Cuadernos del CIMBAGE Nº 14 (2012) 1-16

UNA APLICACIÓN DE MODELOS MULTIVARIADOS PARA DATOS LONGITUDINALES: EVALUACIÓN DEL COMPORTAMIENTO DE INDICADORES DEL MERCADO LABORAL

María del Carmen García, Liliana Koegel, Cecilia Rapelli
Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística
Facultad de Ciencias Económicas y Estadística
Universidad Nacional de Rosario
Bvard Oroño 1261, 2º Piso-Rosario-2000-Argentina
mgarcía@fcecon.unr.edu.ar, lkoegel@fcecon.unr.edu.ar, crapell1@rosario.gov.ar

Recibido 14 de noviembre de 2011, aceptado 24 de febrero de 2012

Resumen

En investigaciones sociales frecuentemente se realizan mediciones de múltiples variables respuestas a lo largo del tiempo a dos o más grupos de unidades. La modelación conjunta de varias variables respuestas es conveniente en muchas situaciones sobre la modelación de las mismas en forma separada. Los datos longitudinales multivariados surgen cuando un conjunto de diferentes respuestas se mide repetidamente en el tiempo sobre una misma unidad. Resulta de interés para tales datos conocer cómo la evolución de una respuesta está relacionada con la evolución de otra respuesta y/o cómo la asociación entre las distintas respuestas evoluciona con el tiempo. Para modelar tanto la correlación entre las mediciones repetidas de una variable respuesta dada como la correlación entre las mediciones de todas las respuestas en una determinada ocasión, se propusieron diferentes estrategias de modelación conjunta de variables. Una de ellas consiste en ajustar un modelo con una estructura de covariancias especial usando la notación "producto Kronecker"; la otra es modelar introduciendo efectos aleatorios, es decir, usando un modelo mixto. En este trabajo se presenta una aplicación de la metodología para el ajuste de modelos multivariados con el objetivo de explicar la evolución conjunta de las variables tasas de actividad y desocupación, para varios aglomerados de la República Argentina durante el tercer trimestre 2006 y el cuarto trimestre 2008.

Palabras clave: datos longitudinales multivariados, modelos mixtos, tasa de desocupación, tasa de actividad.

AN APPLICATION OF MULTIVARIATE MODELS FOR LONGITUDINAL DATA: THE PERFORMANCE ANALYSIS OF LABOUR MARKET INDICATORS

María del Carmen García, Liliana Koegel, Cecilia Rapelli
Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística
Facultad de Ciencias Económicas y Estadística
Universidad Nacional de Rosario
Bvard Oroño 1261, 2º Piso-Rosario-2000-Argentina
mgarcía@fcecon.unr.edu.ar, lkoegel@fcecon.unr.edu.ar, crapell1@rosario.gov.ar

Received November 14th 2011, accepted February 24th 2012

Abstract

Measurements of multiple variable responses over time to two or more groups of units are frequently carried out in social researches. In many situations, the joint modeling of several variable responses is more convenient than the modeling of responses separately. Multivariate longitudinal data arise when a set of different responses is measured repeatedly over time on the same unit. It is of interest to such data to know how the evolution of a response is related to the evolution of another response and/or how the association between the different responses evolves over time. Different strategies of joint modeling of variables were proposed to model both the correlation between repeated measurements of a given variable response and the correlation between measurements of all responses on a particular occasion. One of these strategies consists of adjusting a model with a special covariance structure using the "Kronecker product" notation. Another strategy is to model by introducing random effects, i.e. by using a mixed model. This paper presents an application of the methodology for the adjustment of multivariate models, with the purpose of explaining the joint evolution of the activity and unemployment rate variables for various agglomerates of the Argentine Republic during the third quarter of 2006 and the fourth quarter of 2008.

Keywords: multivariate longitudinal data, mixed models, unemployment rate, activity rate.

---

1. INTRODUCCIÓN

En investigaciones sociales resulta frecuente realizar mediciones de múltiples variables respuestas a lo largo del tiempo a dos o más grupos de unidades. La modelación conjunta de los perfiles longitudinales multivariados, en muchas situaciones, es necesaria o tiene ventajas adicionales sobre la modelación de las diferentes respuestas en forma separada.

Los datos longitudinales multivariados surgen cuando un conjunto de diferentes respuestas sobre la misma unidad se mide repetidamente en el tiempo. Resulta de interés para tales datos conocer cómo la evolución de una respuesta está relacionada con la evolución de otra respuesta y/o cómo la asociación entre las distintas respuestas evoluciona con el tiempo. El análisis de ese tipo de datos, comparado con el análisis longitudinal univariado, resulta más complicado debido a la existencia de correlación tanto entre las medidas tomadas en diferentes tiempos como entre las variables respuestas.

