Measuring the Impact of Noncognitive Skills by Structural Equation Models

Rémi Piatek

University of Konstanz

Mai 16, 2008
Inaugural Conference – Noncognitive Skills:
Acquisition and Economic Consequences

Introduction

Study of **Noncognitive skills (NCS)**: keen interest in the social sciences

- Intuition that personality traits influence individual behaviors and choices
- Predictive power of noncognitive skills (Heckman et al. 2006)
- Data availability opens new research horizons

The methodology used to measure NCS in many empirical applications may be questionable...

Measuring Noncognitive skills

NCS not directly observable, only **indicators** are available.

Two mainstream approaches, and (some of) their drawbacks:

Index construction

Combining the indicators to create a synthetic index.

- no theoretical justification
- specific problems:
  - measurement error
  - endogeneity problem
  - reverse causality

Factor analysis

- strong requirements:
  - factor orthogonality
  - distributional assumptions (normality)
  - factor exogeneity
- interpretation can be problematic

Any other possible method?

Factor Structure Models

Heckman and coauthors (2004, 2006): Factor Structure Models

- Cognitive and noncognitive skills captured by latent factors
- Adress the measurement error and the endogeneity problems
- Make it possible to deal with the reverse causality problem
- Possible to investigate the impact of NCS on outcome variables

Structural Equation Modeling (SEM) is in the same vein, and introduces even more flexibility...

Introducing Structural Equation Model

SEM is a well-known and well-documented approach, extensively implemented in empirical research.

N. Cliff (1983):

SEM approach described as “perhaps the most important and influential statistical revolution to have occurred in the social sciences.”

And yet, SEM has been scarcely applied to the study of NCS until now, though its advantages:

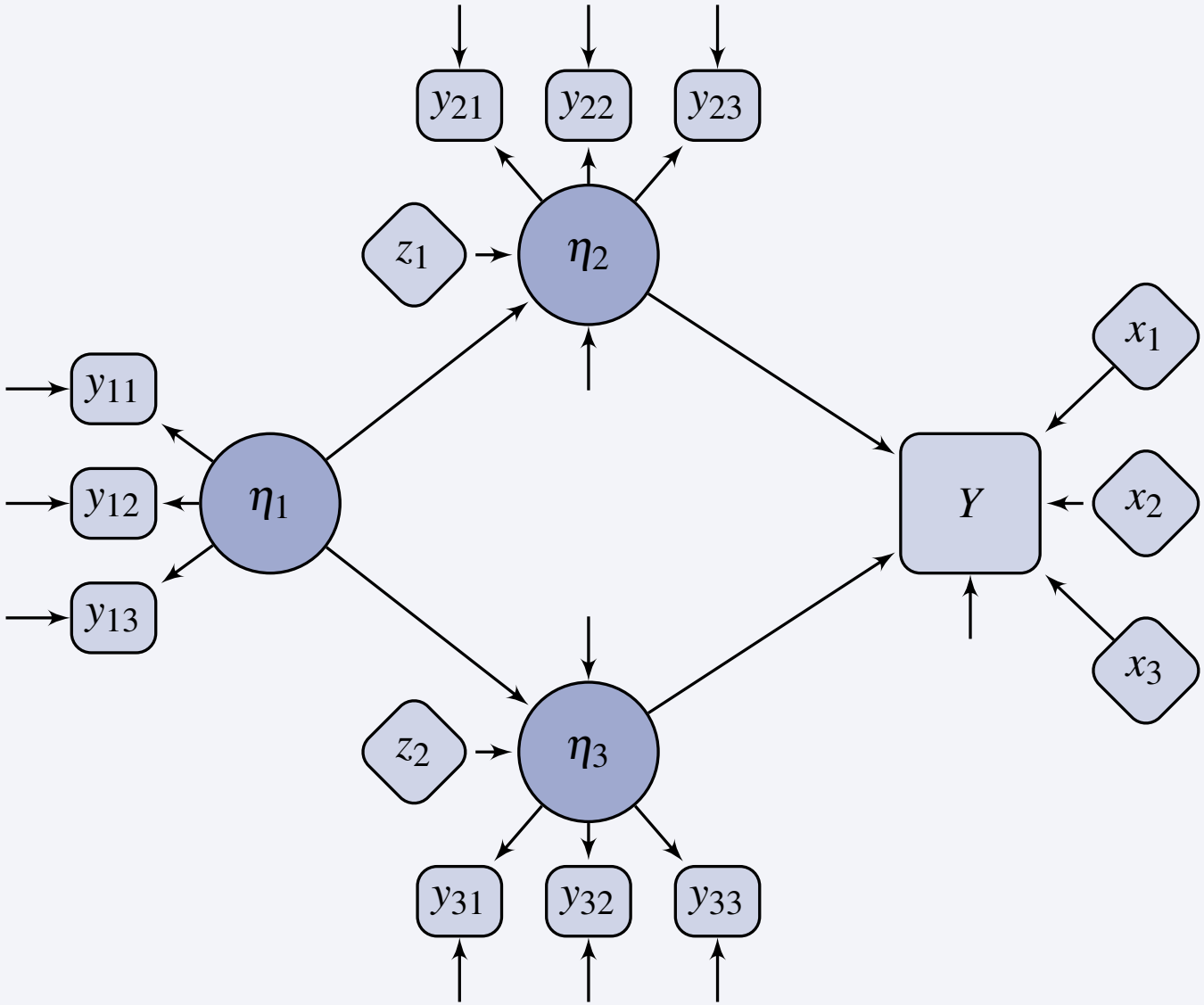
- interrelations between latent constructs can be formalized
- causal mechanisms relating the latent constructs can be disentangled

