

10th

SENSOMETRICS



SEM for small samples

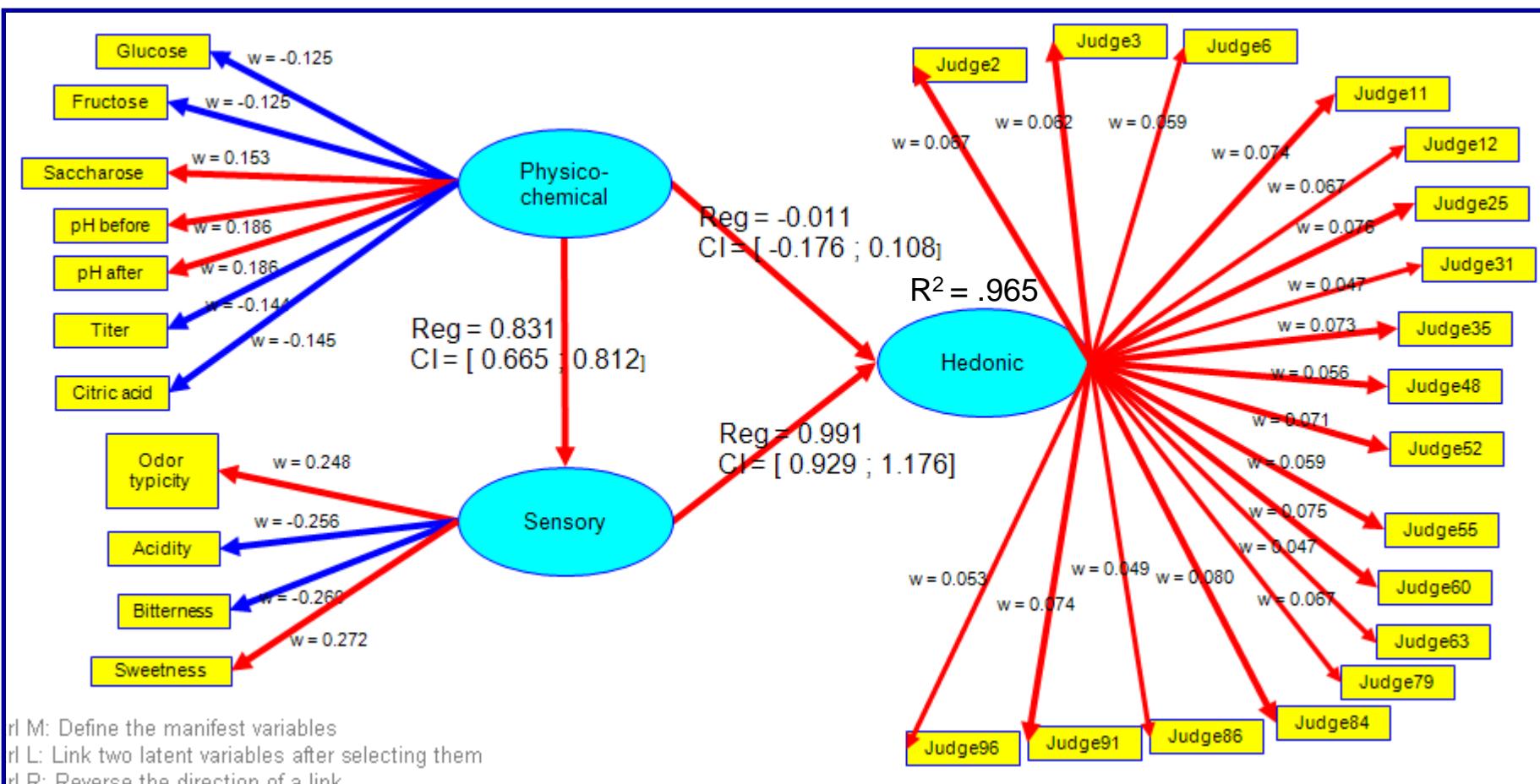
Michel Tenenhaus

Orange juice example (J. Pagès)

	PAMPRYL r.t.	TROPICANA r.t.	FRUIVITA refr.	JOKER r.t.	TROPICANA refr.	PAMPRYL refr.
Glucose	25.32	17.33	23.65	32.42	22.70	27.16
Fructose	27.36	20.00	25.65	34.54	25.32	29.48
Saccharose	36.45	44.15	52.12	22.92	45.80	38.94
Sweetening power	89.95	82.55	102.22	90.71	94.87	96.51
pH before processing	3.59	3.89	3.85	3.60	3.82	3.68
pH after centrifugation	3.55	3.84	3.81	3.58	3.78	3.66
Titer	13.98	11.14	11.51	15.75	11.80	12.21
Citric acid	.84	.67	.69	.95	.71	.74
Vitamin C	43.44	32.70	37.00	36.60	39.50	27.00
Smell Intensity	2.82	2.76	2.83	2.76	3.20	3.07
Odor typicality	2.53	2.82	2.88	2.59	3.02	2.73
Pulp	1.66	1.91	4.00	1.66	3.69	3.34
Taste intensity	3.46	3.23	3.45	3.37	3.12	3.54
Acidity	3.15	2.55	2.42	3.05	2.33	3.31
Bitterness	2.97	2.08	1.76	2.56	1.97	2.63
Sweetness	2.60	3.32	3.38	2.80	3.34	2.90
Judge 1	2.00	2.00	3.00	2.00	4.00	3.00
Judge 2	1.00	3.00	3.00	2.00	4.00	1.00
Judge 3	2.00	3.00	4.00	2.00	3.00	1.00
.						
.						
.						
Judge 96	3.00	3.00	4.00	2.00	4.00	1.00

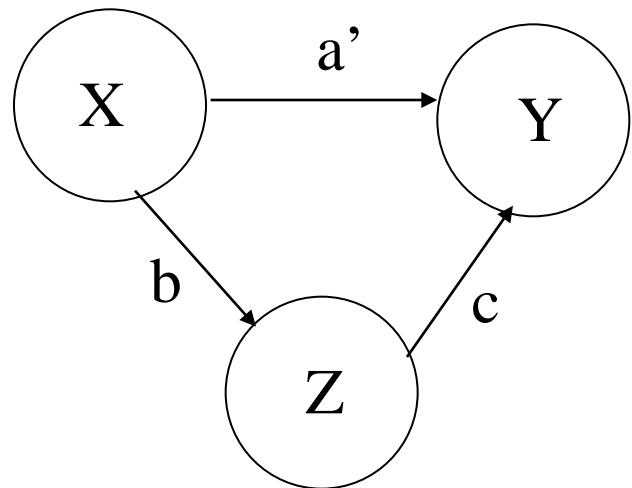
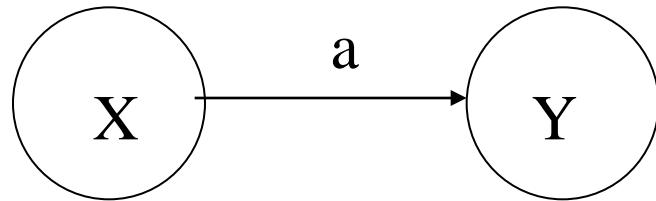
X₁ = Physico-chemical, X₂ = Sensorial, X₃ = Hedonic

A structural regression model



Is Sensory a full mediator between Physico-chemical and Hedonic ?

What is a full mediator ?

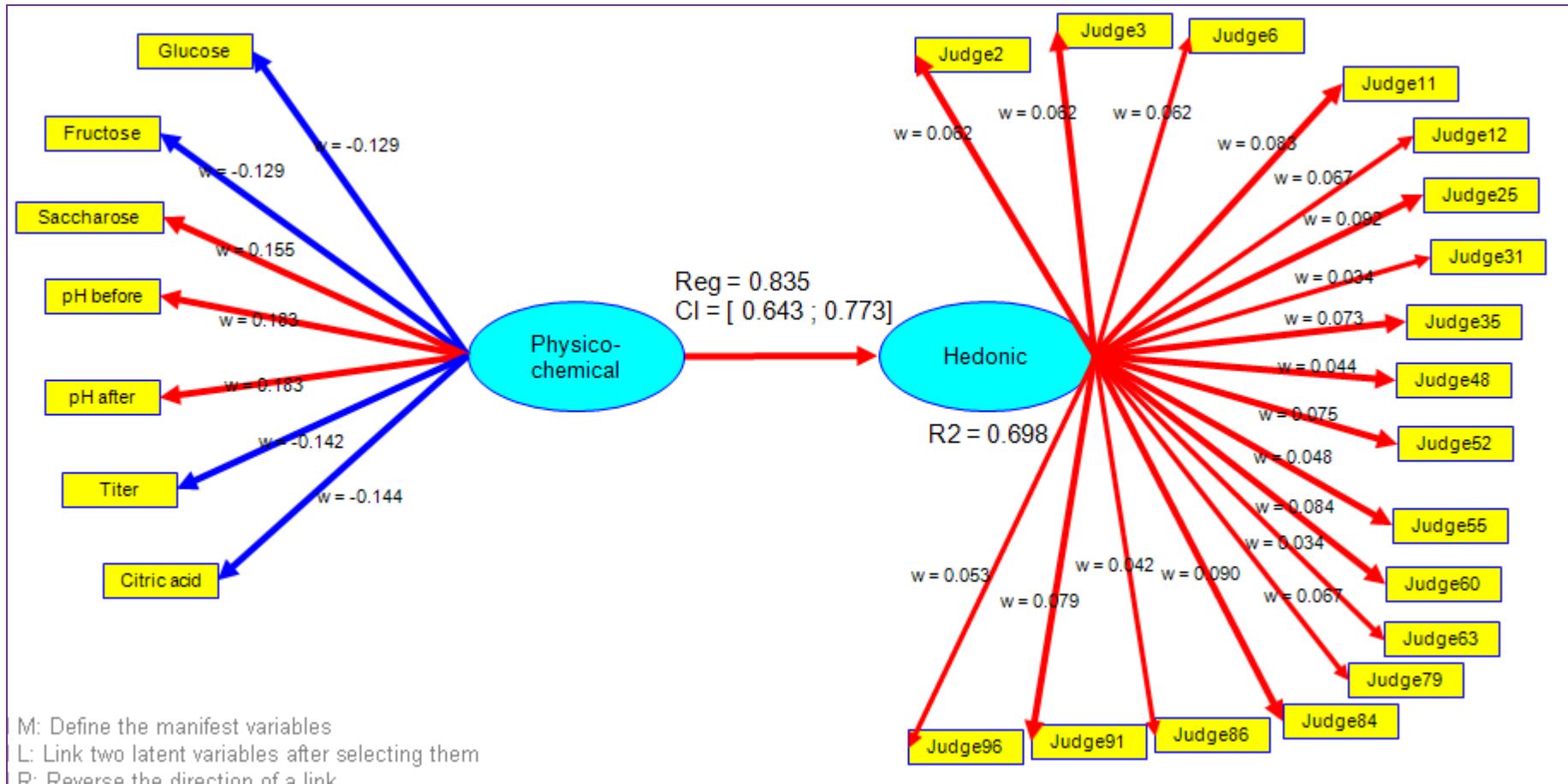


Z is a full mediator between X and Y

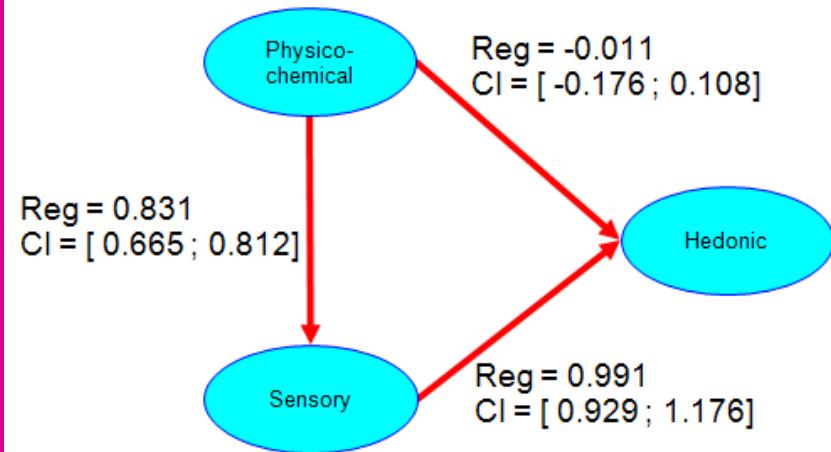
if the following three conditions are satisfied:

- 1) a is significant
- 2) a' is not significant
- 3) Indirect effect $b \times c$ is significant.

Conditions 1 and 2 are satisfied



Indirect effect of Physico-chemical on Hedonic



Is the indirect effect
 $0.831 \times 0.991 = 0.823$
significant ?

Indirect effect:

	Physico-chemical
Hedonic	0.823

Indirect effect /
Standard deviation (Jackknife):

	Physico-chemical
Hedonic	0.049

Indirect effect /
Lower bound (90%) :

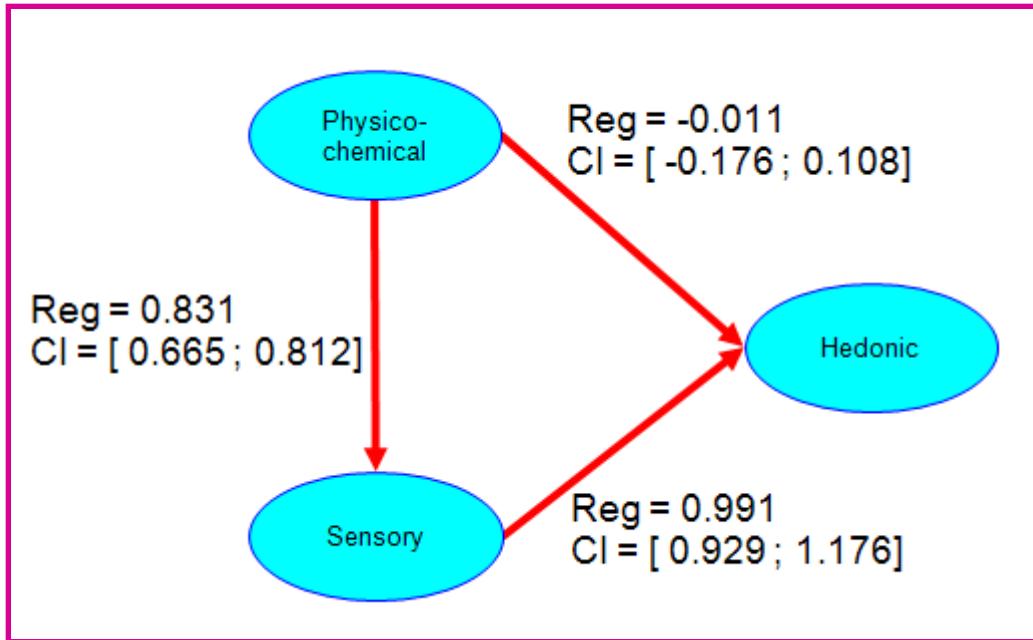
	Physico-chemical
Hedonic	0.731

Indirect effect /
Upper bound (90%):

	Physico-chemical
Hedonic	0.930

Condition 3
is satisfied.

