Mixed Effects Multivariate Adaptive Splines Model for the Analysis of Longitudinal and Growth Curve Data

Heping Zhang

Yale University

NIA Workshop
on Statistical Methods for Longitudinal Data on Aging
June 14, 2007

Multivariate Adaptive Splines Model for the Analysis of Longitudinal Data (MASAL)

Email: Heping.Zhang@Yale.EDU
Website: http://peace.med.yale.edu
Or http://c2s2.yale.edu

Motivation

In ordinary regression models where observations are independent, we face challenges including variable selection, model selection, model diagnostics, …

Do we get a break with the longitudinal data?

NO!

Why don’t we deal with them often?

Too difficult!

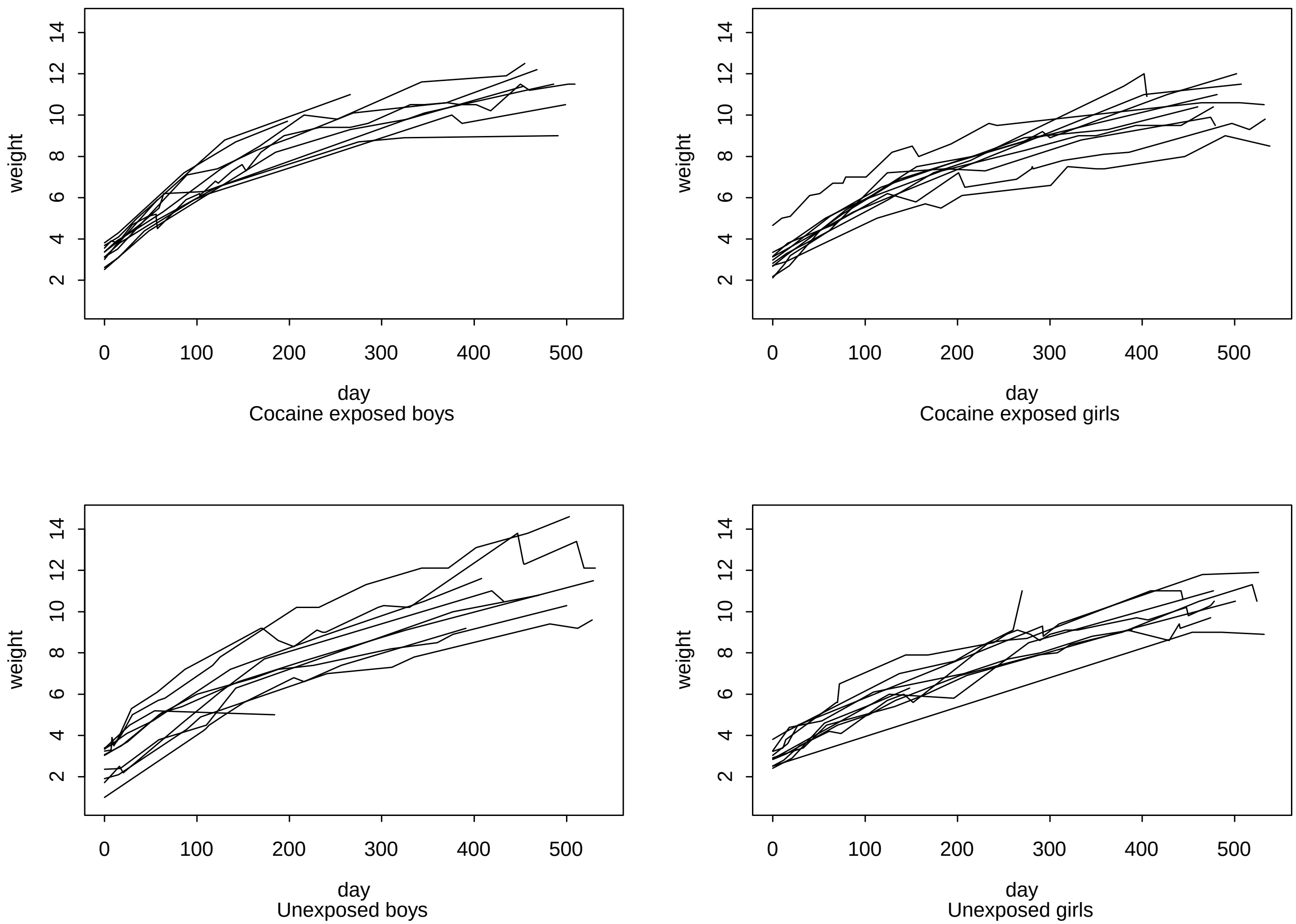
Outline

1. An Example
2. A Brief Review
3. Multivariate Adaptive Splines
4. The MASAL Algorithm
5. Usage of the MASAL Program
6. An Application and Model Interpretation
7. Relation to Other Models
8. Concluding Remarks
9. References on MASAL

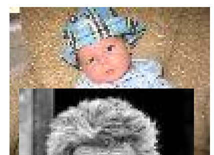
Growth Curve Data

- A retrospective study on infant growth (Dr. J. Leventhal)
- 298 children born at Yale-New Haven Hospital from September 1, 1989 through September 30, 1990.
- To examine the potential impact of the mother's cocaine use during pregnancy on the infant's growth after birth.

