed to have variance one.

To place this model in the context of those previously developed, we examine it for two extreme cases. First, if $K = T$, then each proto-spline is just a rescaled version of the basis function. This is essentially the model proposed in Brumback (1996) and Brumback and Rice (1998), although the form of their model was chosen to produce cubic spline *predictions* for individual curves. If $K = 1$, then we are estimating a single principal function (a smooth principal component) in the presence of noise, and it is a natural cubic spline.[8]

A more useful approach is to choose K to be small in relation to T, so that for equal-sized index sets, T/K bases are available for each latent curve. Equations (14) and (15) still apply, but the "proto-spline" nature of the curve estimates becomes more apparent. This intermediate case is similar to 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 are enforcing a small number of these by our choice of K, and we maintain the orthogonality requirement by construction. The limitations of our approach depend on the unknown curves—if they are well-approximated using a few distinct basis functions each (and we can determine the proper set of bases), then our method should work well. Note that this model differs from a principal functions analysis in that we choose our function spaces with substantive features in mind, rather than simple smoothness constraints. We then build our model directly around these structures.

3.6 Application to Wage Growth

For ease of exposition, we illustrate our model class by fitting a single latent quadratic curve proto-spline model to longitudinal wage data from the NLS. For the fixed effects, $X(t)$, we use a simple quadratic in age; this yields Model III of Section 3.4. In Figure 4, we display the cross-sectional mean of the process. It provides the center from which the curves deviate. In Figure 5, we superimpose 4 simulated realizations from the proto-spline model fit, with the residual process $\epsilon(t)$ suppressed. The fitted curve $\hat{\phi}_1$ used in that simulation is presented in Figure 6. From this figure, one can see that the growth of wages near the college years of 18 to 22 sets the extent of growth for the later years as well. The shape of the single latent curve describes the long-term trend in variation—strong growth in the 20s, followed by steady but diminished growth in the 30s. Each realization is simply the mean curve plus some random multiple (positive or negative) of the latent curve $\hat{\phi}_1$.

One might be concerned that imposing a quadratic latent curve is overly restrictive and essentially forces the decomposition into the shape indicated above. However, within the class of quadratic curves there are pure linear and constant curves, so if there were no change in the *growth rate* at early and later ages, then we would expect a different fitted latent curve. By forming a single latent curve spanning all ages, we are specifying that we want this curve to represent long-term structure, within the quadratic class. More complex spaces may reveal more complicated dependencies, should they exist, and they should be considered in the model selection process.

In Figure 5, we see that the effect of the single latent curve crosses the mean at age 17. The near-zero value for $\hat{\phi}_1$ near this age corresponds to a negligible amount of permanent variance, but even this is not predetermined by our choice of basis. If below-mean wages at those ages

[8] Principal functions are defined and discussed at length in Ramsay and Silverman (1997).

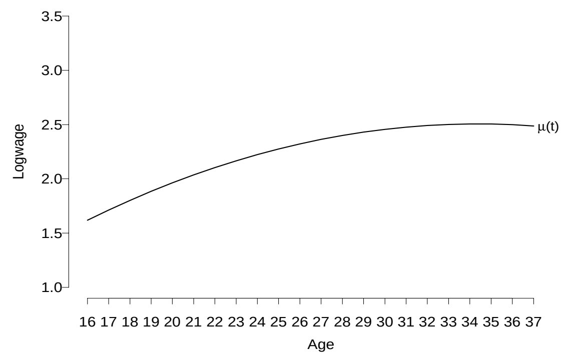

Figure 4: Mean curve for single proto-spline model

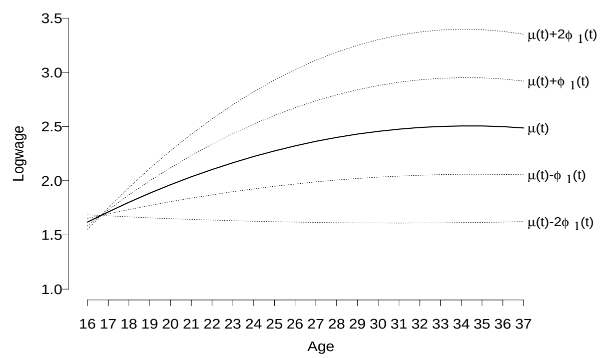

Figure 5: Curves for random coefficients one and two standard deviation from the mean for single proto-spline model

