National College of Ireland

Advancing Player Modelling in Rugby Union: An Evaluation of Path Analysis Techniques

MSc Research Project
Programme Name

Cian Ó Muilleoir
Student ID: 20144717

School of Computing
National College of Ireland

Supervisor: Jorge Basilio

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Cian Ó Muilleoir
Student ID:	20144717
Programme:	Programme Name
Year:	2023
Module:	MSc Research Project
Supervisor:	Jorge Basilio
Submission Due Date:	14/08/2023
Project Title:	Advancing Player Modelling in Rugby Union: An Evaluation of Path Analysis Techniques
Word Count:	936
Page Count:	8

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	14th August 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Advancing Player Modelling in Rugby Union: An Evaluation of Path Analysis Techniques

Cian Ó Muilleoir
20144717

1 Overview

This manual provides instructions to replicate the experimental setup.

2 Hardware and Software Requirements

2.1 Hardware

- Operating System: Microsoft Windows 11 Home Version 10.0.22621 Build 22621
- Machine Type: x64-based PC
- Central Processing Unit (CPU): Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz with 6 cores and 12 logical processors running at a maximum clock speed of 2592 MHz
- Random Access Memory (RAM): 16.0 GB of physical RAM installed.

2.2 Development Environment

Integrated Development Environment: Visual Studio Code version 1.82.0-insider

2.3 Languages and Runtimes

Python version 3.10.6

2.4 Statistical Software

- SEM Software: AMOS version 26 or greater
- PLS-PM Software: SmartPLS version 4 or greater

3 Data

3.1 Data Source

The raw rugby event data used in the research was provided by Oval Insights, a sports technology company focused on generating insights from player tracking and event data.

A bespoke computer vision machine learning system and human event annotation process were used to collect over 70 distinct event types from match footage.

3.2 Dataset Contents

The dataset contains matches played during the 2020/2021 season of a premier northern hemisphere rugby union competition. It includes 16 teams representing 5 countries across 301 matches played in venues in those countries. Over 246,554 distinct events were recorded capturing passes, tackles, scrums, kicks and more.

3.3 File Format

The raw data will be provided in the form of a single comma-separated values (CSV) file named "URC 202101.csv". This CSV file format stores 88 columns of data for each event row.

3.4 Access Restrictions

The data is provided under non-disclosure by Oval Insights and cannot be further distributed. The analysis is intended to be replicated using the provided software configuration and CSV data file only.

4 Environment

Opening the code files in Visual Studio Code with a Python 3.10.6 kernel (or greater) activated. Jupyter Notebooks should also be compatible, though this was not tested.

5 EDA

5.1 Python Environment

Set up the Python 3.10.6 virtual environment as described in Section 2.2. Install the following packages:

- pandas
- numpy
- matplotlib
- seaborn

5.2 Jupyter Notebook

Open the Jupyter Notebook "Rugby Data EDA.ipynb" as provided. This notebook performs EDA on the raw rugby data CSV file.

5.3 File Paths

Ensure that the df variable at the start of the script references the correct raw data CSV path in the pd.read_csv command.

5.4 Summary Statistics

Summary statistics of the over 70 encoded event types are generated to understand their distribution.

5.5 Correlation Analysis

Pearson's correlation coefficient is used to identify relationships between variables like x and y coordinates.

5.6 Event Type Distributions

The distribution of event types is analysed across match outcomes (wins/losses) and visualised.

5.7 Pitch Heatmaps

Hexagonal bin heatmap visualisations are generated to understand the concentration of events across pitch zones.

6 ETL Pipeline

6.1 Python Environment

Set up the Python 3.10.6 virtual environment as described in Section 2.2. Install the following packages:

- pandas
- numpy
- sklearn

6.2 Jupyter Notebook

Open the Jupyter Notebook "Rugby Analytics Pipeline.ipynb" as provided. Running this notebook end-to-end generates the final preprocessed dataset used in subsequent modelling and analysis steps.

6.3 File Paths

Ensure the code references the correct raw CSV input file path. The output preprocessed CSV path can be modified as required.

6.4 Data Extraction

The raw rugby data CSV file is read into a Pandas DataFrame for initial cleaning and manipulation.

