



UNIVERSIDAD
NACIONAL
DE COLOMBIA

ESTIMACIÓN PREDICTIVA DE LA
INNOVACIÓN TECNOLÓGICA EN COLOMBIA
EMPLEANDO SIMULACIONES MONTE CARLO

DANIEL RODRÍGUEZ CÉSEPDES

Universidad Nacional de Colombia
Facultad de Administración
Manizales, Colombia
2022

**ESTIMACIÓN PREDICTIVA DE LA INNOVACIÓN
TECNOLÓGICA EN COLOMBIA EMPLEANDO
SIMULACIONES MONTE CARLO**

DANIEL RODRÍGUEZ CÉSPEDES

Tesis presentada como requisito parcial para optar al título de:

Magister en Administración

Director:

JUAN NICOLÁS MONTOYA MONSALVE

PhD. En administración

Codirectora:

ELISABETH RESTREPO PARRA

Doctora en Ingeniería – Línea Automática

Línea de Investigación:

Análisis de datos enfocado en la administración

Grupos de Investigación:

Grupo de Investigación en Administración Moderna GIAM – PCM Computational
Applications

Universidad Nacional de Colombia

Facultad de Administración

Manizales, Colombia

2022

“Cada experiencia tiene su elemento de magia”

Herman Hesse

Agradecimientos:

La construcción de la presente tesis hubiera sido imposible sin la contribución de un gran número de personas, de las cuales sólo unas pocas pueden mencionarse aquí. Ante todo, mi gratitud más profunda a mi madre Geny, a Guillermo y a mi familia, quienes con su amor y esfuerzo me han permitido alcanzar este sueño. A mi padre, Miguel, quien siempre me animó a alcanzar mis metas. A mis tutores, Los profesores Juan Nicolás y Elisabeth, quienes desde sus conocimientos y experiencia hicieron posible realizar esta investigación.

A la Universidad Nacional de Colombia, principalmente a la Facultad de Administración por las oportunidades que me han otorgado. Al Centro de Desarrollo Tecnológico “Innvestiga”, al Grupo de Investigación en Administración Moderna, al Laboratorio de Física del Plasma y al grupo de investigación PCM Computational Applications por abrirme sus puertas para aprender sobre investigación.

Finalmente, a todas las personas que me han acompañado durante este camino y que de una u otra forma han apoyado el desarrollo de esta tesis.

ESTIMACIÓN PREDICTIVA DE LA INNOVACIÓN TECNOLÓGICA EN COLOMBIA EMPLEANDO SIMULACIONES MONTE CARLO

Resumen

La innovación tecnológica es un proceso que ha adquirido una gran importancia para el desarrollo económico y social de los países; por esta razón, se ha hecho necesario la creación de modelos predictivos que sirvan como base para la estimación del comportamiento de la innovación tecnológica. No obstante, estos modelos son generalizados a nivel mundial y muchas veces los parámetros utilizados para estimar la innovación tecnológica en un país industrializado no se adaptan a los países en desarrollo.

Por lo tanto, en la presente tesis se busca explorar metodologías basadas en Análisis de Componentes Principales y en simulaciones Márkov Chain Monte Carlo, con el fin de realizar modelos predictivos que sirvan como base para la estimación de la innovación tecnológica en Colombia, usando como parámetro de medición las patentes registradas por residentes en la Super Intendencia de Industria y Comercio y utilizando variables macroeconómicas y sociales que se consideran que pueden tener influencia en la producción de patentes.

Luego de realizar y validar el modelo en una simulación de Márkov Chain Monte Carlo basado en procesos de Weiner, se observó que el modelo y la metodología tiene una buena capacidad de predicción para el fenómeno, pese a la alta volatilidad de éste, teniendo una diferencia de menos del 15% de la media de la simulación y los datos reales en los primeros dos años; no obstante, dada la alta volatilidad de las patentes, se recomienda en futuros estudios utilizar modelos donde se contemple la volatilidad variable del parámetro de medición.

Palabras clave: Innovación Tecnológica, Simulación Monte Carlo, Análisis de Componentes Principales, Patentes.

PREDICTIVE ESTIMATION OF TECHNOLOGICAL INNOVATION IN COLOMBIA USING MONTE CARLO SIMULATIONS

Abstract

Technological innovation is a process that has acquired great importance for the economic and social development of countries; for this reason, it has become necessary to create predictive models that serve as a basis for estimating the behavior of technological innovation. However, these models are generalized worldwide and often the parameters used to estimate technological innovation in an industrialized country are not adapted to developing countries.

Therefore, this thesis seeks to explore methodologies based on Principal Component Analysis and Markov Chain Monte Carlo simulations to perform predictive models. These models can serve as a basis for estimating technological innovation in Colombia. It was carried out using, as a measurement parameter, registered patents in the Superintendence of Industry and Commerce and using macroeconomic and social variables that are considered. These considerations may have an influence on the production of patents.

After performing and validating the model in a Markov Chain Monte Carlo simulation based on Weiner processes, it was observed that the model and the methodology have a good prediction capacity for the phenomenon, despite its high volatility, having a difference less than 15% of the average of the simulation and the real data in the first two years; however, given the high volatility of patents, it is recommended, for future studies to use models that take into account the variable volatility of the measurement parameter.

Keywords: Technological Innovation, Monte Carlo Simulation, Principal Components Analysis, Patents.

Tabla de Contenido

CAPÍTULO 1: INTRODUCCIÓN.....	1
1.1 Planteamiento del problema:	1
1.3 Justificación:	2
1.4 Objetivos.....	3
1.4.1 Objetivo general:	3
1.4.2 Objetivos específicos:	3
1.5 Metodología:.....	3
1.5.1 Recolección de datos:	4
1.5.2 Universo de estudio:	5
1.5.3 Instrumentos y técnicas:	5
CAPÍTULO DOS: MARCO TEORÍCO	1
2.1 Innovación tecnológica.....	1
2.1.1 Capacidad de innovación tecnológica:	1
2.1.2 Medición de la innovación tecnológica en Colombia.	2
2.1.3 Medición de la innovación tecnológica en Colombia	2
2.1.4 Medición de la capacidad de innovación tecnológica en Colombia.....	4
2.2 Modelo estadístico bayesiano:.....	7
2.2.1 Teorema de Bayes:	8
2.3 Simulación Monte Carlo:.....	10
2.3.2 Implementación del método Monte Carlo:	11
2.3.3 Ley fuerte de los grandes números:	11
2.3.4 Cadenas de Márkov:	12
2.3.5 Proceso de Wiener:.....	13
2.3.6 Factorización o descomposición de Cholesky:.....	14
CAPÍTULO 3: ESTADO DEL ARTE	18
3.1 Mediciones estadísticas de la innovación tecnológica.....	18
3.2 Método Monte Carlo y la innovación tecnológica	20
3.3 Simulación Monte Carlo en la administración:	21
CAPÍTULO CUATRO: RESULTADOS Y ANALÍISIS.....	24
4.1 Entender y describir el problema a resolver:	24
4.2 Planificación y recolección de datos:	25

4.3 Exploración de los datos obtenidos:	26
4.3.1 Exploración general:	26
Análisis de variables por estadística descriptiva	31
4.4 Postular un modelo que se adapte a los datos:	38
4.5 Ajustar el modelo:	39
4.6 Revisión del modelo y obtención de resultados	40
4.7 Revisión de la evolución de las variables independientes por MCMC:	50
4.7.1 Distribución de simulaciones de PIB per cápita	50
4.7.2 Distribución de las simulaciones de la inversión en I+D:	51
4.7.3 Distribución de las simulaciones de Calidad institucional	52
4.7.4 Distribución de las simulaciones de penetración de internet.....	54
4.8 Validar el modelo	56
CAPITULO 5: CONCLUSIONES Y RECOMENDACIONES	59
5.1 Conclusiones:.....	59
5.2 Recomendaciones:	60
BIBLIOGRAFÍA:	61

LISTA DE FIGURAS

Ilustración 4 -1: Test de correlación inicial, empleando el método de correlación de Pearson	28
Ilustración 4- 2: Correlación grafica entre variables	29
Ilustración 4 - 3: Correlación entre variables	31
Ilustración 4 – 4: Registro de patentes.....	33
Ilustración 4 – 5: PIB per cápita anual	34
Ilustración 4 – 6: Inversión en I+D.....	35
Ilustración 4 – 7: Histórico PIB	36
Ilustración 4 – 8: Calidad institucional.....	37
Ilustración 4 – 9: Penetración de internet	38
Ilustración 4 – 10: Proporción de Varianza Explicada por cada componente.....	40
Ilustración 4 – 11: Tendencia de patentes 2030 (10.000, 20.000, 50.000 y 100.000 simulaciones).....	44
Ilustración 4 – 12: Simulaciones gráficas a 10 años.....	46
Ilustración 4 – 13: Grafica de distribución a 10 años.....	47
Ilustración 4 – 14: Simulaciones de patentes a 2021 y 2025.....	48
Ilustración 4 – 15: Simulaciones de patentes por líneas 2021	48
Ilustración 4 – 16: Distribución de las simulaciones de PIB per cápita	50
Ilustración 4 – 17: Distribución de las simulaciones de I+D a 10 años.....	52
Ilustración 4 – 18: Distribución de las simulaciones de Calidad Institucional.....	53
Ilustración 4 – 19: Distribución de las simulaciones de penetración de internet	54
Ilustración 4 – 20: Camino de simulaciones de penetración de internet	55
Ilustración 4 – 21: Distribución de las simulaciones de patentes (2018 – 2021)	56

Lista de Ecuaciones:

Ecuación 1: Teorema de Bayes	8
Ecuación 2: Teorema de Bayes 2	8
Ecuación 3: Conjunto de datos de β	8
Ecuación 4: Distribución previa	9
Ecuación 5: Distribución muestral	9
Ecuación 6: Distribución a posteriori	9
Ecuación 7: Teorema de Cox-Jaynes	9
Ecuación 8: Distribución predictiva posterior	10
Ecuación 9: Precisión de método Monte Carlo	10
Ecuación 10: Ecuación de muestras de θ	11
Ecuación 11: Ecuación de ley fuerte de los grandes números	11
Ecuación 12: Promedio muestral – ley débil	11
Ecuación 13: Demostración de Ley débil de los grandes números Usando Chebyshev ...	11
Ecuación 14: Valor esperado de S_n	12
Ecuación 15: Probabilidad de obtener el promedio de la muestra a partir de ley débil de los grandes números	12
Ecuación 16: Tendencia al promedio	12
Ecuación 17: Representación cadenas de Markov	12
Ecuación 18: Representación dependencias cadenas de Markov	12
Ecuación 19: Proposición movimiento Browniano	13
Ecuación 20: Vectores	13
Ecuación 21: Función de densidad	13
Ecuación 22: Analisis de componentes principales	15
Ecuación 23: Ecuación de X PCA	15
Ecuación 24: Ecuación de vectores normalizados	15
Ecuación 25: Ecuación de vectores normalizados	16
Ecuación 26: Centrar datos a media 0	39
Ecuación 27: Calculo de variable explicada	39
Ecuación 28: Modelo de medición de innovación tecnológica	40
Ecuación 29: Distribución del logaritmo en calculo del crecimiento	41
Ecuación 30; Tasa de crecimiento	41
Ecuación 31: Tasa de crecimiento	41
Ecuación 32: Volatilidad del modelo	41
Ecuación 32: Ecuación de la volatilidad simplificada	42
Ecuación 33: Factorización de Cholesky	42
Ecuación 34: Ecuación diagonal L	43
Ecuación 35: Ecuación diagonal principal	43
Ecuación 36: Matriz L	43
Ecuación 37: Ecuación Matriz U	43
Ecuación 38: Matriz LU	43

Lista de tablas

Tabla 4 - 1: Exploración inicial de variables – Comportamiento histórico.....	26
Tabla 4 - 2: Histórico de datos variables	31
Tabla 4 - 3: Estadísticas descriptivas.....	32
Tabla 4 – 4: Estadísticas descriptivas de las simulaciones	45
Tabla 4 – 5: Estadísticas descriptivas de patentes por años	49
Tabla 4 – 6: Estadísticas descriptivas del PIB en Colombia por años.....	50
Tabla 4 – 7: Estadísticas descriptivas de la inversión en I+D en Colombia por años	52
Tabla 4 – 8: Estadísticas descriptivas de la calidad institucional en Colombia por años	53
Tabla 4 – 9: Estadísticas descriptivas de la penetración de internet en Colombia por años	55
Tabla 4 – 10: Datos reales de registro de patentes por residentes (2019 – 2021)	56
Tabla 4 – 11: Estadísticas de patentes simuladas (2019 – 2021).....	57

Lista de símbolos y abreviaturas:

Símbolos con letras latinas

Símbolo	Término
n	Número de muestras generadas
p	Probabilidad
W	Proceso estocástico de Wiener
Z	Puntuación estándar

Símbolos con letras griegas

Símbolo	Término
μ	Promedio aritmético
ϵ	Experimento aleatorio
σ^2	Varianza
σ	Desviación Estándar
ϕ	Modelo autorregresivo lineal
λ	Número de eventos esperado por unidad de tiempo
θ	Parámetro

Abreviaturas

Abreviatura Término

<i>MCMC</i>	Márkov Chain Monte Carlo
<i>PCA</i>	Principal Component Analysis
<i>GARCH</i>	Generalized AutoRegressive Conditional Heteroskedasticity
<i>BMA</i>	Bayesian Model Averaging
<i>SIC</i>	Super Intendencia de Industria y Comercio
<i>L</i>	Matriz Triangular Superior
<i>U</i>	Matriz Triangular inferior
<i>PIB</i>	Producto Interno Bruto
<i>I+D</i>	Investigación y desarrollo
<i>VAR</i>	Modelo Vectorial Autorregresivo
<i>iid</i>	Independientes e idénticamente distribuidas

CAPÍTULO 1: INTRODUCCIÓN

1.1 Planteamiento del problema:

La innovación tecnológica es un segmento cada vez más importante en el desarrollo económico y social de los países. Para Artz et al. (2010) la innovación y la productividad han sido reconocidas desde hace mucho tiempo como la fuerza impulsora y dominante para el crecimiento sostenido en la economía de una región, la supervivencia y el éxito de las empresas individuales.

