Analysing repeated measurements whilst accounting for derivative tracking, varying within-subject variance and autocorrelation: the xtiou command

R.A. Hughes*[1], M.G. Kenward[2], J.A.C. Sterne[1], K. Tilling[1]

[1]School of Social and Community Medicine
University of Bristol

[2]Luton, London

* Funded by the Medical Research Council

Introduction

- The linear mixed effects model (Laird and Ware, 1982) is commonly used to model biomarker trajectories
- Linear mixed effects (LME) model for subject i

$$Y_i = X_i\beta + Z_i u_i + e_i$$

  - fixed effects: β
  - random effects: $u_i \sim N(0, G)$
  - measurement errors: $e_i \sim N(0, \sigma^2 I)$
  - u_i and e_i are independent
- LME model assumes:
  - within subject errors are independent
  - variance of within subject errors is constant

Integrated Ornstein Uhlenbeck process

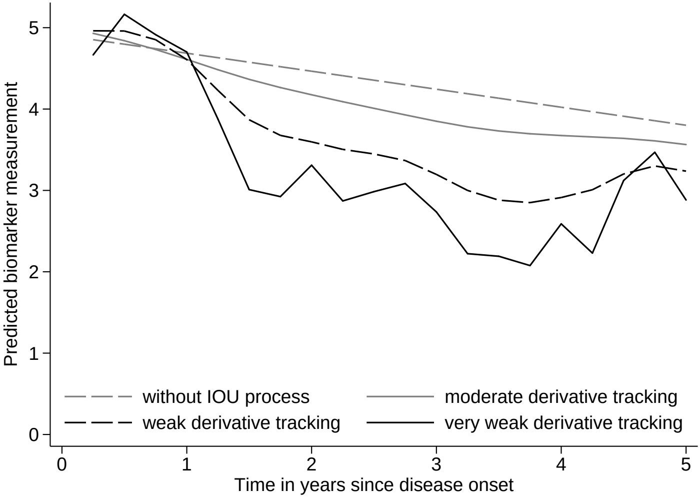
- Taylor et al (1994) proposed LME model with added Integrated Ornstein-Uhlenbeck (IOU) process
  - Linear Mixed Effects IOU (LME IOU) model

- IOU process quantifies the degree of derivative tracking
  - tendency of measurements to maintain the same trajectory
  - estimated from the data

- IOU process indexed by α and τ
  - small α and τ : strong derivative tracking
  - large α and τ : weak derivative tracking

- Special case: $\alpha \to \infty$ with τ/α held constant
  - scaled Brownian Motion (BM) process
  - BM process indexed by ϕ
  - Linear Mixed Effects BM (LME BM) model

Different degrees of derivative tracking



Linear mixed effects IOU (or BM) model

- LME IOU (or BM) model for subject i

$$Y_i = X_i\beta + Z_i u_i + w_i + e_i$$

  - w_i is independent of u_i and e_i
  - $w_i \sim N(0, H_i)$

IOU covariance function at time points s and t

$$\frac{\tau^2}{2\alpha^3}[2\alpha \min(s,t)+\exp(-\alpha s)+\exp(-\alpha t)-1-\exp(-\alpha \mid t-s \mid)]$$

BM covariance function at time points s and t

$$\phi s \qquad \text{if } s \leq t$$

- LME IOU (or BM) model also allows for:
  - correlated within subject error
  - variance of within subject errors can change over time

Estimation of the LME IOU (or BM) model

- Estimate variance parameters
  - components of random effects covariance matrix G
  - IOU parameters α and τ (or BM parameter ϕ)
  - measurement error variance σ^2
- REestricted Maximum Likelihood (REML)
  - Profile REML function with respect to σ^2
- Log-Cholesky parameterization for G
  - To ensure resulting estimate is positive semi-definite
- Optimization using Newton-Raphson type algorithms
  - Mata function `optimize`
- Wolfinger et al (1994)'s method to efficiently calculate log-likelihood and its 1st and 2nd derivatives
- Implemented in MATA

The `xtiou` command

- Fits the linear mixed effects IOU model
  - option to fit the linear mixed effects BM model
- Shares features of a Stata regression command
  - supports factor notation ([U] **11.4.3 Factor variables**)
  - supports maximization options ([R] **maximize**)
  - returns results in `e()`
  - supports `estimates`
- `predict` generates predictions under the fitted model:
  - fixed portion linear prediction
  - standard error of the fixed portion linear prediction
  - fitted values
  - residuals (response minus fitted values)

