Overview

Transition models

Remember that a transition model (a special kind of conditional model) explains correlation between $Y_{i1}, \ldots, Y_{in_i}$ by letting past values $Y_{i1}, \ldots, Y_{ij-1}$ influence the present observation Y_{ij}. More specifically:
Let $\mathcal{H}_{ij} = \{y_{i1}, \ldots, y_{ij-1}\}$ denote the past responses for subject i (the history at time t_{ij}). Then, a transition model assumes for the conditional moments

- $g(\mu_{ij}^C) = g(\mathrm{E}(Y_{ij}|\mathcal{H}_{ij})) = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \sum_{r=1}^{s} \kappa_r(\mathcal{H}_{ij}; \boldsymbol{\alpha}, \boldsymbol{\beta})$
- $v_{ij}^C = \mathrm{Var}(Y_{ij}|\mathcal{H}_{ij}) = \phi v(\mu_{ij}^C)$.

Here, the $\kappa_r(\cdot)$ are known functions and might, for example, include interactions between past responses.

Transition models

The most useful and common transition models are Markov chains of order q, for which the conditional distribution of $Y_{ij}|\mathcal{H}_{ij}$ only depends on the last q observations $y_{ij-1}, \ldots, y_{ij-q}$.

Then, the conditional distribution is assumed to follow a GLM

- $f(y_{ij}|\mathcal{H}_{ij}) = \exp\{\phi^{-1}[y_{ij}\eta_{ij} - \psi(\eta_{ij})] + c(y_{ij}, \phi)\}$ with conditional moments $\mu_{ij}^C = \mathrm{E}(Y_{ij}|\mathcal{H}_{ij}) = \psi'(\eta_{ij})$ and $v_{ij}^C = \mathrm{Var}(Y_{ij}|\mathcal{H}_{ij}) = \phi\psi''(\eta_{ij})$.
- $g(\mu_{ij}^C) = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \sum_{r=1}^{s} \kappa_r(\mathcal{H}_{ij}; \boldsymbol{\alpha}, \boldsymbol{\beta})$
- $v_{ij}^C = \phi v(\mu_{ij}^C)$.

Transition models are most useful for equally space data, i.e. where the t_{ij} are equally spaced, and we will assume this in the following.

Linear transition models

A simple autoregressive model of order $q = 1$ is

$$\begin{aligned} Y_{ij} &= \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \varepsilon_{ij} \\ \varepsilon_{ij} &= \alpha\varepsilon_{ij-1} + Z_{ij}, \quad Z_{ij} \overset{iid}{\sim} N(0, \tau^2), \end{aligned}$$

with $\tau^2 = \sigma^2(1 - \alpha^2)$ and $\alpha = \exp(-\phi) \geq 0$.

We can rewrite this model by substituting $\varepsilon_{ij-1} = Y_{ij-1} - \boldsymbol{x}'_{ij-1}\boldsymbol{\beta}$ as

$$Y_{ij} | Y_{ij-1} \sim N(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \alpha(Y_{ij-1} - \boldsymbol{x}'_{ij-1}\boldsymbol{\beta}), \tau^2).$$

This form shows how the past observation Y_{ij-1} is treated as a predictor of Y_{ij}, the same way as the explanatory variables $\boldsymbol{x}_{ij}$.

We can also derive the marginal covariance structure

$$\mathrm{Cov}(Y_{ij}, Y_{ik}) = \alpha^{|j-k|}\sigma^2,$$

which is the AR(1) structure already mentioned for marginal models.

Linear transition models

Note that in this model, the marginal mean is

$$\mathrm{E}(Y_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta},$$

and the $\boldsymbol{\beta}$ parameters thus have a marginal interpretation.

In the economics literature, the autoregressive model is often written as

$$Y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \alpha Y_{ij-1} + Z_{ij}, \quad Z_{ij} \overset{iid}{\sim} N(0, \tau^2),$$

i.e. Y_{ij} is regressed on the past value Y_{ij-1} itself rather than adjusting for its expectation. While the predictions in both model formulations are the same, note that here

$$\mathrm{E}(Y_{ij}) = \sum_{r=0}^{\infty} \alpha^r \boldsymbol{x}'_{ij-r}\boldsymbol{\beta}.$$

$\boldsymbol{\beta}$ thus does not have the usual marginal interpretation and the interpretation of $\boldsymbol{\beta}$ will change with the correlation model. Thus, the other model formulation is generally preferable.

