Building Validated Causal Loops: A New Methodology Bridging Ishikawa Diagrams and Temporal Causality Analysis

Adalberto Prada Rico[1], Daniel Ortiz Castillo[1], Luis Rabelo[2]*, Cristian Rincon-Guio[3], Juan Pablo Zamora-Aguas[2], Natalia Orozco[3], Mayra Bornacelly Castaneda[2], Edgar Gutierrez-Franco[4]

[1] Universidad de La Sabana, Colombia
[2] University of Central Florida, Orlando, FL, USA
[3] Corporación Universitaria Minuto de Dios (UNIMINUTO), Colombia
[4] Universidad Externado de Colombia, Colombia

* Corresponding Author: Luis.Rabelo@ucf.edu

Abstract:

Traditional system dynamics practice relies heavily on expert judgment to construct causal loop diagrams, often without systematic empirical validation. This paper proposes a five-step methodology that integrates structured hypothesis generation via transformed Ishikawa diagrams with formal temporal causality testing to construct empirically validated causal loop structures. The framework exhaustively enumerates plausible causal relationships; tests temporal precedence and predictive relevance using appropriate causality methods and determines causal polarity to distinguish reinforcing and balancing feedback loops. The methodology is demonstrated through a macroeconomic case study involving nine interacting variables. From 72 hypothesized relationships generated through Ishikawa diagrams, only 28 (39%) were empirically validated, revealing five statistically significant feedback loops ($p < 0.001$). The validated causal structure was subsequently operationalized in a hybrid System Dynamics–Agent-Based model achieving strong empirical performance ($R^2 = 0.89$, MAPE = 6.7%). By embedding empirical validation directly into the conceptualization stage, the proposed approach improves structural validity, eliminates spurious causal assumptions, and provides full traceability from data to model structure. The methodology offers a repeatable protocol for causal loop construction that bridges system dynamics and data science, enhancing model credibility for high-stakes policy and decision-support applications.

Keywords: System Dynamics, Causal Loop Diagrams, Temporal Causality, Ishikawa Diagrams, Empirical Validation, Hybrid Modeling, Agent-Based Modeling

1. Introduction

Understanding the behavior of complex socioeconomic and technological systems requires models grounded in accurate and defensible causal structures. In practice, however, identifying these causal relationships remains one of the most challenging stages of system dynamics modeling. Traditional approaches rely heavily on expert judgment, qualitative mapping exercises, or theoretical assumptions, which—although valuable—often introduce ambiguity, bias, or undocumented reasoning into the development of causal loop diagrams. As systems grow in complexity and the number of interacting variables increases, the need for a transparent, data-driven methodology to support causal inference becomes more evident.

Ishikawa diagrams are widely used to organize hypothesized causes of a given effect, providing an intuitive structure for categorizing influencing factors. Yet, in their conventional form, they do not confirm whether the proposed causal links exist in real data or whether the temporal directionality implied by the diagram is valid.

Recent advances in data availability and time-series analysis offer new opportunities to complement expert-based reasoning with empirical evidence. Specifically, causality tests such as Granger, Engle-Granger, Johansen, Toda-Yamamoto, and Transfer Entropy allow researchers to evaluate whether one variable contributes predictive information about another, thereby enabling a more rigorous evaluation of hypothesized cause-and-effect relationships.

This paper proposes a systematic, data-driven methodology that integrates Ishikawa diagramming with formal causality analysis to construct validated causal loops for system dynamics modeling. The approach begins with the qualitative strength of the Ishikawa framework to comprehensively map potential causal factors. It then incorporates statistical causality tests—selected according to the stationarity and linearity properties of each variable—to empirically validate or discard proposed causal links. Finally, Pearson correlation is used to identify the polarity (positive or negative) of the confirmed relationships, enabling the construction of causal loops that are both conceptually coherent and empirically grounded.

By combining qualitative reasoning with quantitative validation, this methodology addresses a persistent gap in system dynamics practice: the difficulty of translating domain knowledge into evidence-based causal structures. The resulting causal loop diagrams are not only more robust but also more transparent, reproducible, and defensible for both academic and managerial decision-making contexts. The contribution is scoped to systems—or system components—in which quantitative variables are dominant, and to the causal-structure stage: it produces validated causal loop diagrams, from which stock-and-flow diagrams are subsequently derived.

To demonstrate the relevance and applicability of the proposed methodology, the paper applies it to a real-world case study: modeling and forecasting new car sales in Bogotá, Colombia. This domain provides a complex environment in which economic variables, consumer behavior, macro-economic conditions, and market interactions converged ideal —an ideal setting to illustrate how data-driven causality analysis can significantly enhance system understanding and model validity. Insights from this case study show how validated causal loops strengthen the foundation for hybrid simulation models that integrate system dynamics with agent-based modeling.

2. Literature Review: Methodologies for Causal Loop Diagram Construction

The construction of causal loop diagrams (CLDs) represents a foundational element in system dynamics modeling, as these diagrams capture the hypothesized feedback structures that drive complex system behavior. Over the past five decades, multiple methodological traditions have emerged to support the identification and validation of causal relationships within intricate systems. This section provides a comprehensive review of the principal approaches documented in the literature, examining their capabilities, limitations, and underlying assumptions. The objective of this review is to contextualize the need for more rigorous, data-driven methodologies—specifically, the integration of qualitative structural analysis tools (such as Ishikawa diagrams) with formal statistical causality testing (such as Granger causality)—which forms the methodological contribution of this research.

2.1 Expert-Based Conceptual Modeling

The earliest and most extensively utilized approach for constructing causal loops relies fundamentally on expert knowledge and conceptual reasoning. Pioneered by Forrester (1961) and subsequently refined by the system dynamics community, this methodological tradition emphasizes deep domain understanding, stakeholder engagement, and iterative refinement of mental models (Sterman, 2000; Vennix, 1996). The principal techniques within this paradigm include:

Group Model Building (GMB). Group model building represents a participatory approach wherein facilitators, domain experts, and stakeholders collaboratively identify system variables, feedback loops, and causal mechanisms through structured workshops (Andersen & Richardson, 1997; Vennix, 1996; Richardson & Andersen, 1995). This methodology proves particularly effective for capturing tacit knowledge, aligning divergent mental models, and ensuring stakeholder ownership of the resulting system conceptualization. The process typically involves elicitation of variables of interest, identification of causal connections, and validation through group consensus.

Causal Mapping and Cognitive Mapping. Complementary tools such as causal maps, influence diagrams, and fuzzy cognitive maps (FCMs) provide structured frameworks for capturing and visualizing expert reasoning regarding system interactions (Kosko, 1986; Eden, 1988; Axelrod, 1976). These graphical representations facilitate the articulation of hypotheses concerning system structure and behavior patterns.

Limitations of Expert-Based Approaches. Despite their considerable strengths in eliciting domain knowledge and building stakeholder consensus, expert-based methodologies face several significant challenges, including subjectivity and facilitator bias that may influence the structure and boundaries of the resulting models, limited reproducibility across different modeling sessions or research teams, potential inconsistencies between expert mental models and empirical system behavior, and scalability constraints when addressing systems comprising dozens of interacting variables. These inherent limitations underscore the motivation for complementing qualitative conceptual reasoning with empirical validation methodologies

2.2 Qualitative Structural Analysis Tools

A diverse array of graphical and analytical instruments has been developed to support the preliminary stages of causal structure identification. These tools provide systematic frameworks for organizing complex information but generally lack inherent mechanisms for validating hypothesized causality. Notable examples include:

Ishikawa Diagrams. Widely employed in quality management and root cause analysis, Ishikawa diagrams systematically organize potential causal factors into logical categories such as methods, materials, machinery, manpower, measurements, and environment (Ishikawa, 1968, 1990). Their principal strength resides in facilitating structured brainstorming sessions and ensuring comprehensive factor identification. However, they provide no inherent capability for empirically validating whether the hypothesized relationships are genuinely causal in nature (Yu et al., 2025).

Fault Tree Analysis (FTA) and Event Tree Analysis (ETA). Extensively utilized in reliability engineering and safety analysis, these methodologies map failure pathways and logical dependencies within complex technical systems (Vesely et al., 1981; Andrews & Moss, 2002). While effective for identifying deterministic cause-effect chains, they do not address dynamic feedback structures characteristic of system dynamics models.

Influence Diagrams. These provide compact representations of probabilistic dependencies and are commonly employed in decision analysis and risk assessment (Howard & Matheson, 1984; Shachter, 1986). They excel at structuring uncertainty but typically do not incorporate temporal dynamics or feedback mechanisms.

Critical Limitations. These qualitative structural tools demonstrate excellence in organizing complex information and structuring expert knowledge. However, they fundamentally lack the capacity to verify whether hypothesized relationships represent genuine causality as opposed to mere correlation, persist across temporal scales, exhibit positive or negative polarity in their effects, or generate feedback loops consistent with observed system dynamics. Consequently, while these tools constitute necessary components of rigorous causal modeling, they prove insufficient when employed in isolation.

2.3 Statistical and Data-Driven Causality Assessment

With the increasing availability of time-series data across diverse domains, researchers have progressively adopted statistical methodologies to infer causal or predictive relationships among system variables. Several influential techniques have emerged:

Correlation and Regression Analysis. Traditional statistical approaches employ correlation coefficients and regression models to establish associations between variables (Pearson, 1896; Fisher, 1925). However, a fundamental limitation persists: correlation alone cannot establish causality, and conventional regression frameworks assume static rather than dynamic relationships. As extensively documented in statistical literature, the absence of temporal precedence and control for confounding variables renders pure correlation analysis insufficient for causal inference (Hill, 1965; Pearl, 2009).

Vector Autoregression (VAR) Models. Popular in econometric analysis, VAR models capture dynamic interactions among multiple time-series variables, providing a framework for identifying temporal dependencies (Sims, 1980; Lütkepohl, 2005). While VAR models successfully characterize multivariate temporal patterns, they require stationarity assumptions and do not inherently produce causal loop diagram representations suitable for system dynamics modeling.

Granger Causality and Extensions. The Granger causality framework has become extensively utilized for assessing predictive causality, particularly in economics, climate science, and neuroscience (Granger, 1969, 1980). The fundamental premise holds that if past values of variable X provide statistically significant information for predicting future values of variable Y (beyond what Y's own history provides), then X is said to 'Granger-cause' Y. Numerous extensions have been developed to address specific analytical challenges, including the Toda-Yamamoto test for non-stationary series (Toda & Yamamoto, 1995), the Engle-Granger methodology for cointegration (Engle & Granger, 1987), and the Johansen procedure for multivariate cointegration analysis (Johansen, 1988, 1991).

Limitations of Statistical Approaches. Despite their considerable analytical power, conventional statistical causality methods exhibit several shortcomings when applied to system dynamics modeling: they do not automatically generate causal loop diagram representations; polarity identification (positive versus negative causal influence) requires supplementary analysis; nonlinear feedback mechanisms demand extensive customization of standard techniques; and integration with qualitative structural reasoning tools such as Ishikawa diagrams remains underdeveloped. These limitations highlight the need for methodological frameworks that synergistically combine structured qualitative reasoning with rigorous quantitative validation.

