QÜESTIIÓ, vol. 25, 2, p. 225-262, 2001

MODELIZACIÓN DE DATOS LONGITUDINALES CON ESTRUCTURAS DE COVARIANZA NO ESTACIONARIAS: MODELOS DE COEFICIENTES ALEATORIOS FRENTE A MODELOS ALTERNATIVOS

VICENTE NÚÑEZ-ANTÓN*
DALE L. ZIMMERMAN**

Un tema que ha suscitado el interés de los investigadores en datos longitudinales durante las dos últimas décadas, ha sido el desarrollo y uso de modelos paramétricos explícitos para la estructura de covarianza de los datos. Sin embargo, el análisis de estructuras de covarianza no estacionarias en el contexto de datos longitudinales no se ha realizado de forma detallada principalmente debido a que las distintas aplicaciones no hacían necesario su uso. Muchos son los modelos propuestos recientemente, pero la mayoría son estacionarios de segundo orden. Algunos de éstos, sin embargo, son no estacionarios y suficientemente flexibles, de tal forma que es posible modelizar varianzas no constantes y/o correlaciones que no sean sólo función del tiempo que separa a dos observaciones dadas. Estudiaremos algunas de estas propuestas y las compararemos con los modelos de coeficientes aleatorios, evaluando sus ventajas y desventajas e indicando cuándo su uso no es apropiado o útil. Presentaremos dos ejemplos para ilustrar el ajuste de estos modelos y los compararemos entre sí, mostrando de esta forma cómo pueden modelizarse datos longitudinales de forma efectiva y simple. En estos ejemplos, los distintos modelos alternativos, especialmente los modelos antedependientes, fueron superiores a los modelos de coeficientes aleatorios.

Modelling longitudinal data with nonstationary covariance structures: random coefficients models versus alternative models

Palabras clave: Antedependencia, modelos Arima, AIC, BIC, estructuras de covarianza, máxima verosilimitud residual, modelos mixtos

Clasificación AMS (MSC 2000): 62J05, 62F10, 62P10

* Departamento de Econometría y Estadística (E.A. III). Universidad del País Vasco. Avenida Lehendakari Aguirre, 83. 48015 Bilbao. E-mail: vn@alcib.bs.ehu.es.
** Department of Statistics and Actuarial Science. The University of Iowa. Iowa City, Iowa 52242. Estados Unidos.

– Recibido en abril de 2000.
– Aceptado en enero de 2001.

1. INTRODUCCIÓN

Un tema importante en el análisis de modelos de regresión y, especialmente, en su utilización para el análisis de datos longitudinales, ha sido el desarrollo paralelo de los distintos modelos paramétricos para la matriz de varianzas y covarianzas de los datos y de los modelos de coeficientes aleatorios. Recordemos que en el análisis de datos longitudinales, se realizan mediciones a lo largo del tiempo en cada una de las unidades experimentales (normalmente asignadas a diferentes grupos o tratamientos). A este respecto, podemos consultar como referencias bibliográficas, por ejemplo, a Laird y Ware (1982), Crowder y Hand (1990, cap. 5 y 6), Jones (1993, cap. 2 y 3), Diggle, Liang y Zeger (1994, cap. 4 y 5), Wolfinger (1996), y Verbeke y Molenberghs (1997, cap. 3). Sin embargo, uno de los aspectos a resaltar debe ser que los modelos de coeficientes aleatorios se han considerado típicamente distintos de los modelos que especifican de forma paramétrica la matriz de varianzas y covarianzas de los datos. Esto es principalmente debido a que en los modelos de coeficientes aleatorios se ve el origen de la estructura de covarianzas como regresiones que varían entre los individuos o entes en el estudio, en lugar de verla como una consideración relativa a similitud de la estructura intra-individuos. En cualquier caso, en general, los modelos de coeficientes aleatorios pueden usarse y ser de mucha utilidad en el contexto de datos longitudinales.

En cuanto a los modelos que utilizan estructuras paramétricas para la matriz de varianzas y covarianzas de los datos, podemos mencionar que la modelización paramétrica de la estructura de covarianzas tiene algunas ventajas: (*i*) permite obtener de una forma más eficiente los estimadores de los parámetros usados en la modelización de la estructura de medias en los datos; (*ii*) permite obtener estimadores más adecuados para los errores estándar de los estimadores de los parámetros usados en la modelización de la estructura de medias; (*iii*) en muchos casos, permite solucionar de forma efectiva los problemas relativos a datos faltantes o datos perdidos o a datos en los que los tiempos de medición no sean los mismos para todos los individuos; y (*iv*) puede utilizarse aún cuando el número de ocasiones en que se realizan mediciones en los individuos es grande en comparación con el número de individuos.

Uno de los modelos paramétricos más utilizados para la estructura de varianzas y covarianzas de los datos es el modelo de correlación en serie. Es decir, el modelo en el que las correlaciones muestrales para un individuo determinado decrecen a medida que el tiempo de separación entre dichas observaciones aumenta. Entre estos modelos el más utilizado es el modelo estacionario autorregresivo (modelo AR) y otras versiones no muy parametrizadas de modelos estacionarios de segundo orden (véase, por ejemplo, Jennrich y Schluchter, 1986; Diggle, 1988; Jones y Boadi-Boateng, 1991; y Muñoz y otros, 1992). En estos modelos, las varianzas son constantes en el tiempo y las correlaciones entre mediciones equidistantes en el tiempo son las mismas. Cuando éste no sea el caso, el uso de los modelos estacionarios no es aconsejable. Es decir, se deben considerar otros modelos capaces de modelizar esta no estacionariedad.

Si la no estacionariedad se da en las varianzas, se puede optar por transformar los datos para estabilizar las varianzas o por utilizar modelos que permitan tener varianzas heterogéneas (véase, por ejemplo, Wolfinger, 1996). Sin embargo, estas alternativas pueden llegar a no resolver el problema si los datos, además, presentan una no estacionariedad en correlación. A pesar de esto, existen modelos alternativos, aplicables a datos longitudinales, en los que se permite que exista no estacionariedad en varianza y en correlación. Posiblemente el más obvio de ellos es el modelo clásico multivariante completamente no estructurado. En muchos casos, sin embargo, la no estacionariedad de los datos tiene una estructura que puede modelizarse usando un modelo con pocos parámetros e ignorar esto puede, obviamente, eliminar todas las ventajas que hemos mencionado anteriormente sobre la modelización paramétrica de la estructura de covarianzas. Una familia de modelos paramétricos que puede permitir correlaciones no estacionarias es una generalización de los modelos AR, conocida como la familia de los modelos antedependientes: no estructurados (Gabriel, 1962; Kenward, 1987) y estructurados (Núñez Antón y Woodworth, 1994; Núñez Antón, 1997; Zimmerman y Núñez Antón, 1997). Otra generalización no estacionaria de los modelos AR, son los modelos autorregresivos integrados de medias móviles (modelos ARIMA) (véase, por ejemplo, Diggle, 1990). Una posibilidad final, de la que ya hemos hablado brevemente, son los modelos de coeficientes aleatorios. Nuevamente indicamos que a pesar de que estos modelos son típicamente modelos para regresiones de una variable de respuesta en el tiempo (o, posiblemente, en otras covariables) que varía de individuo a individuo, un modelo de coeficientes aleatorios puede usarse para modelizar ciertos tipos de no estacionariedad (véase, por ejemplo, Diggle, Liang y Zeger, 1994).

Cada uno de los modelos descritos anteriormente, además de los modelos estacionarios, han sido propuestos para su uso en contextos de datos longitudinales por al menos un autor, pero nunca han sido comparados ni evaluados en conjunto, menos aún ante la presencia de no estacionariedad en varianza y correlación. La práctica general suele ser que, ante un ajuste erróneo de los modelos, presumiblemente debido a la presencia de no estacionariedad, un modelo de coeficientes aleatorios sería el adecuado y el ajustado (véase, por ejemplo, Jones, 1990). Así, ante la falta de esta comparación entre los distintos modelos en datos reales, intentaremos comparar tanto los modelos estacionarios como los no estacionarios con el más utilizado de estos últimos, el de coeficientes aleatorios, en situaciones de no estacionariedad.

En este trabajo, consideraremos las ventajas y limitaciones de cada uno de los modelos, enfatizando los casos en que sus usos sean inadecuados. Para ello, presentaremos dos ejemplos que ilustrarán el ajuste y la comparación entre los modelos alternativos y el de efectos aleatorios. Los ejemplos demostrarán que es posible modelizar datos longitudinales no estacionarios de forma efectiva y, en algunos casos, utilizando modelos paramétricos estructurados. En la Sección 2 describimos los dos estudios longitudinales que analizaremos posteriormente en la Sección 5. Las hipótesis y notación que utilizaremos se mencionan en la Sección 3. En la Sección 4 realizamos una breve descripción

de los distintos modelos paramétricos para la estructura de covarianzas intra-individuos y de los aspectos computacionales para el ajuste de los mismos. Finalmente, en la Sección 6, evaluamos las distintas propuestas del trabajo y establecemos las conclusiones del mismo.

2. DATOS

Utilizaremos datos provenientes de dos estudios longitudinales para motivar la consideración de modelos no estacionarios distintos al de coeficientes aleatorios. Estos datos también los usaremos, en la Sección 5, para ilustrar el ajuste y comparación de los distintos modelos entre ellos y con el de coeficientes aleatorios.

Los datos que denominaremos *race data*, que amablemente nos ha dejado utilizar Ian Jollife de la Universidad de Kent, corresponden a cada uno de los tiempos parciales (en minutos) para cada uno de los 80 competidores en cada una de las secciones de 10 kilómetros de una carrera con un total de 100 kilómetros, que se llevó a cabo en el Reino Unido en 1984. Además de los tiempos parciales, los datos contienen la edad de 76 de los 80 competidores.

Tabla 1. Medias muestrales (en minutos), varianzas y correlaciones muestrales correspondientes a los datos del estudio longitudinal *race data*.

Sección (t)	1	2	3	4	5	6	7	8	9	10
Corr.:										
	1.0									
	.95	1.0								
	.84	.89	1.0							
	.78	.82	.92	1.0						
	.60	.63	.75	.88	1.0					
	.60	.62	.72	.84	.94	1.0				
	.52	.54	.60	.69	.75	.84	1.0			
	.45	.48	.61	.69	.78	.84	.78	1.0		
	.51	.51	.56	.65	.73	.77	.69	.75	1.0	
	.38	.40	.44	.49	.52	.64	.72	.65	.77	1.0
Medias:	47.8	50.9	49.6	53.2	54.7	60.1	62.4	69.3	68.7	67.4
Var.:	26.9	34.8	49.0	58.9	91.4	149.9	107.9	152.2	145.0	167.2

Los datos que denominaremos *cattle data* corresponden a un experimento descrito por Kenward (1987). Un grupo de vacas fue sometido a uno de los dos posibles tratamientos para parásitos intestinales, a los que denominaremos A y B. Las vacas fueron pesadas 11 veces en un periodo de tiempo de 133 días. Treinta vacas recibieron cada uno de los tratamientos. Las primeras 10 mediciones se realizaron cada dos semanas, mientras que la última medición se realizó una semana después de la décima.

Tabla 2. Medias (en Kg.), varianzas y correlaciones muestrales correspondiente a los datos del estudio longitudinal *cattle data*, tratamiento A.

Tiempo (en días)	0	14	28	42	56	70	84	98	112	126	133
Corr.:											
	1.0										
	.82	1.0									
	.76	.91	1.0								
	.66	.84	.93	1.0							
	.64	.80	.88	.94	1.0						
	.59	.74	.85	.91	.94	1.0					
	.52	.63	.75	.83	.87	.93	1.0				
	.53	.67	.77	.84	.89	.94	.93	1.0			
	.52	.60	.71	.77	.84	.90	.93	.97	1.0		
	.48	.58	.70	.73	.80	.87	.88	.94	.96	1.0	
	.48	.55	.68	.71	.77	.83	.86	.92	.96	.98	1.0
Medias:	226.2	230.3	246.9	265.6	281.2	294.9	304.7	312.9	315.1	324.1	325.5
Var.:	105.6	155.1	165.2	184.9	243.0	283.8	306.6	340.7	389.2	470.1	444.6

Las Tablas 1, 2 y 3 muestran las medias, varianzas y correlaciones muestrales correspondientes a cada uno de los dos conjuntos de datos, mientras que las Figuras 1 y 2 muestran la evolución de la media muestral en el tiempo para cada uno de los estudios longitudinales. En el caso del *cattle data*, el contraste de homogeneidad indicó que no era razonable utilizar la misma matriz de varianzas y covarianzas para los dos grupos de tratamientos. Las medias muestrales correspondientes al *race data* (ver Tabla 1 y Figura 1) indican que los tiempos tienden a incrementarse hasta los primeros 80 kilómetros, pero luego decrecen de forma leve en los últimos 20 kilómetros. Esto indica que un modelo con estructura lineal en t no sería el adecuado y que habría que intentar ajustar un modelo cuadrático o cúbico en t. En el sentido práctico, estos resultados indican que, a medida que la carrera avanza, los corredores se cansan más y, por tanto, tardarán cada vez mas en cubrir cada tramo de diez kilómetros.

Tabla 3. Medias (en Kg.), varianzas y correlaciones muestrales correspondiente a los datos del estudio longitudinal *cattle data*, tratamiento B.