were to lead to much larger gains later on, the "crossover" would be at some later age. Random quadratics do limit us to a single change in the direction of growth (positive or negative), while higher order polynomials would not. Finally, note an important difference between this model and Model II. In Model II, each individual has a uniquely shaped quadratic curve, so it may rise quickly and not level off, or it may level off quickly. In our model, which is basically Model III, every individual's variation beyond the mean has the same shape, given by $\hat{\phi}_1$—only the magnitude of that variation is allowed to vary.

4 Comparisons of models

To illustrate how the models differ in practice, we apply several different covariance models to labor market data from the NLS. After a preliminary analysis, we found that the mean structure in this data resembles a quadratic curve, so we set the columns of the fixed effects design matrix to

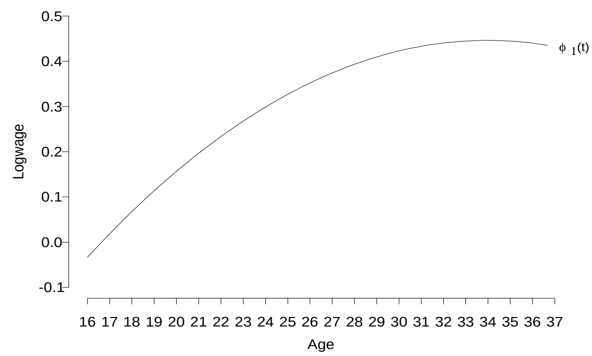

Figure 6: Fitted latent curve for single proto-spline model

correspond to constant, linear and quadratic growth over time.[9] More complex mean structures could describe the influence of additional covariates on aggregate wage growth; our goal in this study is to understand the degree of long-term wage stratification, so the overall divergence of these curves over time in comparable samples yields important substantive information. Next, we must select a form for the structured portion of the variation. For wage data, the structured portion consists of long-term, or permanent, differences between wage trajectories.

We will compare three models. The first is a random quadratic mixed effects model similar to Model II and to that used by Bernhardt, et al. (1997) in their analyses. The second is a single latent curve proto-spline model, similar to Model III, and the third is an extension of proto-spline models that includes a second non-orthogonal latent curve. Beyond the structural variation just described, the residual variation is modeled simply as independent error with constant variance.

4.1 Random Quadratics

The strength of a random quadratic model, such as that given by (12) is the flexibility provided by the three random coefficients, δ_{1i}, δ_{2i} and δ_{3i}. Note that the quadratic basis we use is an orthogonalized and normalized version of $(1, t, t^2)$, which is also the fixed effects basis. These coefficients are globally constrained to come from a multivariate Gaussian density. This choice yields a broad range of curves of various shapes and intensities. This distributional form does, however, require that there are no clusters of curves, or other multimodalities.

We assume that the δ are distributed as $N(0, G)$, where G is a completely unspecified 3×3 covariance matrix defined by six distinct parameters. We fit the model using maximum likelihood estimation, and find that

$$\hat{G} = \begin{pmatrix} +2.019 & +0.899 & -0.265 \\ +0.899 & +1.312 & +0.096 \\ -0.265 & +0.096 & +0.529 \end{pmatrix}$$

[9] Note that we orthogonalize and normalize these basis vectors, so instead of a vector of ones, the first column is a vector consisting entirely of $1/\sqrt{22}$, based on the 22 ages from 16 to 37.