6.5 Data Transformation

Columns are converted to appropriate data types, strings are cleaned, and additional features like aggregated values and one-hot encodings are generated.

6.6 Data Loading

The transformed DataFrame is written out to three new CSV files. The filepath set by UNGROUPED_OUTPUT_FILEPATH provides the processed data without grouping, and the one set by SEQ_GROUPED_OUTPUT_FILEPATH provides that data grouped by an identifier based on the provided "sequence" variable, both of which were used for troubleshooting purposes.

The filepath set by ABS_POSS_GROUPED_OUTPUT_FILEPATH is grouped by custom identifier designed to map chains of possession which outlast either the "sequence" or "possession" variables. The resulting file was used in the modelling which followed.

7 Models & Results

7.1 CB-SEM Model Loading

From the provided "Final Simple CB Model Results.zip", open "Final Simple CB Model.amw" in AMOS 26 or greater for the CB-SEM model. This file specifies the exogenous and endogenous variables based on the conceptual framework and draws paths between constructs based on hypothesised relationships.

7.2 CB-SEM Model Analysis

The necessary analysis settings are saved within the file automatically but will be included here for completeness in Figures 1 to 4. They are set under the menu option View → Analysis Properties. The results reported are provided in the "Final Simple CB Model Results.html" file in the same zip folder.

7.3 PLS-SEM Model Loading

Within an existing workspace in SmartPLS 4, use the menu option Files → "Import from backup file" on the file "Final SmartPLS Models.zip" as provided. This file contains the measurement model linking indicators to latent variables and the structural model paths between latent variables.

7.4 PLS-SEM Model Analysis

The PLS algorithm and bootstrap re-sampling are used to estimate model paths. The settings are saved with the files, but again they will be provided here for completeness' sake in Figures 5 and 6. They are set under the menu option Calculate → Bootstrapping.

