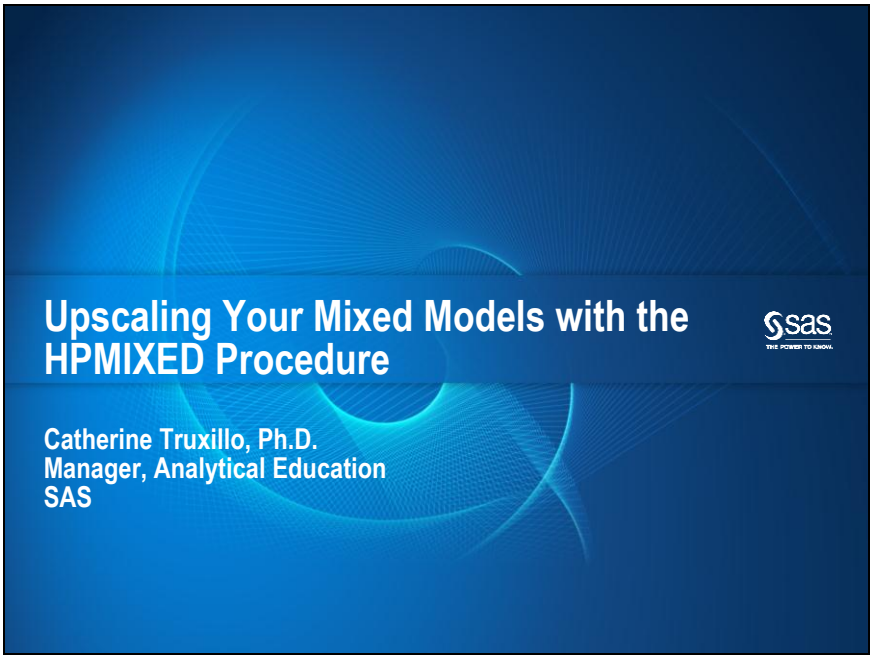
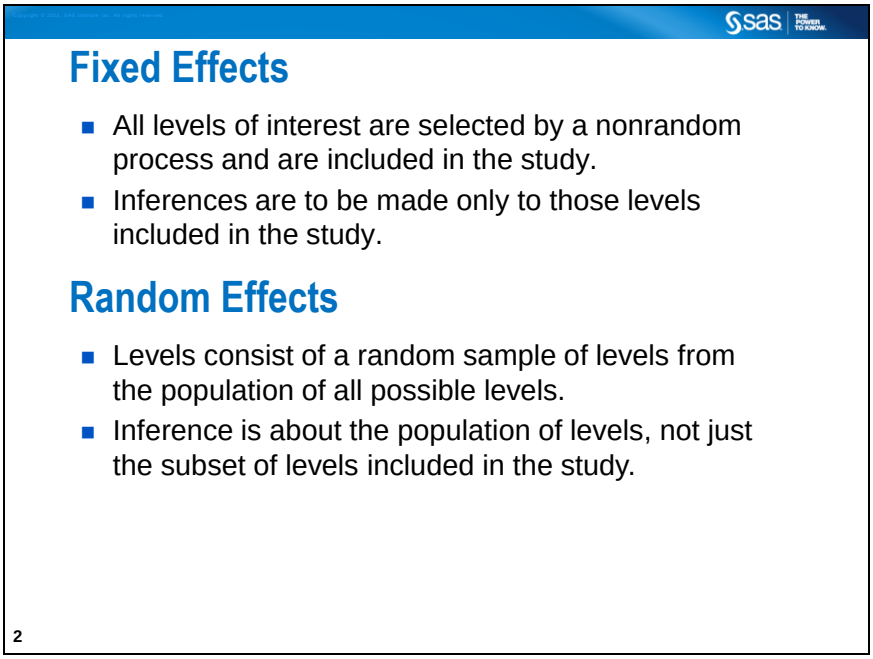




Mixed Models

Models in which some factors are fixed effects and other factors are random effects are called *mixed models*.