and $\hat{\sigma}_\epsilon^2 = 0.0719$.[10] Unfortunately, these results are somewhat hard to interpret. The structured portion of the covariance is given by $Z\hat{G}Z^T$, where Z is the random effects design matrix. Since the rows of Z correspond to the subject's age, this matrix product describes individual wage differences at each age and how they relate to each other. For example, the diagonal of $Z\hat{G}Z^T$ represents the structured, or permanent, wage variance at each age. These values are plotted against age in Figure 7 below. The initially larger variance at the earliest ages indicates some

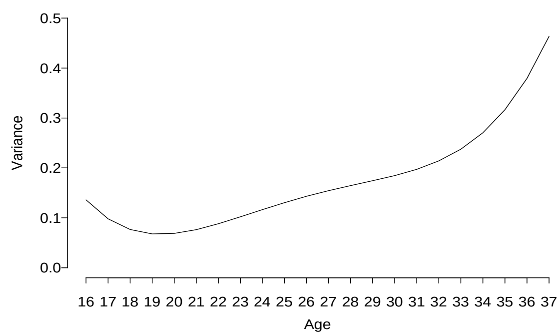

Figure 7: Permanent wage variance for random quadratic model

initial stratification between individual trajectories that seems to diminish by age 20, only to increase substantially from that point forward, with a dramatic rise after age 32. Had permanent differences in trajectories been limited to an intercept shift, this graph would have consisted of a horizontal line some distance above the axis. The result above indicates that wages fan out quite dramatically as individuals age, and gives some indication of how the process accelerates. We can infer that the trajectory fans out as a whole because the partition is based on a model employing a continuous curve for the "signal" portion of the trajectory.

4.2 Single latent curve proto-spline

In this model, we assume that most of the structured variation takes a specific form, but we let the exact shape be determined by the data. The explicit model is

$$Y_i(t) = X(t)\beta + \delta_i\phi_1(t) + \epsilon_i(t), \tag{16}$$

where ϕ_1 is the single latent curve, assumed to lie in the space of quadratic curves, and δ_i is the random coefficient for the i^{th} individual. This is the same model as the one used for our illustrative example in Section 3.6. A look at Figure 5 (previous page) reveals the strength of this model. A wide range of outcomes are easily represented by the mean plus a random multiple of the single latent curve, $\hat{\phi}_1$. Figure 6 (previous page) displays this curve for the recent cohort.

In this figure, the interpretability of these longitudinal data models becomes apparent. The single latent curve reveals most of what we need to know about structured variation. Contrast this to the covariance matrix $\hat{G}$, which along with design matrix Z provides the equivalent

[10] Differences in these findings and those presented by Bernhardt, et al. (1997) are due to a different choice for the quadratic basis, but are otherwise comparable.

information in a less accessible form. The random coefficient on our proto-spline model is standard Gaussian, so we have an immediate sense of the range of impact of the single latent curve.

The restriction to a single latent curve does limit our ability to model more complex structured variation. In Figure 8 below, we see that the permanent variation, the squared version of $\hat{\phi}_1$, describes a very simple growth structure.[11] Two features stand out in comparison to random quadratic models: the permanent variation starts out lower at the youngest ages and it does not grow as dramatically as individuals age. We believe that the initial variation is less important from a likelihood perspective, so it is effectively being ignored in the estimation process. Had we used a higher order polynomial, we might have discovered persistent initial wage differences. If this were the case, we would expect $\hat{\phi}_1$ to begin higher, possibly decrease somewhat and then increase again, in a shape similar to the permanent variance graph from the random quadratic model.