Varios autores proponen diferentes estrategias para la modelación conjunta de variables. Una de ellas consiste en ajustar un modelo con una estructura de covariancias especial usando la notación "producto Kronecker" y la otra, en modelar introduciendo efectos aleatorios; es decir, usando un modelo mixto. Ambos métodos, mediante un enfoque algo diferente, permiten modelar la correlación entre las mediciones repetidas de una variable respuesta dada y la correlación entre las mediciones de todas las respuestas en una determinada ocasión.

En este trabajo se presenta una aplicación de la metodología para el ajuste de modelos multivariados con el objetivo de explicar la evolución conjunta de las variables tasa de actividad y desocupación, a partir de información suministrada por la Encuesta Permanente de Hogares relevada por el Instituto Nacional de Estadística y Censos (INDEC), para diferentes aglomerados desde el tercer trimestre de 2006 hasta el cuarto de 2008.

2. MODELOS CONJUNTOS PARA DATOS LONGITUDINALES MULTIVARIADOS

Se considera para cada unidad p variables respuestas medidas en n_i ocasiones (j=1,..., n_i). La matriz de las respuestas de la unidad i es $\mathbf{Y}_i = \left[\mathbf{Y}_i^1,\ \mathbf{Y}_i^2,\ ...,\ \mathbf{Y}_i^p\right]$, i=1,...,N, donde cada $\mathbf{Y}_i^k$, k=1,...,p, es de dimensión n_i x 1 y $\boldsymbol{\varepsilon}_i = \left[\boldsymbol{\varepsilon}_i^1,\ \boldsymbol{\varepsilon}_i^2,\ ...,\ \boldsymbol{\varepsilon}_i^p\right]$ es la matriz de errores aleatorios de dimensión n_i x p. Al utilizar el operador "vec", que coloca en fila los elementos de una matriz, se obtiene el vector pn_ix1 $\mathbf{y}_i = \text{vec}(\mathbf{Y}_i) = \lfloor \mathbf{Y}_i^{1'}, \ \mathbf{Y}_i^{2'}, \ . \ . \ ., \ \mathbf{Y}_i^{p'} \rfloor$ y $\mathbf{e}_i = \text{vec}(\boldsymbol{\varepsilon}_i)$, donde $\boldsymbol{\varepsilon}_i$ tiene p columnas, siendo la k-ésima el error de la k-ésima característica.

Se presentan dos modelos que se pueden utilizar para analizar datos longitudinales multivariados.

2.1. Modelos con matriz de covariancias "producto Kronecker"

Este modelo permite especificar la matriz de covariancias del error dentro de cada unidad de estudio, y así examinar las correlaciones intra y entre variables.

El modelo lineal para el vector pn_i x1 de las respuestas de la unidad i,

$$\mathbf{y}_i = \mathbf{X}_i^* \boldsymbol{\beta} + \mathbf{e}_i \,, \tag{2.1.1}$$

$\mathbf{X}_i^*$ es una matriz bloque diagonal. Cada uno de los p bloques contiene las matrices $\mathbf{X}_\mathbf{k}$ de dimensión (pn_i x q), de las variables explicativas para cada una de las p variables respuestas, k=1,..., p.

$\boldsymbol{\beta}$ es un vector (q x 1) de parámetros denominados efectos fijos.

$\mathbf{e}_i$ es el vector de errores aleatorios, que se supone independiente, y normalmente distribuido con media cero y matriz de covariancias $\boldsymbol{\Omega}$. Entonces,

$$\mathbf{y}_i \overset{\text{iid}}{\sim} N_{pn_i}(\mathbf{X}_i^* \boldsymbol{\beta}, \boldsymbol{\Omega}) \,. \tag{2.1.2}$$

La estimación de $\boldsymbol{\Omega}$ no es posible cuando el número de unidades es menor o igual a pn_i. Para evitar este problema, se asume alguna estructura de covariancia para $\boldsymbol{\Omega}$. Esta matriz se puede expresar mediante la notación denominada "producto Kronecker", siendo la forma de la misma,

$$\boldsymbol{\Omega}_{pn_i x pn_i} = \mathbf{V}_{pxp} \otimes \boldsymbol{\Sigma}_{n_i x n_i} \,,$$

donde $\mathbf{V}$ y $\boldsymbol{\Sigma}$ son matrices definidas positivas y $\otimes$ es el producto directo. La matriz $\mathbf{V}$ representa la matriz de covariancias entre todas las variables respuestas de una unidad en una ocasión dada. Esta matriz se supone que no depende de una ocasión particular y es la misma para todas las ocasiones. La matriz $\boldsymbol{\Sigma}$ representa la matriz de covariancias entre las mediciones repetidas realizadas sobre una unidad y una variable respuesta, que tiene la misma estructura para todas las variables. Supone que para todas las variables respuestas la estructura de correlación entre las mediciones repetidas es la misma, y que la covariancia entre todas las respuestas no depende del tiempo y es constante para todas las ocasiones.