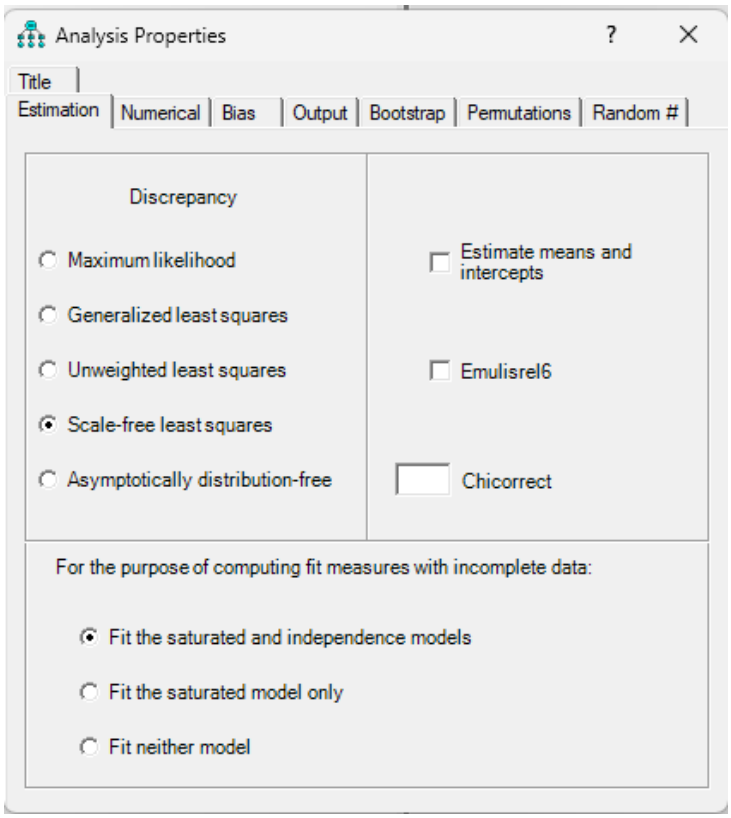

Figure 1: Analysis Properties - Estimation

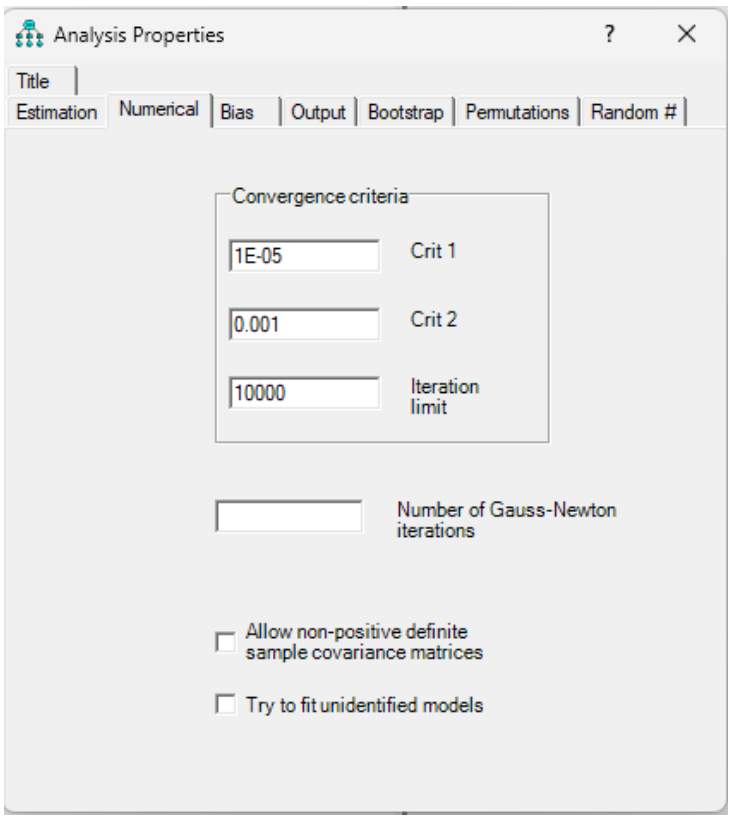

Figure 2: Analysis Properties - Numerical

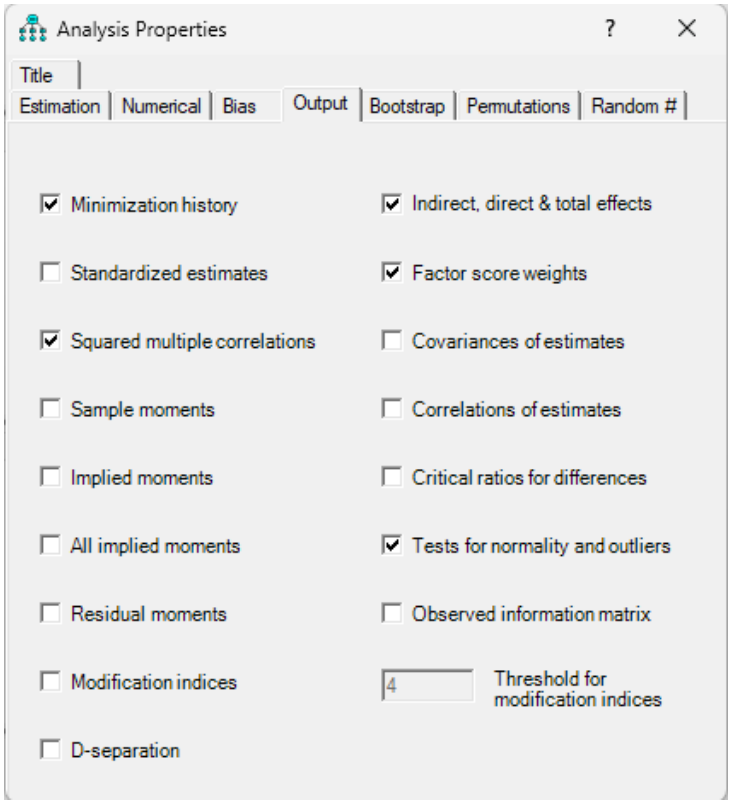

Figure 3: Analysis Properties - Output

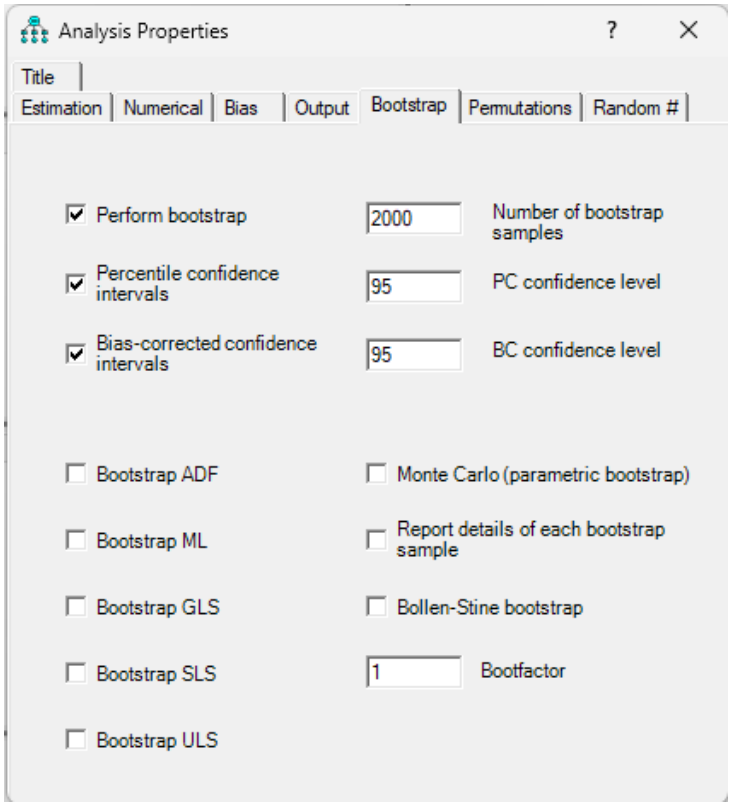

Figure 4: Analysis Properties - Bootstrap