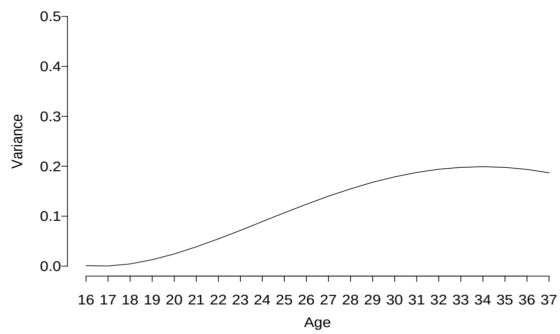

Figure 8: Permanent wage variance for single latent curve model

4.3 Double latent curve model

The limitations of a single curve model prompts us to explore a model with two latent curves. Fitting such a model under the pure proto-spline formulation would require the fitted curves to be orthogonal, and this restricts the function spaces in which each may lie. We propose a new model that effectively "reuses" the basis for each latent curve. The model, abstractly, is given by

$$Y_i(t) = X(t)\beta + \delta_{1i}\phi_1(t) + \delta_{2i}\phi_2(t) + \epsilon_i(t), \tag{17}$$

where ϕ_1 and ϕ_2 are latent curves and δ_{1i} and δ_{2i} are i.i.d. standard Gaussian random coefficients. We construct each curve from the same basis.

$$\phi_\nu(t) = \sum_{j=1}^{T} \eta_{\nu j}\psi_j(t), \tag{18}$$

[11] The permanent variance calculation is straightforward in this case because the random coefficient δ_i is standard Gaussian.

with $\nu = 1$ or 2, effectively doubling the number of parameters used by the single latent curve model. After adding some identifiability constraints to our estimation procedure, we are able to fit this more complex model.

The double latent curve model can best be understood as the combination of a common mean process and two independent "shocks" taking some functional form. We choose to continue to employ the space of quadratic polynomials for ease of exposition. Looking at Figure 9, we find that the fitted curves are quite different from each other. These are the forms for the two shocks, $\hat{\phi}_1$ and $\hat{\phi}_2$. We see that $\hat{\phi}_1$ is quite similar to its counterpart in the single latent curve model, although it starts out further below the origin. The latter feature will induce greater permanent variation at the youngest ages, and then this will subside, as the curve crosses the origin between ages 18 and 19 (contrast this to the crossing at age 17 in the single curve model). The second curve introduces a whole new feature to the covariation. It appears that individuals who start out earning more are penalized as they age. This is indicated by $\hat{\phi}_2$'s initially positive level of about 0.2 at age 16, which sinks to -0.3 by age 37. Of course, negative random coefficients are just as likely as positive ones, so this curve could also represent later growth for young workers who initially accept lower wages. There is mild evidence that this is capturing an "education effect," in which individuals who defer fully entering the labor market (and possibly pursue education or training) benefit with larger wage growth in the long run.[12]

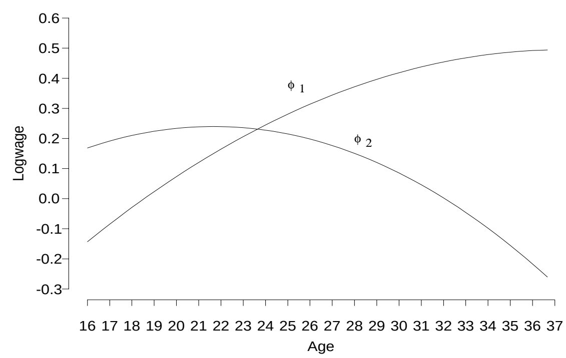

Figure 9: Fitted proto-splines for double latent curve model

The permanent variance partitioning for this model is given in Figure 10, below. By including two curves additively and independently, this model allows for larger early and later year variation. The effects are permanent, in that they persist over the lifetime of a worker, but their independence points to a subtlety of these variance decompositions. Two curves, along with their coefficients describe the systematic portion of a trajectory, but the independence of the coefficients severs any link between the two. In terms of generating mechanisms, this only makes sense if two different features of the wage growth process are being captured, such as an overall growth (often attributed to returns to job tenure and experience), and an education effect.