La matriz **Σ**, para una variable respuesta dada, puede tener cualquier estructura incluso ser no estructurada. Las dos estructuras más utilizadas son simetría compuesta (SC), que asume que todas las mediciones repetidas están equicorrelacionadas, no dependiendo de la separación entre las ocasiones de medida; o autorregresiva de orden 1 (AR(1)) para la cual la correlación disminuye cuando aumenta la separación entre las ocasiones. Este modelo presenta varias limitaciones: una correlación común intra variable para las diferentes variables, una correlación constante intra variable para variables medidas en la misma ocasión y ocasiones de medida equiespaciadas.

2.2. Modelos con efectos aleatorios: Modelo lineal mixto multivariado

Se modela la asociación entre las distintas respuestas utilizando efectos aleatorios. En lugar de modelar la variación dentro de una unidad como en los modelos con sólo efectos fijos, los modelos mixtos asumen que los coeficientes de regresión son una muestra aleatoria de alguna población y permiten modelar variaciones entre unidades. Estos modelos describen la evolución media de una respuesta determinada utilizando una función de tiempo y se introducen los llamados efectos aleatorios para representar las desviaciones específicas de la unidad de esta evolución promedio. En un enfoque común de modelado utilizando modelos mixtos se suponen efectos aleatorios para cada respuesta e imponiendo una distribución multivariada sobre los mismos, las diferentes variables están asociadas. Este enfoque tiene muchas ventajas y es aplicable en una amplia variedad de situaciones.

La expresión del modelo mixto multivariado (Shah, Laird y Schoenfeld, 1997) para la unidad i es de la forma

$$\mathbf{y}_i = \mathbf{X}_i^* \boldsymbol{\beta} + \mathbf{Z}_i^* \boldsymbol{\gamma}_i + \mathbf{e}_i \,, \quad i=1,\ldots,N, \qquad (2.2.1)$$

donde

$\mathbf{X}_i^*$ y $\mathbf{Z}_i^*$ son matrices bloque diagonal. Cada uno de los p bloques contiene las matrices $\mathbf{X}_k$ y $\mathbf{Z}_k$, de dimensión (pn_i x q) y (pn_i x r) respectivamente, de las variables explicativas para cada una de las p variables respuestas, k=1,…, p.

$\boldsymbol{\beta}$ es un vector (q x 1) de parámetros denominados efectos fijos.

$\boldsymbol{\gamma}_i$ es un vector (r x 1) cuyas componentes se denominan efectos aleatorios.

$\mathbf{e_i}$ es una matriz (pn_i x pn_i) de errores dentro de cada unidad.

Generalmente se asume que,

$$\mathbf{e}_i \overset{iid}{\sim} N_{pn_i}(\mathbf{0}, \mathbf{R}_i) \quad \text{y} \quad \boldsymbol{\gamma}_i \overset{iid}{\sim} N_r(\mathbf{0}, \mathbf{D}).$$

Además, se supone que siendo i e i' dos unidades distintas, las componentes de error están no correlacionadas y no asociadas con los efectos aleatorios, $\text{cov}(\mathbf{e}_i, \mathbf{e}_{i'}) = \mathbf{0}$, $\text{cov}(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_{i'}) = \mathbf{0}$ y $\text{cov}(\mathbf{e}_i, \boldsymbol{\gamma}_i) = \mathbf{0}$.

La matriz de covariancias $\mathbf{D}$ contiene tanto la covariancia entre los efectos aleatorios dentro de una respuesta dada como la covariancia entre los efectos aleatorios de las distintas variables respuestas. La matriz $\mathbf{R}_i$ tiene una estructura específica para reflejar la correlación entre las mediciones repetidas. Si no se hacen supuestos sobre la forma de la estructura de covariancias para cada fila y columna de $\boldsymbol{\varepsilon}_i$, los parámetros de covariancias de $\mathbf{R}_i$ serán numerosos para ser estimados correctamente. Además, si se sospecha la existencia de una estructura, utilizarla conduce a una estimación e inferencia más eficiente. Generalmente se considera que los vectores filas de $\boldsymbol{\varepsilon}_i$ son independientes y tienen distribución $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, siendo $\boldsymbol{\Sigma}_{pxp}$ una matriz con estructura arbitraria. En este caso, la Var($\mathbf{e_i}$) tiene estructura $\boldsymbol{\Sigma} \otimes \mathbf{I}_i$, con $\mathbf{I_i}$ una matriz identidad de dimensión n_i. La independencia de las filas implica que cada columna de $\boldsymbol{\varepsilon}_i$ se distribuya normalmente con media $\mathbf{0}$ y matriz de covariancias $\boldsymbol{\sigma}^2_{kk}\, \mathbf{I}_i$, para k=1,…,p. La estimación de los parámetros $\boldsymbol{\beta}$ y $\boldsymbol{\sigma}^2$ y los elementos de $\mathbf{D}$ y $\boldsymbol{\Omega}$ se realiza minimizando la función objetivo, menos dos veces el logaritmo de la función de verosimilitud (l=-2 log(L), siendo L la función de verosimilitud) mediante el algoritmo de Newton-Raphson.

