Structural Equation Modeling as an Alternate Method for Estimating Variations in Fish Stock Abundance

by

Jamil Hayward-Kabani

B.Sc., University of British Columbia, 2010

Research Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Resource Management

in the
School of Resource and Environmental Management
Faculty of Environment

© Jamil Hayward-Kabani 2015

SIMON FRASER UNIVERSITY

Fall 2015

Approval

Name:	**Jamil Hayward-Kabani**
Degree:	**Master of Resource Management**
Report No.:	**631**
Title:	***Structural Equation Modeling as an Alternate Method for Estimating Variations in Fish Stock Abundance***
Examining Committee:	**Chair:** Neil Ladell
	Doctor of Philosophy Candidate

Andy Cooper
Senior Supervisor
Associate Professor

Anne Salomon
Supervisor
Associate Professor

Date Defended/Approved: November 25, 2015

Abstract

Using Indian Ocean swordfish-specific longline data, this paper explores some of the possible deficiencies of the most commonly used method (GLM method) for estimating an index of abundance and then answers two questions: 1) When estimating an index of abundance, can a structural equation model provide a viable alternative to the common GLM method? 2) How do the estimates derived from the GLM method and SEM method contrast? We discover that, at least for this data set, SEM-based methods consistently produce estimates for abundance that are significantly different from those produced by GLM-based methods. Considering the fundamental importance currently ascribed to the GLM-based methods for fisheries management, it is argued that further investigation of SEM-based methodologies is of high priority.

Acknowledgements

I would like to thank my supervisors, Dr. Andrew Cooper and Dr. Anne Salomon, for their encouragement and support throughout my master's program at SFU. In particular, I would like to thank Andrew for his invaluable guidance as I navigated unforeseen challenges with my research.

Table of Contents

List of Tables

List of Figures

1 Introduction

When making management decisions, a fisheries stock assessment is one of the key pieces of information utilized by a fisheries manager. Stock assessments describe the past status of a fish stock and make predictions about how a stock will respond to current and future management measures. With so much riding on these assessments, it is crucial that they be as accurate as possible. This requires the ongoing evolution of fisheries modeling techniques in the pursuit of increasingly accurate stock assessments.

The process of stock assessment typically involves the use of population dynamics models, which function to estimate stock abundance through time (Maunder, 2001). In addition to specific biological parameters (e.g. length, age, reproduction) unique to the population dynamics model being used, these models generally include an index of abundance and data on removals due to harvesting (Hilborn & Walters, 2013).

An index of abundance is a relative measure of the population size for a fish species. It can be applied to the whole population of a species or a sub-unit of the population and is generally calculated using the number (or weight) of fish caught per standard unit of fishing effort (e.g. hooks, tows, days fishing) (Maunder & Punt, 2004). The catch per unit effort (CPUE) can be calculated based on fishery independent data such as scientific surveys or, as it is most commonly done, it can be calculated using fishery dependent data (catch and effort recorded by a fishery) (Ye & Dennis, 2009). At small spatial scales it is generally assumed that CPUE is related to density (N) through a constant of proportionality:

$$CPUE = q\mathrm{N} \tag{1}$$

Where q is known as the catchability coefficient (the fraction of the population that is captured by one unit of effort). Equation (1) can be generalized to an entire population in which case N represents the population size. In order to use this relationship it is important that CPUE be first standardized over time in order to correct for systemic factors influencing the data such as fishing area, season, vessel type, and gear type. This process is referred to as CPUE standardization and without this process it is difficult to know whether changes in CPUE are indicative of changes in abundance or are due to other factors.

Due to its importance in the stock assessment process, CPUE standardization has been the focus of significant academic study over the last 30 years. During this time, many different methods for standardization have evolved. When choosing a method, the choice should be based on an evaluation of the underlying assumptions of the models and the use of appropriate statistical tests and diagnostics (Maunder & Punt, 2004). An understanding of the fishery dynamics being modeled can also provide insight into which method should be used.

Generalized linear models (GLMs) are the most commonly used method for standardizing catch and effort data (Maunder & Punt, 2004). The popularity of GLMs is due to the power of these models, their relative ease of use, and their familiarity to most scientists working in fisheries and the biological sciences (Maunder & Punt, 2004; Myers et al., 2012). However, GLMs as a method to standardize CPUE have a number of weaknesses and many of the alternative methods for standardization exist to address one shortcoming or another (Glaser et al., 2011; Maunder, 2001; Maunder & Punt, 2004; Venables & Dichmont, 2004). For example, many fishery systems are inherently nonlinear, but linearity is a key assumption of GLMs, so a generalized additive model (GAM), which is better suited to handle nonlinear relationships between catch rate and potential independent variables, may be more appropriate. Despite there often being a more nuanced standardization option available to fisheries scientists, GLMs continue to be widely applied to CPUE standardization due to the factors listed above.

This paper further investigates two potentially problematic assumptions made during the standard process of developing an index of abundance using GLM-based methods for standardizing CPUE and then explores, through the use of a case study, an alternate approach for estimating abundance based on Structural Equation Modeling (SEM).

The first potentially problematic assumption is the presupposition of an exogenous relationship between the nominally "independent" observables and the nominally "dependent" variable, CPUE. A prerequisite for any GLM is that the independent variables must be exogenous (i.e., there can be no causal influence of a dependent variable on an independent variable) (Breslow, 1996). For example, fishing equipment can affect the CPUE, but the resulting CPUE must not affect fishers' decisions about which equipment to use and/or purchase. Similarly, fishing area can affect CPUE, but the CPUE should not influence the fishing area in which effort is expended. One can imagine scenarios in which the assumption of exogeneity could be called into question. For instance, fishers may be inclined to utilize more expensive equipment when CPUE is high and less expensive equipment when it is low. If, in fact, the assumption of exogeneity does not hold for all independent variables, estimates obtained through a GLM-based methodology may exhibit an endogeneity bias (Hicks, 2013; Marchal et al., 2006).

A second problem with the standard process of developing an index of abundance using GLM-based methods for standardizing CPUE is that this process is rooted in the assumption that CPUE is proportional to fish density, N, through equation (1). When referring to the CPUE of a fleet fishing uniformly over the entire range of a particular species, this implies an assumption of proportionality between CPUE and stock abundance.

