

# **Structural Equation Modeling of the Global Ecological Footprint:**

**Interplay Between Human Development, Resource Consumption, and Environmental Sustainability**

Ayisha C. O. K.

Student ID: 1378611

STAT\*6821: Multivariate Analysis

Department of Mathematics and Statistics

University of Guelph

December 2024

## Abstract

This study employs structural equation modeling (SEM) to examine the relationships between human development, ecological footprint, biocapacity, and environmental sustainability using data from 139 countries. Three latent constructs were specified: *DEVELOPMENT* (SDG index, life expectancy, HDI), *FOOTPRINT* (cropland and carbon footprints), and *BIOCAPACITY* (forest land and fishing grounds). Results indicated that development strongly predicts footprint ( $\gamma = 0.72$ ,  $R^2 = 0.52$ ), footprint negatively affects ecological reserve ( $\beta = -0.34$ ), and biocapacity positively predicts reserve ( $\beta = 1.03$ ). The model achieved excellent fit (CFI = 0.945, TLI = 0.910). Biocapacity's effect was approximately three times larger than footprint's, indicating that natural resource endowments currently constrain sustainability more than consumption patterns. These findings emphasize the need for the simultaneous transformation of development pathways and the preservation of biocapacity.

## 1 Introduction

The global ecological crisis presents one of the most pressing challenges of the 21st century. As nations pursue economic development and improved living standards, the resulting environmental footprint threatens planetary boundaries (Rockström et al., 2009). Understanding the complex relationships between human development, resource consumption patterns, and environmental capacity is crucial for sustainable development policy. The relationship between human development and environmental impact has been central to ecological economics for decades. The ecological footprint concept was introduced as a standardized measure of human demand on nature, quantifying how much biologically productive land and water area is required to produce resources and absorb waste (Wackernagel and Rees, 1996). This metric enables comparison of human consumption against planetary regenerative capacity, providing a clear indicator of ecological sustainability or overshoot.

The global footprint network has refined this methodology over three decades, producing comprehensive national accounts that track both footprint (demand) and biocapacity (supply) across multiple land use categories (Global Footprint Network, 2023). Recent iterations incorporate improved data quality and methodological advances, making cross-national comparisons increasingly robust. These accounts reveal that humanity currently consumes resources at 1.7 times earth's regenerative capacity, with substantial variation across nations driven by development levels and consumption patterns.

The environmental kuznets curve (EKC) hypothesis suggests that economic development initially increases environmental degradation before eventually declining at higher income levels, producing an inverted U-shaped relationship (Stern, 2004). Early industrialization generates pollution and resource extraction, but wealth accumulation eventually enables investments in cleaner technologies and environmental protection. However, empirical evidence for global-scale decoupling remains contentious, particularly when consumption-based rather than

production-based accounting is employed. High-income nations may appear to reduce environmental impacts domestically while effectively outsourcing degradation through international trade.

The I=PAT framework ( $\text{Impact} = \text{Population} \times \text{Affluence} \times \text{Technology}$ ) provides another influential theoretical lens, positing that environmental impact results from population size, per capita consumption (affluence), and production technology efficiency (Dietz and Rosa, 1997). This framework predicts that development, by increasing both population and especially affluence, will intensify environmental pressures unless offset by dramatic technological improvements. Empirical tests generally confirm affluence (measured through GDP per capita or similar indicators) as the dominant driver of carbon emissions and resource consumption, with elasticities often exceeding 1.0.

The structural equation modeling (SEM) has emerged as a powerful tool for testing complex environmental hypotheses involving multiple interconnected processes. A comprehensive review by Fan et al. (2016) demonstrated SEM's advantages for modeling latent constructs (unobservable phenomena measured through multiple indicators), testing mediating pathways, and accounting for measurement error. These capabilities make SEM particularly suited for investigating development-environment relationships, where constructs like "human development" and "environmental capacity" are inherently multidimensional.

While existing literature establishes that development increases the environmental footprint and that this footprint contributes to ecological deficits, some questions remain underexplored. This study employs SEM with confirmatory factor analysis (CFA) to investigate how human development influences ecological footprint and how both interact with biocapacity to determine ecological sustainability. Three primary research questions guide this analysis: (1) Does human development (measured through sustainable development goals, life expectancy, and human development index) predict increased ecological footprint consumption? (2) How do consumption patterns (cropland and carbon footprint) relate to ecological balance outcomes? (3) What is the relative importance of ecological footprint versus biocapacity in determining ecological reserve or deficit?