3. APLICACIÓN

En este trabajo se presenta una aplicación de la metodología para evaluar el comportamiento del mercado laboral a partir de información suministrada por la Encuesta Permanente de Hogares (EPH) relevada por el INDEC.

La información relativa a los diferentes aglomerados relevada mediante esa encuesta ofrece importantes ventajas para el análisis empírico de variables registradas en la misma. En este trabajo, resulta de interés la modelación conjunta de las variables tasa de actividad (ACT) y tasa de desocupación (DES) en el período tercer trimestre de 2006 hasta el cuarto trimestre de 2008. La tasa de actividad se define como el cociente entre la población económicamente activa (personas que trabajan y personas que buscan trabajo) y la población total, mientras que la de desocupación es el cociente entre la población que busca trabajo y la económicamente activa. Se agruparon algunos de los aglomerados de acuerdo al número de habitantes:

Grupo A (los aglomerados con más de 500.000 habitantes) constituido por Ciudad Autónoma de Buenos Aires, Partidos del Gran Buenos Aires, Gran Mendoza, Gran Tucumán- Tafí Viejo, Gran Córdoba, Gran La Plata, Gran Rosario, Mar del Plata-Batán.

Grupo B (los aglomerados con menos de 500.000 habitantes) constituido por San Luis- El Chorrillo, Bahía Blanca-Cerri, Gran Paraná, Gran Santa Fe, Río Cuarto, San Nicolás-Villa Constitución, Comodoro Rivadavia-Rada Tilli y Ushuaia-Río Grande.

Las respuestas se simbolizan $\mathbf{Y}^1$= tasa de actividad e $\mathbf{Y}^2$= tasa de desocupación. Resulta de interés conocer cómo evoluciona con el tiempo la asociación entre las variables ACT y DES, y cómo están asociadas las evoluciones de esas variables. Los perfiles promedio de las variables se presentan en el siguiente gráfico:

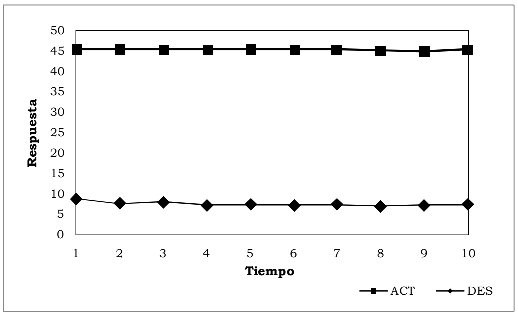

Gráfico 3.1. Perfiles promedio para las variables tasas de actividad y de desocupación

En el grafico se observa que la trayectoria de ambas variables es lineal y presenta un comportamiento similar en el tiempo. Los gráficos 3.2 y 3.3 muestran el comportamiento de las variables en los grupos de aglomerados.

Grupo A *Grupo B*

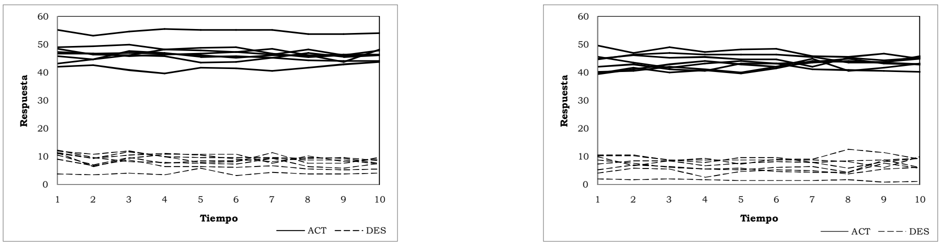

Gráfico 3.2. Perfiles individuales para las variables tasas de actividad y de desocupación por grupo

Grupo A *Grupo B*

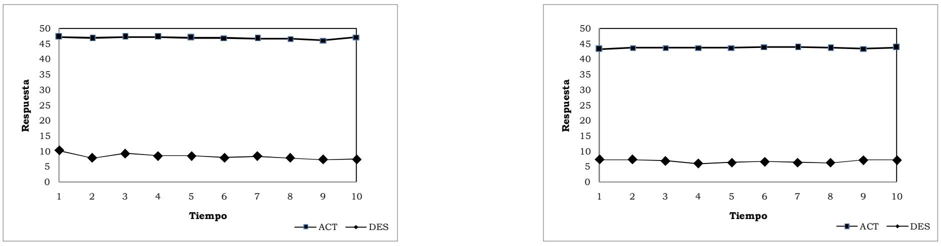

Gráfico 3.3. Perfiles promedio para las variables tasas de actividad y de desocupación por grupo