Infant Growth Curves

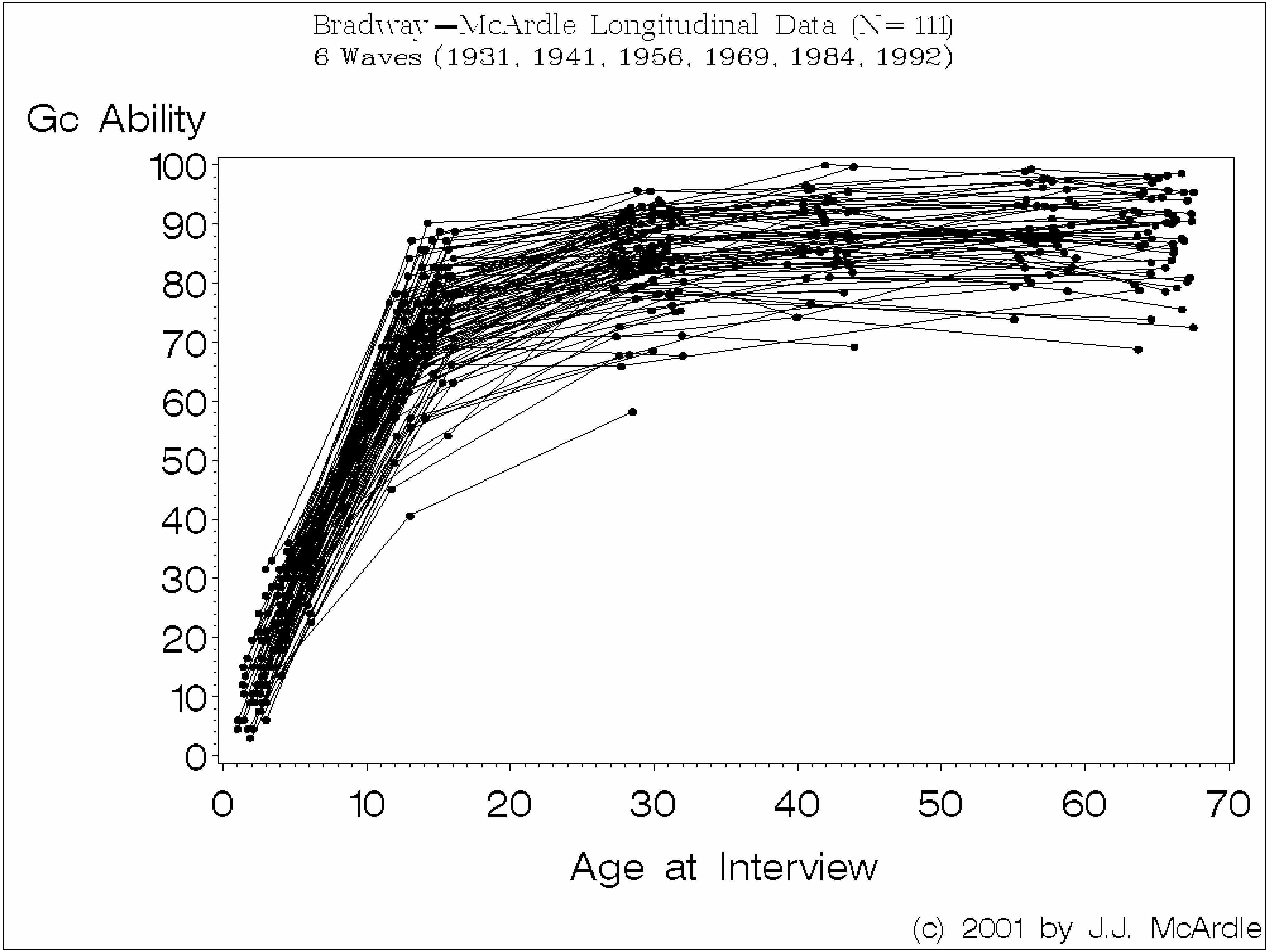


Aging?



ars

150 years


Bradway—McArdle Longitudinal Data (N= 111)
6 Waves (1931, 1941, 1956, 1969, 1984, 1992)
Gc Ability
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
Age at Interview
(c) 2001 by J.J. McArdle

Data Configuration

Subject	Occasion (visit)		
	1	…	q
1	$t_{11}, x_{1,11}, \ldots, x_{p,11}, y_{11}$		$t_{1,q1}, x_{1,1q1}, \ldots, x_{p,1q1}, y_{1q1}$
⋮	⋮	⋮	⋮
i	$t_{i1}, x_{1,i1}, \ldots, x_{p,i1}, y_{i1}$	…	$t_{i,qi}, x_{1,iqi}, \ldots, x_{p,iqi}, y_{iqi}$
⋮	⋮	⋮	⋮
n	$t_{n1}, x_{1,n1}, \ldots, x_{p,n1}, y_{n1}$	…	$t_{n,qn}, x_{1,nqn}, \ldots, x_{p,nqn}, y_{nqn}$

General Model

$$y_{ij} = f(x_{1,ij}, \ldots, x_{p,ij}, t_{ij}) + e_{ij}$$

y_{ij} : Response

$x_{k,ij}$: The *k*-th covariate (can be time dependent)

t_{ij} : Time of measurements

e_{ij} : Measurement error (can be correlated).

Fixed Effects

Laird and Ware (1982) : $f(x_{1,ij},\ldots,x_{p,ij}) = X_{ij}\beta$

Zeger and Diggle (1994) : $f(x_{1,ij},\ldots,x_{p,ij}) = X_{ij}\beta + \mu(t_{ij})$

Zhang (1997) : $f(x_{1,ij},\ldots,x_{p,ij})$ – smooth function

Hoover, Rice, Wu, & Yang (1998) : $f(x_{1,ij},\ldots,x_{p,ij}) = X_{ij}\beta(t_{ij})$

Random Effects

Laird and Ware (1982) : $Z_{ij}b_i + e_{ij}$

Zeger and Diggle (1994) : $W_i(t_{ij}) + e_{ij}$

Zhang (1997) : $e_i(t_{ij})$

Hoover, Rice, Wu, & Yang (1998) : $e_i(t_{ij})$

Brumback & Rice (1998); Rice & Wu (2001) :

$e_i(t_{ij}) = Z_i b_i(t_{ij}) + e_{ij}$

Meredith & Tisak (1990); Staniswalis & Lee (1998);

Zhang (2004) : $\sum_{u=1}^{v} b_{iu}\phi_u(t_{ij}) + e_{ij}$

Random Effect Structure

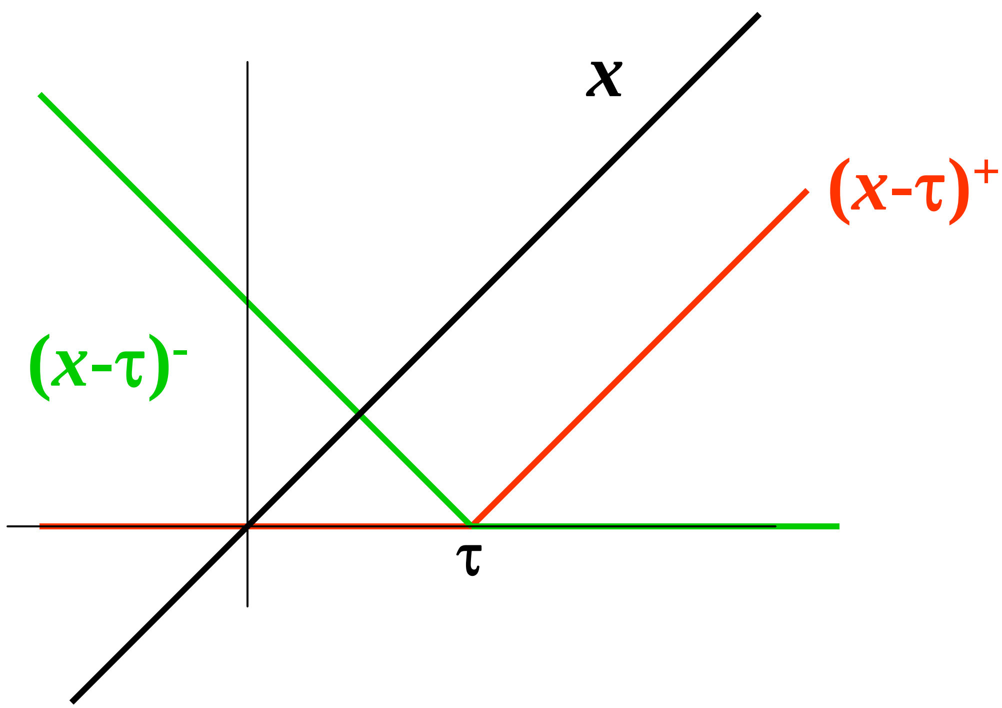
$$Z_{ij}b_i = \sum_{u=1}^{v} b_{iu}\phi_u(t_{ij}) + e_{ij}$$

where each ϕ_u is a simple function of time t such as a linear trend, which results in a quadratic variance function.