Tiempo (en días)	0	14	28	42	56	70	84	98	112	126	133
Corr.:											
	1.0										
	.86	1.0									
	.83	.94	1.0								
	.68	.89	.93	1.0							
	.67	.84	.88	.95	1.0						
	.66	.84	.87	.95	.98	1.0					
	.61	.78	.82	.91	.94	.97	1.0				
	.63	.81	.84	.92	.92	.95	.95	1.0			
	.63	.80	.79	.90	.93	.95	.93	.96	1.0		
	.48	.65	.67	.78	.78	.82	.76	.78	.83	1.0	
	.44	.57	.62	.73	.68	.74	.71	.71	.75	.92	1.0
Medias:	224.6	227.9	243.5	262.5	276.4	290.1	299.2	317.7	319.7	326.9	320.5
Var.:	105.3	108.4	147.1	198.5	217.7	250.4	248.2	234.1	287.0	404.7	598.6

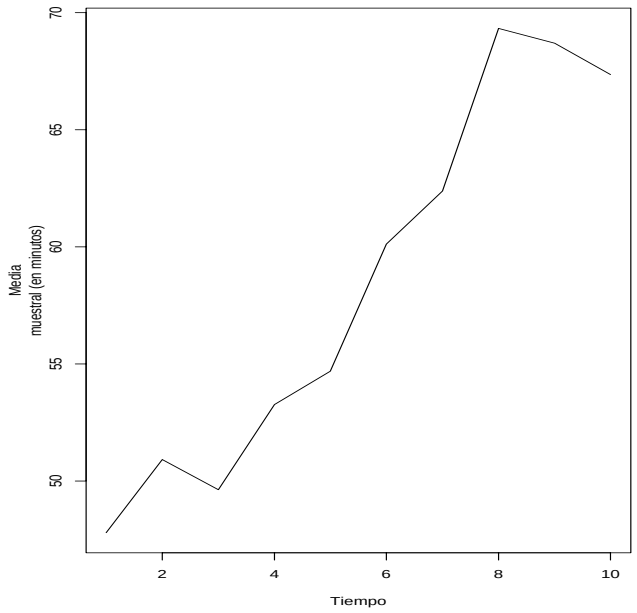

Figura 1. Medias muestrales para el estudio longitudinal *race data* en función de la sección de la carrera (tiempo).

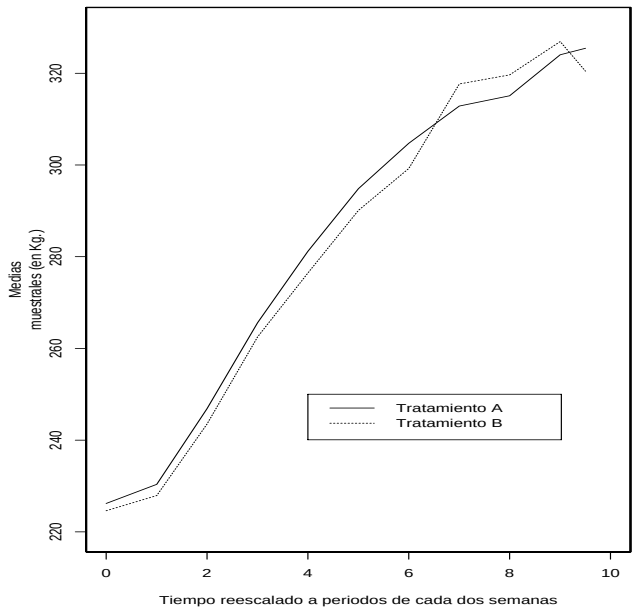

Figura 2. Medias muestrales para el estudio longitudinal *cattle data*, grupos A y B, en función del tiempo reescalado a periodos de cada dos semanas (tiempo=t/14).

En el caso del estudio *cattle data* (ver Tablas 2 y 3 y Figura 2), se observa que las medias correspondientes a ambos tratamientos tienden a incrementarse con el tiempo, aunque no linealmente. Las medias del grupo correspondiente al tratamiento A son ligeramente superiores hasta aproximadamente la séptima medición ($t = 6$, que corresponde a una medición efectuada a las 12 semanas o a los 84 días, en la Figura 2). A partir de ese instante, las medias muestrales del tratamiento B superan a las del tratamiento A, aunque al final del estudio, a partir de la medición efectuada a las 18 semanas, la situación vuelve a invertirse. Nuevamente, estas tendencias indican que las vacas tienden a ganar peso a medida que el tratamiento al que se les somete empieza a hacer efecto o se empieza a observar sus resultados.

Estas matrices muestran algunas características interesantes, comunes a muchos datos en el coxtexto de datos longitudinales:

1. Las varianzas no son homogéneas y tienden a incrementarse con el tiempo (es decir, en el caso del *race data* a medida que la carrera avanza). Este incremento es bastante monótono para el *cattle data* y no tanto para el caso del *race data*. Una forma de interpretar esta característica para el caso del *race data* es pensar que a medida que la carrera avanza, los corredores tenderán a diferenciarse cada vez más unos de

otros en cuanto al tiempo necesario para completar cada tramo de la misma, lo que incrementará la varianza a medida que la carrera avanza. En el caso del *cattle data* un razonamiento similar, pero aplicado a los pesos de las vacas y su variabilidad a medida que el estudio avanza, permite comprender este incremento de varianza con el tiempo.

2. Las correlaciones son todas positivas.
3. Existe correlación en serie dado que las correlaciones dentro de una columna específica tienden a decrecer hacia cero (a menos que estén cercanas a cero inicialmente).
4. Las correlaciones entre observaciones separadas por la misma distancia (es decir, subdiagonales o superdiagonales de la matriz) no son constantes. Por el contrario, en el caso del tratamiento A del *cattle data* tienden a incrementarse al principio del estudio antes de nivelarse o, en algunos casos (por ejemplo, tratamiento B para el *cattle data*) decrecer levemente al final del estudio. Por otro lado, en el caso del *race data*, son menores al final del estudio que al principio del mismo. Una interpretación práctica de este hecho en el caso del *race data* podría ser que los tiempos utilizados en una sección de la carrera serán predictores más fiables de los tiempos en las secciones siguientes al principio de la carrera que al final de la misma.

Para estabilizar las varianzas, transformamos las respuestas para cada uno de los datos utilizando diversas alternativas, sin conseguir el resultado esperado. Esto era previsible dado que en estos casos las varianzas no parecían ser funciones suaves de la media. Aún en los casos en que la transformación estabilizaba las varianzas, la no estacionariedad de las correlaciones permanecía presente en los mismos.

3. EL MODELO GENERAL: HIPÓTESIS Y NOTACIÓN

Supondremos que estamos en el contexto de datos longitudinales. Es decir, se realizan mediciones a lo largo del tiempo en cada uno de los m individuos o entes bajo estudio. Sea $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})'$ el vector de las n_i mediciones que se han realizado en el i-ésimo individuo y sea $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})'$ el vector que contiene los tiempos correspondientes en los que se han realizado dichas mediciones. Además suponemos que observamos un vector p-variante de covariables, $\mathbf{x}_{ij}$, asociado con y_{ij}. Así utilizando una notación matricial más compacta, sean $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_m')'$, $\mathbf{t} = (\mathbf{t}_1', \ldots, \mathbf{t}_m')'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})'$, $\mathbf{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_m')'$, y $N = \sum_{i=1}^{m} n_i$.

Denominaremos diseño de medición al conjunto de los tiempos en que las mediciones se han realizado. No imponemos ninguna restricción en el diseño de medición. En general, los tiempos en los que se realizan las mediciones para un mismo individuo pueden estar espaciados de forma irregular y pueden ser distintos para los diferentes individuos. Si los tiempos de medición son los mismos para todos los individuos, entonces tendre-

mos un diseño de medición rectangular, como es el caso en los dos ejemplos descritos en la Sección 2. Analizaremos, posteriormente, la influencia de tener o no tener este tipo de diseño en la modelización de la estructura de covarianzas.

Las hipótesis generales que tendremos presentes en nuestra modelización paramétrica de estructuras de covarianzas son:

1. Las mediciones realizadas en individuos distintos son independientes entre sí, aunque las mediciones realizadas en un mismo individuo podrán ser dependientes.
2. La respuesta media (para los individuos) es una combinación lineal de funciones conocidas de las covariables y/o del tiempo.
3. Las respuestas están normalmente distribuidas (posiblemente después de hacer alguna transformación).
4. La naturaleza de la variación de segundo orden intra-individuos es la misma para todos los individuos.
5. Si hay datos perdidos, éstos se asumen ignorables (Laird, 1988).

Estas hipótesis nos dan el modelo $\mathbf{y} \sim MVN\,(\mathbf{X}\beta, \Sigma)$, donde β es un vector p-dimensional de parámetros fijos, desconocidos y, típicamente, no restringidos y Σ es la matriz de covarianzas desconocida de $N \times N$, diagonal en bloques, con elementos no nulos en los bloques respectivos dados por $\mathbf{V}_i = \mathrm{var}(\mathbf{y}_i)$. En el caso de diseño de medición rectangular, las $\mathbf{V}_i$ son todas iguales.

A continuación, para tener una modelización paramétrica de la estructura de covarianzas, tendremos que establecer un modelo para Σ:

$$\mathbf{y} \sim MVN\,(\mathbf{X}\beta, \Sigma(\mathbf{t}, \theta)), \tag{1}$$

donde θ es un vector q-dimensional de parámetros desconocidos, restringido a un espacio de parámetros Θ que es o el conjunto de todos los vectores θ para los que Σ es definida positiva o algún subconjunto de este conjunto. Los elementos de Σ pueden ser funciones de $\mathbf{t}$, pero no de $\mathbf{X}$. Además, Σ es definida positiva si y sólo si $\mathbf{V}_i$ es definida positiva para cada i. La estimación de los parámetros β y θ en este modelo se realiza utilizando el método de máxima verosimilitud o el de máxima verosimilitud residual (MVR), que a menudo debe implementarse utilizando subrutinas de optimización numérica. Los detalles adicionales y más específicos sobre cómo se ajustan los distintos modelos y la comparación entre los mismos se mencionarán en la Sección 5.

A continuación describiremos la mayoría de los modelos que se utilizan frecuentemente para la modelización de la estructura de covarianzas en el contexto de datos longitudinales. Estos modelos serán posteriormente ajustados y comparados con el modelo de coeficientes aleatorios. En cada caso, mencionaremos sus propiedades principales, ventajas y limitaciones. Dado que los modelos se utilizan para la estructura de covarianzas

intra-individuos, eliminaremos, al menos inicialmente, el subíndice i (denotando al individuo) en y_{ij}, n_i, x_{ij}, y $\mathbf{V}_i$. Además, $\{v_{ju}\}$ denotará a los elementos de $\mathbf{V}$ y $\{\rho_{ju}\}$ a los elementos correspondiente de la matriz de correlaciones.

4. MODELOS PARAMÉTRICOS PARA LA ESTRUCTURA DE COVARIANZAS INTRA-INDIVIDUOS

En esta sección describimos los modelos más utilizados para la estructura de covarianzas intra-individuos: (*i*) de coeficientes aleatorios (CA), (*ii*) de simetría compuesta (SC), (*iii*) autorregresivo de primer orden (AR(1)), (*iv*) Huynh-Feldt (HF), (*v*) Toeplitz (TOEP y TOEPH), (*vi*) no estructurado (NE), (*vii*) antedependientes no estructurados (AD), (*viii*) antedependientes estructurados (ADE) y (*ix*) ARIMA. En la descripción de cada uno de estos modelos, ilustraremos la estructura del modelo para el caso en que todos los individuos tengan $n_i = n = 4$ $(i = 1, \ldots, m)$ observaciones igualmente espaciadas en los tiempos $t = 1, 2, 3, 4$. A continuación realizamos una descripción de los modelos y luego, al final de esta sección, mencionaremos algunos aspectos computacionales básicos que indican la forma de ajustar estos modelos.

4.1. Modelo de Coeficientes Aleatorios (CA)

Los modelos de coeficientes aleatorios fueron introducidos por Rao (1959) y su popularidad ha ido en aumento debido a su interpretabilidad intuitiva y su relación cercana con las técnicas Bayesianas (véase, por ejemplo, Laird y Ware, 1982; Reinsel, 1982; o Rutter y Elashoff, 1994).

El modelo general de coeficientes aleatorios viene dado por la ecuación

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i \qquad (i = 1, \ldots, m),$$

donde las $\mathbf{Z}_i$'s representan matrices conocidas, cuya estructura describiremos en breve, los $\mathbf{u}_i$ son vectores de coeficientes aleatorios distribuidos, independientes entre sí, como $MVN(\mathbf{0}, \mathbf{G}_i)$, las $\mathbf{G}_i$'s son matrices definidas positivas y, aparte de esto, matrices no restringidas, y los $\mathbf{e}_i$'s están distribuidos, independientes de los $\mathbf{u}_i$'s y entre sí, como $MVN(\mathbf{0}, \sigma^2\mathbf{I}_{n_i})$. En general, las $\mathbf{G}_i$ son las mismas para todos los individuos y, por tanto, la matriz de covarianzas de $\mathbf{y}_i$ será $\mathbf{V}_i = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \sigma^2\mathbf{I}_{n_i}$. Entre los casos especiales tenemos el modelo de coeficientes aleatorios lineales (CAL) y el modelo de coeficientes aleatorios cuadráticos (CAC). En el caso cuadrático, $\mathbf{Z}_i = [\mathbf{1}_{n_i}, \mathbf{t}_i, (t_{i1}^2, t_{i2}^2, \ldots, t_{in_i}^2)']$, y

$$\mathbf{G} = \begin{pmatrix} \sigma_{00} & \sigma_{01} & \sigma_{02} \\ \sigma_{01} & \sigma_{11} & \sigma_{12} \\ \sigma_{02} & \sigma_{12} & \sigma_{22} \end{pmatrix}$$

En el caso lineal, $\mathbf{Z}_i = [\mathbf{1}_{n_i}, \mathbf{t}_i]$. Otro caso especial que se obtiene si $\mathbf{Z}_i = \mathbf{1}_{n_i}$, es equivalente al comúnmente utilizado modelo de simetría compuesta.