In fact, CPUE may not be proportional to abundance. Even if we assume that equation (1) is valid, fishermen are incented to expend effort in regions with high fish density. If a fleet is able to identify such regions (e.g., through improved intra-fleet communication) and expend disproportionate effort in those regions, it is possible for

3

CPUE to remain high even as the overall stock is depleted. Simulation studies confirm such "hyper-stability" when communication in a fleet is modeled (Gaertner & Dreyfus-Leon, 2004). This phenomenon can be seen in many fisheries but is most evident for schooling fisheries. Paloheimo and Dickie (1964) have stressed the importance of understanding both the spatial distribution of fish and the spatial distribution fishing effort when interpreting CPUE data. Whether standardized commercial CPUE should be used as an index of abundance, is a topic of debate in the fisheries literature (Gaertner & Leon, 2004; Harley et al., 2001; Marchal et al., 2007; Richards & Schnute, 1986). Nevertheless, despite these well-documented shortcomings, CPUE remains a simple and attractive index of abundance and is commonly used for stock assessment by fishery agencies around the world.

An alternate approach to estimating abundance may be available through Structural Equation Modeling (SEM). SEM deploys a fundamentally different modeling paradigm than GLM-based methods and, notably, does not suffer from the two theoretical shortcomings listed above. SEMs can incorporate endogeneity between observables within any given model. Indeed, SEMs not only allow for endogenous relationships between modeled observables, they provide a theoretical framework by which to estimate endogeneity quantitatively. SEMs also allow for the estimation of latent observables, such as fish stock abundance, directly without the need for any assumption of proportionality to CPUE.

A SEM is designed to examine a set of relationships between independent observables (often called exogenous in SEM literature) and dependent (endogenous) variables. Endogenous variables can be either measured (directly observed), or latent (not directly observed) (Ullman, 2006). SEMs are generally validated through confirmatory factor analysis (CFA). In such cases, the causal pattern of inter-variable relationships within the theory is specified based on prior experience with the modeled system. The SEM then enables confirmation through multivariate analysis. The goal is to determine whether newly collected data is consistent with a hypothesized model. Consistency is evaluated through model-data fit, which indicates the extent to which the postulated network of relations among variables is plausible (Lei & Wu, 2007).

It is also possible to arrive at a SEM through exploratory factor analysis (EFA). With EFA, several plausible models may be compared and evaluated for quality of fit with the data set. However, when a model is derived, even partly, through EFA (as we shall do in this paper), much caution must be taken to validate the model with an independent data set and thereby demonstrate that the quality of its fit cannot be ascribed to over-tuning.

A further important difference between GLMs and SEM is worthy of note, and, in fact, will be of central importance in the analysis we perform in this paper. GLM-based methodologies seek to identify linear correlations between independent and dependent variables through a purely statistical methodology. No presumption of causation is implied or even relevant to the process. By contrast, a SEM is an expression of presumed causal relationships within the system being modeled. A SEM is therefore, fundamentally an expression of theoretical considerations deduced through a process not grounded in the data set. As a consequence, a SEM will often be framed in terms of observables that would not normally be selected for a GLM-based approach. For example, whereas with a GLM one might seek to remove systemic "area" effects with arbitrarily selected regions of the ocean, with a SEM one would attempt to define an observable for the fishing area that is sensitive to known regional dynamics of the fishery.

2 Materials and Methods

2.1 Data Set

This study utilizes publicly available longline data from the Indian Ocean Tuna Commission (IOTC). The original data set can be found at: http://www.iotc.org/English/data/databases.php.

In order to better associate effort with catch for a specific species, we decided to exclude all but data for which "Gear Type" is identified as "ELL" (i.e., swordfish specific longline). Effort expended with this gear has swordfish as the targeted species. Acting on this premise and the fact that swordfish catch accounted for between 50-70% of the total catch for other European fleets, we associated all effort with the recorded swordfish catch in our models.

The data was further filtered exclusively for swordfish specific longline gear deployed by the Spanish fleet. This provides three advantages. First, unlike several other fleets, the Spanish fleet consistently had a reporting quality ranking of 3 (i.e., good). These quality scores (on a scale of 0-3) indicate the IOTC's confidence that the data represents the effort and catch in the stratum concerned. Second, unlike some other fleets, throughout the study period, the Spanish fleet consistently reported swordfish catch using a single metric (i.e., tonnage as opposed to number of fish caught). This allowed us to avoid having to posit a methodology for converting between fish number and tonnage for the study period. Third, throughout the study period, the Spanish fleet had consistently higher than average effort for swordfish longline. In fact,

the Spanish fleet alone accounted for 15% of total swordfish catch and 40% of catch with a reporting quality ranking of 2 or greater.

It should be noted that, due largely to its public and international nature, this dataset does suffer from some important deficiencies. First, the data set is not vessel specific. As a consequence, our models account for dynamics and systemic effects at the fleet level only. Second, effort and catch are aggregated into relatively large grid cells (i.e., 5° latitude by 5° longitude) and into month-long intervals. Correspondingly, our analysis attempts to model only macro-dynamics of the Spanish fleet. Third, in the first eight years of the study period, effort as recorded by month and grid cell for the Spanish fleet was relatively sparse.

Acknowledging these deficiencies, the goal of this paper is not to deduce specific results for swordfish abundance, but rather to contrast paradigmatic elements of GLM and SEM-based methodologies, therefore specifics of our selected data set are ultimately of lesser significance.

2.2 GLM-based Method

Adopting common practice for evaluating fish stock abundance indices (Maunder et al., 2004; Maunder et al., 2006) and, in particular, for swordfish in the Indian Ocean (Kolody et al., 2010; Mejuto et al., 2013; Uozumi, 1998; Wang & Nishida, 2011) we standardized CPUE through a GLM. Using the software "R" and the steps detailed below, we compared GLMs to find the most appropriate model (Richards et al. 2011; Wagenmakers & Farrell, 2004):

1. We identified all possible predictor variables. Three predictor variables were reported in the source data set and were suitable for use in the GLM:
   a. Year - A categorical variable ranging from 1993-2011. The variable was converted to a factor.

b. Season – The year was divided into four quarters: January-March, April-June, July-September, and October-December. The variable is categorical and was converted to a factor.

c. Area – In the dataset fishing area was divided into 5˚ by 5˚ grid cells. We regrouped area into fours quadrants that cover contiguous areas:

   o *Southeast*: 20˚-45˚ south and 30˚-70˚ east.
   o *Southwest*: 20˚-45˚ south and 70˚-105˚ east.
   o *Northeast*: 20˚ south-10˚ north and 30˚-70˚ east.
   o *Northwest*: 20˚ south -10˚ north and 70˚-105˚ east.

   The variable is categorical and was converted to a factor.