Este ámbito está muy relacionado con la competitividad de sus instituciones y la generación de valor que éstas proveen a la sociedad. Para Tejada-Estrada, et al., (2019) el desarrollo de un país está marcado por el desarrollo de sus instituciones. En la medida que éstas se tornen competitivas, innovando en sus métodos, procesos, productos y/o servicios, se contribuirá a alcanzar su desarrollo sostenido.

Por esta razón, gran parte de los países realizan fuertes inversiones en el desarrollo de las actividades de I+D (investigación y Desarrollo) y otros relacionados con el incremento de la innovación tecnológica, a fines de, con este incremento, incentivar la economía y aumentar la competitividad del país. Sin embargo, las decisiones de las cuantías que se invierten en el desarrollo de la innovación no siempre cuentan con el conocimiento necesario para generar los máximos beneficios.

A pesar de la alta importancia de la innovación tecnológica en el desarrollo económico, hay pocos modelos predictivos que se centren en estimar cómo evolucionará la innovación tecnológica. Además, la revisión bibliográfica no revela estudios que cuenten con variables predictoras independientes para crear un modelo sobre la innovación de un país o región.

Para Aguirre-Ramírez (2010) la falta de metodologías para medir y evaluar las capacidades de innovación tecnológica se debe principalmente a la falta de integración de diferentes tipos de variables y un análisis integral matemático formal entre ellas.

“Las aproximaciones teóricas y metodológicas existentes tanto para la medición como para la evaluación de la innovación, persiste en la falta de modelos formales y la evidencia empírica que ayuden a comprender el fenómeno”. (Grundey 1998 y Tschang 2003 en Aguirre-Ramírez, 2010),

Es bien sabido que, el desarrollo económico de un país depende fuertemente de su desarrollo científico, tecnológico y de la transferencia del conocimiento hacia sectores productivos y a la sociedad, lo que en últimas puede conllevar a generar innovación en la región. Por tal razón, si se quiere escalar en los índices de competitividad y de innovación, es necesario desarrollar metodologías que permitan identificar caminos y estrategias para incrementar la transferencia del conocimiento y para generar propiedad intelectual, que represente nuevos o mejorados productos y procesos.

De no desarrollar métodos para medir la innovación de un país, y su proyección hacia el futuro, es posible que el país no esté transitando en el camino adecuado hacia el crecimiento científico, tecnológico y por ende económico.

Por lo tanto, el estudio tiene como objetivo comprender el fenómeno social de la innovación tecnológica en Colombia, de forma cuantitativa y comprobar la influencia que tiene en la innovación, el PIB per cápita, la inversión en I+D, la calidad institucional, la penetración del internet y las exportaciones como porcentaje del PIB. Siendo medida la innovación a través del registro de patentes.

De esta forma, se espera crear un modelo predictivo, utilizando análisis de componentes principales y el método de simulación Monte Carlo, buscando que el modelo genere herramientas para mejorar la toma de decisiones sobre innovación tecnológica a nivel nacional y regional.

1.3 Justificación:

Debido a que la innovación tecnológica de los países es causa directa de la productividad y competitividad de un país o región, la presente investigación se enfocará en predecir el comportamiento que tendrá la innovación tecnológica en Colombia desde el punto de vista científico y tecnológico, ya que actualmente hay poca bibliografía relacionada a estimarla y predecirla.

Por ende, se espera que el estudio incremente el conocimiento que se tiene sobre innovación tecnológica en Colombia y sirva como base para futuros estudios que busquen estimar y predecir la innovación tecnológica en países no industrializados.

Por otro lado, el desarrollo de la presente tesis permitirá fortalecer líneas de investigación en los grupos de trabajo académicos participantes, en temas de innovación tecnológica y en el campo de las simulaciones Monte Carlo aplicadas a temas diferentes a la física teórica.

Económicamente este trabajo es pertinente, ya que puede tomarse como un punto de partida para tomar decisiones de inversión con miras a generar un incremento en la innovación y productividad del país.

Así, mediante simulaciones Monte Carlo y Análisis de Componentes Principales (PCA) la presente investigación permitirá mostrar los cambios que la innovación tecnológica en Colombia tendrá en los próximos diez años y profundizar en las variables más influyentes en la innovación tecnológica. Además de ofrecer una revisión histórica del comportamiento de la innovación tecnológica en Colombia desde el año 2000 hasta el año 2020 contrastando con diferentes variables.

1.4 Objetivos

1.4.1 Objetivo general:

Desarrollar un modelo predictivo para estimar la tendencia de la innovación tecnológica en Colombia empleando simulaciones Monte Carlo.

1.4.2 Objetivos específicos:

1. Identificar y clasificar variables que influyen en la innovación tecnológica de Colombia.
2. Diseñar e implementar un modelo de simulación Monte Carlo en Python para determinar relaciones entre las variables identificadas
3. Analizar los resultados obtenidos en las estimaciones predictivas de la innovación tecnológica de Colombia.

1.5 Metodología:

La presente tesis está basada en un modelo cuantitativo, con un diseño no experimental de correlación predictiva, por la cual se pretende determinar que probabilidades de variación hay en la innovación tecnológica de Colombia en el año 2030, utilizando el modelo de simulación Monte Carlo en el lenguaje de programación Python.

El modelo utilizará diferentes variables claves consideradas en la literatura sobre innovación y se propondrá un modelo basado en el Método de Análisis de Componentes Principales.

Así, inicialmente se hacen estadísticas descriptivas de todas las variables buscando entender de mejor forma el comportamiento de la innovación tecnológica en Colombia y las variables que la componen. Estas son validadas por pruebas de correlación de Pearson, tanto gráficas como numéricas que permitan observar su correlación.

Posteriormente se establece un modelo buscando definir el comportamiento del parámetro de medición, las patentes, utilizando las diversas variables postuladas. Sirviendo como base para la implementación del método Monte Carlo, basado en los principios de los Procesos de Weiner y la Descomposición de Cholesky. Esto arrojará los datos que se usarán para predecir las probabilidades del comportamiento de la innovación tecnológica en Colombia.

Posteriormente se valida el modelo, para lo cual se utilizarán únicamente los datos obtenidos desde el año 2000 al año 2018 y se simula el registro de patentes para los años 2019, 2020 y 2021. Posteriormente se compararán estos datos con los registrados en la realidad y se establece así un margen de error.

Así, el proceso de simulación que se realizará se puede resumir en 6 pasos claves:

1. Entender y describir el problema a resolver
2. Planificación y recolección de datos
3. Exploración de los datos obtenidos
4. Postular un modelo que se adapte a los datos
5. Ajustar el modelo
6. Revisar el modelo

1.5.1 Recolección de datos:

Los datos sobre patentes y modelos de utilidad son obtenidos registros públicos de la Superintendencia de Industria y Comercio de Colombia.

Los datos de las variables son obtenidos del Departamento Administrativo Nacional de Estadística DANE, el Departamento Nacional de Planeación DNP, el Banco Mundial y otras fuentes de datos nacionales e internacionales.

1.5.2 Universo de estudio:

Datos de las patentes de residentes registradas, el PIB per cápita promedio anual, la inversión en I+D, el índice de calidad institucional y la tasa de penetración de internet. Todos los datos referentes a Colombia y tomados desde el año 2000 hasta el año 2020.

1.5.3 Instrumentos y técnicas:

Para validar las variables que más han influido históricamente en la innovación tecnológica de Colombia, primeramente, se hacen estadísticas descriptivas y matrices de correlación de Pearson, utilizando el lenguaje de programación Python.

Los números pseudoaleatorios son generados a través del objeto “Random” y se utiliza la semilla (171239) de la librería “Numpy” misma librería que es utilizada como apoyo para la realización de estadísticas descriptivas. Así mismo, la librería “Scipy” se utiliza como base para algunos análisis descriptivos e inferenciales.

Las gráficas y dibujos se realizarán a través de las librerías “Matplotlib” y “Seaborn” en Python. Las bases de datos se crean en el programa Excel y son importadas a Python mediante la librería “Pandas”, También se utiliza para aplicar todos los análisis de datos.

La librería Sklearn se utiliza de apoyo para la validación de las variables y obtención del modelo por Análisis de Componentes Principales.

CAPÍTULO DOS: MARCO TEORÍCO

2.1 Innovación tecnológica.

La innovación es un concepto muy amplio, pero usualmente se refiere a la creación o mejoramiento de un bien o servicio que brinde facilidades, comodidades o que permita realizar procesos que antes presentaban dificultades. La innovación también supone renovación, nuevos conocimientos y creación de soluciones para un entorno.

“La introducción de un nuevo, o significativamente mejorado, producto (bien o servicio), de un proceso, de un nuevo método de comercialización o de un nuevo método organizativo, en las prácticas internas de la empresa, la organización del lugar de trabajo o las relaciones exteriores”. Pérez-Padrón et. Al, (2018)

La innovación tecnológica es una clasificación dentro del amplio concepto de innovación, donde por medio de la aplicación de la ciencia a la resolución de problemas, se busca crear o mejorar productos o servicios que satisfacen necesidades, aspiraciones y facilitan la adaptación al medio ambiente del ser humano.

Por lo tanto, la innovación tecnológica se puede dividir en (i) innovación incremental, que busca mejorar un producto previamente existente, como podría ser la adaptación de un módulo de cámaras a un celular o (ii) la innovación radical, que se refiere a la creación de un producto totalmente novedoso, como el invento de la cámara o del celular.

2.1.1 Capacidad de innovación tecnológica:

La capacidad de innovación tecnológica está fuertemente ligada con la predicción del comportamiento de la innovación tecnológica. Puesto que este parámetro no mide la innovación, sino las variables que pueden influir en ésta, como son la inversión, el entorno, los recursos científicos y tecnológicos, entre otros.

“La capacidad de innovación tecnológica se refiere al proceso independiente de desarrollo e innovación de los sujetos (incluidos el gobierno, las instituciones de investigación científica, las empresas de alta tecnología y las instituciones financieras) de una ciudad sobre la base de sus recursos científicos y tecnológicos, las características relevantes del entorno de innovación y su propia inversión integral en innovación”. (Wang G., Wang Y., 2019).

Así, la capacidad de innovación tecnológica es clave para la medición de la innovación tecnológica, ya que intenta englobar las diferentes variables que influyen en ésta, su impacto y de esta forma, determinar el entorno de la innovación en una empresa, región o país.

2.1.2 Medición de la innovación tecnológica en Colombia.

Mundialmente se utilizan diversos recursos para medir la capacidad de innovación tecnológica. Entre los más destacados se encuentran la producción de artículos científicos y técnicos, documentos citados, productividad laboral, certificados de reconocimiento de calidad ISO 9001, complejidad de producción e importación, exportaciones de alta tecnología, entre otras. La innovación tecnológica es usualmente medida con el uso de patentes de invención y modelos de utilidad como norma general.

Para World Intellectual Property Organization (2021), Colombia se encuentra en el ranking 67 de innovación, entre los 132 países evaluados, mientras que, únicamente considerando la innovación tecnológica, el país se encuentra en el puesto 80 de este ranking.

Por otro lado, Gómez-Hurtado (2018) realizó un análisis de las patentes concedidas a instituciones o inventores de origen colombiano, registradas en las bases de datos de las oficinas de patentes de Estados Unidos, Europa, WIPO, Alemania y Colombia. En el estudio, la autora analizó las tendencias de innovación tecnológica que rigen los procesos y productos en Colombia y comprueba que las patentes son indicadores de innovación tecnológica. La autora resaltó que Colombia se ha vuelto más competitiva en innovación y ha disminuido la brecha tecnológica con otros países de la región.

2.1.3 Medición de la innovación tecnológica en Colombia

Usualmente la innovación tecnológica de una región es medida utilizando su registro de patentes. Poop (2005) defiende el uso de las patentes en este proceso puesto que:

“Al ser un resultado directo del proceso inventivo, especialmente de las invenciones que se espera que tengan impacto; son muy apropiadas para comprender la competencia dentro de la dimensión tecnológica, puesto que las patentes son desglosadas en campos técnicos, de manera que indican no solo la intensidad de la actividad inventiva, sino también, su dirección; las patentes son documentos

públicos y por lo tanto es muy sencillo acceder a ellas; sus estadísticas están disponibles en grandes cantidades y durante largos períodos de tiempo. Y a diferencia de otros indicadores como la inversión en I+D, las patentes tienen la ventaja de ofrecer detalles de cada invención”. (Poop, 2005).

Más recientemente, Igami & Subrahmanyam (2019) concluyen que el número de patentes se correlacionan positivamente con las innovaciones de una manera estadísticamente significativa y, por lo tanto, pueden "predecir" las innovaciones mejor que una suposición aleatoria.

Por lo tanto, la innovación tecnológica de un país puede ser estimada gracias a su registro de patentes. Este registro provee valiosa información, no sólo de las patentes de invención y de modelos de utilidad registrados, sino también de las áreas donde más se patenta y los tipos de patente, entre otros. Esto permite medir la innovación tecnológica y sus particularidades.

En Colombia las patentes son concedidas por la Superintendencia de Industria y Comercio, quien las clasifica en patentes de invención y patentes de modelos de utilidad.