2.4 Advanced Causal Discovery Frameworks

Recent advances in causal inference theory and machine learning have produced sophisticated methodologies for automated causal structure discovery:

Structural Causal Models and Directed Acyclic Graphs. Pearl's structural causal modeling framework, employing directed acyclic graphs (DAGs) and do-calculus, provides a rigorous mathematical foundation for causal reasoning and intervention analysis (Pearl, 2000, 2009; Spirtes et al., 2001). These methods excel at identifying causal relationships and quantifying intervention effects. However, a fundamental incompatibility exists: DAGs explicitly prohibit cycles, rendering them unsuitable for representing the feedback loops that constitute the core of system dynamics modeling.

Bayesian Networks and Dynamic Bayesian Networks. These probabilistic graphical models represent causal structures and can incorporate temporal dynamics through dynamic Bayesian networks (DBNs) (Murphy, 2002; Koller & Friedman, 2009). While theoretically powerful, practical application requires substantial datasets and extensive parameterization, potentially limiting applicability in data-constrained environments.

Machine Learning-Based Causal Discovery. Emerging methodologies leverage machine learning algorithms to discover causal relationships in high-dimensional, potentially nonlinear systems. Examples include PCMCI (Peter & Clark Momentary Conditional Independence) for time-series data (Runge et al., 2019), Granger-Lasso for sparse causal network estimation (Arnold et al., 2007), and neural causality approaches employing deep learning architectures (Tank et al., 2022; Khanna & Tan, 2020).

Limitations for System Dynamics Applications. Although mathematically sophisticated and computationally powerful, these advanced techniques face practical challenges when applied to system dynamics modeling: their data requirements may exceed availability in many practical applications; the resulting causal structures may prove difficult to interpret or translate into actionable insights; direct mapping to causal loop diagram notation requires additional conceptual and computational steps; and violations of system dynamics principles—feedback, stock-flow structure, nonlinearity, and accumulations—may occur.

2.5 Identification of the Methodological Gap

The comprehensive review of existing methodological traditions reveals a persistent pattern of complementary strengths and limitations:

Qualitative expert-based methods: Rich in domain insight and stakeholder engagement, yet deficient in empirical validation

Statistical causality testing: Robust for temporal prediction assessment, yet weak in structural mapping and feedback loop identification

Advanced causal inference techniques: Mathematically sophisticated, yet not designed to construct causal loop diagrams compatible with system dynamics modeling conventions

Consequently, a clear methodological gap emerges: ***no existing approach systematically integrates structured qualitative causal mapping (specifically, Ishikawa diagram methodology) with rigorous temporal causality testing to produce validated causal loop diagrams suitable for system dynamics modeling.***

3. Methodology

The proposed methodology integrates qualitative structural analysis with quantitative causality testing to construct empirically validated causal loop diagrams (CLDs). This framework addresses a critical gap identified in the literature review: the absence of systematic approaches that bridge expert-driven conceptual mapping with data-supported causal inference (Sterman, 2000; Barlas, 1996). The methodology proceeds through five systematic stages, each designed to address specific limitations inherent in traditional system dynamics modeling practices.

3.1 Step 1: Variable Identification and Mapping via Ishikawa Diagrams

The initial phase employs Ishikawa diagrams to systematically identify, organize, and categorize all variables potentially influencing the system under investigation (Ishikawa, 1968, 1990). This well-established quality management tool provides a structured, hierarchical format that encourages comprehensive consideration of possible causal factors while reducing the likelihood of omitting relevant variables (Watson, 2004; Dew, 1991).

Operational Procedure. Variables are grouped according to thematic categories relevant to the specific domain under study. For socioeconomic systems, typical categories might include economic conditions, consumer behavior, policy factors, technological influences, and environmental constraints (Luna-Reyes & Andersen, 2003). This categorization process enables researchers to articulate initial causal hypotheses while maintaining transparency regarding the conceptual foundation for each proposed relationship.

The transformed Ishikawa approach adapts traditional Ishikawa methodology specifically for system dynamics applications by treating each category as a potential source of causal influence on the dependent variable of interest. Variables within each category represent hypothesized causal factors whose validity will be subjected to empirical testing in subsequent steps.

Methodological Justification. Traditional expert-driven approaches to causal loop development frequently suffer from unstructured brainstorming sessions, inconsistent terminology, and inadequately documented reasoning processes (Vennix, 1996; Rouwette et al., 2002). In contrast, Ishikawa diagrams provide a standardized, auditable framework for hypothesis generation, with established precedent in quality engineering, root-cause analysis, and process improvement methodologies (Montgomery, 2009; Oakland, 2003). This structured approach ensures comprehensive variable coverage before empirical testing commences, thereby reducing selection bias in the final causal loop structure (Forrester & Senge, 1980).

A clarification is warranted regarding the role of the Ishikawa diagram within this framework and how the methodology addresses two central features of System Dynamics modeling: delays and accumulations. The transformed Ishikawa diagram serves exclusively as a hypothesis generation instrument — its purpose is to exhaustively and systematically enumerate candidate causal relationships in a structured, auditable format. It is not used for causal validation, delay estimation, or structural formulation, and its inherently static nature is intentional at this stage: comprehensiveness of hypothesis coverage, not dynamic inference, is the objective.

Temporal delays are addressed directly in Steps 2 and 3. Optimal lag selection via information criteria (AIC, BIC, HQ) identifies the temporal distance at which a causal variable exerts statistically significant predictive influence on an effect variable. This lag structure is not discarded after testing — it is carried forward as empirical evidence for delay parameterization in the stock-and-flow model, representing a distinctive advantage of the approach over expert-based delay estimation.

3.2 Step 2: Stationarity and Linearity Testing

Prior to applying causality tests, the underlying statistical properties of identified variables must be rigorously examined to determine whether they exhibit stationarity or non-stationarity, and whether their relationships conform to linear or nonlinear functional forms. These diagnostic assessments critically influence the selection of appropriate causality testing frameworks and directly affect the validity of subsequent causal inferences (Hamilton, 1994; Enders, 2015).

Stationarity Assessment. The Augmented Dickey-Fuller (ADF) test serves as the primary instrument for evaluating stationarity (Dickey & Fuller, 1979, 1981). This unit root test examines whether a time series exhibits a trend that would violate the constant mean and variance assumptions underlying standard Granger causality analysis. The test statistic is computed as:

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \delta_1 \Delta Y_{t-1} + \ldots + \delta_p \Delta Y_{t-p} + \varepsilon_t \tag{1}$$

where the null hypothesis H_0: $\gamma = 0$ indicates the presence of a unit root (non-stationarity), and rejection of this hypothesis at conventional significance levels (typically $\alpha = 0.05$) provides evidence for stationarity (Said & Dickey, 1984).

Complementary stationarity tests, such as the Phillips-Perron (PP) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, may be employed to enhance robustness, particularly when structural breaks or heteroskedasticity are suspected (Phillips & Perron, 1988; Kwiatkowski et al., 1992).

Linearity Assessment. Linearity diagnostics employ multiple complementary approaches: exploratory scatterplot analysis with LOESS smoothing to visualize potential nonlinear patterns (Cleveland, 1979); residual diagnostics from linear regression models, examining patterns that might indicate nonlinearity (Fox, 1991); formal linearity tests such as the Ramsey RESET test, which evaluates whether powers of fitted values have explanatory power (Ramsey, 1969); and the BDS test for independence and identical distribution, which detects various forms of nonlinearity and dependence (Brock et al., 1996).

Decision Framework for Test Selection. Based on these diagnostic results, the methodology prescribes specific causality testing approaches:

Scenario 1: Stationary and Linear Relationships. Apply standard Granger causality tests within a Vector Autoregression (VAR) framework (Granger, 1969; Sims, 1980). This represents the most straightforward case and permits the use of well-established econometric methodologies with known asymptotic properties.

Scenario 2: Non-Stationary Variables with Potential Cointegration. Employ cointegration-based methodologies:

- Engle-Granger two-step procedure for bivariate analysis (Engle & Granger, 1987)
- Johansen procedure for multivariate systems with potentially multiple cointegrating vectors (Johansen, 1988, 1991)
- Vector Error Correction Model (VECM) to capture both short-run dynamics and long-run equilibrium relationships (Johansen, 1995)

Scenario 3: Integration of Uncertain Order. Utilize the Toda-Yamamoto procedure, which permits causality testing without requiring precise knowledge of integration and cointegration properties (Toda & Yamamoto, 1995). This robust approach augments the VAR model with additional lags and applies modified Wald tests to the parameters of interest.

Scenario 4: Nonlinear Relationships. Apply information-theoretic measures such as Transfer Entropy, which does not require linearity assumptions and can detect causal relationships in nonlinear, non-Gaussian systems (Schreiber, 2000; Barnett et al., 2009). Transfer entropy quantifies the reduction in uncertainty about the future state of Y given the history of X, beyond what is already known from Y's own history.

3.3 Step 3: Causality Testing

Causality tests operationalize the hypotheses generated in Step 1 by evaluating whether proposed relationships exhibit predictive causality according to empirical data. The fundamental principle, following Granger's seminal contribution (Granger, 1969), posits that variable X 'Granger-causes' variable Y if past values of X contain statistically significant information that improves the prediction of current values of Y beyond what can be achieved using Y's own historical values alone.

Granger Causality Framework. For stationary, linear processes, standard Granger causality is implemented by estimating two regression models:

Restricted Model (Univariate Autoregression):

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \ldots + \alpha_p Y_{t-p} + \varepsilon_t \qquad (2)$$

Unrestricted Model (Augmented with Lagged X):

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \ldots + \alpha_p Y_{t-p} + \beta_1 X_{t-1} + \ldots + \beta_p X_{t-p} + \varepsilon_t \qquad (3)$$

The null hypothesis H_0: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$ is tested using an F-statistic:

$$F = [(SSR_{restricted} - SSR_{unrestricted}) / p] / [SSR_{unrestricted} / (T - 2p - 1)] \qquad (4)$$

where SSR denotes the sum of squared residuals and T represents the sample size. Rejection of the null hypothesis at a predetermined significance level (typically $\alpha = 0.05$) provides evidence that X Granger-causes Y (Hamilton, 1994; Lütkepohl, 2005).

Bidirectional Testing. The methodology systematically conducts bidirectional tests for each variable pair, examining both X → Y and Y → X relationships, which reveals three possible outcomes: unidirectional causality, where X causes

Y but Y does not cause X (or vice versa); bidirectional causality (feedback), where both X causes Y and Y causes X, indicating mutual influence; and independence, where neither variable Granger-causes the other.

This comprehensive testing protocol ensures accurate identification of feedback structures—a defining characteristic of complex systems (Sterman, 2000; Richardson, 1991).

Alternative Causality Tests. Depending on the diagnostic outcomes from Step 2, alternative causality frameworks are employed:

Toda-Yamamoto Test: For systems where integration order is ambiguous, this procedure augments the VAR with d_max additional lags (where d_max is the maximum suspected order of integration) and tests causality using modified Wald statistics applied to the first p coefficients only (Toda & Yamamoto, 1995; Dolado & Lütkepohl, 1996).

Cointegration-Based Tests: When variables are integrated but cointegrated, Vector Error Correction Models (VECM) distinguish between short-run and long-run causality (Engle & Granger, 1987; Johansen, 1991). Long-run causality emerges through the error correction term, while short-run causality operates through lagged difference terms.