The above comment points to a limitation of the random quadratic model. Namely, it is hard to describe an underlying process (often thought of as a latent characteristic) that is driving the

[12]This effect was based on an analysis of the final level of education attained by each individual and the predicted value of the coefficient $\hat{\delta}_{2i}$ of $\hat{\phi}_2$.

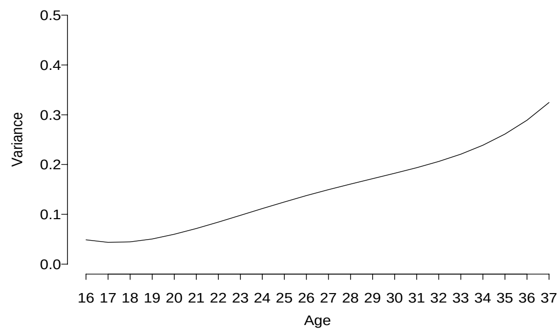

Figure 10: Permanent wage variance for double latent curve model

three coefficients forming the curves. It is hard to imagine that a social or economic generating mechanism involves intercept, slope and acceleration components. The latent curve models provide simpler explanations, which is an advantage in this case.

4.4 Comparing variance partitions

While related, these three models provide different variance decompositions. We display the permanent variation plots in Figure 11, below, and include 95% confidence intervals at each age. We discuss the construction of those intervals in Section A.1 of the appendix. Notable differences exist for the youngest and oldest ages, with strong agreement in the middle range. The single latent curve model does not pick up structured wage variation at the youngest ages. If initial differences in wages persist during the youngest ages, but then diminish, then this model will have to choose between the initial and later year effects, and since the latter are larger, they tend to dominate. The double proto-spline model picks up this extra variation in $\hat{\phi}_2$, and this is reflected in larger permanent variance for the younger ages. The random quadratic model picks up more variation in both younger and older ages and labels it permanent. We contend that the additional flexibility of the random quadratic model allows it to follow the raw data more closely, capturing less rigid forms of variation. This is indirectly confirmed by examining the residual variation, which is 0.072 for random quadratics, and 0.078 and 0.098 for double and single latent curve models, respectively.

Below we present the variance decomposition for each cohort to address an important question. While each model partitions the variance differently, do these differences have substantive impact? That is, how sensitive are the answers to the substantive questions to the choice of model? Even though any model we use will only be an approximation, the social phenomena may manifest themselves in such a way that their detection and measurement is fairly insensitive to the choice of model.

In Figures 12 through 14 below, we display the model-based permanent variance for each model along with 95% confidence intervals at each age. All of the models indicate a significantly larger permanent variance in the recent cohort, starting sometime in the mid-twenties. The difference is most dramatic in the random quadratic model and least so in the single latent curve model. There is some between-model discrepancy in what portion of the variance is

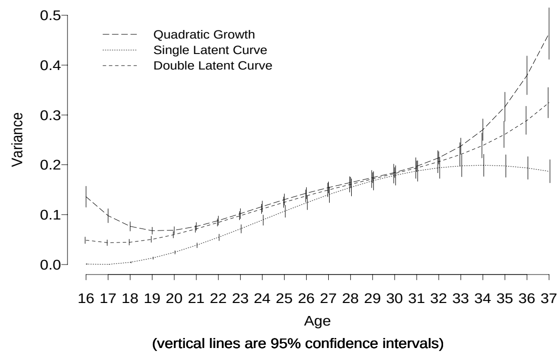

Figure 11: Permanent wage variance for all three models

permanent at the youngest ages, and in how the cohorts differ. Both latent curve models describe a crossover, in which the original cohort starts out more stratified until the early twenties, at which point the opposite it true. The random quadratic model posits that both cohorts are more permanently stratified initially and that the magnitude is comparable. If we were interested in the absolute magnitude of permanent wage stratification, we would look more closely at these models and determine which is more justified on substantive grounds. If we were concerned about wage stratification at the younger ages, then these models tell different stories, so a deeper understanding of each model's characteristics is warranted.