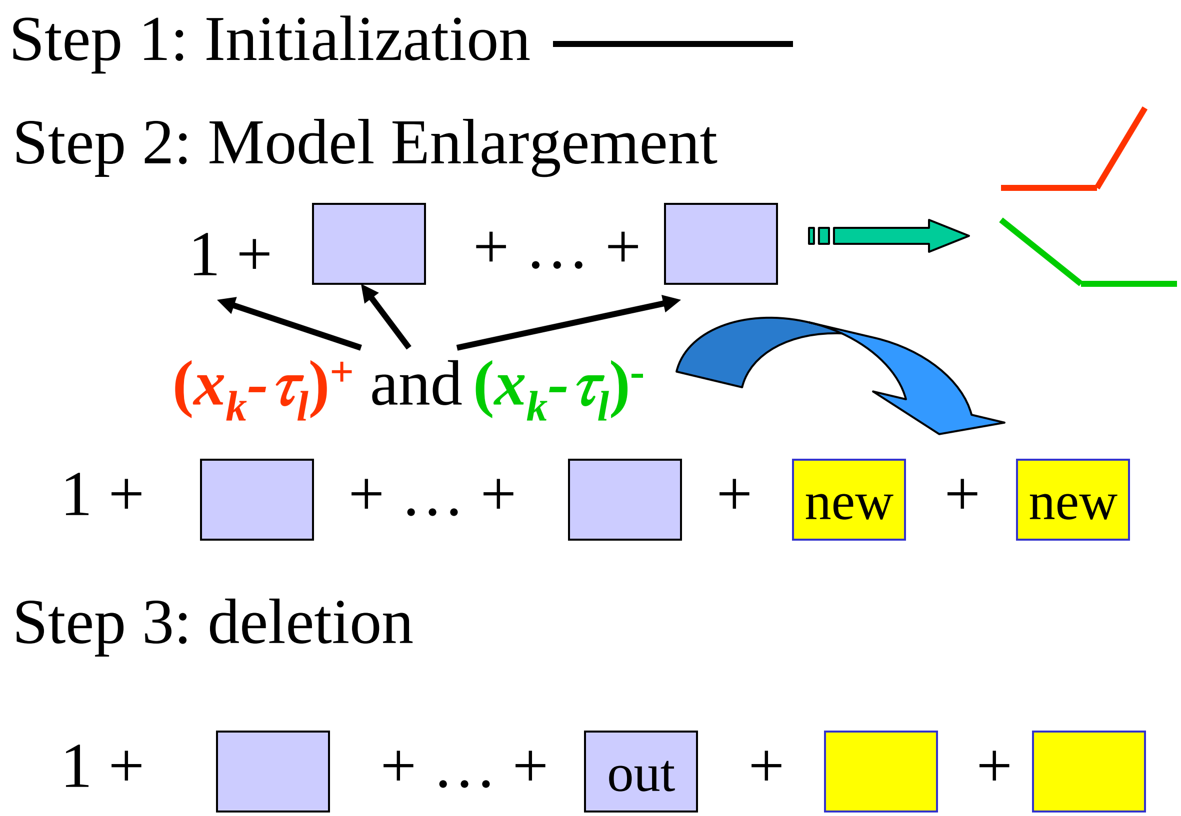
Multivariate Adaptive Splines

The function *f* is approximated and estimated by a function from the following class of functions:

$$\beta_0 + \sum \beta_{kl}(x_k - \tau_l)^* + \sum \beta_{klqs}(x_k - \tau_l)^*(x_q - \tau_s)^* + \cdots$$