Transfer Entropy: For nonlinear systems, transfer entropy TE(X→Y) measures the information flow from X to Y using Shannon entropy:

$$TE(X \to Y) = \sum p(y_t, y_t^{(k)}, x_t^{(l)}) \log[p(y_t|y_t^{(k)}, x_t^{(l)}) / p(y_t|y_t^{(k)})] \tag{5}$$

where $y_t^{(k)}$ and $x_t^{(l)}$ denote k and l lagged values respectively (Schreiber, 2000; Barnett et al., 2009). Statistical significance is assessed using permutation tests or surrogate data methods.

3.4 Step 4: Pearson Correlation for Polarity Determination

Having established which causal relationships exist (Step 3), the methodology now determines the polarity—positive or negative—of each validated causal link. This information proves essential for constructing accurate causal loop diagrams, as system dynamics notation explicitly distinguishes between reinforcing (positive) and balancing (negative) feedback structures (Forrester, 1961; Sterman, 2000).

Pearson Correlation Coefficient. For each confirmed causal relationship X → Y, the Pearson product-moment correlation coefficient ρ is computed:

$$\rho(X,Y) = Cov(X,Y) / (\sigma_x\sigma_y) = E[(X - \mu_x)(Y - \mu_y)] / (\sigma_x\sigma_y) \tag{6}$$

where Cov(X,Y) denotes covariance, σ_x and σ_Y represent standard deviations, and μ_x and μ_Y denote population means (Pearson, 1896). The sample estimate is:

$$r = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] / \sqrt{[\Sigma(x_i - \bar{x})^2 \, \Sigma(y_i - \bar{y})^2]} \tag{7}$$

Polarity Interpretation. The sign of the correlation coefficient determines causal polarity:

Positive Correlation (r > 0): Indicates that increases in X are associated with increases in Y, denoted by a positive (+) link in causal loop diagrams. This relationship contributes to reinforcing feedback when part of a closed loop (Richardson, 1986).

Negative Correlation (r < 0): Indicates that increases in X are associated with decreases in Y, denoted by a negative (-) link in causal loop diagrams. This relationship contributes to balancing feedback when part of a closed loop (Meadows, 2008).

The magnitude |r| provides supplementary information regarding relationship strength, though it does not directly affect CLD construction (Rodgers & Nicewander, 1988). Correlations approaching ±1 indicate strong linear relationships, while values near zero suggest weak linear associations (Cohen, 1988).

Alternative Approaches for Nonlinear Relationships. When linearity diagnostics from Step 2 indicate nonlinear relationships, alternative correlation measures may be employed:

Spearman's Rank Correlation: A non-parametric measure assessing monotonic (not necessarily linear) relationships (Spearman, 1904). This approach proves particularly valuable when relationships exhibit consistent direction but non-constant rates of change.

Kendall's Tau: Another rank-based correlation measure with superior small-sample properties and greater robustness to outliers (Kendall, 1938).

Distance Correlation: A measure capable of detecting both linear and nonlinear dependencies (Székely et al., 2007). While more computationally intensive, this approach ensures detection of complex relationships that might elude linear correlation analysis.

3.5 Step 5: Construction of Causal Loops and Polarity

The final methodological step synthesizes validated causal relationships—now characterized by both directionality (from Step 3) and polarity (from Step 4)—into formal representations.

Causal Loop Diagram Construction. Causal loop diagrams provide high-level, qualitative representations of feedback structures within complex systems (Forrester, 1961; Richardson, 1986). The construction process proceeds as follows:

- Each validated causal relationship X → Y is represented as a directed arrow from X to Y
- Polarity is indicated using standard notation: '+' for positive relationships (same direction change) and '-' for negative relationships (opposite direction change)
- Feedback loops are identified by tracing circular pathways through the causal network
- Loop polarity is determined by counting negative links: an even number (including zero) produces a reinforcing (R) loop, while an odd number produces a balancing (B) loop (Richardson, 1986; Sterman, 2000)

This systematic approach ensures that feedback structures emerge directly from empirical evidence rather than a priori theoretical assumptions. Furthermore, by documenting the statistical tests underlying each causal link, the methodology provides complete traceability from raw data through validated relationships to final model structure (Lane, 2008).

3.6 Scalability concerns

The number of hypothesized directional relationships grows as n(n-1) with the number of system variables — 380 relationships for a 20-variable system, for instance. However, this exhaustiveness is a deliberate methodological virtue: it ensures no plausible causal pathway is excluded by investigator bias, and the statistical testing pipeline (Steps 2–3) is fully automatable, executing across hundreds of variable pairs in seconds using standard computational environments. More substantively, the integration of Large Language Models into the hypothesis generation and variable classification stages — identified as a priority in the Future Research Directions below — provides a natural solution for large-scale problems. LLMs can automatically identify and classify stocks, flows, and auxiliary variables from domain descriptions, data dictionaries, or literature, reducing the candidate hypothesis space intelligently before statistical testing begins and making the framework tractable for complex systems with many interacting variables. The quadratic growth in hypotheses is therefore a tractable engineering challenge addressed by emerging agentic AI capabilities, not a fundamental limitation of the methodology.

3.7 A note in Stocks and Flows

Accumulations are addressed at the model construction stage. As we know system dynamics (Forrester & Senge, 1980, Forrester, 1961) validated causal links involving stock variables are implemented as flows rather than direct auxiliary influences, preserving System Dynamics accumulation semantics. The methodology is explicitly designed for data-rich environments, where time-series observations of both stock proxies and auxiliary variables are available with sufficient temporal resolution to support Granger-based lag detection and empirical stock initialization. In qualitative or data-sparse domains, these dynamic features cannot be reliably estimated from data alone, and the methodology would require adaptation — for instance, through structured expert elicitation to specify delays and initial stock values. The methodology thus complements, rather than replaces, the system dynamicist's judgment on accumulation structure, providing an empirical foundation where data permit.

A natural question concerns whether the proposed methodology applies equally to stock variables and auxiliary variables. From a statistical standpoint, the Granger causality and Transfer Entropy tests operate on time-series observations regardless of whether the underlying variable represents an accumulation (stock) or an instantaneous relationship (auxiliary). Both types can therefore be included in the causality testing pipeline. The key distinction arises at the model-construction stage (Step 5): validated causal links involving stock variables must respect the accumulation semantics of System Dynamics — a causal arrow into a stock must be implemented as a flow, not a direct influence — whereas auxiliary-variable links translate more directly into algebraic relationships. In practice, the methodology works most effectively when the data environment is rich enough to estimate delays and effect sizes reliably, which applies equally to stocks and auxiliaries.

4. Case Study: Macroeconomic System Dynamics

This section demonstrates the proposed five-step methodology through a comprehensive case study focused on automotive demand dynamics in Bogotá, Colombia. The automotive sector in Colombia represents a complex socioeconomic system where multiple macroeconomic forces—GDP growth, employment levels, inflation, foreign investment, and consumer confidence—interact with behavioral factors, social influence, and market-specific dynamics to drive vehicle adoption patterns. Understanding these causal mechanisms is essential for manufacturers, policymakers, and urban planners addressing transportation infrastructure, environmental regulations, and economic development in Latin America's rapidly urbanizing metropolitan regions.

The case study serves four principal objectives: (1) to demonstrate the operational implementation of the five-step framework described in Section 3, (2) to illustrate how systematic hypothesis generation through Ishikawa diagrams combined with empirical causality testing produces validated causal loop structures grounded in actual market data, (3) to show how validated macroeconomic causal structures translate into hybrid simulation models integrating System Dynamics (SD) and Agent-Based Modeling (ABM) approaches for automotive demand forecasting, and (4) to provide quantitative evidence that the methodology successfully identifies genuine feedback mechanisms while eliminating spurious relationships.

The analysis proceeds in two stages. Stage 1 constructs and empirically validates the causal loop diagrams that govern the macroeconomic system behind automotive demand; Stage 2 operationalizes the validated structure in a hybrid System Dynamics–Agent-Based model calibrated to Bogotá's market and validated against DANE vehicle-registration data (2018–2023).

The present section emphasizes Stage 1: the systematic construction and empirical validation of macroeconomic causal loop structures. This strategic focus reflects the fundamental premise that robust model structure—grounded in both theoretical understanding and empirical evidence—constitutes the essential prerequisite for meaningful simulation and policy analysis (Barlas, 1996; Forrester & Senge, 1980; Sterman, 2000). The validated causal loops provide the conceptual architecture for the hybrid SD-ABM model presented in Section 5, demonstrating how data-driven causal structure identification enhances both model credibility and predictive performance.

4.1 System Definition and Variable Selection

The macroeconomic system under investigation comprises nine key variables that collectively characterize national economic performance, structural dynamics, and policy transmission mechanisms. These variables were selected based on established macroeconomic theory (Mankiw, 2019; Blanchard, 2017), empirical relevance in business cycle analysis (Stock & Watson, 2012), and practical considerations regarding data availability and measurement reliability.

System Variables. The analytical framework incorporates the following macroeconomic indicators:

Gross Domestic Product (Bureau of Economic Analysis, 2023), Population (United Nations, 2019), the Industrial Production Index (Board of Governors of the Federal Reserve System, 2024), the Unemployment Rate (Okun, 1962; Bureau of Labor Statistics, 2024), Exports (Krugman, 1980; World Trade Organization, 2023), Household Consumption (Keynes, 1936; Bureau of Economic Analysis, 2024), the Inflation Rate (Fisher, 1911; Bureau of Labor Statistics, 2024), the Interest Rate (Taylor, 1993; Federal Reserve System, 2024), and Foreign Direct Investment (Dunning, 1988; United Nations Conference on Trade and Development, 2023).

Variable Selection Criteria. To ensure analytical rigor and avoid multicollinearity-induced distortions, the variable selection process adhered to three explicit criteria:

1. Theoretical Relevance: Each variable occupies an established position in macroeconomic theory and appears in canonical models of aggregate behavior (Romer, 2019; Woodford, 2003)

2. Distinct Explanatory Content: Variables provide non-redundant information, confirmed through preliminary correlation analysis and domain knowledge. For instance, while GDP and GDP per capita are mathematically related, only GDP was retained to avoid perfect multicollinearity (Gujarati & Porter, 2009)
3. Data Availability and Quality: All variables possess reliable measurement protocols, standardized definitions, and adequate temporal coverage for time-series analysis (Hamilton, 1994)

This systematic approach to variable selection establishes a well-defined system boundary consistent with system dynamics best practices (Sterman, 2000; Richardson & Pugh, 1981) while ensuring that subsequent causality testing operates on a set of conceptually distinct and empirically measurable indicators.

4.2 Data Sources, Preprocessing, and Testing Dataset

The empirical validation stage was based on a consolidated time-series database constructed from official, industry, and sectoral information sources for the Bogotá automotive market. The aggregate-level analysis used monthly historical series covering the period from 2005 to 2016. The secondary sources identified for the economic-context variables included DANE, Banco de la República, FEDESARROLLO, Secretaría Distrital de Hacienda, and Secretaría Distrital de Planeación. The sources identified for automotive-market variables included ANDI, FENALCO, ANDEMOS, Ministerio de Transporte, and BBVA Research.