Random: physicians, agents, clinics, financial advisors

Fixed: drugs, specialties, training programs, automobile type

3

The General Linear Model

$$y = X\beta + \varepsilon$$

Design matrix

Random errors

Response

Fixed effects

Assume $E[\varepsilon] = 0, Var[\varepsilon] = \sigma^2 I_n$

Therefore, $E[y] = X\beta, Var[y] = \sigma^2 I_n$

4

The Linear Mixed Model

Random effects

$$y = \mathbf{X}\beta + \mathbf{Z}\gamma + \varepsilon$$

Design matrix for random effects

No longer required to be independent and homogeneous

Assume $\gamma \sim N(0, \mathbf{G})$ and $\varepsilon \sim N(0, \mathbf{R})$

$$E(y) = \mathbf{X}\beta, \quad Var(y) = \mathbf{ZGZ}' + \mathbf{R} = \mathbf{V}$$

5

Statements in the MIXED and HPMIXED Procedures

Specified in the REPEATED statement for non-default structures

$$y = \mathbf{X}\beta + \mathbf{Z}\gamma + \varepsilon$$

Specified in the MODEL statement

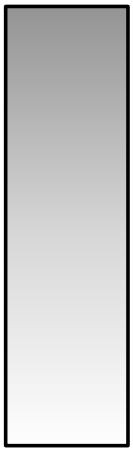
Specified in the RANDOM statement

6

Massive Mixed Models

A mixed model can be large because of:

- Many observations (huge N)

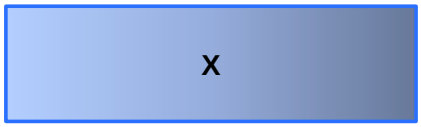


7

Massive Mixed Models

A mixed model can be large because of:

- Many observations (huge N)
- Lots of fixed effects or lots of levels of fixed effects (Wide **X**)



8

Massive Mixed Models

A mixed model can be large because of:

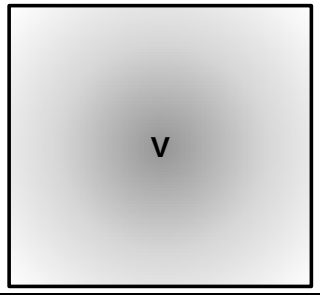
- Many observations (huge N)
- Lots of fixed effects or lots of levels of fixed effects (Wide **X**)
- Lots of levels of random effects (Wide **Z**)

9

Massive Mixed Models

A mixed model can be large because of:

- Many observations (huge N)
- Lots of fixed effects or lots of levels of fixed effects (Wide **X**)
- Lots of levels of random effects (Wide **Z**)
- Lots of different covariance parameters (very complex **G** and/or **R**)



10

Massive Mixed Models

A mixed model can be large because of:

- Many observations (huge N)
- Lots of fixed effects or lots of levels of fixed effects (Wide **X**)
- Lots of levels of random effects (Wide **Z**)
- Lots of different covariance parameters (very complex **G** and/or **R**)

HPMIXED is made for this

11

Massive Mixed Models

A mixed model can be large because of:

- Many observations (huge N)
- Lots of fixed effects or lots of levels of fixed effects (Wide **X**)
- Lots of levels of random effects (Wide **Z**)
- Lots of different covariance parameters (complex **G** and/or **R**)

MIXED works well in this situation, and for a large variety of model types. MIXED was originally designed for small experimental studies and has been improved to meet customer needs.

12

In most massive mixed models, many of the matrices have relatively few unique values and lots of zeroes

$$\begin{bmatrix}
V & C_1 & C_2 & C_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & V & C_2 & C_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & V & C_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & V & C_1 & C_2 & C_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & V & C_2 & C_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & V & C_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & & V & C_1 & C_2 & C_2 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & & & V & C_2 & C_2 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & & & & V & C_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & & & & & V & C_1 & C_2 & C_2 & 0 & 0 & 0 \\
 & & & & & & & & & & V & C_2 & C_2 & 0 & 0 & 0 \\
 & & & & & & & & & & & V & C_1 & 0 & 0 & 0 \\
 & & & & & & & & & & & & V & C_1 & C_2 & C_2 \\
 & & & & & & & & & & & & & V & C_2 & C_2 \\
 & & & & & & & & & & & & & & V & C_1 \\
 & & & & & & & & & & & & & & & V
\end{bmatrix}$$

where

$V = \sigma^2 + \sigma_b^2 + \sigma_{ab}^2$

$C_1 = \sigma_b^2 + \sigma_{ab}^2$

$C_2 = \sigma_b^2$

13

Massive Mixed Models

- Specialized algorithms in PROC HPMIXED can take advantage of this sparseness and repetitiveness to perform certain calculations much more quickly.