2. We fit all combinations of predictor variables to the natural log of CPUE. We further investigated any interactions between the predictor variables. Interactions involving the year effect would invalidate the year effect as an index of abundance; therefore we did not investigate any possible interactions including the year variable (Hinton and Maunder, 2003). In total, 9 distinct models were evaluated.

3. We calculated AICs and AIC weights for each model. The preferred GLM was determined by selecting the model with the lowest AIC value and by following the procedure for model selection using Akaike weights outlined by Wagenmakers and Farell (2004). We further calculated residual plots, QQ plots, and Cook's Distance plots for each model. Plots for the preferred GLM are discussed in further detail in Appendix A. The plots suggest no significant violation of standard GLM statistical assumptions.

4. The preferred model's resulting coefficients for the factor "Year", were used to derive the nominal and standardized CPUE indices using equations (2) and (3) below:

$$\text{Nominal CPUE Index}_i = \left(\frac{\text{Catch}}{\text{Effort}}\right)_i - \left(\frac{\text{Catch}}{\text{Effort}}\right)_0 \qquad (2)$$

$$\text{Standardized CPUE Index}_i = e^{\beta_{(\text{Year}_1)} - \beta_{(\text{Year}_0)}} \qquad (3)$$

Where $\beta_{(Year_1)}$ is the coefficient for the current year and $\beta_{(Year_0)}$ is the coefficient for the base year (1993).

2.3 SEM-based Method

SEM analysis is categorized as either exploratory or confirmatory (Lei & Wu, 2007). To our knowledge, no SEM has yet been constructed that predicts Swordfish abundance, so SEM analysis conducted here is necessarily exploratory. Correspondingly, the model developed here should be viewed with caution. As with all exploratory models, ours must be validated with independent data sets before it can be viewed as a legitimate foundation for prediction. Having said this, the focus of this paper is not to defend any particular model nor any estimates it might generate, but rather to address foundational questions about SEM and GLM-based methodologies for estimating fish stock abundance.

2.3.1 Definition of Observables

It is neither necessary nor even desirable that GLM and SEM models reference a common set of observables. Unlike a GLM, a SEM is rooted in a theoretical model of causation. Observables for the SEM should be selected to best isolate causal relationships within the system being modeled. It is important that the selected observables lead to a model that is both simple and predictive. Meanwhile, when selecting variables for GLM-based analysis, we have opted to mimic methodologies most widely used in standard CPUE studies. Typically, these are selected without sensitivity to the dynamics of the modeled system. For instance, fishing area is most often treated as a categorical variable, grouped into contiguous regions of ocean having approximately comparable size (Garcia-Cortez et al., 2012; Mejuto et al., 2013; Wang & Nishida, 2011). Similarly, fishing season is most frequently decomposed into contiguous three month intervals according to the Julian calendar. .

When defining SEM observables, we initially thought to isolate factors that would determine whether, where, when, and how intensively each captain would fish. An immediate challenge we confronted, however, is that the source data set provides no information that can identify individual captains or vessels. A further challenge is that data is not available for the Spanish fleet in every grid cell for every month or for that matter, for every year. Therefore, in our exploratory factor analysis, we have attempted to identify observables that illuminate causal relations evident in the macro-dynamics of the entire Spanish fleet. We identified the following observables as worthy of exploration:

Observables:

1. $\text{Catch}_{T,X}$: Swordfish catch in grid cell X during month T as measured in metric tonnes.
2. $\text{Effort}_{T,X}$: Effort expended in grid cell X during month T as measured in number of fish hooks.
3. $\text{Abundance}_{T,X}$: A latent variable indicating the abundance in grid cell X during month T.
4. Seasonality_T: Catch for the entire study region (i.e., the Indian Ocean) during the month T of a given calendar year normalized by the maximum catch for any month during that same year. This variable is intended to provide a measure of annual cyclicality of the fishery. It should be noted, however, that considerable change occurred over the course of the nineteen years covered by the data set. It is for this reason that we opted to normalize relative to the current calendar year rather than relative to a mean measure for the entire study period.
5. Locality_X: Catch in grid cell X for the entire calendar year normalized by the maximum catch for the same year in any one grid cell. This variable is intended to provide a measure for the spatial distribution patterns of swordfish catch in the Indian Ocean. Again, we note that the distribution of catch and effort changed considerably over the course of the nineteen years in the study period. Much of this change was likely influenced by factors other than abundance (e.g., piracy, exploration, weather, convenience, etc.). Therefore, we again opted to normalize

with reference to a maximum for the current calendar year rather than with reference to a single measure defined for the study period as a whole.

6. $\text{Catch}_{(T-1M)}$: Catch for the Spanish fleet over the entire Indian Ocean during the past calendar month.

7. $\text{Catch}_{(T-1Y)}$: Catch for the Spanish fleet over the entire Indian Ocean during the past calendar year.

8. $\text{CPUE}_{(T-1M)}$: Nominal CPUE for the Spanish fleet over the entire Indian Ocean during the past month.

9. $\text{CPUE}_{(T-1Y)}$: Nominal CPUE for the Spanish fleet over the entire Indian Ocean during the past calendar year.

10. $\text{Seasonality}_{(T-1Y)}$: Catch for the entire Indian Ocean during the same calendar month 1 year ago normalized by the maximum catch for any month during that same calendar year.

11. $\text{Locality}_{(T-1Y,X)}$: Catch in grid cell "X" for the entire previous calendar year normalized by the maximum catch for the same year in any one grid cell.

2.3.2 Model Specification

A priori we identified the following causal relationships as worthy of exploration:

1. $[\text{Catch}_{T,X} \leftrightarrow \text{Effort}_{T,X}]$: We postulated that changes in Effort result in correlated changes in Catch. Conversely, we postulated that changes in Catch lead to correlated changes in Effort as the fleet attempts to maximize overall return. This is an example of an endogenous relationship that cannot be modeled by standard GLM-based analysis.

2. $[\text{Abundance}_{T,X} \rightarrow \text{Catch}_{T,X}]$: We postulated that a change in abundance within grid cell X during month T results in correlated changes in catch within the same grid cell and time period.

3. $[\text{Abundance}_{T,X} \rightarrow \text{Locality}_X]$ and $[\text{Abundance}_{T,X} \rightarrow \text{Seasonality}_T]$: We postulated that changes in abundance in grid cell X during month T give rise to correlated changes in both normalized catch for that grid cell and normalized catch for that month.