Conclusion



Sensory is a full mediator between Physico-chemical and Hedonic.

Structural Equation Modeling for small samples

- Study of a system of linear relationships between latent variables (unobservable variables).
- Each latent variable is *described* by a set of manifest variables (observable variables).
- MV's can be numerical, ordinal or nominal (no need for normality assumptions).
- The number of observations can be small compare to the number of variables.

Orange juice example

Measurement model for MV's

Glucose
Fructose
Saccharose
Sweetening power
pH before processing
pH after centrifugation
Titer
Citric acid
Vitamin C

Physico-chemical

Structural regression Model for LV's

Exogenous
latent variable

Smell intensity
Odor typicality
Pulp
Taste intensity
Acidity
Bitterness
Sweetness

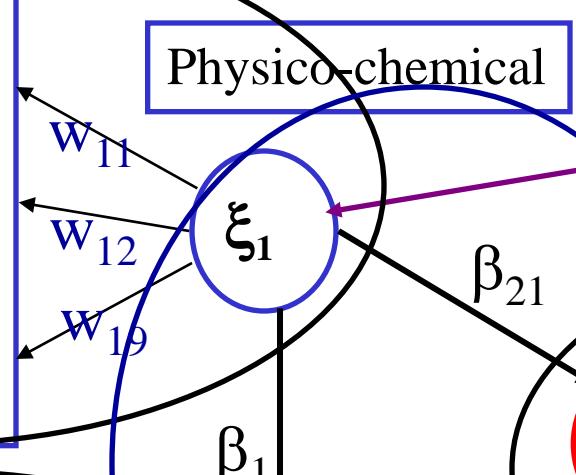
Sensorial

Judge 2
Judge 3
⋮
Judge 96

Endogenous
latent variables

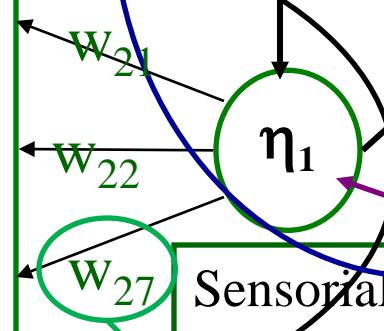
Manifest variables

MV weight or loading



β_{21}

β_1



β_{22}

β_2

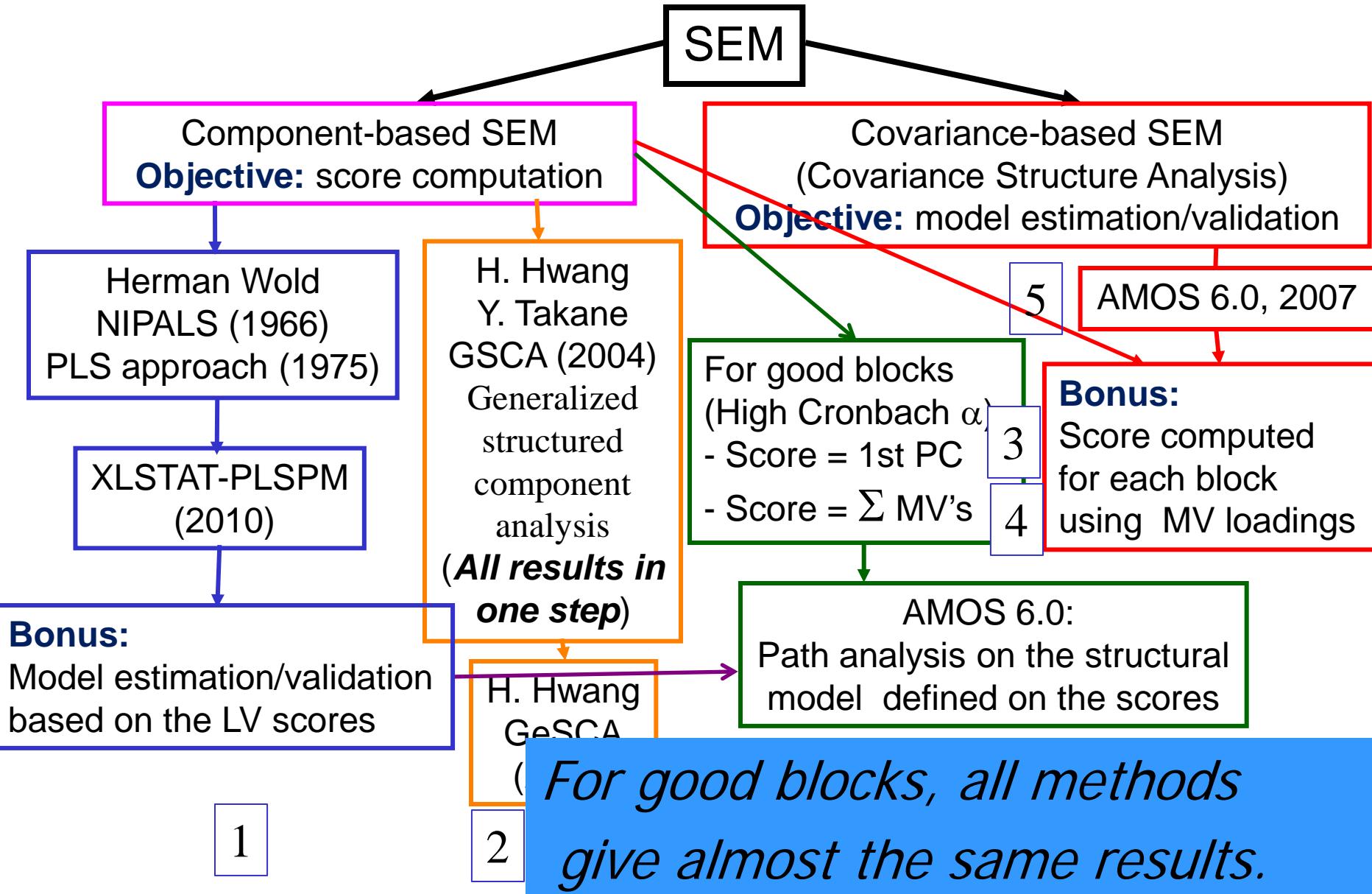


w_{32}
 w_{33}
⋮
 w_{396}

w_{21}
 w_{22}
 w_{27}

Exogenous group of judges

A SEM tree

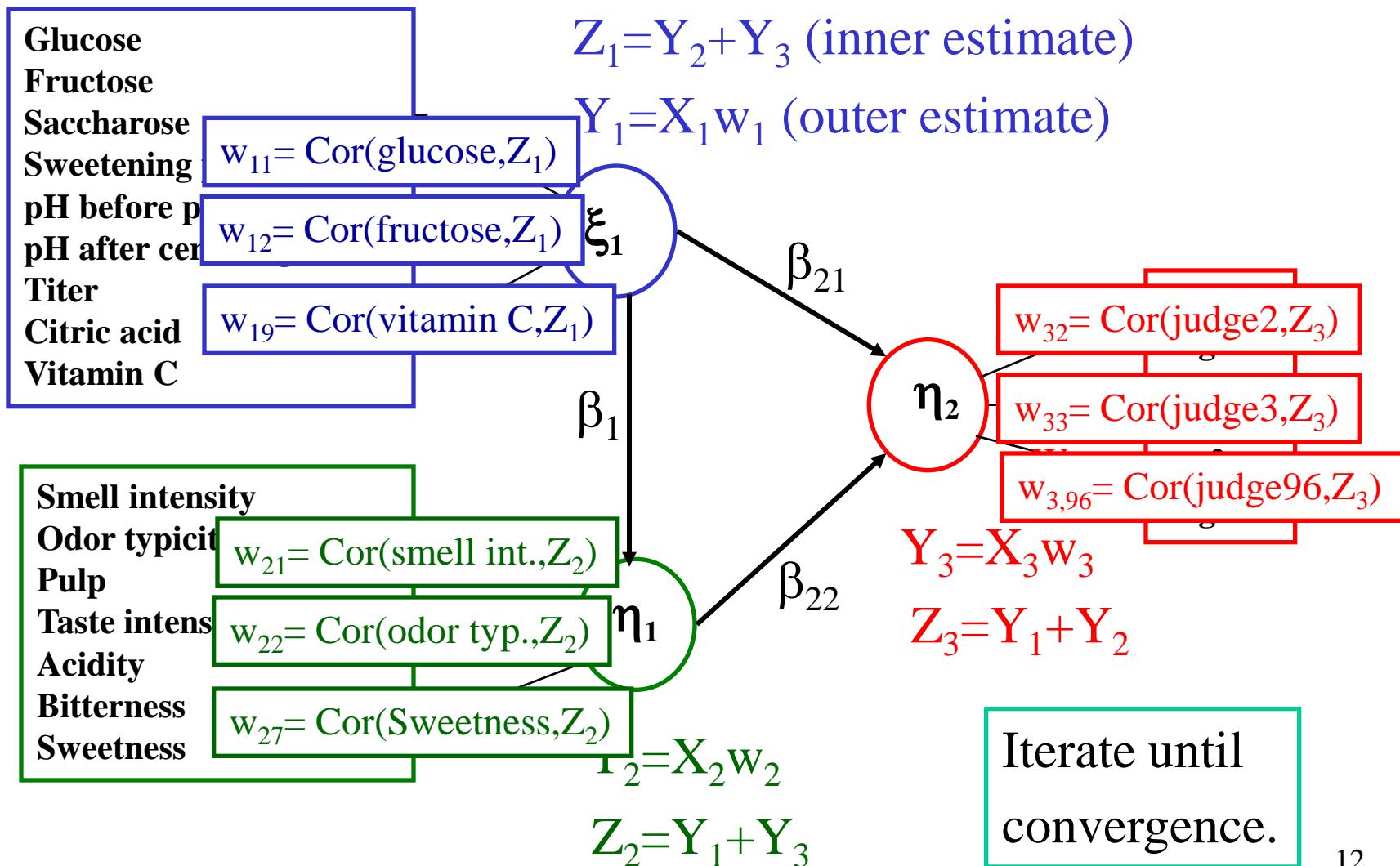


Results

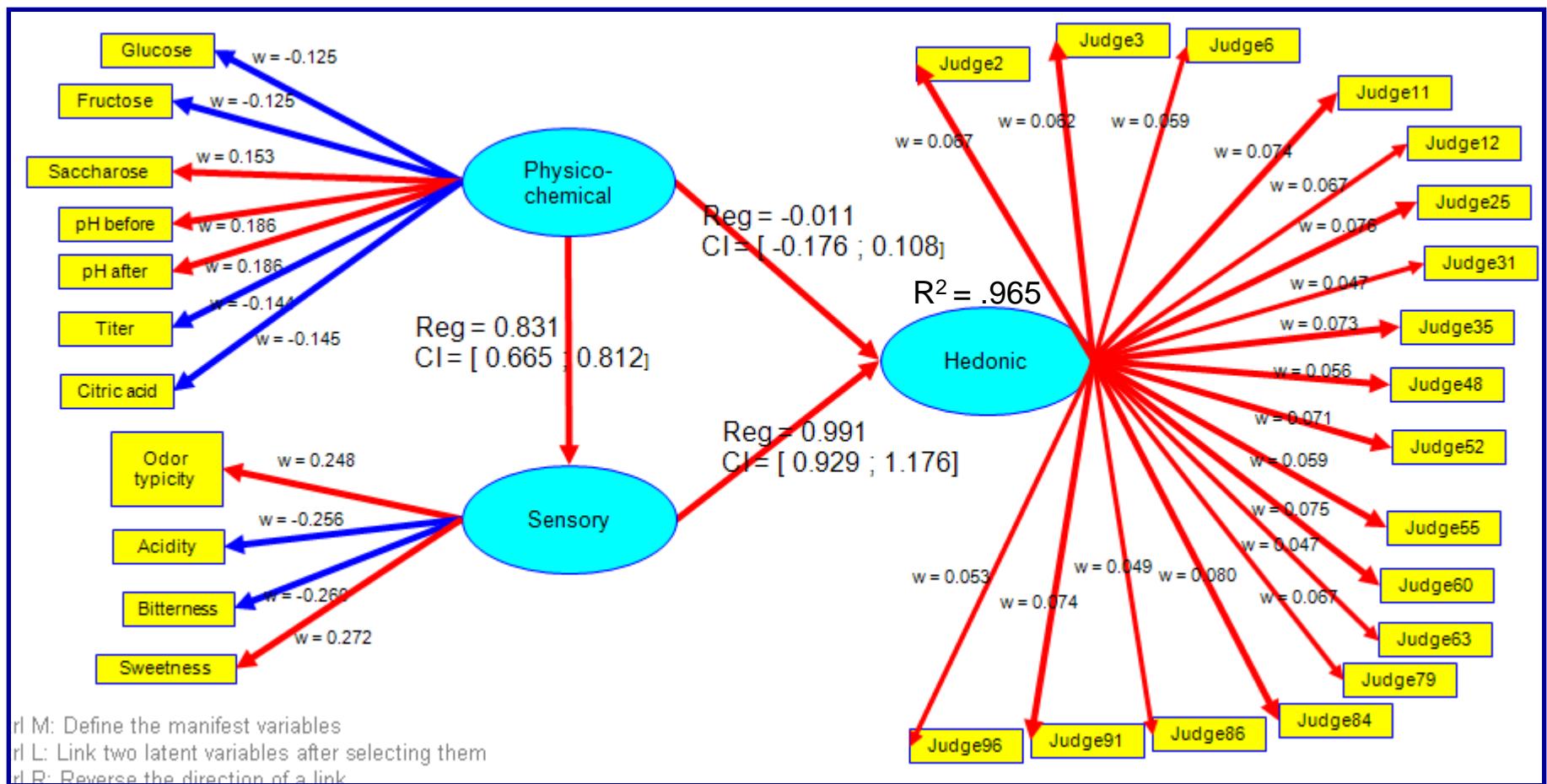
- When all blocks are good, all the methods give practically the same results:
M. Tenenhaus : Component-based SEM
Total Quality Management, 2008
- For all data, PLS and covariance-based SEM yield highly correlated LV scores:
M. Tenenhaus : SEM for small samples
HEC Working paper, 2008.