La evolución de cada respuesta se modela como una función lineal del tiempo, permitiendo que la ordenada y la tasa de cambio difieran entre variables y grupos. Se evaluaron tres modelos: con estructura de covariancia usando la notación producto Kronecker (modelo 1) y con covariancia inducida por los efectos aleatorios en dos formulaciones, univariada (modelo 3) y bivariada (modelo 2). El modelo 3 considera a las dos variables respuestas independientes. La comparación entre los modelos se realiza usando el criterio de información de Akaike (AIC), que penaliza el logaritmo de la verosimilitud ($\log \hat{L}$) restando el número de parámetros de cada modelo (s), $AIC = -2\log \hat{L} + 2s$.

Modelo 1

El modelo utilizado para las tasas de actividad y de desocupación del aglomerado i en la ocasión j tiene la expresión,

Grupo A	Grupo B
$Y_{ij}^1 = \beta_{01}^1 + \beta_1^1 t_{ij} + \varepsilon_{ij}^1$	$Y_{ij}^1 = \beta_{02}^1 + \beta_2^1 t_{ij} + \varepsilon_{ij}^1$
$Y_{ij}^2 = \beta_{01}^2 + \beta_1^2 t_{ij} + \varepsilon_{ij}^2$	$Y_{ij}^2 = \beta_{02}^2 + \beta_2^2 t_{ij} + \varepsilon_{ij}^2$,

Se supone, $\mathbf{e}_i = \begin{pmatrix} \boldsymbol{\varepsilon}_i^1 \\ \boldsymbol{\varepsilon}_i^2 \end{pmatrix} \overset{iid}{\sim} N_{20}(\mathbf{0}, \boldsymbol{\Omega})$.

Los parámetros del modelo representan,

$\beta_{01}^1, \beta_{01}^2, \beta_{02}^1, \beta_{02}^2$ las ordenadas de las variables ACT y DES para cada grupo.

$\beta_1^1, \beta_1^2, \beta_2^1, \beta_2^2$ las pendientes de las variables ACT y DES para cada grupo.

Se ajusta el modelo propuesto, usando el procedimiento "MIXED" del software SAS, utilizando dos estructuras de covariancia para $\boldsymbol{\Omega}$: arbitraria (UN) para $\mathbf{V}$ y simetría compuesta y autorregresiva para $\boldsymbol{\Sigma}$, cuya notación es $\boldsymbol{\Omega}$=UN@CS y $\boldsymbol{\Omega}$=UN@AR(1). A partir del valor del criterio de Akaike (AIC) (1173.2 vs 1204), se elige la primera cuya forma estimada es

$$\hat{\mathbf{\Omega}} = \hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}} = \begin{pmatrix} \hat{\sigma}^2_{Y^1} & \hat{\sigma}_{Y^1Y^2} \\ \hat{\sigma}_{Y^1Y^2} & \hat{\sigma}^2_{Y^2} \end{pmatrix} \otimes \begin{pmatrix} 1 & \hat{\rho} & \hat{\rho} & . & \hat{\rho} \\ \hat{\rho} & 1 & \hat{\rho} & . & \hat{\rho} \\ . & . & . & . & . \\ . & . & . & . & . \\ \hat{\rho} & \hat{\rho} & . & \hat{\rho} & 1 \end{pmatrix}$$

$$\hat{\mathbf{\Omega}} = \begin{pmatrix} 9.141 & 1.217 \\ 1.217 & 8.099 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0.823 & 0.823 & . & 0.823 \\ 0.823 & 1 & 0.823 & . & 0.823 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0.823 & 0.823 & . & 0.823 & 1 \end{pmatrix}$$

El modelo supone que las dos variables comparten una correlación común entre las mediciones repetidas dada por $\mathbf{\Sigma}$ y las correlaciones entre las variables $\rho^* = \dfrac{\sigma_{Y^1Y^2}}{\sqrt{\sigma^2_{Y^1}\ \sigma^2_{Y^2}}}$ son las mismas en cada ocasión.

Los datos no muestran una fuerte correlación positiva entre las variables en cada ocasión $\hat{\rho}^* = 0.1414$. La correlación entre dos mediciones consecutivas para las variables ACT y DES es $\hat{\rho} = 0.8230$. Los parámetros estimados y la significación de los mismos se presentan en la Tabla 3.1.