Los modelos de coeficientes aleatorios se han considerado típicamente distintos de los modelos que especifican de forma paramétrica la matriz de varianzas y covarianzas de los datos. Esto es principalmente debido a que en los modelos de coeficientes aleatorios se ve el origen de la estructura de covarianzas como regresiones que varían entre los individuos o entes en el estudio, en lugar de verla como una consideración relativa a similitud de la estructura intra-individuos. En cualquier caso, los modelos de coeficientes aleatorios pueden, en general, usarse y ser de mucha utilidad en el contexto de datos longitudinales para modelizar estructuras que presentan varianzas no constantes y correlaciones no estacionarias, con la excepción del modelo de simetría compuesta. Este es un hecho muy poco conocido y, obviamente, poco utilizado. Consideremos, por ejemplo, la estructura CAL para un individuo en el que se han realizado mediciones en tiempos igualmente espaciados, digamos $t_1 = 1, \ldots, t_n = n$. Para este modelo, tendremos que $\text{var}(y_{ij}) = \sigma^2 + \sigma_{00} + 2\sigma_{01} j + \sigma_{11} j^2$ y

$$\text{corr}(y_{ij}, y_{ik}) = \frac{\sigma_{00} + \sigma_{01}(j+k) + \sigma_{11} jk}{\sqrt{\sigma^2 + \sigma_{00} + 2\sigma_{01} j + \sigma_{11} j^2}\sqrt{\sigma^2 + \sigma_{00} + 2\sigma_{01} k + \sigma_{11} k^2}}$$

Como puede observarse, éste es un modelo muy flexible que permite modelizar diversos tipos de comportamientos en las varianzas y correlaciones, entre los que se incluyen varianzas que crecen o decrecen, así como correlaciones que pueden ser negativas o positivas. Sin embargo, no permite que la varianza sea una función cóncava hacia abajo del tiempo o que la varianza sea constante si las correlaciones entre observaciones igualmente espaciadas no lo son. Además, el número de parámetros en estos modelos no está relacionado con el número de ocasiones en las que se realizan mediciones.

4.2. Modelos de Simetría Compuesta (SC) y Autorregresivos de Primer Orden (AR(1))

Estos dos modelos paramétricos para la estructura de covarianzas se consideran modelos homogéneos. Es decir, la varianza a lo largo de la diagonal principal de esta matriz permanece constante. Sin embargo, difieren en el trato de la covarianza. Para el modelo SC, ésta permanece constante y para el modelo AR(1), ésta decrece exponencialmente. Las matrices de varianzas y covarianzas para los modelos SC y AR(1) son

$$v\begin{bmatrix} 1 & \rho & \rho & \rho \\ & 1 & \rho & \rho \\ & & 1 & \rho \\ & & & 1 \end{bmatrix} \quad \text{y} \quad v\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix},$$

respectivamente, en donde $v = \text{Var}(y_{ij})$. Como puede verse, estos modelos tienen dos parámetros cada uno. Es decir, son modelos muy restringidos que deben utilizarse sólo

en los casos en que se tenga correlaciones constantes o que decrezcan exponencialmente, además de varianzas constantes. Sin embargo, es posible obtener una generalización directa de los modelos SC y AR(1) permitiendo que las varianzas en la diagonal principal de las matrices de covarianzas sean distintas. Denominaremos a estos modelos, que son una extensión heterogénea de los anteriores, SCH y ARH(1), respectivamente. Las matrices de varianzas y covarianzas para estos modelos son, por tanto,

$$\begin{bmatrix} v_{11} & \sqrt{v_{11}v_{22}}\rho & \sqrt{v_{11}v_{33}}\rho & \sqrt{v_{11}v_{44}}\rho \\ & v_{22} & \sqrt{v_{22}v_{33}}\rho & \sqrt{v_{22}v_{44}}\rho \\ & & v_{33} & \sqrt{v_{33}v_{44}}\rho \\ & & & v_{44} \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} v_{11} & \sqrt{v_{11}v_{22}}\rho & \sqrt{v_{11}v_{33}}\rho^2 & \sqrt{v_{11}v_{44}}\rho^3 \\ & v_{22} & \sqrt{v_{22}v_{33}}\rho & \sqrt{v_{22}v_{44}}\rho^2 \\ & & v_{33} & \sqrt{v_{33}v_{44}}\rho \\ & & & v_{44} \end{bmatrix},$$

respectivamente. Como puede verse, estos modelos tienen cinco parámetros cada uno. Son modelos menos restringidos que sus versiones homogéneas en varianza, pero siguen siendo válidos sólo para los casos de correlaciones constantes o exponencialmente decrecientes.

4.3. Modelos Huynh-Feldt (HF) y Toeplitz (TOEP y TOEPH)

Los contrastes de esfericidad de Huynh y Feldt (1970) han sido muy relevantes para decidir si optar por un análisis univariante o multivariante para datos de medidas repetidas o longitudinales. Sin embargo, pocas veces hemos visto que la estructura de covarianzas Huynh-Feldt haya sido ajustada directamente a los datos principalmente debido a que los análisis estándar utilizan el hecho de que un conjunto de contrastes ortogonales de datos con esta estructura de covarianzas tienen de por si una estructura de covarianzas diagonal. El modelo HF es similar al modelo SCH, tanto en el número de parámetros como en el hecho de poseer una heterogeneidad no estructurada para las varianzas en la diagonal principal. Sin embargo, la estructura HF construye los elementos fuera de la diagonal principal utilizando medias aritméticas en lugar de medias geométricas. Así, la estructura de varianzas y covarianzas Huynh-Feldt (HF), que tiene 5 parámetros, puede expresarse como:

$$\begin{bmatrix} v_{11} & (v_{11}+v_{22})/2-\lambda & (v_{11}+v_{33})/2-\lambda & (v_{11}+v_{44})/2-\lambda \\ & v_{22} & (v_{22}+v_{33})/2-\lambda & (v_{22}+v_{44})/2-\lambda \\ & & v_{33} & (v_{33}+v_{44})/2-\lambda \\ & & & v_{44} \end{bmatrix}$$

Por otro lado, Los modelos TOEP y TOEPH generalizan, respectivamente, a los modelos AR(1) y ARH(1). En estos modelos no se asume que las correlaciones decrezcan de forma exponencial y, por el contrario, se las deja variar de forma no estructurada. Sin embargo, seguimos asumiendo que las correlaciones entre observaciones igualmente espaciadas son las mismas. Estos modelos tienen 4 y 7 parámetros y sus matrices de varianzas y covarianzas son

$$v\begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ & 1 & \rho_1 & \rho_2 \\ & & 1 & \rho_1 \\ & & & 1 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} v_{11} & \sqrt{v_{11}v_{22}}\rho_1 & \sqrt{v_{11}v_{33}}\rho_2 & \sqrt{v_{11}v_{44}}\rho_3 \\ & v_{22} & \sqrt{v_{22}v_{33}}\rho_1 & \sqrt{v_{22}v_{44}}\rho_2 \\ & & v_{33} & \sqrt{v_{33}v_{44}}\rho_1 \\ & & & v_{44} \end{bmatrix},$$

respectivamente, en donde $v = \text{Var}(y_{ij})$.

4.4. Modelo No Estructurado (NE)

El modelo no estructurado de covarianzas es el caso extremo de modelización paramétrica de estructuras de covarianza en el que θ tiene $n(n+1)/2$ varianzas y covarianzas, o equivalentemente $n(n+1)/2$ varianzas y correlaciones. Dado que es un modelo completamente general para la estructura de covarianzas intra-individuo, no requiere que las observaciones se encuentren igualmente espaciadas entre sí. Para este modelo, el espacio de parámetros $\Theta = \{\theta\text{: } \mathbf{V} \text{ es definida positiva}\}$ no puede ser expresado como restricciones de desigualdades lineales en cada uno de los parámetros, lo que origina que estas restricciones sean muy difíciles de cumplir. El modelo NE tiene 10 parámetros y puede expresarse como:

$$\begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} \\ & v_{22} & v_{23} & v_{24} \\ & & v_{33} & v_{34} \\ & & & v_{44} \end{bmatrix}$$

4.5. Modelos Antedependientes no Estructurados (AD)

Las observaciones normales multivariantes $y_1, \ldots, y_n$ son antedependientes de orden s [AD(s)] si y_j e y_{j+k+1}, condicionadas a las observaciones intermedias y_{j+1}, ..., y_{j+k}, son independientes, para todo $j = 1, \ldots, n-k-1$ y todo $k \geq s$. La definición original de antedependencia fue propuesta por Gabriel (1962), aunque en la misma nunca se hizo referencia alguna a una estructura paramétrica. Una definición equivalente para el modelo AD(s), pero especificada de forma paramétrica, puede expresarse con las ecuaciones:

$$\begin{aligned} y_1 &= \mathbf{x}_1'\beta + \varepsilon_1, \\ y_j &= \mathbf{x}_j'\beta + \sum_{k=1}^{s^*} \phi_{jk}(y_{j-k} - \mathbf{x}_{j-k}'\beta) + \varepsilon_j \quad (j = 2, \ldots, n) \end{aligned} \tag{2}$$

donde $s^* = \min(s, j-1)$, los ε_j's son variables aleatorias normales independientes con media cero y varianzas, $\sigma_j^2 > 0$, que puede depender del tiempo. Los ϕ_{jk}'s son parámetros no estructurados. Es importante indicar que σ_j^2 y v_{jj} no son las mismas cantidades. De la especificación paramétrica en (2), podemos observar cómo los modelos AD generalizan a los modelos estacionarios AR: tal y como ocurre en los modelos AR, los modelos AD permiten la existencia de correlación en serie en las observaciones realizadas en cada individuo pero, a diferencia de los modelos AR, los modelos AD no exigen que las varianzas sean constantes o que las correlaciones entre observaciones igualmente espaciadas en el tiempo sean las mismas. En resumen, podríamos decir que el modelo AR(s) es un caso especial del modelo AD(s) en el que: (a) $\phi_{jk} \equiv \phi_k$ para $j = s+1, \ldots, n$ y $k = 1, \ldots, s$; (b) las s raíces de la ecuación característica del modelo AR(s), $1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_s x^s = 0$, todas son mayores que uno en valor absoluto; (c) $\sigma_{s+1}^2 = \sigma_{s+2}^2 = \cdots = \sigma_n^2 > 0$; y (d) los «valores iniciales» $\{\phi_{jk}\colon\ j = 2, \ldots, s;\ k = 1, \ldots, j-1\}$ y $\sigma_1^2, \sigma_2^2, \ldots, \sigma_s^2$ se eligen de una forma adecuada. Además, mencionaremos que el modelo AD($n-1$) es equivalente al modelo NE y que el modelo AD(0) es equivalente al modelo de independencia heterogéneo (es decir, $\mathbf{V} = \text{diag}(v_{11}, \ldots, v_{nn})$). Por ejemplo, podemos escribir la matriz de varianzas y covarianzas para el modelo AD(1), que tiene 7 parámetros.

$$\begin{bmatrix} v_{11} & \sqrt{v_{11}v_{22}}\rho_1 & \sqrt{v_{11}v_{33}}\rho_1\rho_2 & \sqrt{v_{11}v_{44}}\rho_1\rho_2\rho_3 \\ & v_{22} & \sqrt{v_{22}v_{33}}\rho_2 & \sqrt{v_{22}v_{44}}\rho_2\rho_3 \\ & & v_{33} & \sqrt{v_{33}v_{44}}\rho_3 \\ & & & v_{44} \end{bmatrix}$$

Nos referiremos al modelo (2) como un modelo AD no estructurado de orden s [AD(s)], donde al decir «no estructurado» nos referimos al hecho de que los parámetros ϕ_{jk} y σ_j^2 no se pueden expresar como funciones de un número menor de parámetros. Un modelo de covarianzas AD(s) tiene $(s+1)(2n-s)/2$ parámetros. Gabriel (1962) además propuso un contraste de razón de verosimilitudes que permite determinar el orden adecuado de antedependencia para un conjunto de datos específicos.