Data structures are stronger than statistical methods.

I. PLS algorithm (Mode A, Centroid scheme)



Use of XLSTAT-PLSPM (Mode A, centroid scheme)



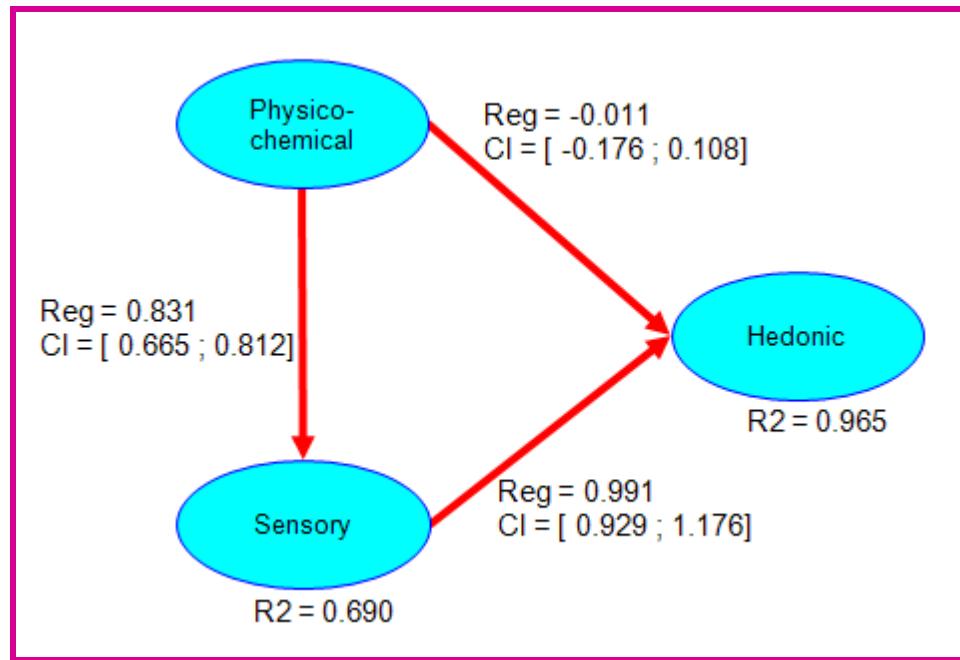
Latent variable scores

	Physico-chemical	Sensory	Hedonic
fruvita refr.	0.842	1.072	1.282
tropicana refr.	0.62	1.175	0.982
tropicana r.t.	1.214	0.645	0.662
<hr/>			
joker r.t.	-1.662	-0.794	-0.738
pampryl refr.	-0.229	-0.715	-1.009
pampryl r.t.	-0.785	-1.383	-1.18

Correlations (Latent variable):

	Physico-chemical	Sensory	Hedonic
Physico-chemical	1.000	0.830	0.824
Sensory	0.830	1.000	0.980
Hedonic	0.824	0.980	1.000

Structural regression model



The structural regressions are computed
by OLS regressions on the latent variables (!).
Why not path analysis ?

SPECIAL CASES OF PLS PATH MODELLING

- Principal component analysis
- Multiple factor analysis (J. Pagès)
- Canonical correlation analysis
- Redundancy analysis
- PLS regression
- Generalized canonical correlation analysis (Horst)
- Generalized canonical correlation analysis(Carroll)
- Multiple co-inertia analysis (Chessel & Hanafi)
- and so on

II. Covariance-based SEM

Latent variables :

$$\eta = \underbrace{\begin{bmatrix} \eta_1 = Sensorial \\ \eta_2 = Hedonic \end{bmatrix}}_{\text{Endogenous LV}}$$

$$\xi = \underbrace{\begin{bmatrix} \xi_1 = Physico-chemical \end{bmatrix}}_{\text{Exogenous LV}}$$

Structural regression model (inner model) :

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 \\ \beta_{22} & 0 \end{bmatrix}}_{\mathbf{B}} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \underbrace{\begin{bmatrix} \beta_1 \\ \beta_{21} \end{bmatrix}}_{\boldsymbol{\Gamma}} \underbrace{\begin{bmatrix} \xi_1 \\ \xi \end{bmatrix}}_{\boldsymbol{\xi}} + \underbrace{\begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}}_{\boldsymbol{\zeta}}$$



$$\eta = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\xi} + \boldsymbol{\zeta}, \quad \Phi = Cov(\boldsymbol{\xi}), \Psi = Cov_{\boldsymbol{\lambda}}(\boldsymbol{\zeta})$$

Structural Equation Modeling

Measurement model (outer model) :

$$\begin{bmatrix} \text{Pulp} \\ \vdots \\ \text{Sweetness} \\ \text{Judge 1} \\ \vdots \\ \text{Judge 4} \end{bmatrix} = \begin{bmatrix} \lambda_{21} & 0 \\ \vdots & \vdots \\ \lambda_{24} & 0 \\ 0 & \lambda_{31} \\ \vdots & \vdots \\ 0 & \lambda_{34} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{21} \\ \vdots \\ \varepsilon_{24} \\ \varepsilon_{31} \\ \vdots \\ \varepsilon_{34} \end{bmatrix}$$

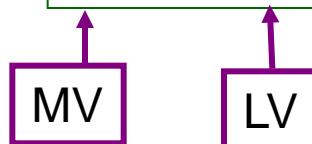
$$\begin{bmatrix} \text{Glucose} \\ \vdots \\ \text{Citric acid} \end{bmatrix} = \begin{bmatrix} \lambda_{11} \\ \vdots \\ \lambda_{14} \end{bmatrix} \begin{bmatrix} \xi \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{14} \end{bmatrix}$$

$$\mathbf{y} = \boldsymbol{\eta} \Lambda_y + \boldsymbol{\varepsilon}, \quad \Theta_{\varepsilon} = Cov(\boldsymbol{\varepsilon})$$



Endogenous

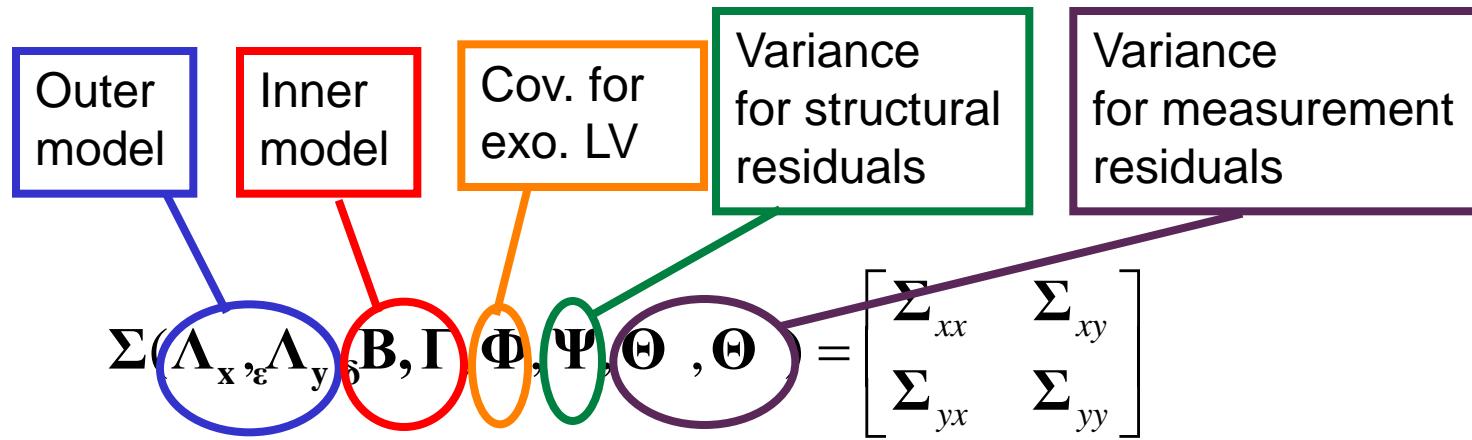
$$\mathbf{x} = \boldsymbol{\xi} \Lambda_x \boldsymbol{\delta}, \quad \Theta_{\delta} = Cov(\boldsymbol{\delta})$$



Exogenous

Structural Equation Modeling

MV covariance matrix :



$$= \begin{bmatrix} \Lambda_x \Phi \Lambda_x^T + \Theta & \Lambda_x \Phi \Gamma^T [(\mathbf{I}_y - \mathbf{B})^T]^{-1} \Lambda^T \\ \Lambda_y [(\mathbf{I} - \mathbf{B})]^{-1} \Gamma \Phi \Lambda_x^T & \Lambda_y [(\mathbf{I} - \mathbf{B})^{-1} (\Gamma \Phi \Gamma^T + \Psi)] [(\mathbf{I} - \mathbf{B})^T]^{-1} \Lambda_y^T + \Theta \end{bmatrix}$$

Covariance-based SEM

ULS algorithm (Unweighted Least Squares) :

S = Observed covariance matrix for MV's

$\Sigma(\Lambda_x, \Lambda_y, B, \Gamma, \Phi, \Psi, \Theta_{\text{ex}}, \Theta_{\text{res}})$ = Implied covariance matrix

$$\underset{\Lambda_x, \Lambda_y, B, \Gamma, \Phi, \Psi, \Theta_{\text{ex}}, \Theta_{\text{res}}}{\text{Minimize}} \quad \| S - \Sigma(\Lambda_x, \Lambda_y, B, \Gamma, \Phi, \Psi, \Theta_{\text{ex}}, \Theta_{\text{res}}) \|^2$$

Generalization of PCA

Goodness-of-fit Index (Jöreskog & Sorbum):

$$GFI = 1 - \frac{\| S - \Sigma(\hat{\Lambda}_x, \hat{\Lambda}_y, \hat{B}, \hat{\Gamma}, \hat{\Phi}, \hat{\Psi}, \hat{\Theta}_{\text{ex}}, \hat{\Theta}_{\text{res}}) \|^2}{\| S \|^2}$$

Should be > 0.9

Use of AMOS 6.0

Method = ULS

Results are similar to PLS.

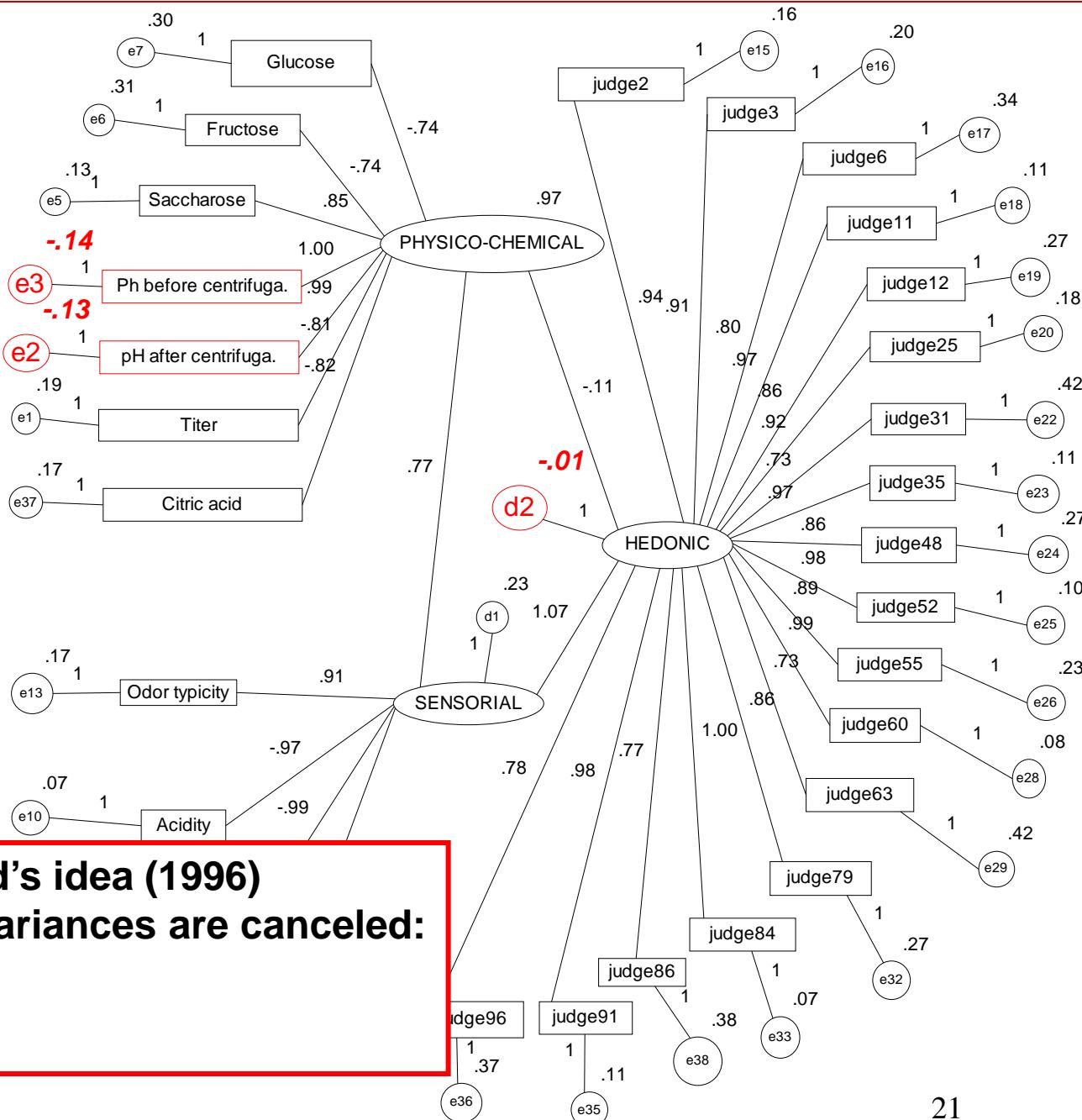
But:

The following variances
are negative:

d2	e2	e3
-.005	-.128	-.139

First Roderick McDonald's idea (1996)
Measurement residual variances are canceled:

$$\hat{\Theta}_\epsilon = \hat{\Theta}_\delta = 0$$



Covariance-based SEM

ULS algorithm with the McDonald's constraints:

S = Observed covariance matrix for MV

$\Sigma(\Lambda_x, \Lambda_y, B, \Gamma, \Phi, \Psi, \Theta_\epsilon = \mathbf{0}, \Theta_\delta = \mathbf{0})$ = Implied cov. matrix

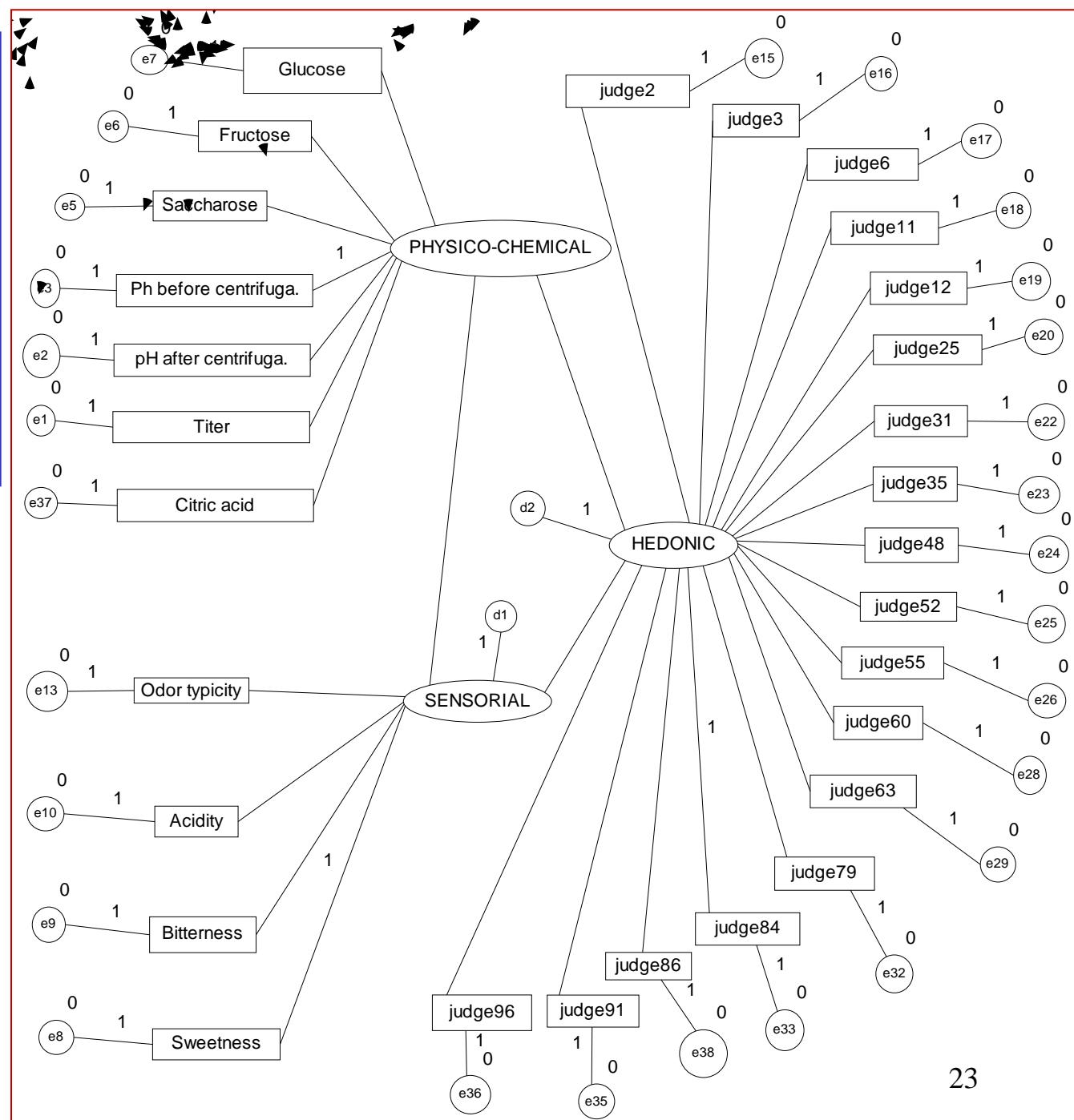
$$\underbrace{\text{Minimize}_{\Lambda_x, \Lambda_y, B, \Gamma, \Phi, \Psi}}_{\text{Outer model}} \left\| S \Sigma \Lambda \underbrace{\Lambda_x \quad B, \Gamma \quad \Phi}_{\text{Inner model}}, \Psi, \Theta_\epsilon = \mathbf{0}, \Theta_\delta = \mathbf{0} \right\|^2$$

Goodness-of-fit Index (Jöreskog & Sorbum):

$$GFI = 1 - \frac{\left\| S \Sigma \Lambda \hat{\Lambda} \hat{B}, \hat{\Gamma}, \hat{\Phi}, \hat{\Psi}, \hat{\Theta}_\epsilon, \hat{\Theta}_\delta \right\|^2}{\| S \|^2}$$

Use of AMOS 6.0

- Method = ULS
- Measurement residual variances = 0



Results

GFI = .965

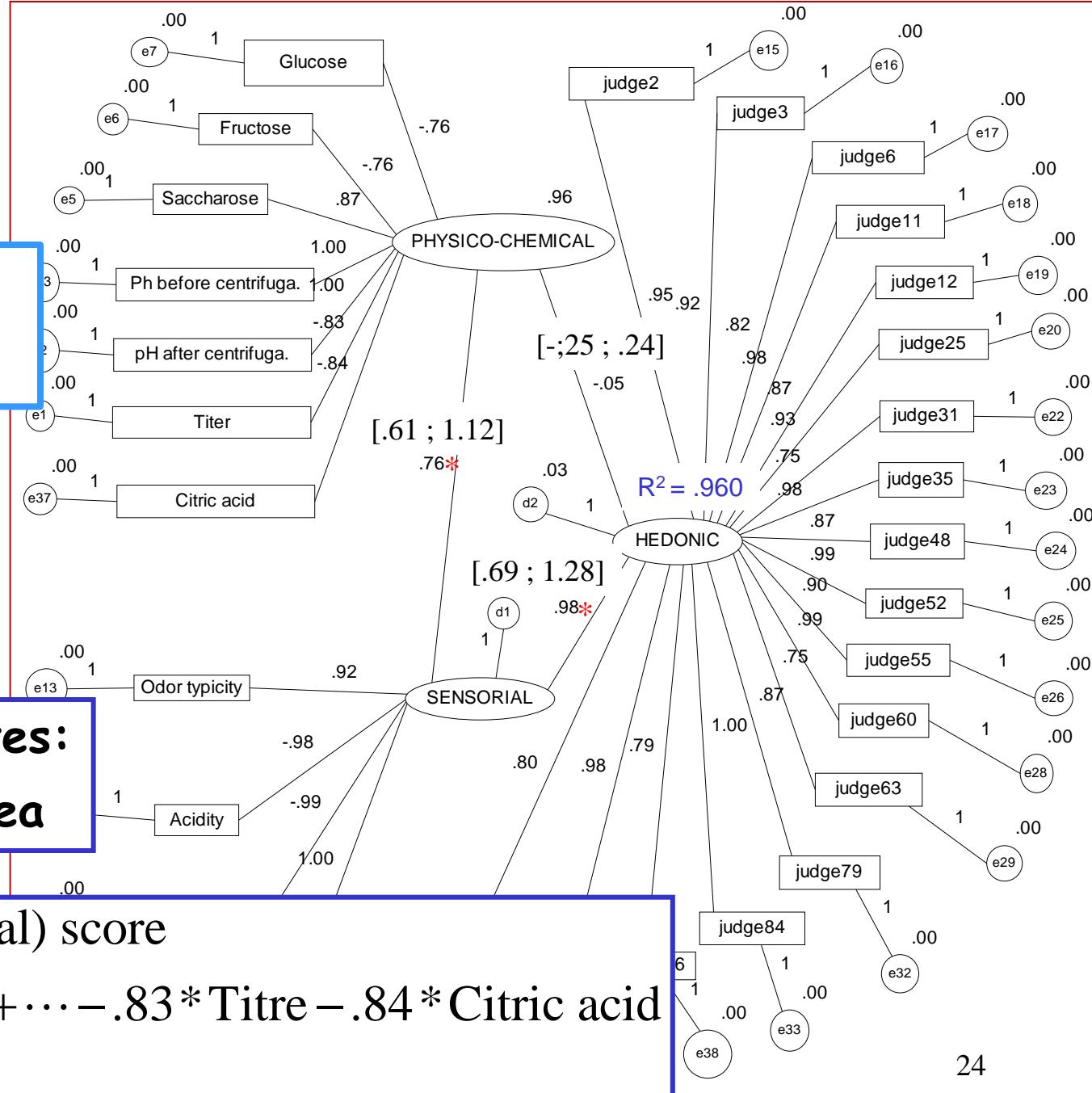
Results are similar to PLS.

90% CI computed by bootstrap.

Outer LV Estimates:
2nd McDonald's idea

LV(Physico-chemical) score

$$\propto -.76^* \text{Glucose} + \dots - .83^* \text{Titre} - .84^* \text{Citric acid}$$



Use of ULS-SEM

Latent variable estimates (Scores)

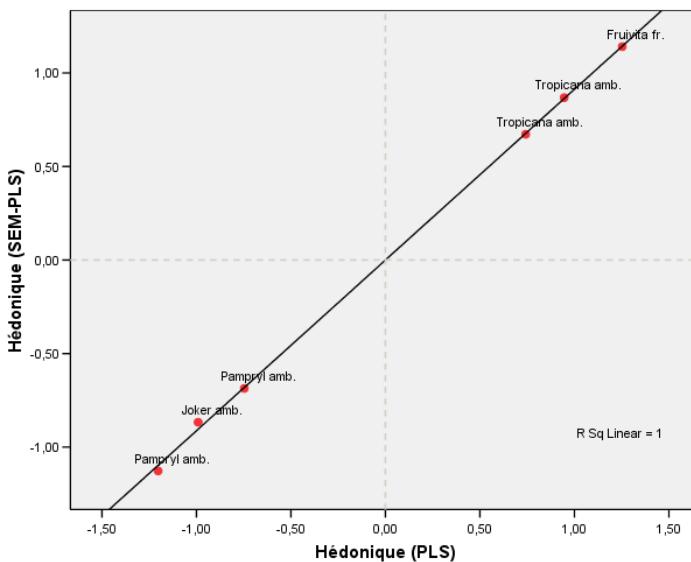
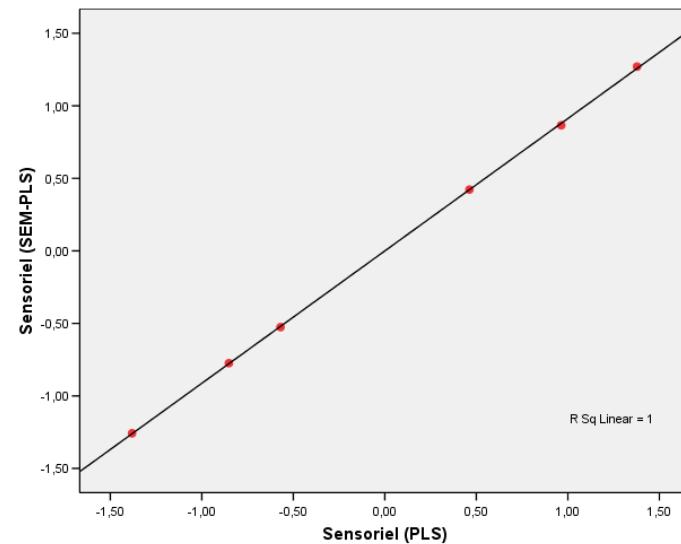
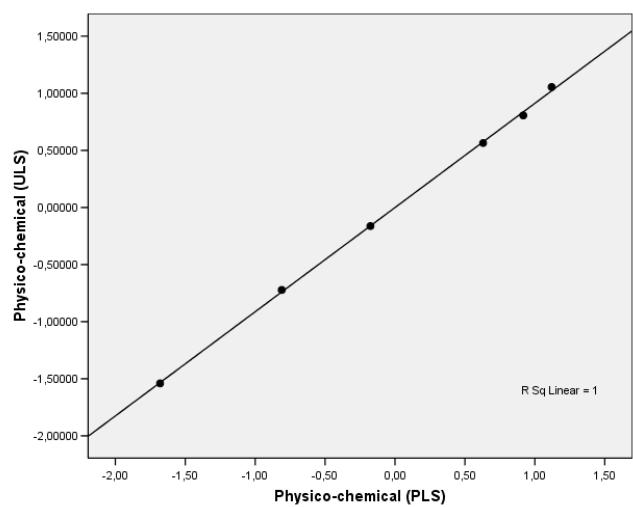
	Physico-chemical	Sensory	Hedonic
Fruivita refr.	0.76	0.98	1.22
Tropicana refr.	0.56	1.07	0.87
Tropicana r.t.	1.11	0.59	0.57

Joker r.t.	-1.53	-0.72	-0.69
Pampryl refr.	-0.21	-0.66	-0.93
Pampryl r.t.	-0.69	-1.26	-1.04

Latent variable score correlation matrix

	Physico-chemical	Sensory	Hedonic
Physico-chemical	1.000	.823	.804
Sensorial	.823	1.000	.978
Hedonic	.804	.978	1.000

Comparison between the PLS and ULS-SEM scores



Conclusion 1: ULS-SEM > PLS

- When mode A is chosen, outer LV estimates using Covariance-based SEM (ULS or ML) or Component based SEM (PLS) are always very close.
- It is possible to mimic PLS with a covariance-based SEM software (McDonald, 1996, Tenenhaus, 2008).
- Covariance-based SEM authorizes to implement non recursive models (feedback loops) and constraints on the model parameters. This is impossible with PLS.