Para la Super Intendencia de Industria y Comercio (n.d) una patente de modelos de utilidad es un privilegio que le otorga el Estado al inventor como reconocimiento de la inversión y esfuerzos realizados, a aquellas invenciones que consisten en una nueva forma, configuración o disposición de elementos de un artefacto, herramienta, instrumento, mecanismo u otro objeto o parte de los mismos, que permita un mejor o diferente funcionamiento, utilización o fabricación del objeto que lo incorpora o que le proporcione alguna utilidad, ventaja o efecto técnico que antes no tenía.

Las patentes de modelo de utilidad únicamente protegen invenciones de un producto, mientras que las patentes de invención protegen tanto el producto como el procedimiento.

En los últimos años las patentes de invención concedidas a residentes de Colombia por la Superintendencia de Industria y Comercio han tenido un incremento sostenido. Mientras en el año 2000 se concedieron 24 patentes de invención de 84 presentadas. En el año 2018 las patentes concedidas aumentaron a 215 de 415 presentadas.

Las patentes de modelos de utilidad concedidas a residentes de Colombia han tenido una gran variabilidad entre los años, y aunque parece que esta tiene una tendencia al alza, no

hay un patrón claro que permita estimar cuántas patentes de modelos de utilidad se presentarán y se concederán por la Superintendencia de Industria y Comercio a residentes de Colombia en un año dado.

A pesar de que se observa un leve aumento en el número de patentes presentadas y registradas en Colombia, aún hay una gran diferencia con un país desarrollado como Estados Unidos, donde, según U.S Patent and Trademark Office (2019), en el año 2018 fueron presentadas 285.095 solicitudes de patentes por residentes de ese país, de las cuales fueron concedidas 144.413. Esto es un total de 0,44 patentes de invención concedidas por cada mil habitantes.

Para este mismo año, en Colombia se presentaron un total de 581 solicitudes de patente, de las cuales se concedieron 306. Lo que equivale a 0,00616 patentes concedidas a residentes por cada mil habitantes.

Sin embargo, es de mencionar que los requisitos para la obtención de patentes varían según el país o región donde se quiera obtener la patente. Para S. Chang et al (2018), estos requisitos afectan la calidad de las patentes y por ende la capacidad para medir la innovación tecnológica de las mismas.

2.1.4 Medición de la capacidad de innovación tecnológica en Colombia

Las variables más determinantes de la innovación tecnológica han sido propuestas en diversos estudios. Entre las más destacadas se encuentran:

El Producto Interno Bruto (PIB) per cápita: Para Ulku (2004) el valor del PIB per cápita y la innovación tecnológica para los países desarrollados y en desarrollo están relacionados positivamente. Además, existe una relación entre el stock de I + D y la innovación en los países desarrollados con grandes mercados.

Phoonk S & Phoonk S (2017), comprueban empíricamente esta relación en Malasia concluyendo que existe una fuerte relación positiva entre el PIB y gasto en innovación tecnológica. Con base en los resultados de un modelo VAR (Modelo Vectorial Autorregresivo) aplicado en su estudio. Los autores validan que los cambios en el PIB tienen un efecto positivo con el gasto público en innovación tecnológica.

Inversión en I + D: La inversión gubernamental en investigación y desarrollo de un país tiene una relación directa con la capacidad de innovación tecnológica del mismo, puesto que estos recursos se suelen ver reflejados en el incremento de la productividad y competitividad de las empresas, así como en la disminución de problemas ambientales derivados de las grandes industrias. Esto permite que las grandes empresas sean más viables y que a su vez, sean más innovadoras, generen más conocimiento e información.

La inversión del gobierno en ciencia y tecnología tiene un impacto positivo en la capacidad de innovación tecnológica. La inversión del gobierno en ciencia y tecnología transmite información sobre la orientación política del país y el apoyo del gobierno a las empresas industriales grandes y medianas. Orienta la entrada activa de capital financiero y capital social y alivia eficazmente las limitaciones financieras de la empresa, lo que refleja la perspectiva de desarrollo industrial.
(Zhang M., 2018)

Zhang M. (2018) muestra en su estudio como la inversión en I+D es la variable que más influencia tiene sobre la innovación tecnológica en China. Así, se puede determinar la alta influencia que tiene la inversión en ciencia y tecnología por parte del estado en el desarrollo de la innovación tecnológica regional. La influencia tan directa de la inversión en I+D ha sido estudiada por diversos autores habiendo un consenso sobre los beneficios que genera el incremento de inversión en la innovación.

Zhang M., realiza el modelo para la simulación utilizando una técnica llamada Bayesian Model Averaging (BMA) de donde obtiene los parámetros necesarios para realizar la simulación.

Para Chi M. et al (2021) Existe un consenso común entre los investigadores sobre los beneficios generalizados de la I + D a la hora de mejorar el grado de innovación tecnológica. Los gastos en I + D se asocian positivamente con un mayor grado de innovación tecnológica.

Calidad institucional: Para Molano-Rojas et al (2018) la calidad institucional puede definirse como la capacidad con la que cuentan las instituciones para establecer un sistema

de incentivos que reduzca el margen de incertidumbre y los costos de transacción promoviendo la eficiencia y facilitando mejores resultados económicos.

Así, se establece que la calidad institucional tiene una relación directa con el desarrollo de innovación tecnológica. En los estudios empíricos realizados por D'Ingiullo & Evangelista (2020) se concluye que los efectos de la calidad institucional sobre el desempeño innovador de las provincias italianas. En particular, la calidad general de las instituciones contribuye significativamente a explicar las brechas regionales de innovación.

Acceso a internet: En la era actual el acceso a internet de las comunidades toma gran importancia en el desarrollo de patentes, y aunque el internet ya es un servicio casi estandarizado, aún hay muchas regiones que no cuentan con acceso a éste, principalmente en los países en vías de desarrollo.

Wang, et al., (2021) defienden que las empresas con sede en ciudades con altas tasas de penetración de Internet tienden a ser más innovadoras; esto debido a que invierten más en investigación y desarrollo.

Exportaciones realizadas: Usualmente se considera que las exportaciones realizadas influyen significativamente en la innovación tecnológica, entre más productos sean exportados se considera que el país tiene mayor innovación; no obstante, en países con poca capacidad tecnológica esta relación no es muy fuerte puesto que la mayor parte de los productos exportados son materias primas y productos con baja tecnología.

“El análisis histórico permite inferir que la relación entre la capacidad de innovación y el comercio internacional en América Latina es muy débil, pues su estructura productiva se concentra en bienes primarios y en manufacturas de baja tecnología y sus procesos de innovación tecnológica son incipientes comparados con los de potencias como Estados Unidos, China, Alemania y Japón”. (Díaz E, 2019, pp. 44)

Registro de software: En muchos escenarios se relaciona directamente el registro de software con la innovación tecnológica; esto debido a las repercusiones que tiene este dentro de las organizaciones que lo desarrollan.

Andersson et al., (2021) muestra que las empresas que se involucran en el desarrollo de software, especialmente aquellas con desarrolladores de software internos, reportan niveles más altos de producción de innovación y pueden atribuir una mayor parte de sus ventas a la innovación.

No obstante, se debe tener en cuenta que el registro de software en Colombia es considerado como una obra literaria por la Ley 23 de 1982 sobre Derechos de Autor, por lo que es registrado en la Dirección Nacional de Derechos de Autor y se pueden registrar varios softwares que cumplan con exactamente la misma función.

Modelos de utilidad: Un modelo de utilidad es una mejora que se le aplica a una invención ya realizada, a fin de mejorar su utilidad.

“Un modelo de utilidad son aquellas invenciones que consisten en una nueva forma, configuración o disposición de elementos de un artefacto, herramienta, instrumento, mecanismo u otro objeto o parte de estos, que permita un mejor o diferente funcionamiento, utilización o fabricación del objeto que lo incorpora o que le proporcione alguna utilidad, ventaja o efecto técnico que antes no tenía”.
Super Intendencia De Industria y Comercio (n.d)

Para Molina & Gómez (2017) las innovaciones protegidas por modelos de utilidad impactan positivamente en el desarrollo de la innovación, aunado a esto, las innovaciones de modelos de utilidad pueden ser un insumo importante en la generación de futuras innovaciones patentables.

2.2 Modelo estadístico bayesiano:

La estadística se encarga de todo el proceso de planeación, recolección y análisis de datos para que los resultados sean científicamente defendibles. Para Barreto-Villanueva (2012), La Estadística es la ciencia cuyo objetivo es reunir información tanto cuantitativa como cualitativa concerniente a individuos, grupos, series de hechos, etc., para deducir de ello, gracias al análisis de estos datos, significados precisos o previsiones para el futuro. También se puede definir como el estudio de la incertidumbre.

Por su parte, la estadística bayesiana expresa sus resultados en términos de grados de creencia, o más específicamente en probabilidades bayesianas. Para Dienes (2011) PP. 276 La estadística bayesiana parte de la premisa de que podemos asignar grados de plausibilidad a las teorías, y lo que queremos que hagan nuestros datos es decirnos cómo ajustar estas plausibilidades.

2.2.1 Teorema de Bayes:

La estadística bayesiana soporta parte de sus preceptos sobre la estadística en el teorema de Bayes, este expresa la probabilidad condicional de un evento aleatorio A dado B, en términos de la distribución de la probabilidad condicional de B dado A y la distribución de probabilidad marginal de A, sobre la probabilidad de B.

se expresa con la fórmula de:

Ecuación 1: Teorema de Bayes

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

Considerando que P es la probabilidad, A y B son dos sucesos aleatorios.

Lo que también se puede expresar como:

Ecuación 2: Teorema de Bayes 2

$$p(A|B) = \frac{p(B|A)p(A)}{\sum p(B|A)p(A)}$$

En el caso de que las observaciones obtenidas de B sobre A sean enteros.

Axioma: Todas las evaluaciones del estado de información del fenómeno están bajo supuestos teóricos.

Axioma: Estos supuestos y juicios de valor son una forma de información que se puede expresar mediante el conjunto de proposiciones.

Ecuación 3: Conjunto de datos de β

$$\beta = \{\beta_1, \dots, \beta_n\}$$

Siendo β el ente de interés compuesto por un conjunto de observaciones entre 1 y n

Usualmente el ente de interés se denomina parámetro y se denota con el símbolo θ

Axioma: Existe una o varias fuentes de información que me permiten disminuir la incertidumbre sobre el parámetro.

Axioma: existe una fuente de datos “ y ” que disminuye la incertidumbre de θ . Así, la existencia de y hace que pueda dividir la información de θ en dos: Información interna y externa.

Así, inicialmente la información externa a “ y ” se denomina con la distribución previa:

Ecuación 4: Distribución previa

$$p(\theta|B) \equiv p(\theta)$$

Donde $p(\theta|B)$ son los supuestos y $p(\theta)$ es toda la información externa a la distribución previa “ y ” .

Posteriormente la información acerca de θ interna a “ y ” se representa con la distribución muestral:

Ecuación 5: Distribución muestral

$$p(y|\theta, B) \equiv p(y|\theta)$$

Cuando se combina la información interna y externa se obtiene la distribución a posteriori:

Ecuación 6: Distribución a posteriori

$$p(\theta|y, B) \equiv p(\theta|y)$$

La unión óptima de las dos fuentes de información, interna y externa de θ está basada en la distribución posterior y se realiza haciendo uso del teorema de Cox-Jaynes (1946), siendo este:

Ecuación 7: Teorema de Cox-Jaynes

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int_{\Theta} p(y|\theta)p(\theta)d\theta} \propto p(y|\theta)p(\theta)$$

La predicción óptima de un nuevo conjunto de datos denominado y^* , estaría basada en la distribución predictiva posterior.

Ecuación 8: Distribución predictiva posterior

$$p(y^*|y) = \int_{\Theta} p(y^* \cdot \theta|y)d\theta = \int_{\Theta} p(y^*|\theta)p(\theta|y)d\theta$$

Donde Θ representa todos los valores que puede tomar θ y en el caso de los parámetros tener valores continuos, $p(y)$ se representa como una integral. En el caso de tener valores discretos se usa una sumatoria.

2.3 Simulación Monte Carlo:

La simulación Monte Carlo es una herramienta estadística no determinista que, con base en la generación de números aleatorios, permite modelar expectativas probabilísticas. Para esto se usa como base el comportamiento histórico de las variables y la probabilidad de ocurrencia.

El objetivo del método Monte Carlo consiste en aproximar cualquier cantidad de los parámetros (θ) asociada con la distribución posterior de los parámetros $P(\theta|y)$ por medio de una simulación, donde se utilizan métodos aleatorios.

Si podemos obtener muestras del parámetro θ procedentes de la distribución posterior, cualquier cantidad posterior de interés se puede aproximar con un grado de precisión arbitrario usando métodos de Monte Carlo (Stanislaw Ulam, John Von Neumann, Nicholas Metrópolis en Sosa 2021).

Dicho método es usado para encontrar la expectativa matemática de una o varias variables, que serían complejas de evaluar con exactitud, haciendo posible identificar estructuras dentro de las distintas variables que permitan determinar una relación causa - efecto. Para esto, se deben generar diversas muestras independientes de la variable aleatoria y posteriormente, se debe tomar el promedio empírico de la muestra como una estimación puntual de la expectativa. La precisión de este método es proporcional a:

Ecuación 9: Precisión de método Monte Carlo

$$\frac{\sigma^2}{\sqrt{n}}$$

donde σ^2 denota la varianza de cada muestra y n indica el número de muestras generadas.

2.3.2 Implementación del método Monte Carlo:

Sea θ un parámetro de interés y, $y = (y_1, \dots, y_n)$ un conjunto de observaciones realizadas sobre este parámetro.

