A Step Beyond Psychology ——

Exploring The Utility of Factor Analysis in Social Media Survey

Yudong Cao
Factor Analysis Final Project
Vanderbilt University Peabody College
April 14, 2016

Social Media Analytics

+ Gather data from blogs and social media websites

+ Analyze data to make informed business decisions

+ Common Objectives:
  + Mine customer sentiment
  + Support marketing & customer service
  + Increase revenues & reduce customer service costs
  + Get feedback on products & services →SURVEYS!
  + Improve public opinion of product or business division

Factor Analysis in Business

+ *From Encyclopedia Of Business:*

+ "Data reduction" technique

+ Develop scales to measure
  + Customer satisfactions
  + Employee work attitudes

  ➔ Principal Component Analysis?

Objectives

+ Factor Analysis ⬅➡ Social Media Analysis

+ Explore the underlying factors behind social media surveys

+ Evaluate the effectiveness of the surveys

+ Deal with systematically missing data

+ Include latent variables & all other variables
  + Factor Analysis + Principal Component Analysis

Data

+ Facebook Omnibus Survey

+ Pew Institute (2012)

+ Habits & attitudes towards

+ 27-item, 15-question

+ 1006 responses
  + Sample 1 n=601
  + Sample 2 n=505

27 Items (15 Q+12)

+ Ownership: ebook, tablet, MP3, PSP

+ Internet: use, email, mobile

+ Facebook: now, past, start (again), break, why break

+ Facebook: impact in life, time spent, future spend

+ Demo/Bio: state, region, gender, age, height, weight, parents, race, hispanic?

+ Employment, Education, Income

To Explore...

+ Number of factors

+ Best-fit model

+ Interpretation of results

+ Limitations, weaknesses, etc.

+ Missing data

+ Factor analysis on Question-items ONLY
 + VS. all items TOGETHER

Missing Data (Systematic)

- Mplus: convergence failed (too many missing data)

- Reduce dependent missing data:

- FB use (=no) Past? Stopped? Start (again)? → DELETED

- 12+12, Spearman's Rank Correlations

- EFA & CFA on 12 vs 24 items

EFA on 12 Qs

+ Software: CEFA, Mplus

+ CEFA: MWL, Correlation matrix, Oblique rotation, CF-QUARTIMAX, random starts=5
  + → Fix indices

+ Mplus: ULS (default for categorical variables), Oblique rotation, GEOMIN, random starts=30
  + → Factor loadings & correlations matrices

EFA on 12 Qs: 4-Factor!