Parámetro	Variable	Grupo	Estimación	t Student	p=Pr>|t|
β^1_{01}	ACT	A	47.2442	46.44	<.0001
β^1_{02}	ACT	B	43.5117	42.77	<.0001
β^2_{01}	DES	A	9.5292	9.95	<.0001
β^2_{02}	DES	B	6.8350	7.14	<.00010
β^1_1	ACT	A	-0.06553	-1.32	0.1867
β^1_2	ACT	B	0.02379	0.48	0.6313
β^2_1	DES	A	-0.2242	-4.81	<.0001
β^2_2	DES	B	-0.02205	-0.47	0.6366

Tabla 3.1. Estimación y significación de los parámetros

En ambos grupos el cambio promedio de las variables decrece levemente con el tiempo, excepto en el Grupo B, donde el cambio promedio de la tasa de actividad incrementa en el tiempo. El coeficiente correspondiente a la tasa de desocupación para el Grupo A es el único que resulta significativo (p<.0001). Tanto la tasa de desocupación como la de actividad se mantienen estables en el periodo estudiado.

De esto se infiere que los grandes aglomerados (Grupo A) tienen una disminución en la tasa de desocupación que no fue acompañada de una variación de la tasa de actividad. Sin embargo, en los pequeños aglomerados ambas se mantuvieron estables.

Se comprueba que las pendientes de cada variable para ambos grupos son las mismas. Existe evidencia para rechazar esa hipótesis (χ^2_{OBS} = 10.12, p=0.0063).

Modelo 2

Se plantea el mismo modelo anterior para la parte media. Para modelar la covariancia se introducen efectos aleatorios. El vector de los efectos fijos describe la evolución promedio de las variables y el vector de los efectos aleatorios describe cómo el perfil del aglomerado i se desvía del perfil promedio.

El modelo utilizado para la tasa de actividad y tasa de desocupación del aglomerado i en la ocasión j tiene la expresión

Grupo A	Grupo B
$Y^1_{ij} = \beta^1_{01} + \beta^1_1 t_{ij} + \gamma^1_{0i} + \gamma^1_{1i} t_{ij} + \varepsilon^1_{ij}$	$Y^1_{ij} = \beta^1_{02} + \beta^1_2 t_{ij} + \gamma^1_{0i} + \gamma^1_{1i} t_{ij} + \varepsilon^1_{ij}$
$Y^2_{ij} = \beta^2_{01} + \beta^1_1 t_{ij} + \gamma^2_{0i} + \gamma^2_{1i} t_{ij} + \varepsilon^2_{ij}$	$Y^2_{ij} = \beta^2_{02} + \beta^2_2 t_{ij} + \gamma^2_{0i} + \gamma^2_{1i} t_{ij} + \varepsilon^2_{ij}$,

$\gamma^1_{0i}, \gamma^2_{0i}, \gamma^1_{1i}, \gamma^2_{1i}$ son efectos aleatorios para cada variable.

Se supone que

$$\mathbf{e}_i = \begin{pmatrix} \boldsymbol{\varepsilon}_i^1 \\ \boldsymbol{\varepsilon}_i^2 \end{pmatrix} \overset{iid}{\sim} N_{20}(\mathbf{0}, \mathbf{R}_i), \qquad \boldsymbol{\gamma}_i = \begin{pmatrix} \gamma_{0i}^1 \\ \gamma_{1i}^1 \\ \gamma_{0i}^2 \\ \gamma_{1i}^2 \end{pmatrix} \overset{iid}{\sim} N_4(\mathbf{0}, \mathbf{D}), \qquad \mathbf{R_i} = \sigma^2 \mathbf{I},$$

$$\mathbf{D} = \begin{pmatrix} \sigma^2_{\gamma_{0i}^1} & \sigma_{\gamma_{0i}^1\gamma_{1i}^1} & \sigma_{\gamma_{0i}^1\gamma_{0i}^2} & \sigma_{\gamma_{0i}^1\gamma_{1i}^2} \\ \sigma_{\gamma_{1i}^1\gamma_{0i}^1} & \sigma^2_{\gamma_{1i}^1} & \sigma_{\gamma_{1i}^1\gamma_{0i}^2} & \sigma_{\gamma_{1i}^1\gamma_{1i}^2} \\ \sigma_{\gamma_{0i}^2\gamma_{0i}^1} & \sigma_{\gamma_{0i}^2\gamma_{1i}^1} & \sigma^2_{\gamma_{0i}^2} & \sigma_{\gamma_{0i}^2\gamma_{1i}^2} \\ \sigma_{\gamma_{1i}^2\gamma_{0i}^1} & \sigma_{\gamma_{1i}^2\gamma_{1i}^1} & \sigma_{\gamma_{1i}^2\gamma_{0i}^2} & \sigma^2_{\gamma_{1i}^2} \end{pmatrix}.$$

La evolución de cada respuesta se modela como una función lineal del tiempo mediante un modelo cuya pendiente refleja el cambio promedio de la variable en el tiempo.