4. $\left[\text{Catch}_{(T-1M)} \rightarrow \text{Effort}_{T,X}\right]$ and $\left[\text{Catch}_{(T-1Y)} \rightarrow \text{Effort}_{T,X}\right]$: We postulated that changed catch during month T results in a correlated change in effort both in the following month and in the same month of the immediately following year.

5. $\left[\text{Catch}_{(T-1M)} \rightarrow \text{Abundance}_{T,X}\right]$ and $\left[\text{Catch}_{(T-1Y)} \rightarrow \text{Abundance}_{T,X}\right]$: We postulated that changes to total catch throughout the Indian Ocean in a given calendar month results in anti-correlated changes in abundance in each grid cell both in the immediately following calendar month and in the following calendar year.

6. $\left[\text{Seasonality}_{(T-1Y)} \rightarrow \text{Effort}_{T,X}\right]$: We postulated that changes in catch throughout the Indian Ocean during a given calendar month result in correlated changes in effort in any given grid cell the same calendar month of the immediately following year.

7. $\left[\text{Locality}_{(X,T-1Y)} \rightarrow \text{Effort}_{T,X}\right]$: We postulated that changes in catch within a given grid cell and calendar month result in correlated changes in effort in the same grid cell and month of the immediately following year.

8. $\left[\text{CPUE}_{(T-1Y)} \rightarrow \text{Effort}_{T,X}\right]$ and $\left[\text{CPUE}_{(T-1M)} \rightarrow \text{Effort}_{T,X}\right]$: We postulated that changes in CPUE throughout the Indian Ocean within a given month result in correlated changes in effort in each grid cell within the following calendar month and in the same month of the immediately following calendar year.

These relationships result in the initial exploratory SEM path diagram shown in Figure 1.

Figure 1. Initial exploratory SEM for Spanish swordfish fishery in the Indian Ocean

2.3.3 Model Identification

Before a model is estimated, it must be properly identified. A model is identified if there is a unique numerical solution for each of the parameters in the model (Ullman, 2006). In general, each model in the study should be over justified (enough independently measured variables to estimate the parameters), have the variance (or one of the coefficients) of the latent variable fixed at one, and have at least 3 indicators for the latent variable. A discussion of the model identification process used in this study can be found in Ullman (2006) and Bollen (1989).

2.3.4 Criteria and Methods for Model Evaluation

There is no universally accepted methodology for exploratory SEM model evaluation - the process of refining and selecting a SEM that best represents the system

under study. The process can involve some judgment calls, so not all researchers will arrive at the same exact model given the same data. Grace (2006) and Hooper et al. (2008) offer excellent reviews of best practices for SEM model evaluation.

In this study, our EFA process was governed by three major practices. First, we evaluated each distinct SEM model relative to the following four criteria:

1. Overall fit of the model as indicated through Chi Square tests, Root Mean Square Error of Approximation (RMSEA) values, and Standardized Root Mean Square Residual (SRMR) values,
2. Individual parameter significance as indicated through p-values,
3. Modification indices indicating changes in Chi Square values resulting from freeing fixed parameters,
4. Theoretical considerations: to help avoid over-fitting to the source data set, changes to a model suggested by some other factors can only be implemented if there is a strong theoretical argument supporting the change.

Second, when the above criteria suggested several modifications to a given model, we made such modifications in isolation from each other while evaluating the fit of all resulting models. We followed this practice to avoid missing a preferred model that might otherwise be hidden through making several concurrent changes to an antecedent model (Grace, 2006).

Third, throughout our EFA we "trained" all models using only data from the first 15 years of the 19 year data set. We followed this practice so that the residual data could be used to provide evidence to help confirm or invalidate models selected through the exploratory modeling process. 15 years equates to approximately 75% percent of the data and also represents the minimum amount of data required to make the model run reliably.

3 Results

3.1 GLM Results

Table 1 displays the AIC values and AIC weights for each of the generalized linear models investigated. The following GLM had the lowest AIC value and was the only model that had a significant Akaike weight. All other AIC weights were << 0.01.

$$\log(\text{CPUE}) = \beta_0 + \beta_1(\text{Year}) + \beta_2(\text{Season}) + \beta_3(\text{Area}) + \beta_4(\text{Season} * \text{Area}) + \varepsilon \qquad (4)$$

Table 1.   **AIC values and AIC weights for the 9 GLMs examined**

Independent Variables for the Model	AIC Value	AIC Weight
Year	1572	4.58×10^{-31}
Area	1737	6.78×10^{-67}
Month	1735	2.38×10^{-66}
Year and Month	1518	3.03×10^{-19}
Year and Area	1539	6.84×10^{-24}
Area and Month	1690	1.07×10^{-56}
Area and Month with Interaction	1614	3.01×10^{-40}
Year, Area and Month	1488	8.16×10^{-13}
Year, Area and Month with Area/Month Interaction	1432	1.00

Figure 2 shows nominal and standardized CPUE indices as calculated using the GLM methodology of section 2.2. The gap at year 5 corresponds to an absence of data for the Spanish fleet in calendar year 1997.

Figure 2.   **Nominal versus standardized CPUE index obtained with the preferred GLM.**

3.2 SEM Results

The SEM in Figure 1 is properly identified and therefore we were able to estimate the model parameters, resulting in Figure 3. Figure 3 shows the path coefficients for this model as calculated using the LAVAAN package in *R*. Detailed results for path

coefficients, variances, and fit measures for Model 1 and all other models evaluated in this study are shown in Appendix B.

Figure 3.   **Initial SEM with path coefficients and error estimates.**

Upon estimation of Model 1 it was determined that the model had a poor overall fit to the data ($\chi^2 = 427$ p-value = 0.00; RMSEA: 0.170 p-value= 0.00; SRMR: 0.082).[1] The decision metrics associated with the modification indices did not indicate that any new relationships should be added to the model. However, there were 5 relationships with high p-values and relatively low path coefficient values (0.00, 0.00, -0.05, 0.02, and 0.00). After re-examining the 5 relationships from a theoretical perspective, in spite of our a priori expectations, it is plausible that current effort within any given grid cell is not significantly influenced by catch or CPUE throughout the Indian Ocean in the previous month or in the corresponding month of the previous year. Similarly, given that most studies conclude that swordfish catch was below the Maximum Sustainable Yield (MSY)

---

[1] Note that with EFA for SEM the null hypothesis is that the model fits the data, so larger p-values confirm the fit. Common practice is to accept p-values > 0.1 as confirmation of a good model fit.

for most of the study period (IOTC Swordfish report, 2014), it is plausible that current abundance within any given grid cell is not significantly affected by catch throughout the Indian Ocean in the previous month or in the corresponding month of the previous year.