Y, asumiendo que es posible obtener una muestra aleatoria de B valores asociados con la distribución posterior $p(\theta | y)$, se obtiene la siguiente ecuación:

Ecuación 10: Ecuación de muestras de θ

$$\theta(1), \dots, \theta(B) \sim p(\theta|y)$$

Donde desde $\theta(1), \dots, \theta(B)$ son variables aleatorias, independientes e idénticamente distribuidas.

La distribución empírica de $\theta(1), \dots, \theta(B)$ aproxima la distribución posterior y la aproximación incrementa su precisión a medida que se aumenta el valor de B .

Esta aproximación se conoce como la aproximación de Monte Carlo de $p(\theta|y)$

2.3.3 Ley fuerte de los grandes números:

La ley de números grandes muestra como una variable aleatoria converge al promedio a largo plazo al hacer un muestreo repetitivo y está dada por la formula:

Ecuación 11: Ecuación de ley fuerte de los grandes números

$$P(\lim_{n \rightarrow \infty} X^n = \mu) = 1$$

2.3.4 Ley débil de los grandes números:

Sea $X_1 \dots X_n$ una secuencia de variables aleatorias independientes e idénticamente distribuidas (iid) con media μ y varianza finita σ^2 . Entonces, para $\epsilon > 0$. Siendo ϵ cualquier número positivo y S una muestra de un conjunto de datos, el promedio muestral es igual a:

Ecuación 12: Promedio muestral – ley débil

$$p(|S - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

Sea

Ecuación 13: Demostración de Ley débil de los grandes números Usando Chebyshev

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Entonces:

Ecuación 14: Valor esperado de Sn

$$E(Sn) = \mu \text{ y } Var(Sn) = \frac{\sigma^2}{n}$$

Por lo tanto:

Ecuación 15: Probabilidad de obtener el promedio de la muestra a partir de ley débil de los grandes números

$$p\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| > \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2} \rightarrow 0$$

Se obtiene que:

Ecuación 16: Tendencia al promedio

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu$$

Cuando $n \rightarrow \infty$

2.3.4 Cadenas de Márkov:

Una cadena de Márkov es un tipo de proceso estocástico discreto en el que la probabilidad que ocurra un evento depende únicamente del estado anterior.

Matemáticamente, las cadenas de Márkov se pueden representar con la formula:

Ecuación 17: Representación cadenas de Markov

$$p(X_{t+1}|X_t, X_{t-1}, \dots, X_2, X_1) = p(X_{t+1}|X_t)$$

Para todos los $t = 2, \dots, n$ donde t representa el tiempo. Así, esta expresión se puede representar de forma más clara como:

Ecuación 18: Representación dependencias cadenas de Markov

$$p(X_1, X_2, \dots, X_n) = p(X_1) \cdot p(X_2|X_1) \cdot p(X_3|X_2) \cdot p(X_4|X_3) \dots \cdot p(X_n|X_{n-1})$$

Estas transiciones de estado se pueden realizar hasta que el estado de la cadena se encuentre estable, es decir, que la distribución de la matriz seguirá siendo la distribución de los estados y no tendrá cambios a partir de entonces. Así, se puede determinar el comportamiento de los diferentes estados a largo plazo. Esto es conocido como distribución estacionaria.

El método de cadenas de Márkov es muy usado para la simulación de Monte Carlo y este es denominado como Márkov Chain Monte Carlo (MCMC). Donde su principal característica es que las diferentes muestras están auto correlacionadas.

2.3.5 Proceso de Wiener:

El proceso de Wiener es un proceso estocástico de tiempo continuo utilizado en diversas ramas de la ciencia, en las finanzas y en las matemáticas.

El Movimiento Browniano o proceso estocástico de Wiener estándar, es denotado por $\{W(t): \geq 0\}$ con valores en el conjunto de R, es un proceso estocástico que cumple las siguientes propiedades:

1. $W_0 = 0$
2. W_t es casi seguro continuo
3. W_t tiene un incremento independiente
4. $W_t - W_s \sim N(0, t - s)$ (para $0 \leq s \leq t$)

“Hoy en día, el movimiento browniano ha adquirido gran influencia en numerosas áreas de las ciencias puras y aplicadas. Entre estas últimas, cabe citar la electrónica, la propia cibernética, la biología y la economía.” (Vélez Ibarrola, 2021) esto se puede expresar con la proposición:

Ecuación 19: Proposición movimiento Browniano

$$Si 0 < t_1 < t_2 < \dots < t_n$$

Entre el vector

Ecuación 20: Vectores

$$(W_{t_1}, W_{t_2}, \dots, W_{t_n})$$

Tiene una función de densidad

Ecuación 21: Función de densidad

$$f(x_1, x_2, \dots, x_n) = p(t_1, 0, x_1) \\ p(t_2 - t_1, x_1, x_2) \dots p(t_n - t_{n-1}, x_{n-1}, x_n)$$

Para la aplicación en la simulación de Monte Carlo se considera a W_n como una Cadena de Márkov.

2.3.6 Factorización o descomposición de Cholesky:

La descomposición de Cholesky es una manera de resolver procesos matriciales. Para esto la matriz debe ser positiva definida y simétrica.

Dentro del método de Monte Carlo, la factorización de Cholesky es utilizada para simular sistemas que tengan múltiples variables correlacionadas, buscando usualmente obtener la evolución de la variable independiente en función del tiempo.

Para realizar este procedimiento, la matriz de correlación entre variables se descompone, en búsqueda de obtener la triangular inferior “ L ”. Aplicando ésta a un vector de ruidos simulados cuando $r = 0$. El valor “ u ”, produce un vector “ Lu ” con las propiedades de covarianza del sistema a ser modelado.

2.3.7 Análisis de componentes principales (PCA):

El análisis de componentes principales, por sus siglas en inglés PCA es una técnica matemática utilizada para describir un conjunto de datos en términos de nuevas variables no correlacionadas. El PCA busca una proyección en la cual los datos obtenidos queden representados en términos de mínimos cuadrados. Convirtiendo un conjunto de observaciones correlacionadas en un conjunto de variables sin correlación denominadas componentes principales.

“Podría decirse que el método lineal utilizado con más frecuencia es el análisis de componentes principales (PCA). Es más efectivo si las dimensiones de los datos de entrada están correlacionadas, lo cual es común. PCA utiliza esta propiedad y encuentra las direcciones de los datos que contienen la varianza más grande. Esto es logrado al realizar la descomposición de valores propios de la matriz de covarianza de la muestra que se estima a partir de la entrada”. (Gortler J, et al, 2020)

El Análisis de componentes principales se emplea en el análisis exploratorio de datos y para construir modelos predictivos en diversas áreas del conocimiento debido a su fácil adaptabilidad.

Para calcular los componentes principales se debe considerar una serie de variables (x_1, x_2, \dots, x_p) de un grupo de objetos y a partir de estas se calcula un nuevo conjunto de

variables incorreladas (y_1, y_2, \dots, y_p) las varianzas de este nuevo conjunto deben decrecer progresivamente. Cada y_j (donde $j = 1, \dots, p$), es una combinación lineal de las variables originales (x_1, x_2, \dots, x_p)

Así se obtiene que:

Ecuación 22: Analisis de componentes principales

$$y_j = a_{j1}x_1 + a_{j2}x_2 + \dots + a_{jp}x_p = a'_j x$$

Siendo $a'_j = (a_{1j}, a_{2j}, \dots, a_{pj})$ un vector de constantes y

Ecuación 23: Ecuación de X PCA

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Teorema:

Sean t_1, t_2, \dots, t_n Los p vectores propios normalizados de una matriz de covarianza S, con la ecuación:

Ecuación 24: Ecuación de vectores normalizados

$$St_i = \lambda_i t_i \quad t_i t_i = 1, i = 1, \dots, p$$

Entonces:

1. Las variables compuestas $Y_i = X t_i, i = 1, 2, \dots, p$, son las componentes principales.
2. Las varianzas son los valores propios de S,

$$Var(Y_i) = \lambda_i \quad i = 1, 2, \dots, p$$

3. Las componentes principales son variables incorrelacionadas:

$$cov(Y_i, Y_j) = 0, i \neq j = 1, \dots, p$$

2.3.8 Prueba de Kolmogorov - Smirnov:

El estadístico de Kolmogorov–Smirnov, es la máxima distancia vertical que tienen las funciones de distribución acumulada o empíricas de dos muestras, o entre una función de distribución empírica y una función de distribución acumulada teórica tomada como muestra para la medición. La prueba de Kolmogorov-Smirnov es que es sensible a diferencias tanto en la localización de las funciones de distribución como en la forma de la distribución de estas.

“El procedimiento Prueba de Kolmogorov-Smirnov para una muestra compara la función de distribución acumulada observada de una variable con una distribución teórica determinada, que puede ser la normal, la uniforme, la de Poisson o la exponencial. La Z de Kolmogorov-Smirnov se calcula a partir de la diferencia mayor (en valor absoluto) entre las funciones de distribución acumuladas teórica y observada. Esta prueba de bondad de ajuste contrasta si las observaciones pudieran razonablemente proceder de la distribución especificada”. IBM (2021).

Definiendo una función de distribución acumulada (CDF) de una variable aleatoria de valor X o la función de distribución de X evaluado en x es la probabilidad de que X tome un valor igual o menor que x.

Así, para el caso de una función de distribución acumulada (CDF) dado F(x) la prueba Kolmogorov-Smirnov está dada por la ecuación:

Ecuación 255: Ecuación de vectores normalizados

$$D_n = \sup_x |F_n(x) - F(x)|$$

Donde \sup_x es el “supremum” del conjunto de distancias. Siendo el “Supremum” de un conjunto “S” de un conjunto parcialmente ordenado “P” el último elemento de “P” que es mayor o igual a cada elemento de “S”. Así, el Supremum puede definirse como el límite superior mínimo. Así, se toma la mayor diferencia absoluta entre las dos funciones de distribución en todos los valores x.

CAPÍTULO 3: ESTADO DEL ARTE

3.1 Mediciones y predicciones de la innovación tecnológica.

Se han realizado diversos estudios donde se intenta medir o predecir el comportamiento de la innovación tecnológica, usualmente con fines de conocer mejor este fenómeno y poder gestionarlo de manera eficaz. Entre las mediciones más relevantes se encuentran:

El ranking mundial de innovación realizado por “World Intellectual Property Organization” (WIPO) donde se intenta medir la innovación desde los enfoques institucional, capital humano e investigación, infraestructura, sofisticación del mercado, sofisticación de los negocios, conocimiento y productos tecnológicos y productos creativos.

Jiménez-Gómez. et al., (2016), que basado en la subjetividad actual de la medición de la innovación tecnológica y entendiendo que, en cada país, los factores relevantes son diferentes, proponen una serie de variables y un índice para medir de forma estadística la innovación tecnológica en Colombia, basando su estudio en datos de la Organización para la Cooperación y Desarrollo Económicos. (OCDE)

Posteriormente, Ding. S et al (2021), estudia de forma estadística la relación que hay entre la innovación tecnológica y la internacionalización. En éste se hace una revisión bibliográfica y se sustenta de forma empírica la relación positiva que hay entre estas dos variables en los países estudiados.

También, se han realizado diversos estudios utilizando como base para la medición de la innovación tecnológica el desarrollo de patentes de invención en una determinada nación. Pudiendo servir esto como base para la toma de decisiones con menor incertidumbre.

“La gestión de la innovación tecnológica analizada a partir del estudio de la información proveniente de las patentes de invención es un proceso clave para hacer el debido seguimiento al desarrollo tecnológico de los países y, por ende, fundamental, ya que a través de su análisis e interpretación se logran establecer tendencias tanto desde el punto de vista de áreas tecnológicas, como de mercados potenciales, competidores y posibles aliados en las diferentes etapas del proceso de gestión de innovación tecnológica” (Aponte-Figueroa, 2016)

Las patentes constantemente han sido usadas como criterio base de la medición tecnológica y para ellos se han usado diferentes parámetros para medir su impacto. Se ha considerado la cantidad de citas, las áreas del conocimiento, la entropía de Shannon y otras formas de discriminación para darles un valor determinado a las patentes en búsqueda de realizar un cálculo preciso del comportamiento de la innovación tecnológica. Así, se han logrado realizar mediciones y predicciones con un alto nivel de precisión.

“Las patentes tienen una gran cantidad de información sobre las tendencias del mercado de la tecnología y los flujos de conocimiento tecnológico. Un tipo de información de patentes es la citación entre patentes. Al explotar esta información de citas, no solo se pueden analizar las relaciones entre patentes sino también entre tecnologías” (Oh & Ayoung, 2017)

Debido al alto impacto que la innovación tecnológica tiene para el economía y bienestar social, se han realizado diversos estudios buscando parámetros para su medición estadística, en búsqueda de establecer parámetros que disminuyan la incertidumbre en la toma de decisiones sobre innovación tecnológica. Entre las variables más utilizadas para su medición se encuentran las patentes, debido a que son el reflejo del desarrollo de la ciencia y la tecnología.

Hay otros ítems que se pueden considerar para la medición de la innovación, En el manual de Frascati de la OECD (2015) se mencionan, aparte de las patentes. Los prototipos, plantas pilotos, diseños industriales, pruebas de producción, desarrollos de preproducción, desarrollo de servicios de posventa y solución de problemas y recopilaciones de datos como parámetros válidos para medir la innovación tecnológica.

Sin embargo, el uso de patentes y modelos de utilidad es el método más consistente y confiable debido a el soporte administrativo y legal que tiene el proceso así como lo refiere la OECD (2015) en el manual de Frascati.

