*Philip Morris International: PMI IMPACT Project*

IT for Illicit Trade Risk Management (IT2RM)

Data Science Delivery Report 7:

**Scenario Builder Overview**

**Prepared by: Prof. Anton Gerunov, Ph.D., PMP**

Delivery Date: 31. May 2022

**Contents**

[**Executive Summary** 3](#_Toc76554022)

[**Background** 5](#_Toc76554023)

[**Overview of the Scenario Builder** 5](#_Toc76554024)

[Part 1: Quantitative Modelling and Simulations Setup 6](#_Toc76554025)

[Part 2: Scenario Results and Visualization 7](#_Toc76554026)

[**Functionalities of the Scenario Builder** 7](#_Toc76554027)

[Relationships Panel 8](#_Toc76554028)

[Illicit Incidence Panel 8](#_Toc76554029)

[Forecast Dynamics Panel 9](#_Toc76554030)

[Distribution Panel 10](#_Toc76554031)

[**Conclusion** 10](#_Toc76554032)

**List of Figures**

Figure 1: Overall Logic of the Scenario Builder 6

Figure 2: The PMI Illicit Incidence Scenario Builder App Home Screen, Running on Local Host 7

Figure 3: The PMI Illicit Incidence Scenario Builder Simulation Module 8

Figure 4: The PMI Illicit Incidence Scenario Builder Forecast Dynamics Module 9

Figure 5: The PMI Illicit Incidence Scenario Builder Distribution Module 10

# **Executive Summary**

The current delivery report builds upon the insights gleaned from the previous modeling and risk management exercises and aims to use the outlined process drivers to construct a Scenario Builder – a simulation modeling tool that enables to analyst to quickly generate different scenarios for illicit trade incidence given dynamics in underlying process drivers.

More specifically, the Scenario Builder tools should provide the following functionalities:

* Ability to **track and review the relationship between the main drivers** of illicit cigarette incidence and their connection to the resulting percent of contraband cigarettes.
* Functionality to **generate simulations for illicit incidence** for different values of the drivers; in this respect this enables one to forecast the expected prevalence of contraband in case of economic crisis or boom, under different social conditions, and over alternative demographic regimes.
* Module for **forecasting dynamics of illicit incidence** following the simulated period; once the objective conditions of the economy and society change, this changes not merely the next period, but the overall trajectory of incidence and the Scenario Builder must be able to forecast that.
* Visualization for **the distribution of expected outcomes** that gives an idea of overall statistical properties of the actual and forecasted values, outline the range of variation, and allows for iterative model improvement.



Figure: Overall Logic of the Scenario Builder

A schematic representation of the overall logic of the Scenario Builder app is shown in the figure aboveFigure 1. The Scenario Builder development was thus divided into two major phases:

* Quantitative Modelling and Simulations Setup
* Scenario Results and Visualization

The Illicit Incidence Scenario Builder is a tool that simulates what different levels of illegal incidence can be expected for specific value of the process drivers. It builds upon the data science results of the IT2RM Project and parameterizes an adaptive regression spline structural model. Coefficient values indicate the strength of each specific driver.



Figure: The PMI Illicit Incidence Scenario Builder App Home Screen

The Scenario Builder comprises four panels that cover its main functionalities (see Figure above):

* Relationships Panel
* Illicit Incidence Panel
* Forecast Dynamics Panel
* Distribution Panel

Scenario building takes place at the country level and considers each specific country’s peculiarities and different influences. Specifically, the effects of economic and social factors of contraband can be explored in depth.

Overall, **the data science component is proceeding according to the timeline**. Data curation is complete and needs only updating as new data becomes available. The structural forecasting model is estimated, and the risk management forecasting model are used to construct risk scores and show which ones of a country regions present the greatest risk for illicit cigarette incidence. The Scenario Builder provides easy capability to understand and visualize the major insights from the structural model and enable a wide range of stakeholders to easily peruse major results and generate alternative scenarios.

**With the completion of this project part, the substantive data science work is also completed according to the initial and the expanded project scope.** Results and code are made available so that they can be used either as-is, expanded, or used as components of further modeling, forecasting and risk management projects.

# **Background**

This PMI Impact project – IT for Illicit Trade Risk Management (IT2RM) aims at utilizing publicly and privately available data, link them in a unified data warehouse and develop sophisticated analytic capabilities on top of it. Leveraging data on crime, socio-economic development, consumer sentiment, legitimate trade, consumer behavior, illicit cigarette and tobacco market and intercepted illegal imports the project will create a unified database that can be used to visualize and analyze key trends in illicit trade and outline the main drivers at a regional level. This will be used to gain insight into the connection between illicit trade in cigarettes and other criminal activities at a detailed level of granularity. Furthermore, a sophisticated forecasting and risk management system is to be built on top of that, dynamically showing increases in the risk of illicit cigarette trade in different regions that can guide both producers and law enforcement authorities.

