Measuring Subordination of Women in the Home

A Confirmatory Factor Analysis for Ordinal Variables

Kaylee Hodgson

Brigham Young University

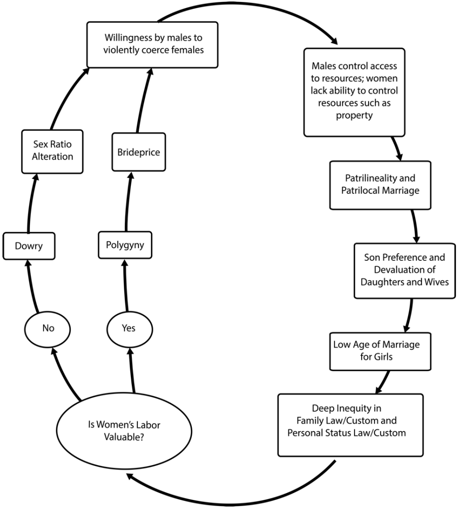
December 13, 2018

Introduction

- The WomanStats Project released a multivariate country-level scale in 2017: Patrilineality/Fraternity Syndrome Scale
- Combines 11 indicators that measure subordination of women in the household
- Scale was built for large-scale analyses - test theory that societies with systematic gender suppression and inequality in the home perform worse on macro-level indicators of success

Purpose: Identify whether a single factor is sufficient for the 11 variables.

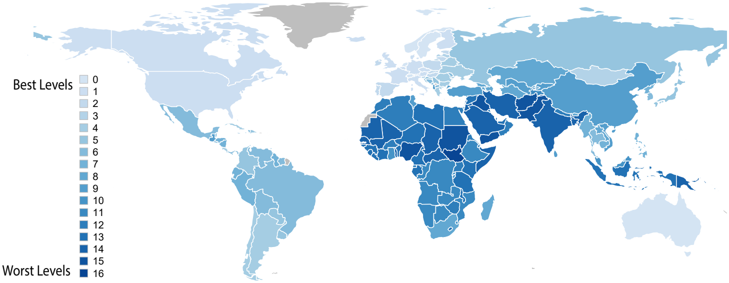
The Theory of the Syndrome Scale



The Syndrome Scale

Combines 11 indicators of the subordination of women in the home:

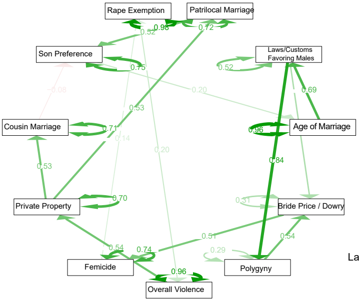
- prevalence of patrilocal marriage
- prevalence of brideprice or dowry
- prevalence and legality of polygyny
- presence of cousin marriage
- age of marriage for girls
- laws and practices surrounding women's property rights
- presence of son preferences or sex ratio alteration
- presence of inequity in family law/custom that favors males
- overall level of violence against women in society
- presence of societal sanction for femicide
- whether there is legal exoneration for rapists who offer to marry their victims



Path Analysis

- essentially an extension of multiple regression analysis, but on multiple levels
- special case of structural equation modeling (SEM) - no latent variable
- used to evaluate the "causal" theory behind the scale - should be cautious in interpretation

Path Analysis Results



Correlation Plot

	Laws/Customs Favor Males	Age of Marriage	Bride Price/Dowry	Polygyny	Overall Violence	Femicide	Private Property	Cousin Marriage	Son Preference	Rape Exemption
Son Preference										0.09
Cousin Marriage									0.11	0.14
Private Property								0.53	0.32	0.25
Femicide							0.47	0.33	0.33	0.14
Overall Violence						0.58	0.55	0.38	0.37	0.2
Polygyny					0.56	0.44	0.72	0.56	0.21	0.28
Bride Price/Dowry				0.81	0.55	0.51	0.71	0.54	0.33	0.27
Age of Marriage			0.46	0.53	0.44	0.48	0.53	0.27	0.2	0.26
Laws/Customs Favor Males		0.69	0.78	0.84	0.64	0.57	0.73	0.52	0.31	0.26
Patrilocal Marriage	0.61	0.35	0.61	0.61	0.55	0.43	0.53	0.36	0.49	0.18