Default syntax of `xtiou`

`xtiou` *depvar* [*indepvars*] [*if*] [*in*] ,
`id(`*levelvar*`)` `time(`*timevar*`)` [`other_options`]

- Data required to be in long format
  - subjects at level-2
  - measurements at level-1
- Required options
  - `id(`*levelvar*`)` identifies subjects
  - `time(`*timevar*`)` defines the time variable for the measurements
- By default:
  - includes a constant term in the fixed portion
  - includes only a random intercept
  - includes an IOU process

Options for model structure

- `reffects`(*varlist*) defines the random-effects of the model
  - assumes an unstructured covariance matrix
  - factor variables not allowed
- `brownian` specifies a scaled Brownian Motion process
  - fits a LME BM model

Option for the starting values

- By default starting values derived assuming strong derivative tracking
    - fits linear mixed effects model using `mixed`
    - EM estimates used as starting values for random-effects covariance matrix and measurement error variance
    - IOU or BM parameters set to small positive values
- `svdataderived` derives starting values making no assumptions about derivative tracking
    - including IOU or Brownian Motion parameters
    - derived from variances and covariances of the observed measurements across subjects
    - assumes random effects includes either a random intercept and/or a random linear slope

Option for the IOU process

- `iou`(*ioutype*) specifies the parameterization of the IOU process used during estimation
- where *ioutype* is

ioutype	Description
`at`	alpha and tau, the default
`ao`	alpha and omega $=$ (tau $\div$ alpha)^2
`et`	eta $=$ ln(alpha) and tau
`eo`	eta $=$ ln(alpha) and omega $=$ (tau $\div$ alpha)^2
`it`	iota $=$ alpha^{-2} and tau
`eo`	iota $=$ alpha^{-2} and omega $=$ (tau $\div$ alpha)^2

- Changing IOU parameterization may improve convergence

Options for maximization

- By default uses modified Newton-Raphson algorithm
- `algorithm`(*algorithm_spec*) specifies one or more optimization algorithms
  - Newton-Raphson algorithm
  - Fisher-Scoring algorithm
  - Average-Information algorithm
- Includes maximize options ([R] **maximize**) common to Stata regression commands
  - `iterate(#)`, `nolog`, `trace`, `gradient`, `showstep`, `hessian`, `difficult`

Example

- Simulated data based on characteristics of a HIV cohort study (UK CHIC study 2004)
- Patient's CD4 cell counts measured every 3 months
- CD4 cell counts used to monitor a patient's:
  - response to therapy
  - HIV disease progression
- Patient characteristics
  - sex
  - age at start of therapy
  - ethnicity (white, black African, other)
  - risk for HIV infection (homosexual, heterosexual, other)
  - pre-therapy CD4 cell count group (0 to 99, 100 to 199, 200 to 349 and $\geq$ 350 cells/mm^3)

Simulated Data

- Unbalanced data of 1000 patients with up to 5 years of follow-up
- Patient characteristics simulated under general location model
  - categorical variables: multinomial distribution
  - continuous given categorical variables: Normal distribution
- Simulated repeated CD4 counts (natural log scale) under LME BM model
  - population $\ln CD4$ trajectory: fractional polynomial with powers 0 and 0.5
  - patient characteristics included as fixed effects
  - intercept and fractional powers included as random effects
  - BM process

Comparisons

- Fit LMEs with differing variance structures

Comparisons

- Fit LMEs with differing variance structures:
  ri: random intercept

  rfp: random intercept and fractional polynomial powers

Comparisons

- Fit LMEs with differing variance structures:
  ri: random intercept

  rfp: random intercept and fractional polynomial powers

  riiou: random intercept and IOU process

  ribm: random intercept and BM process

Comparisons

- Fit LMEs with differing variance structures:
  ri: random intercept

  rfp: random intercept and fractional polynomial powers

  riiou: random intercept and IOU process

  ribm: random intercept and BM process

  rfpiou: random intercept and fractional polynomial powers, and IOU process

  rfpbm: random intercept and fractional polynomial powers, and BM process

Comparisons

- Fit LMEs with differing variance structures:
  ri: random intercept

  rfp: random intercept and fractional polynomial powers

  riiou: random intercept and IOU process

  ribm: random intercept and BM process

  rfpiou: random intercept and fractional polynomial powers, and IOU process

  rfpbm: random intercept and fractional polynomial powers, and BM process

- All models have the same, correct mean structure

- Compare model fit and accuracy of patient-level predictions

Random intercept IOU model

- Fit the LME IOU model

```
xtiou lncd4 time_ln time_05 age sex i.risk ///
i.ethnicity ib2.baselinecd4, id(patid) time(time) svdata
```