The current delivery report builds upon the insights gleaned from the previous modeling and risk management exercises and aims to use the outlined process drivers to construct a Scenario Builder – a simulation modeling tool that enables to analyst to quickly generate different scenarios for illicit trade incidence given dynamics in underlying process drivers.

More specifically, the Scenario Builder tools should provide the following functionalities:

* Ability to **track and review the relationship between the main drivers** of illicit cigarette incidence and their connection to the resulting percent of contraband cigarettes.
* Functionality to **generate simulations for illicit incidence** for different values of the drivers; in this respect this enables one to forecast the expected prevalence of contraband in case of economic crisis or boom, under different social conditions, and over alternative demographic regimes.
* Module for **forecasting dynamics of illicit incidence** following the simulated period; once the objective conditions of the economy and society change, this changes not merely the next period, but the overall trajectory of incidence and the Scenario Builder must be able to forecast that.
* Visualization for **the distribution of expected outcomes** that gives an idea of overall statistical properties of the actual and forecasted values, outline the range of variation, and allows for iterative model improvement.

In short, this delivery report outlines the process of creating the Scenario Builder tool, using it to generate simulations, and visualizing key results.

# **Overview of the Scenario Builder**

The Scenario Builder follow an intuitive structure. It is developed along the logic of the data science analytic pipeline along the following steps:

1. Use optimal structural forecasting model to find most important incidence drivers across all four countries.
2. Select those driver variable variables that have a sufficiently long time series (data availability) that can be suitably used for modeling and forecasting (<50% missing observations).
3. Include those drivers as modifiable features in the Scenario Builder specification.
4. Provide simulation capability for illicit incidence given driver values.
5. Provide forecasting capability given a simulated incidence value.
6. Code an intuitive app that wraps those capabilities. The current version of the Scenario Builder is coded in the R library Shiny but can be ported to alternative languages if needed as part of project expansion or in other projects.
7. A visualization module is added to the app, allowing the display of simulations, forecasts and statistical distributions.
8. An additional module for investigating the relationship between variable drivers is added to further aid understanding the dynamics, reliability, and limitations of results.



Figure 1: Overall Logic of the Scenario Builder

A schematic representation of the overall logic of the Scenario Builder app is shown in Figure 1. The Scenario Builder development was thus divided into two major phases described below. Those are the following:

* Quantitative Modelling and Simulations Setup;
* Scenario Results and Visualization.

## **Part 1: Quantitative Modelling and Simulations Setup**

**Scenario Builder** – the small-scale simulation model aims to make an interactive simulation model that enables analysts, law-enforcement, decision-makers, and members of the public to explore how different social, political and economic development scenarios will influence the total level of illicit cigarette trade. The Scenario Builder integrates all of the relationships uncovered in previous stages of the project and enable the exploration of their quantitative effect on illicit cigarette influence.

More particularly, the Scenario Builder will see how Non-domestic Cigarette Incidence in response to changes in the following variables:

* Age Dependency Ratio
* Agricultural Employment
* Total Employment
* Real Growth
* Remittances
* Population Growth

Setting desired values for each of those (or letting them take default values) enable the simulation of how the results of different policy interventions affect the level of contraband. Specifically, the effects of economic and social factors of contraband can be explored in depth. Scenario building takes place at the country level and considers each specific country’s peculiarities and different influences.

## **Part 2: Scenario Results and Visualization**

**Scenario Builder Visualization Functionality** – this part of the work package focuses on creating appropriate visualizations for the Scenario Builder that include both visuals of the variables that are amenable to changes, as well as the resultant dynamics of illicit cigarette incidence. Intuitive user interface through sliders enables a wide range of stakeholders to take advantage of the tool. Special care is be taken to provide clear results of how policy intervention affects total contraband. The visualization functionality strictly follows the logic of the Scenario Builder and provides visualizations at the level of a specific country which can aid the direct comparison between the same interventions across the diverse countries that are in the project scope. visualization functionality is developed in such a way that can be deployed even outside the standard results dashboard currently included in the project scope, thus making it a useful standalone utility that can be made available to a wide range of stakeholders.

# **Functionalities of the Scenario Builder**

The Illicit Incidence Scenario Builder is a tool that simulates what different levels of illegal incidence can be expected for specific value of the process drivers. It builds upon the data science results of the IT2RM Project and parameterizes an adaptive regression spline structural model. Coefficient values indicate the strength of each specific driver. Its home screen is shown in Figure 2.



Figure 2: The PMI Illicit Incidence Scenario Builder App Home Screen, Running on Local Host

The simulator allows for a detailed examination of contraband dynamics as different objective conditions change. It is composed of four key modules that allow the visualization of relationships, the simulation of values, the forecast dynamics given a simulated value, and one that visualizes the spread of realizations.