Some papers implementing SEM

- Guay, Marsch & Boivin (2003): “Academic Self-Concept and Academic Achievement: Developmental Perspectives on their Causal Ordering”
- Marsch, Trautwein, Lüdtke, Köller & Baumert (2005): “Academic Self-Concept, Interest, Grades, and Standardized Test Scores: Reciprocal Effects Models of Causal Ordering”
- Ruban & McCoach (2005): “Gender Differences in Explaining Grades Using Structural Equation Modeling”

Outline

An example: SEM Path diagram



Formal presentation

A SEM consists of two parts:

- Measurement part (p observed variables):

$$y_i = \nu + \Lambda\eta_i + Kx_i + \varepsilon_i, \qquad \varepsilon_i \sim N_p(0, \Theta)$$

- Latent part (m latent variables):

$$\eta_i = \alpha + B\eta_i + \Gamma z_i + \zeta_i, \qquad \zeta_i \sim N_m(0, \Psi)$$

- Covariates x and z can have common components
- Factors have to be scaled for identification (typical in factor analysis)

Some cautions

SEM requires special attention, many possible traps!
cf. Kline (2004) “How to Fool Yourself with SEM”

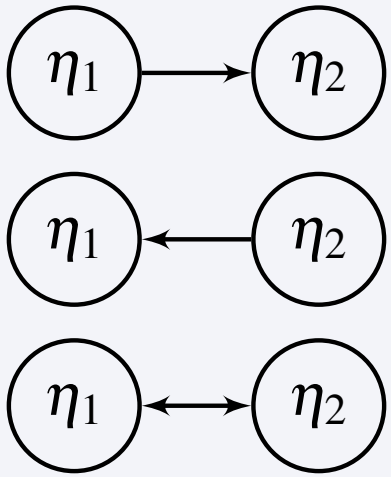
Some burning issues:

- Indicators measuring the latent constructs should be selected carefully and in accordance with the psychological theory
- Model specification should rely on psychological and economic developments
- Omission of causes correlated with some latent variables
- Identification can be problematic
- Interpretation should be done cautiously
- ...

Causality: the problem of equivalent models

Fundamental difficulty in SEM (Williams et al. 1996)

SEMs are said to be ‘equivalent’ when they equally fit the data: identical predicted covariances, same goodness of fit values. Example:



Statistically, impossible to decide between those three latent models!

Goal of this empirical application

Investigate the **impact of personality traits on wages** in Germany

Analysis based on the **Big Five** personality traits approach: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, Neuroticism

The data set

- 2005 wave of the German Socio Economic Panel
- Sample population:
  - German Males in West Germany, aged 25 to 65
  - Full-time employed, excluding self-employed and people on vocational training
- Big Five personality traits: 15 indicators (3 for each trait)
- Outcome variable: log of the gross hourly wage
- Sample size: 3,477 males
- Control variables:
  - Personal characteristics: age, education, assets
  - Family situation: partner, children
  - Job Characteristics: tenure, firm size, civil service

Big Five personality traits indicators

Items in the GSOEP questionnaire, and their corresponding Big Five personality traits:

Traits	Items	I see myself as someone who…	Scores
Openness to Experience	O_1	is original, comes up with new ideas	$\oplus$
	O_2	values artistic experiences	$\oplus$
	O_3	has an active imagination	$\oplus$
Conscientiousness	C_1	does a thorough job	$\oplus$
	C_2	tends to be lazy	$\ominus$
	C_3	does things effectively and efficiently	$\oplus$
Extraversion	E_1	is communicative, talkative	$\oplus$
	E_2	is outgoing, sociable	$\oplus$
	E_3	is reserved	$\ominus$
Agreeableness	A_1	is sometimes somewhat rude to others	$\ominus$
	A_2	has a forgiving nature	$\oplus$
	A_3	is considerate and kind to others	$\oplus$
Neuroticism	N_1	worries a lot	$\oplus$
	N_2	gets nervous easily	$\oplus$
	N_3	is relaxed, handles stress well	$\ominus$

Preliminary Factor analysis on the 15 items

Items	Open.	Consc.	Extra.	Agree.	Neuro.
O_1	0.6670	0.2620	0.2491	-0.1376	-0.0933
O_2	0.6534	-0.0206	-0.0021	0.1663	0.0118
O_3	0.7213	0.0055	0.1563	0.0866	0.0472
C_1	0.0530	0.8146	0.0218	0.0930	0.0139
C_2	0.1758	-0.6584	-0.1392	-0.1754	0.0163
C_3	0.2713	0.7213	0.0494	0.0666	-0.1154
E_1	0.2522	0.2212	0.7245	0.1424	0.0033
E_2	0.3120	0.0800	0.7102	0.2115	0.0074
E_3	0.1163	0.0790	-0.7681	0.2281	0.1654
A_1	0.2014	-0.0668	0.0714	-0.7212	0.2132
A_2	0.1881	0.0487	0.1209	0.6169	0.0188
A_3	0.2049	0.2247	0.0601	0.7387	-0.0168
N_1	0.1724	0.1736	-0.0668	0.0322	0.7236
N_2	-0.0089	-0.1318	-0.0567	-0.0762	0.7899
N_3	0.3371	0.1568	0.0614	0.2039	-0.6444

Principal-component factor analysis, Quartimin rotation.

The question of orthogonality

Is the orthogonality assumption of the Big Five traits realistic?

J. Block (1995): "A contrarian view of the five-factor approach to personality description"

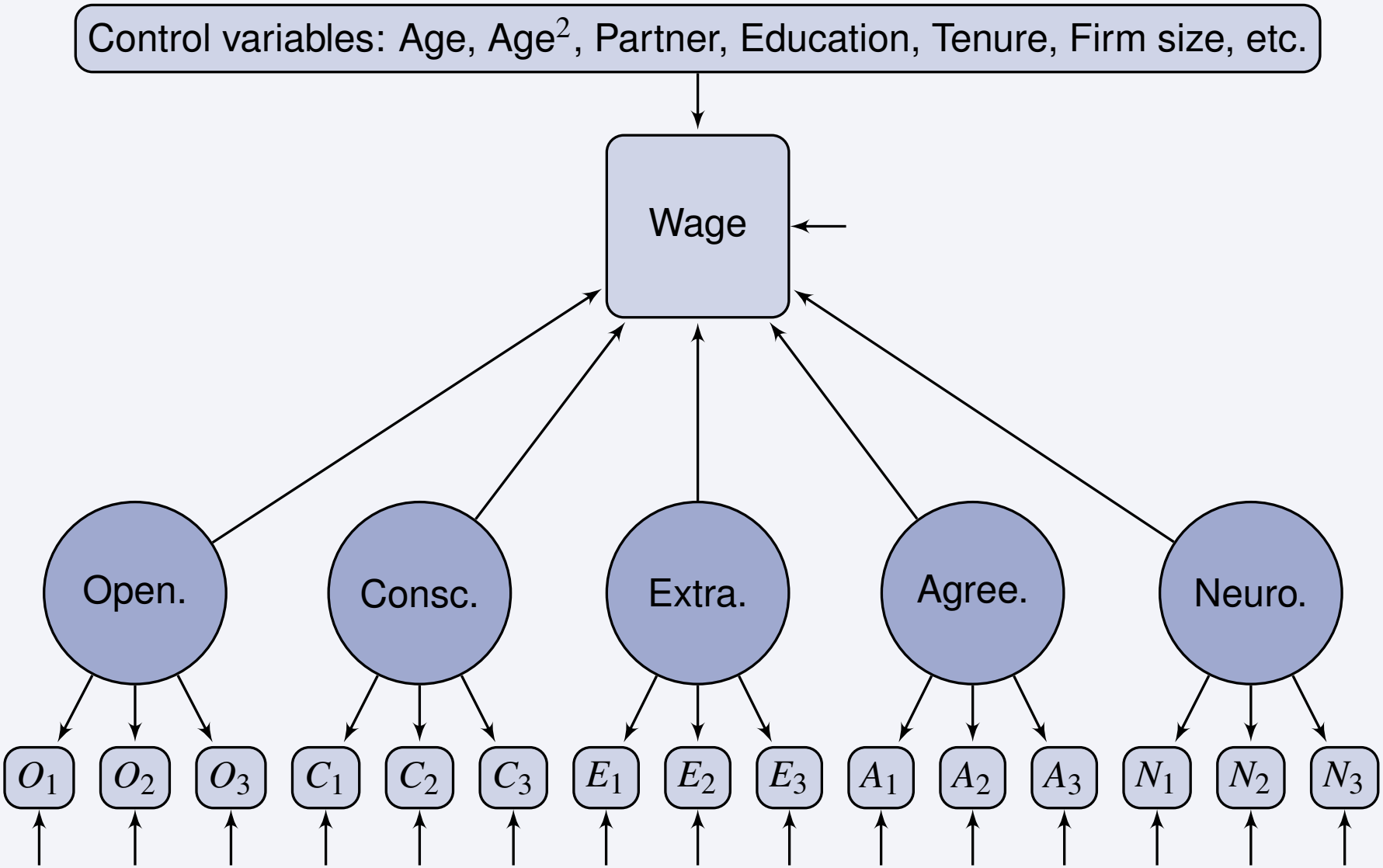
"Repeatedly, the lexical Big Five factors have been described as orthogonal or 'nearly orthogonal' to each other. However, the empirical research findings indicate that the five factors are frequently importantly correlated with each other, usually to reflect an overriding evaluative component."