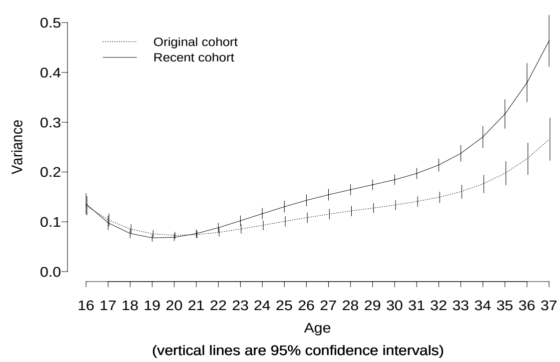

Figure 12: Permanent wage variance for random quadratic model

4.5 Discussion of findings

All three of these models found a significant increase in permanent wage variation in the recent cohort for the older ages. But the magnitude of these differences varied greatly between models, and strong differences in comparative partitions exist at the younger ages.

Since the random quadratic model labeled more variation as permanent, it may be overfitting, in some sense. The flexibility of random quadratics admits even a U-shaped curve (for some

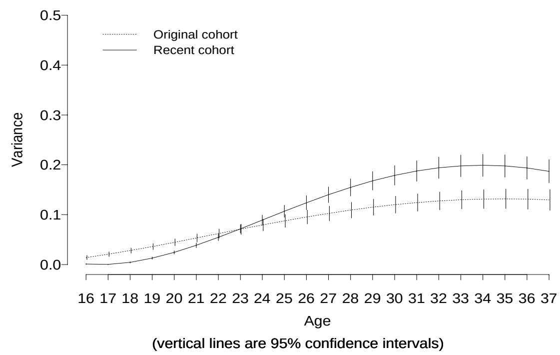

Figure 13: Permanent wage variance for single latent curve model

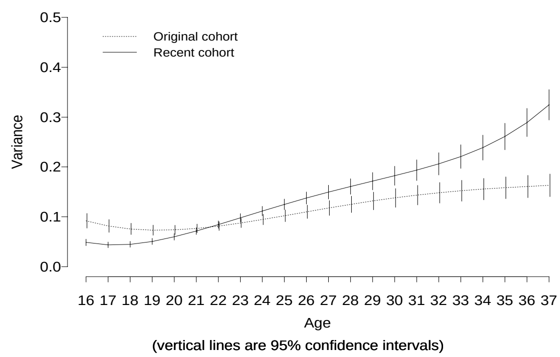

Figure 14: Permanent wage variance for double latent curve model

individuals and not others), but is it desirable to use such a shape to describe *permanent* wage gains? U-shaped curves, in which initial and final wages are nearly identical, involve a shift in the direction of wage change, from loss to gain, so in what sense is this indicative of a permanent, or lasting, trend? We must understand how the choice of model is reflected in the variance partition if we intend to make an informed assessment of social phenomena.

Latent curve models stand out as conservative approaches to variance partitioning. They are highly interpretable, with independent components acting as shocks to the process. The shocks may be readily interpretable in the context of the generating mechanisms for the social processes under study. They offer a handy form of rigidity compared to random quadratic models, yet are inherently adaptive to overall patterns of structured variation. Thus, these models can be viewed as excellent foils to the classical random quadratic model.

All three of the models describe the structured portion of variance in such a way that "permanent" is a reasonable label to apply. That is, the model describes smooth versions of the curves in space that are reasonable attempts to separate signal from noise, and the signal is non-stochastic, conditional on the parameters that describe it. The differences in these definitions allow different aspects of variation to be identified.