SEM provides a powerful framework for testing these relationships as it allows simultaneous estimation of measurement models (latent constructs) and structural relationships, while accounting for measurement error (Kline, 2015). Unlike traditional regression approaches, SEM enables modeling of unobserved constructs such as "human development" and "biocapacity" through multiple indicators, providing more robust estimates of theoretical relationships. This multivariate technique is particularly suited for complex ecological systems where multiple interconnected processes operate simultaneously.

## 2 Data overview

The data come from a Kaggle dataset compiling the Global Footprint Network's 2023 National Footprint and Biocapacity Accounts ([Jain, 2023](#)), which provides comprehensive country-level measurements of ecological resource consumption and capacity for 182 countries worldwide. The original data were collected and maintained by the Global Footprint Network's 2023 National Footprint and Biocapacity Accounts ([Global Footprint Network, 2023](#)). After listwise deletion of cases with missing data, the final analytical sample comprised 139 countries (76.4% retention), exceeding the recommended minimum of 100 cases for SEM analysis ([Kline, 2015](#)) and providing adequate statistical power for detecting medium-sized effects. Table 1 presents descriptive statistics for all key variables.

Three variable sets corresponded to theoretical constructs. **Human Development Indicators** (exogenous latent variable) included: (1) Sustainable Development Goals Index (SDGi; 0-100 scale) measuring UN development goal progress, (2) Life Expectancy (years) indicating health outcomes, and (3) Human Development Index (HDI; 0-1 scale), a composite measure of health, education, and income. **Ecological Footprint Indicators** (endogenous latent variable) included: (1) Cropland Footprint (global hectares per capita) for crop production, and (2) Carbon Footprint (global hectares per capita) for carbon sequestration from fossil fuel use. **Biocapacity Indicators** (exogenous latent variable) included: (1) Forest Land (global hectares per capita) measuring forest biocapacity, and (2) Fishing Ground (global hectares per capita) measuring marine biocapacity. **Ecological Reserve/Deficit** (observed outcome) measured the difference between biocapacity and ecological footprint, with positive values indicating sustainability and negative values indicating deficit.

All variables were continuous and required minimal preprocessing. The primary preprocessing step was handling missing data through listwise deletion, which is appropriate when data are missing at random and sample size remains adequate ([Kline, 2015](#)). No transformations were applied to maintain interpretability of standardized coefficients.

Table 1: Descriptive statistics for key variables (N=139)

Variable	Mean	SD	Min	Max
SDGi	66.58	11.77	0.00	86.50
Life expectancy (years)	71.12	7.69	53.00	84.00
HDI	0.72	0.16	0.39	0.96
Cropland footprint (gha)	0.61	0.35	0.10	1.90
Carbon footprint (gha)	1.62	1.82	0.00	11.60
Forest land (gha)	3.05	5.34	0.00	31.54
Fishing ground (gha)	0.92	1.50	0.00	11.30
Ecological reserve (gha)	0.18	6.78	-7.62	46.61

Note: gha = global hectares per capita

### 3 Methodology: Structural equation modeling framework

SEM combines factor analysis and path analysis to test complex theoretical models (Byrne, 2016). SEM consists of two integrated components: (1) the measurement model (CFA) tests whether observed variables adequately measure hypothesized latent constructs, and (2) the structural model tests directional relationships among latent variables. The general framework can be expressed as:

**Measurement model:**

$$\mathbf{x} = \Lambda_x \boldsymbol{\xi} + \boldsymbol{\delta}, \quad \mathbf{y} = \Lambda_y \boldsymbol{\eta} + \boldsymbol{\epsilon} \quad (1)$$

**Structural model:**

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad (2)$$

where  $\boldsymbol{\xi}$  represents exogenous latent variables,  $\boldsymbol{\eta}$  represents endogenous latent variables,  $\Lambda_x$  and  $\Lambda_y$  are factor loading matrices, and  $\boldsymbol{\delta}$ ,  $\boldsymbol{\epsilon}$ , and  $\boldsymbol{\zeta}$  are error terms.