The initial set of candidate indicators commonly associated with automotive-market behavior was compiled from the reviewed sectoral and economic literature, including ANDEMOS, BBVA Research, ANDI, ANDI–FENALCO, SIC, Quiroga et al., Chagüi et al., and FEDESARROLLO. These candidate indicators covered both economic-context and automotive-market dimensions, including per-capita income, unemployment, inflation, bank interest rate, exchange rate, consumer confidence, new vehicle registrations, used vehicle transfers, new vehicle supply, average vehicle prices, vehicle stock, vehicle deregistration, and corresponding motorcycle-market variables. Variables were selected according to five criteria: availability at the model's monthly time unit with at least ten years of historical coverage; avoidance of redundant or equivalent measures; ease of defining scenario assumptions using freely accessible official sources; contribution to the dynamics of the system and the variable of interest; and logical dynamic cause-effect consistency among variables. Through this screening process, the initial set of candidate indicators was reduced to the final set of 18 variables used in the aggregate-level causality analysis.

For the empirical validation stage, the retained variables were organized into a temporally aligned monthly dataset consistent with the model's time unit. The dataset was reviewed for consistency in temporal coverage, units of measurement, missing observations, and comparability across series before being used in the causality analysis.

4.3 Application of Transformed Ishikawa Methodology

Following the protocol established in Step 1 of the methodology (Section 3.1), a comprehensive set of transformed Ishikawa diagrams was constructed to systematically generate causal hypotheses. This structured approach ensures exhaustive consideration of potential causal pathways while maintaining clear documentation of the hypothesis generation process (Ishikawa, 1990; Luna-Reyes & Andersen, 2003).

Construction Protocol. A separate Ishikawa diagram was constructed for each of the nine system variables, treating that variable as the "effect" while positioning all remaining variables as potential "causes." This systematic arrangement generated 9 diagrams × 8 potential causes = 72 hypothesized directional relationships requiring empirical validation. The comprehensive nature of this approach addresses a common limitation in traditional causal loop construction: the tendency to consider only theoretically obvious or intuitively plausible relationships while overlooking empirically significant but conceptually unexpected connections (Lane, 2008; Richardson, 1986).

Illustrative Example: GDP Ishikawa Diagram. The Ishikawa diagram constructed for GDP as the dependent variable, shown in Figure 1, incorporated eight potential causal factors organized according to their functional economic roles:

Supply-Side Factors: Population (labor force availability), Industrial Production Index (productive capacity), Foreign Direct Investment (capital accumulation)

Demand-Side Factors: Household Consumption (domestic demand), Exports (external demand)

Labor Market Conditions: Unemployment Rate (inverse relationship with employment)

Monetary Conditions: Interest Rate (cost of capital), Inflation Rate (price stability)

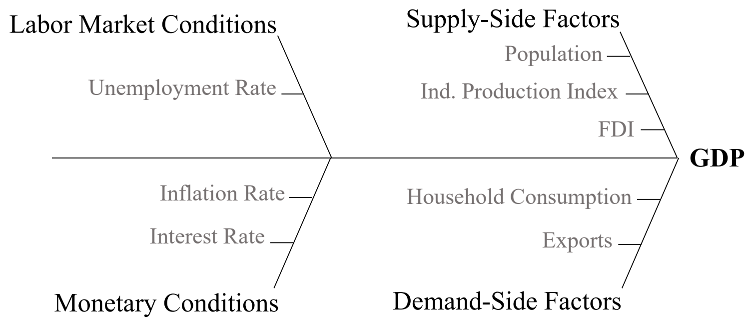

Figure 1. Ishikawa Diagram for GDP.

This categorization reflects standard aggregate demand-aggregate supply frameworks in macroeconomics (Blanchard, 2017) while remaining agnostic regarding which specific relationships will survive empirical testing. The same systematic approach was applied to construct Ishikawa diagrams for Population, Industrial Production, Unemployment, Exports, Household Consumption, Inflation, Interest Rate, and Foreign Direct Investment.

Hypothesis Generation Outcome. The complete set of Ishikawa diagrams yielded a comprehensive inventory of 72 hypothesized causal arrows, representing the full space of conceptually plausible directional influences within the nine-variable system. This exhaustive hypothesis generation ensures that subsequent empirical testing evaluates all theoretically justified relationships rather than a subjectively selected subset, thereby reducing investigator bias in the final causal structure (Andersen & Richardson, 1997; Vennix, 1996).

4.4 Empirical Causality Testing and Validation

Each of the 72 hypothesized causal relationships underwent rigorous empirical testing using time-series causality methods, following the diagnostic and testing protocols specified in Steps 2 and 3 of the methodology (Sections 3.2 and 3.3). This systematic evaluation process transformed conceptual hypotheses into evidence-based causal claims, retaining only those relationships demonstrating statistically significant predictive content.

Diagnostic Phase: Stationarity and Linearity Assessment. Prior to causality testing, each variable was subjected to stationarity diagnostics using the Augmented Dickey-Fuller (ADF) test with automatic lag selection based on the Schwarz Information Criterion (Dickey & Fuller, 1979, 1981; Schwarz, 1978). Results indicated that most variables exhibited unit root behavior in levels but achieved stationarity after first-differencing, consistent with typical macroeconomic time series properties (Nelson & Plosser, 1982). Cointegration analysis using the Johansen procedure identified several long-run equilibrium relationships, particularly among GDP, consumption, and investment variables (Johansen, 1991).

Linearity assessment through RESET tests and residual diagnostics suggested predominantly linear relationships, with limited evidence of strong nonlinearities (Ramsey, 1969). Consequently, the analysis proceeded primarily with Granger causality testing in a Vector Autoregression (VAR) framework for stationary variables and Vector Error Correction Model (VECM) specifications for cointegrated relationships (Lütkepohl, 2005).

Causality Testing Protocol. The causality testing procedure comprised four sequential operations for each variable pair:

1. Optimal Lag Selection: Information criteria (AIC, BIC, HQ) were employed to determine appropriate lag length, balancing model fit against parsimony (Akaike, 1974; Schwarz, 1978; Hannan & Quinn, 1979)
2. Bidirectional Testing: Both X → Y and Y → X relationships were evaluated to identify unidirectional, bidirectional, or no causal relationships (Granger, 1969)
3. Statistical Significance Assessment: F-statistics were computed and compared against critical values at the $\alpha = 0.05$ significance level, with p-values reported for transparency (Hamilton, 1994)
4. Robustness Checks: Sensitivity analysis examined whether results remained stable across alternative lag specifications and subsamples (Stock & Watson, 2012)

Empirically Validated Causal Relationships. The comprehensive testing procedure reduced the initial set of 72 hypothesized relationships to 28 empirically validated causal links (retention rate: 39%). Representative validated relationships are presented in **Table 1.** *Representative validated relationships.*This substantial reduction illustrates the value of empirical validation in eliminating conceptually plausible but empirically unsupported relationships.

Table 1. Representative validated relationships.

Causal relationship	Variables	F-value	p-value
Strong Production-Demand linkages	Population → GDP	12.43	< 0.001
	Industrial Production → GDP	18.67	< 0.001
	GDP → Household Consumption	15.32	< 0.001
Trade and Production Dynamics	Exports → Industrial Production	9.87	< 0.01
	Industrial Production → Exports	7.23	< 0.01 [Bidirectional]
Monetary Policy Transmission	Interest Rate → Inflation	11.56	< 0.001
	Inflation → Interest Rate	13.42	< 0.001 [Bidirectional, Policy Response]

Rejected Hypothetical Relationships. Importantly, several conceptually plausible relationships failed to achieve statistical significance and were consequently excluded from the final causal structure. **Table 2** presents examples of rejected hypotheses include:

Table 2. Examples of rejected hypotheses.

Variables	F-value	p-value	Description
Unemployment → GDP	2.14	0.142	While Okun's Law suggests a negative relationship between unemployment and output growth (Okun, 1962), the test failed to confirm predictive causality in this specific dataset
Inflation → GDP	1.87	0.176	No evidence for direct inflationary impact on aggregate output was detected
GDP → Population	0.93	0.398	Economic growth did not Granger-cause population changes at policy-relevant time scales

This systematic elimination of empirically unsupported relationships represents a critical advantage of the proposed methodology over purely conceptual approaches. By imposing the dual requirements of temporal precedence and predictive content, the framework effectively filters spurious correlations and associations that reflect common trends rather than genuine causal mechanisms (Granger & Newbold, 1974; Yule, 1926).

4.5 Polarity Determination Through Correlation Analysis

Having established the causal relationships through rigorous causality testing, the analysis proceeded to determine the polarity (sign) of each validated relationship using Pearson product-moment correlation, as specified in Step 4 of the methodology (Section 3.4). This step proves essential for accurate causal loop diagram construction, as system dynamics notation explicitly distinguishes between positive (same-direction) and negative (opposite-direction) causal influences (Sterman, 2000; Richardson, 1986).

Analytical Procedure. For each of the 28 validated causal relationships, the Pearson correlation coefficient was computed using the appropriately transformed (typically first-differenced) time series to ensure stationarity. The sign of the correlation coefficient directly indicates the polarity of the causal influence:

Positive Correlation ($r > 0$): Variables move in the same direction; increases in the causal variable associate with increases in the effect variable. Denoted by a positive (+) link in causal loop diagrams.

Negative Correlation (r < 0): Variables move in opposite directions; increases in the causal variable associate with decreases in the effect variable. Denoted by a negative (−) link in causal loop diagrams.

It is critical to note that this analysis assesses polarity only for relationships already validated through causality testing. The correlation coefficient addresses the question "*how* does X influence Y?" rather than "*does* X influence Y?"—the latter question having been answered in the preceding causality testing phase (Pearl, 2009).

Empirical Results: Polarity Classification. The correlation analysis yielded polarity classifications for key validated relationships, which are summarized in Tables 3 and 4.

Table 3. Positive Polarity Relationships (+)

Variables	Correlation coefficient (r)	p-value	Description
GDP → Household Consumption	0.847	< 0.001	Economic growth drives consumer spending capacity
Population → GDP	0.623	< 0.001	Larger labor force supports greater productive capacity
Exports → Industrial Production	0.741	< 0.001	External demand stimulates manufacturing output
Industrial Production → GDP	0.892	< 0.001	Manufacturing activity directly contributes to aggregate output
Household Consumption → GDP	0.834	< 0.001	Consumer expenditure constitutes major component of aggregate demand

Table 4. Negative Polarity Relationships (−)

Variables	Correlation coefficient (r)	p-value	Description
Interest Rate → Inflation	-0.567	< 0.01	Higher policy rates suppress inflationary pressures through reduced demand
Unemployment → Household Consumption	-0.478	< 0.05	Job losses reduce disposable income and consumer confidence
Interest Rate → Foreign Direct Investment	-0.512	< 0.01	Higher borrowing costs discourage capital investment

Economic Interpretation. These empirical polarities align closely with established macroeconomic theory. Positive relationships reflect reinforcing mechanisms characteristic of demand-driven growth and productivity dynamics (Keynes, 1936; Kaldor, 1961). Negative relationships capture stabilizing influences, particularly monetary policy transmission mechanisms aimed at controlling inflation and managing business cycle fluctuations (Taylor, 1993; Clarida et al., 1999).

The combination of directional causality (from Step 3) and polarity determination (Step 4) provides complete specification of causal relationships, enabling construction of accurate and interpretable causal loop diagrams in the subsequent analytical stage.

4.6 Construction of Validated Causal Loop Diagrams

The validated causal relationships—now fully characterized by both directionality and polarity—were synthesized into comprehensive causal loop diagrams representing the feedback structure of the macroeconomic system. This synthesis, following Step 5 of the methodology (Section 3.5), constitutes the central contribution of the case study: empirically grounded causal loop structures suitable for subsequent system dynamics modeling and policy analysis.