5 Conclusion

5.1 Summary

We embarked on this analysis to determine how different models for covariance affect variance component partitioning. Along the way, we introduced a new class of latent curve models, proto-splines, that offer an interpretable paradigm for describing covariation, which is well-suited to formulating substantive questions directly. These models locate persistent covariance structure and reflect it in the shape and size of latent curves. We view proto-splines as covariance function smoothers; they are non-parametric in the sense that the estimated curve lies in a function space, yet the model formulation provides a straightforward interpretation of the curves that is often missing in other non-parametric techniques. In the model formulation, the researcher imposes a class of functions to capture substantively meaningful structure. The restriction to a particular class of functions forces proto-spline models to be conservative in the way they fit the data, making them much less susceptible to "following noise." This makes them invaluable in comparisons with more traditional models; the ways in which they differ point out characteristics of each, with the clearly defined behavior of our models acting as a foil for the others.

5.2 Future work

In future work, we will consider relaxing the independence assumption for the proto-spline model class. For example, our double latent curve model could include a term for the correlation between curves. This extension would open up the possibility of very different latent curves, since the independence constraint ultimately lowers the likelihood of certain shapes for the fitted curves. Including more complex residual structures, such as age-specific variances, as a check on the homogeneous variance assumption could prove useful.

Relaxing the Gaussianality assumption is worth investigating, but we would limit this to forms that remain interpretable, such as parametric forms. One approach that has been suggested by several researchers is a latent class, or mixture formulation (see Clogg 1995, Banfield and Raftery 1993, Muthén and Shedden 1998, Xu, et al. 1996). Under this paradigm, individuals belong to *one* latent class, and then conditional on class membership they follow a certain structure. This remaining structure could be flexibly captured in the proto-spline models just introduced.

In work in progress, we are examining diagnostics for these models in greater detail. Model selection criteria such as AIC (Akaike 1974) and BIC (Schwarz 1978) can be applied here. These are discussed in Vonesh and Chinchilli (1998) and Pinheiro et al. (1994). Recent extensions to the AIC discussed in Simonoff and Tsai (1999) appear to be especially promising in the context of these variance component models. An alternative to model selection is the use of Bayesian model averaging (Hoeting et al. 1998). A developed set of diagnostic techniques will add to our understanding of how each model captures and partitions variation.

A Appendix

A.1 Construction of confidence intervals

Confidence intervals for the model-based variances, such as the permanent variation, are constructed from the asymptotic covariance matrix of the model parameters. For proto-spline models, explicit forms for these covariance matrices are given in Scott and Handcock (1998). The construction begins by finding the variance associated with each point on the latent curve. Each curve is a linear combination of a set of basis functions, with the coefficients specified by the model parameters. If these coefficients are given by column vector $\ell = (\eta_{\nu 1}, \ldots, \eta_{\nu_T})^T$, the basis functions by matrix $Z = [\psi_1, \ldots, \psi_T]$ and the asymptotic covariance matrix of ℓ by H, then the resulting latent curve is given by $\phi_\nu = Z\ell$ and the covariance of $Z\ell$ is ZHZ^T. So the variance of the estimate of the curve $\hat{\phi}_\nu$ at each time point (and their covariances) are contained in the diagonals (and off-diagonal elements) of ZHZ^T. Confidence intervals for the permanent variance (the squared value of the curve estimate) at a particular time point can be constructed via a delta-method approximation involving that same asymptotic covariance matrix.[13]

References

[1] Akaike, H. (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control,* **19**, 716-723.

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

[3] Banfield, J. D. and Raftery, A. E. (1993). Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics,* **49**, 803-821.

[4] Bernhardt, A., Morris, M., Handcock, M. and Scott, M. (1997). Work and Opportunity in the Post-Industrial Labor Market. Final report to the Russell Sage and Rockefeller Foundations. Institute on Education and the Economy, Columbia University New York.

[5] Brumback, B. A., and Rice, J. A. (1998). Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves. *Journal of American Statistical Association* **93**, 961-976.

[6] Brumback, B. A. (1996). Statistical Models for Hormone Data. Ph.D Thesis, University of California, Berkeley.