The hypothesized model (Figure 2) specifies three latent factors: DEVELOPMENT ( $\xi_1$ , exogenous) measured by SDGi, Life expectancy, and HDI; FOOTPRINT ( $\eta_1$ , endogenous) measured by cropland and carbon footprint; and BIOCAPACITY ( $\xi_2$ , exogenous) measured by forest land and fishing ground. Three hypotheses were tested: H1: DEVELOPMENT positively predicts FOOTPRINT ( $\gamma_{11} > 0$ ); H2: FOOTPRINT negatively predicts ecological reserve ( $\beta_{21} < 0$ ); H3: BIOCAPACITY positively predicts ecological reserve ( $\beta_{22} > 0$ ). The structural equations are: FOOTPRINT =  $\gamma_{11} \cdot$  DEVELOPMENT +  $\zeta_1$  and Ecological reserve =  $\beta_{21} \cdot$  FOOTPRINT +  $\beta_{22} \cdot$  BIOCAPACITY +  $\epsilon$ .

The model was estimated using maximum likelihood estimation with robust standard errors (MLR estimator) in lavaan version 0.6-20 (Rosseel, 2012). Model fit was evaluated using multiple indices: comparative fit index (CFI  $\geq 0.90$  acceptable,  $\geq 0.95$  excellent), Tucker-Lewis index (TLI  $\geq 0.90$  acceptable,  $\geq 0.95$  excellent), root mean square error of approximation (RMSEA  $< 0.08$  acceptable), and standardized root mean square residual (SRMR  $< 0.08$  good fit) (Hu and Bentler, 1999).

The analysis proceeded through iterative refinement: Model 1 (initial) included all available indicators including total biocapacity and forest product footprint (CFI = 0.883, RMSEA = 0.226); Model 2 removed total biocapacity due to extreme multicollinearity with forest land ( $r = 0.98$ ), causing estimation problems (CFI = 0.896, RMSEA = 0.199); Model 3 (final) removed forest product footprint due to weak loading ( $\lambda = 0.33$ ) and non-significance ( $p = 0.074$ ) in alternative testing (CFI = 0.945, RMSEA = 0.166). This iterative process follows standard SEM practice (Byrne, 2016) of balancing statistical fit with theoretical validity.

Table 2: Selected correlations among key variables (N=139)

Variable pair	r	p
<i>Within DEVELOPMENT Factor</i>		
Life expectancy – HDI	0.914	<.001
SDGi – HDI	0.824	<.001
SDGi – Life expectancy	0.771	<.001
<i>Within FOOTPRINT factor</i>		
Cropland – Carbon	0.379	<.001
<i>Within BIOCAPACITY factor</i>		
Forest land – Fishing ground	0.642	<.001
<i>Outcome correlations</i>		
Forest land – Ecological reserve	0.947	<.001
Carbon footprint – Ecological reserve	-0.263	.002

## 4 Results

Correlation analysis (Figure 1) revealed patterns supporting the hypothesized factor structure (Table 2). Within-construct correlations were strong: development indicators (mean  $r = 0.836$ ), biocapacity indicators ( $r = 0.642$ ), and footprint indicators ( $r = 0.379$ ). The correlation matrix demonstrated discriminant validity, with between-construct correlations lower than within-construct correlations. The strongest predictor of ecological reserve was forest land ( $r = 0.947$ ), while carbon footprint showed a moderate negative relationship ( $r = -0.263$ ).

The final measurement model included three latent factors with seven indicators. All factor loadings were statistically significant ( $p < .001$ ) and substantively strong (Table 3). The DEVELOPMENT factor was exceptionally well-defined (loadings 0.823-1.002), with HDI's near-perfect loading (1.002) indicating it is an almost pure measure of the construct. The FOOTPRINT factor demonstrated adequate to strong measurement, with carbon footprint as the dominant indicator ( $\lambda = 0.929$ ,  $R^2 = 0.863$ ). The BIOCAPACITY factor showed strong properties, with forest land as the primary indicator ( $\lambda = 0.942$ ,  $R^2 = 0.887$ ).