La ecuación (2) es una especificación *autorregresiva* de un modelo AD(s), es decir una especificación paramétrica en función de los coeficientes autorregresivos $\{\phi_{jk}\colon\ j = 2\ldots, n;\ k = 1, \ldots, s^*\}$ y la varianzas específicas $\{\sigma_j^2\colon\ j = 1, \ldots, n\}$. Una segunda forma equivalente de especificar (o reparametrizar) el modelo AD(s) se denomina especificación *de covarianza*, la que es una especificación paramétrica en función de las varianzas y covarianzas (o correlaciones) de las observaciones. Es fácilmente demostrable que las varianzas y las covarianzas en las primeras s subdiagonales (or superdiagonales) de la matriz de covarianzas de un modelo AD(s) permanecen no estructuradas, mientras que las restantes covarianzas se encuentran completamente determinadas por estos valores. Por ejemplo, en el caso de un modelo de primer orden, la matriz $\mathbf{V}$ puede escribirse de la siguiente forma:

$$
(3)\qquad
\begin{bmatrix}
v_{11} & & & & \\
\sqrt{v_{11}v_{22}}\rho_1 & v_{22} & & & \\
\sqrt{v_{11}v_{33}}\rho_1\rho_2 & \sqrt{v_{22}v_{33}}\rho_2 & v_{33} & & \\
\sqrt{v_{11}v_{44}}\rho_1\rho_2\rho_3 & \sqrt{v_{22}v_{44}}\rho_2\rho_3 & \sqrt{v_{33}v_{44}}\rho_3 & \ddots & \\
\vdots & \vdots & \vdots & \ddots & v_{n-1,n-1} \\
\sqrt{v_{11}v_{nn}}\prod_{j=1}^{n-1}\rho_j & \sqrt{v_{22}v_{nn}}\prod_{j=2}^{n-1}\rho_j & \cdots & \cdots & \sqrt{v_{n-1,n-1}v_{nn}}\rho_{n-1}\ \ v_{nn}
\end{bmatrix}
$$

donde $\rho_j \equiv \rho_{j,j+1}$. En el caso en que $s > 1$, las correlaciones en las subdiagonales $s+1,\ldots,n-1$ de $\mathbf{V}$ dependerán funcionalmente de las correlaciones en las primeras s subdiagonales, pero esta dependencia es mucho más complicada que la estructura multiplicativa evidente en (3).

Una tercera especificación posible para un modelo AD(s) se denomina la especificación de *concentración*. En este caso, se especifican los elementos en la diagonal principal y las primeras s subdiagonales (y superdiagonales) de la matriz que, comúnmente, se denomina matriz de concentración $\mathbf{V}^{-1}$. Los elementos restantes de $\mathbf{V}^{-1}$ son cero (Gabriel, 1962, Núñez-Antón y otros, 1995, Zimmerman y otros, 1998). Se pueden encontrar más detalles sobre los tres tipos de especificación de estos modelos en Zimmerman y Núñez-Antón (1997).

En general no es posible establecer el espacio de parámetros en las especificaciones de covarianza y de concentración en función de restricciones simples sobre cada uno de los parámetros. Una excepción es el caso de especificación de covarianza para el modelo AD(1) en (3), para el que $\Theta = \{\theta\text{: } -1 < \rho_j < 1 \text{ para } j = 1,\ldots,n-1;\ v_{jj} > 0 \text{ para } j = 1,\ldots,n\}$. Es decir, con la excepción del modelo de primer orden, la especificación autorregresiva, que no impone restricciones en los coeficientes autorregresivos, es la más conveniente desde el punto de vista de imponer restricciones en los parámetros. De hecho, la especificación autorregresiva puede incluso evitar imponer todo tipo de restricciones en los parámetros si se reparametriza las varianzas específicas de forma que $\gamma_j = \log\sigma_j^2$. Cualquier especificación del modelo AD(s) se puede utilizar sin importar el espaciamiento temporal entre las mediciones.

4.6. Modelos Antedependientes Estructurados (ADE)

Aunque sabemos que los modelos AD tienen menos parámetros que el modelo NE, aún pueden tener demasiados parámetros en situaciones reales. Por ejemplo, si $n = 10$ y tenemos un diseño de medición rectangular, el modelo NE tiene 45 parámetros, mientras que los modelos AD(1) y AD(2) tienen 19 y 27 parámetros, respectivamente. De hecho, el modelo NE tiene $O(n^2)$ parámetros, mientras que los modelos AD tienen $O(n)$ parámetros y los modelos estacionarios más utilizados sólo tienen $O(1)$ parámetros.

Esta situación originó que Zimmerman y Núñez-Antón (1997) propusiesen versiones estructuradas y menos generales de los modelos AD, a las que denominaron modelos antedependientes estructurados (ADE). En estos modelos, los coeficientes autorregresivos, las correlaciones, o las correlaciones parciales (dependiendo de si la especificación del modelo es autorregresiva, de covarianza, o de concentración) siguen una ley de potencias de Box-Cox, y las varianzas específicas, varianzas, o varianzas parciales (nuevamente dependiendo de la especificación utilizada) son funciones polinomiales o por tramos (es decir, que toma valores distintos en distintos tramos de la escala temporal) del tiempo de medición. Como ejemplo, escribiremos de forma detallada las especificaciones estructuradas autorregresiva y de covarianza. En ambos casos, f es una función dada por (Núñez-Antón y Woodworth, 1994):

$$f(t;\lambda) = \begin{cases} (t^{\lambda}-1)/\lambda & \text{si } \lambda \neq 0 \\ \log t & \text{si } \lambda = 0 \end{cases} \tag{4}$$

y g, para que en la práctica tenga una especificación útil, es una función de pocos parámetros (e.g., una función polinomial de bajo orden).

Especificación autorregresiva (ADE-EA):

$$\begin{aligned} \phi_{jk} &= \phi_k^{f(t_j;\lambda_k)-f(t_{j-k};\lambda_k)} \quad (j = s+1,\ldots,n;\ k = 1,\ldots,s), \\ \sigma_j^2 &= \sigma^2 g(t_j;\psi) \quad (j = s+1,\ldots,n). \end{aligned} \tag{5}$$

Especificación de covarianza (ADE-EC):

$$\begin{aligned} \rho_{j,j-k} &= \rho_k^{f(t_j;\lambda_k)-f(t_{j-k};\lambda_k)} \quad (j = k+1,\ldots,n;\ k = 1,\ldots,s), \\ v_{jj} &= \sigma^2 g(t_j;\psi) \quad (j = 1,\ldots,n). \end{aligned} \tag{6}$$

Indicaremos que en el caso del modelo ADE-EA, $\{\phi_{jk} : j = 2,\ldots,s;\ k = 1,\ldots,j-1\}$ y $\sigma_1^2,\ldots,\sigma_s^2$ se dejan sin estructurar (es decir, como parámetros a estimar en el modelo). Además, la forma funcional de Box-Cox que tiene f especifica la estructura del modelo de tal forma que, los coeficientes autorregresivos de orden k (es decir, ϕ_k en el modelo ADE-EA) o las correlaciones en la k-ésima diagonal (en el modelo ADE-EC) son monótonas creciente si $\lambda_k < 1$, monótonas decreciente si $\lambda_k > 1$, o constantes si $\lambda_k = 1$ ($k = 1,\ldots,s$). Explicado de una forma más simple, el efecto de f es el de transformar la escala temporal de una forma no lineal para poder lograr que los coeficientes autorregresivos (en el modelo ADE-EA) o las correlaciones entre mediciones equidistantes en el tiempo (en el modelo ADE-EC), en la escala transformada, sean constantes.

Un modelo ADE tiene bastante menos parámetros que un modelo AD del mismo orden. Por ejemplo, si g es cuadrática, entonces, los modelos ADE-EA(1) y ADE-EA(2) tienen

6 y 10 parámetros, respectivamente, y los modelos ADE-EC(1) y ADE-EC(2) tienen 5 y 7 parámetros, respectivamente. Además, Cualquier especificación del modelo ADE(s) se puede utilizar sin importar el espaciamiento temporal entre las mediciones. La estimación de los parámetros en estos modelos precisa del uso de optimización numérica. Sin embargo, a diferencia de los modelos AD, los parámetros en estos modelos tienen restricciones, aún en el caso de la especificación autorregresiva. En particular, las restricciones asociadas al modelo (5) son $\phi_k > 0$, $\sigma^2 > 0$, y $\{\psi\colon\ g(t_j;\psi) > 0\}$. Las restricciones para el modelo (6) son similares pero sustituyendo $\phi_k > 0$ por $\rho_k > 0$.

4.7. Modelos ARIMA

Un modelo ARIMA(s,d,q) generaliza un modelo autorregresivo de medias móviles (ARMA), ya que especifica que las diferencias de orden d entre mediciones consecutivas, en lugar de las propias mediciones, siguen un modelo estacionario ARMA(s,q). Un caso particular es el modelo ARIMA(0,1,0) o modelo de paseo aleatorio

$$y_j - \mathbf{x}_j'\beta = \sum_{t=1}^{j} a_t \qquad (j = 1\ldots,n) \tag{7}$$

donde $a_1,\ldots,a_n$ son variables aleatorias independientes y con una distribución $N(0,v_a)$. Para este proceso, tenemos que $\text{var}(y_j) = jv_a$, $\text{cov}(y_j,y_u) = jv_a$ para $1 \le j \le u \le n$ y $\text{corr}(y_j,y_u) = \sqrt{j/u}$ para $1 \le j \le u \le n$. Por tanto, las varianzas crecen (linealmente) con el tiempo y las correlaciones entre observaciones igualmente espaciadas también crecen (no linealmente) con el tiempo. Este comportamiento es típico en los modelos ARIMA (véase Cryer, 1986, cap. 5). Existen otros dos casos dignos de mencionar en los modelos ARIMA: el modelo ARIMA(0,1,1) [o IMA(1,1)]

$$y_j - \mathbf{x}_j'\beta = y_{j-1} - \mathbf{x}_{j-1}'\beta + a_t - \gamma a_{t-1}$$

y el modelo ARIMA(1,1,0) [o ARI(1,1)]

$$y_j - \mathbf{x}_j'\beta - (y_{j-1} - \mathbf{x}_{j-1}'\beta) = \phi[y_{j-1} - \mathbf{x}_{j-1}'\beta - (y_{j-2} - \mathbf{x}_{j-2}'\beta)] + a_t.$$

Una desventaja de este tipo de modelos es que, para que puedan utilizarse en un contexto de datos longitudinales, el diseño de medición debe ser rectangular y los tiempos de medición equiespaciados entre sí. Sin embargo, existen modelos aplicables al caso de tiempos continuos y no discretos, lo que permitiría que estas restricciones se puedan relajar (véase, por ejemplo, Bergstrom, 1985). En este trabajo, consideraremos sólo uno de estos casos, el proceso de Wiener (WI), que es el análogo en tiempo continuo del modelo de paseo aleatorio. La función de covarianza para un proceso de Wiener está dada por $\text{cov}(y_j,y_k) = v\min(t_j,t_k)$, que coincide con la función de covarianza en (7), para el caso de datos equiespaciados en el tiempo, tomando $v = v_a$ y $t_j = j$. Así

tendremos que, para el modelo WI, la estructura de covarianzas intra-individuo, que tiene un solo parámetro ($v > 0$), se puede escribir como:

$$\mathbf{V} = v \begin{pmatrix} t_1 & t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & t_2 & \cdots & t_2 \\ t_1 & t_2 & t_3 & \cdots & t_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & t_3 & \cdots & t_n \end{pmatrix} \equiv v\mathbf{H}. \tag{8}$$

4.8. Aspectos Computacionales del Ajuste

La primera desventaja de los modelos que utilizan estructuras paramétricas para la matriz de covarianzas intra-individuos ha sido, hasta hace poco, la ausencia de paquetes estadísticos que permitiesen ajustar los mismos. Recientemente, el paquete estadístico SAS, con su PROC MIXED (SAS Institute Inc., 1996) ha logrado que muchos de estos modelos se puedan ajustar y comparar a otros modelos alternativos. Sin embargo, existen algunos modelos que deben ajustarse utilizando otros métodos.

Los modelos SC, AR(1), SCH, ARH(1), HF, TOEP, TOEPH, CA, NE, AD y ADE permiten tener cualquier diseño de medición y el hecho de tener observaciones igualmente espaciadas o diseño de medición rectangular no tiene mayores ventajas, aparte de las que garanticen su ajuste. Así, por ejemplo, para el ajuste del modelo NE, es necesario que el diseño no se aleje mucho de la rectangularidad. Si el diseño de medición es rectangular, y m es suficientemente grande, existen expresiones explícitas para los estimadores MV y MVR de $\mathbf{V}$. Sin embargo, si el diseño de medición no es rectangular, no existen expresiones explícitas para los estimadores MV o MVR. En dicho caso, y dependiendo del alejamiento del diseño rectangular, puede ser difícil o imposible maximizar la verosimilitud. En particular, la función de verosimilitud puede ser muy plana (debido al gran número de parámetros), lo que puede causar, y de hecho causa, problemas de convergencia. Los modelos ARIMA, por otro lado, sólo pueden utilizarse en un contexto de datos longitudinales en el caso de diseño rectangular.

En el caso de los modelos AD, la rectangularidad en el diseño simplifica de manera sustancial el proceso de estimación. Si el diseño de medición es rectangular y m es suficientemente grande, existen expresiones para los estimadores MV o MVR de $\mathbf{V}$ en el caso de un modelo de primer orden (véase Byrne y Arnold, 1983). En el caso de un diseño rectangular y modelos de órdenes superiores a uno, no existen expresiones simples para obtener los estimadores MV y MVR pero se pueden obtener a partir de los elementos de $\mathbf{S}$, utilizando un proceso recursivo, que no requiere del uso de técnicas de optimización numéricas (Johnson, 1989). Sin embargo, en los casos de diseños no rectangulares, debemos utilizar optimización numérica. En estos casos, el modelo AD(1) con especificación de covarianza tiene ventajas computacionales sobre las otras

especificaciones dado que existen fórmulas explícitas para la inversa y el determinante de la matriz de covarianzas de los modelos AD(1) y ADE(1) (Núñez-Antón y otros, 1995, Núñez-Antón, 1997 y Zimmerman y otros, 1998).

Los modelos de coeficientes aleatorios, de simetría compuesta, de simetría compuesta heterogéneo, autorregresivos de primer orden, autorregresivos de primer orden heterogéneo, Huynh-Feldt, Toeplitz, no estructurado, antedependiente no estructurado de orden uno y ciertos modelos ARIMA se pueden ajustar utilizando PROC MIXED, aunque en los modelos CA suelen existir problemas de convergencia en el algoritmo. Sin embargo, mencionaremos que el ajuste de los modelos AR(1), ARH(1), TOEP y TOEPH en SAS sólo se puede realizar si los tiempos en los que se realizan las mediciones son equidistantes y, además, van desde $t_1 = 1$ hasta $t_n = n$.