3.2 Método Monte Carlo y la innovación tecnológica

Por el gran impacto que la innovación tecnológica tiene sobre la economía y la calidad de vida en las regiones, esta ha sido medida y estimada con diferentes herramientas estadísticas, entre ellas la simulación Monte Carlo.

P. Koh & D. Reeb (2015), exploran por medio de Simulaciones de Monte Carlo diferentes métodos para manejar la I+D, buscando formas de maximizar la innovación tecnológica dentro de las empresas por medio de la inversión y la gestión de la investigación y el desarrollo.

Posteriormente Azad & Ghodsypour (2017) realizan una simulación de la innovación tecnológica en el sector petróleo y gas usando el método Monte Carlo, en este se proponen y validan variables concretas que pueden afectar la dinámica del sistema de innovación tecnológica en este sector, además de proponer alternativas para su desarrollo.

También se han realizado estudios que contemplan diferentes parámetros para la estimación del modelo, incluyendo el propuesto por Zhang M., (2018) donde por medio de MCMC (Márkov Chain Monte Carlo) se analizan los factores que más influyen en la capacidad de innovación tecnológica.

“hay seis variables, inversión en I + D, aporte de personal de I + D, valor de importación y exportación, dependencia del comercio de exportación y aporte de ciencia y tecnología del gobierno Entre los 15 posibles factores de influencia, tienen una gran influencia en la capacidad de innovación tecnológica de las empresas industriales. Excepto por la entrada de personal de I+D y los volúmenes de importación, las otras cuatro variables están correlacionadas positivamente. Los nueve factores restantes no tienen un impacto significativo.” (Zhang M.,2018)

Zhang M. determina que existen 6 variables con un alto grado de significancia en la capacidad de innovación tecnológica entre 15 escogidas inicialmente. Todo lo anterior a partir de un modelo bayesiano (BMA) e incorporando MCMC.

El método Monte Carlo se ha utilizado anteriormente para el cálculo de la capacidad de innovación tecnológica y para intentar medir la influencia de diversas variables en la

misma. Sin embargo, aún hay un amplio campo de estudio de aplicación de la estadística bayesiana en la capacidad de innovación tecnológica y en la innovación tecnológica.

3.3 Simulación Monte Carlo en la administración:

Usualmente se relaciona la simulación Monte Carlo con las ciencias naturales y las predicciones que este método puede hacer en las mismas, no obstante, Monte Carlo es una simulación que puede ser aplicada en diversas áreas del conocimiento incluyendo la administración.

Dentro de la administración, la aplicación en las finanzas es el más usual, y cada vez se hace más popular; para Bolia & Jueneja (2005), los métodos de Monte Carlo se han convertido cada vez más en una herramienta computacional popular para valorar opciones financieras complejas, especialmente cuando el espacio subyacente de los activos tiene una gran dimensionalidad.

La simulación Monte Carlo puede y ha sido utilizada en diversos campos de estudio de la administración, lo que ha llevado a comprender de una forma más amplia los fenómenos que esta estudia.

“En las finanzas, la economía y los negocios, las aplicaciones más significativas (refiriéndose a la simulación Monte Carlo) se han dado en términos de la gestión de proyectos financieros, el cálculo del valor de la empresa, la construcción de escenarios financieros, la evaluación de decisiones de inversión de la empresa o de los derivados financieros, incluso para identificar dificultades y proponer acciones correctivas en una línea de producción en marcha, determinar políticas de ordenamiento en un sistema de inventario y otros sistemas de producción”.

(Salazar, Álzate, 2018)

Así, el método de simulación Monte Carlo tiene un gran potencial en el estudio de diversas áreas de la administración y tiene un gran potencial para su incorporación en nuevas temáticas.

No obstante, en la macroeconomía no hay muchos artículos disponibles que utilicen métodos bayesianos para la estimación de modelos, puesto que el método más usado para la predicción de parámetros es el frecuentista.

“La mayoría de los artículos disponibles han utilizado un enfoque frecuentista que adopta algoritmos basados en probabilidades o estimadores de momentos simulados. Por el contrario, los enfoques de estimación bayesianos se pueden encontrar solo en muy pocos artículos. Esto es sorprendente en la medida en que los métodos bayesianos se han convertido en el paradigma dominante en la estimación de modelos contemporáneos de equilibrio general dinámico en macroeconomía”. (Herbst, EP y Schorfheide, F en Lux T., 2021)

Aunque el método de la estadística bayesiana está altamente en concordancia con el método científico, este no es frecuentemente utilizado en la economía, por lo que los métodos bayesianos, incluyendo la simulación Monte Carlo, a pesar de que posee gran potencial dentro de las temáticas de la administración, no es frecuentemente utilizado.

3.4 Procesos de Weiner en la administración:

Los procesos de Weiner o Movimiento browniano han sido utilizados ampliamente dentro del campo de las finanzas y la economía, principalmente en el cálculo de variación en el precio de acciones, debido a ser un proceso de Márkov a tiempo continuo capaz de relacionar múltiples variables.

Se han realizado diversos estudios dentro del campo de la administración donde son utilizados los procesos de Weiner, entre los más destacados están:

Kim S (2022) utiliza procesos de Weiner en un modelo ARMA – GARCH para simular el comportamiento de diversos índices bursátiles y su volatilidad, concluyendo que este tipo de modelos describe con una alta aceptabilidad el comportamiento del proceso residual de retornos de alta frecuencia que el proceso hiperbólico generalizado no fraccionario o proceso gaussiano.

Sinha A., (2021) realiza estimaciones de acciones utilizando el movimiento browniano geométrico. Donde estima términos de deriva y difusión pronosticados estimados por separado y recursivamente se conectan al marco para pronosticar los valores del índice S&P500.

Di Scorio & Matiozzi (2020) desarrollan un método para calcular el precio de la opción de compra europea utilizando movimiento browniano multifraccional con función determinista

de Hurst. Los autores construyen un marco matemático utilizando un estudio de Lebovits et al. (2014) Los autores obtienen el precio de opción de compra europea, a través de la simulación del precio logarítmico.

A pesar del frecuente uso de los procesos de Weiner en la economía y finanzas, posterior a la revisión de la literatura, no se han encontrado estudios que utilicen procesos de Weiner para la estimación de la innovación tecnológica.

3.5 Análisis de componentes principales (PCA) en construcción de modelos económicos:

El análisis de componentes principales ha sido utilizado ampliamente en la construcción de modelos para la construcción de modelos predictivos.

Solano-Benavides et al (2019) construyen un modelo basado en Análisis de Componentes Principales para la proyección de la competitividad en Colombia. Para esto consideran diferentes factores que se integran para componer un indicador que sirve como base para crear un modelo predictivo para la competitividad.

Los modelos producidos por PCA mejoran su capacidad predictiva cuando se utiliza identificación previa que sirve como base para la elaboración del modelo. Este tipo de modelos tienen una mayor precisión que los de identificación como afirma Zhang L. (2019) quien propone una construcción de un modelo basado en Análisis de Componentes Principales usando información previa para realizar una identificación mejorada de subespacios de bucle. Aunque esta simulación se realiza por regresión de mínimos cuadrados, en el artículo realizado se evidencia como el método de PCA propuesto es un método preciso y estable, incluso más que otros modelos como el de identificación.

No obstante, el artículo también refiere como una información previa con mucho ruido o mal formulada puede dar lugar a la generación de modelos mal formulados que posteriormente ocasionan problemas en las simulaciones.

3.6 Simulación de Monte Carlo que utiliza la descomposición de Cholesky en economía:

La descomposición de Cholesky se ha utilizado dentro del método Monte Carlo para simular un parámetro utilizando diversas variables correlacionadas entre sí. Esto ha sido

particularmente útil en la administración dentro de la economía, la producción y las finanzas.

Medina & Restrepo (2013) hacen un estudio donde utilizando diversas variables económicas que busca obtener una estimación de la utilidad en riesgo de una empresa de energía eléctrica de Colombia. Para esto realizan simulaciones Monte Carlo utilizando descomposición de Cholesky y aplicando metodologías GARCH para la construcción del modelo.

Posteriormente, Shirota S, et al., (2021) aplican descomposición de Cholesky y modelos de volatilidad estocástica multivariada en modelos financieros aplicados en simulaciones Márkov Chain Monte Carlo. Este proceso tiene como objetivo disminuir problemas derivados de tener demasiados parámetros para estimar el riesgo de los activos del mercado.

En la economía y las finanzas es común encontrar modelos que apliquen descomposición de Cholesky para la integrar múltiples variables simulaciones Márkov Chain Monte Carlo, esto debido a su capacidad para simular múltiples variables correlacionadas en función del tiempo.

CAPÍTULO CUATRO: RESULTADOS Y ANALÍISIS

Para plantear de forma correcta el modelo que nos permita realizar el proceso de simulación empleando el método Monte Carlo de forma adecuada, es necesario dividir el proceso en siete pasos que permitan comprender de una forma más profunda el problema y las diferentes variables que lo componen. Siendo estos pasos: Entender y describir el problema a resolver, planificación y recolección de datos, exploración de los datos obtenidos, postular un modelo que se adapte a los datos, ajustar el modelo, revisar el modelo y validar el modelo.

4.1 Entender y describir el problema a resolver:

Para proponer un modelo que sirva para predecir la innovación tecnológica en Colombia, inicialmente se establecen como parámetro de medida el registro de patentes realizado por la Superintendencia de Industria y Comercio; posteriormente se establecen las variables que más influyen en la innovación tecnológica en Colombia se determina cómo se

correlacionan y, con base en los datos, en la bibliografía y en métodos de modelado previamente utilizados en la economía se establece un modelo para la predicción de la variación de la innovación tecnológica que, en este caso se realizará por PCA dada la naturaleza de los datos y de la predicción.

Posteriormente, para realizar las simulaciones de Monte Carlo se establecen matrices basadas en procesos de Markov, que evolucionan teniendo como parámetro de entrada la base de datos; las transiciones de las cadenas representan un año de evolución, ya que los parámetros de entrada son ingresados con este mismo intervalo de tiempo.

4.2 Planificación y recolección de datos:

1. Las patentes de invención y los modelos de utilidad concedidos serán obtenidos a partir de los datos registrados y publicados por la Superintendencia de Industria y Comercio de Colombia.
2. El producto interno bruto per cápita (PIB per cápita) desde el año 2000 hasta el año 2020 registrado por el Banco Mundial y llevado a moneda corriente en USD.
3. Las inversiones en Investigación y Desarrollo (I+D), dentro de Colombia desde el año 2000 hasta el año 2018 serán obtenidas del Banco Mundial como porcentaje del PIB total, los años 2019 y 2020 son obtenidos del Observatorio Colombiano de Ciencia y Tecnología (2020).
4. Los registros de calidad institucional serán obtenidos de Worldwide Governance Indicators (2020) tomando como referencia para la interpretación de los datos que dan Barbier y Joanne (2021) PP. 5 para la medición de la calidad institucional.
5. La penetración del internet será obtenida de estadísticas del Banco Mundial como porcentaje de la población.
6. Los datos de exportaciones son obtenidos de Expansión – Datos Macro como porcentaje del PIB.

Todos los datos anteriormente mencionados serán obtenidos a partir del año 2000 hasta el año 2020.

La variable de software es descartada debido a la alta varianza que existe en los datos; además, debido a las políticas actuales para registrar software en Colombia, es difícil

estimar cuantos de estos desarrollos son efectivamente derivados de una actividad innovadora.

4.3 Exploración de los datos obtenidos:

Con el fin de postular un modelo utilizando Análisis de Componentes Principales, que represente correctamente los datos recogidos, primero se debe hacer una exploración de éstos, buscando la correlación existente entre los diferentes parámetros establecidos. De esta forma se espera determinar que variables impactan sobre la innovación y en qué grado lo hacen y comprender, de mejor forma, el impacto que tienen estas variables sobre la innovación y sobre la construcción del modelo.