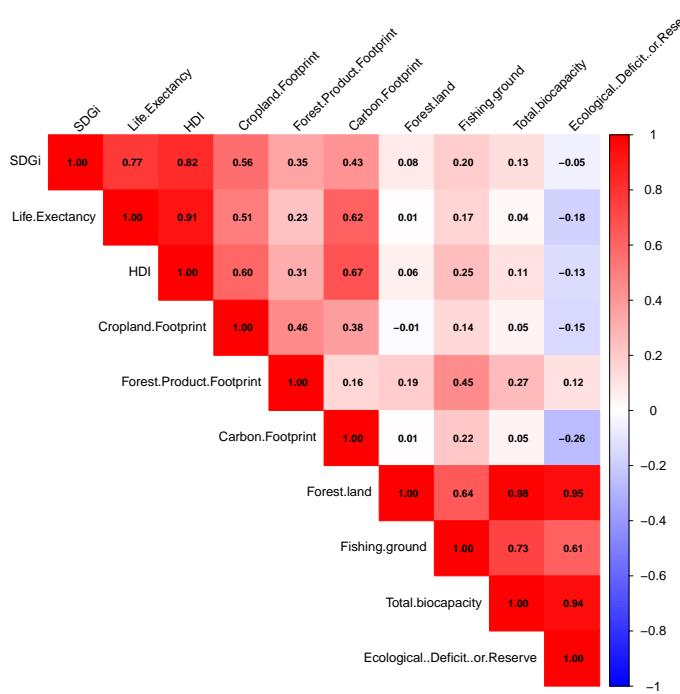


Figure 1: Correlation matrix heatmap for SEM variables (N=139). Warmer colors indicate stronger positive correlations, cooler colors indicate stronger negative correlations.

Table 3: Standardized factor loadings and R<sup>2</sup> Values (Final model)

Factor	Indicator	$\lambda$	SE	R <sup>2</sup>
DEVELOPMENT	SDGi	0.823	–	0.677
	Life expectancy	0.912	0.040	0.832
	HDI	1.002	0.001	1.000
FOOTPRINT	Cropland footprint	0.485	–	0.235
	Carbon footprint	0.929	2.659	0.863
BIOCAPACITY	Forest land	0.942	–	0.887
	Fishing ground	0.671	0.016	0.450

All  $p < .001$ . First indicator per factor fixed to 1.0.

All three hypothesized structural paths were statistically significant and in predicted directions (Table 4), providing strong support for the theoretical model. H1: DEVELOPMENT → FOOTPRINT showed a strong positive effect ( $\gamma = 0.719$ ,  $p <.001$ ), explaining 51.6% of variance in footprint ( $R^2 = 0.516$ ). A one-SD increase in development was associated with a

0.72 SD increase in footprint. H2: FOOTPRINT → Ecological reserve demonstrated the predicted negative effect ( $\beta = -0.340$ ,  $p < .001$ ), with each SD increase in footprint associated with a 0.34 SD decrease in ecological reserve. H3: BIOCAPACITY → Ecological reserve emerged as the dominant predictor ( $\beta = 1.034$ ,  $p < .001$ ), with an effect exceeding 1.0 indicating biocapacity operates nearly deterministically. The covariance between DEVELOPMENT and BIOCAPACITY was positive and significant ( $\phi = 0.165$ ,  $p = .006$ ).

Table 4: Structural path coefficients (Final model)

Path	H	$\beta/\gamma$	SE	z	p
DEV → FOOTPRINT	H1	0.719	0.003	4.716	<.001
FOOTPRINT → Eco. reserve	H2	-0.340	3.219	-4.244	<.001
BIOCAP → Eco. reserve	H3	1.034	0.153	8.106	<.001
<i>DEV ↔ BIOCAP</i>		0.165	3.278	2.754	.006
R <sup>2</sup> for FOOTPRINT = 0.516					