RMSEA=0.205 & SRMR=0.3 (standard is that they should both be below 0.08)

- Possible reason for poor fit: situations described in the theory are sometimes only partially covered by a single variable, and some variables describe multiple phenomena at once

Confirmatory Factor Analysis (CFA)

- All variables are ordinal (scales with 2-5 levels)
- CFA using maximum likelihood estimators performs poorly with ordinal data that has few levels
- Other options considered: Weighted least square (WLS), Diagonally weighted least square (DWLS), and Unweighted Least Square (ULS)
- WLS - poor performance when sample size is small
- DWLS - less accurate estimates and less precise standard errors

Unweighted Least Squares

The general structure for a least squares estimator is:

$$q(\boldsymbol{\theta}; \mathbf{S}) = (\text{vech}\mathbf{S} - \text{vech}\boldsymbol{\Sigma}[\boldsymbol{\theta}])'\hat{\mathbf{V}}^{-1}(\text{vech}\mathbf{S} - \text{vech}\boldsymbol{\Sigma}[\boldsymbol{\theta}]), \tag{1}$$

where the method finds $\hat{\boldsymbol{\theta}}_{LS}$ that minimizes $q(\boldsymbol{\theta}; \mathbf{S})$. $q(\hat{\boldsymbol{\theta}}_{LS}; \mathbf{S})$ approaches a chi-squared distribution ($\chi^2_{p*-q,\alpha}$, $p* = \frac{1}{2}p(p+1)$). In the case of the unweighted least square model, $\hat{\mathbf{V}}^{-1} = \mathbf{I}$.

Linear Combination Equations for CFA Models

	One-Factor Model	Two-Factor Models
Polygyny	$y_1 = \lambda_1 f_1 + \epsilon_1$	$y_1 = f_2 + \epsilon_1$
Laws/Customs Favoring Males	$y_2 = \lambda_2 f_1 + \epsilon_2$	$y_2 = \lambda_{21} f_1 + \lambda_{22} f_2 + \epsilon_2$
Bride Price/Dowry	$y_3 = \lambda_3 f_1 + \epsilon_3$	$y_3 = \lambda_{31} f_1 + \lambda_{32} f_2 + \epsilon_3$
Property Rights	$y_4 = \lambda_4 f_1 + \epsilon_4$	$y_4 = \lambda_{41} f_1 + \lambda_{42} f_2 + \epsilon_4$
Cousin Marriage	$y_5 = \lambda_5 f_1 + \epsilon_5$	$y_5 = \lambda_{51} f_1 + \lambda_{52} f_2 + \epsilon_5$
Age of Marriage	$y_6 = \lambda_6 f_1 + \epsilon_6$	$y_6 = \lambda_{61} f_1 + \lambda_{62} f_2 + \epsilon_6$
Legal Exoneration for Rapists	$y_7 = \lambda_7 f_1 + \epsilon_7$	$y_7 = \lambda_{71} f_1 + \lambda_{72} f_2 + \epsilon_7$
Son Preference	$y_8 = \lambda_8 f_1 + \epsilon_8$	$y_8 = \lambda_{81} f_1 + \lambda_{82} f_2 + \epsilon_8$
Patrilocal Marriage	$y_9 = f_1 + \epsilon_9$	$y_9 = \lambda_{91} f_1 + \lambda_{92} f_2 + \epsilon_9$
Overall Violence	$y_{10} = \lambda_{10} f_1 + \epsilon_{10}$	$y_{10} = \lambda_{101} f_1 + \lambda_{102} f_2 + \epsilon_{10}$
Societal Santion of Femicide	$y_{11} = \lambda_{11} f_1 + \epsilon_{11}$	$y_{11} = f_1 + \epsilon_{11}$
Only Second-Order CFA		$(f_1 = \beta_1 f_3 + e_1,\ f_2 = \beta_2 f_3 + e_2)$