The above considerations caused us to evaluate several models resulting from removal of different combinations of the five suspect relationships in Model 1. Following the procedures detailed in section 2.3.4, we determined that all five of the suspect relationships should indeed be removed. We labeled the resulting SEM as "Model 2". Figure 4 shows the path coefficients estimated through LAVAAN for this model.

Figure 4. SEM for "Model 2" obtained by removing five suspect relationships from Model 1.

While the fit of Model 2 improves relative to Model 1, it is still poor (χ^2 = 190 p-value = 0.00; RMSEA: 0.147 p-value= 0.00; SRMR: 0.075). Using the methodology detailed in section 2.3.4, we identified seven additional models (models 3-9) to evaluate, each with at least one relationship from Model 2 removed. Results of this analysis are detailed in Appendix B.

From this new set of models, "Model 9" shown in Figure 5 is preferred.

Model 9

1

Abundance$_{T,X}$

.39 Seasonality$_T$ — .85

.51

.70 Locality$_X$ — .74

Effort$_{T,X}$ —.17— Catch$_{T,X}$
            —.56—

.57   .31

Figure 5.   SEM for Model 9 obtained by removing 3 relationships from Model 2.

This new model has three more relationships removed (Locality$_{(X,T-1Y)}$ → Effort$_{T,X}$, Catch$_{(T-1Y)}$ → Effort$_{T,X}$, and CPUE$_{(T-1Y)}$ → Effort$_{T,X}$), all of which had low or negative path coefficients in Model 2.

Removal of these three relationships to produce Model 9 is theoretically consistent with the removal of the five relationships from Model 1. The model suggests that CPUE, Locality, Seasonality, and Catch results from the immediately preceding month and the corresponding month of the previous year do not significantly affect current Effort or Abundance in a given grid cell. Model 9 is preferred because it is the most parsimonious model and it has a good overall fit: χ^2 = 0.008 p-value = 0.93; RMSEA: 0.000 p-value= 0.98; SRMR: 0.001.

To validate the EFA above, we re-trained Model 9 and all other Models using all data from the entire 19 year study. When trained with the entire data set, path coefficients and variances can change by up to ~10% for each model. Nevertheless, the overall preferred fit for Model 9 (as well as the respective deficiencies of the other models) remains unchanged. When fit with the entire 19 year data set, Model 9 path coefficients and variances are as shown in Figure 6.

Figure 6.   **Path coefficients and variances for Model 9 when fit using all 19 years of data.**

Estimates for variation in abundance are shown in Figure 7. The figure shows curves derived from Model 2 through Model 9 as trained on the entire 19 year data set. All estimates are in units of standard deviation and are obtained using the "predict" function of LAVAAN (Muthen, 2004 p.47). Excluding 1997 (for which there is no data), estimates for all models agree to within an average of 0.042 standard deviations (1.05% relative to the total range for Model 9). The greatest disagreement between the models

occurs in the year 2000, where the average deviation from the estimate obtained through Model 9 is 0.292 standard deviations (7.30% relative to the total range for Model 9).

Variation in Abundance for Models 2-9

Figure 7.   **Estimates for variation in abundance derived from Model 2 through Model 9 trained on the entire 19 year data set.**

3.3 Comparison of GLM and SEM results

The "predict" function of LAVAAN produces estimates for variation of the Abundance latent observable in units of standard deviations from the mean. Comparing this abundance estimate to estimates for variation in Nominal and Standardized CPUE (expressed in units of standard deviations from the mean Nominal CPUE) produces Figure 8.

Variation of Nominal CPUE, Standardized CPUE and SEM Abundance

Figure 8. **Comparison of variations in Nominal CPUE and Standardized CPUE as derived through the GLM of section 2.2, and variations in the abundance latent observable derived through the SEM of section 2.3. Note that all variations are measured in units of standard deviation from the mean for the relevant observable.**

4 Discussion

4.1 GLM and Standardized CPUE

As discussed in section 2.2, the GLM for this study adopts common practices for correcting CPUE to account for systematic effects attributable to variations in the independent observables. Series values for nominal and standardized CPUE shown in Figure 2 are qualitatively similar. In fact, nominal and standardized CPUE agree within statistical error for 14 of the 18 years included in the study. This differs from many studies, which often show statistically significant differences between nominal and standardized CPUE (Bigelow et al., 1999; Nakano et al., 2005; Punt et al., 2000).

The unusual degree of similarity between nominal and standardized CPUE in this case could trace to the absence of information specific to vessel and captain in the source data set. After filtering by fleet and gear type, the only residual information available to further characterize catch/effort is spatial (i.e., latitude and longitude in 5 degree increments) and temporal (i.e., month of the year). As is common practice in GLM-based studies, our model associates this spatial/temporal data with categorical variables corresponding to contiguous regions of the ocean and seasons of the year. Since these categorical variables are defined without sensitivity to dynamics of the fishery (i.e., similarly sized sectors of the Southern Indian Ocean with arbitrary boundaries and quarters of the year defined relative to the Julian calendar), there is no reason to expect that systemic corrections to the nominal CPUE arising through these variables would be significant.

The GLM estimates for standardized CPUE (Figure 2) show a range of values more than twice the magnitude of the CPUE for the initial reference year. Beyond this high variability, the most obvious pattern is that a significant increase in standardized CPUE occurs in the year 2000 followed by relative stability thereafter. The average standardized CPUE index between 1993 and 2000 (excluding 1997, for which there is no data) is 0.88. Meanwhile for the years from 2001 to 2011 the average standardized CPUE index nearly doubles to 1.61. Assuming a constant catchability coefficient as per equation (1), the implication is that stock abundance nearly doubled after 2000.

However, alternate explanations for the increase in standardized CPUE after the year 2000 may be more plausible. Operations of the Spanish surface longline fleet in the Indian Ocean started in 1993. Before 2000, data was mostly obtained from surveys targeting swordfish in unknown fishing areas (Fernández-Costa et al., 2014). After this preliminary period, the Spanish fleet consolidated its operations and began specifically targeting swordfish in the Indian Ocean. The increase in CPUE post 2000 may, therefore, reflect maturation of techniques used within the fishery rather than any absolute increase in the abundance of the stock. Relatively low CPUEs between 1993 and 2000 may also have been impacted by high incidence of piracy in the south western Indian Ocean during these years (Santos et al., 2012).