Scree Plot of Eigenvalues 1.0
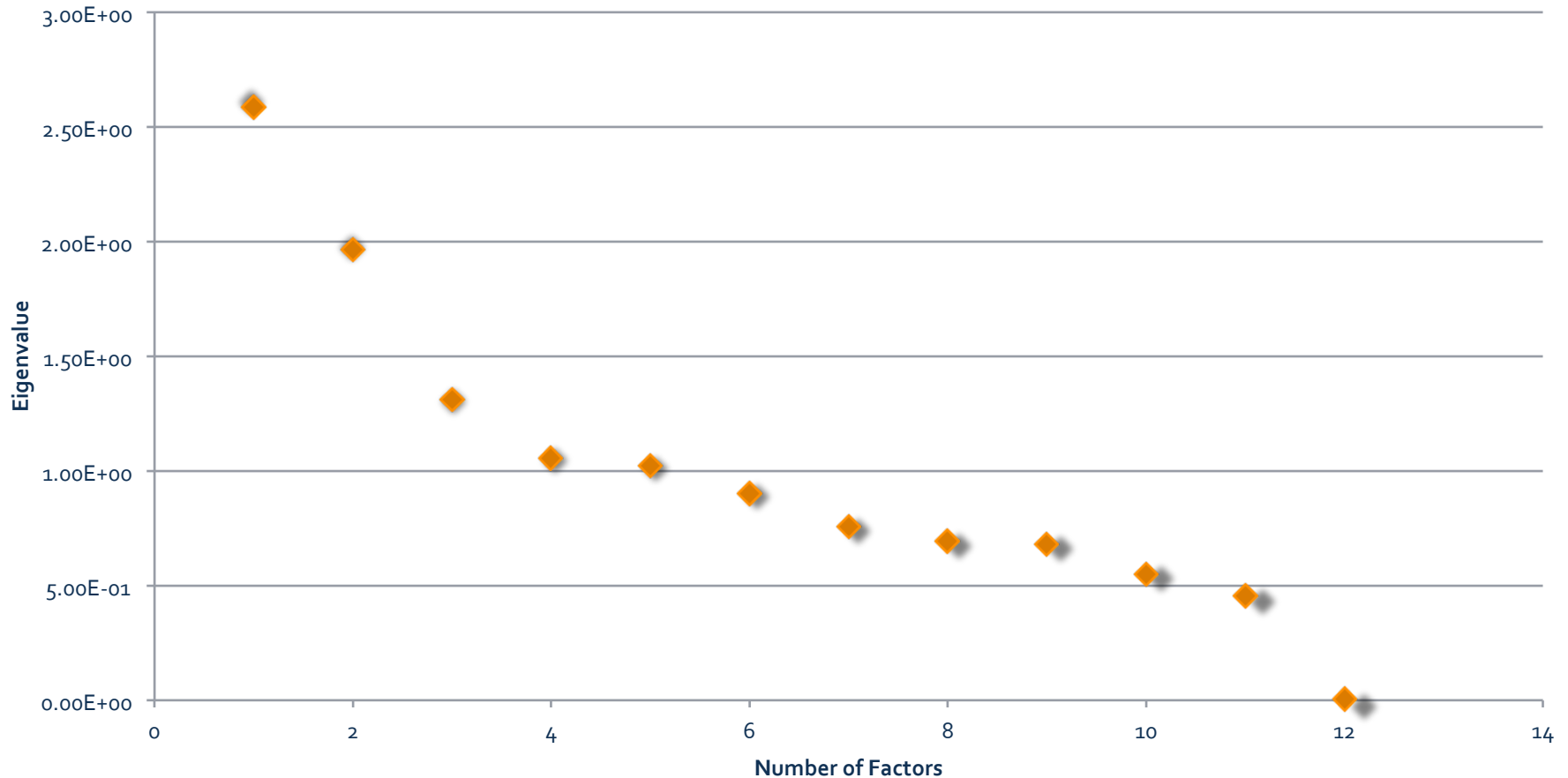
EFA on 12 Qs: 4-Factor!

+ Fit indices from CEFA, n/a in Mplus

+ May be biased but better than n/a

m	$\hat{F}$	χ^2	df	q	p perfect fit	p close fit	$\hat{\varepsilon}$	90% CI for ε
2	0.402	98.998	43	35	0	0.025	0.073	(0.054; 0.092)
3	0.19	46.766	33	45	0.057	0.687	0.041	(0; 0.067)
4	0.101	24.915	24	54	0.41	0.925	0.012	(0; 0.054)
5	0.0559	13.763	16	62	0.616	0.946	0	(0; 0.051)

Factor Loadings Matrix

From Mplus

GEOMIN ROTATED LOADINGS	Ldevice	FBbreak	Mdevice	FBusage
eBook	0.369	0.029	0.002	0.067
Tablet	0.683	-0.011	0.048	0.213
MP3	0.263	-0.038	0.399	-0.097
PS	0.024	0.028	0.619	0.018
IntUse	0.003	-0.029	-0.072	-0.089
Email	0.436	0.003	-0.258	-0.269
IntMobile	0.408	0.029	0.268	0.023
FBbreak?	-0.008	0.998	0.007	0.007
yFBbreak	0.008	0.996	-0.005	-0.005
FBimpact	-0.007	0.015	-0.044	0.618
timeuse	0.031	-0.019	0.023	0.536
futuretim	0.079	0.064	-0.027	0.402

?

Factor Correlations Matrix

From Mplus

GEOMIN FACTOR CORRELATIONS				
	Ldevice	FB break	Mdevice	FBusage
Ldevice	1			
FB break	0.055	1		
Mdevice	0.484	0.017	1	
FBusage	0.239	0.284	0.163	1