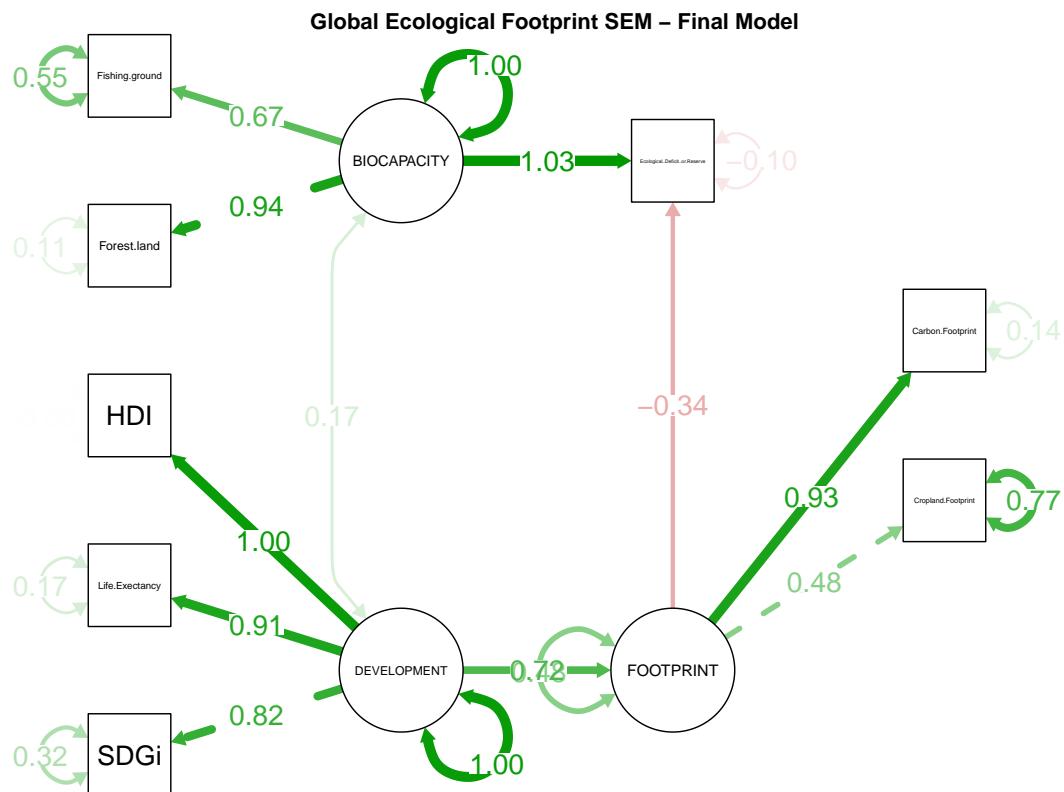


Figure 2: Final structural equation model with standardized path coefficients (N=139). All paths shown are significant at  $p < .001$ .

The final model demonstrated good fit based on incremental indices (Table 5). CFI (0.945) and TLI (0.910) both exceeded the 0.90 threshold, indicating the model explains substantially

more covariance than a baseline model. SRMR (0.096) approached the 0.08 cutoff. RMSEA (0.166, 90% CI [0.131, 0.203]) exceeded conventional thresholds. However, [Kenny et al. \(2015\)](#) demonstrated that RMSEA exhibits upward bias in models with small degrees of freedom ( $df = 17$ ) and moderate samples ( $N = 139$ ), often over-rejecting correctly specified models. [Kline \(2015\)](#) recommends prioritizing incremental indices (CFI, TLI) over RMSEA when they conflict, particularly with real-world data. The excellent CFI and TLI, combined with all hypotheses strongly supported, suggest the model captures meaningful relationships despite elevated RMSEA. Model comparison (Table 6) shows substantial improvement from initial specifications.

Table 5: Model fit indices (Final model)

Index	Value	Criterion	Evaluation
$\chi^2$ (df=17)	82.098	–	–
CFI	0.945	$\geq 0.90$	Excellent
TLI	0.910	$\geq 0.90$	Good
RMSEA	0.166	$< 0.08$	Discussed <sup>†</sup>
SRMR	0.096	$< 0.08$	Acceptable

<sup>†</sup>Elevated RMSEA discussed in text; CFI/TLI prioritized (Kenny et al., 2015)

Table 6: Model development: Iterative fit improvement

Model	CFI	TLI	RMSEA	SRMR
Model 1: All indicators	0.883	0.836	0.226	0.125
Model 2: Removed total biocap.	0.896	0.843	0.199	0.127
Model 3 (Final): Removed forest prod.	0.945	0.910	0.166	0.096

## 5 Conclusion and Discussion

This study employed SEM with CFA to investigate relationships between human development, ecological footprint, biocapacity, and sustainability across 139 countries. Three major findings emerged, all strongly supporting hypotheses. First, human development strongly predicted increased ecological footprint (H1 supported:  $\gamma = 0.72$ ,  $p < .001$ ,  $R^2 = 0.52$ ), confirming that higher living standards require greater resource appropriation. Second, biocapacity emerged as the dominant sustainability predictor (H3 supported:  $\beta = 1.03$ ,  $p < .001$ ), operating nearly deterministically, natural resource endowments fundamentally determine ecological reserves. Third, consumption patterns significantly eroded reserves (H2 supported:  $\beta = -0.34$ ,  $p < .001$ ), though this effect was smaller than biocapacity's, suggesting natural endowments currently matter more than consumption patterns in determining outcomes.