4.2 SEM Discussion

We identified Model 9 as a preferred candidate SEM through an EFA process involving model specification, identification, and evaluation. Statistical considerations favour Model 9 over models 1-8. Principally, Model 9 is the most parsimonious, yet it still explains the data approximately equally to or better than the other models. Despite the good fit of Model 9 to the data, it is prudent that we retain some scepticism. As discussed in section 2.3, when performing EFA there is always the risk of over-fitting a model to a specific data set.

Although Model 9 certainly needs to be confirmed, it is unlikely that any over training of the model had a significant impact on the abundance variation estimates, as models 2-9 all produce very similar estimates (section 3.2). A comparison of the path coefficients in all 9 models shows that the magnitudes of the path coefficients were quite consistent. The relationships from the exogenous variables to effort and abundance (listed as relationships 4-8 in 2.3.2) had weak coefficients. At the same time, the relationships between catch and effort and between abundance and catch, seasonality and locality, all had fairly strong coefficients.

Model 9 should also be assessed relative to theoretical considerations. It includes 4 relationships:

1. $\left[\text{Catch}_{T,X} \leftrightarrow \text{Effort}_{T,X}\right]$: Within this endogenous relationship, catch has a relatively strong influence on effort (path coefficient of 0.59) while effort has a weaker influence on catch (path coefficient of 0.15). A plausible explanation for the high influence of catch on effort can be deduced from the simulations performed by Gaertner and Leon (2004). They demonstrated that exchange of information within a fishing fleet can cause effort to rapidly converge on areas of high catch/abundance. If the majority of fleet effort is influenced by such information exchanges, a pattern emerges that catch is a strong predictor of effort while the converse is less true. All of our SEM models are consistent with collaboration being an important factor determining the locality and temporality of effort within the Spanish fleet.

2. $\left[\text{Abundance}_{T,X} \rightarrow \text{Catch}_{T,X}\right]$: As expected, stock abundance within a particular grid cell and during particular month is a strong predictor (path coefficient of 0.80) of catch in the same grid cell and month.

3. $\left[\text{Abundance}_{T,X} \rightarrow \text{Locality}_X\right]$ As expected, abundance within a particular grid cell and month influences (with a path coefficient of 0.43) catch within that grid cell for the calendar year.

4. $\left[\text{Abundance}_{T,X} \rightarrow \text{Seasonality}_T\right]$ Correspondingly, abundance within a particular grid cell and month is a predictor (with a coefficient of 0.37) of catch for the entire study region during that month.

From the model evaluation process, we deduce that the selected exogenous variables have minimal influence on effort and abundance (see relationships 4-8 in section 2.3.2). This result is consistent with:

1. Intra-fleet communication dominating over historical factors as a determinant of effort.
2. The catch being under MSY through much of the study period.

The relatively high variances for the latent and observable parameters are expected given that there are a number of key exogenous variables that are not accessible to us in this data set (e.g., factors specific to individual vessels and captains, factors related to evolving fishing gear and methodology, stock migratory patterns, environmental effects, oceanic conditions, prey patterns, etc.). In general, these high variances are suggestive that our SEM could be improved if more detailed data about the system were provided.

Ultimately, Model 9 needs to be considered confirmed with independent validation using new data sets. For instance, data from another fleet in the Indian Ocean Swordfish fishery could be used to verify the model derived in this study. Partial confirmation of Model 9 was achieved by training it using only the first 15 years of the data set and then comparing the results against those obtained using the entire data set. Estimates for path coefficients and variances obtained with the partial and complete data sets agree to within ~10%.

4.3 Comparison of SEM and GLM Results

While there is good agreement in Figure 8 between Nominal and Standardized CPUE, the estimate for Abundance generated by the SEM, is markedly different. For instance, consider the period from 1997-2005. While the SEM abundance estimate peaks in 1999 (1.3 standard deviations from the mean), the standardized CPUE estimate drops to a minimum (-1.8 standard deviations from the mean). The SEM

abundance estimate then falls and levels out at approximately -0.5 standard deviations from the mean until 2005. Alternatively, during the same period, the standardized CPUE estimate rises sharply before levelling out at approximately 1 standard deviation from the mean. Outside 1997-2005, the two series appear to qualitatively agree. Although the values differ, both series peak in 1994-1995 and then both decrease from 2005 to 2006 before rising to their highest values in 2011. Despite the differences, both curves suggest an increase in abundance over the study period. However, in reference to studies conducted by Shono (2008) and Abeare (2009), where different CPUE standardization methods are compared, it is clear that the two data series do significantly differ, especially during the period of 1997-2005.

Given the very different premises for the GLM and SEM, it is not surprising that they should arrive at divergent estimates for variations in stock abundance. The predictions of each modeling paradigm are consistent with its own set of assumptions. Whereas the GLM based approach for estimating abundance presumes that abundance relates to Standardized CPUE via a constant of proportionality, the SEM implies that local abundance is a strong predictor of local catch, which, in turn, is a strong predictor of local effort.

It is natural to question whether any independent evidence (e.g. a research survey) is available on swordfish abundance in the Indian Ocean during the study period that would validate either of the models under investigation here. Unfortunately, all current abundance estimates for this stock appear to use a standardized CPUE as part of their population dynamics models. However, one factor to consider is that the Nominal CPUEs for different fleets during the study period differ significantly one to another (IOTC Swordfish report, 2014). Noting that reasonably close alignment is expected between Standardized and Nominal CPUE when using GLM-based methods with this data set (for the reasons discussed in section 4.1), these differing fleet CPUEs raise doubt about whether the assumption of proportionality between Standardized CPUE and stock abundance can hold true in this case. Of course, if more detailed data were available (e.g., on variances of fishing gear, vessel types, etc.), the differences between the Nominal CPUEs recorded by the different fleets might still be explained through GLM-based methods.

4.4 Policy Implications

Acknowledging that the work in this paper requires independent validation, it is important to recognize the potential policy implications that these results may have. Stock assessments are essential pieces of information that not only help fisheries managers protect the ocean's resources, but also help managers to make tough choices when balancing the social, economic, and environmental interests of the stakeholders involved. When considering stock assessments in conjunction with other information, fisheries managers have many policy options available to them. These can include anything from government regulations (like quotas, licenses, and season openings) to educational materials. Resultant fisheries policy decisions can impact our society and environment in many ways, both directly (fishing jobs, processing jobs, subsistence, species health, etc) and indirectly (tourism, ecosystem health, general economy, drilling and mining activities, shipping routes, etc).