CFA on 4-Factor Model

From Mplus: w/o CATEGORICAL ARE

# MVs	AIC	BIC	chisq	df	p
12	7630.061	7709.869	110.846	51	0
11	7610.5	7697.563	82.781	42	0.0002

RMSEA	90% LB	90% UP	CFI	TLI	SRMR
0.065	0.048	0.082	0.948	0.933	0.082
0.059	0.04	0.078	0.964	0.953	0.067

Path Diagram

+11 var

+4 factors
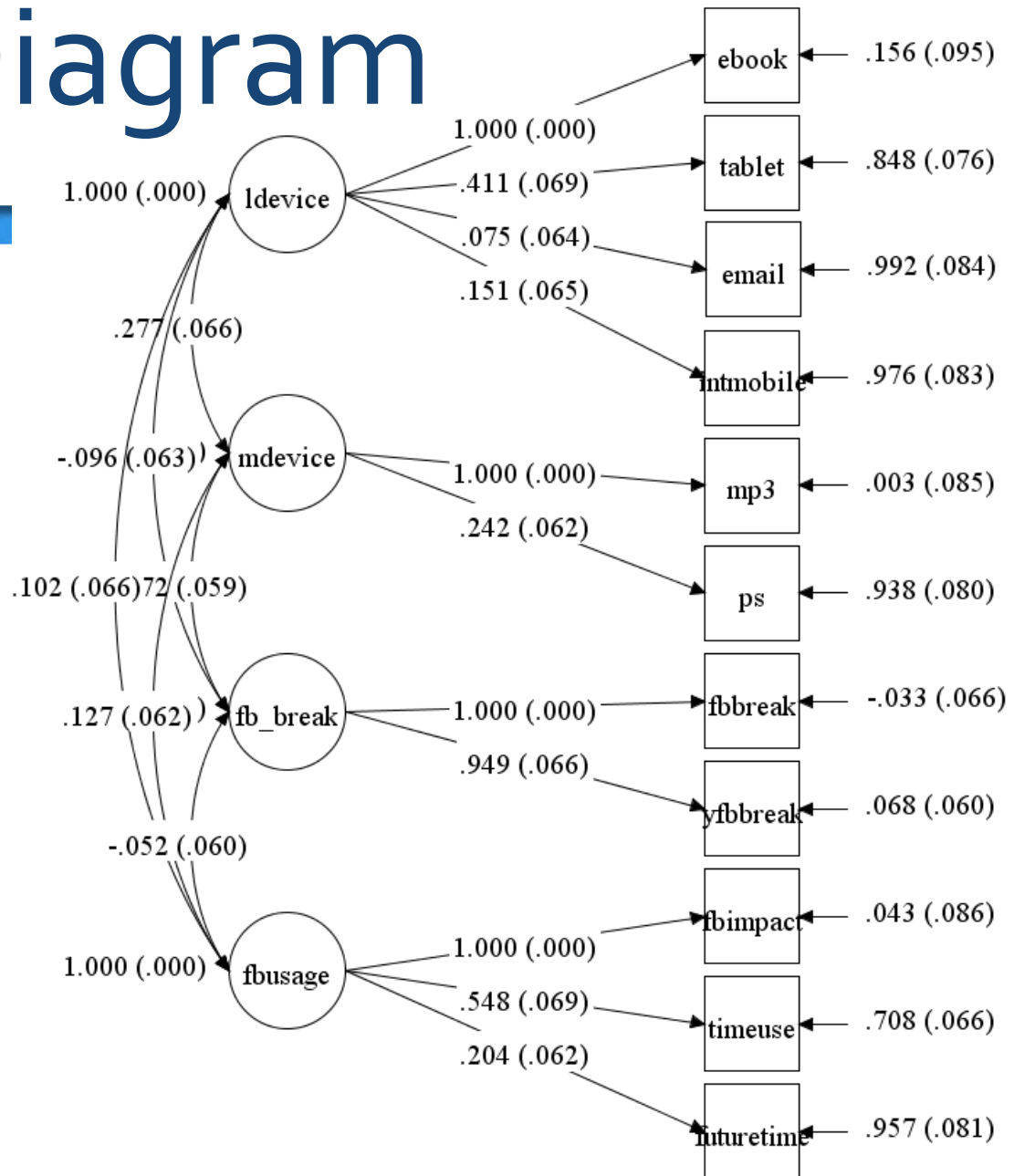
FA on 12 Qs

+ 4-factor model fits close to the survey data

+ InternetUse is not necessary, b/c assumed for all items

+ Factor loadings structure cleaner & more interpretable after dropping InternetUse

+ Clost fit for CFA too!