- Post estimation

```
estimates store riiou_model
predict riiou_fit, fitted
predict riiou_res, residuals
```

Linear mixed IOU REML regression	Number of obs	=	15526
	Number of groups	=	1000
	Obs per group : min	=	2
	avg	=	15.5
Restricted log likelihood = -6169.4427	max	=	26

lncd4	Coef.	Std. Err.	z	P >|z|	[95% Conf.	Interval]
time_ln	.1232436	.0223509	5.51	0.000	.0794366	.1670506
time_05	.077378	.0500194	1.55	0.122	-.0206582	.1754142
age	-.0000926	.0014625	-0.06	0.950	-.002959	.0027738
sex	.0923211	.0441723	2.09	0.037	.0057449	.1788972
risk						
heterosexual	-.1314315	.0452229	-2.91	0.004	-.2200668	-.0427961
other risk	-.1403481	.0555603	-2.53	0.012	-.2492443	-.0314519
_cons	4.151499	.0803116	51.69	0.000	3.994091	4.308907

Variance parameters	Estimate	Std. Err.	[95% Conf.	Interval]
Random-effects:				
Var(_cons)	.1320698	.0080314	.1172301	.148788
IOU-effects:				
alpha	.9403315	.1105896	.7467442	1.184105
tau	.4873562	.0409801	.4133049	.5746751
Var(Measure. Err.)	.0747382	.0011132	.0725879	.0769522

```
Linear mixed IOU REML regression                Number of obs      =      15526
                                                Number of groups   =       1000

                                                Obs per group : min =          2
                                                                avg =       15.5
Restricted log likelihood = -6169.4427                          max =         26
```

lncd4	Coef.	Std. Err.	z	P >|z|	[95% Conf. Interval]	
time_ln	.1232436	.0223509	5.51	0.000	.0794366	.1670506
time_05	.077378	.0500194	1.55	0.122	-.0206582	.1754142
age	-.0000926	.0014625	-0.06	0.950	-.002959	.0027738
sex	.0923211	.0441723	2.09	0.037	.0057449	.1788972
risk						
heterosexual	-.1314315	.0452229	-2.91	0.004	-.2200668	-.0427961
other risk	-.1403481	.0555603	-2.53	0.012	-.2492443	-.0314519
_cons	4.151499	.0803116	51.69	0.000	3.994091	4.308907

Variance parameters	Estimate	Std. Err.	[95% Conf. Interval]	
Random-effects:				
Var(_cons)	.1320698	.0080314	.1172301	.148788
IOU-effects:				
alpha	.9403315	.1105896	.7467442	1.184105
tau	.4873562	.0409801	.4133049	.5746751
Var(Measure. Err.)	.0747382	.0011132	.0725879	.0769522

Linear mixed IOU REML regression

Number of obs	=	15526
Number of groups	=	1000
Obs per group : min	=	2
avg	=	15.5
max	=	26

Restricted log likelihood = -6169.4427

lncd4	Coef.	Std. Err.	z	P >|z|	[95% Conf.	Interval]
time_ln	.1232436	.0223509	5.51	0.000	.0794366	.1670506
time_05	.077378	.0500194	1.55	0.122	-.0206582	.1754142
age	-.0000926	.0014625	-0.06	0.950	-.002959	.0027738
sex	.0923211	.0441723	2.09	0.037	.0057449	.1788972
risk						
heterosexual	-.1314315	.0452229	-2.91	0.004	-.2200668	-.0427961
other risk	-.1403481	.0555603	-2.53	0.012	-.2492443	-.0314519
_cons	4.151499	.0803116	51.69	0.000	3.994091	4.308907