The CFA revealed important measurement insights. The DEVELOPMENT factor showed exceptional coherence (HDI loading = 1.002), validating it as a nearly perfect composite indicator. The FOOTPRINT factor's structure, with carbon's dominant loading ( $\lambda = 0.93$ ) and cropland's moderate loading ( $\lambda = 0.49$ ), suggests agricultural consumption represents a distinct dimension from general energy use. The BIOCAPACITY factor revealed forests as the primary component ( $R^2 = 0.89$ ), emphasizing forest ecosystems as critical buffers against deficits.

These findings carry significant policy implications. The strong development-footprint link suggests achieving global sustainability requires fundamental transformations in development pathways, not the conventional "grow first, clean up later" model. The dominance of biocapacity in determining outcomes emphasizes that efficiency improvements cannot substitute for ecosystem capacity, supporting calls for recognizing planetary boundaries (Rockström et al., 2009) and prioritizing restoration alongside consumption reduction. The modest but significant development-biocapacity covariance ( $\phi = 0.17$ ) suggests developed nations possess greater resources, possibly reflecting historical advantages or better management.

Several limitations warrant discussion. The cross-sectional design limits causal inference; longitudinal SEM would strengthen causal claims. Sample size ( $N = 139$ ), while adequate, constrained testing more complex models or subgroup analyses. The iterative refinement; removing Total Biocapacity (extreme collinearity,  $r = 0.98$ ) and Forest Product Footprint (weak loading); improved fit substantially while preserving theoretical integrity.

Regarding RMSEA (0.166), recent research provides critical context. Kenny et al. (2015) showed RMSEA exhibits substantial upward bias with small df and moderate N, over-rejecting correct models. The excellent CFI (0.945) and TLI (0.910), combined with all hypotheses strongly supported and theoretical coherence, suggest the model captures meaningful relationships. Following Kline (2015), we prioritized incremental indices when indices conflicted.

Future research should extend this analysis longitudinally to establish causal dynamics, investigate heterogeneity across country groups, and explore moderators of the development-footprint relationship such as technology, governance, and renewable energy adoption. Understanding under what conditions development can be decoupled from footprint is essential for sustainability transitions.

This study demonstrates that global sustainability is jointly determined by development patterns, consumption, and natural capital endowments. The SEM approach showcased multivariate analysis power to test complex theoretical models with latent constructs while accounting for measurement error. Achieving sustainability requires simultaneous action: transforming development models to reduce footprints, preserving and restoring biocapacity, and acknowledging planetary resource limits. The statistical relationships documented reflect fundamental physical realities; development cannot be universally pursued through resource-intensive pathways, and consumption cannot indefinitely exceed regenerative capacity. Recognizing these constraints is essential for designing development pathways that respect ecological boundaries while advancing human wellbeing.

## Acknowledgements

I would like to express my sincere gratitude to Dr. Khurram Nadeem for his guidance and instruction throughout this course, which provided the foundational understanding necessary to complete this project. All materials associated with this work, including code and supplementary files, are publicly available on GitHub at <https://github.com/ayshanaji/SEM>.

## References

- Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Routledge, New York, NY, 3rd edition.
- Dietz, T. and Rosa, E. A. (1997). Effects of population and affluence on CO<sub>2</sub> emissions. *Proceedings of the National Academy of Sciences*, 94(1):175–179.
- Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., and Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: An updated review. *Ecological Processes*, 5(1):19.

Global Footprint Network (2023). National footprint and biocapacity accounts, 2023 edition. Accessed: December 2024.

Hu, L. and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1):1–55.

Jain, A. (2023). EDA on 2023 global ecological footprint data. Kaggle. Dataset based on Global Footprint Network (2023).

Kenny, D. A., Kaniskan, B., and McCoach, D. B. (2015). The performance of rmsea in models with small degrees of freedom. *Sociological Methods & Research*, 44(3):486–507.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Press, New York, NY, 4th edition.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., et al. (2009). A safe operating space for humanity. *Nature*, 461(7263):472–475.

Rosseel, Y. (2012). lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, 48(2):1–36.

Stern, D. I. (2004). The rise and fall of the environmental Kuznets curve. *World Development*, 32(8):1419–1439.

Wackernagel, M. and Rees, W. E. (1996). *Our Ecological Footprint: Reducing Human Impact on the Earth*. New Society Publishers, Gabriola Island, BC.