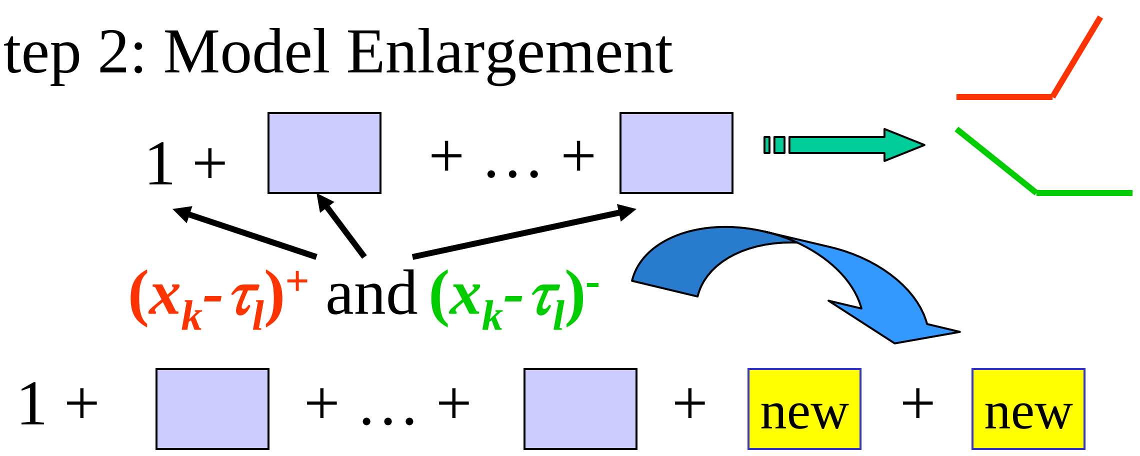
Schematic MASAL Algorithm



Schematic MASAL Algorithm

Step 1: Initialization

Step 2: Model Enlargement

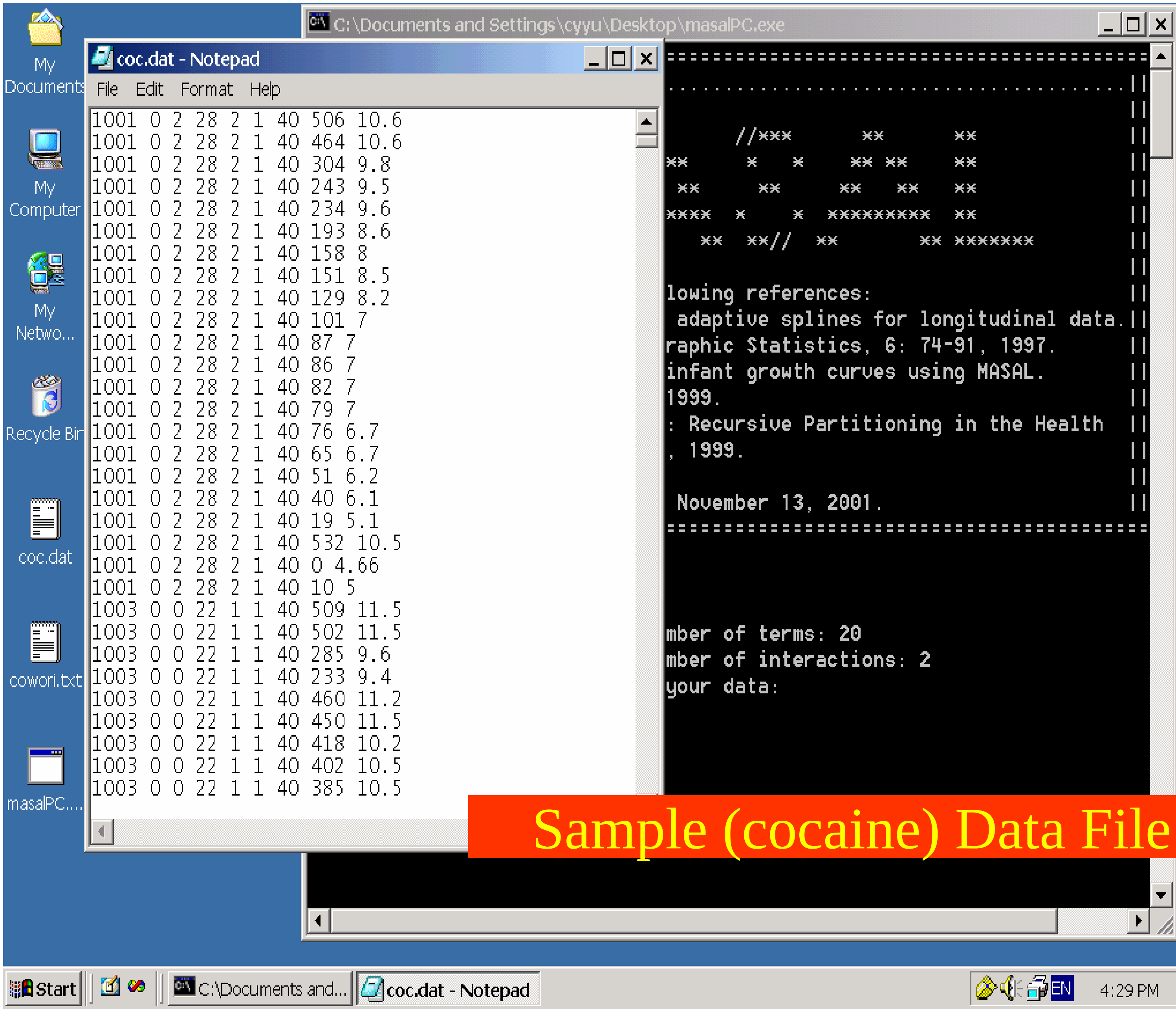


Step 3: deletion

Estimation of Random Effects

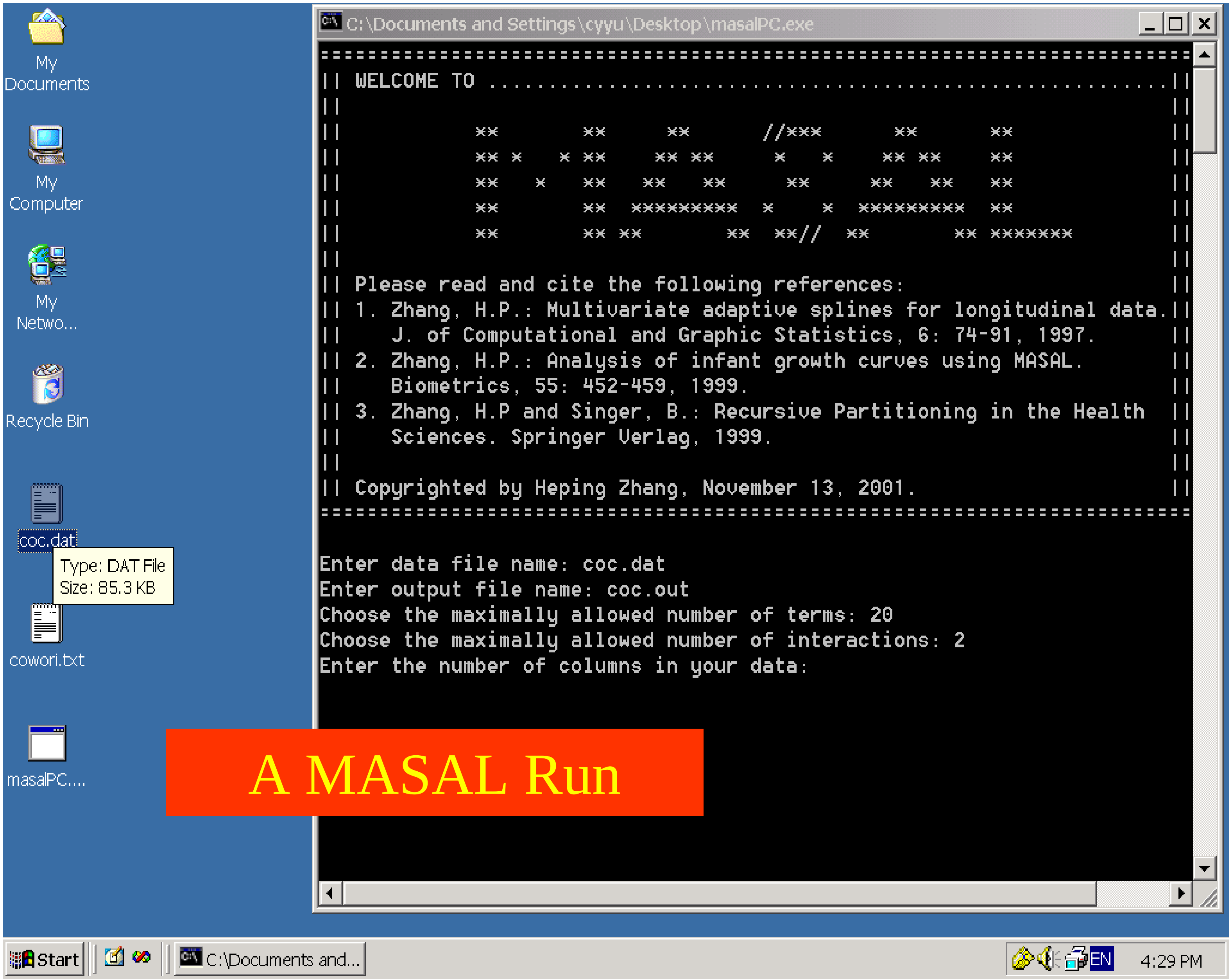
1. Begin with assuming independent data and proceed with the MASAL estimation of the fix effect, *i.e.*, the function *f*.
2. Obtain the residuals and use them to find the MLE of the covariance parameters, *e.g.*, pretending that the residuals are centered and normal.
3. Use the estimated covariance parameters as if known and decompose the residuals to obtain the estimates for b_{iu}.