Se obtuvo la siguiente estimación de σ^2**,** $\hat{\sigma}^2 = 1.2907$.

La matriz de covariancia de los efectos aleatorios (**D**) muestra cómo la evolución de la variable ACT está asociada con la evolución de la variable DES. Los valores de las correlaciones obtenidos a partir de la matriz **D** se resumen en la siguiente tabla.

	Ordenada ACT	Pendiente ACT	Ordenada DES	Pendiente DES
Ordenada ACT	1			
Pendiente ACT	-0.752	1		
Ordenada DES	0.148	0.005	1	
Pendiente DES	-0.869	0.995	-0.368	1

Tabla 3.2. Matriz de correlaciones estimada del modelo

Se observa:

- Una alta correlación negativa entre la ordenada y pendiente de la tasa de actividad, $\hat{\rho}_{\gamma_{1i}^1\gamma_{0i}^1} = -0.752$. Los aglomerados que al comienzo del estudio tienen una tasa de actividad alta presentan en el tiempo fuertes disminuciones de la tasa de actividad promedio.
- Una alta correlación negativa entre la ordenada de la tasa de actividad y la pendiente de la tasa de desocupación,

$\hat{\rho}_{\gamma_{1i}^2\gamma_{0i}^1} = -0.869$. Cuando la tasa de actividad es alta al comienzo del estudio, en promedio la tasa de desocupación tiene un fuerte descenso en el tiempo.

- Una alta correlación entre las pendientes de las dos variables indica fuerte correlación entre las evoluciones de ambas variables.
- Ordenada y pendiente de desocupación correlacionadas negativamente, $\hat{\rho}_{\gamma_{1i}^2\gamma_{0i}^2} = -0.368$.

La tabla siguiente presenta la estimación de los parámetros y la significación de los mismos.

Parámetro	Variable	Grupo	Estimación	t Student	p=Pr>|t|
β_{01}^1	ACT	A	47.2442	36.06	<.0001
β_{02}^1	ACT	B	43.5117	32.28	<.0001
β_{01}^2	DES	A	9.5292	10.31	<.0001
β_{02}^2	DES	B	6.8350	7.40	<.0001
β_1^1	ACT	A	-0.0655	-0.74	0.4587
β_2^1	ACT	B	0.0238	0.27	0.7879
β_1^2	DES	A	-0.2242	-3.86	<.0001
β_2^2	DES	B	-0.0221	-0.38	0.7045

Tabla 3.3. Estimación y significación de los parámetros

El patrón de cambio promedio de ambas variables decrece con el tiempo para los dos grupos de aglomerados excepto en el Grupo B, donde el cambio promedio de la tasa de actividad aumenta. Esos cambios resultan no significativos salvo para la tasa de desocupación del Grupo A.

Se realizan pruebas de hipótesis para comprobar si existen diferencias entre grupos y pendientes, considerando un nivel de significación de 0.05.

Se comprueba que la evolución de las respuestas no tiene el mismo comportamiento en los dos grupos (χ^2_{OBS} = 7.26, p=0.0265) y, además, las dos variables del Grupo A no tienen el mismo comportamiento (χ^2_{OBS} = 5.40, p=0.0201).

Modelo 3

Se ajusta el modelo 2 considerando que las dos variables respuestas están no correlacionadas. El modelo y los supuestos expresados de una forma alternativa pero equivalente al anterior son

$$\begin{aligned}\mathbf{Y}_i^1 &= \mathbf{X}_i^1\boldsymbol{\beta} + \mathbf{Z}_i^1\boldsymbol{\gamma}_i^1 + \mathbf{e}_i^1\\ \mathbf{Y}_i^2 &= \mathbf{X}_i^2\boldsymbol{\beta} + \mathbf{Z}_i^2\boldsymbol{\gamma}_i^2 + \mathbf{e}_i^2\end{aligned},$$

$$\begin{array}{ll}\mathbf{e}_i^1 \sim N_{10}(0,\boldsymbol{\sigma}^2_{\mathbf{e}^1}\mathbf{I}_{10}) & \mathbf{e}_i^1 \sim N_{10}(0,\boldsymbol{\sigma}^2_{\mathbf{e}^2}\mathbf{I}_{10})\\ \boldsymbol{\gamma}_i^1 \sim N_2(\mathbf{0},\mathbf{D}^1) & \boldsymbol{\gamma}_i^2 \sim N_2(\mathbf{0},\mathbf{D}^2)\end{array},$$

donde los parámetros fueron definidos en (2.2.1) e $\mathbf{y}_i = \left[\mathbf{Y}_i^1, \mathbf{Y}_i^2\right]$ el vector de respuesta para el aglomerado i=1,..,N siendo n_i=10 la dimensión de cada subsector.

En este modelo, cada aglomerado varía no sólo en su nivel de respuesta al comienzo del estudio sino también en términos de cambios de sus respuestas en el tiempo.