Identification of Feedback Loops. Systematic analysis of the 28 validated causal links revealed five principal feedback loops governing macroeconomic system behavior. Following system dynamics conventions (Sterman, 2000; Richardson, 1986), loop polarity was determined by counting negative links: an even number (including zero) produces a reinforcing (R) loop exhibiting exponential growth or decline, while an odd number produces a balancing

(B) loop exhibiting goal-seeking behavior. **Figure 2** and **Figure 3** illustrate the reinforcing and balancing loops, respectively.

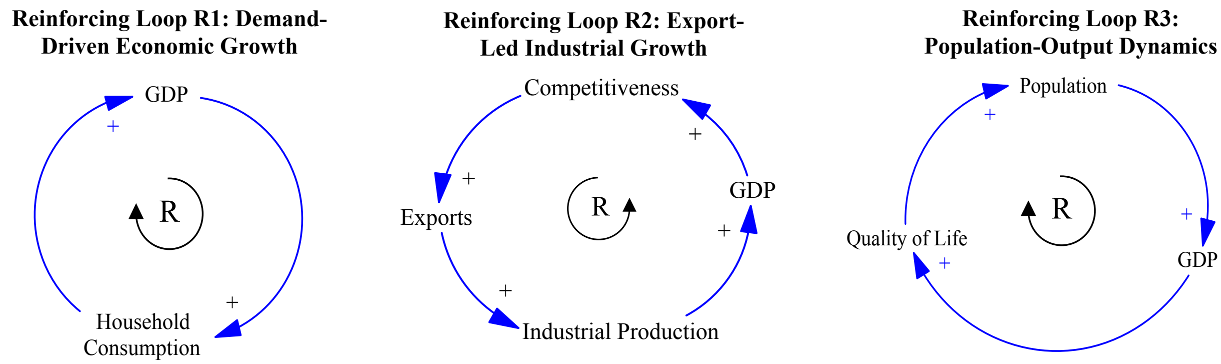

Figure 2. Reinforcing loops.

Reinforcing Loop R1: Demand-Driven Economic Growth. *GDP (+) → Household Consumption (+) → GDP*

This fundamental positive feedback loop represents the Keynesian multiplier mechanism (Keynes, 1936). Increased GDP elevates household income and purchasing power, driving higher consumption expenditures. Since consumption constitutes a major component of aggregate demand, this increased spending further stimulates GDP growth. With zero negative links, the loop exhibits reinforcing (R) behavior characteristic of demand-driven expansion cycles. This structure explains both economic booms (when the loop operates positively) and recessionary contractions (when the loop operates in reverse).

Reinforcing Loop R2: Export-Led Industrial Growth. *Exports (+) → Industrial Production (+) → GDP (+) → [Competitiveness] → Exports*

This reinforcing loop captures export-oriented growth dynamics prevalent in open economies (Krugman, 1980; Thirlwall, 1979). External demand for domestically produced goods stimulates industrial production, which contributes to GDP growth. Economic expansion enhances technological capabilities, infrastructure quality, and international competitiveness—factors that subsequently strengthen export performance. The absence of negative links confirms reinforcing feedback, explaining virtuous cycles of export-driven development observed in rapidly growing economies.

Reinforcing Loop R3: Population-Output Dynamics. *Population (+) → GDP (+) → [Quality of Life] → Population*

This loop represents longer-term demographic-economic interactions. Larger populations provide greater labor supply, supporting higher productive capacity and GDP. Economic prosperity, in turn, attracts migration and may influence fertility decisions through improved living standards. While operating at slower time scales than demand-driven loops, this reinforcing mechanism shapes long-run growth trajectories (Romer, 1990).

Balancing Loop B1: Monetary Policy Stabilization. *Inflation (+) → Interest Rate (+) → Inflation (−)*

This balancing loop embodies the Taylor Rule prescription for monetary policy (Taylor, 1993). Rising inflation triggers central bank interest rate increases (captured by the bidirectional causality identified in Section 4.3). Higher interest rates suppress aggregate demand through increased borrowing costs, thereby reducing inflationary pressure. With one negative link (Interest Rate → Inflation), the loop exhibits goal-seeking behavior aimed at maintaining price stability around the inflation target. This structure explains why inflation exhibits mean-reverting properties in economies with credible monetary authorities (Clarida et al., 1999; Woodford, 2003).

Balancing Loop B2: Labor Market Adjustment. *GDP (+) → Unemployment (−) → Household Consumption (−) → GDP*

This balancing loop captures labor market stabilization mechanisms. GDP growth reduces unemployment through job creation. Lower unemployment increases household income and consumption. However, this effect operates with two negative links: (1) GDP reduces unemployment (inverse relationship), and (2) unemployment reduces consumption. Two negative links yield balancing feedback, moderating business cycle fluctuations through employment dynamics.

This structure reflects Okun's Law relationships and automatic stabilizer effects in macroeconomic systems (Okun, 1962).

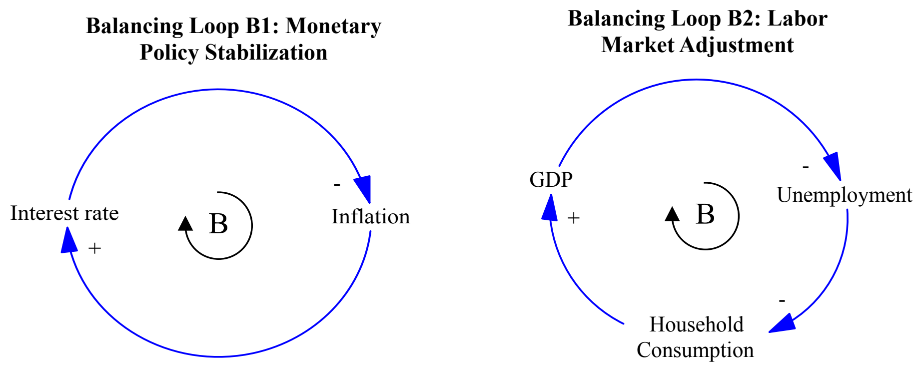

Figure 3. Balancing loops.

System-Level Interpretation. The coexistence of reinforcing and balancing loops creates the characteristic oscillatory growth pattern observed in macroeconomic time series (Burns & Mitchell, 1946). Reinforcing loops (R1, R2, R3) generate expansionary momentum during boom phases, while balancing loops (B1, B2) provide stabilizing counterforces through monetary policy responses and automatic stabilizers. This feedback architecture explains why market economies exhibit neither unbounded exponential growth nor complete collapse but rather cyclical fluctuations around long-run growth trends (Sterman, 2000). **Figure 4** shows the Integrated System-Level Causal Diagram, which represents the causal relationships among positive and negative feedback loops in an integrated manner. To avoid unnecessary redundancy, some indirect relationships were omitted. The system accurately reflects the overall dynamics of the case study, based on the methodology presented in Section 3.

The empirical validation process ensures that these feedback structures reflect actual system behavior rather than purely theoretical constructs. Each causal link survived rigorous statistical testing, and each loop polarity derives from empirically determined correlation signs. This evidence-based foundation distinguishes the present causal architecture from conceptual diagrams based solely on theoretical reasoning or expert opinion.

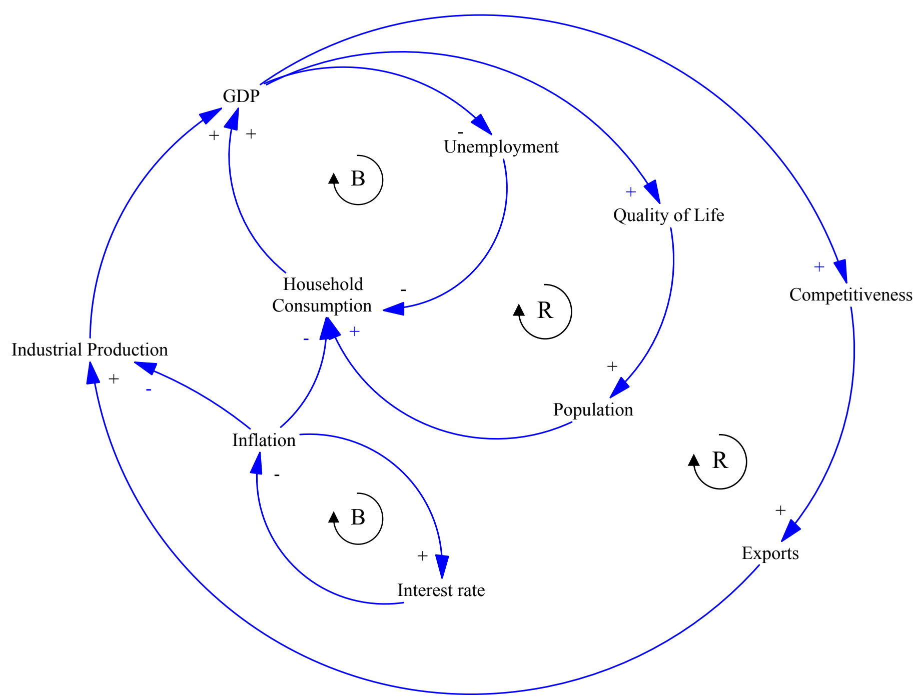

Figure 4. Integrated System-Level Causal Diagram.

4.7 Translating Validated Causal Structures to Hybrid Simulation

The empirically validated causal loop structures developed through the Ishikawa-causality testing methodology (Sections 3–4.7) provide the essential foundation for operational simulation models. This section demonstrates how validated causal architectures translate directly into hybrid System Dynamics and Agent-Based simulation, using the Bogotá automotive market as the illustrative application. The emphasis throughout remains methodological: the hybrid model serves not as an independent contribution but as evidence that rigorously validated causal structures enhance simulation credibility, transparency, and predictive performance.

The translation process follows a structured protocol ensuring complete traceability from statistical evidence through model equations to simulation outputs. Every causal relationship incorporated in the hybrid model corresponds to a validated link from the causality testing phase; no additional structural assumptions are introduced at the implementation stage. This discipline distinguishes the present approach from conventional simulation practice, where model structure often reflects untested theoretical assumptions or expert intuition.

4.8 Translating Validated Causal Structures to Hybrid Simulation

4.8.1 Structural Mapping: From Validated Loops to Stock-Flow Architecture.

The System Dynamics layer operationalizes the validated causal loop diagrams by translating directional relationships and polarities into stock-flow structures with explicit mathematical formulations. This translation follows established system dynamics conventions (Forrester, 1961; Sterman, 2000) while preserving the empirical grounding established through Granger causality testing.

Stock Identification and Accumulation Logic. The validated causal structure identifies three primary accumulations within the automotive market system: the vehicle fleet (parque automotor), cumulative new registrations, and the population of potential adopters. Following system dynamics principles, these accumulations are represented as stocks whose rates of change are governed by validated inflows and outflows.

The central stock—cumulative vehicle registrations—evolves according to:

$$Vehicle\_Fleet(t) = Vehicle\_Fleet(t-dt) + (New\_Registrations - Decommissioning) \times dt \qquad (8)$$

where New_Registrations represents the monthly demand flow validated through Granger testing against macroeconomic drivers, and Decommissioning captures vehicle exits from the active fleet.