Step 3 gives the BLUP, and the process is referred to as empirical Bayes estimation.

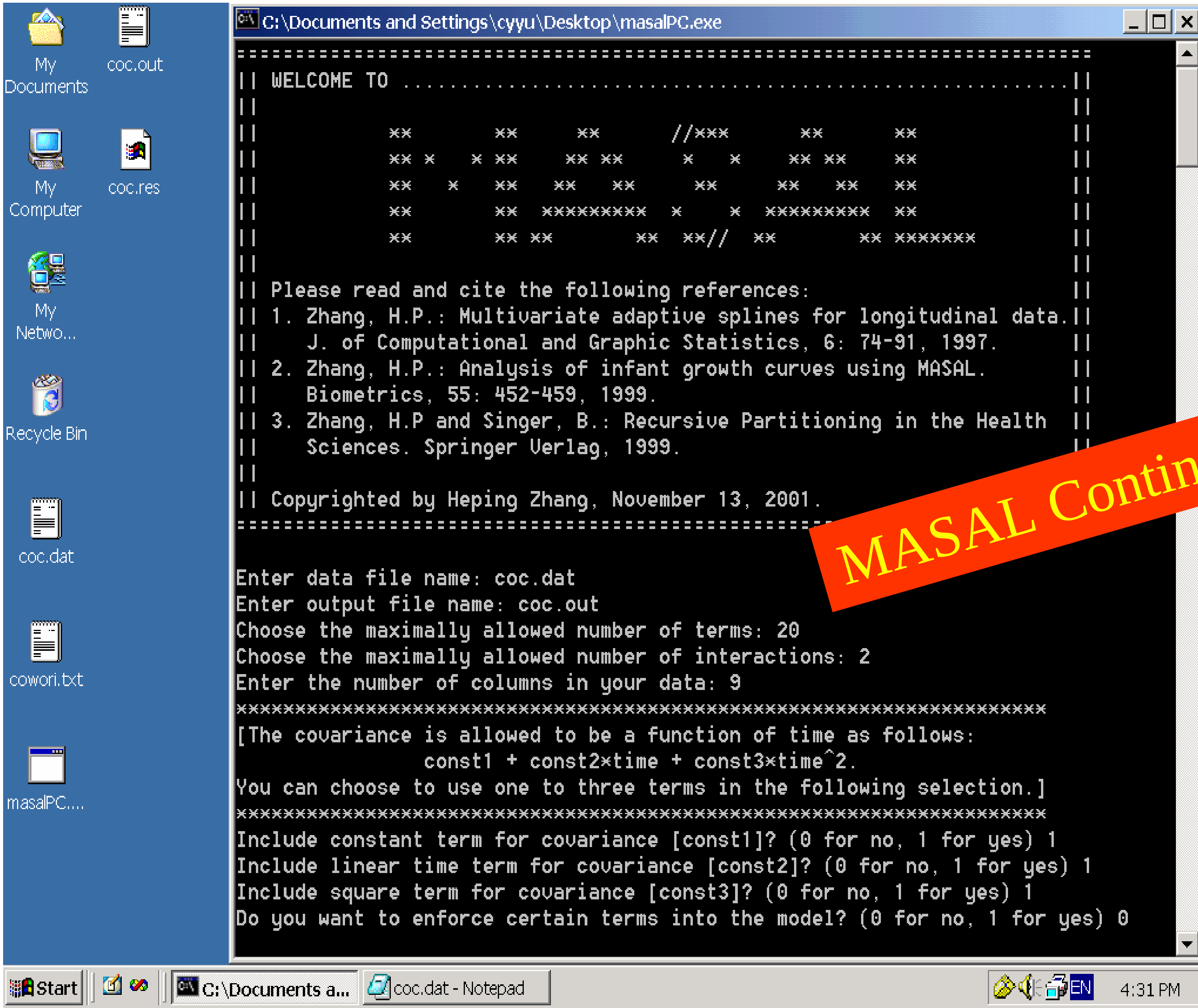


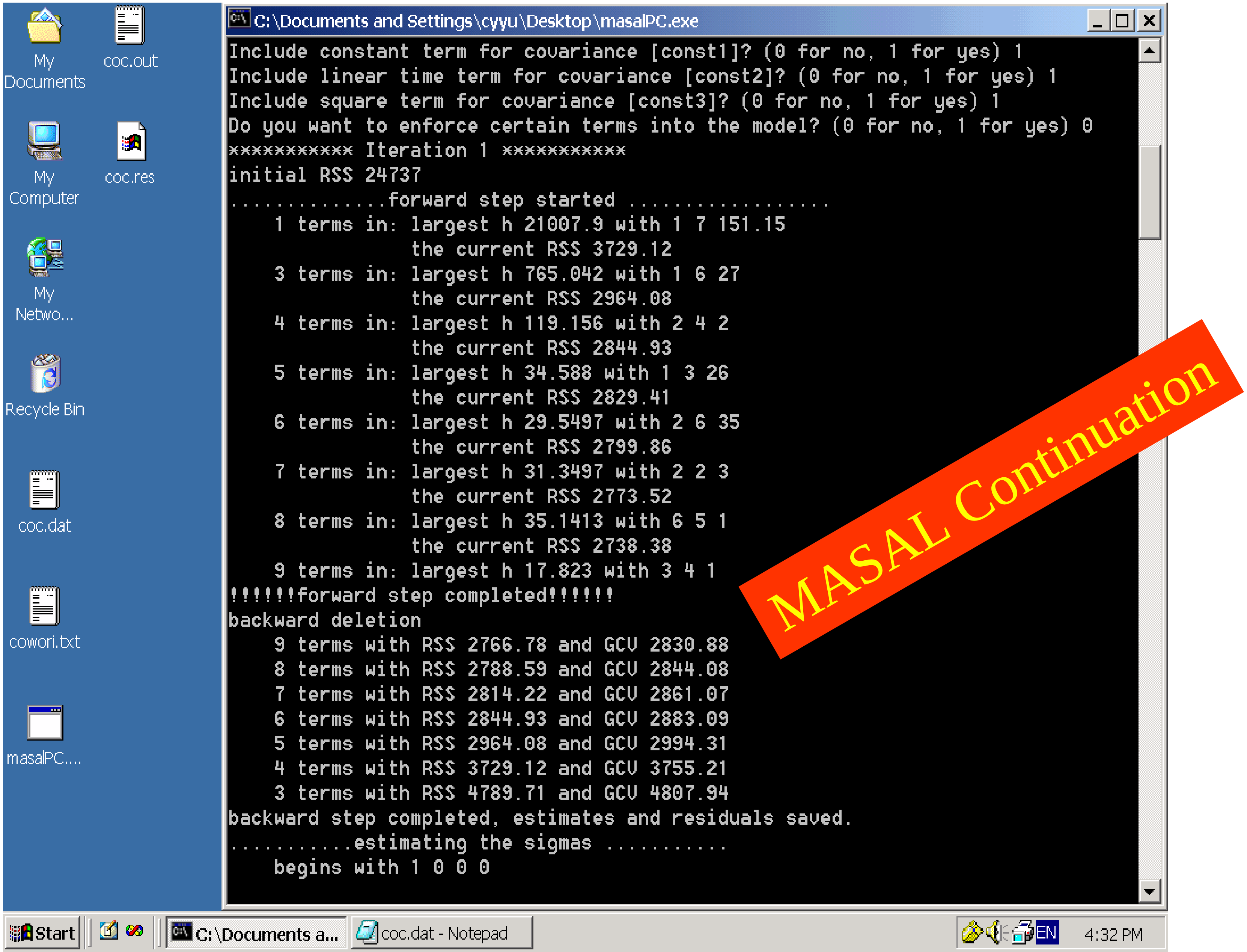
Sample (cocaine) Data File

A MASAL Run

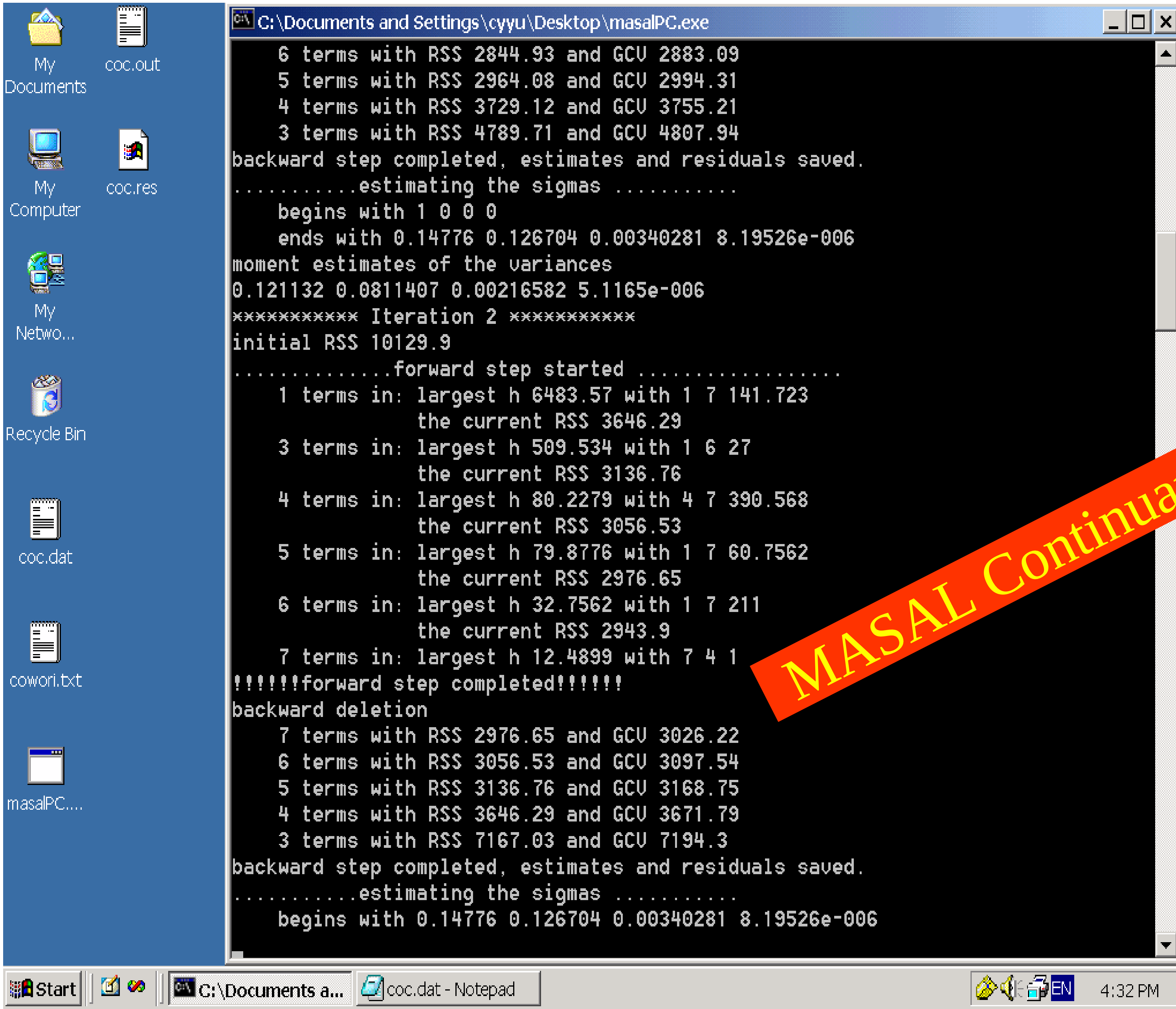

C:\Documents and Settings\cyyu\Desktop\masalPC.exe
	WELCOME TO	
	Please read and cite the following references:	
	1. Zhang, H.P.: Multivariate adaptive splines for longitudinal data.	
	J. of Computational and Graphic Statistics, 6: 74-91, 1997.	
	2. Zhang, H.P.: Analysis of infant growth curves using MASAL.	
	Biometrics, 55: 452-459, 1999.	
	3. Zhang, H.P and Singer, B.: Recursive Partitioning in the Health	
	Sciences. Springer Verlag, 1999.	
	Copyrighted by Heping Zhang, November 13, 2001.	
Enter data file name: coc.dat
Enter output file name: coc.out
Choose the maximally allowed number of terms: 20
Choose the maximally allowed number of interactions: 2
Enter the number of columns in your data:
My Documents
My Computer
My Netwo...
Recycle Bin
coc.dat
Type: DAT File
Size: 85.3 KB
cowori.txt
masalPC....
Start
C:\Documents and...
EN
4:29 PM