$$\begin{bmatrix} V \\ C_1 \\ C_2 \end{bmatrix}$$

where

$V = \sigma^2 + \sigma_b^2 + \sigma_{ab}^2$

$C_1 = \sigma_b^2 + \sigma_{ab}^2$

$C_2 = \sigma_b^2$

nonzero elements are are indexed and stored

14

Simulated Example: Health Insurance Payout

Fixed: 5 specialty areas, 100 different diagnosis codes (nested)
Random: 3000 providers, need eblups and CLs for each.
DV: Payout (4000 payouts per provider)

$$payout_{ijkm} = \mu + \alpha_i + (\alpha\beta)_{ij} + b_k + \varepsilon_{ijkm}$$

specialty effect, fixed

specialty*diagnosis effect, fixed

physician effect, random

$$b_k \sim N(0, \sigma_b^2),\ \varepsilon_{ijkm} \sim N(0, \sigma^2)$$

therefore,

$$E(y_{ijkm}) = \mu + \alpha_i + (\alpha\beta_{ij})$$
$$Var(y_{ijkm}) = \sigma_b^2 + \sigma^2$$

15

Syntax

```
proc hpmixed data=wiilus.Sim;
   class Specialty Diagnosis Physician;
   model Payout = Specialty Diagnosis*Specialty;
   random Physician/cl;
   test Specialty Diagnosis*Specialty;
run;
```

```
proc mixed data=wiilus.Sim;
   class Specialty Diagnosis Physician;
   model Payout = Specialty Diagnosis*Specialty;
   random Physician/cl;
run;
```

16

Example

```
proc hpmixed data=Sim;
...
NOTE: PROCEDURE HPMIXED used (Total process time):
      real time          2.71 seconds
      cpu time           2.66 seconds

proc mixed data=Sim;
...
  random physician/cl;
NOTE: PROCEDURE MIXED used (Total process time):
      real time          4:05.12
      cpu time           5:04.71
...
```

17

HPMIXED Output

Model Information	
Data Set	WORK.SIM
Response Variable	Payout
Estimation Method	Restricted Maximum Likelihood (REML)
Degrees of Freedom Method	Residual

Class Level Information		
Class	Levels	Values
Specialty	5	1 2 3 4 5
Diagnosis	100	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 ...
Physician	3000	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 ...

Number of Observations Read	40000
Number of Observations Used	40000

19

MIXED Output

Model Information	
Data Set	WORK.SIM
Dependent Variable	Payout
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Specialty	5	1 2 3 4 5
Diagnosis	100	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
Physician	3000	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541

3000 physician IDs

20

Output

HPMIXED

Dimensions	
G-side Cov. Parameters	1
R-side Cov. Parameters	1
Columns in X	505
Columns in Z	3000
Subjects (Blocks in V)	1

Number of Observations Read	40000
Number of Observations Used	40000

MIXED

Dimensions	
Covariance Parameters	2
Columns in X	505
Columns in Z	3000
Subjects	1
Max Obs Per Subject	40000

Number of Observations	
Number of Observations Read	40000
Number of Observations Used	40000
Number of Observations Not Used	0

21

Output

HPMIXED

Covariance Parameter Estimates	
Cov Parm	Estimate
Physician	38.7042
Residual	70.2725

Fit Statistics	
-2 Res Log Likelihood	287422
AIC (smaller is better)	287426
AICC (smaller is better)	287426
BIC (smaller is better)	287438
CAIC (smaller is better)	287440
HQIC (smaller is better)	287431

MIXED

Covariance Parameter Estimates	
Cov Parm	Estimate
Physician	38.7036
Residual	70.2726

Fit Statistics	
-2 Res Log Likelihood	287422.3
AIC (smaller is better)	287426.3
AICC (smaller is better)	287426.3
BIC (smaller is better)	287438.3

22

HPMIXED Output (Partial)

Solution for Random Effects									
Effect	Physician	Estimate	Std Err Pred	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Physician	1	10.1387	4.6518	39501	2.18	0.0293	0.05	1.0211	19.2563
Physician	2	7.7243	2.8116	39501	2.75	0.0060	0.05	2.2135	13.2352
Physician	3	0.2596	3.1002	39501	0.08	0.9333	0.05	-5.8170	6.3361
Physician	4	2.1841	4.0981	39501	0.53	0.5941	0.05	-5.8482	10.2164
Physician	5	-8.4203	2.8863	39501	-2.92	0.0035	0.05	-14.0774	-2.7632
Physician	6	0.6139	3.3146	39501	0.19	0.8531	0.05	-5.8829	7.1107
Physician	7	1.2282	2.8618	39501	0.43	0.6678	0.05	-4.3810	6.8373
Physician	8	-1.2713	3.2358	39501	-0.39	0.6944	0.05	-7.6134	5.0709
Physician	9	-4.1698	4.7089	39501	-0.89	0.3759	0.05	-13.3993	5.0598
Physician	10	5.7538	3.2293	39501	1.78	0.0748	0.05	-0.5757	12.0834
Physician	11	1.7662	2.9558	39501	0.60	0.5501	0.05	-4.0272	7.5596
Physician	12	15.4249	3.7319	39501	4.13	<.0001	0.05	8.1102	22.7396