La estimación del mismo para cada grupo es

Grupo A	Grupo B
$\hat{Y}_i^1 = 47.24 - 0.0655\,t_i$	$\hat{Y}_i^1 = 43.51 + 0.0238\,t_i$
$\hat{Y}_i^2 = 9.529 - 0.2242\,t_i$	$\hat{Y}_i^2 = 6.835 - 0.022\,t_i$,

AIC = 1162.9.

La comparación entre los modelos mixtos univariado y bivariado (AIC=1162.9 vs. 1155.6) muestra la ventaja de considerar una asociación entre las dos variables. Además, se puede observar que el modelo mixto fue mejor que el modelo que incluye una estructura de simetría compuesta, modelada mediante la notación producto de Kronecker (AIC = 1155.6 vs. 1173.2).

El examen de residuos para comprobar el cumplimiento de los supuestos muestra que el modelo mixto provee un ajuste adecuado del patrón de cambio de las respuestas en el tiempo.

4. CONSIDERACIONES FINALES

En este trabajo, se considera el problema de modelar la evolución de varias variables respuestas en el tiempo. Se utilizan dos estrategias que difieren en la caracterización de la estructura de covariancia. Los modelos con estructura de covariancia producto Kronecker posibilitan examinar las correlaciones entre y dentro de las variables. Sin embargo, presenta varias limitaciones: una correlación común intra variable para las diferentes variables y una correlación constante intra variable para variables medidas en la misma. En la modelación conjunta mediante un modelo mixto multivariado, ambas trayectorias están vinculadas a través de la distribución conjunta de los efectos aleatorios. Los modelos con coeficientes aleatorios permiten examinar correlaciones entre las trayectorias de las variables en el tiempo y obtener la evolución de esas variables respuestas. Con este enfoque, se pueden manejar datos no equiespaciados, varias variables respuestas y datos perdidos.

En la aplicación del ajuste de los modelos y su comparación se puede concluir que:

. El modelo con estructura de covariancia producto Kronecker no muestra una fuerte correlación positiva entre las variables en cada ocasión, pero sí entre dos mediciones consecutivas para ACT y DES. Del análisis se puede inferir que los grandes aglomerados tienen una disminución en la tasa de desocupación que no fue acompañada de una variación de la tasa de actividad. Sin embargo, en los pequeños aglomerados ambas se mantienen estables.

. El modelo mixto permite estimar la matriz de correlaciones entre las pendientes individuales para cada variable y, de esta manera, determinar cómo la evolución de la variable ACT está asociada con la evolución de la DES. Para estos datos se obtiene una fuerte correlación entre las evoluciones de ambas variables. Se observa que aquellos aglomerados con tasa de actividad alta al comienzo del estudio presentan un patrón de cambio en la misma que disminuye en el tiempo. Además, cuando la tasa de actividad es alta al comienzo del estudio, el cambio promedio de la tasa de desocupación disminuye en el tiempo.

Se comprueba que las dos variables en cada grupo tienen diferente comportamiento.

Para este conjunto de datos el uso del modelo mixto resulta preferible a los otros modelos considerados.

BIBLIOGRAFÍA

Fieuws, S.; Verbeke, G. (2004). “Joint modeling of multivariate longitudinal profiles: pitfalls of the random-effects approach”. *Statistical in Medicine*, vol. 23, pp.3093-3104.

Galecki, A. T. (1994). “General class of covariance structures for two or more repeated factors in longitudinal data analysis”. *Communication Statististics-Theory Meth*, vol. 23, pp.3105-19.

Littell, R. C.; Milliken, G.A.; Stroup, W.W.; Wolfinger, R.D. (1996). *SAS System for Mixed Models*, SAS Institute, Cary, NC.

Naik, D. R.; Rao, S. (2001). “Analysis of multivariate repeated Measures Data with a Kronecker Product Structured Covariance Matrix”. *Journal of Applied Statistics*, vol. 28, pp.91-105.

SAS/IML Software: Usage and Reference. Version 6. (1990). SAS Institute Inc., Carry, NC.

Shah, A.; Laird, N.; Schoenfeld, D. (1997). “A random-effects model for multiple characteristics with possibly missing data”. *Journal of the American Statistical Association*, vol. 92, pp.775-9.

Sy, J. P.; Taylor, J. M.; Cumberland, W. G. (1997). “A stochastic model for the analysis of bivariate longitudinal AIDS data”. *Biometrics,* vol. 53, pp.542-55.

Thiébaut, R.; Jacqmin-Gadda, H.; Chêne, G.; Leport, C.; Commenges, D. (2002). “Bivariate linear mixed models using SAS proc MIXED”. *Computer Methods and Programmes in Biomedicine*, vol. 69, pp.249-256.