## **Relationships Panel**

This panel of the simulator shows the relationship between the key model variables (see Figure 2). The user can select the variables for the graph axes and note whether they would like to see a smoothed fit. The visualization is reactive and changes as the different variables for the x and y axis are selected. Likewise, the visualization reacts to the type of smoothed relationship chose. The options here are:

* None
* Linear
* Generalized Linear Model
* Loess

The ability to see the smoothed trend provides for a clearer understanding of how the two selected process drivers are related.

## **Illicit Incidence Panel**

This panel show the simulation results. It displays the values for the country selected and estimates the expected level of illicit incidence, given the parameter values entered in the left-side panel.



Figure 3: The PMI Illicit Incidence Scenario Builder Simulation Module

The incidence panel is shown in Figure 3. It displays only the value of the country selected in the left-hand pane. As the user inputs the desired values for remittances, real economic growth, quarter, employment rate, rate of agricultural employment, age dependency ration, and the rate of population growth, the underlying model automatically generates a simulated value for the expected illicit incidence. This simulation value is then visualized as a next period (T+1) forecast. This simulation module includes necessary functionality for devising alternative courses of action and, through imputing their possible effects, enables the analyst to quantitatively estimate the results.

## **Forecast Dynamics Panel**

This panel simulates the long-term dynamics following the simulated period. This panel leverages an advanced ARIMA model. Hundreds of alternative time series ARIMA models are generated in real-time and the best one is chosen based on the Bayesian Information Criterion (BIC). This is then used to generated forecasts for the expected illicit incidence for the four quarters (one year) ahead after the simulated period. Note that the simulated number is now treated as an actual and used for the forecasting.



Figure 4: The PMI Illicit Incidence Scenario Builder Forecast Dynamics Module

As the data input for the left-hand panel is updated, so is incidence value for the simulation period. It is then included as a realization and the whole realization series is used to generate the four-period-ahead forecast. In practice, the user can merely change the inputs and then the visualization updates in real time to reflect both the new simulated values and the resulting forecast.

## **Distribution Panel**

This panel shows the distribution of both the realized and the forecasted incidence values. It gives an idea of the statistical properties of the illicit incidence realizations and provides insight into their range of variation. The Distribution panel is shown visually in Figure 5.



Figure 5: The PMI Illicit Incidence Scenario Builder Distribution Module

The visualized histogram also makes the distinction between actual incidence values and the forecasted ones – they are colored accordingly. This module also allows the analyst to select the appropriate number of bins for the histogram, thus making it more or less granular, depending on the needs. As with all inputs, the default selection is either a generally accepted convention, or roughly reflects the mean value of the data samples.

# **Conclusion**

The modeling exercise of the first eight quarters of the **project has achieved the following**:

* Collated, procured and organized a large database of socio-economic, demographic, business, and crime statistics for the purposes of the project.
* Performed initial data analysis and visualization of the data.
* Outlined possible drivers of non-domestic cigarette incidence and interpreted their associations within the framework of correlational analysis.
* Selected best predictors and constructed an optimal structural forecasting model and estimated it for the national level using data at hand.
* Investigated a wide variety of approaches for making granular regional-level forecasts that can be used for risk management.
* Selected an ARIMA-based automatic forecasting and optimal model determination approach.
* Estimated forecasts for 55 regions in Bulgaria, Serbia, and Turkey.
* Generated visualizations and exported forecasts in machine-readable form for inclusion in the other parts of the project.
* Used risk management forecasts to generate risk scores and prioritize regions.
* Proposed appropriate color coding and intuitive visualization that can be used by stakeholders with varied backgrounds and interests.
* Created a Scenario Builder that includes the insights from the structural modeling parts of the data science work and enables the seamless generation of simulations for illicit cigarette incidence.

Overall, **the data science component is proceeding according to the timeline**. Data curation is complete and needs only updating as new data becomes available. The structural forecasting model is estimated and the risk management forecasting model are used to construct risk scores and show which ones of a country regions present the greatest risk for illicit cigarette incidence. The Scenario Builder provides easy capability to understand and visualize the major insights from the structural model and enable a wide range of stakeholders to easily peruse major results and generate alternative scenarios.

All results from the data science part of the IT2RM project are made available to other project participants on time so that overall delivery is assured. Should updated date become available, the database can easily be populated with the new values to be used for re-estimating both the national-level structural forecasting model as well as regional level ARIMA-models and the risk scores. The Scenario Builder tool can also be swiftly updated to incorporate updated data and re-estimated structural model coefficients. **With the completion of this project part, the substantive data science work is also completed according to the initial and the expanded project scope.** Results and code are made available so that they can be used either as-is, expanded, or used as components of further modeling, forecasting and risk management projects.