[7] Clogg, C. C. (1995). Latent Class Models. In G. Arminger, C. C. Clogg and M. E. Sobel (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, 311-359. Plenum Press.

[8] Cressie, N. (1991). Statistics for Spatial Data. New York: Wiley.

[13] Confidence intervals for extra-mean variation for random quadratic models involve a slightly different calculation and will not be discussed in this paper.

[9] Davidian, M., and Gallant, A. R. (1993). The Nonlinear Mixed Effects Model with a Smooth Random Effects Density. *Biometrika*, **80**, 475-488.

[10] Diggle, P. J. and Verbyla, A. P. (1998). Nonparametric Estimation of Covariance Structure in Longitudinal Data. *Biometrics*, **54**, 401-415.

[11] Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford University Press.

[12] Flury, B. D. and Tarpey, T. (1993). Representing a Large Collection of Curves: A Case for Principal Points *The American Statistician*, **47**, 304-306.

[13] Gasser, T., Hall, P. and Presnell, B. (1998). Nonparametric Estimation of the Mode of a Distribution of Random Curves. *Journal of the Royal Statistical Society, Series B*, **60**, 681-691.

[14] Gottschalk, Peter and Moffit, Robert (1994). The Growth of Earnings Instability in the US Labor Market. *Brookings Papers on Economic Activity*, **2**, 217-272.

[15] Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* London: Chapman and Hall.

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

[17] Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian Model Averaging. *Statistical Science*, forthcoming.

[18] Jones, M. C. and Rice, J. A. (1992). Displaying the Important Features of a Large Collection of Similar Curves. *The American Statistician*, **46**, 140-145.

[19] Jones, M. C. (1990). Serial Correlation or Random Subject Effects? *Communications in Statistics Series B*, **19**(3), 1105-1123.

[20] Longford, N. T. (1993). *Random Coefficient Models* New York: Oxford.

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

[22] Muthén, B. and Shedden, K. (1998). Finite Mixture Modeling with Mixture Outcomes using the EM Algorithm. *Biometrics (forthcoming)*

[23] Pinheiro, J. C., Bates, D. M., and Lindstrom, M. J. (1994). Model Building for Nonlinear Mixed Effects Models Technical Report 931, University of Wisconsin, Madison, Dept. of Statistics.

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

[25] Rice, J. 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.

[26] Schwarz, G. (1978) Estimating the Dimension of a Model. *The Annals of Statistics.* **6**, 461-464.

[27] Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components.* Wiley.

[28] Scott, M. A. and Handcock, M. S. (1998) Latent Curve Covariance Models for Longitudinal Data. *Journal of the American Statistical Association (under review)*

[29] Scott, M. A. (1998). Statistical Models for Heterogeneity in the Labor Market. Ph.D Thesis, New York University.

[30] Simonoff, J. S. and Tsai, C. (1999). Semiparametric and Additive Model Selection Using an Improved AIC Criterion. *Journal of Computational and Graphical Statistics* **8**(1), 22-40.

[31] Topel, R. H. and Ward, M. P. (1992). Job Mobility and the Careers of Young Men. *Quarterly Journal of Economics,* **107**, 439-479.

[32] Verbeke, G. and Lesaffre, E. (1997). The Effect of Misspecifying the Random-Effects Distribution in Linear Mixed Models for Longitudinal Data. *Computational Statistics and Data Analysis* **23**, 541-556.

[33] Verbeke, G. and Lesaffre, E. (1996). A Linear Mixed-Effects Models with Heterogeneity in the Random-Effects Population. *Journal of American Statistical Association* **91**, 217-221.

[34] Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements.* Marcel Dekker.

[35] Xu, W., Hedeker, D., and Ramakrishnan, V. (1996). Mixtures in Random-effects Regression Models. Manuscript, School of Public Health, University of Illinois at Chicago.