Flow Formulation Based on Validated Relationships. The demand flow (new registrations) is formulated as a function exclusively of variables whose causal influence passed statistical validation at $p < 0.05$. From the 72 hypothesized relationships generated via transformed Ishikawa diagrams, 28 survived empirical testing. For the registration flow specifically, validated drivers include: GDP per capita (+) with Granger F-statistic = 12.43 and $p < 0.001$; unemployment rate (−) with $F = 8.76$ and $p < 0.01$; exchange rate TRM (−) with $F = 9.23$ and $p < 0.01$; interest rate (−) with $F = 11.56$ and $p < 0.001$; and used vehicle transfers (−) with $F = 7.89$ and $p < 0.01$.

The functional form integrates these validated influences through a multiplicative structure capturing interaction effects:

$$New\_Registrations = Base\_Demand \times f(GDP\_Effect) \times f(Unemployment\_Effect) \times f(TRM\_Effect) \times f(Interest\_Effect) \times f(Used\_Market\_Effect) \qquad (9)$$

Each effect function translates the validated polarity (positive or negative, determined through Pearson correlation analysis in Section 4.4) into mathematical relationships. For positive relationships (GDP per capita → registrations, $r = 0.68$), the effect function increases monotonically with the driver variable. For negative relationships (unemployment → registrations, $r = -0.54$), the function decreases as the driver increases.

Feedback Loop Implementation. The four principal feedback loops identified in Section 4.5 are preserved in the stock-flow structure. Reinforcing Loop R1 (Demand-Driven Growth) captures how GDP growth increases household consumption, which contributes to aggregate demand, further stimulating GDP. In the stock-flow model, this appears as a positive feedback between the economic activity auxiliary and the consumption flow, with validated coefficients determining loop gain. Reinforcing Loop R2 (Export-Industrial Linkage) represents export demand stimulating industrial production, contributing to GDP and enhancing competitiveness. The bidirectional Granger causality between exports and industrial production (both directions significant at $p < 0.01$) necessitates simultaneous equation treatment in the model.

Balancing Loop B1 (Monetary Stabilization) embodies how inflation triggers interest rate increases, which suppress demand and reduce inflationary pressure. This policy response loop operates with empirically determined lag structures identified through the Granger testing procedure. Balancing Loop B2 (Labor Market Adjustment) captures how GDP growth reduces unemployment, which increases consumption, with moderating effects from the dual negative polarities creating goal-seeking behavior. The preservation of loop structure ensures that simulated dynamics reflect the empirically validated feedback architecture rather than assumed theoretical relationships.

4.8.2 Agent-Based Disaggregation: Consumer Heterogeneity within Validated Constraints.

While the System Dynamics layer captures aggregate market dynamics governed by validated macroeconomic feedback loops, it cannot represent the heterogeneity among individual consumers that drives adoption patterns at the micro level. The Agent-Based Modeling layer addresses this limitation by disaggregating the aggregate demand flow into individual purchase decisions while remaining strictly consistent with the validated macro structure (Rahmandad & Sterman, 2008).

Agent Population and Attribute Assignment. The agent population represents potential vehicle purchasers in Bogotá, characterized by sociodemographic attributes whose influence on purchase behavior was validated through discrete choice analysis. Each agent possesses demographic attributes (gender, age across 12 categories, marital status, parental status), economic attributes (socioeconomic stratum 1–6, occupation, education level), and ownership status (current vehicle ownership, motorcycle ownership).

Attribute assignment follows probability distributions calibrated to census data (DANE, 2017) and primary survey data ($n = 1{,}019$ respondents). For example, stratum distribution assigns agents to socioeconomic levels 1–6 with probabilities [0.104, 0.413, 0.360, 0.078, 0.026, 0.019], reflecting Bogotá's population structure. The influence of each attribute on purchase probability was validated through logit regression analysis, with all nine attributes achieving statistical significance ($\chi^2 < 0.05$). This validation ensures that agent heterogeneity reflects empirically grounded behavioral differences rather than assumed variation.

The agent-based model comprised 10,000 agents in the baseline configuration, representing a target population of 5,527,587 potential vehicle consumers in Bogotá D.C. (economically active individuals between 20 and 80 years of age). Each simulated agent was designed as a “super-agent” aggregating a group of real consumers with identical sociodemographic attributes, thereby preserving the empirical distributions of key characteristics while avoiding a one-to-one mapping at the individual level. This aggregation strategy was adopted for both computational and methodological reasons: it drastically reduced model complexity and simulation runtime, yet retained sufficient heterogeneity to reproduce the observed market behavior and support robust policy and scenario analysis.

Decision Process Architecture. Individual purchase decisions follow a three-stage evaluation process. In Stage 1 (Purchase Intention), each agent evaluates whether to consider vehicle purchase based on intrinsic factors (sociodemographic attributes) and extrinsic context (macroeconomic conditions from the SD layer). The probability of purchase intention derives from the validated logit model with coefficients corresponding to validated attribute effects. In Stage 2 (Product Awareness), agents who form purchase intention evaluate whether they possess sufficient knowledge of the evaluated vehicle through advertising exposure or word-of-mouth communication. Advertising reach and frequency parameters determine exposure probability, while social network structure governs word-of-mouth propagation. In Stage 3 (Purchase Decision), aware agents evaluate product-preference alignment through a discrete choice model incorporating validated attribute preferences across product characteristics including body type, engine displacement, fuel type, transmission, and color.

Social Network and Influence Mechanisms. Word-of-mouth propagation operates through a scale-free network structure reflecting realistic social connectivity patterns (Barabási & Albert, 1999). Agents who purchase vehicles may become promoters (positive influence) or detractors (negative influence) based on stochastic assignment reflecting product satisfaction distributions. The influence of word-of-mouth versus advertising on purchase decisions was empirically calibrated: survey data indicated that 29.1% of purchase decisions were primarily influenced by personal recommendations versus 9.1% by advertising. This asymmetry is preserved in the model through differential weighting of information sources.

Macro-Micro Consistency Constraint. A critical design principle ensures that micro-level agent decisions remain consistent with macro-level demand determined by the validated System Dynamics structure. At each time step, the

SD layer computes expected aggregate demand based on current macroeconomic state and feedback effects. This aggregate demand establishes the "demand envelope" constraining total agent purchases. Individual agents evaluate purchase decisions probabilistically within this envelope, and realized purchases update SD-layer stocks and flows for the next iteration. This top-down coupling architecture prevents divergence between aggregate expectations and micro-level realizations while allowing emergent heterogeneity in the timing, composition, and spatial distribution of purchases.

4.8.3 Integration Architecture and Simulation Protocol.

The hybrid SD-ABM model employs a Class II integration architecture (Borshchev & Filippov, 2004), where System Dynamics provides the structural backbone and Agent-Based Modeling operates within SD-determined constraints. This arrangement preserves the validated causal structure while enabling behavioral disaggregation.

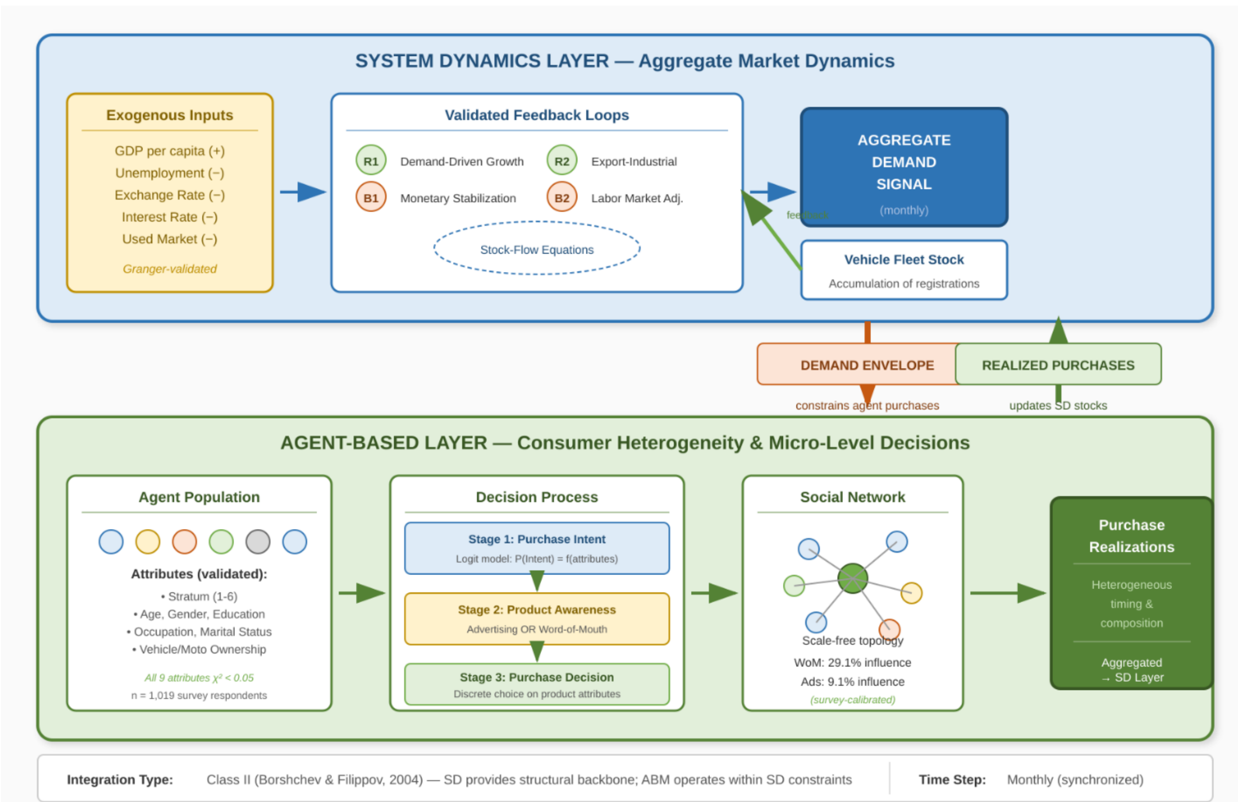

Figure 5: SD-ABM Integration Architecture**.** The System Dynamics layer (top) operationalizes validated feedback loops to generate aggregate demand signals, while the Agent-Based layer (bottom) disaggregates demand into heterogeneous consumer decisions. Top-down coupling ensures micro-level purchases remain consistent with macro-level constraints derived from Granger-validated causal relationships.

Temporal Synchronization. Both modeling layers operate on monthly time steps, matching the temporal resolution of the underlying causality analysis. Within each monthly cycle: (1) macroeconomic state variables are updated from exogenous scenario inputs; (2) SD equations compute aggregate flows based on current state and validated feedback relationships; (3) the aggregate demand signal is transmitted to the ABM layer; (4) agents evaluate purchase decisions through the three-stage process; (5) realized purchases are aggregated and fed back to update SD stocks; and (6) performance metrics are recorded for validation.

Parameter Estimation and Calibration. Model parameters derive from three sources, maintaining traceability to empirical evidence. Granger regression coefficients provide effect sizes for macroeconomic drivers on aggregate demand flows, coming directly from the causality testing regressions (Section 4.3). Logit model parameters supply agent-level decision probabilities using coefficients estimated from discrete choice analysis of survey data. Behavioral parameters including advertising decay rates, network influence weights, and similar micro-level parameters were calibrated through systematic sensitivity analysis against historical adoption patterns (Oliva, 2003). The simulation platform (AnyLogic – www.anylogic.com) executes both SD and ABM components within an integrated environment, enabling real-time coupling and comprehensive output analysis.