Es posible ajustar el modelo AD(1) utilizando PROC MIXED, para lo cual es necesario utilizar la opción ANTE(1). La especificación utilizada es la de covarianza, con las restricciones que aseguran que la matriz sea definida positiva. Para que se dé la convergencia, se permite que haya ciertos tipos de no rectangularidad en el diseño de medición. Los modelos AD de órdenes superiores no pueden ajustarse utilizando PROC MIXED. Sin embargo, SAS no puede utilizarse para ajustar los modelos ADE y programas específicos que permitan estos ajustes deben ser utilizados.

Algunos modelos ARIMA pueden ajustarse en PROC MIXED, utilizando los comandos adecuados para que sean aplicados a los datos diferenciados. Por ejemplo, los modelos ARI(1,1) e IMA(1,1) pueden ajustarse utilizando los comandos para ajustar los modelos AR(1) o TOEP(2) en PROC MIXED, respectivamente, a las diferencias de primer orden de las observaciones originales. El modelo WI se puede ajustar directamente a los datos no diferenciados utilizando el comando LIN y especificando $\mathbf{H}$ en la ecuación (8). De forma alternativa, el modelo WI se puede ajustar usando mínimos cuadrados ordinarios para las diferencias de primer orden. Sin embargo, esto no es recomendable debido a que reduce el número de observaciones y, además, elimina cualquier variable explicativa que sea constante dentro de un mismo individuo, como, por ejemplo, los efectos de uno o varios tratamientos.

5. EJEMPLOS: AJUSTE DE MODELOS

5.1. Aspectos Generales del Ajuste

Para ilustrar el ajuste de los modelos introducidos en la Sección 4 y, además, para poder compararlos entre sí y con el modelo de coeficientes aleatorios, en esta sección ajustaremos algunos de estos modelos a los datos introducidos en la Sección 2. A continuación, mencionaremos algunos aspectos generales que utilizaremos en el ajuste de los modelos paramétricos de la estructura de covarianza en el contexto de datos longitudinales:

1. Dado que nuestro principal interés en este trabajo es la modelización de la estructura de covarianzas, utilizaremos un modelo tan saturado como nos sea posible para la estructura de la respuesta media. Puesto que para el *cattle data* tenemos dos posibles tratamientos, utilizaremos el siguiente modelo para la estructura de la respuesta media:

$$E(y_{ij}) = \begin{cases} \mu_{Aj} & \text{si el individuo } i \text{ recibe el tratamiento A} \\ \mu_{Bj} & \text{si el individuo } i \text{ recibe el tratamiento B,} \end{cases} \tag{9}$$

para $i = 1,\ldots,60$ y $j = 1,\ldots,7,8,9,10,11$. Los tiempos han sido reescalados dividiéndolos por 14, que son los días que hay en un periodo de dos semanas. Es decir, los tiempos que utilizaremos en estos datos son $t = 1,\ldots,7,8,9,9.5$. Mencionaremos que, en base a lo que vemos en la Figura 2, es posible que un modelo cuadrático en el tiempo sea una buena alternativa para modelizar la respuesta media en estos datos. Sin enbargo, análisis previos de estos datos que también centraron su interés en la modelización de la estructura de covarianzas (véase, por ejemplo, Kenward, 1987 y Zimmerman y Núñez-Antón, 1997), también utilizaron un modelo similar al de la ecuación (9).

En el caso del *race data*, utilizaremos un modelo parecido al anterior pero considerando que en este caso no tenemos ningún tratamiento y simplemente estamos midiendo la variable de respuesta para cada uno de los tiempos o tramos de la carrera. Así, el modelo utilizado será

$$E(y_{ij}) = \mu_j, \tag{10}$$

para $i = 1,\ldots,80$ y $j = 1,\ldots,10$. Los tiempos que utilizaremos representarán a cada una de las diez secciones (cronológicamente hablando) de la carrera. Es decir, $t = 1,\ldots,10$. Estos datos han sido analizados previamente de forma exploratoria y gráfica por Everitt (1994*a*, 1994*b*). Además, Zimmerman y otros (1998) ajustaron inicialmente un modelo paramétrico cúbico en el tiempo para la estructura de medias y un modelo no estacionario ADE para la estructura de covarianzas intra-individuos. En ninguno de estos análisis previos el objetivo principal de los mismos era el de comparar diferentes estructuras de covarianzas, por lo que, dado que este es el objetivo de este trabajo, hemos decidido mantener el modelo (10) para la estructura de medias.

En cualquier caso, creemos que al utilizar una estructura de medias saturada, y concentrarnos en la modelización de la estructura de covarianzas, no estamos realizando un análisis todo lo riguroso que querríamos. Pensamos que esta propuesta debe complementarse con una posible reducción en la estructura de medias de forma que se

reduzca el número de parámetros presentes en la misma. Por ejemplo, podría plantearse el uso de un modelo paramétrico para esta estructura en cualquiera de los dos conjuntos de datos o, como alternativa, podría contrastarse hipótesis sobre el modelo propuesto que permitan su reducción.

2. Estimaremos los parámetros utilizando el método de máxima verosimilitud residual o restringida (MVR). En este contexto, menos dos veces el logaritmo de la verosimilitud residual para los datos será igual a:

$$\begin{aligned} -2l_R(\theta) &= \sum_{i=1}^{m} \log|\mathbf{V}_i(\theta)| + \sum_{i=1}^{m} [\mathbf{y}_i - \mathbf{X}_i\hat{\beta}(\theta)]'[\mathbf{V}_i(\theta)]^{-1}[\mathbf{y}_i - \mathbf{X}_i\hat{\beta}(\theta)] \\ &\quad + \log|\sum_{i=1}^{m} \mathbf{X}_i'[\mathbf{V}_i(\theta)]^{-1}\mathbf{X}_i| + (N-p)\log 2\pi \end{aligned}$$

   donde $\hat{\beta}(\theta) = (\sum_{i=1}^{m} \mathbf{X}_i'[\mathbf{V}_i(\theta)]^{-1}\mathbf{X}_i)^{-1} \sum_{i=1}^{m} \mathbf{X}_i'[\mathbf{V}_i(\theta)^{-1}]\mathbf{y}_i$.

3. Dado que en la Sección 2 hemos indicado que existe una no estacionariedad en varianza y en correlación, ajustaremos los distintos modelos no estacionarios mencionados en la Sección 4 (es decir, modelo NE, modelo WI, uno o más modelos AD, y uno o más modelos ADE) y los compararemos con los modelos de coeficientes aleatorios lineales (CAL) y cuadráticos (CAC). Usaremos PROC MIXED para ajustar los modelos NE, WI, CAL y CAC. Los modelos ADE y AD de órdenes mayores los ajustaremos utilizando programas en FORTRAN escritos por los autores y usados conjuntamente con las subrutinas IMSL (IMSL, Inc., 1991*a*, 1991*b*). Estos programas, en su proceso de optimización de la función de verosimilitud residual $l_R(\theta)$, utilizan el algoritmo simplex de Nelder y Mead (Nelder y Mead, 1965). Todos los modelos AD y ADE ajustados a los datos tienen una especificación autorregresiva. Además, para poder comparar todos estos modelos entre sí y con los modelos de coeficientes aleatorios, hemos ajustado los modelos Toeplitz (TOEP) y su extensión heterogénea (TOEPH) (sólo para el caso de race data), Huynh-Feldt (HF), además de los modelos estacionarios de simetría compuesta (SC) y autorregresivo de primer orden (AR(1)), y sus extensiones heterogéneas (SCH y ARH(1)).

   Los modelos TOEP y TOEPH, HF, SC y SCH se ajustaron utilizando PROC MIXED. Los modelos AR(1) y ARH(1), casos especiales de los modelos ADE, además del modelo WI, se ajustaron (sólo en el caso del *cattle data* por existir observaciones no equiespaciadas entre sí) utilizando versiones adaptadas de nuestros programas. Los modelos AD, especificación autorregresiva, se ajustaron siguendo las pautas que a este efecto se mencionan en Macchiavelli y Arnold (1994) y Zimmerman y Núñez-Antón (1997).

4. Los ajustes de los modelos paramétricos utilizados para la estructura de covarianzas intra-individuo se compararon a través de dos criterios comúnmente propuestos para este fin. Ambos se encuentran especificados de tal forma que un valor mayor de estos criterios indica un mejor modelo de covarianzas. Los criterios son: el criterio de

información de Akaike AIC $= l_R(\hat{\theta}) - q$, y el criterio de información Bayesiana de Schwarz BIC $= l_R(\hat{\theta}) - \frac{q}{2}\log(N-p)$. En estas ecuaciones, $\hat{\theta}$ es el estimador MVR de θ, y q es la dimensión de θ. Como ya es sabido, el criterio BIC tiene una penalización mayor por los parámetros adicionales en el modelo, por lo que tenderá a favorecer modelos con menos parámetros que los seleccionados utilizando AIC. Otro criterio de comparación de estos ajustes es el contraste de razón de vesosimilitudes residuales (CRVR). Este contraste se realiza haciendo la diferencia entre los valores de $-2l_R$ para dos modelos anidados y comparando este valor con el percentil correspondiente de una distribución χ^2 con grados de libertad iguales a la diferencia en el número de parámetros en las dos estructuras de covarianza que se estén contrastando.

5.2. Análisis del Race Data

Los primeros trabajos que realizan un análisis de tipo gráfico y exploratorio para estos datos son Everitt (1994*a* y 1994*b*). Zimmerman y otros (1998) ajustaron un modelo de pocos parámetros para describir la relación entre los tiempos que cada corredor utilizaba en recorrer cada sección de 10 kilómetros de la carrera y el número de la sección correspondiente ($t = 1,\ldots,10$) y la edad del corredor. El modelo final que se utilizó para la estructura de medias era un modelo lineal en el tiempo y que eliminó la variable edad. Su modelo para la estructura de covarianzas intra-individuos fue un modelo ADE(1) con especificación de covarianza (es decir, ADE-EC(1)), y una varianza que cambiaba de forma lineal con el tiempo (véase el modelo (6)). Este análisis previo, aunque muy limitado, nos puede dar alguna pista sobre los modelos que podrían ser adecuados para estos datos. Hay una clara presencia de no estacionariedad en varianza y en correlación, por lo que los modelos no estacionarios, entre los que se encuentra el de coeficientes aleatorios, serán seguramente las mejores opciones a ajustar.

El contraste de verosimilitudes de Gabriel (1962) para determinar el orden adecuado de antedependencia sugiere que un modelo de orden dos será suficiente para estos datos. Es decir, un modelo AD(2) será mejor que modelos de órdenes inferiores y tan bueno como modelos de órdenes superiores (es decir, mayores que dos). Así, tendremos que al menos ajustar modelos estructurados y no estructurados de orden dos y uno. Sin embargo, dado que el contraste de Gabriel es un contraste aproximado, también ajustaremos modelos no estructurados de órdenes tres y cuatro.

Para las varianzas específicas, en la especificación autorregresiva estructurada, utilizaremos modelos lineales (ADEL) y cuadráticos (ADEQ) en el tiempo (véase el modelo (5)). Es decir, para el caso del modelo ADEQ(1), tendremos que:

$$\begin{aligned} \phi_j &= \phi_1^{f(t_j;\lambda_1)-f(t_{j-k};\lambda_1)} \qquad (j=2,\ldots,n), \\ \sigma_j^2 &= \sigma^2\left(1+\psi_1 t_j+\psi_2 t_j^2\right) \qquad (j=2,\ldots,n). \end{aligned}$$

En este caso, tendremos que el modelo tiene 6 parámetros, de tal forma que $\theta = (\phi_1, \lambda_1, \sigma^2, \sigma_1^2, \psi_1, \psi_2)$. Así, en el caso del modelo ADEL(1), $\psi_2 = 0$ y, por tanto, el modelo tendrá 5 parámetros.

Tabla 4. Resultados del análisis de las distintas propuestas de modelos para la estructura de covarianzas intra-individuo para el *race data*.

Estructura	q	AIC	BIC	$-2l_R$	MC	ν^*	χ^2	Pr $> \chi^2$
SC	2	−2673.6	−2678.2	5343.1				
AR(1)	2	−2550.6	−2555.3	5097.2				
AD(0)	10	−2897.8	−2921.1	5775.5				
WI	1	−2532.5	−2534.9	5063.1				
SCH	11	−2558.6	−2584.3	5095.2	SC	9	247.9	0.00
CAL	4	−2566.6	−2575.9	5125.1	SC	2	218.0	0.00
CAC	7	−2525.0	−2541.4	5036.1	CAL	3	89.0	0.00
HF	11	−2634.7	−2660.4	5247.5	SC	9	95.6	0.00
ARH(1)	11	−2395.7	−2421.4	4769.4	AR(1)	9	327.8	0.00
TOEP	10	−2539.3	−2562.7	5058.6	AR(1)	8	38.6	0.00
TOEPH	19	−2396.9	−2441.3	4755.8	ARH(1)	9	13.6	0.09
ADEL(1)	5	−2504.5	−2516.2	4999.0	AD(1)	14	283.0	0.00
ADEQ(1)	6	−2443.3	−2457.3	4874.6	AD(1)	13	158.6	0.00
ADEL(2)	9	−2469.5	−2490.5	4920.9	ADEL(1)	4	78.1	0.00
ADEQ(2)	10	−2422.0	−2445.3	4823.9	ADEL(2)	1	97.0	0.00
ADEL(5)	27	−2445.3	−2508.4	4836.3	ADEL(2)	18	84.6	0.00
ADEQ(5)	28	−2421.0	−2486.4	4786.0	ADEL(5)	1	50.3	0.00
AD(1)	19	−2377.0	−2421.4	4716.0	ARH(1)	8	53.4	0.00
AD(2)	27	−2361.1	−2424.2	4668.2	AD(1)	8	47.8	0.00
AD(3)	34	−2357.3	−2436.7	4646.6	AD(2)	7	21.6	0.00
AD(4)	40	−2361.0	−2454.4	4642.0	AD(3)	6	4.6	0.60
NE	55	−2365.2	−2493.7	4620.4	AD(3)	21	26.2	0.20

En la Tabla 4 y posteriores, q es el número de parámetros en la estructura de covarianzas, MC es el modelo anidado con el que se compara el modelo en cuestión, ν^* son los grados de libertad para el contraste de verosimilitudes residuales, y la última columna contiene el valor de probabilidad que permitirá o no rechazar el modelo en la hipótesis nula. Para el cálculo de BIC, $N = 800$ y $p = 10$.