EFA on 24 Items

+ Software: CEFA, Mplus

+ CEFA: MWL, Correlation matrix, Oblique rotation, CF-QUARTIMAX, random starts=5
  + → Fix indices

+ Mplus: ULS (default for categorical variables), Oblique rotation, GEOMIN, random starts=30
  + → Factor loadings & correlations matrices

EFA on 24 Items

Scree Plot of Eigenvalues 2.0
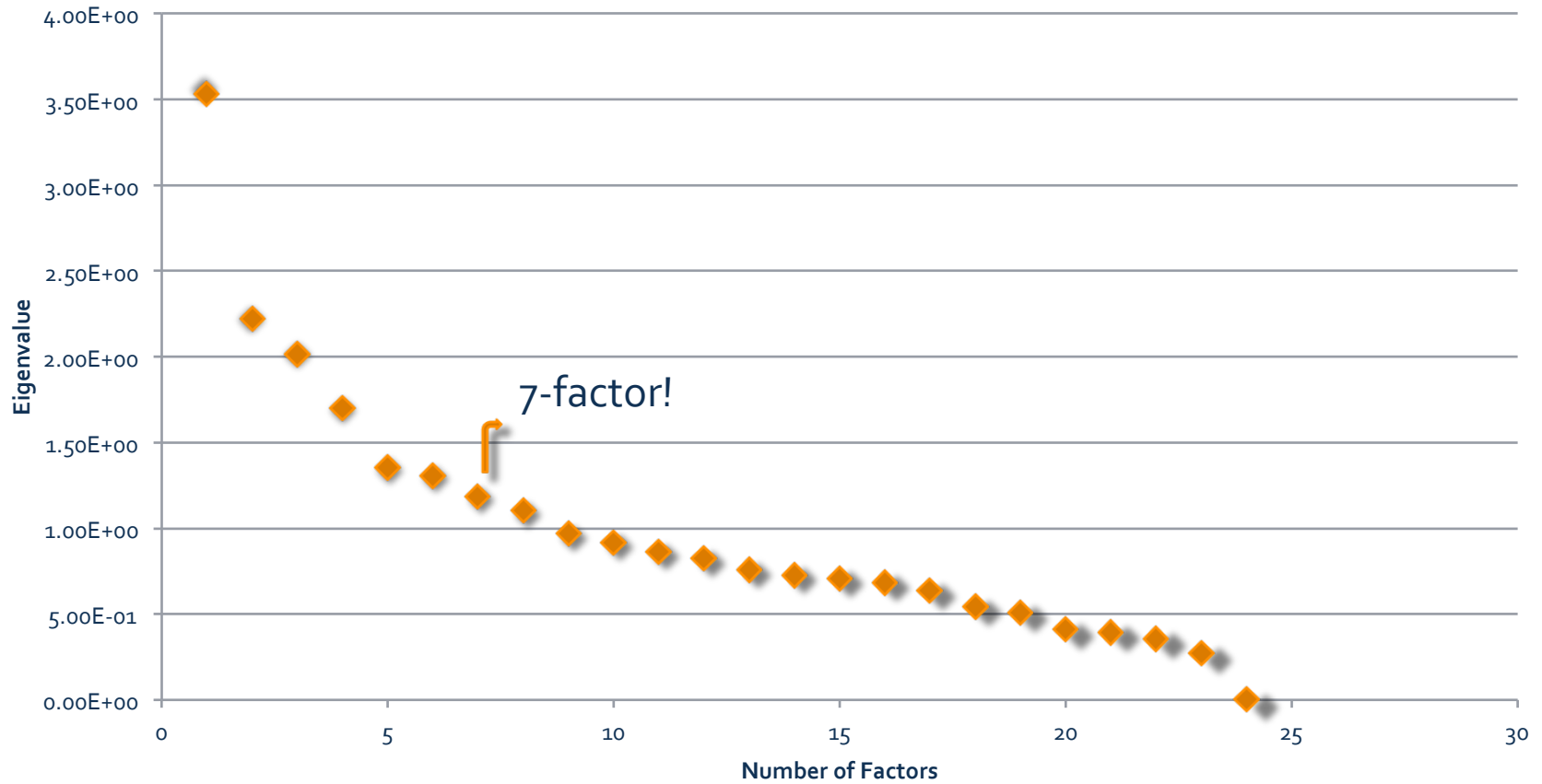
EFA on 24 Items

+ Fit indices from CEFA, n/a in Mplus

+ May be biased but better than n/a

m	F	χ^2	df	q	p perfect fit	p close fit	$\hat{\varepsilon}$	90% CI for ε
3	1.95	479.824	207	93	0	0	0.073	(0.065; 0.082)
4	1.58	388.729	186	114	0	.002	0.067	(0.057; 0.076)
5	1.25	306.546	166	134	0	.083	.059	(0.048; 0.069)
6	1	247.704	147	153	0	.334	.053	(0.041; 0.064)
7	.75	184.889	129	171	.001	.836	.042	(0.027; 0.055)
8	.55	134.807	112	188	.07	.987	.029	(0; .045)