(3000 physician eblups)

23

MIXED Output (Partial)

Solution for Random Effects

Effect	Physician	Estimate	Std Err Pred	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Physician	1	10.1387	4.6517	37E3	2.18	0.0293	0.05	1.0211	19.2562
Physician	2	7.7243	2.8116	37E3	2.75	0.0060	0.05	2.2135	13.2352
Physician	3	0.2596	3.1002	37E3	0.08	0.9333	0.05	-5.8170	6.3361
Physician	4	2.1841	4.0980	37E3	0.53	0.5941	0.05	-5.8482	10.2163
Physician	5	-8.4203	2.8862	37E3	-2.92	0.0035	0.05	-14.0774	-2.7632
Physician	6	0.6139	3.3146	37E3	0.19	0.8531	0.05	-5.8829	7.1106
Physician	7	1.2282	2.8618	37E3	0.43	0.6678	0.05	-4.3810	6.8373
Physician	8	-1.2713	3.2357	37E3	-0.39	0.6944	0.05	-7.6134	5.0709
Physician	9	-4.1698	4.7089	37E3	-0.89	0.3759	0.05	-13.3993	5.0598
Physician	10	5.7538	3.2293	37E3	1.78	0.0748	0.05	-0.5757	12.0834
Physician	11	1.7662	2.9558	37E3	0.60	0.5501	0.05	-4.0272	7.5596
Physician	12	15.4249	3.7319	37E3	4.13	<.0001	0.05	8.1102	22.7395

3000 physician eblups

24

Output

HPMIXED

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Specialty	4	39501	262.49	<.0001
Diagnosis(Specialty)	494	39501	11.49	<.0001

MIXED

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Specialty	4	37E3	262.50	<.0001
Specialty*Diagnosis	494	37E3	11.49	<.0001

25

Parameterization in HPMIXED

Parameterization of (fixed) effects

- MIXED uses MODEL, CLASS, and class level information to order terms and to compute generalized inverses.
  - Last level of class variables set to 0 by default, or you can specify the reference level.
- HPMIXED automatically reorders terms to find singularities quickly and to compute generalized inverses.
  - Must look at parameter estimates to see how model is parameterized.
- **Hypothesis tests of effects and predicted values are invariant to parameterization.**

26

Add This Syntax:

```
proc hpmixed data=wiilus.Sim;
   …
   model Payout = Specialty Diagnosis*Specialty/S;
   …
run;
```

```
proc mixed data=wiilus.Sim;
   …
   model Payout = Specialty Diagnosis*Specialty/S;
   …
run;
```

27

HPMIXED and MIXED Parameterizations

HPMIXED

Solution for Fixed Effects							
Effect	Specialty	Diagnosis	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			306.72	2.3361	39501	131.30	<.0001
Specialty	1		6.6865	3.2145	39501	2.08	0.0375
Specialty	2		-3.8301	4.0741	39501	-0.94	0.3472
Specialty	3		0	.	.	.	.
Specialty	4		16.6819	3.2265	39501	5.17	<.0001
Specialty	5		14.5886	3.5735	39501	4.08	<.0001
Specialty*Diagnosis	1	1	-10.5412	3.6961	39501	-2.85	0.0043
Specialty*Diagnosis	1	2	-7.8958	3.5207	39501	-2.24	0.0249
Specialty*Diagnosis	1	3	-10.8295	3.4975	39501	-3.10	0.0020

…+ 500 Specialty*Diagnosis estimates

Reference level is determined by speed of processing.