Si la selección del modelo se basa en el criterio AIC, el mejor modelo ajustado es el AD(3), con $q = 34$ parámetros. Los modelos AD(2) y AD(4), que tienen $q = 27$ y $q = 40$ parámetros, respectivamente, son también modelos con un buen ajuste. Si, por el contrario, usamos el criterio BIC, los mejores modelos ajustados son ARH(1) y AD(1), que tienen, respectivamente $q = 11$ y $q = 19$ parámetros. El contraste CRVR con 19-11=8

Tabla 5. Resultados del análisis de las distintas propuestas de modelos para la estructura de covarianzas intra-individuo para el *cattle data*-tratamiento A.

Estructura	q	AIC	BIC	$-2l_R$	MC	ν^*	χ^2	P
SC	2	−1192.2	−1196.0	2380.4				
AR(1)	2	−1052.9	−1056.7	2101.8				
AD(0)	11	−1366.0	−1386.7	2710.0				
WI	1	−1053.7	−1055.6	2105.4				
SCH	12	−1172.3	−1194.9	2320.6	SC	10	59.8	0.00
HF	12	−1190.0	−1212.6	2355.9	SC	10	24.5	0.01
CAL	4	−1080.8	−1088.3	2153.6	SC	2	226.8	0.00
CAC	7	−1051.2	−1064.4	2088.3	CAL	3	65.3	0.00
ARH(1)	12	−1057.0	−1079.6	2089.9	AR(1)	10	11.8	0.30
ADE(1)	4	−1048.9	−1056.4	2089.7	AR(1)	2	12.0	0.00
ADE(2)	8	−1051.2	−1066.2	2086.4	ADE(1)	4	3.4	0.49
AD(1)	21	−1055.9	−1095.4	2069.8	ARH(1)	9	20.2	0.02
					ADE(1)	17	19.9	0.28
AD(2)	30	−1055.6	−1112.1	2051.2	AD(1)	9	18.6	0.03
					ADE(2)	22	35.2	0.04
NE	66	−1075.7	−1200.0	2019.4	AD(2)	36	31.7	0.67

grados de libertad rechaza el modelo ARH(1) en favor del modelo AD(1). Los distintos contrastes realizados en la Tabla 4 sugieren que un modelo adecuado para estos datos sería el modelo AD(3) con $q = 34$ parámetros, muchos menos parámetros de los que tiene el modelo NE. Para la realización de estos contrastes, hemos comparado sucesivamente el modelo AD(3) con los modelos NE, AD(4) and AD(2), de tal forma que el modelo AD(3) fue el seleccionado. Mencionaremos que, en este caso, los modelos estructurados no proporcionaron un buen ajuste y no creemos necesario compararlos con los modelos no estructurados. Como ilustración, Zimmerman y Núñez-Antón (1997) escriben de forma explícita las ecuaciones recursivas para el ajuste de un modelo AD(2) con especificación autorregresiva.

En resumen, debemos mencionar que: (*i*) Los modelos de coeficientes aleatorios CAL y CAC no son adecuados para estos datos y no deben utilizarse; (*ii*) si los comparamos con los modelos AD o con otros modelos alternativos, ninguno de los modelos ADE da un ajuste adecuado para estos datos. Sin embargo, podríamos pensar que, dada la naturaleza de los modelos de coeficientes aleatorios, una comparación más justa entre estos modelos y los demás habría tenido que considerar un modelo de coeficientes aleatorios con media estructurada (por ejemplo, lineal o cuadrática en el tiempo). Por simple comparación, hemos ajustado los modelos CAL y CAC estructurando la media con una dependencia lineal en el tiempo y, dado que estos modelos tienen un peor

ajuste que los modelos con medias no estructuradas, no pueden competir con los mejores modelos ajustados para estos datos. Finalmente, indicaremos a nivel interpretativo que las estructuras antedependientes de primer orden son fácilmente motivables por el simple hecho de observar si las primeras subdiagonales o superdiagonales en la matriz de correlaciones presentan correlaciones muestrales que crecen o decrecen con el paso del tiempo. Hemos comprobado que esta intuición no es fácilmente extendible a los modelos de órdenes superiores. A nuestro entender, la única forma de «evaluar» si el modelo realmente ajusta los datos es obtener la matriz de varianzas y covarianzas para el modelo seleccionado y compararla con la empírica correspondiente. No hemos incluido esta comparación aquí, pero al igual que en los trabajos previos en este área, la hemos realizado y el ajuste, creemos, es bastante aceptable.

Tabla 6. Resultados del análisis de las distintas propuestas de modelos para la estructura de covarianzas intra-individuo para el *cattle data*-tratamiento B.

Estructura	q	AIC	BIC	$-2l_R$	MC	ν^*	χ^2	P
SC	2	−1188.9	−1192.7	2373.9				
AR(1)	2	−1104.8	−1108.6	2205.6				
AD(0)	11	−1345.9	−1366.6	2669.9				
WI	1	−1097.9	−1099.8	2193.9				
SCH	12	−1140.9	−1163.5	2257.8	SC	10	116.1	0.00
HF	12	−1184.6	−1207.1	2345.1	SC	10	28.8	0.00
CAL	4	−1128.7	−1136.2	2249.4	SC	2	124.5	0.00
CAC	7	−1099.8	−1113.0	2185.6	CAL	3	63.8	0.00
ARH(1)	12	−1054.7	−1077.2	2085.2	AR(1)	10	120.4	0.00
ADE(1)	4	−1074.0	−1081.5	2139.9	AR(1)	2	66.7	0.00
ADE(2)	8	−1077.3	−1092.3	2138.5	ADE(1)	4	1.4	0.84
ADE(3)	13	−1076.4	−1100.9	2126.9	ADE(2)	5	11.7	0.04
AD(1)	21	−1043.9	−1083.5	2045.8	ARH(1)	9	39.4	0.00
					ADE(1)	17	94.1	0.00
AD(2)	30	−1046.8	−1103.3	2033.6	AD(1)	9	12.3	0.20
					ADE(2)	22	105.0	0.00
AD(3)	38	−1045.0	−1116.5	2013.9	AD(2)	8	19.6	0.01
					AD(1)	17	31.9	0.02
					ADE(3)	25	113.0	0.00
NE	66	−1055.6	−1179.9	1979.2	AD(3)	28	34.7	0.18

Concluimos este análisis mencionando que los modelos antedependientes no estructurados son modelos adecuados para estos datos lo que, de alguna forma, ya fue indicado en Zimmerman y otros (1998). Si nos fijamos en el número de parámetros y el ajuste dado por los criterios AIC y BIC, éstos son alternativas válidas y relevantes para el

modelo NE u otros modelos. Indicaremos además que todos los modelos con muchos parámetros son, en este caso, más adecuados que lo modelos más simples, tales como los modelos SC, AD(0) o AR(1).

5.3. Análisis del Cattle Data

Contrastamos, en primer lugar, la hipótesis de igualdad de las dos matrices de covarianzas intra-individuos para los distintos grupos utilizando el contraste clásico de razón de verosimilitudes. La hipótesis de igualdad se rechaza de forma clara ($P = 0.02$). Consecuentemente, utilizamos estructuras paramétricas distintas para modelizar las matrices de covarianzas intra-individuo para cada uno de los dos grupos en estos datos. Los análisis posteriores que se realicen con estos datos (por ejemplo, análisis que permitan reducir la estructura de medias), deben utilizar, para cada grupo, la estructura de covarianzas intra-individuo que haya dado el «mejor» ajuste. Dado que éste no es el enfoque de este trabajo, proponemos este análisis para futuros trabajos y, por tanto, no lo incluimos en el presente.

Kenward (1987) utilizó un modelo AD para analizar estos datos pero a lo largo de todo su análisis utilizó una estructura de covarianzas común para los dos grupos, lo cual no es correcto. Es por eso que no haremos mayores comentarios a los resultados o modelos ajustados, aunque si mencionaremos que sus conclusiones indicaron que un modelo AD de orden dos es el adecuado para la estructura común de covarianzas en los dos grupos para estos datos. Es decir, usando el contraste de Gabriel (1962), los modelos de órdenes inferiores a dos no son adecuados, y el modelo de orden dos lo hace tan bien como los de órdenes superiores. Además, Kenward (1987) encontró que existía una diferencia significativa entre los distintos tratamientos.

Zimmerman y Núñez-Antón (1997) han ajustado modelos AD y ADE a estos datos. En este trabajo extenderemos el de Zimmerman y Núñez-Antón (1997) al centrarnos en ajustar estructuras de covarianzas alternativas, incluyendo las estructuras de coeficientes aleatorios y algunas otras alternativas que ya hemos mencionado en la Sección 4. El contraste de Gabriel (1962) sugiere que los órdenes adecuados de antedependencia para los modelos UAD que pueden usarse en los tratamientos A y B son dos y tres (al igual que modelos de órdenes superiores a éstos), respectivamente. Por esto, ajustamos modelos AD y ADE de órdenes menores o iguales a los sugeridos por este contraste. Los modelos ADE utilizan una especificación autorregresiva con varianzas específicas constantes (ver ecuación (5)). Es decir, en este caso tendremos que, por ejemplo, para el modelo ADE(1):

$$\begin{aligned} \phi_j &= \phi_1^{f(t_j;\lambda_1)-f(t_{j-1};\lambda_1)} \qquad (j = 2,\ldots,n) \\ \sigma_j^2 &= \sigma^2 \qquad (j = 2,\ldots,n) \end{aligned}$$

En este caso, tenemos que el modelo anterior tiene 4 parámetros, de tal forma que $\theta = (\phi_1, \lambda_1, \sigma^2, \sigma_1^2)$.

En las Tablas 5 y 6, tenemos la información de los modelos ajustados para el *cattle data*, grupos A y B, respectivamente. Para el cálculo de BIC, $N = 330$ y $p = 11$. En el caso del grupo A, si la selección del modelo se basa en el criterio AIC, el mejor modelo ajustado es el ADE(1), con $q = 4$ parámetros, aunque también los modelos AR(1) y ADE(2) . Si, por el contrario, usamos el criterio BIC, el mejor modelo ajustado es el WI, con $q = 1$ parámetro, aunque también los modelos ADE(1) y AR(1) tienen un buen ajuste. El contraste de CRVR entre estos modelos también sugiere el uso del modelo ADE(1). Sin embargo, este modelo no puede compararse con el modelo WI usando este contraste, ya que no son modelos anidados. Además, los contrastes escalonados que se realizan en la Tabla 5 sugieren que el mejor modelo para estos datos es el AD(2), con $q = 30$ parámetros. Un contraste entre el modelo ADE(1) y el modelo AD(2) rechaza el primero de ellos. Para la realización de estos contrastes, hemos comparado sucesivamente el modelo AD(2) con los modelos NE, AD(1) y ADE(2), de tal forma que el modelo AD(2) fue el seleccionado.

En el caso del grupo B, si la selección del modelo se basa en el criterio AIC y CRVR, los mejores modelos ajustados son los modelos AD, mientras que los modelos ARH(1), ADE(1) y AD(1) son los mejores si nos basamos en el criterio BIC. Los contrastes escalonados que se realizan en la Tabla 6 sugieren que el mejor modelo para estos datos es el AD(3), con $q = 38$ parámetros. Para la realización de estos contrastes, hemos comparado sucesivamente el modelo AD(3) con los modelos NE, AD(2) y ADE(3), de tal forma que el modelo AD(3) fue el seleccionado. Además, debemos mencionar que, en este caso, los modelos de coeficientes aleatorios CAL y CAC no son adecuados para estos datos y no deben utilizarse.

En resumen, los modelos antedependientes no estructurados son modelos adecuados para estos datos lo que, de alguna forma, ya fue indicado en Zimmerman y Núñez Antón (1997). Si nos fijamos en el número de parámetros y el ajuste dado por los criterios AIC y BIC, éstos son alternativas válidas y relevantes para el modelo NE u otros modelos. Indicaremos además que, en este caso, todos los modelos con muchos parámetros son más adecuados que lo modelos más simples.

Finalmente y como detalle adicional comentaremos que si utilizamos como modelos de covarianzas para cada uno de los grupos los modelos AD(2) y AD(3), respectivamente, y llevamos a cabo un contraste de diferencias significativas en las respuestas de los distintos grupos (tratamientos), concluiremos que existe una diferencia estadísticamente significativa ($P = 0.00$) en las respuestas de los distintos tratamientos.