Conclusion 2: PLS > ULS-SEM

- When ULS-SEM does not converge or does not give an admissible solution, PLS is an attractive alternative.
- PLS offers many optimization criterions for the LV search (*Tenenhaus & Tenenhaus, submitted*).
- PLS still works when the number of MV is very high and the number of cases very small (for example 100 MV's and 6 cases).
- PLS allows to use *formative* LV in a much easier way than ULS-SEM.

III. Multi-block data analysis

*Data driven second order
Confirmatory Factor Analysis*

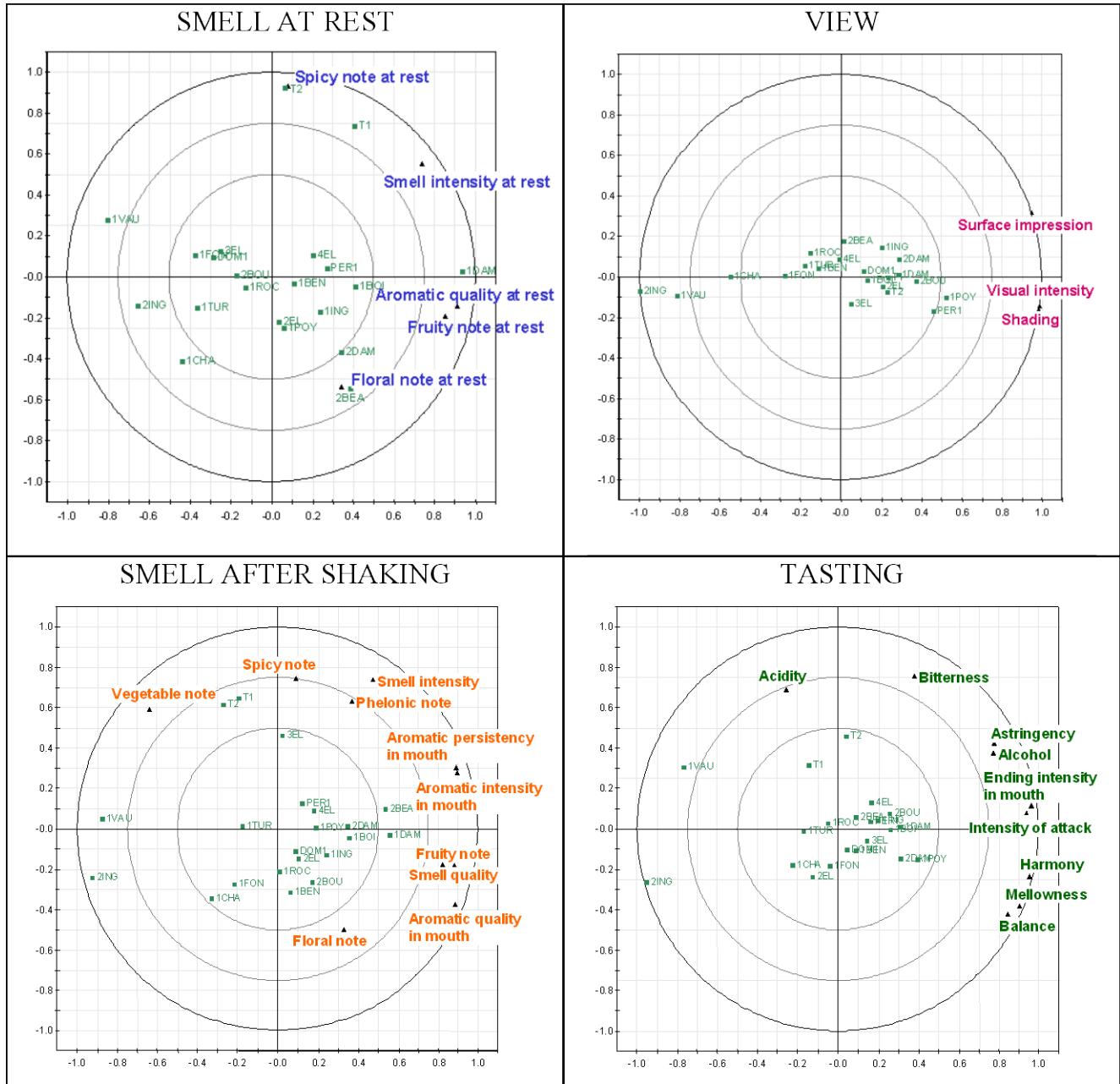
Sen 3 Appellations 4 Soils Red Wines (J. Pagès)

4 blocks of variables

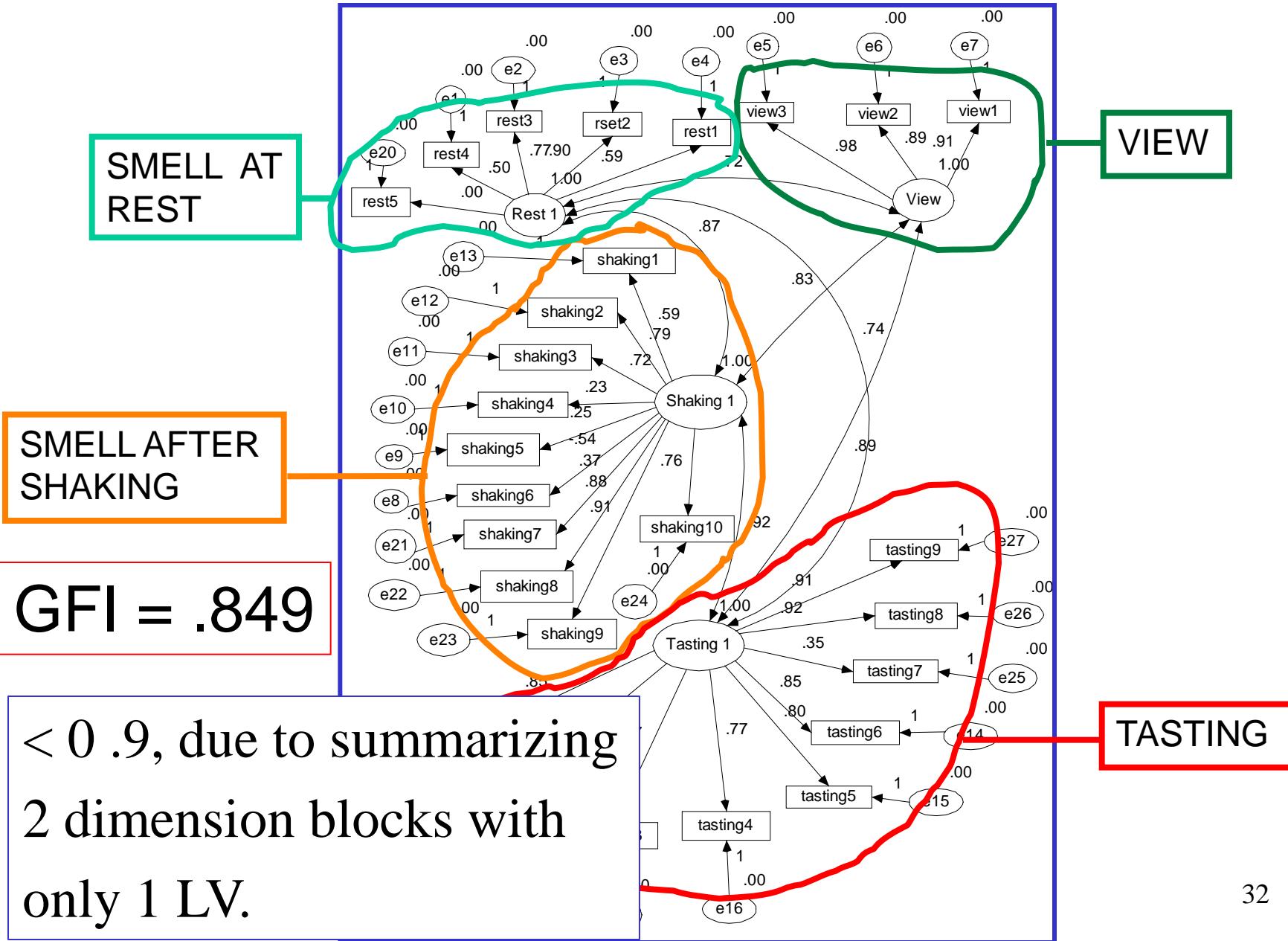
	2el (Saumur),1	1cha (Saumur),1	1fon (Bourgueil),1	1vau (Chinon),3	...	t1 (Saumur),4	t2 (Saumur),4
X₁	Smell intensity at rest Aromatic quality at rest Fruity note at rest Floral note at rest Spicy note at rest	3.07 3.00 2.71 2.28 1.96	2.96 2.82 2.38 2.28 1.68	2.86 2.93 2.56 1.96 2.08	2.81 2.59 2.42 1.91 2.16	...	3.70 3.19 2.83 1.83 2.38
X₂	Visual intensity Shading (orange to purple) Surface impression	4.32 4.00 3.27	3.22 3.00 2.81	3.54 3.39 3.00	2.89 2.79 2.54	...	4.32 4.00 3.33
X₃	Smell intensity after shaking Smell quality after shaking Fruity note after shaking Floral note after shaking Spicy note after shaking Vegetable note after shaking Phenolic note after shaking Aromatic intensity in mouth Aromatic persistence in mouth Aromatic quality in mouth Intensity of attack Acidity Astringency Alcohol Balance (Acid., Astr., Alco.) Mellowness Bitterness Ending intensity in mouth Harmony	3.41 3.31 2.88 2.32 1.84 2.00 1.65 3.26 3.26 3.26 2.96 2.11 2.43 2.50 3.25 2.73 1.93 2.86 2.96	3.37 3.00 2.56 2.44 1.74 2.00 1.38 2.96 2.96 2.96 3.22 2.11 2.18 2.65 2.25 3.32 2.68 2.00 3.07 3.14	3.25 2.93 2.77 2.19 2.25 1.75 1.25 3.08 3.08 3.08 2.70 3.18 2.18 2.18 2.50 2.33 1.68 1.96 2.46 2.04	3.16 2.88 2.39 2.08 2.17 2.30 1.48 2.54 2.54 2.54	3.74 3.08 2.83 1.77 2.44 2.29 1.57 3.44 3.44 3.44 2.96 2.41 2.64 2.96 2.57 2.07 2.22 3.04 2.74 3.00
X₄	Global quality	4.9	3.21	3.54	2.46	...	2.64
	Illustrative variable						2.85

X₁ = Smell at rest, X₂ = View, X₃ = Smell after shaking, X₄ = Tasting

PCA of each block: Correlation loadings



CFA with unit variance LV and zero residual variances



Confirmatory Factor Analysis: Decomposition of the correlation matrix

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & R_{14} \\ R_{21} & R_{22} & R_{23} & R_{24} \\ R_{31} & R_{32} & R_{33} & R_{34} \\ R_{41} & R_{42} & R_{43} & R_{44} \end{bmatrix} \approx \hat{\mathbf{R}} = \begin{bmatrix} \lambda_1 \lambda_1^T & \varphi_{12} \lambda_1 \lambda_2^T & \varphi_{13} \lambda_1 \lambda_3^T & \varphi_{14} \lambda_1 \lambda_3^T \\ \varphi_{21} \lambda_2 \lambda_1^T & \lambda_2 \lambda_2^T & \varphi_{23} \lambda_2 \lambda_3^T & \varphi_{24} \lambda_2 \lambda_4^T \\ \varphi_3 \lambda_{13} \lambda_1^T & \varphi_3 \lambda_{23} \lambda_2^T & \lambda_3 \lambda_3^T & \varphi_3 \lambda_{43} \lambda_4^T \\ \varphi_{41} \lambda_4 \lambda_1^T & \varphi_{42} \lambda_4 \lambda_2^T & \varphi_{43} \lambda_4 \lambda_3^T & \lambda_4 \lambda_4^T \end{bmatrix}$$

Observed Implied

where :

taking into account
the other blocks

Similar to four PCA

$$\lambda_1 = \begin{bmatrix} \lambda_{rest1} \\ \lambda_{rest2} \\ \vdots \\ \lambda_{rest5} \end{bmatrix} = \text{Vector of factor loadings for LV } \xi_1, \text{ idem for } \lambda_2, \lambda_3, \lambda_4$$

$$\varphi_{ij} = Cor(\xi_i, \xi_j)$$

Confirmatory Factor Analysis

ULS algorithm with the McDonald's constraints:

\mathbf{R} = observed correlation matrix for MV's

$\hat{\mathbf{R}}$ = implied correlation matrix

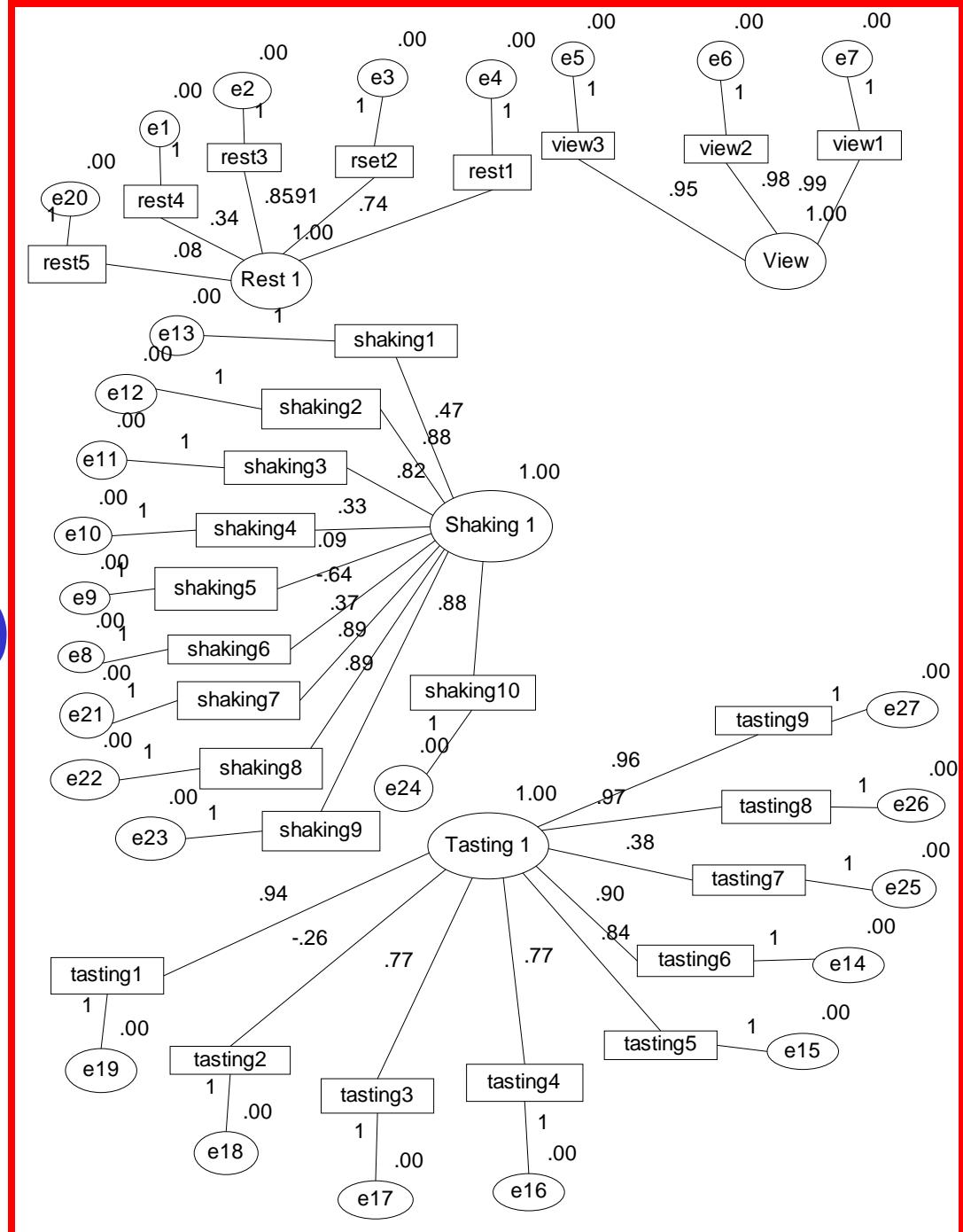
$$\underbrace{\text{Minimize}}_{\lambda_1, \dots, \lambda_4, \varphi_{12}, \dots, \varphi_{34}} \|\mathbf{R} - \hat{\mathbf{R}}\|^2$$

Goodness-of-fit Index (Jöreskog & Sorbum):

$$GFI = 1 - \frac{\|\mathbf{R} - \hat{\mathbf{R}}\|^2}{\|\mathbf{R}\|^2}$$

1st PC's of
each block
are obtained
when LV's are
constrained to
be orthogonal
 $(\varphi_{ij} = 0 \text{ for all } i \neq j)$

GFI = .301

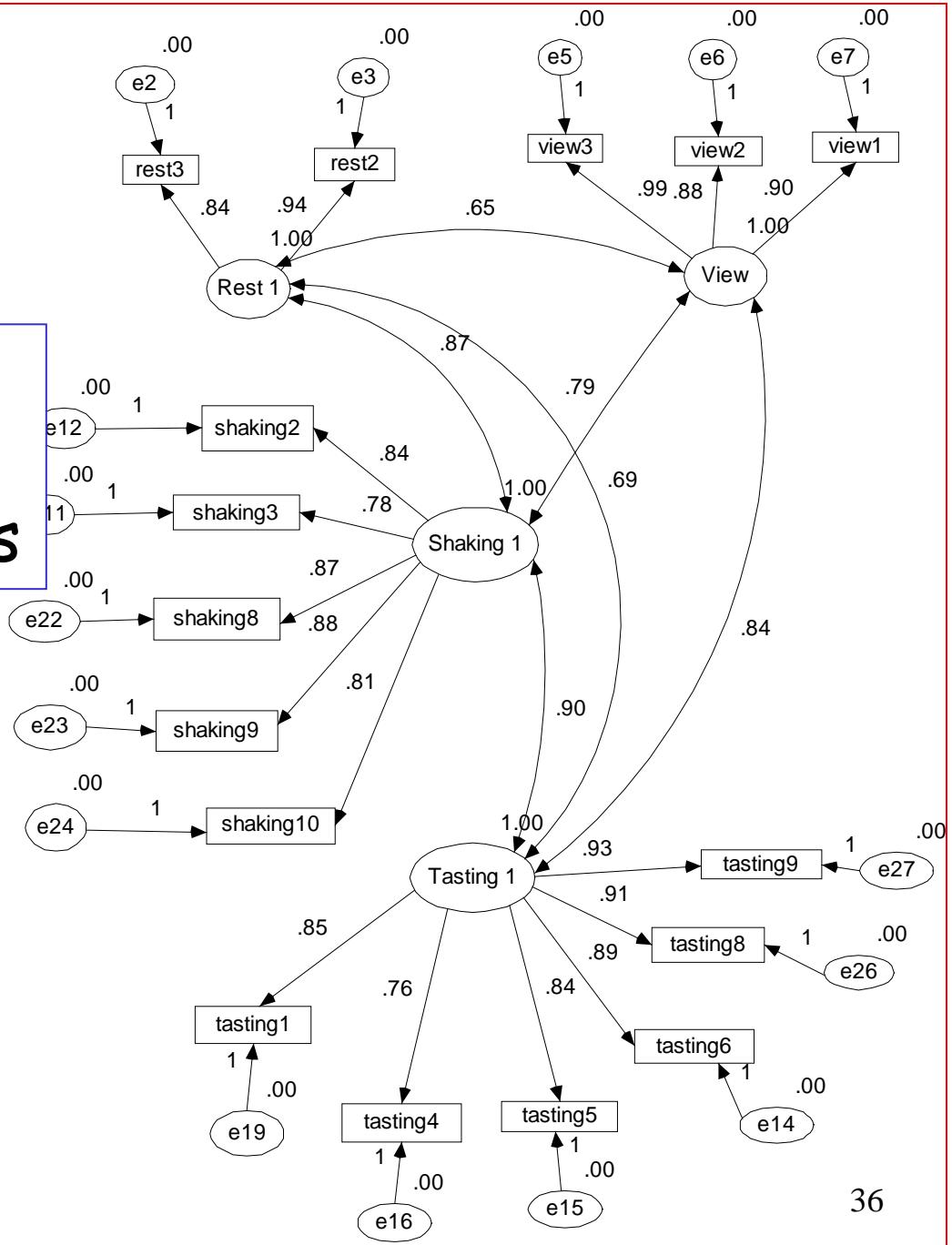


First dimension

Using MV's with significant loadings in the CFA on all MV's

GFI = .983

High correlations between the LV's.



First global score

2nd order CFA

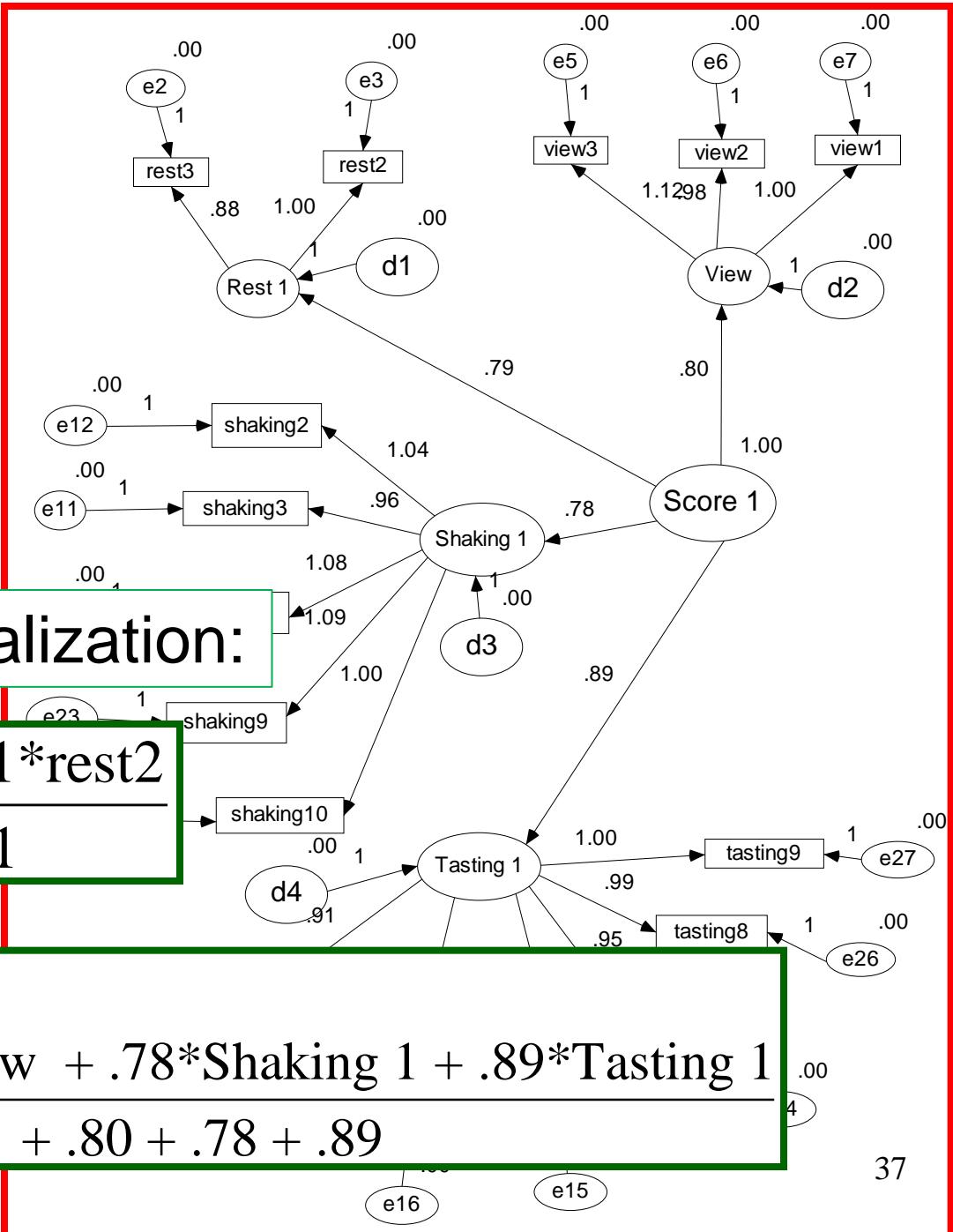
$$GFI = .973$$

Using Fornell normalization:

$$\text{Rest 1} = \frac{.88 * \text{rest3} + 1 * \text{rest2}}{.88 + 1}$$

Score 1

$$= \frac{.79 * \text{Rest 1} + .80 * \text{View} + .78 * \text{Shaking 1} + .89 * \text{Tasting 1}}{.79 + .80 + .78 + .89}$$



Validation of the first dimension

Correlations

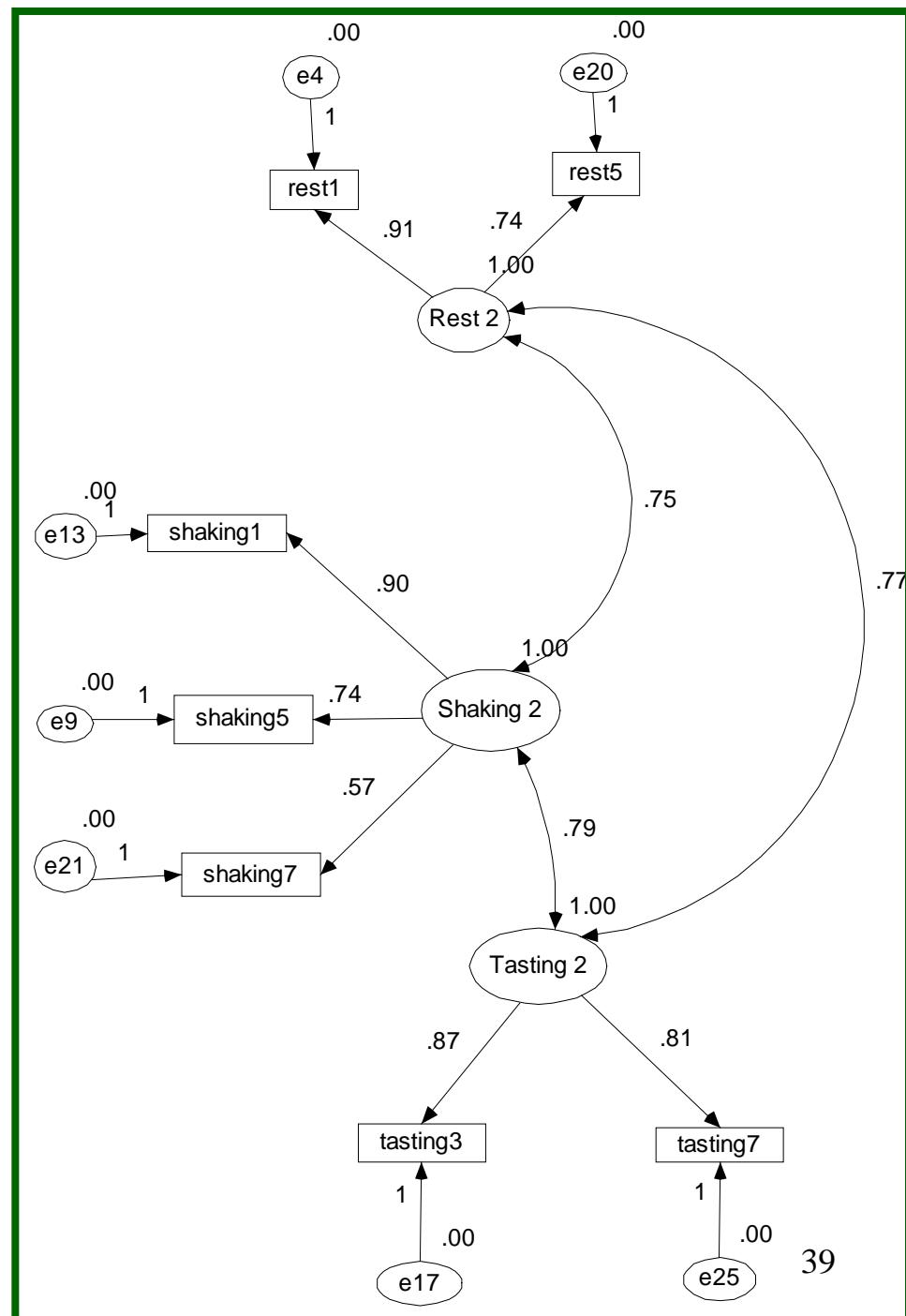
	Rest1	View	Shaking1	Tasting1
Rest1	1			
View	.621	1		
Shaking1	.865	.762	1	
Tasting1	.682	.813	.895	1
Score1	.813	.920	.942	.944

Second dimension

Using MV's with non significant loadings in the CFA on all MV's

GFI = .919

High correlations between the LV's.