Linear transition models

An autoregressive model of order q (AR(q)) is

$$Y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sum_{r=1}^{q} \alpha_r (Y_{ij-r} - \boldsymbol{x}'_{ij-r}\boldsymbol{\beta}) + Z_{ij}, \quad Z_{ij} \overset{iid}{\sim} N(0, \tau^2).$$

This is a transition model with $g(\mu_{ij}^C) = \mu_{ij}^C$, $v(\mu_{ij}^C) = 1$, and $\kappa_r(\mathcal{H}_{ij}; \alpha, \boldsymbol{\beta}) = \alpha_r (y_{ij-r} - \boldsymbol{x}'_{ij-r}\boldsymbol{\beta})$.

Logistic transition models

A first-order Markov chain for binary responses is

$$\operatorname{logit} P(Y_{ij} = 1 | \mathcal{H}_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \alpha y_{ij-1}. \qquad (1)$$

Here, $g(\mu^C_{ij}) = \operatorname{logit}(\mu^C_{ij}) = \log\left(\frac{\mu^C_{ij}}{1-\mu^C_{ij}}\right)$, $v(\mu^C_{ij}) = \mu^C_{ij}(1-\mu^C_{ij})$, and $\kappa_1(\mathcal{H}_{ij}; \alpha, \boldsymbol{\beta}) = \alpha y_{ij-1}$.

For example: the probability of a child having a respiratory infection at time t_{ij} depends not only on explanatory variables, but also on the infection status at the previous visit.

- $\exp(\alpha)$ then is the ratio of the odds of infection among children with and without an infection at the previous visit.
- $\boldsymbol{\beta}_k$ is the change per unit change in $\boldsymbol{x}_k$ in the log odds of infection among children free of infection at the last visit.

Note that $\boldsymbol{\beta}$ does no longer have a marginal interpretation as it had in the linear transition model.

Logistic transition models

This model is a first-order Markov chain and can be described by its transition matrix with entries $P(Y_{ij} = y_{ij} | Y_{ij-1} = y_{ij-1}),\ y_{ij}, y_{ij-1} \in \{0, 1\}$.

$$
\begin{array}{cccc}
 & & \multicolumn{2}{c}{y_{ij}} \\
 & & 0 & 1 \\
\multirow{2}{*}{y_{ij-1}} & 0 & \frac{1}{1+\exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta})} & \frac{\exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta})} \\
 & 1 & \frac{1}{1+\exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\alpha)} & \frac{\exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\alpha)}{1+\exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\alpha)}
\end{array}
$$

Note that the rows of this matrix sum to 1, as
$P(Y_{ij} = 1 | Y_{ij-1} = y_{ij-1}) + P(Y_{ij} = 0 | Y_{ij-1} = y_{ij-1}) = 1$.

The transition probabilities depend on the explanatory variables $\boldsymbol{x}_{ij}$ and can thus vary from subject to subject.

Logistic transition models

An extension of this model to a model of order q is

$$\text{logit}P(Y_{ij} = 1|\mathcal{H}_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_q + \sum_{r=1}^{q} \alpha_r y_{ij-r}.$$

The notation $\boldsymbol{\beta}_q$ is chosen to indicate that the interpretation (and value) of $\boldsymbol{\beta}_q$ will change with the Markov order q.

Another extension would be to let $\boldsymbol{\beta}$ differ by y_{ij-1}. This can be achieved by letting

$$\text{logit}P(Y_{ij} = 1|\mathcal{H}_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + y_{ij-1}\boldsymbol{x}'_{ij}\boldsymbol{\alpha}.$$

In this model we could then test $\boldsymbol{\alpha} = (\alpha, \mathbf{0})$ to see whether (1) holds and the covariate effects are independent of $y_{ij-1} = 1$ or $y_{ij-1} = 0$.

Similarly, we could extend the model of order q to allow for covariate effects to differ by past values.

Transition models for categorical data

Similar models can be derived for categorical data without natural ordering, by modeling transition probabilities

$$\pi_{ab} = P(Y_{ij} = b | Y_{ij-1} = a).$$

For ordered categorical data, the proportional odds model can also be extended to include functions of past values. See Diggle et al (2002) for more details.

Log-linear transition models

A first-order Markov model for count data analogous to the one for binary data would be

$$\mu_{ij}^{C} = \mathrm{E}(Y_{ij}|\mathcal{H}_{ij}) = \exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \alpha y_{ij-1}).$$

However, if $\alpha > 0$, this model causes the conditional expectation to grow exponentially over time. Thus, this model can only be reasonably used to describe negative associations ($\alpha < 0$).