Variance parameters	Estimate	Std. Err.	[95% Conf.	Interval]
Random-effects:				
Var(_cons)	.1320698	.0080314	.1172301	.148788
IOU-effects:				
alpha	.9403315	.1105896	.7467442	1.184105
tau	.4873562	.0409801	.4133049	.5746751
Var(Measure. Err.)	.0747382	.0011132	.0725879	.0769522

```
Linear mixed IOU REML regression              Number of obs      =      15526
                                              Number of groups   =       1000

                                              Obs per group : min =          2
                                                              avg =       15.5
Restricted log likelihood = -6169.4427                        max =         26
```

lncd4	Coef.	Std. Err.	z	P >|z|	[95% Conf.	Interval]
time_ln	.1232436	.0223509	5.51	0.000	.0794366	.1670506
time_05	.077378	.0500194	1.55	0.122	-.0206582	.1754142
age	-.0000926	.0014625	-0.06	0.950	-.002959	.0027738
sex	.0923211	.0441723	2.09	0.037	.0057449	.1788972
risk						
heterosexual	-.1314315	.0452229	-2.91	0.004	-.2200668	-.0427961
other risk	-.1403481	.0555603	-2.53	0.012	-.2492443	-.0314519
_cons	4.151499	.0803116	51.69	0.000	3.994091	4.308907

Variance parameters	Estimate	Std. Err.	[95% Conf.	Interval]
Random-effects:				
Var(_cons)	.1320698	.0080314	.1172301	.148788
IOU-effects:				
alpha	.9403315	.1105896	.7467442	1.184105
tau	.4873562	.0409801	.4133049	.5746751
Var(Measure. Err.)	.0747382	.0011132	.0725879	.0769522

```
Linear mixed IOU REML regression                Number of obs      =      15526
                                                Number of groups   =       1000

                                                Obs per group : min =          2
                                                                avg =       15.5
Restricted log likelihood = -6249.6745                          max =         26
```

lncd4	Coef.	Std. Err.	z	P >|z|	[95% Conf.	Interval]
time_ln	.1283745	.0226364	5.67	0.000	.0840079	.1727412
time_05	.0690668	.0467146	1.48	0.139	-.0224921	.1606258
age	-.0001694	.0014558	-0.12	0.907	-.0030227	.0026839
sex	.0946172	.044012	2.15	0.032	.0083553	.1808791
risk						
heterosexual	-.1316994	.0450399	-2.92	0.003	-.219976	-.0434228
other risk	-.1305444	.05534	-2.36	0.018	-.2390088	-.02208
_cons	4.162428	.0797391	52.20	0.000	4.006142	4.318714

Variance parameters	Estimate	Std. Err.	[95% Conf.	Interval]
Random-effects:				
Var(_cons)	.1110791	.0079717	.0965037	.1278559
BM-effects:				
phi	.1377509	.0038615	.1303865	.1455313
Var(Measure. Err.)	.0597721	.0010262	.0577943	.0618177

Compare model fit

```
. estimates stats ///
> ri_model riiou_model ribm_model ///
> rfp_model rfpbm_model rfpiou_model
```

(*output omitted*)

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)

Model	AIC	BIC
random intercept only	22481	22589
random intercept & IOU	12371	12493
random intercept & BM	12529	12644
random fractional powers	12793	12938
random fractional powers & IOU	12130	12267
random fractional powers & BM	12128	12258

Compare model fit

```
. estimates stats ///
> ri_model riiou_model ribm_model ///
> rfp_model rfpbm_model rfpiou_model
```

(*output omitted*)

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)

Model	AIC	BIC
random intercept only	22481	22589
random intercept & IOU	12371	12493
random intercept & BM	12529	12644
random fractional powers	12793	12938
random fractional powers & IOU	12130	12267
random fractional powers & BM	12128	12258

Compare model fit

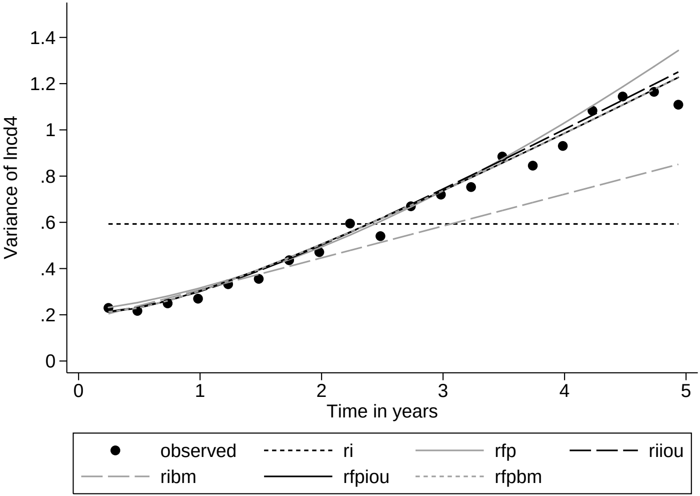
```
. estimates stats ///
> ri_model riiou_model ribm_model ///
> rfp_model rfpbm_model rfpiou_model
```