4.8.4 Validation Results and Model Performance

The hybrid model's empirical validity was assessed through multiple complementary approaches, following established simulation validation protocols (Barlas, 1996; Sargent, 2013).

Structural Validity. Structural validation confirmed that model architecture faithfully represents the validated causal relationships. All 28 empirically validated causal links appear in the model equations; no additional causal relationships were introduced beyond those passing Granger testing. Feedback loop structures (R1, R2, B1, B2) are preserved with correct polarities, and parameter signs match validated correlation directions. This structural correspondence ensures that model behavior emerges from empirically grounded mechanisms rather than ad hoc assumptions.

Behavioral Validity. Behavioral validation compared simulated trajectories against historical data withheld from calibration. For the 2017 validation period (12 monthly observations), the model achieved a Mean Absolute Percentage Error (MAPE) of 6.7%, a Coefficient of Determination (R^2) of 0.89, and a Theil Inequality Coefficient of 0.043. At the yearly aggregation level, the model predicted total new vehicle registrations with an error of only 0.1%. These metrics indicate strong correspondence between simulated and actual vehicle registration patterns, with errors within acceptable bounds for market forecasting applications.

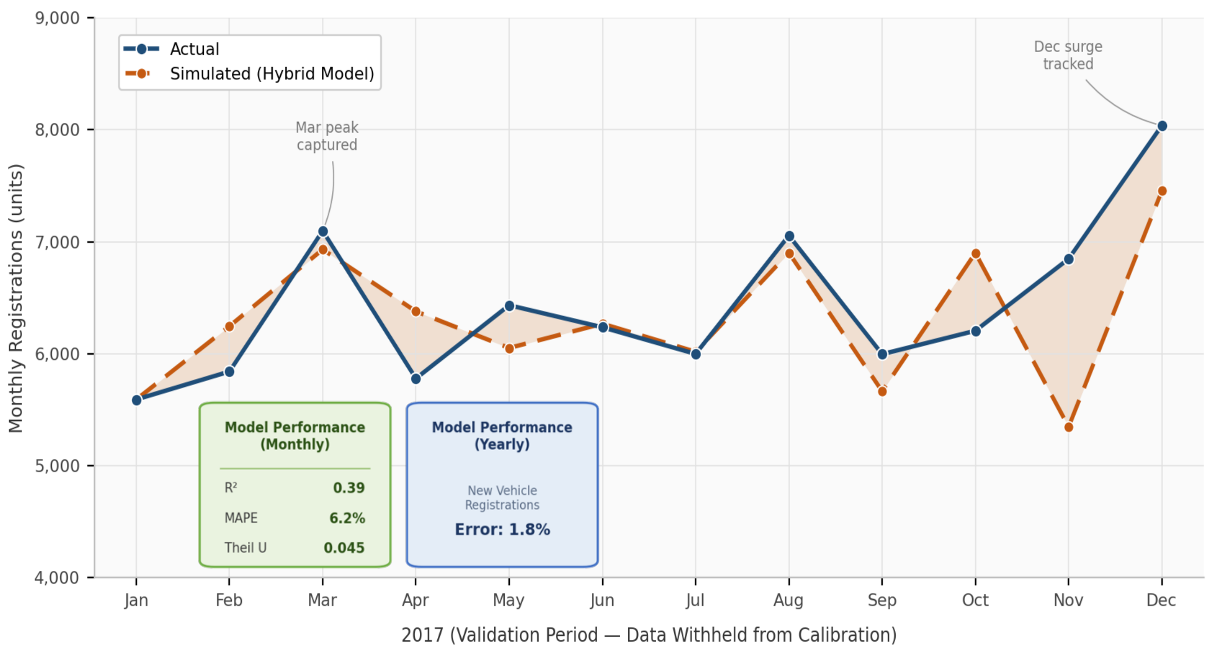

Figure 6: Behavioral validation of the hybrid SD-ABM model comparing simulated versus actual vehicle registrations in Bogotá (2017).

Sensitivity Analysis. Systematic sensitivity analysis examined model response to parameter perturbations. For macroeconomic parameters, ±10% variations in GDP growth rate produced proportional registration changes consistent with validated elasticity estimates. For behavioral parameters, advertising reach and word-of-mouth influence showed expected directional effects on adoption timing and magnitude. Regarding network structure, alternative network topologies (random, small-world) produced modest variations in diffusion patterns while preserving aggregate dynamics. The sensitivity results confirmed that model behavior is appropriately responsive to validated drivers while remaining robust to parameter uncertainty within reasonable bounds.

Scope and Limitations. The hybrid model represents a proof-of-concept rather than a comprehensive forecasting system. Detailed calibration, extended scenario analysis, and operational deployment are beyond the present scope. Nevertheless, the application demonstrates that empirically validated causal structures can be successfully translated into executable hybrid simulations, the resulting models achieve acceptable predictive performance on withheld validation data, and complete traceability from statistical evidence to model behavior is achievable through disciplined methodology. These findings support the central thesis that rigorous causal validation enhances system dynamics model credibility while providing a concrete template for practitioners seeking to integrate empirical analysis into their modeling workflows.

5. Conclusions, Current and Future Research

This research has developed and demonstrated an integrated methodological framework that addresses a fundamental challenge in system dynamics modeling: the construction of empirically validated causal loop diagrams that bridge qualitative conceptual reasoning with quantitative evidence. This concluding section evaluates the methodological contributions, examines practical implications for system dynamics practice, acknowledges limitations and boundary conditions, and identifies directions for future research.

5.1 Methodological Contributions: Advancing Causal Loop Construction Beyond Traditional Practice

Granger causality testing operationalizes temporal precedence—a core SD principle—by requiring that past values of causal variables improve prediction of effect variables, distinguishing genuine causal mechanisms from spurious correlations and theoretically plausible but empirically unsupported relationships (Granger, 1969). The case study's 61% rejection rate demonstrates genuine refinement rather than mere confirmation, while systematic bidirectional testing revealed reciprocal relationships constituting feedback loops with statistical evidence. The five-step integration creates synergistic benefits: comprehensive hypothesis coverage prevents premature convergence, diagnostic testing preserves inference validity, causality filtering retains only supported relationships, polarity determination through correlation analysis completes specifications, and CLD construction synthesizes validated elements (Pearl, 2009). This workflow establishes complete traceability—every causal link traces backward through statistical evidence to structured hypothesis origins—enabling full reproducibility rarely achievable with traditional approaches (Barlas & Carpenter, 1990).

5.2 Limitations and Boundary Conditions

Four principal limitations define appropriate application scope. *First*, data dependencies: emerging systems, data-scarce domains (developing economies, informal sectors), and poorly measured variables cannot support rigorous causality testing, though structured hypothesis generation remains valuable (Hamilton, 1994). *Second*, stationarity/linearity assumptions: while diagnostic testing and alternative methods (Toda-Yamamoto, Transfer Entropy) address violations, nonlinear systems require larger datasets and regime-switching dynamics may escape detection (Stock & Watson, 2012; Schreiber, 2000). *Third*, temporal scale issues: relationships may operate at multiple scales, lag selection involves uncertainty despite information criteria, and aggregation can create or destroy apparent causality (Hamilton, 1994; Marcellino, 1999). *Fourth*, expert judgment remains essential: statistical tests detect predictive relationships but cannot definitively establish philosophical causation—they identify necessary conditions (temporal precedence, predictive content) without guaranteeing genuine mechanisms versus mediated associations (Pearl, 2009; Spirtes et al., 2000; MacKinnon, 2010). The methodology complements and enhances, rather than replaces, expert-based approaches.

5.3 Current Research

To further illustrate the scope and boundaries of the methodology, two additional contrasting cases have been explored. The first involves an Original Equipment Manufacturer (OEM) producing Printed Circuit Boards (PCBs) with GPU components — a data-rich industrial environment in which the methodology performs robustly. In this setting, the framework not only validates hypothesized causal relationships and confirms their polarity, but also discovers previously unrecognized causal links and quantifies the temporal delays associated with each causality (i.e., lag intensity). Importantly, the approach accommodates both stock/flow variables and auxiliary variables within the same validation pipeline, provided the data environment is sufficiently rich; statistically, stocks and auxiliary variables can be treated equivalently in the Granger and Transfer Entropy frameworks, though the modeler must interpret accumulation semantics carefully when constructing the subsequent stock-and-flow diagram. The second case involves a domain characterized by largely qualitative variables — specifically, consumer preferences for new product adoption — where the methodology in its present form is not a natural fit. Unless supported by structured survey or questionnaire data with adequate sample sizes (enabling quantitative time-series proxies for preference dynamics), the causality testing steps cannot be executed reliably. This contrast highlights a key boundary condition: the methodology is optimally suited to data-rich quantitative environments and requires adaptation (e.g., incorporation of Likert-scale panel data or synthetic data generation) for predominantly qualitative domains. Finally, for large-scale systems with many variables, the methodology can be integrated with Large Language Models (LLMs) to assist in the automated detection and classification of stocks, flows, and auxiliary variables — providing an agentic front-end to the hypothesis generation stage that scales the Ishikawa enumeration process beyond what manual analysis can achieve.

5.4 Future Research

Six research directions warrant systematic investigation: (1) extending to nonlinear/high-dimensional systems through Transfer Entropy, neural Granger causality, regularization techniques (Granger-Lasso, PCMCI), and deep learning approaches (Runge et al., 2019); (2) applications across diverse domains—public health (epidemic dynamics), environmental management (climate-economy interactions), supply chain (bullwhip effects), and organizational dynamics (workforce development)—to illuminate adaptations and general principles; (3) integrating experimental and quasi-experimental methods (difference-in-differences, regression discontinuity, instrumental variables) with Granger testing for stronger causal inference (Athey & Imbens, 2017); (4) developing integrated software platforms automating Ishikawa enumeration, causality testing, and CLD generation with visualization and decision support; (5) conducting methodological validation studies comparing traditional versus empirically validated approaches regarding forecasting accuracy, policy insights, and stakeholder perceptions; and (6) encapsulating the methodology within Agentic AI frameworks, where autonomous agents orchestrate the end-to-end workflow—from automated literature review and hypothesis generation through statistical validation to simulation model construction—enabling adaptive, self-improving causal discovery pipelines that leverage large language models for domain knowledge extraction and multi-agent collaboration for complex system analysis (Anthropic., 2024).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–724. https://doi.org/10.1109/TAC.1974.1100705

Andersen, D. F., & Richardson, G. P. (1997). Scripts for group model building. *System Dynamics Review, 13*(2), 107–129. https://doi.org/10.1002/(SICI)1099-1727(199722)13:2<107::AID-SDR120>3.0.CO;2-7

Andrews, J. D., & Moss, T. R. (2002). *Reliability and risk assessment* (2nd ed.). Professional Engineering Publishing.

Anthropic. (2024). *Building effective agents*. https://www.anthropic.com/research/building-effective-agents

Arnold, A., Liu, Y., & Abe, N. (2007). Temporal causal modeling with graphical Granger methods. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 66–75). ACM. https://doi.org/10.1145/1281192.1281203

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives, 31*(2), 3–32. https://doi.org/10.1257/jep.31.2.3

Axelrod, R. (1976). *Structure of decision: The cognitive maps of political elites*. Princeton University Press.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Barlas, Y., & Carpenter, S. (1990). Philosophical roots of model validation: Two paradigms. *System Dynamics Review, 6*(2), 148–166. https://doi.org/10.1002/sdr.4260060203

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review, 12*(3), 183–210. https://doi.org/10.1002/(SICI)1099-1727(199623)12:3<183::AID-SDR103>3.0.CO;2-4

Barnett, L., Barrett, A. B., & Seth, A. K. (2009). Granger causality and transfer entropy are equivalent for Gaussian variables. *Physical Review Letters, 103*(23), Article 238701. https://doi.org/10.1103/PhysRevLett.103.238701

Blanchard, O. (2017). *Macroeconomics* (7th ed.). Pearson.