MASAL Continuation

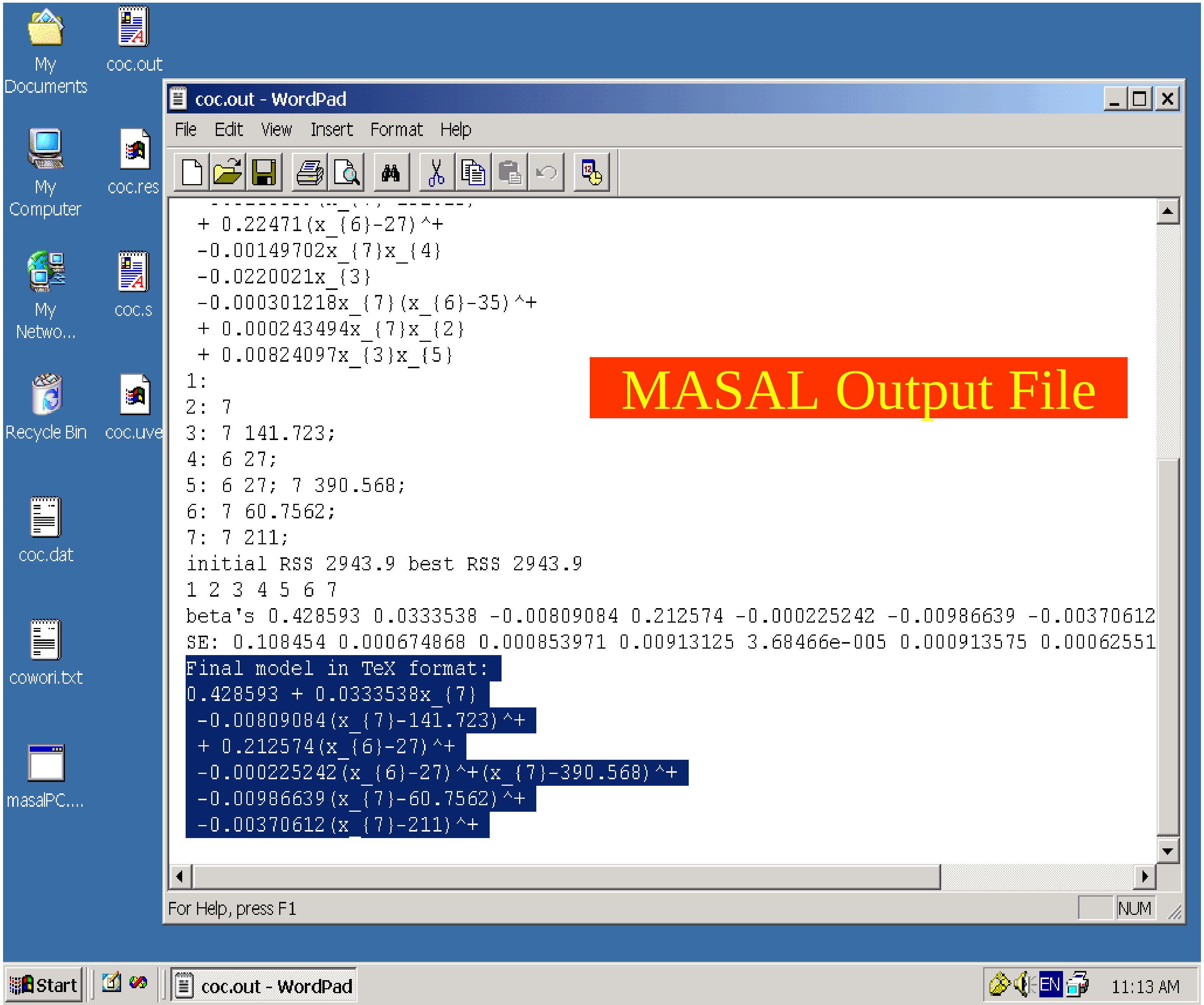



MASAL Continuation
C:\Documents and Settings\cyyu\Desktop\masalPC.exe
Include constant term for covariance [const1]? (0 for no, 1 for yes) 1
Include linear time term for covariance [const2]? (0 for no, 1 for yes) 1
Include square term for covariance [const3]? (0 for no, 1 for yes) 1
Do you want to enforce certain terms into the model? (0 for no, 1 for yes) 0
*********** Iteration 1 ***********
initial RSS 24737
..............forward step started ....
1 terms in: largest h 21007.9 with 1 7 151.15
the current RSS 3729.12
3 terms in: largest h 765.042 with 1 6 27
the current RSS 2964.08
4 terms in: largest h 119.156 with 2 4 2
the current RSS 2844.93
5 terms in: largest h 34.588 with 1 3 26
the current RSS 2829.41
6 terms in: largest h 29.5497 with 2 6 35
the current RSS 2799.86
7 terms in: largest h 31.3497 with 2 2 3
the current RSS 2773.52
8 terms in: largest h 35.1413 with 6 5 1
the current RSS 2738.38
9 terms in: largest h 17.823 with 3 4 1
!!!!!!forward step completed!!!!!!
backward deletion
9 terms with RSS 2766.78 and GCV 2830.88
8 terms with RSS 2788.59 and GCV 2844.08
7 terms with RSS 2814.22 and GCV 2861.07
6 terms with RSS 2844.93 and GCV 2883.09
5 terms with RSS 2964.08 and GCV 2994.31
4 terms with RSS 3729.12 and GCV 3755.21
3 terms with RSS 4789.71 and GCV 4807.94
backward step completed, estimates and residuals saved.
...........estimating the sigmas ....
begins with 1 0 0 0
My Documents
coc.out
My Computer
coc.res
My Netwo...
Recycle Bin
coc.dat
cowori.txt
masalPC....
Start
C:\Documents a...
coc.dat - Notepad
EN
4:32 PM