6. CONCLUSIONES

Hemos estudiado los diferentes modelos que se pueden proponer para la estructura de covarianzas intra-individuos en el contexto de datos longitudinales. Uno de nuestras principales motivaciones era la de ilustrar la posible existencia de no estacionariedad en varianza y/o en correlación en esta matriz de covarianzas y los posibles modelos que permiten explicar estos comportamientos no estacionarios. Entre ellos, el más utilizado en la práctica es el modelo de coeficientes aleatorios, aunque no se le ha reconocido el mérito adecuado en la modelización de comportamientos de tipo no estacionario. Así, comparamos los ajustes de estos modelos con los de otros que permiten la modelización de estos comportamientos y con los modelos estacionarios más utilizados a través de dos estudios longitudinales que han sido frecuentemente citados en la literatura de datos longitudinales.

Hemos motivado el uso de todos los modelos en la Sección 4 a través de las características presentes en la matriz de covarianzas intra-individuos para cada uno de ellos, lo que nos lleva a aconsejar o desaconsejar su uso en algunos casos. Hemos hablado de modelos simples y sencillos, como los modelos SC, WI o AR(1), entre otros, además de hablar de modelos flexibles y en cierto sentido simples, como los modelos ADE, que pueden tener una utilidad importante en ciertas aplicaciones. También hemos mencionado los modelos más complicados o generales, como los modelos NE o AD, entre otros, que siempre representarán una posibilidad ya conocida de modelizar esta matriz en datos longitudinales. Esto es, si no se tiene una idea clara del modelo que se puede utilizar para modelizar el comportamiento de esta estructura en los datos, estos modelos son suficientemente generales y, por tanto, pueden utilizarse, aunque, obviamente, el precio a pagar es la alta dimensionalidad del vector de parámetros.

No hemos modelizado la estructura de medias de forma paramétrica explícita debido al interés especial que tenemos en este trabajo en la modelización de la estructura de covarianzas intra-individuos. A este respecto indicaremos que, en nuestra opinión, un análisis completo de este tipo de datos requiere plantear un modelo inicial bastante general para la estructura de medias, modelo que se puede basar en un análisis previo de las medias en los distintos tiempos. A continuación, y con esta estructura de medias, se deben proponer distintas estructuras de covarianza intra-individuos y seleccionar la que mejor ajuste los datos. Finalmente, con la estructura seleccionada se realizan contrastes de reducción en la estructura de medias, obteniendo de esta forma un modelo final para ambas estructuras. En este trabajo hemos utilizado una estructura bastante general para la media y nos hemos centrado en la modelización de la estructura de covarianzas intra-individuos de los datos.

Hemos ajustado distintos modelos a dos conjuntos de datos en los que los de coeficientes aleatorios demostraron no ser adecuados. Estas mismas conclusiones se han obtenido previamente para otros conjuntos de datos distintos (Núñez-Antón, 1993).

En los distintos modelos ajustados tenemos que mencionar su comparación no sólo en términos de ajuste, sino también en términos de flexibilidad y número de parámetros. El modelo NE es el más flexible y es el que más parámetros tiene, con $O(n^2)$ parámetros. Los modelos AD son un poco menos flexibles pero tienen menos parámetros (es decir, $O(n)$ parámetros). Los modelos ADE, ARIMA y de coeficientes aleatorios (CA), al igual que todos los modelos estacionarios, son muy estructurados (con $O(1)$ parámetros) y muy poco flexibles.

El espaciamiento irregular entre observaciones y la no rectangularidad en el diseño de medición no presentan problema alguno para los modelos ADE o CA, pero hay que tener cuidado con una o ambas condiciones cuando se ajustan los modelos NE, AD o ARIMA. Hay que mencionar que existe un claro problema para ajustar algunos de estos modelos, especialmente cuando se trata de observaciones irregularmente espaciadas. Cuando éste es el caso, PROC MIXED sólo se puede usar para los modelos en los que la estructura no dependa de los tiempos de observación (por ejemplo, CA, HF y NE). El resto de modelos deben ajustarse a través de programas que se deben escribir en algún lenguaje específico, en nuestro caso FORTRAN.

Los modelos antedependientes son pocos conocidos, muy útiles y claramente superiores a los modelos de coeficientes aleatorios en contextos no estacionarios para datos longitudinales. Su uso debe al menos empezar a ser considerado por los estadísticos como una alternativa real, cuando así lo sean, a los modelos CA.

AGRADECIMIENTOS

El trabajo de Vicente Núñez Antón ha sido financiado por los proyectos de investigación PB98-0149 de la Dirección General de Enseñanza Superior e Investigación Científica del Ministerio Español de Educación y Cultura, UPV 038.321-HA129/99 de la Universidad del País Vasco/Euskal Herriko Unibertsitatea y PI-1999-46 del Gobierno Vasco. El trabajo de Dale L. Zimmerman ha sido financiado parcialmente por el proyecto de investigación 9628612 de la National Science Foundation. Los autores agradecen los comentarios del editor y de un evaluador que han mejorado de forma importante la presentación de este trabajo.

REFERENCIAS

Bergstrom, A. R. (1985). «The estimation of parameters in nonstationary higher-order continuous-time dynamic models». *Econometric Theory*, 1, 369-385.

Byrne, P. J. & Arnold, S. F. (1983). «Inference about multivariate means for a nonstationary autoregressive model». *Journal of the American Statistical Association*, 78, 850-855.

Crowder, M. J. & Hand, D. J. (1990). *Analysis of Repeated Measures*. London: Chapman & Hall.

Cryer, J. D. (1986). *Time Series Analysis*. Boston: PWS-Kent.

Diggle, P. J. (1988). «An approach to the analysis of repeated measures». *Biometrics*, 44, 959-971.

Diggle, P. J. (1990). *Time Series: A Biostatistical Introduction*. Oxford: Oxford University Press.

Diggle, P. J., Liang, K. Y. & Zeger, S. L. (1994). *Analysis of Longitudinal Data*. New York: Oxford University Press.

Everitt, B. S. (1994*a*). «Exploring multivariate data graphically: a brief review with examples». *Journal of Applied Statistics*, 21, 63-94.

—— (1994*b*). *A Handbook of Statistical Analysis Using S-Plus*. London: Chapman & Hall.

Gabriel, K. R. (1962). «Ante-dependence analysis of an ordered set of variables». *Annals of Mathematical Statistics*, 33, 201-212.

Huynh, H. & Feldt, L. S. (1970). «Conditions under which mean square ratios in repeated measurements designs have exact $\mathcal{F}$-distributions». *Journal of the American Statistical Association*, 65, 1582-1589.

IMSL, Inc. (1991*a*). *Fortran Subroutines for Mathematical Applications. MATH/LIBRARY Version 2.0*. Houston, Texas: IMSL, Inc.

—— (1991*b*). *Fortran Subroutines for Mathematical Applications. STAT/LIBRARY Version 2.0*. Houston, Texas: IMSL, Inc.

Jennrich, R. L. & Schluchter, M. D. (1986). «Unbalanced repeated-measures models with structured covariance matrices». *Biometrics*, 42, 805-820.

Johnson, K. L. (1989). «Higher-order antedependence models». Unpublished Ph. D. Thesis. Department of Statistics, Pennsylvania State University.

Jones, R. H. (1990). «Serial correlation or random subject effects?» *Communication in Statistics, Simulation and Computation*, 19, 1105-1123.

—— (1993). *Longitudinal Data with Serial Correlation: A State-Space Approach*. London: Chapman & Hall.

Jones, R. H. & Boadi-Boateng, F. (1991). «Unequally spaced longitudinal data with AR(1) serial correlation». *Biometrics*, 47, 161-175.

Kenward, M. C. (1987). «A method for comparing profiles of repeated measurements». *Applied Statistics*, 36, 296-308.

Laird, N. M. (1988). «Missing data in longitudinal studies». *Statistics in Medicine*, 7, 305-315.

Laird, N. M. & Ware, J. H. (1982). «Random effects models for longitudinal data». *Biometrics*, 38, 963-974.

Macchiavelli, R. E. & Arnold, S. F. (1994). «Variable order ante-dependence models». *Communications in Statistics, Theory and Methods*, 23, 2683-2699.

Muñoz, A., Carey, V., Schouten, J. P., Segal, M. & Rosner, B. (1992). «A parametric family of correlation structures for the analysis of longitudinal data». *Biometrics*, 48, 733-742.

Nelder, J. A. & Mead, R. (1965). «A simplex method for function minimization». *The Computer Journal*, 7, 308-313.

Núñez-Antón, V. (1993). «Analysis of longitudinal data with unequally spaced observations and time-dependent correlated errors». Unpublished Ph. D. Thesis. Department of Statistics and Actuarial Science, The University of Iowa.

—— (1997). «Longitudinal data analysis: non-stationary error structures and antedependent models». *Applied Stochastic Models and Data Analysis*, 13, 279-287.

Núñez-Antón, V. & Woodworth, G. G. (1994). «Analysis of longitudinal data with unequally spaced observations and time-dependent correlated errors». *Biometrics*, 50, 445-456.

Núñez-Antón, V., El Barmi, H. & Zimmerman, D. L. (1995). «Una nota sobre matrices de covarianzas con inversas tridiagonales». *Estadística Española*, 37, 139, 201-215.

Rao, C. R. (1959). «Some problems involving linear hypotheses in multivariate analysis». *Biometrika*, 46, 49-58.

Reinsel, G. (1982). «Multivariate repeated-measurement or growth models with multivariate random-effects covariance structure». *Journal of the American Statistical Association*, 77, 190-195.

Rutter, C. M. & Elashoff, R. M. (1994). «Analysis of longitudinal data: random coefficient regression modelling». *Statistics in Medicine*, 13, 1211-1231.

SAS Institute Inc. (1996). *SAS/STAT Software: Changes and Enhancements through Release 6.11*. Cary, North Carolina: SAS Institute Inc.

Verbeke, G. & Molenberghs, G. (1997). *Linear Mixed Models in Practice. A SAS-Oriented Approach*. Lecture Notes in Statistics Nº. 126. New York: Springer-Verlag.

Wolfinger, R. D. (1996). «Heterogeneous variance-covariance structures for repeated measures». *Journal of Agricultural, Biological, and Environmental Health*, 1(2), 205-230.

Zimmerman, D. L. & Núñez-Antón, V. (1997). «Structured antedependence models for longitudinal data». In *Modelling Longitudinal and Spatially Correlated Data. Methods, Applications, and Future Directions*. (T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russell-Cohen, W. G. Warren, and R. Wolfinger, Eds.) Lecture Notes in Statistics Nº. 122, 63-76. New York: Springer-Verlag.

Zimmerman, D. L., Núñez-Antón, V. & El Barmi, H. (1998). «Computational aspects of likelihood-based estimation of first-order antedependence models». *Journal of Statistical Computation and Simulation*, 60, 67-84.

ENGLISH SUMMARY

MODELLING LONGITUDINAL DATA WITH NONSTATIONARY COVARIANCE STRUCTURES: RANDOM COEFFICIENTS MODELS VERSUS ALTERNATIVE MODELS

VICENTE NÚÑEZ-ANTÓN*
DALE L. ZIMMERMAN**

An important theme of longitudinal data analysis in the past two decades has been the development and use of explicit parametric models for the data's variance-covariance structure. However, nonstationary covariance structures had not been analyzed in detail for longitudinal data mainly because the existing applications did not require their use. There has been a large amount of recently proposed models but most of them are second-order stationary. A few, however, are flexible enough to accommodate nonstationarity, that is, nonconstant variances and/or correlations which are not only a function of the elapsed time between measurements. We study some of these proposed models and compare them to the random coefficients models, evaluating the relative strengths and limitations of each model, emphasizing when it is inappropriate or unlikely to be useful. We present two examples to illustrate the fitting and comparison of the models and to demonstrate that nonstationary longitudinal data can be modelled effectively and, in some cases, quite parsimoniously. In these examples the antedependence models generally prove to be superior and the random coefficients models prove to be inferior.

Keywords: Antedependence, Arima models, AIC, BIC, covariance structures, residual maximum likelihood, mixed models

AMS Classification (MSC 2000): 62J05, 62F10, 62P10

* Departamento de Econometría y Estadística (E.A. III). Universidad del País Vasco. Avenida Lehendakari Aguirre, 83. 48015 Bilbao. E-mail: vn@alcib.bs.ehu.es.

** Department of Statistics and Actuarial Science. The University of Iowa. Iowa City, Iowa 52242. Estados Unidos.

– Received April 2000.
– Accepted January 2001.

1. INTRODUCTION

An important theme of models in regression analysis and, specially, about their use to analyze longitudinal data, have been the parallel development and use of explicit parametric models for the data's variance-covariance structure and random coefficients models. Longitudinal data consists of measuring along time a given characteristic on each of the experimental units, normally allocated to different groups or treatments.

However, one of the main issues we want to mention has to do with the fact that random coefficients models have been usually regarded as models for regressions of response on time (and, possibly, on other covariates) that vary across subjects, instead of using the possibility of considering these models as a way to explain different phenomena in the within-subject structure of the data. In any case, random coefficients models can be used and, indeed, be very useful in longitudinal data analysis.

Compared to various analysis-of-variance methods, which ignore the covariance structure, and to the classical multivariate approach, which estimates the covariance matrix but imposes no structure on it (beyond that required for positive definiteness), parametric covariance modelling has several advantages. First, it generally results in more efficient estimation of parameters in the data's mean structure, which are usually of primary interest. Second, it yields more appropriate estimates of the standard errors of those estimated mean parameters. Third, in many cases it can deal effectively with missing data and with data for which the measurement times are not common across subjects. Finally, it can be employed even when the number of measurement times is large relative to the number of subjects.