Log-linear transition models

A possible first-order Markov chain for count data is a model where $Y_{ij}|\mathcal{H}_{ij}$ follows a Poisson distribution, and where $\kappa_1(\mathcal{H}_{ij};\alpha,\boldsymbol{\beta}) = \alpha\{\log(y^*_{ij-1}) - \boldsymbol{x}'_{ij-1}\boldsymbol{\beta}\}$ with $y^*_{ij} = \max(y_{ij}, d), 0 < d < 1$, such that

$$\mu^C_{ij} = \mathrm{E}(Y_{ij}|\mathcal{H}_{ij}) = \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})\left(\frac{y^*_{ij-1}}{\exp(\boldsymbol{x}'_{ij-1}\boldsymbol{\beta})}\right)^{\alpha}.$$

Zeger & Qaqish, Biometrics, 1988 propose this model, where d prevents $y_{ij-1} = 0$ from being an absorbing state, which forces all future observations to be 0 as well. For $\alpha > 0$, μ^C_{ij} is increased when y_{ij-1} exceeds $\exp(\boldsymbol{x}'_{ij-1}\boldsymbol{\beta})$ (positive correlation), and reversed for $\alpha < 0$ (negative correlation).

Note that while in the linear model $\kappa_r(\mathcal{H}_{ij};\alpha,\boldsymbol{\beta}) = \alpha_r(y_{ij-r} - \boldsymbol{x}'_{ij-r}\boldsymbol{\beta})$ could be used to obtain marginal interpretations for $\boldsymbol{\beta}$, i.e. $\mathrm{E}(Y_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}$ independent of the order q, this is difficult in log-linear or logistic models. The interpretation of $\boldsymbol{\beta}$ will usually depend on the model for the time dependence.

Fitting transition models

In a Markov model of order q, the likelihood contribution from the ith subject is

$$\begin{aligned}\mathcal{L}(y_{i1},\ldots,y_{in_i}) &= f(y_{i1},\ldots,y_{iq})\prod_{j=q+1}^{n_i} f(y_{ij}|\mathcal{H}_{ij}) \\ &= f(y_{i1},\ldots,y_{iq})\prod_{j=q+1}^{n_i} f(y_{ij}|y_{ij-1},\ldots,y_{ij-q}).\end{aligned}$$

If our model is correctly specified, then conditional on the past, the transitions for a person are independent. Thus, each term in the product contributes an independent univariate GLM.

For $f(y_{i1},\ldots,y_{iq})$, a separate model has to be specified to obtain the full likelihood, as this is not described by the conditional model.
For the Gaussian AR(q) models, assuming a multivariate normal distribution for $Y_{i1},\ldots,Y_{iq}$ nicely produces a multivariate normal distribution for $Y_{i1},\ldots,Y_{in_i}$ without additional unknown parameters. Inference can then proceed using the full likelihood.

Fitting transition models

Generally, however, $f(y_{i1}, \ldots, y_{iq})$ is not determined from the conditional model, and cannot be specified without additional unknown parameters.

An alternative (less efficient than ML, but best option without additional assumptions on $f(y_{i1}, \ldots, y_{iq})$) then is to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ using the conditional likelihood

$$\prod_{i=1}^{I} f(y_{iq+1}, \ldots, y_{in_i} | y_{i1}, \ldots, y_{iq}) = \prod_{i=1}^{I} \prod_{j=q+1}^{n_i} f(y_{ij} | \mathcal{H}_{ij}).$$

Now, in the case where $\kappa_r(\mathcal{H}_{ij}; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \alpha_r \kappa_r(\mathcal{H}_{ij})$ (as in the discussed logistic models), we have

$$g(\mu_{ij}^C) = \boldsymbol{x}_{ij}' \boldsymbol{\beta} + \sum_{r=1}^{s} \alpha_r \kappa_r(\mathcal{H}_{ij})$$

and $g(\mu_{ij}^C)$ is a linear function of both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_s)$. Estimation can then proceed using independent GLMs with extended explanatory variables $(\boldsymbol{x}_{ij}, \kappa_1(\mathcal{H}_{ij}), \ldots, \kappa_s(\mathcal{H}_{ij}))$.