4.3.1 Exploración general:

Tabla 4 - 1: Exploración inicial de variables – Comportamiento histórico

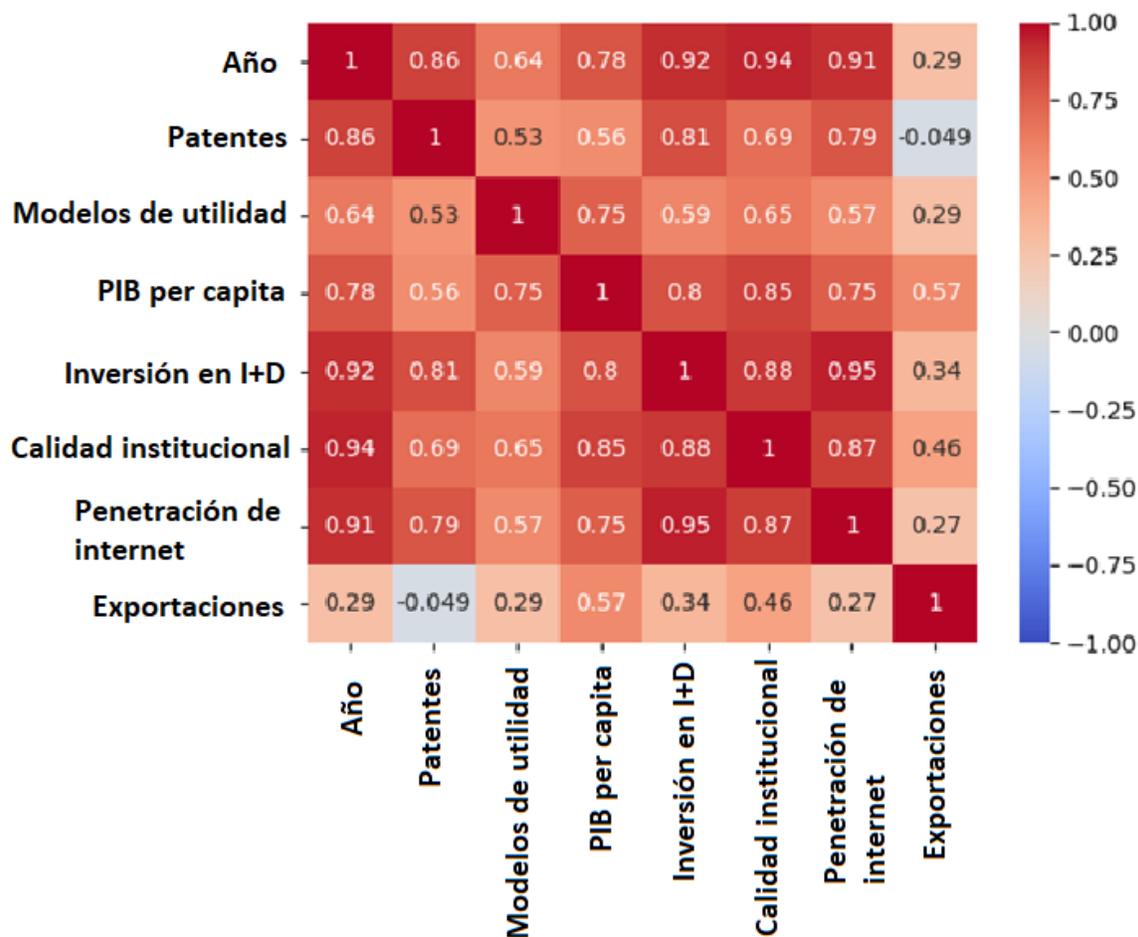
Año	Patentes de invención concedidas	Modelos de utilidad Concedidas	PIB per capita - US a precios actuales	Calidad institucional	Penetración de internet (% de la población)	Inversión I+D (% del PIB)	Exportaciones (% del PIB)
2000	23	35	\$ 2.520,48	34,94	2,208%	0,13	11,42%
2001	13	39	\$ 2.439,68	35,27	2,854%	0,13	12,21%
2002	13	40	\$ 2.396,63	35,60	4,600%	0,14	12,50%
2003	7	52	\$ 2.281,40	37,11	7,389%	0,16	11,83%
2004	10	39	\$ 2.782,62	38,11	9,119%	0,15	11,23%
2005	8	28	\$ 3.414,47	38,34	11,007%	0,15	12,16%
2006	11	37	\$ 3.741,09	40,46	15,342%	0,15	14,39%
2007	13	26	\$ 4.714,07	42,06	21,80%	0,19	15,40%
2008	31	62	\$ 5.472,54	42,49	25,60%	0,20	16,22%
2009	20	58	\$ 5.193,24	40,44	30,00%	0,20	16,99%
2010	26	50	\$ 6.336,71	41,68	36,50%	0,20	13,86%

2011	34	127	\$ 7.335,17	45,67	40,351%	0,21	14,13%
2012	109	124	\$ 8.050,26	44,01	48,98%	0,23	15,52%
2013	146	145	\$ 8.218,35	43,54	51,70%	0,27	14,54%
2014	117	75	\$ 8.114,34	44,16	52,57%	0,31	15,08%
2015	88	82	\$ 6.175,88	45,92	55,905%	0,29	14,55%
2016	103	61	\$ 5.870,78	45,85	58,136%	0,27	13,95%
2017	170	116	\$ 6.376,71	44,80	62,26%	0,24	13,91%
2018	219	97	\$ 6.729,58	45,01	64,126%	0,24	12,24%
2019	314	51	\$ 6.424,98	46,23	65,007%	0,32	12,60%
2020	244	118	\$ 5.334,55	45,81	68,00%	0,29	13,14%

En la tabla 4-1 se presentan los datos de los parámetros que se emplearán para este estudio, en función del año. A partir de estos datos, se buscarán las correlaciones entre variables

Correlación entre variables: Inicialmente se realiza una matriz de correlación de Pearson donde 1 es el valor de correlación positivo más alto, -1 el valor de correlación negativa más alto y 0 refleja que las variables no están correlacionadas. Las correlaciones se muestran en la siguiente gráfica de la ilustración 4-1

Ilustración 4 -1: Grafico de correlación inicial, empleando el método de correlación de Pearson



En la correlación de Pearson de datos anteriores se observa cómo, en concordancia con la bibliografía encontrada, las exportaciones en Colombia no tienen correlación alguna con el registro de patentes ni con los registros de modelos de utilidad. Esto se puede deber a que las exportaciones en Colombia son principalmente materias primas. Sin embargo, por lo que la variable queda descartada ya que se asume que los valores que esta toma no afectan de forma directa la innovación tecnológica.

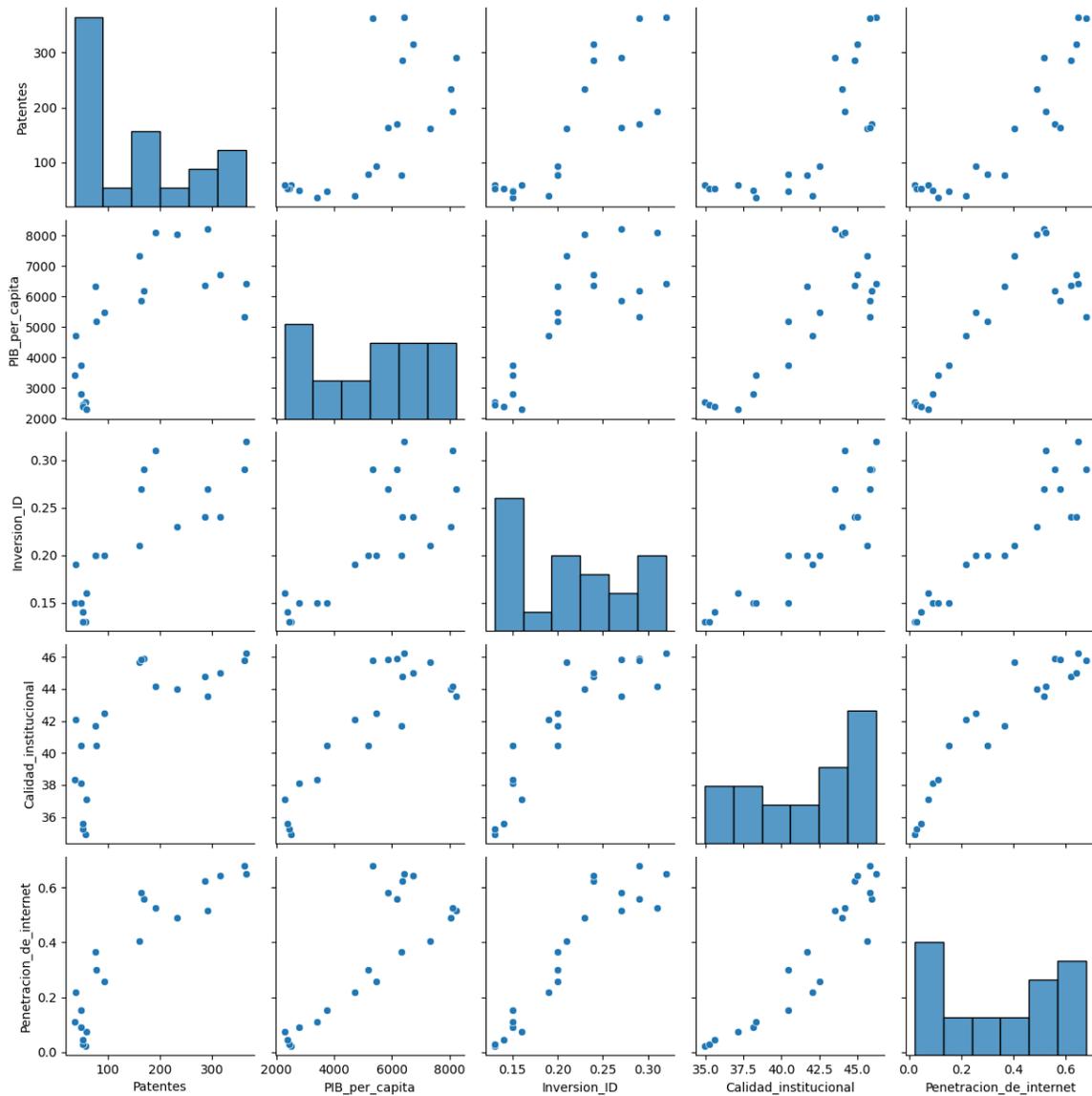
Por otro lado, se observa como las otras variables tienen altos coeficientes de correlación por lo que, sumado a la evidencia encontrada en la bibliografía, se asume que influyen de forma significativa en el desarrollo de innovación tecnológica en Colombia. Por lo que

serán analizadas y descritas para posteriormente incluirlas en el modelo planteado para la simulación.

Teniendo en cuenta que tanto las patentes de invención como las patentes de modelos de utilidad son resultados de actividades de innovación, se decide sumarlas a fin de establecer el parámetro de patentes como referencia final para medir la innovación tecnológica.

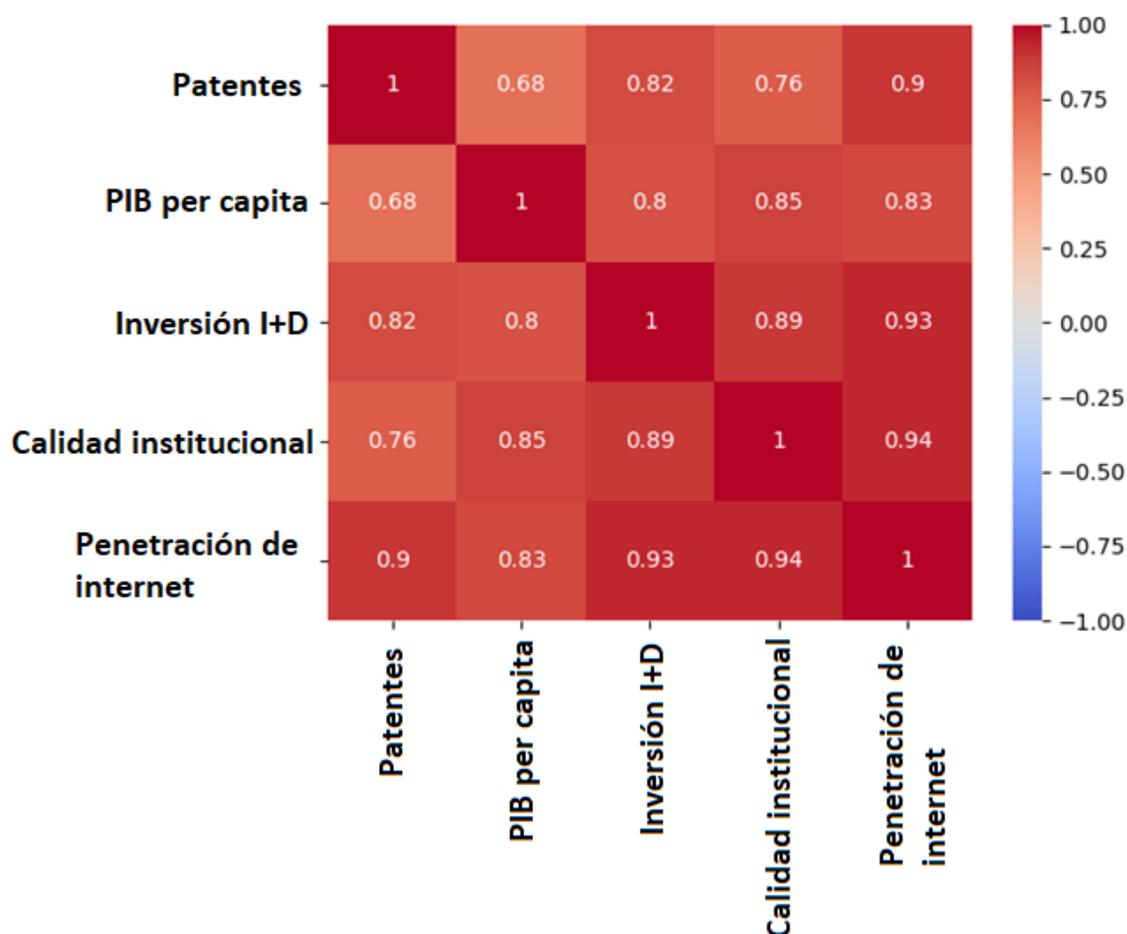
Par evidenciar el tipo de correlación y la forma que toman los datos se realiza la matriz de correlación de forma gráfica a fin de tener una mejor exploración de los datos. En la Ilustración 4-2, se presenta la matriz de correlación, en donde se descarta usar la variable de índice de exportación, pues, como se mencionó anteriormente en la ilustración 4-1, no se observó correlación con el número de patentes.

Ilustración 4- 2: Correlación grafica entre variables



El diagrama de correlación y dispersión muestra cómo se comporta una variable en relación con la otra. Entre más cercanos estén los puntos de los datos representa una correlación mayor; la forma que toman estos gráficos también expresa si la correlación es positiva/negativa, exponencial o correlación nula. En este caso, se observa como todos tienen una correlación positiva/negativa y los datos obtenidos muestran una evidente correlación. La correlación numérica, luego de la reducción y agrupación de variables, se muestra en la Ilustración 4-3.:

Ilustración 4 - 3: Correlación entre variables



Así, se evidencia la fuerte correlación positiva que hay entre las diversas variables seleccionadas para el desarrollo del modelo; posteriormente se describe el comportamiento individual de cada una de ellas, buscando conocer a fondo su comportamiento a lo largo del tiempo.