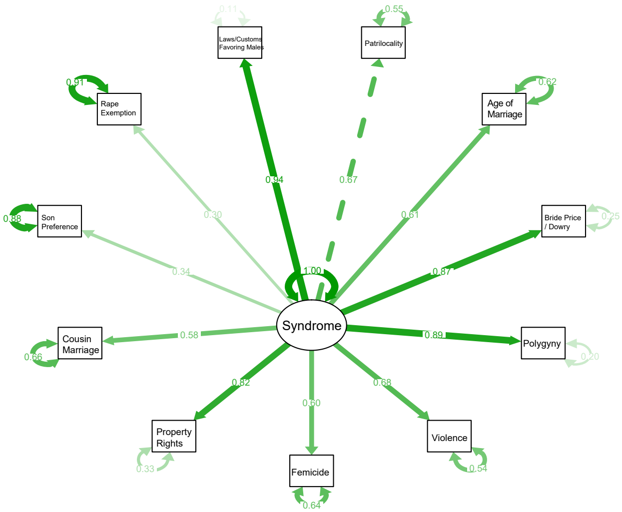
Model Fit Diagnostics

- AGFI: measures the "proportion of variance accounted for by the estimated population covariance"
- NFI: gives the percentage that the model improves the fit from the null model
- SRMR measures the difference between the **S** matrix for the data and the $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ estimated from the model
- AVE: the average of the the R^2 values in the factor

	Diagnostics			
CFA	AGFI	NFI	SRMR	AVE
Rule for Good Fit	AGFI $\geq$ 0.90	NFI $\geq$ 0.95	SRMR $<$ 0.08	AVE $>$0.5
One-Factor Model	0.989	0.990	0.053	0.499
Two-Factor Model	0.994	0.993	0.042	0.537
Two-Factor Second	0.993	0.993	0.044	0.548

- All three models fit well, two-factor models fit only marginally better than one-factor

The One-Factor CFA Model



Variable Estimates

RMSEA = 0.000, CFI = 1.000

Variable	Estimate	Standard Error	z-value	p-value	R^2
Patrilocality	1.00				0.46
Age of Marriage	1.33	0.09	15.16	0.00	0.38
Bride Price/Dowry	1.519	0.10	15.70	0.00	0.75
Polygyny	2.32	0.14	16.58	0.00	0.80
Violence	0.956	0.072	13.37	0.00	0.46
Femicide	0.621	0.05	10.43	0.00	0.36
Property Rights	1.49	0.10	15.632	0.00	0.67
Cousin Marriage	1.27	0.09	14.95	0.00	0.34
Son Preference	0.45	0.06	8.20	0.00	0.12
Rape Exemption	0.23	0.05	4.59	0.00	0.09
Laws/Customs Favoring Males	1.84	0.11	16.28	0.00	0.89

Simulation Study

Common issue with country-level data: 1) data is missing altogether or 2) data is available but poorly measured or estimated

- 1%, 5%, 10%, 25%, 50%, and 90% of the countries randomly selected using random draws from Bernoulli, p=each percentage
- Create two data sets for each percentage value. If Bernoulli draw = 1...
- First data set: delete entire row
- Second data set: replace observed values in row with randomly drawn values from Binomial distribution, with number of trials equal to number of values in each scale.

Simulation Study Results

- RMSEA, CFI, and SRMR estimates don't change
- Missing data: R^2 values similar and estimates remain significant until 50% missing, when model will no longer estimate
- Poor-quality data: R^2 values decrease as percentage reassigned increases, but estimates remain significant. At 50%, negative R^2 values are estimated and all variable estimates become insignificant

Discussion and Conclusion

- One-factor CFA model fits the 11 indicators well
- Path analysis does not confirm the theory of the causal relationship
- Future research: find variables that better separate to fit the theory, evaluate whether the actual scale fits the data well

Overall, I find sufficient evidence to conclude that the single factor fits the data well and significantly describes variation in the individual variables. These results are robust to low to moderate levels of missing or poor-quality data.