An index of abundance is a critical component of a stock assessment and inaccuracies in the index can lead to inaccuracies in the assessment (Walters & Maguire, 1996). This occurrence can sometimes result in grave consequences. For example, after investigating the northern cod collapse Walters and Maguire (1996) concluded that: 1) There were inaccuracies in the indices of abundance because commercial CPUE was used as an index of abundance, leading to stock size over-estimation. 2) Stock assessment errors likely contributed to overfishing by creating optimistic long-term forecasts, which lead to total allowable catches being set higher than they should have been. The collapse of the Atlantic northern cod stock and simultaneous closing of the fishery has resulted in long lasting impacts to both the environment and livelihoods of people that depended on the fishery (Walters & Maguire, 1996).

Our study suggests that the index of abundance estimated using the SEM differs significantly from the index obtained using the GLM. This could potentially have far reaching consequences if it is discovered that SEMs do indeed give a more accurate index of abundance. However, it should be cautioned that this study was performed only

using data for the Spanish longline Swordfish fleet. There needs to be significant additional research in order to determine: 1) whether the results from this study can be confirmed 2) whether this work can have a larger applicability 3) whether the SEM is a preferred tool over a GLM for estimating an index of abundance. It should also be noted that there are many alternative CPUE standardization methods, most with more fine-tuning than a GLM, and these methods should be considered when assessing the use of a SEM for abundance estimation.

5 Conclusion

In the face of problematic assumptions made by standard GLM-based abundance estimation, we have explored the use of SEMs as an alternative. In particular, we have focused on the problem of deducing indices of abundance for the swordfish stocks in the Indian Ocean based on data for the Spanish fleets' swordfish longline fishery. We have found evidence for significant differences between the GLM-based and SEM-based measures of abundance. Further, we have found evidence for an endogenous relationship between catch and effort. More specifically, we find that for all SEMs we studied, the influence of catch on effort was relatively strong (~0.6) while the influence of effort on catch was significantly weaker (~0.15). These observations are consistent with collaboration between fishers through information exchange. Future research on swordfish abundance with both GLM and SEM approaches would be required to either validate or invalidate these tentative observations.

References

Abeare, S. M. (2009). Comparisons of boosted regression tree, glm and gam performance in the standardization of yellowfin tuna catch-rate data from the gulf of Mexico longline fishery. (Doctoral dissertation, University of Pretoria).

Bigelow, K. A., Boggs, C. H., & He, X. (1999). Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. *Fisheries Oceanography*, 8(3): 178-198.

Bollen, K. A. (2014). Structural equations with latent variables. John Wiley & Sons.

Breslow, N. E. (1996). Generalized linear models: checking assumptions and strengthening conclusions. *Statistica Applicata*, 8: 23-41.

Fernández-Costa, J., Ramos-Cartelle, A., García-Cortés, B., Mejuto, J. (2014). Standardized catch rates for swordfish (Xiphas gladius) caught by the Spanish longline in the indian ocean during the 2001-2012 period. IOTC Proceeding 2014-WPB12-20.

Gaertner, D. & Dreyfus-Leon, M. (2004). Analysis of non-linear relationships between catch per unit effort and abundance in a tuna purse-seine fishery simulated with artificial neural networks. *Journal of Marine Science*, 61: 812-820.

García-Cortés, B., Ramos-Cartelle, A., González-González, I., Mejuto, J. (2012). Biological observations of oceanic whitetip shark (Carcharhinus longimanus) on Spanish surface longline fishery targeting swordfish in the Indian Ocean over the period 1993-2011. IOTC-WPEB-2012.

Glaser, S., Ye, H., Maunder, M., MacCall, A., Fogarty, M., Sugihara, G. (2011). Detecting and forecasting complex nonlinear dynamics in spatially structured catch-per-unit-effort time series for North Pacific albacore (Thunnus alalunga). *Canadian Journal of Fisheries and Aquatic Science*, 68: 400-412.

Grace, J. (2006). Structural Equation Modeling and Natural Systems. New York, Cambridge University Press.

Harley, S. J., Myers, R. A., & Dunn, A. (2001). Is catch-per-unit-effort proportional to abundance? *Canadian Journal of Fisheries and Aquatic Sciences*, 58(9): 1760-1772.

Hicks, A. C. (2013). The utility of catch-per-unit-effort when assessing and managing long-lived fish stocks. (Doctoral dissertation).

Hilborn, R., & Walters, C. J. (Eds.). (2013). Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Springer Science & Business Media.

Hinton, M. G., & Maunder, M. N. (2004). Methods for standardizing CPUE and how to select among them. *Col. Vol. Sci. Pap. ICCAT*, *56*(1), 169-177.

Hooper, D., Coughlan, J., Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determing Model Fit. *Journal of Business Research Methods*, 6: 53-60.

IOTC Secretariat (2014). Status of Indian Ocean Swordfish Executive Summary. IOTC-2014-SC17-ES16.

Kolody, D., Poisson, F., Bourjea, J. (2010). Reunion Longline Swordfish Catch Rate Standardization. Indian Ocean Tuna Commission Working Party on Billfish Working Paper. IOTC-2010-WPB-03.

Lei, P. & Wu, Q. (2007). Introduction to Structural Equation Modeling: Issues and Practical Considerations. *Educational Measurement: Issues and Practice*, 33-43.

Marchal, P., Andersen, B., Caillart, B., Eigaard, O., Guyader, O., Hovgaard, H., Iriondo, A., Fur, F., Sacchi, J., Santurtun, M. (2007). Impact of technological creep on fishing effort and fishing mortality, for a selection of European fleets. *Journal of Marine Science*, 64: 192-209.

Maunder, M. N. (2001). A general framework for integrating the standardization of catch per unit of effort into stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(4): 795-803.

Maunder, M. & Punt, A. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*, 70: 141-159.

Maunder, M., Sibert, J., Fonteneau, A., Hampton, J., Kleiber, P., Harley, S. (2006). Interpreting catch per unit effort data to assess the status of individual stocks and communities. *Journal of Marine Science*, 63: 1373-1385.

Mejuto, J., Garcia-Cortes, B., Ramos-Cartelle, A., De le Serma, J., Gonzalez-Gonzalez, I., Fernandez-Costa, J. (2013). Standardized catch rates of Shortfin Mako (Isurus oxyrinchus) caught by the Spanish surface longline fishery targeting swordfish in the Atlantic ocean during the period 1990-2010. *SCI. PAP. ICCAT*, 69(4): 1657-1669.

Muthén, K., & Muthén, O. (2010). Mplus User's Guide. Los Angeles, Muthén & Muthén, 6th edition.