What would be the consequence(s) if those traits were assumed to be correlated?

Three different approaches

- **Benchmark**: OLS regression on the simple indices (standardized sum of the corresponding indicators)
- **SEM 1**: Latent factors are assumed to be **uncorrelated**
- **SEM 2**: Latent factors are assumed to be **correlated**

Path diagram of the model



Empirical Results

	Benchmark OLS	SEM 1: uncorrelated factors	SEM 2: correlated factors
Open.	$.0242^{***}$	0.0334^{***}	0.0427^{***}
Consc.	$-.0128^{**}$	-0.0169^{**}	-0.0188^{*}
Extra.	$.0117^{*}$	0.0137^{*}	0.0039
Agree.	$-.0071$	-0.0125^{*}	-0.0139
Neuro.	$-.0149^{**}$	-0.0182^{**}	-0.0094
Age	$.3203^{***}$	0.3264^{***}	0.3229^{***}
Age^2	$-.2950^{***}$	-0.3004^{***}	-0.2982^{***}
Educ	$.0687^{***}$	0.0675^{***}	0.0676^{***}
Tenure	$.1856^{***}$	0.1871^{***}	0.1878^{***}
$Tenure^2$	$-.0268^{***}$	-0.0271^{***}	-0.0272^{***}

significance levels: * 10%, ** 5%, *** 1%

Empirical Results

SEM 2: Correlation matrix of the factors:

	Open.	Consc.	Extra.	Agree.	Neuro.
Open.	1.0000				
Consc.	0.4048	1.0000			
Extra.	0.6276	0.3834	1.0000		
Agree.	0.3029	0.4708	0.3299	1.0000	
Neuro.	−0.3172	−0.3168	−0.3190	−0.3625	1.0000

Empirical Results

Apparently, the latent constructs are somehow correlated...

Next step of the analysis would be to **disentangle** the relations between the 5 factors:

- Causal relations between some factors?
- Some underlying factor(s) driving these 5 ones?
- Some covariates influencing the factors? (education, age, tenure?)

Relaxing some assumptions

Relaxing overly restrictive assumptions such as:

- Linearity
- Normality of the factors

Modeling nonlinearities between latent constructs

Parametric approach: Introduce multiple interaction and quadratic effects into the latent part of the model (Klein & Muthén 2007)

Semiparametric approach: Use Structural Equation Mixture Model (SEMM) to approximate nonlinearities component-wise (Bauer 2005)

Relaxing the Normality Assumption

Mixtures of normals can approximate a wide range of distributions (Ferguson 1983)

Assuming normality of NCS can be problematic and yield biased estimates

Idea: use SEMM to relax the normality assumption of the factors

Conclusion / Discussion

- SEM is a powerful tool for the study of NCS
- Implementation requires caution and should rely on psychological and economic theory
- Many possible extensions