(*output omitted*)

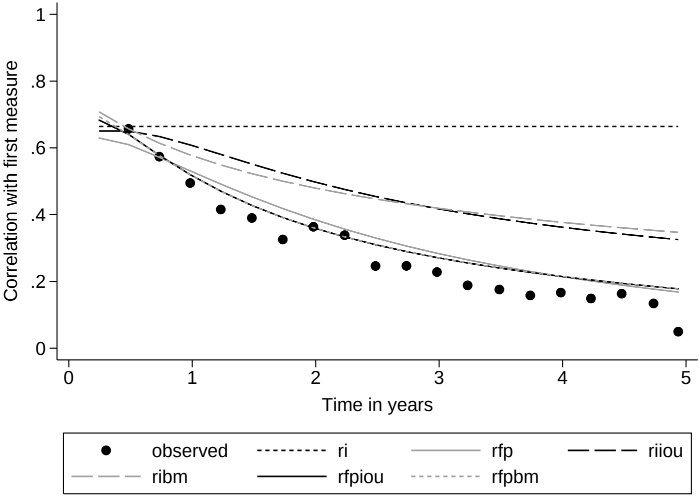
- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)

Model	AIC	BIC
random intercept only	22481	22589
random intercept & IOU	12371	12493
random intercept & BM	12529	12644
random fractional powers	12793	12938
random fractional powers & IOU	12130	12267
random fractional powers & BM	12128	12258

Changes in variance over time



Changes in correlation over time



Comparison of the fitted values

- Average squared difference between predicted and observed measurements
  - Mean Squared Error (MSE)
- Number of predicted measurements within 5% of the observed

Model	MSE	Within 5%
random intercept only	0.1867	5970
random intercept & IOU	0.0597	8844
random intercept & BM	0.0382	10441
random fractional powers	0.0727	8227
random fractional powers & IOU	0.0491	9522
random fractional powers & BM	0.0465	9738

Comparison of the fitted values

- Average squared difference between predicted and observed measurements
  - Mean Squared Error (MSE)

- Number of predicted measurements within 5% of the observed

Model	MSE	Within 5%
random intercept only	0.1867	5970
random intercept & IOU	0.0597	8844
random intercept & BM	0.0382	10441
random fractional powers	0.0727	8227
random fractional powers & IOU	0.0491	9522
random fractional powers & BM	0.0465	9738

Comparison of the fitted values

- Average squared difference between predicted and observed measurements
  - Mean Squared Error (MSE)
- Number of predicted measurements within 5% of the observed

Model	MSE	Within 5%
random intercept only	0.1867	5970
random intercept & IOU	0.0597	8844
random intercept & BM	0.0382	10441
random fractional powers	0.0727	8227
random fractional powers & IOU	0.0491	9522
random fractional powers & BM	0.0465	9738

Discussion

- `xtiou` fits LME IOU model or LME BM model
- These models allow for
  - autocorrelation
  - changing within subject variance
  - incorporation of derivative tracking
- Options available to solve convergence problems
  - `svdataderived`
  - `iou(`*ioutype*`)`
  - `algorithm(`*algorithm_spec*`)`
  - `difficult`
- Accompanying `predict` command
  - Does not provide BLUPs of random effects nor realizations of IOU (or BM) process
- Hope our command will help statisticians apply the LME IOU model and LME BM model to their data

References

- Laird N and Ware J (1982) Random-Effects Models for Longitudinal Data *Biometrics* **38**: 963-974.
- Taylor JMG, Cumberland WG and Sy PJ (1994) A stochastic model for analysis of longitudinal AIDS data *Journal of the American Statistical Association* **89**: 727-736.
- UK Collaborative HIV Cohort Steering Committee (2004) The creation of a large UK based multicentre cohort of HIV-infected individuals: the UK Collaborative HIV Cohort (UK CHIC) Study *HIV Medicine* **5**: 115-124.
- Wolfinger R, Tobias R and Sall J (1994) Computing Gaussian likelihoods and their derivatives for the general linear mixed model *SIAM J Sci Comput* **15**: 1294-1310.