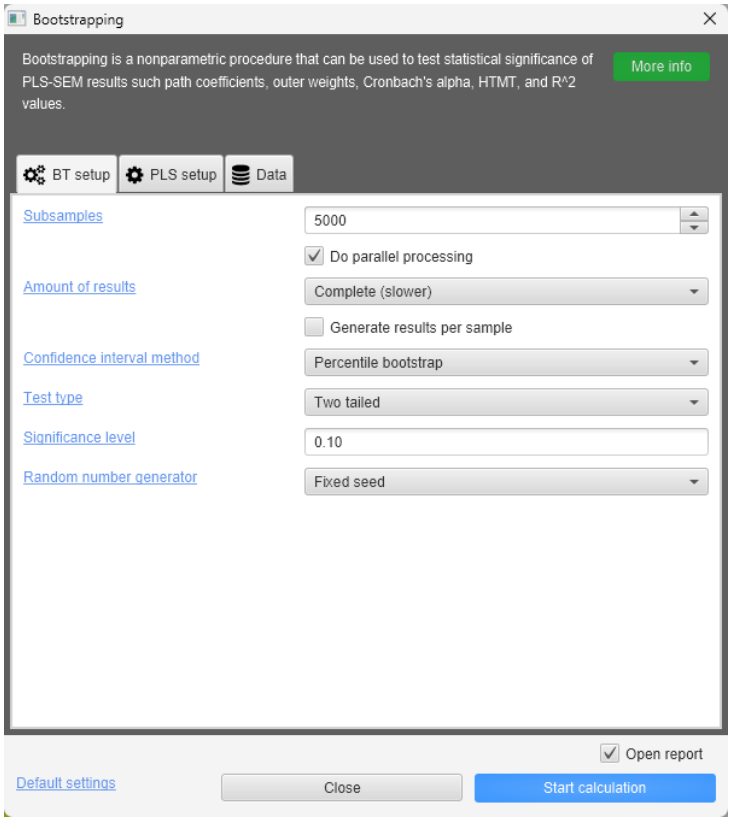

Figure 5: Bootstrapping - BT Setup

The results reported are provided within the SmartPLS backup, and separately as "Final Simple PLS Model Results.html" and "Final Complex PLS Model Results.html" files in a separate zip folder.

Figure 6: Bootstrapping - PLS Setup