Análisis de variables por estadística descriptiva

Es necesario realizar una reinterpretación de los datos, posterior a descartar las variables que no tienen influencia sobre el modelo; además de agrupar las patentes de invención y modelos de utilidad. Las variables seleccionadas para realizar la simulación se muestran en la tabla 4-2:

Tabla 4 - 2: Histórico de datos variables

Año	Patentes	PIB per cápita	Inversión I+D	Calidad institucional	Penetración de internet
2000	58	\$ 2.520,48	0,13	34,94	2,21%

2001	52	\$ 2.439,68	0,13	35,27	2,85%
2002	53	\$ 2.396,63	0,14	35,6	4,60%
2003	59	\$ 2.281,40	0,16	37,11	7,39%
2004	49	\$ 2.782,62	0,15	38,11	9,12%
2005	36	\$ 3.414,47	0,15	38,34	11,01%
2006	48	\$ 3.741,09	0,15	40,46	15,34%
2007	39	\$ 4.714,07	0,19	42,06	21,80%
2008	93	\$ 5.472,54	0,2	42,49	25,60%
2009	78	\$ 5.193,24	0,2	40,44	30,00%
2010	76	\$ 6.336,71	0,2	41,68	36,50%
2011	161	\$ 7.335,17	0,21	45,67	40,35%
2012	233	\$ 8.050,26	0,23	44,01	48,98%
2013	291	\$ 8.218,35	0,27	43,54	51,70%
2014	192	\$ 8.114,34	0,31	44,16	52,57%
2015	170	\$ 6.175,88	0,29	45,92	55,91%
2016	164	\$ 5.870,78	0,27	45,85	58,14%
2017	286	\$ 6.376,71	0,24	44,8	62,26%
2018	316	\$ 6.729,58	0,24	45,01	64,13%
2019	365	\$ 6.424,98	0,32	46,23	65,01%
2020	362	\$ 5.334,55	0,29	45,81	68,00%

A continuación, se realiza una exploración descriptiva de cada variable para así observar cómo se comportan los datos obtenidos y obtener una mejor comprensión de las variables, tanto la variable dependiente como las independientes que influyen en el modelo.

Tabla 4 - 3: Estadísticas descriptivas

.	Patentes	PIB per cápita	Inversión en ID	Calidad institucional	Penetración de internet
mean	151,476	5234,454	0,213	41,786	0,349
Std	111,7	1962,09	0,06	3,759	0,229
min	36	2281,400	0,130	34,940	0,022
25%	53	3414,470	0,150	38,340	0,110
50%	93	5472,540	0,200	42,490	0,365
75%	233	6424,980	0,270	45,010	0,559
Max	365	8218,350	0,320	46,230	0,680

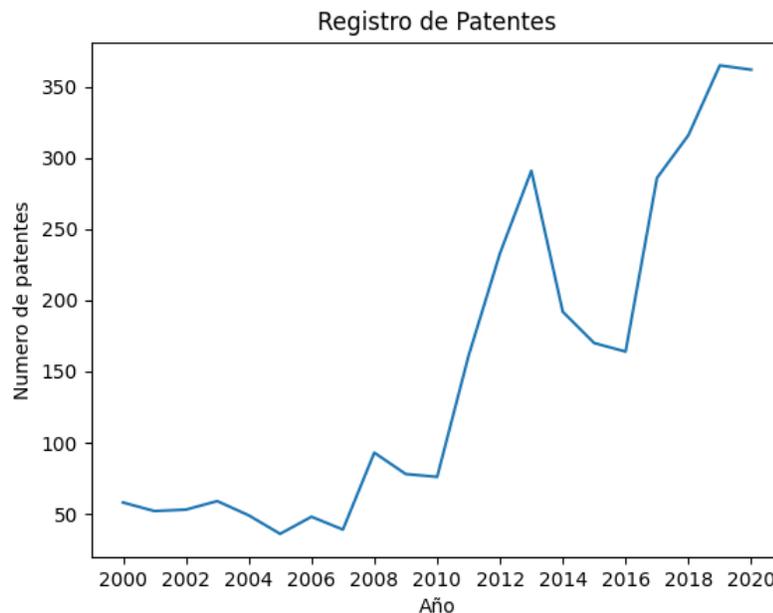
En la tabla 4-3 se presentan datos descriptivos sobre las variables que se consideran influyen en la innovación tecnológica en Colombia, presentando la media, desviación estándar, mínimos, máximos y los cuartiles. Siendo los cuartiles la división de los datos en el punto de 25%, 50% y 75%. Además, se evidencia cómo únicamente la penetración de internet ha tenido un crecimiento sostenido. Se observa que el año 2020 siempre presenta

los valores máximos en las variables. Esto se puede deber principalmente a la pandemia causada por el COVID 19, ya que el gobierno debió invertir recursos y esfuerzos en contrarrestar los efectos negativos que generó la pandemia en todos los aspectos, incluso en la investigación y el desarrollo.

En el Documento Conpes 4023 (2021) del Gobierno de Colombia donde se trata la política para la reactivación, repotenciación y el crecimiento sostenible e incluyente se menciona que el gobierno invertirá arduamente en la reactivación de la ciencia, la tecnología e innovación en el periodo postpandemia en búsqueda de la reactivación económica y social.

El registro de patentes en Colombia tiene una desviación estándar muy amplia, lo que implica que en general, los registros de patentes se encuentran muy lejos de la media, lo que a su vez implica una gran fluctuación en los datos. No obstante, y aunque fluctuante entre años, se observa que en general hay una tendencia creciente en el registro de patentes de Colombia que se acentúa a partir del año 2010.

Ilustración 4 – 4: Registro de patentes



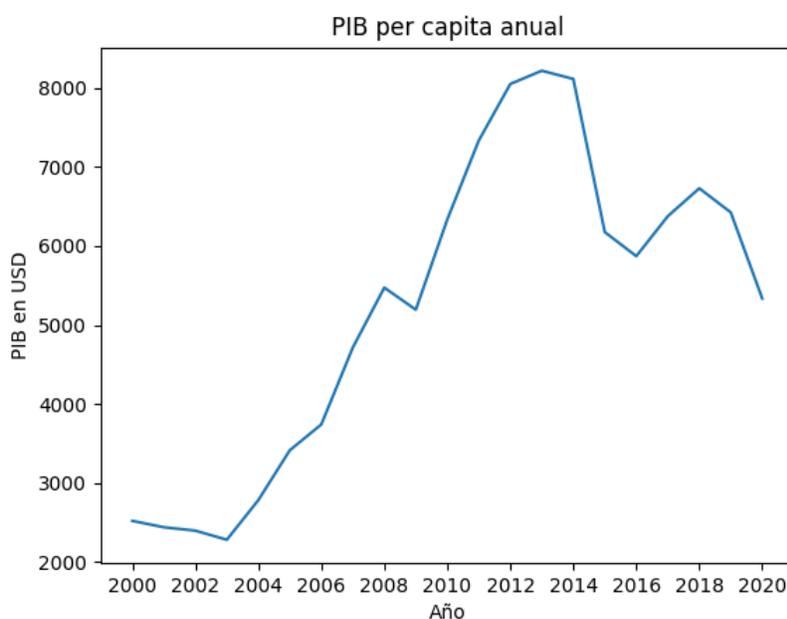
En la ilustración 4-4 se muestra la tendencia de las patentes en Colombia, en función del tiempo, medido en años. La disminución en el número de patentes sucede en el año 2014 de

forma casi simultánea que la caída significativa del PIB per cápita. Sin embargo, el registro de patentes volvió a tener una tendencia al aumento posterior al 2016 mientras que el PIB per cápita siguió disminuyendo. Para CEPAL (2016) la disminución en el PIB per Cápita se debe a la reducción del ritmo de crecimiento económico entre los años 2014 y 2016 situándose en torno al 2%. Además de los precios bajos de los hidrocarburos, se produjo una reducción de la oferta de energía y de bienes agrícolas a causa del fenómeno de El Niño y del paro del transporte de carga en julio del año 2016.

En la ilustración 4-5 se evidencia esta disminución en el PIB per cápita y como a partir de 2014, misma fecha donde se registró la disminución en la producción de patentes se registra una disminución en el PIB per cápita. Sin embargo, aunque la producción de patentes vuelve a incrementar a partir del año 2016, el PIB per cápita sigue disminuyendo.

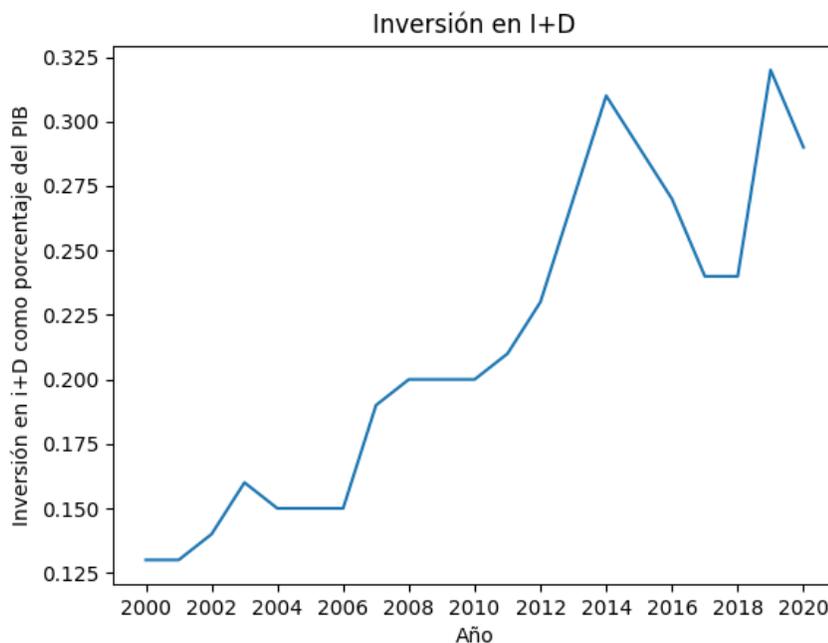
Esto quiere decir, que, aunque puede existir una relación entre el número de patentes por año y el PIB, es necesario tener en cuenta otras variables; además, es de recordar que el proceso de patentes puede tomar de 1 a 5 años, posterior a la invención y a su registro. Lo que puede significar que una de las condiciones favorables para patentar no necesariamente genera buenos resultados de inmediato.

Ilustración 5 – 5: PIB per cápita anual



La tendencia a la disminución alrededor de 2016, también se ve reflejada en la inversión en I+D que hace Colombia anualmente, como se presenta en la Ilustración 4-6. El impacto sobre esta inversión se ve acrecentado considerando, que el PIB también disminuye durante los años 2014 – 2016 y la inversión en I+D está representada en porcentaje del PIB. Esta desaceleración no se ve reflejada de forma simultánea en la inversión en I+D (Ilustración 4-6), pero, si se observa dos años después, la inversión en I+D vuelve a aumentar mientras que el PIB en USD sigue decreciendo. Esto se puede deber a que, en los últimos años, se han desarrollado estrategia para invertir recursos en I+D, de otras fuentes diferentes a las provenientes del PIB, como es el caso de los recursos provenientes de las regalías departamentales. Es así, se observa un poco de inversión, en el año 2019, cuando se creó el Ministerio de Ciencia, Tecnología e Innovación mediante la ley 1955 de 2019, y se inició el desarrollo de convocatorias con recursos de Sistema General de Regalías (SGR), por parte de este ministerio. Posiblemente la disminución en el 2020 de la inversión en I+D, disminuyó a razón de la pandemia, y a que los recursos de regalías son bianuales y la mayor parte de la inversión se lleva a cabo en el primer año, es decir, en el 2019, quedando unos remanentes para el 2020.

Ilustración 6 – 6: Inversión en I+D



Para el caso de la inversión en I+D como % del PIB, este tiene un comportamiento que tiende al alza, pero es de notar como al igual que el PIB per cápita la inversión en I+D tiene una disminución importante en el año 2014. La gran correlación que tienen estas variables se evidencia en el índice de correlación de Pearson que es de 0,8.

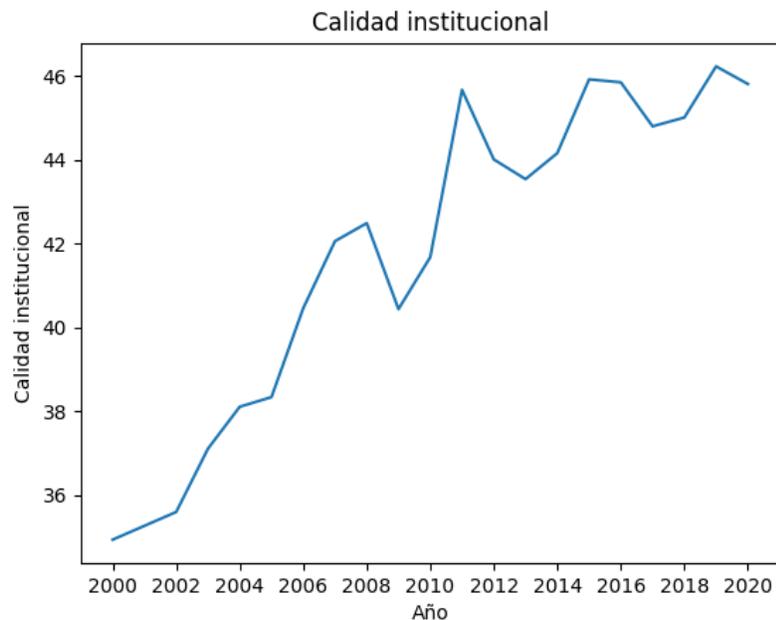
Ilustración 7 – 7: Histórico PIB



Fuente: DANE en Ciguenza Riaño (2018)

En la ilustración 4-7 se presenta el crecimiento del PIB entre 2009 y 2017. Se observa que, después de tener un aumento en el 2011, presentando un valor de 6.6%, inició una disminución hasta llegar a valores aproximados del 2%. La disminución en el PIB, puede ser producto de la disminución en el precio del petróleo, pero puede explicar también el decrecimiento en las variables de registro de patentes, PIB per cápita e inversión en I + D.