Fitting transition models

A somewhat more complicated case arises when $\kappa_r(\mathcal{H}_{ij};\boldsymbol{\alpha},\boldsymbol{\beta})$ depends on both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (as in the linear and log-linear models discussed). In this case, the derivative of the conditional likelihood gives the conditional score equation for $\boldsymbol{\delta}=(\boldsymbol{\beta},\boldsymbol{\alpha})$

$$S^C(\boldsymbol{\delta})=\sum_{i=1}^{I}\sum_{j=q+1}^{n_i}\frac{\partial\mu_{ij}^C}{\partial\boldsymbol{\delta}}{v_{ij}^C}^{-1}(Y_{ij}-\mu_{ij}^C)=\mathbf{0},$$

the conditional analogue of the GLM score equation. $\frac{\partial\mu_{\boldsymbol{ij}}^{\boldsymbol{C}}}{\partial\boldsymbol{\delta}}$ can depend on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Letting $\boldsymbol{Y}_i=(Y_{iq+1},\ldots,Y_{in_i})$, $\boldsymbol{\mu}_i^C=(\mu_{iq+1}^C,\ldots,\mu_{in_i}^C)$, $\boldsymbol{W}_i=\mathrm{diag}(v_{iq+1}^C,\ldots,v_{in_i}^C)$ and $\boldsymbol{X}_i^*$ the $(n_i-q)\times(p+s)$ matrix with kth row $\frac{\partial\mu_{\boldsymbol{iq+k}}^{\boldsymbol{C}}}{\partial\boldsymbol{\delta}}$, gives

$$S^C(\boldsymbol{\delta})=\sum_{i=1}^{I}\boldsymbol{X}_i^*\boldsymbol{W}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}_i^C)=\mathbf{0}.$$

We can solve this equation for $\widehat{\boldsymbol{\delta}}$ using Fisher scoring, iteratively regressing the working variate $\boldsymbol{Z}$, where $\boldsymbol{Z}_i=\boldsymbol{X}_i^*\widehat{\boldsymbol{\delta}}+(\boldsymbol{Y}_i-\widehat{\boldsymbol{\mu}}_i^C)$, on $\boldsymbol{X}^*$ using weights $\boldsymbol{W}^{-1}$.

Inference in transition models

- If the model for conditional mean and variance is correctly specified, $\widehat{\boldsymbol{\delta}}$ is asymptotically (as $I \to \infty$) normal with mean $\boldsymbol{\delta}$ and $(p+s) \times (p+s)$ variance matrix

$$\boldsymbol{V}_{\delta} = \left(\sum_{i=1}^{I} \boldsymbol{X}_i^{*\prime} \boldsymbol{W}_i^{-1} \boldsymbol{X}_i^{*} \right)^{-1},$$

which can be consistently estimated by replacing $\boldsymbol{\alpha}$ by $\widehat{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}$.

- If the conditional mean is correctly specified, but the conditional variance is not, the asymptotic variance changes to

$$\left(\sum_{i=1}^{I} \boldsymbol{X}_i^{*\prime} \boldsymbol{W}_i^{-1} \boldsymbol{X}_i^{*} \right)^{-1} \left(\sum_{i=1}^{I} \boldsymbol{X}_i^{*\prime} \boldsymbol{W}_i^{-1} \boldsymbol{V}_i \boldsymbol{W}_i^{-1} \boldsymbol{X}_i^{*} \right) \left(\sum_{i=1}^{I} \boldsymbol{X}_i^{*\prime} \boldsymbol{W}_i^{-1} \boldsymbol{X}_i^{*} \right)^{-1},$$

which can be consistently estimated by replacing $\boldsymbol{V}_i = \mathrm{Var}(\boldsymbol{Y}_i | \mathcal{H}_{ij})$ by the estimate $(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i^C)(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i^C)'$, and $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ by $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$.

Inference in transition models

- When the Markov property is violated, using the robust variance will often still give consistent confidence intervals for $\widehat{\delta}$. However, the interpretation of $\widehat{\delta}$ becomes questionable, as then $\mu_{ij}^{C} = \mathrm{E}(Y_{ij}|y_{ij-1}, \ldots, y_{ij-q}) \neq \mathrm{E}(Y_{ij}|\mathcal{H}_{ij})$.

Transition Models in R

See `examples_transition_models.R` for example code for

- setting up a data set to fit first- or second-order transition models
- fit first- or second-order transition models for binary data

The data sets mentioned in the Longitudinal Data Analysis book Diggle et al (2002), and in particular the example used here, can be obtained from Patrick Heagerty's website:
`http://faculty.washington.edu/heagerty/Books/AnalysisLongitudinal/`