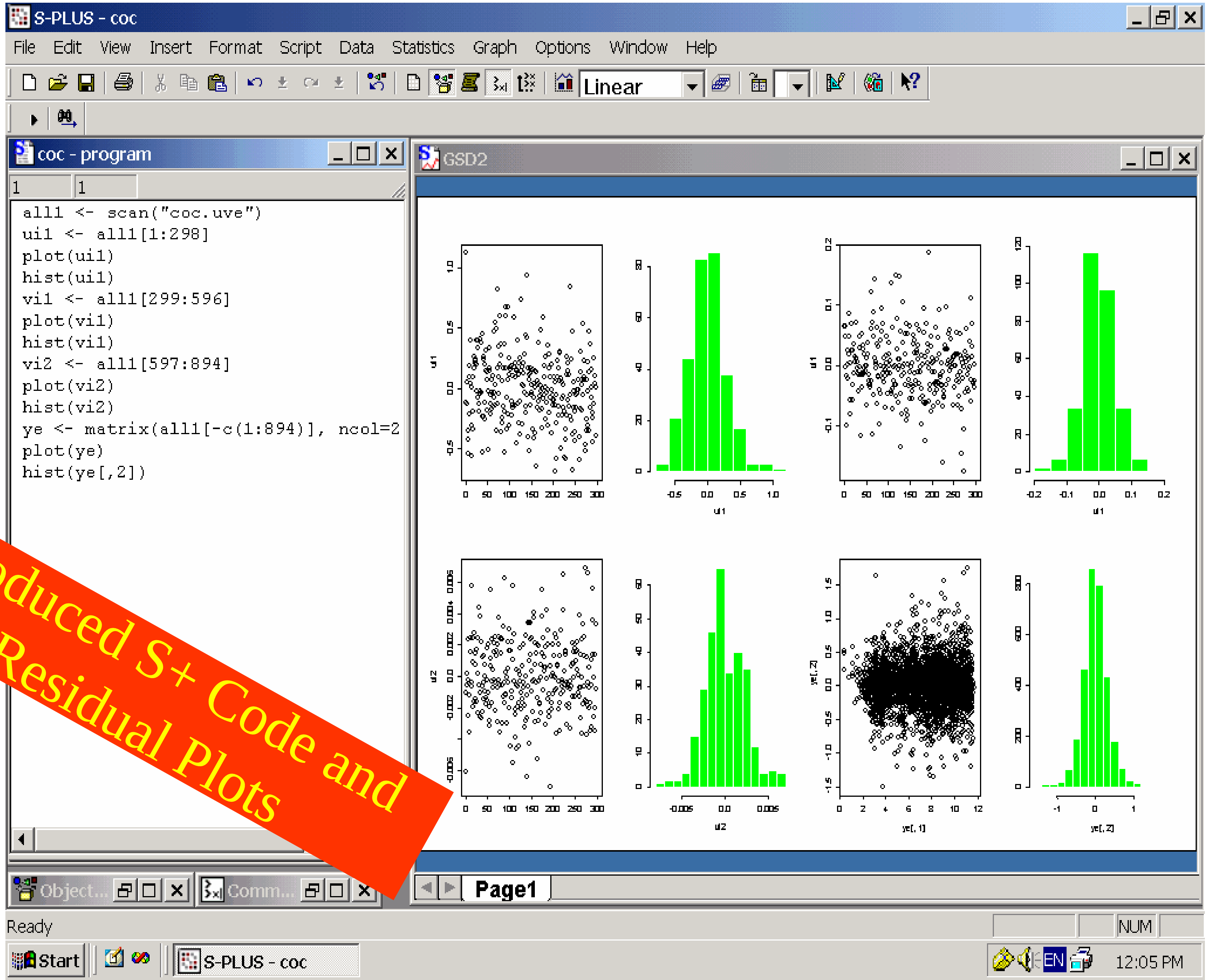
MASAL Continuation



MASAL Output File



Produced S+ Code and Residual Plots



Reformulation

MASAL final

```
0.43 + 0.03x_{7}-0.008(x_{7}-142)^+
+0.21(x_{6}-27)^+
-0.0002(x_{6}-27)^+(x_{7}-391)^+
-0.01(x_{7}-61)^+-0.004(x_{7}-211)^+
```

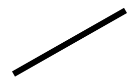
Where x_{7}: day; x_{6} : gestational age

$0.43 + 0.03\, t - 0.01(t-2)^{+} - 0.008(t-5)^{+} - 0.004(t-7)^{+}$
$+ (g_a-27)^{+}[0.21-0.0002\,(t-13)^{+}],$
where *t* is represented in month.

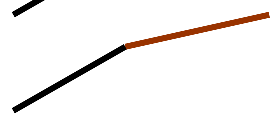
Interpretation

$$0.43 + 0.03\,t - 0.01(t-2)^{+} - 0.008(t-5)^{+} - 0.004(t-7)^{+}$$
$$+ (g_a - 27)^{+}[0.21 - 0.0002\,(t-13)^{+}],$$

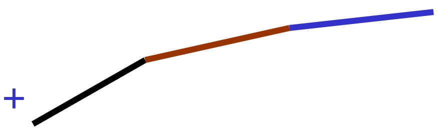
0.03 t



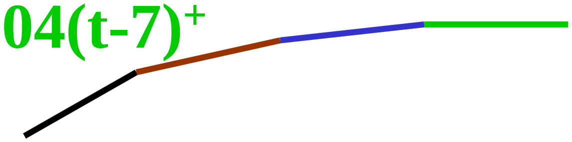
0.03 t - 0.01(t-2)$^{+}$



0.03 t - 0.01(t-2)$^{+}$ - 0.008(t-5)$^{+}$



0.03 t - 0.01(t-2)$^{+}$ - 0.008(t-5)$^{+}$ -0.004(t-7)$^{+}$



Interpretation

$$0.43 + 0.03\, t - 0.01(t-2)^{+} - 0.008(t-5)^{+} - 0.004(t-7)^{+}$$
$$+ (g_a - 27)^{+}[0.21 - 0.0002\,(t-13)^{+}],$$

$(g_a-27)^{+}[0.21-0.0002\,(t-13)^{+}]$ applies to those born 27 weeks or later:

- **$(g_a-27)^{+}\ 0.21$: A net gain linear to the gestational age.**
- **$(g_a-27)^{+}[-0.0002\,(t-13)^{+}]$: A slower growth after 13 months.**

A Different Representation

MASAL Model

$0.43 + 0.03\,t - 0.01(t-2)^+ - 0.008(t-5)^+ - 0.004(t-7)^+$
$+ (g_a-27)^+[0.21-0.0002\,(t-13)^+]$,
where *t* is represented in month.

Varying Coefficient Model

$$\beta_0(t) + X\,\beta_1(t)$$

Concluding Remarks

A general and flexible multivariate nonparametric model for longitudinal and curve data analysis

- Model multifactorial nature of growth trajectories or disease progression
- Allow irregular time points of observations
- Identify groups in lack of change (or slow rate of change)
- Allow different growth patterns at different ages
- Allow different growth patterns for different individuals (in group)

Concluding Remarks

• A user-friendly, free program (ready to execute or DLL for S-PLUS, R, …)
• Intuitive, readily interpretable models.

References on MASAL

- Zhang, H.P. Maximum correlation and splines. *Technometrics*, 36:196-201, 1994.
- Zhang, H.P. Multivariate adaptive splines for longitudinal data. *Journal of Computational and Graphic Statistics*, 6: 74-91, 1997.
- Zhang, H.P. Analysis of infant growth curves using MASAL. *Biometrics*, 55: 452-459, 1999.
- Zhang, H.P. Multivariate adaptive splines in the analysis of longitudinal and growth curve data. *Statistical Methods in Medical Research*, 13, 63-82, 2004.
- Zhang, H.P. Multivariate adaptive splines in the analysis of longitudinal data. *Encyclopedia of Biostatistics*, 2nd Edition, 5, 3463–3466, Wiley, Chichester, England, 2004.