Perhaps the most prevalent kind of covariance structure exhibited by longitudinal data is serial correlation, i.e. within-subject sample correlations that decrease as the elapsed time between measurements increases. The most popular parametric models for serial correlation are stationary autoregressive (AR) models and other parsimonious second-order stationary models. In these models, variances are constant over time and correlations between measurements equidistant in time are equal. When the sample variances and correlations comport with these assumptions, stationary models are generally very useful. When this is not so, however, the use of a stationary model is inadvisable. Instead, the researcher should consider a model flexible enough to accommodate nonstationarity.

If nonstationarity is manifested by nonconstant variances only, options for analysis include transforming the data to stabilize the variance or generalizing stationary models to allow for heterogeneous variances. Heterogeneous extensions of several stationary models are described and fit to data by Wolfinger (1996). These options may not be sufficient, however, when nonstationarity is also manifested by the correlations. There

are, nevertheless, several alternative models that are applicable to longitudinal data that exhibit nonstationarity in their correlations and variances. Perhaps the most obvious of these is the completely unstructured model of the classical multivariate approach. In many cases, however, the data's nonstationarity may possess an structure capable of being modelled with relatively few parameters, and to ignore this would forego the advantages of parametric modelling noted previously. One family of parametric models that can accommodate nonstationary correlations is a generalization of AR models known as antedependence models: general or unstructured and structured. Another, more well-known, nonstationary generalization of AR models are the autoregressive integrated moving average (ARIMA) models (e.g., see Diggle, 1990). One final possibility, we have already briefly mentioned, are the random coefficients models. These models are typically used for regressions on a continuous response variable that changes with time and that varies between individuals. Thus these models can actually explain certain types of nonstationarity.

Each of the models just described, besides the stationary models, has been proposed for use with nonstationary longitudinal data by at least one author, but such proposals have almost always occurred in isolation, apart from consideration of the other models. The general idea has been that, when there is a problem with the fitting of a given proposed model possibly due to the existing nonstationarity in the data, a random coefficients model is automatically adjusted instead. Consequently, the merits of each have never been systematically evaluated and virtually no guidelines exist as to their relative usefulness. In this article, we examine and compare these models. We consider their strengths and limitations, emphasizing when each is inappropriate or problematic and, specially, comparing these alternative models to the random coefficients model. We present two examples that illustrate the fitting and comparison of the models. The examples also demonstrate that nonstationary longitudinal data can be modelled effectively, and in some cases quite parsimoniously, with appropriate parametric models.

2. DATA SETS

Data from two longitudinal studies serve to motivate the consideration of nonstationary models. These data will also be used to illustrate the fitting and comparison of the different models for the within-subjects data structure. The *race data* consist of the «split» times for each of 80 competitors in each 10-km section of a 100-km race held in 1984 in the United Kingdom. Measurement times are evenly spaced and common to all subjects in the study. Thus, the data are rectangular. The objective of our analysis is to find a parsimonious model that adequately describes how competitor's performance on each 10-km section is related to the section number $(t = 1, 2, \ldots, 10)$ and to the performance on previous sections.

The *cattle data* come from an experiment reported by Kenward (1987). Cattle receiving one of two intestinal parasite treatments, say A and B, were weighed 11 times over a 133-day period. Thirty animals received treatment A and thirty received treatment B. The first 10 measurements on each animal were made at two-week intervals and the final measurement was made one week after the tenth. No observations are missing. Although times are not equally spaced (due to the shorter interval before the last measurement), the measurement schedule is rectangular. We wish to study how cattle growth is affected by the treatments.

The sample variances and correlations corresponding to these longitudinal data sets show several interesting characteristics. First, the variances are not homogeneous, but instead tend to increase over time. Second, the correlations are all positive. Third, serial correlation appears to be present, as correlations within any given column tend to decrease towards zero (unless they are close to zero initially). Finally, correlations lagged the same number of observations apart are not constant. Rather, they tend to increase early in the study before levelling off, or in some cases (e.g. treatment B cattle data) decreasing slightly, later in the study.

A battery of power transformations was attempted for each data set with the aim of variance stabilization. These efforts met with only limited success, which is not surprising given that in several cases the variances do not appear to be smooth functions of the mean. Even in those cases where a transformation successfully stabilized the variance, the nonstationary behavior of the correlations persisted after transformation.

3. THE GENERAL MODEL

Suppose that repeated measurements of a continuous response variable are observed over time on each of m «subjects». Let $\mathbf{y}_i$ be the vector of n_i measurements on the ith subject and let $\mathbf{t}_i$ be the corresponding vector of measurement times. Suppose also that we observe a p-vector of covariates, $\mathbf{x}_{ij}$, associated with y_{ij}. Put $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_m')'$, $\mathbf{t} = (\mathbf{t}_1', \ldots, \mathbf{t}_m')'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})'$, $\mathbf{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_m')'$, and $N = \sum_{i=1}^{m} n_i$. We refer to the set of measurement times in the study as the measurement schedule. We impose no restrictions on the measurement schedule; in general the measurement times may be unequally spaced within a subject and may differ across subjects. If measurement times are common across subjects, we call the measurement schedule *rectangular*. Thus, the measurement schedule of the race and the cattle data is rectangular. The extent of the measurement schedule's departure from rectangularity has important implications for modelling the covariance structure, as will be seen subsequently.

Several general modelling assumptions now provide a framework for parametric modelling of the covariance structure. These assumptions yield the model $\mathbf{y} \sim MVN\,(\mathbf{X}\beta, \Sigma)$, where β is a p-vector of fixed, unknown, and typically unrestricted parameters, and

the $N \times N$ unknown covariance matrix Σ is block diagonal, with non-zero blocks $\mathbf{V}_i = \text{var}(\mathbf{y}_i)$. In the case of a rectangular measurement schedule, the $\mathbf{V}_i$ are all equal. The parametric modelling approach now proceeds with the postulation of a parametric model for Σ: $\mathbf{y} \sim MVN(\mathbf{X}\beta, \Sigma(\mathbf{t}, \theta))$, where θ is a q-vector of unknown parameters, restricted to a parameter space Θ which is either the set of all θ-vectors for which Σ is positive definite or some subset of that set. Note that the elements of Σ are permitted to be functions of $\mathbf{t}$ but not of $\mathbf{X}$. Note further that Σ is positive definite if and only if $\mathbf{V}_i$ is positive definite for every i. Estimation of the parameters β and θ of this model is carried out by maximum likelihood (ML) or residual maximum likelihood (REML), which often must be implemented using numerical optimization routines. Further specifics on the fitting of the models and on comparing the fitted models are deferred to the examples. We now briefly describe the random coefficients and the antedependence models.

4. PARAMETRIC MODELS FOR THE WITHIN-SUBJECTS COVARIANCE STRUCTURE

Random Coefficients (RC) Models

A rather general random coefficients model (Laird and Ware, 1982), is $\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i \quad (i = 1, \ldots, m)$, where the $\mathbf{Z}_i$ are specified matrices, the $\mathbf{u}_i$ are vectors of random coefficients distributed independently as $MVN(\mathbf{0}, \mathbf{G}_i)$, the $\mathbf{G}_i$ are positive definite but otherwise unstructured matrices, and the $\mathbf{e}_i$ are distributed independently (of the $\mathbf{u}_i$ and of each other) as $MVN(\mathbf{0}, \sigma^2\mathbf{I}_{n_i})$. Typically the $\mathbf{G}_i$ are assumed to be equal; hence the covariance matrix of $\mathbf{y}_i$ is taken as $\mathbf{V}_i = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \sigma^2\mathbf{I}_{n_i}$. Special cases include the linear random coefficients (RCL) and quadratic random coefficients (RCQ) models. In the linear case, $\mathbf{Z}_i = [\mathbf{1}_{n_i}, \mathbf{t}_i]$. In the quadratic case, $\mathbf{Z}_i = [\mathbf{1}_{n_i}, \mathbf{t}_i, (t_{i1}^2, t_{i2}^2, \ldots, t_{in_i}^2)']$.

Random coefficients models have often been considered as distinct from parametric covariance models, probably because the orogenesis of the covariance structure is typically a consideration of regressions that vary across subjects rather than a consideration of within-subject similarity. Nevertheless, they yield parametric covariance structures that generally have nonconstant variances and nonstationary correlations, a fact that does not appear to be widely appreciated.

Antedependence Models

The unstructured antedependence model of order s [UAD(s)] model is defined as: $y_1 = \mathbf{x}_1'\beta + \varepsilon_1$ and $y_j = \mathbf{x}_j'\beta + \sum_{k=1}^{s^*} \phi_{jk}(y_{j-k} - \mathbf{x}_{j-k}'\beta) + \varepsilon_j \ (j = 2, \ldots, n)$, where $s^* = \min(s, j-1)$, the ε_j's are independent normal random variables with zero means and possibly time-dependent variances $\sigma_j^2 > 0$, and the autoregressive coefficients $\{\phi_{jk}\}$ are completely unrestricted parameters. By the term «unstructured,» we mean that the parame-

ters ϕ_{jk} and σ_j^2 cannot be expressed as functions of a smaller number of parameters. The UAD(s) model generalizes the stationary AR(s) model by allowing the innovation variances and autoregressive coefficients to be time-varying. This greater generality makes UAD models useful for situations in which measurement times are unequally spaced or there is clear evidence of nonstationarity in the data's correlation structure. The cost of this extra flexibility is an increase in the number of parameters, which are $O(n)$ rather than $O(1)$. Furthermore, because the number of parameters increases with the number of distinct measurement times, rectangularity or approximate rectangularity is a practical necessity. The equation used to define an UAD(s) model is both a response equation specification and an autoregressive specification. The corresponding variance-correlation specification is interesting in its own right. In it, response variances and correlations between observations lagged s or less observations apart are arbitrary (subject to positive definiteness constraints), but correlations between observations lagged more than s observations apart are completely determined by those corresponding to lags s or less.

Although the UAD(s) model is more flexible than more specialized AD models, such as stationary AR models, its drawback is that sometimes it may have too many parameters to be useful. There is a way to reduce the number of parameters using the structured AD (SAD) models introduced by Zimmerman and Núñez-Antón (1997). In these models, the autoregressive coefficients or correlations (depending on whether an autoregressive or variance-correlation specification is used) of the UAD(s) model follow a Box-Cox power function of time, and the innovation variances or response variances (again depending on the specification) are polynomial or step functions of time. For example, the variance-correlation specification of this UAD(s) model is given by $v_{jj} = \sigma^2 g(t_j;\psi)$ $(j = 1,\ldots,n)$ and $\rho_{j,j-k} = \rho_k^{f(t_j;\lambda_k)-f(t_{j-k};\lambda_k)}$ $(j = k+1,\ldots,n;\ k = 1,\ldots,s)$, where g, to be most useful in practice, is a function of relatively few parameters (e.g. a low-order polynomial function), and

$$f(t;\lambda) = \begin{cases} (t^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\ \log t & \text{if } \lambda = 0. \end{cases}$$

Exponentiation of the ρ_k's in by the Box-Cox form of f given above prescribes that the correlations on the kth-subdiagonal of the correlation matrix are monotone increasing if $\lambda_k < 1$, monotone decreasing if $\lambda_k > 1$, or constant if $\lambda_k = 1$ $(k = 1,\ldots,s)$. From another point of view, f effects a nonlinear deformation upon the time axis such that correlations between measurements equidistant in the deformed scale are constant.

5. EXAMPLES: FITTING OF THE MODELS

We fitted saturated models for the mean in each of the two data sets and the random coefficients models and several alternative models for the within-subject variance-

covariance structure. For the race data set, if we base our selection on AIC, the best fitting model was the UAD(3) with 34 parameters. However, if we use BIC as the selection criteria, the best fitting models were the ARH(1) and UAD(1). After carrying out several restricted likelihood ratio tests, the selected model was the UAD(3). For the cattle data set, using AIC, BIC and several restricted likelihood ratio tests led us to select as best fitting models for the groups A and B were, respectively, the UAD(2) and the UAD(3) models. For this data set, there was a significant group difference, as indicated by the corresponding test.

6. CONCLUSIONS

We have compared the RC models to several alternative models in the context of longitudinal data and when nonstationarity is present. First, there is the tradeoff between model flexibility and parsimony. The unstructured model with its $O(n^2)$ parameters is the most flexible and least parsimonious. The UAD model has $O(n)$ parameters and is the next most flexible. The SAD, ARIMA, and RC models all are highly structured, with $O(1)$ parameters, and thus are not as flexible as the others. Second, irregular spacing of measurements and non-rectangularity of the measurement schedule present no problems for the SAD and RC models, but one or both of these may require special care for the UN, UAD, and ARIMA models. A final comparison pertains to the existence of widely available software for fitting the models. The UN and RC models, and certain low-order UAD and ARIMA models, have the advantage on this score, for they can be fitted in PROC MIXED. Of all the parametric covariance structures which have been proposed for longitudinal data, stationary autoregressive and random coefficient models seem to receive the most attention; antedependence models, in contrast, get very little press. In our examples, however, stationary models generally did not fit as well as an antedependence model of some kind, and random coefficients models were not competitive at all. Thus, in these examples at least, it appears that some kind of antedependence model strikes the right balance between model flexibility and parsimony. Partly as a result of this, and partly due to their nice properties, we believe that antedependence models should be more routinely fit to longitudinal data exhibiting nonstationarity than they presently are.