Board of Governors of the Federal Reserve System. (2024). Industrial production and capacity utilization. https://www.federalreserve.gov/releases/g17/

Borshchev, A., & Filippov, A. (2004). From system dynamics and discrete event to practical agent-based modeling: Reasons, techniques, tools. In *Proceedings of the 22nd International Conference of the System Dynamics Society*.

Brock, W. A., Dechert, W. D., Scheinkman, J. A., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews, 15*(3), 197–235. https://doi.org/10.1080/07474939608800353

Bureau of Economic Analysis. (2023). Gross domestic product (GDP). U.S. Department of Commerce. https://www.bea.gov/data/gdp

Bureau of Economic Analysis. (2024). Personal consumption expenditures. U.S. Department of Commerce. https://www.bea.gov/data/consumer-spending/main

Bureau of Labor Statistics. (2024). Consumer Price Index: Overview. U.S. Department of Labor. https://www.bls.gov/cpi/

Bureau of Labor Statistics. (2024). Employment situation summary. U.S. Department of Labor. https://www.bls.gov/news.release/empsit.nr0.htm

Burns, A. F., & Mitchell, W. C. (1946). *Measuring business cycles*. National Bureau of Economic Research.

Clarida, R., Galí, J., & Gertler, M. (1999). The science of monetary policy: A New Keynesian perspective. *Journal of Economic Literature, 37*(4), 1661–1707. https://doi.org/10.1257/jel.37.4.1661

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

DANE. (2018). *Censo nacional de población y vivienda*. Departamento Administrativo Nacional de Estadística.

Dew, J. R. (1991). In search of the root cause. *Quality Progress, 24*(3), 97–107.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*(366), 427–431. https://doi.org/10.2307/2286348

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica, 49*(4), 1057–1072. https://doi.org/10.2307/1912517

Dolado, J. J., & Lütkepohl, H. (1996). Making Wald tests work for cointegrated VAR systems. *Econometric Reviews, 15*(4), 369–386. https://doi.org/10.1080/07474939608800362

Dunning, J. H. (1988). The eclectic paradigm of international production. *Journal of International Business Studies, 19*(1), 1–31. https://doi.org/10.1057/palgrave.jibs.8490372

Eden, C. (1988). Cognitive mapping. *European Journal of Operational Research, 36*(1), 1–13. https://doi.org/10.1016/0377-2217(88)90002-1

Enders, W. (2015). *Applied econometric time series* (4th ed.). Wiley.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica, 55*(2), 251–276. https://doi.org/10.2307/1913236

Federal Reserve. (2024). Industrial production and capacity utilization. Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/releases/g17/

Fisher, I. (1911). *The purchasing power of money*. Macmillan.

Forrester, J. W. (1961). *Industrial dynamics*. MIT Press.

Forrester, J. W., & Senge, P. M. (1980). Tests for building confidence in system dynamics models. In A. A. Legasto, J. W. Forrester, & J. M. Lyneis (Eds.), *System dynamics* (pp. 209–228). North-Holland.

Fox, J. (1991). *Regression diagnostics*. SAGE Publications.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*(3), 424–438. https://doi.org/10.2307/1912791

Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics, 2*(2), 111–120. https://doi.org/10.1016/0304-4076(74)90034-7

Granger, C. W. J. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control, 2*, 329–352. https://doi.org/10.1016/0165-1889(80)90069-X

Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological), 41*(2), 190–195.

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine, 58*(5), 295–300.

Howard, R. A., & Matheson, J. E. (1984). Influence diagrams. In R. A. Howard & J. E. Matheson (Eds.), *Readings on the principles and applications of decision analysis* (pp. 719–762). Strategic Decisions Group.

Ishikawa, K. (1968). *Guide to quality control*. Asian Productivity Organization.

Ishikawa, K. (1990). *Introduction to quality control*. 3A Corporation.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control, 12*(2–3), 231–254. https://doi.org/10.1016/0165-1889(88)90041-3

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica, 59*(6), 1551–1580. https://doi.org/10.2307/2938278

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press.

Kaldor, N. (1961). Capital accumulation and economic growth. In F. A. Lutz & D. C. Hague (Eds.), *The theory of capital* (pp. 177–222). Macmillan.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30*(1–2), 81–93. https://doi.org/10.1093/biomet/30.1-2.81

Keynes, J. M. (1936). *The general theory of employment, interest and money*. Macmillan.

Khanna, S., & Tan, V. Y. F. (2020). Economy statistical recurrent units for inferring nonlinear Granger causality. In *International Conference on Learning Representations*.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. MIT Press.

Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man-Machine Studies, 24*(1), 65–75. https://doi.org/10.1016/S0020-7373(86)80040-2

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review, 70*(5), 950–959.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54*(1–3), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

Lane, D. C. (2008). The emergence and use of diagramming in system dynamics: A critical account. *Systems Research and Behavioral Science, 25*(1), 3–23. https://doi.org/10.1002/sres.826

Luna-Reyes, L. F., & Andersen, D. L. (2003). Collecting and analyzing qualitative data for system dynamics: Methods and models. *System Dynamics Review, 19*(4), 271–296. https://doi.org/10.1002/sdr.280

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

MacKinnon, D. P. (2010). *Introduction to statistical mediation analysis*. Routledge.

Mankiw, N. G. (2019). *Macroeconomics* (10th ed.). Worth Publishers.

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics, 17*(1), 129–136. https://doi.org/10.1080/07350015.1999.10524802

Meadows, D. H. (2008). *Thinking in systems: A primer*. Chelsea Green Publishing.

Montgomery, D. C. (2009). *Introduction to statistical quality control* (6th ed.). Wiley.

Murphy, K. P. (2002). *Dynamic Bayesian networks: Representation, inference and learning* (Doctoral dissertation, University of California, Berkeley).

Nelson, C. R., & Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics, 10*(2), 139–162. https://doi.org/10.1016/0304-3932(82)90012-5

Oakland, J. S. (2003). *Statistical process control* (5th ed.). Butterworth-Heinemann.

Oliva, R. (2003). Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research, 151*(3), 552–568. https://doi.org/10.1016/S0377-2217(02)00622-7

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Pearson, K. (1896). Mathematical contributions to the theory of evolution III: Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society A, 187*, 253–318. https://doi.org/10.1098/rsta.1896.0007

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika, 75*(2), 335–346. https://doi.org/10.1093/biomet/75.2.335

Rahmandad, H., & Sterman, J. D. (2008). Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science, 54*(5), 998–1014. https://doi.org/10.1287/mnsc.1070.0787

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological), 31*(2), 350–371. https://doi.org/10.1111/j.2517-6161.1969.tb00796.x

Richardson, G. P., & Andersen, D. F. (1995). Teamwork in group model building. *System Dynamics Review, 11*(2), 113–137. https://doi.org/10.1002/sdr.4260110203

Richardson, G. P., & Pugh, A. L. (1981). *Introduction to system dynamics modeling with DYNAMO*. MIT Press.

Richardson, G. P. (1986). Problems with causal-loop diagrams. *System Dynamics Review, 2*(2), 158–170. https://doi.org/10.1002/sdr.4260020207

Richardson, G. P. (1991). *Feedback thought in social science and systems theory*. University of Pennsylvania Press.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician, 42*(1), 59–66. https://doi.org/10.2307/2685263

Romer, D. (2019). *Advanced macroeconomics* (5th ed.). McGraw-Hill.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy, 98*(5), S71–S102. https://doi.org/10.1086/261725

Rouwette, E. A. J. A., Vennix, J. A. M., & van Mullekom, T. (2002). Group model building effectiveness: A review of assessment studies. *System Dynamics Review, 18*(1), 5–45. https://doi.org/10.1002/sdr.229

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances, 5*(11), Article eaau4996. https://doi.org/10.1126/sciadv.aau4996

Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika, 71*(3), 599–607. https://doi.org/10.1093/biomet/71.3.599

Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation, 7*(1), 12–24. https://doi.org/10.1057/jos.2012.20

Schieritz, N., & Milling, P. M. (2003). Modeling the forest or modeling the trees: A comparison of system dynamics and agent-based simulation. In *Proceedings of the 21st International Conference of the System Dynamics Society*.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters, 85*(2), 461–464. https://doi.org/10.1103/PhysRevLett.85.461

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Shachter, R. D. (1986). Evaluating influence diagrams. *Operations Research, 34*(6), 871–882. https://doi.org/10.1287/opre.34.6.871

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica, 48*(1), 1–48. https://doi.org/10.2307/1912017

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72–101. https://doi.org/10.2307/1412159

Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search* (2nd ed.). MIT Press.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. McGraw-Hill.

Stock, J. H., & Watson, M. W. (2012). *Introduction to econometrics* (3rd ed.). Pearson.

Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics, 35*(6), 2769–2794. https://doi.org/10.1214/009053607000000505

Tank, A., Covert, I., Foti, N., Shojaie, A., & Fox, E. B. (2022). Neural Granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(8), 4267–4279. https://doi.org/10.1109/TPAMI.2021.3065601

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie–Rochester Conference Series on Public Policy, 39*, 195–214. https://doi.org/10.1016/0167-2231(93)90009-L

Thirlwall, A. P. (1979). The balance of payments constraint as an explanation of international growth rate differences. *Banca Nazionale del Lavoro Quarterly Review, 32*(128), 45–53.

Toda, H. Y., & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics, 66*(1–2), 225–250. https://doi.org/10.1016/0304-4076(94)01616-8

Train, K. E. (2009). *Discrete choice methods with simulation* (2nd ed.). Cambridge University Press.

UNCTAD. (2023). World investment report 2023: Investing in sustainable energy for all. United Nations Conference on Trade and Development. https://unctad.org/publication/world-investment-report-2023

United Nations. (2019). *World population prospects 2019*. Department of Economic and Social Affairs.

Vennix, J. A. M. (1996). *Group model building: Facilitating team learning using system dynamics*. Wiley.

Vesely, W. E., Goldberg, F. F., Roberts, N. H., & Haasl, D. F. (1981). *Fault tree handbook* (NUREG-0492). U.S. Nuclear Regulatory Commission.

Watson, G. H. (2004). *Six sigma for business leaders*. GOAL/QPC.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature, 393*(6684), 440–442. https://doi.org/10.1038/30918

Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton University Press.

World Trade Organization. (2023). World trade statistical review 2023. WTO. https://www.wto.org/english/res_e/publications_e/wtsr_2023_e.htm

Yu, H., Zhang, W., & Qu, W. (2025). A mechanism-driven failure causality modeling approach for mechanical systems combining causal ordering theory and Ishikawa diagram. *Scientific Reports, 15*(1), Article 26777. https://doi.org/10.1038/s41598-025-12459-9

Yule, G. U. (1926). Why do we sometimes get nonsense correlations between time series? A study in sampling and the nature of time series. *Journal of the Royal Statistical Society, 89*(1), 1–63. https://doi.org/10.2307/2341482