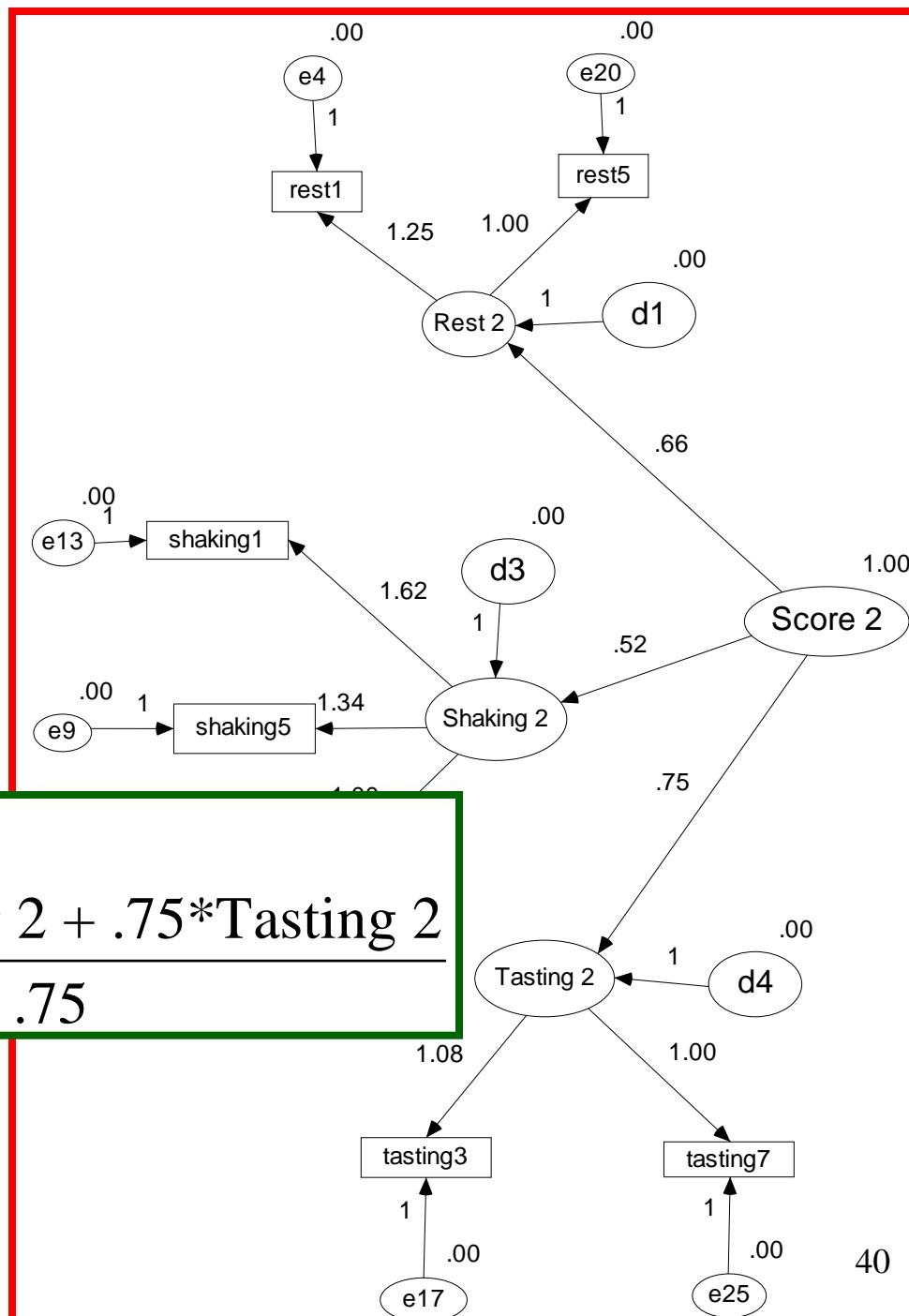


2nd global score

GFI = .905

Score 2

$$= \frac{.66 * \text{Rest 2} + .52 * \text{Shaking 2} + .75 * \text{Tasting 2}}{.66 + .52 + .75}$$



Validation of the second dimension

Correlations

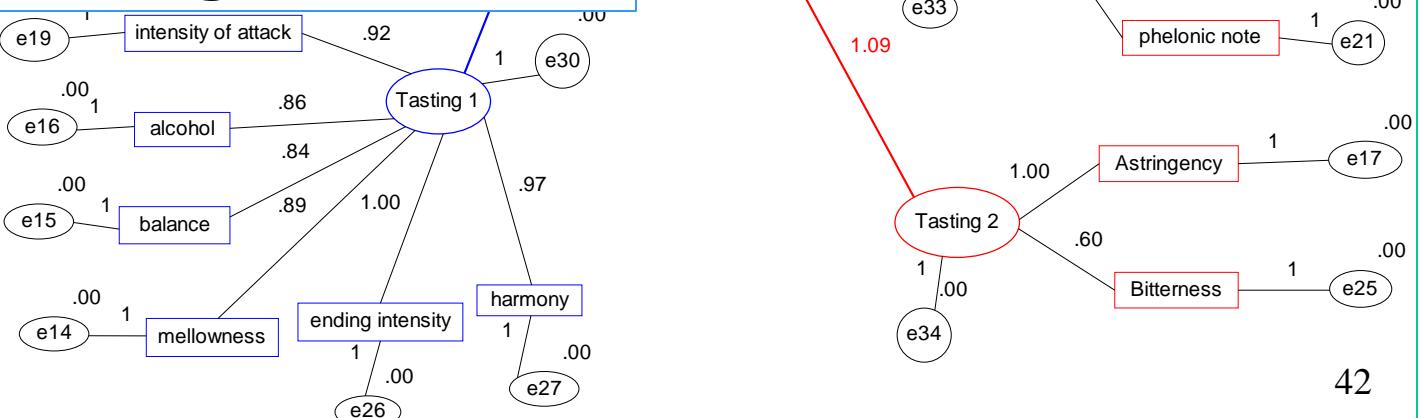
	Rest2	Shaking2	Tasting2
Rest2	1		
Shaking2	.789	1	
Tasting2	.782	.803	1
Score2	.944	.904	.928

A two dimension global model

GFI = .914

But Cor(Score 1, Score 2) = .55 is not significant.

And "spicy note at rest" becomes non significant.



The two previous 2nd order models are recovered.

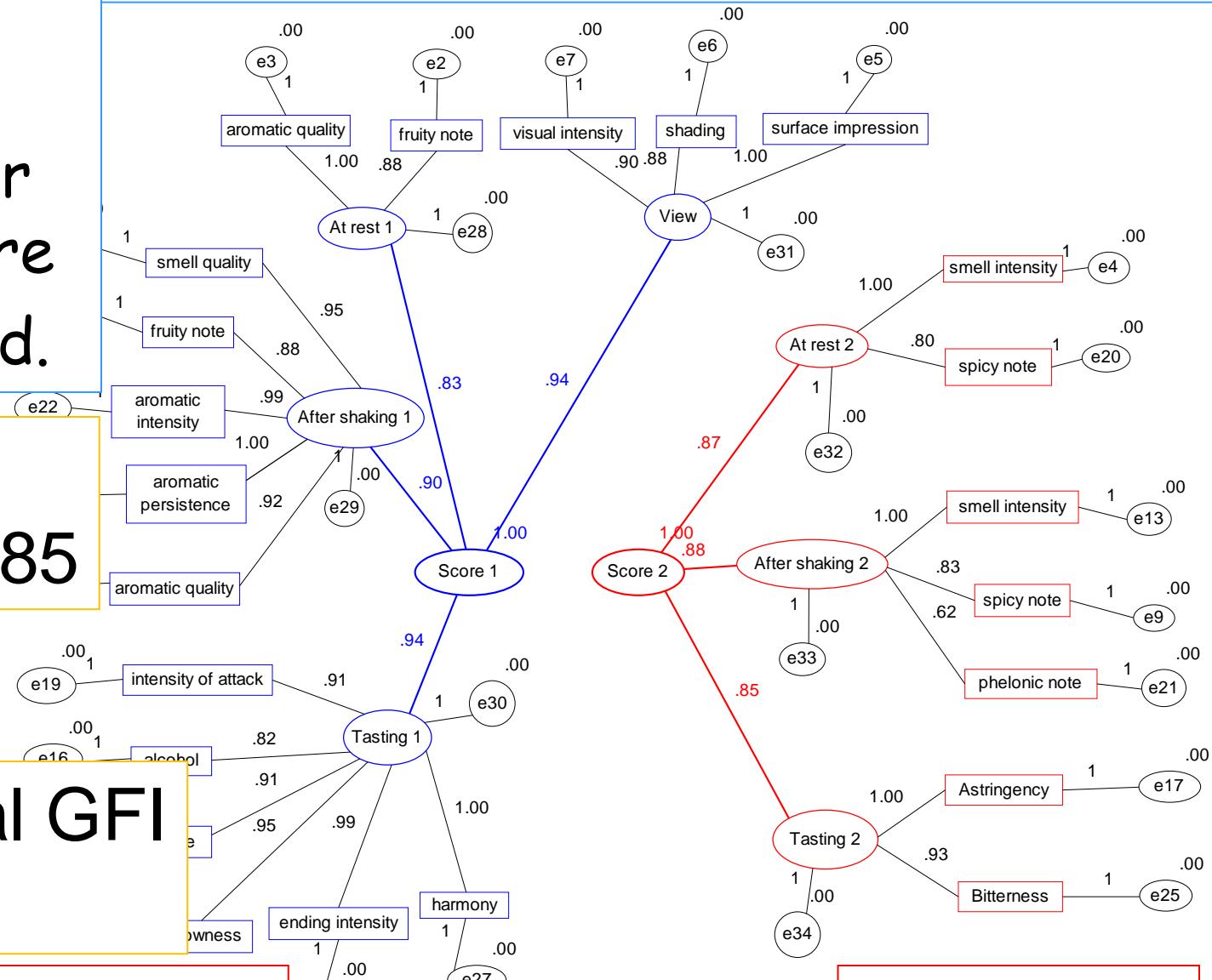
Global GFI = .785

But local GFI are OK:

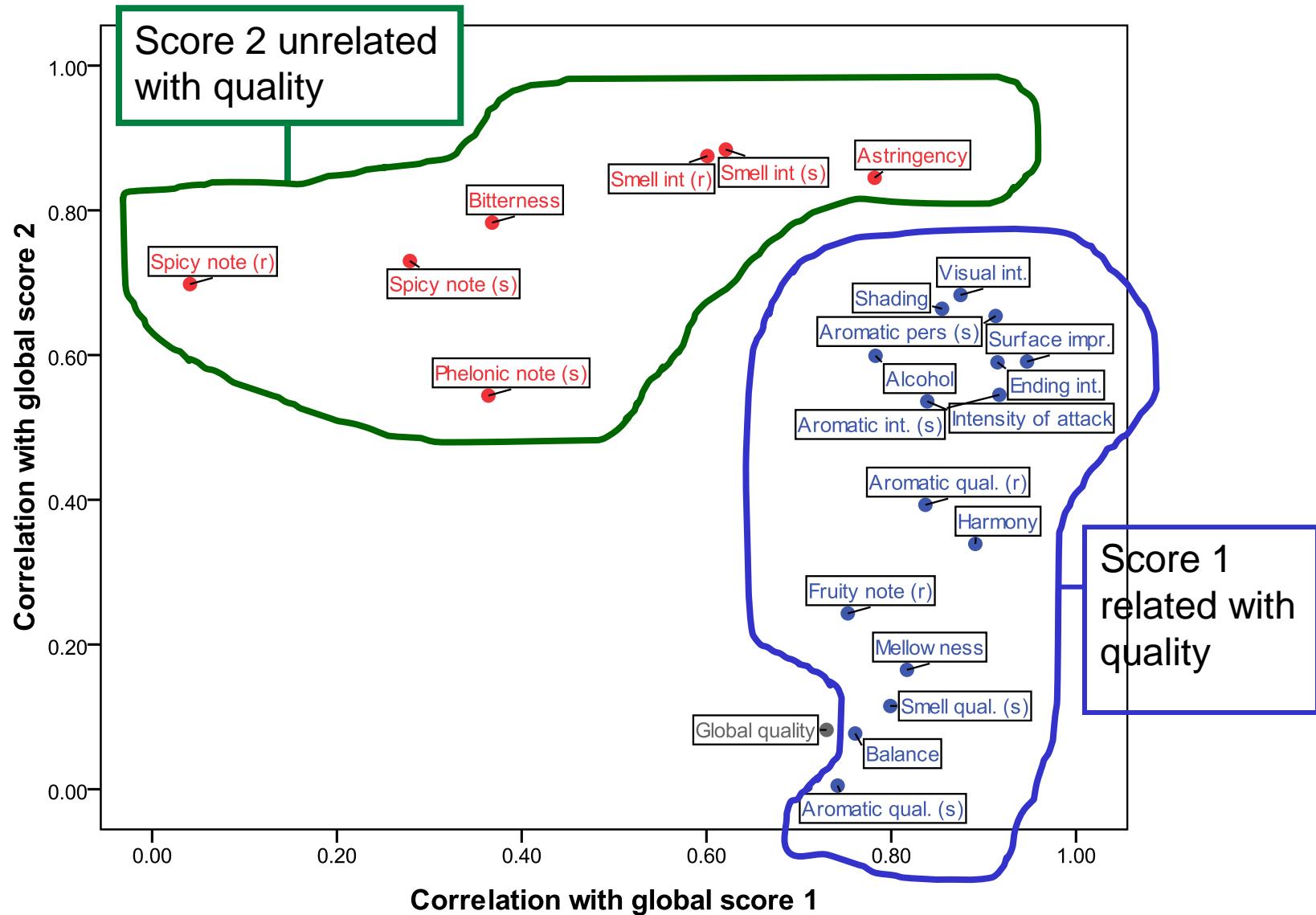
GFI(1) = .973

GFI(2) = .905

Final model



Mapping of the correlations with the global scores



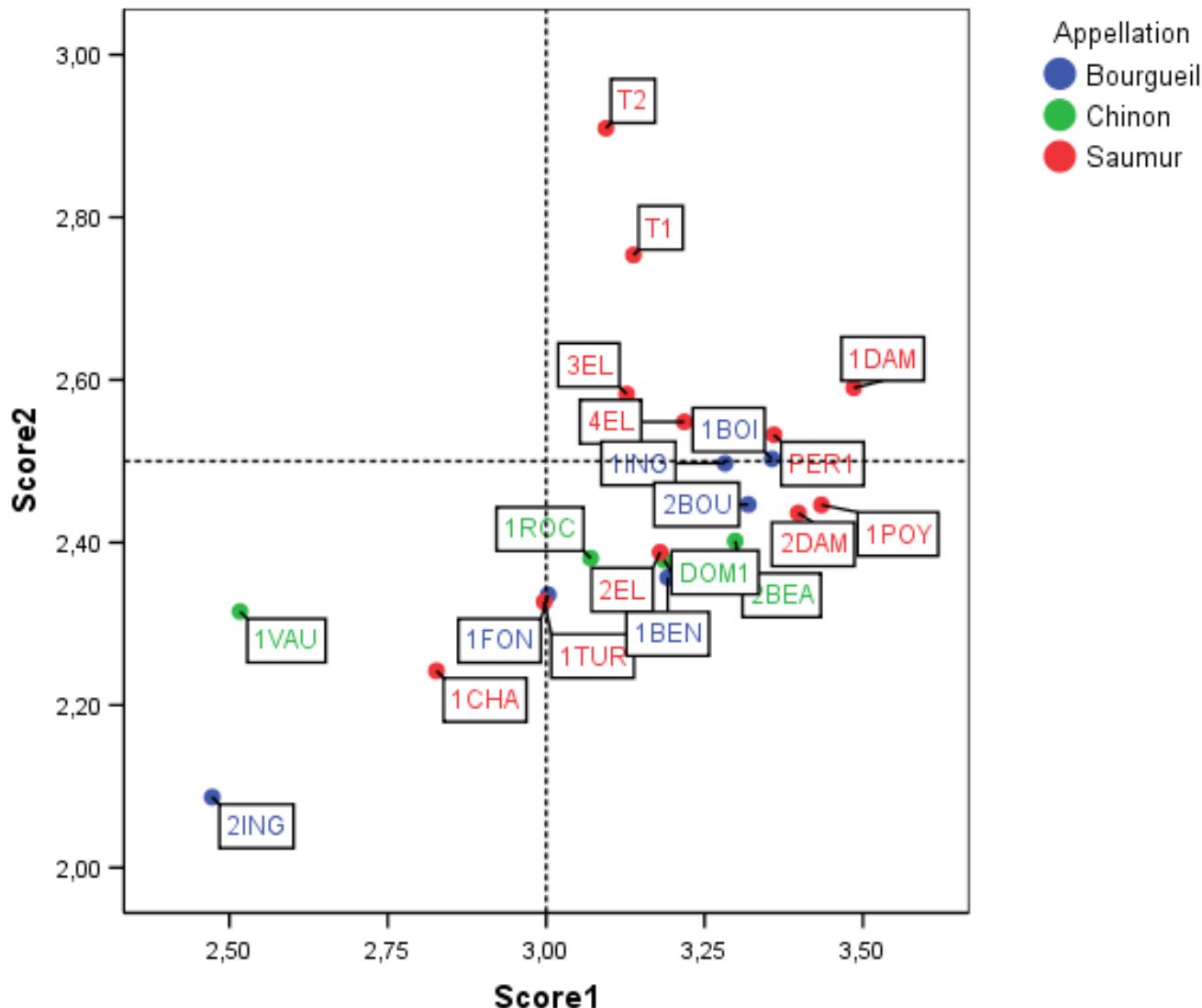
Correlation with global quality

Variables related to dimension 1	Global quality
Aromatic quality at rest	0.62
Fruity note at rest	0.50
Visual intensity	0.54
Shading (from orange to purple)	0.51
Surface impression	0.67
Smell quality	0.76
Aromatic intensity in mouth	0.61
Aromatic persistence in mouth	0.68
Aromatic quality in mouth	0.85
Intensity of attack	0.77
Alcohol	0.52
Balance (acidity, astringency, alcohol)	0.95
Mellowness	0.92
Ending intensity in mouth	0.80
Harmony	0.88

Variables related to dimension 2	Global quality
Smell intensity at rest	0.04
Spicy note at rest	-0.31
Smell intensity after shaking	0.17
Spicy note after shaking	-0.08
Phelonic note	0.09
Astringency	0.41
Bitterness	0.05

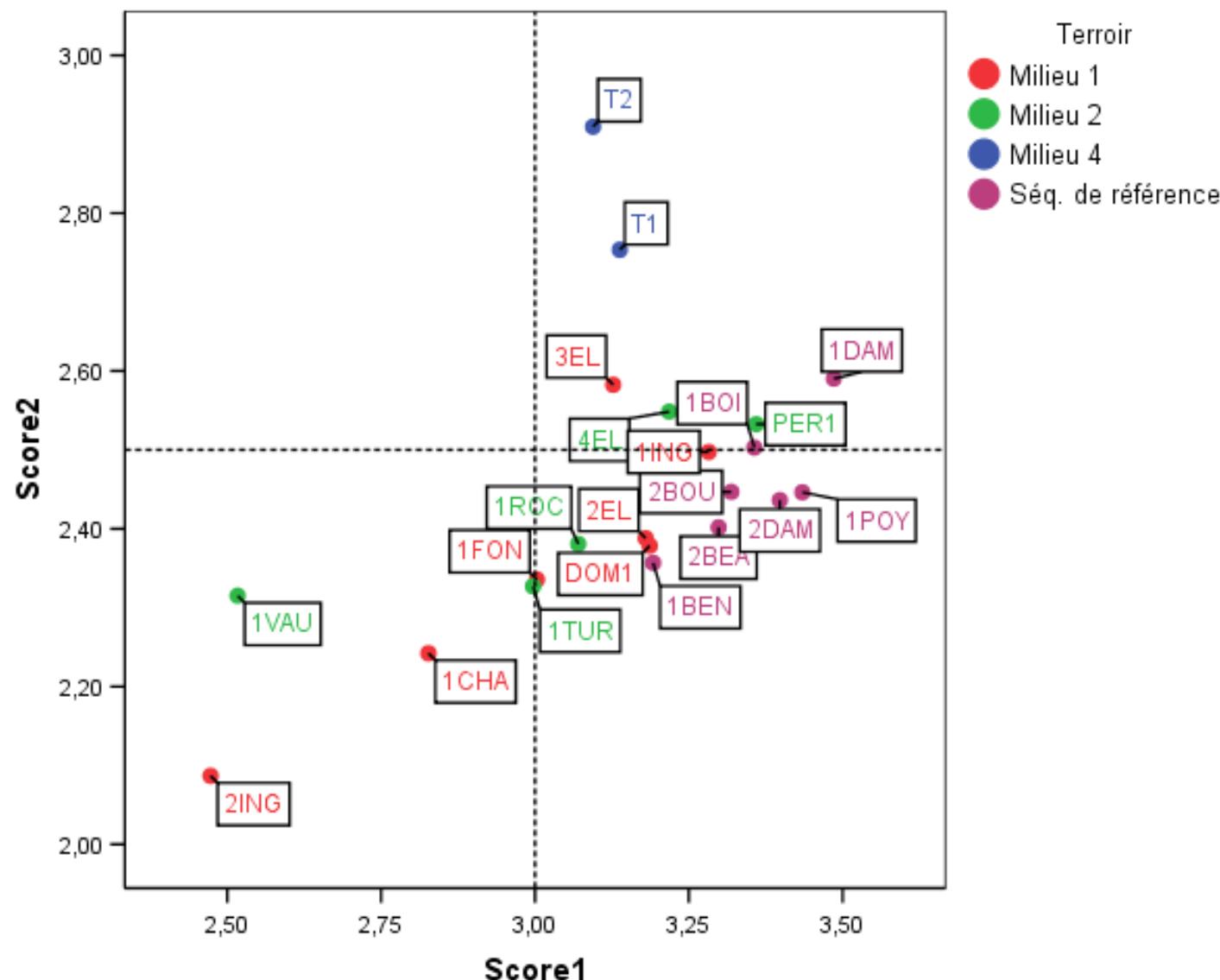
Wine visualization in the global score space

Wines marked by Appellation



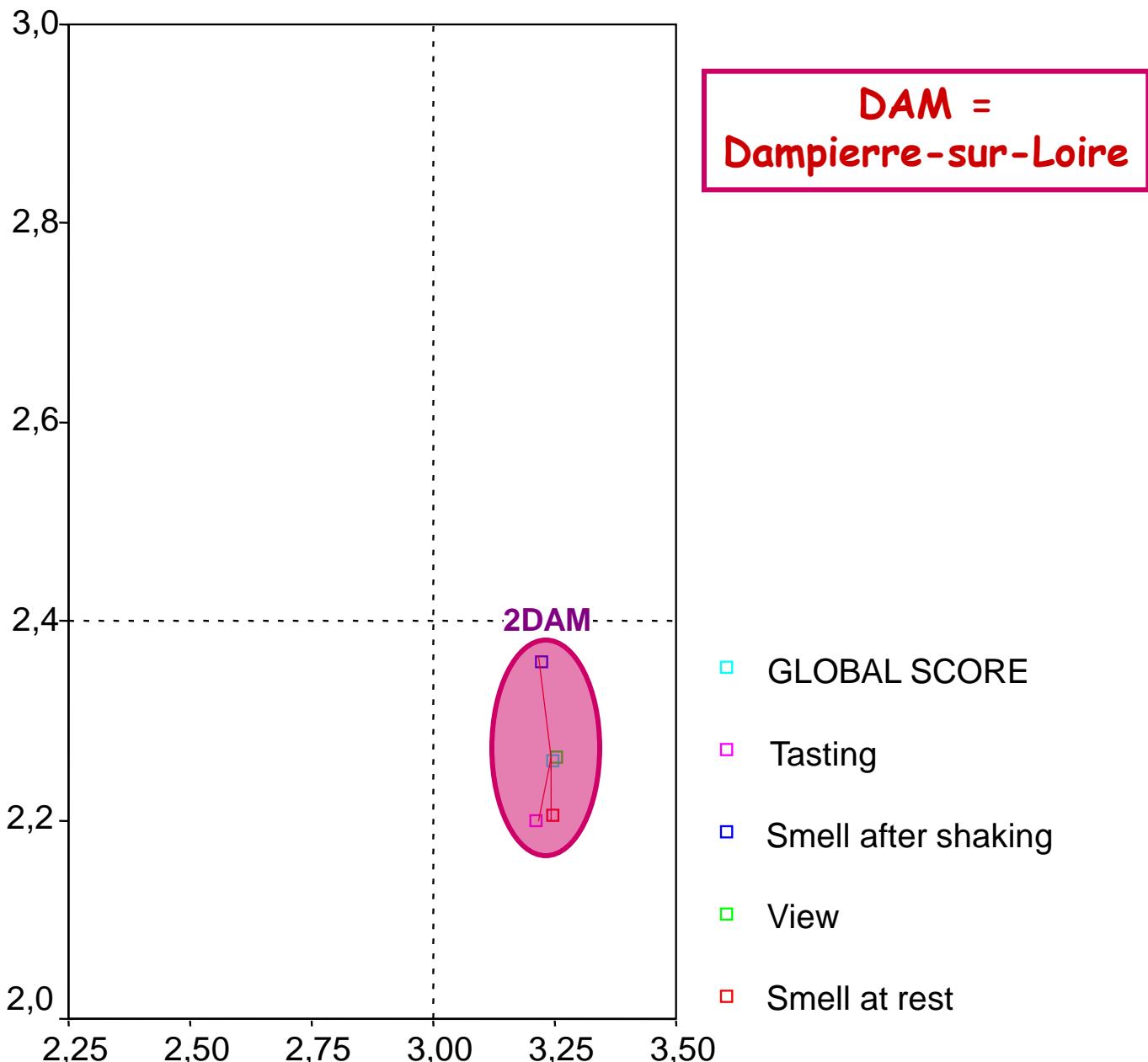
Wine visualization in the global score space

Wines marked by Soil



Visualization of wine variability among the blocks

Star-plot of the “best wine” – 2DAM SAUMUR





Cuvée Lisagathe 1995

A soft, warm, blackberry nose. A good core of fruit on the palate with quite well worked tannin and acidity on the finish; Good length and a lot of potential.

DECANTER (mai 1997)

(DECANTER AWARD ***** : Outstanding quality, a virtually perfect example)

Final conclusion



« All the proofs of a pudding are in the eating, not in the cooking ».

William Camden (1623)