Mplus

	Ldevice	FBbreak	MDBio	GeoDemo	BodyMsr	FBuse	EmpEdu	
STATE	0.012	-0.002	-0.127	0.46	0.042	0.009	0.037	
REGION	0.003	0.045	0.003	-0.49	0.004	0.052	-0.115	
EBOOK	0.403	0.055	0.005	-0.067	-0.062	0.025	0.042	
TABLET	0.664	0.001	0.259	0.016	0.017	0.144	0.003	
MP3	0.215	-0.049	0.37	-0.217	-0.065	-0.115	0.257	
PS	0.174	-0.028	0.609	0.046	0.011	-0.019	-0.079	
INTUSE	0.003	-0.024	-0.099	0.057	-0.067	-0.073	-0.007	x
EMAIL	0.211	0.03	-0.023	0.03	0.135	-0.275	0.198	x
INTMOBIL	0.33	0.017	0.396	-0.03	-0.015	0	0.134	
FBBREAK	-0.008	0.996	0.002	-0.003	-0.011	0.007	0.006	
YFBBREAK	0.007	0.998	0.005	-0.004	0	-0.009	-0.012	
FBIMPACT	0.053	0.022	-0.037	0.011	0.029	0.609	0.038	
TIMEUSE	0.003	-0.02	0.063	-0.059	-0.024	0.551	0.086	
FUTURETI	0.17	0.068	-0.019	0.052	0.005	0.373	-0.124	
HEIGHT	0.004	-0.02	-0.012	-0.013	1.001	-0.018	-0.022	
EMPLOY	-0.278	0.042	0.32	0.03	0.023	-0.025	0.448	
PARENT	0.026	0.1	0.547	0.038	0.068	0.065	0.165	
SEX	-0.195	-0.013	0.167	0.083	-0.032	0.051	0.187	x
AGE	-0.169	0.014	0.744	-0.091	-0.013	0.195	-0.017	
EDUC	-0.001	0.027	-0.013	-0.045	-0.098	-0.116	-0.598	
HISP	-0.033	0.001	0.24	0.488	-0.105	-0.057	-0.006	
RACE	-0.06	-0.006	-0.241	-0.257	0.056	-0.088	0.172	
INCOME	-0.068	0.008	0.105	0.021	0.208	-0.018	-0.663	
WEIGHT	-0.011	0.017	-0.441	-0.018	0.376	0.031	0.14	

Factor Correlations Matrix 2.0

From Mplus

	Device	FBbreak	MDBio	GeoDemo	BodyMsr	FBuse	EmpEdu
Device	1						
FBbreak	0.023	1					
MDBio	0.14	0.116	1				
GeoDemo	-0.122	0.007	0	1			
BodyMsr	-0.099	0.04	-0.294	0.01	1		
FBuse	0.186	0.284	0.164	-0.075	0.053	1	
EmpEdu	0.234	-0.048	-0.006	0.035	0.112	0.044	1