Myers, R. H., Montgomery, D. C., Vining, G. G., & Robinson, T. J. (2012). Generalized linear models: with applications in engineering and the sciences (Vol. 791). John Wiley & Sons.

Nakano, H., & Clarke, S. (2005). Standardized CPUE for blue sharks caught by the Japanese longline fishery in the Atlantic Ocean, 1971–2003. *ICCAT Collective Volume of Scientific Papers*, 58(3), 1127-1134.

Paloheimo, E., & Dickie, M. (1964). Abundance and fishing success. *Rapports et Procès-verbaux des Réunions du Conseil International pour Exploration de la Mer*, 155: 152-163.

Punt, A. E., Walker, T. I., Taylor, B. L., & Pribac, F. (2000). Standardization of catch and effort data in a spatially-structured shark fishery. *Fisheries Research*, 45(2): 129-145.

Richards, A., Whittingham, M., Stephens, A. (2011). Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework. *Behavioral Ecology and Sociobiology*, 65(1): 77-89.

Richards, L.J. & Schnute, J.T. (1986). An Experimental and Statistical Approach to the Question: Is CPUE an Index of Abundance? *Canadian Journal of Fisheries and Aquatic Science*, 43: 1214-1227.

Santos, M., Coelho, R., Lino, P. (2012). A brief overview of the swordfish catches by the portugese pelagic longline fishery in the indian ocean: catch, effort, CPUE and catch-at-size. IOTC Proceeding 2012-WPB10-11.

Shono, H. (2008) Application of the Tweedie distribution to zero-catch data in CPUE analysis. *Fisheries Research*, 93: 154-162.

Ullman, J. (2006). Structural Equation Modelling: Reviewing the Basics and Moving Forward. *Journal of Personality Assessment*, 87: 35-50.

Uozumi, Y. (1998). Standardization of catch per unit effort for swordfish and billfishes caught by the Japanese longline fishery in the Indian Ocean. National Research Institute of Far Seas Fisheries, Fisheries Agency of Japan.

Venables, W.N. & Dichmont, C.M. (2004). GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fisheries Research*, 70: 319-337.

Wagenmakers, E., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1): 192-196.

Walters, C., & Maguire, J. J. (1996). Lessons for stock assessment from the northern cod collapse. *Reviews in fish biology and fisheries*, 6(2): 125-137.

Wang, S. & Nishida, T. (2011). CPUE standardization of swordfish (Xiphias gladius) caught by Taiwanese longline fishery in the Indian Ocean. IOTC Proceeding 2011–WPB09–16.

Ye, Y. & Dennis, D. (2009). How reliable are the abundance indices derived from commercial catch-effort standardization? *Canadian Journal of Fisheries and Aquatic Science*, 66: 1169-1178.

Appendix A.

Model Diagnostics

Model diagnostics were performed for each GLM to ensure that model assumptions were not violated. Below are three diagnostic plots for the preferred GLM.

Cook's Distance Plot

Figure A1. Cook's Distance Plot.

This diagnostic plot gives an indication of the influence of estimates on the outcome of the model. All values in the plot are less than 0.1, so no one estimate is having a much larger influence on the outcome than any other.

Figure A2.   Q-Q Plot.

The Q-Q plot is used to test if the data is normally distributed. This plot does suggest that the model may not be perfectly normal, but the deviation is not significant enough to result in a violation of the assumptions.

Residuals vs. Fitted Plot

Figure A3.   Residuals vs. Fitted.

The residuals vs. fitted plot is used to detect non-linearity, unequal error variances, and outliers. The residuals should; bounce randomly around the 0 line, form a horizontal band around the 0 line, and no one residual should "stand out" from the basic random pattern. This plot suggests that there may be a few outliers, but that the assumption of linearity has not been broken.

Appendix B.

Fit for Models 2-9

Model 1

Figure B1. Initial SEM path coefficients and variances estimated with 15 years of data.

Fit Measures		
Chi Square	427	p-value = 0.00
RMSEA	0.170	p-value = 0.00
SRMR	0.082	
AIC	2572	

Model 2

Figure B2.   Model 2 path coefficients and variances estimated with 15 years of data.

Fit Measures		
Chi Square	190	p-value = 0.00
RMSEA	0.147	p-value = 0.00
SRMR	0.075	
AIC	3683	

Model 3

Figure B3. Model 3 path coefficients and variances estimated with 15 years of data.

Fit Measures		
Chi Square	174	p-value = 0.00
RMSEA	0.169	p-value = 0.00
SRMR	0.085	
AIC	4415	

Model 4

1

Abundance$_{T,X}$

Locality$_{T-1Y,X}$

Catch$_{T-1Y}$

Seasonality$_T$ — .85

Locality$_X$ — .74

.39

.51

.08

.69

-.11

Effort$_{T,X}$

.18
.55

Catch$_{T,X}$

.56

.31

Figure B4. Model 4 path coefficients and variances estimated with 15 years of data.

Fit Measures		
Chi Square	134	p-value = 0.00
RMSEA	0.147	p-value = 0.00
SRMR	0.070	
AIC	3595	

Model 5

Figure B5. **Model 5 path coefficients and variances estimated with 15 years of data.**

Fit Measures		
Chi Square	40.76	p-value = 0.00
RMSEA	0.080	p-value = 0.02
SRMR	0.045	
AIC	3021	

43

Model 6

Figure B6. Model 6 path coefficients and variances estimated with 15 years of data.

Fit Measures		
Chi Square	54	p-value = 0.00
RMSEA	0.122	p-value = 0.00
SRMR	0.056	
AIC	4034	

Model 7

Figure B7. Model 7 path coefficients and variances estimated with 15 years of data.

Fit Measures		
Chi Square	120	p-value = 0.00
RMSEA	0.186	p-value = 0.00
SRMR	0.081	
AIC	4324	

Model 8

Figure B8. Model 8 path coefficients and variances estimated with 15 years of data.

Fit Measures		
Chi Square	3.25	p-value = 0.52
RMSEA	0.048	p-value = 0.96
SRMR	0.011	
AIC	3251	

Model 9

Figure B9. Model 9 path coefficients and variances estimated with 15 years of data.

Fit Measures		
Chi Square	0.008	p-value = 0.93
RMSEA	0.000	p-value = 0.98
SRMR	0.001	
AIC	3942	

Model 9 (19 years)

Figure B10.   Model 9 path coefficients and variances estimated with 19 years of data.

Fit Measures		
Chi Square	0.507	p-value = 0.48
RMSEA	0.000	p-value = 0.84
SRMR	0.004	
AIC	4929	