28

HPMIXED and MIXED Parameterizations

MIXED

Solution for Fixed Effects							
Effect	Specialty	Diagnosis	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			351.28	4.6396	2501	75.71	<.0001
Specialty	1		-22.1085	5.7156	37E3	-3.87	0.0001
Specialty	2		-19.0010	5.2856	37E3	-3.59	0.0003
Specialty	3		-13.9090	5.0178	37E3	-2.77	0.0056
Specialty	4		-7.2728	5.2861	37E3	-1.38	0.1689
Specialty	5		0	.	.	.	.
Specialty*Diagnosis	1	1	-26.2992	4.4640	37E3	-5.89	<.0001
Specialty*Diagnosis	1	2	-23.6539	4.3199	37E3	-5.48	<.0001
Specialty*Diagnosis	1	3	-26.5876	4.3010	37E3	-6.18	<.0001

…+ 500 Specialty*Diagnosis estimates

Reference level is the **last** CLASS level by default

(can optionally specify **first** level)

29

Degrees of Freedom Estimates in HPMIXED

Denominator degrees of freedom for fixed effects are usually different

- MIXED uses different defaults depending on the model, with options to use *containment, Kenward-Roger*, *Satterthwaite*, *residual*, or *between-within*.
- HPMIXED uses *residual* by default, and has option to use *none* (t-tests are equivalent to z-tests)
- **p-values for fixed effect tests can be different.**

30

HPMIXED and MIXED Output

HPMIXED

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Specialty	4	39501	262.49	<.0001
Diagnosis(Specialty)	494	39501	11.49	<.0001

MIXED

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Specialty	4	37E3	262.50	<.0001
Specialty*Diagnosis	494	37E3	11.49	<.0001

31

Covariance Structures in HPMIXED

- The covariance structures permitted in HPMIXED are a subset of the structures in MIXED.
- The most common **G**-side structures are supported (for random coefficient and multilevel models).
- In SAS 9.3 and newer, limited **R**-side structures (REPEATED) are also supported.

32

Post-Processing in HPMIXED

- HPMIXED supports many of the same post-fitting statements as MIXED, including CONTRAST, ESTIMATE, and LSMEANS.
- Not all MIXED functionality is in PROC HPMIXED.
- There are no ODS graphics in HPMIXED.
- You can save the covariance parameter estimates from HPMIXED and use them in PROC MIXED to leverage additional functionality.

33

Post-Processing in HPMIXED

```
proc hpmixed data=Sim;
   class Specialty Diagnosis Physician;
   model Payout = Specialty Diagnosis*Specialty;
   random Physician;
   ods output covparms=a;
run;

proc mixed data=Sim plots=…;
   class Specialty Diagnosis Physician;
   model Payout = Specialty Diagnosis*Specialty;
   random Physician;
   parms /pdata=a hold=1,2 noiter;
   lsmeans specialty / pdiff=all adjust=tukey;
run;
```

34

What About PROC HPLMIXED?

- HP**L**MIXED is a **new** high-performance procedure, based on PROC MIXED that uses distributed processing to solve large problems.
- Its options are a subset of PROC MIXED functionality.
- At this time, HP**L**MIXED uses dense matrix calculations; HPMIXED uses sparse matrix calculations.
- HPLMIXED is still a very young procedure and will have new functionality added *in future releases*.

35

What About PROC HPLMIXED?

- This paper only addresses SAS/STAT 12.2 functionality.
- In some cases, HPMIXED is faster than HP**L**MIXED; in other cases, HP**L**MIXED is faster than HPMIXED.
- Consider the type of massive mixed model you are fitting and the distributed environment you are using.
- Look for new developments in HPLMIXED in the future!

36

Concluding Points

- PROC MIXED is designed to be an all-purpose procedure for fitting linear mixed models, with a wide variety of options, covariance structures, and post-fitting inference and graphs.
- Some models involve very large matrix calculations, and those matrices only involve a limited number of variance-covariance parameters.
- In many of these situations, PROC HPMIXED makes it possible to fit linear mixed models rapidly and efficiently.
- Results from PROC HPMIXED can be used as input to PROC MIXED for additional options.

37

To Learn More

Special Topics in Mixed Models course from SAS (premiers in 2014)

http://support.sas.com/training

PROC HPMIXED documentation

http://support.sas.com/documentation/cdl/en/statug/65328/HTML/default/viewer.htm#statug_hpmixed_toc.htm

SAS Global Forum 2012 paper by Kiernan, et al.

http://support.sas.com/resources/papers/proceedings12/332-2012.pdf

38

Acknowledgments

Thanks to the following for their feedback on an earlier draft of this paper:

Jill Tao
Kathleen Kiernan
Lee Bennett
Mike Patetta
Chip Wells

39