CFA on 24 Items 7 Factors

From Mplus: w/o CATEGORICAL ARE

# var	AIC	BIC	CHISQ	DF	P
24	17868.1	18089.38	619.558	239	0
21	15501.22	15700.74	402.831	176	0

RMSEA	90% CI LB	90% CI UB	CFI	TLI	SRMR
0.076	0.068	0.083	0.748	0.709	0.092
0.068	0.059	0.077	0.832	0.799	0.084

Path Diagram

+ 21 var

+ 7 factors
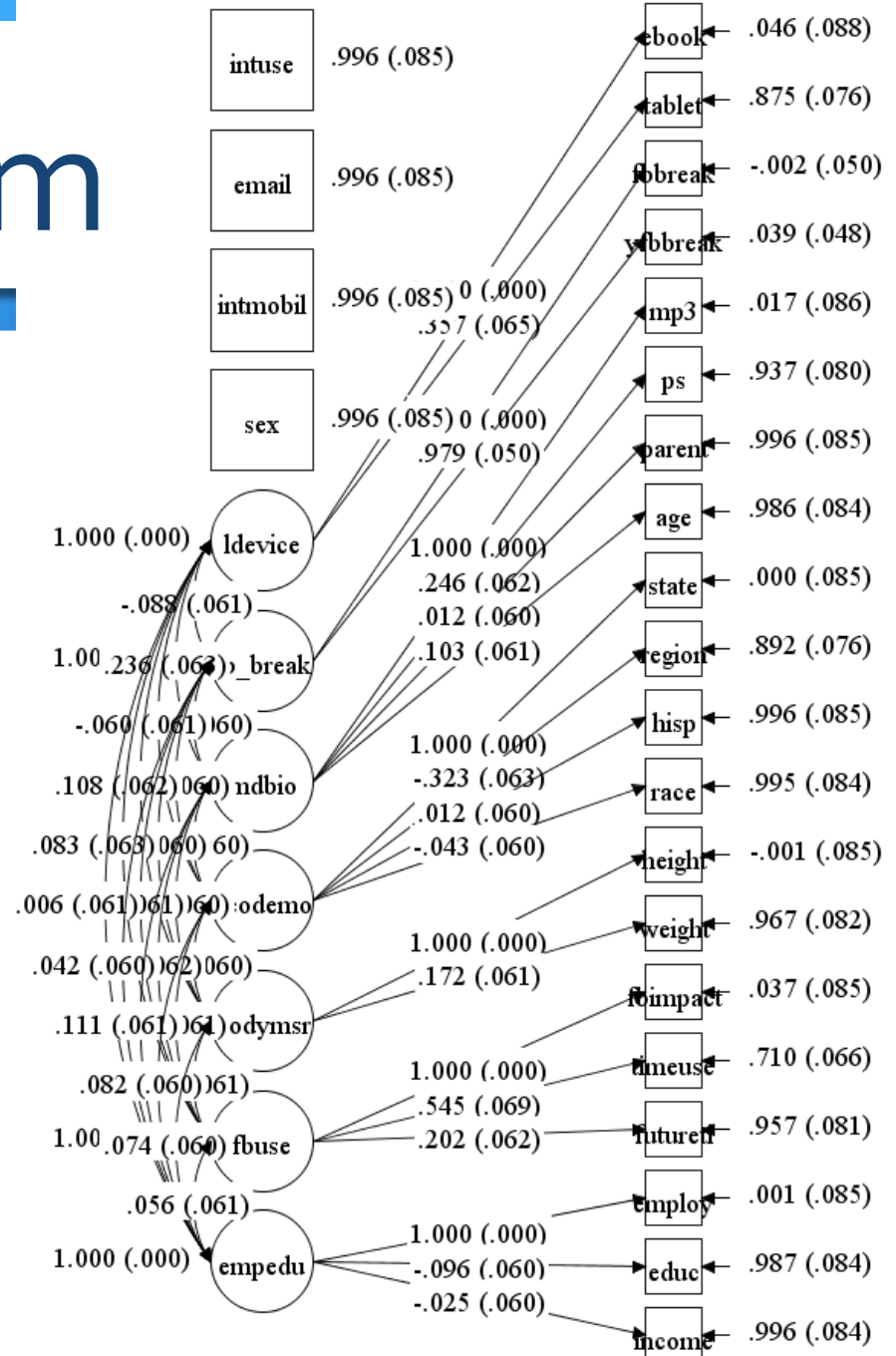
FA on 24 Items

+ 7-factor model has close-to-mediocre fit to the data

+ CFA (mediocre) fit worse than EFA

+ FA *not* very applicable for 24 variables

+ InternetUse, Email & Sex do *not* load on any factor

    + Survey modification suggested

Summary

+ Factor Analysis in social media surveys:

+ Works for "data reduction"
  + Problematic questions/items can be detected

+ Explains question-type items better than all items combined
  + (Factor Analysis x Principal Component Analysis) needs further research

Future social media surveys

+ Can adopt the questions in this survey to survey customers' technology use & attitude towards Facebook

+ Need more specific items about specific products, market research goals, etc.

+ Another Factor Analysis!

References

+ http://searchbusinessanalytics.techtarget.com/definition/social-media-analytics

+ http://www.referenceforbusiness.com/encyclopedia/Ent-Fac/Factor-Analysis.html

+ http://www.pewinternet.org/datasets/december-2012-facebook-omnibus/

+ http://www.noupe.com/essentials/freebies-tools-templates/exclusive-vector-freebie-new-social-media-icon-set.html

+ CEFA + Mplus

+ Dr. Preacher's suggestions