Ilustración 8 – 8: Calidad institucional



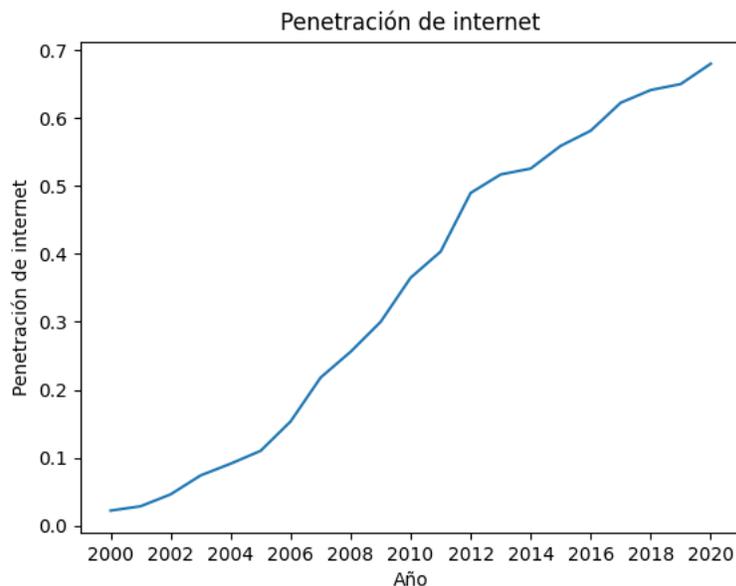
En la ilustración 4-8 se presenta el gráfico del índice de calidad institucional en función del tiempo, la calidad institucional se refiere al impacto de las instituciones en el desarrollo de la sociedad, y en la seguridad que se tiene en las instituciones para el resguardo de los habitantes. Para la medición de esta variable se toma como base la interpretación de Barbier y Joanne (2021).

Este incremento en la calidad institucional de Colombia puede deberse a la reducción de los índices de terrorismo, un mayor nivel de control a casos de corrupción y a un mayor nivel de participación de las personas para seleccionar a sus gobernantes y en la toma de decisiones en general. Esto ha realizado un incremento del nivel de calidad institucional entre los años 2000 y 2020.

Por último, se muestra como el acceso a internet de las diversas regiones de Colombia ha tenido permanentemente una pendiente positiva, de acuerdo con la Ilustración 4-9. El incremento permanente del acceso a internet contribuye fuertemente a la capacidad innovadora de las diversas entidades que tienen la posibilidad de investigar y realizar desarrollo tecnológico. Para el caso de las patentes, el acceso a internet es muy importante, pues, una de las partes principales de este proceso, es hacer un adecuado “estado de la

técnica”, que no es más que una búsqueda bibliográfica que permita determinar la novedad del desarrollo.

Ilustración 9 – 9: Penetración de internet



Después a realizar un análisis descriptivo de las diversas variables que tienen impacto sobre la innovación tecnológica en Colombia, se puede observar que existe una tendencia al crecimiento. Sin embargo, se debe considerar que, pese a la alta correlación que tienen las variables, estas no siempre responden a los mismos estímulos macroeconómicos y sociales. Por esta razón, es difícil encontrar por métodos convencionales patrones para predecir el comportamiento de la innovación tecnológica en próximos años. Así, se hace necesario realizar un análisis por métodos numéricos. Para esto, se realizará un modelo sencillo, tomando como base un modelo donde el peso de cada variable independiente y dependiente son establecidos por el método de Análisis de Componentes Principales (PCA) que se explicará en la sección 4.5.

A continuación, se propondrá un proceso de Monte Carlo basado en modelos de Weiner y factorización de Cholesky.

4.4 Postular un modelo que se adapte a los datos:

Se propone un modelo realizado por el método de Análisis de Componentes Principales (PCA). A partir de este método, se propone un modelo basado en las covarianzas de todas

las variables, respecto a la variable principal. Así, la base para realizar el modelo será la proporción de varianza explicada que se obtiene a partir del método de PCA.

Además, cabe destacar que el tiempo “t” en la simulación esta dado por la trasposición de estados matriciales en diseño de cadenas de Márkov; así que, los resultados obtenidos para un año “n”, únicamente depende del año “n – 1”.

4.5 Ajustar el modelo:

La proporción de varianza explicada, dentro del método PCA, es utilizada para definir cuanta información sobre el parámetro captura cada una de las variables del componente principal.

Para realizar el cálculo de la proporción de varianza explicada, los datos para el análisis tienen que estar centrados a media 0; para esto, se resta la media a cada columna.

Así, en las ecuaciones 25 y 26 t_a representa las proyecciones de X en p_a y T es la matriz transpuesta. El error producido luego del ajuste de los datos se acumula en una matriz E.

Ecuación 266: Centrar datos a media 0

$$X = \sum_{a=1}^l t_a p_a^T + E$$

Cabe destacar que la sumatoria de la proporción de varianza explicada acumulada es siempre 1. La proporción de la variable explicada del componente se calcula con la siguiente ecuación:

Ecuación 277: Calculo de variable explicada

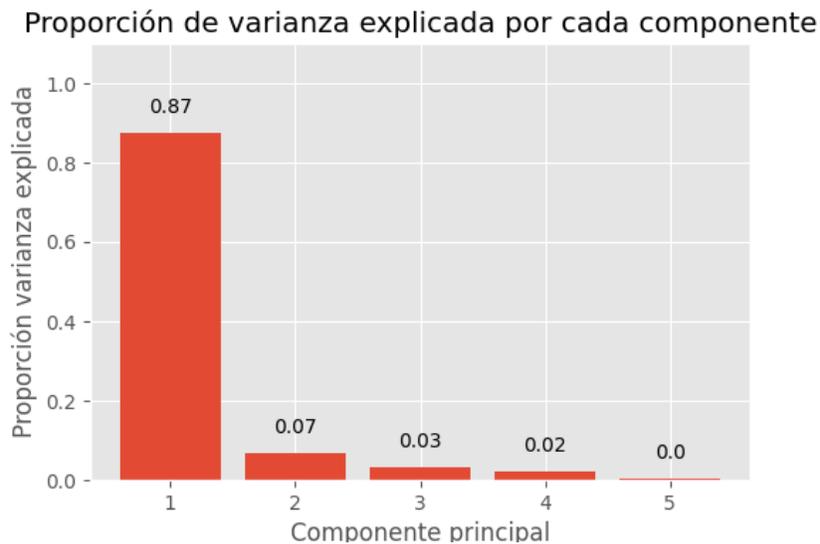
$$\sigma^2(X) = \frac{X^T X}{n - 1}$$

$$\sigma^2(X)p_a = \lambda_a p_a$$

Siendo λ_a la cantidad de varianza capturada que será utilizada para representar el modelo.

Aplicando las fórmulas en un algoritmo de se obtiene la gráfica de la ilustración 4-10

Ilustración 10 – 10: Proporción de Varianza Explicada por cada componente



La cual es una representación visual del modelo representando cuanta información captura cada una de las variables del componente principal. Siento 1 las patentes, 2 el PIB per cápita, 3 la inversión en I+D, 4 la calidad institucional y 5 la penetración de internet. Así, se obtiene el siguiente modelo:

Ecuación 288: Modelo de medición de innovación tecnológica

$$\text{Innovación Tecnológica} = 0.87381757 \text{ Patentes} + 0.06854012 \text{ PIB per cápita} + 0.03221508 \text{ Inversión ID} + 0.02044584 \text{ Calidad institucional} + 0.00498138 \text{ Penetración de internet}$$

Los valores indicados en el modelo representan que tanto los datos de cada una de las variables sirven para explicar el parámetro que se quiere medir. Así, un coeficiente más cercano al 1 significa que los datos de la variable logran explicar muy bien el parámetro. Así, es normal que los registros de patentes anteriores expliquen en gran medida la variabilidad de estas y por esto se le asigna un valor tan cercano al uno, las demás variables consideradas complementan el modelo.

Este modelo será utilizado para el desarrollo de la simulación.

4.6 Revisión del modelo y obtención de resultados

Después de obtener el modelo se realiza la simulación de Monte Carlo, tomando como base el modelo propuesto, buscando simular el número de patentes al año 2030.

4.6.1 Revisión de las ecuaciones del proceso de Weiner:

Los procesos de Weiner se aplican al método Monte Carlo en dos partes, para realizar el cálculo de la tasa de crecimiento y la volatilidad del modelo:

Cálculo de la tasa de crecimiento:

El cálculo de la tasa de crecimiento se obtiene por las propiedades de la distribución Lognormal donde:

Siendo X una variable aleatoria cuyo logaritmo se distribuye normalmente, es decir:

Ecuación 299: Distribución del logaritmo en calculo del crecimiento

$$\log X \sim N(\mu, \sigma^2)$$

Se pueden calcular su tasa de crecimiento exactamente, específicamente, dado que la variable $Y \equiv \log X$ tiene la función generadora de momento

$MY(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$, se tiene que:

Ecuación 30: Tasa de crecimiento

$$E[X] = \text{Exp}\left(\mu + \frac{\sigma^2}{2}\right)$$

Considerando que se calcula con base en una distribución Lognormal, se tiene que la tasa de crecimiento está dada mediante la ecuación:

Ecuación 31: Tasa de crecimiento

$$\text{Tasa de crecimiento} = \mu - \frac{\sigma^2}{2}$$

Cálculo de la volatilidad:

La volatilidad es una medida para la frecuencia e intensidad de los cambios en un conjunto de datos, este se calcula como la desviación estándar de dicha base de datos en un horizonte temporal. Para el presente modelo se define como:

Ecuación 32: Volatilidad del modelo

$\text{Volatilidad} = \sigma * Z(0.87381757 \text{ Patentes} + 0.06854012 \text{ PIB per cápita} + 0.03221508 \text{ Inversión ID} + 0.02044584 \text{ Calidad institucional} + 0.00498138 \text{ Penetración de internet})$

Ecuación 33: Ecuación de la volatilidad simplificada

$$\text{Volatilidad} = \sigma * Z(\text{ModeloPCA})$$

Aplicados al modelo propuesto anteriormente en la ecuación 27 y buscando la correlación de la aplicación del modelo mediante Factorización de Cholesky con la ecuación 33:

Ecuación 34: Factorización de Cholesky

$$[A] * [X] = [B]$$

$$[A] = [L] * [U]$$

Es decir

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & a_{1,5} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & a_{2,5} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & a_{3,5} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} & a_{4,5} \\ a_{5,1} & a_{5,2} & a_{5,3} & a_{5,4} & a_{5,5} \end{bmatrix} = \begin{bmatrix} l_{1,1} & 0 & 0 & 0 & 0 \\ l_{2,1} & l_{2,2} & 0 & 0 & 0 \\ l_{3,1} & l_{3,2} & l_{3,3} & 0 & 0 \\ l_{4,1} & l_{4,2} & l_{4,3} & l_{4,4} & 0 \\ l_{5,1} & l_{5,2} & l_{5,3} & l_{5,4} & l_{5,5} \end{bmatrix} * \begin{bmatrix} 1 & u_{2,1} & u_{3,1} & u_{4,1} & u_{5,1} \\ 0 & 1 & u_{3,2} & u_{4,2} & u_{5,2} \\ 0 & 0 & 1 & u_{4,3} & u_{5,3} \\ 0 & 0 & 0 & 1 & u_{5,4} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Donde a es una matriz de covarianzas de las variables definida así:

	Patentes	PIB per cápita	Inversión I+D	Calidad institucional	Penetración de internet
Penetración de internet	0.11458	0.00930	-0.00075	0.00156	-0.00336
Calidad institucional	0.00930	0.01777	0.00241	0.00082	0.00479
Inversión I+D	-0.00075	0.00241	0.01185	0.00071	0.00300
PIB per cápita	0.00156	0.00082	0.00071	0.00104	0.00108
Patentes	-0.00336	0.00479	0.00300	0.00108	0.02095

Para encontrar los elementos que están por debajo de la diagonal principal de [L] se utiliza la ecuación:

Ecuación 35: Ecuación diagonal L

$$l_{k,i} = \frac{[a_{k,i} - (\sum_{j=1}^{i-1} l_{i,j} * l_{k,j})]}{l_{i,i}}$$

$$k = 1 \rightarrow N$$

Para hallar los elementos dentro de la diagonal principal se utiliza la ecuación:

Ecuación 36: Ecuación diagonal principal

$$l_{k,k} = \sqrt{a_{k,k} - \sum_{j=1}^{k-1} l_{k,j}^2}$$

Así se obtiene la matriz:

Ecuación 37: Matriz L

$$L = \begin{bmatrix} 0.3384 & 0 & 0 & 0 & 0 \\ 0.0274 & 0.13044 & 0 & 0 & 0 \\ -0.0022 & 0.01892 & 0.1071 & 0 & 0 \\ 0.0046 & 0.0053 & 0.0057 & 0.0309 & 0 \\ -0.0099 & 0.0388 & 0.0209 & 0.0258 & 0.1350 \end{bmatrix}$$

Se debe tener en cuenta que la matriz [U] es la transpuesta de la matriz [L], es decir:

Ecuación 38: Ecuación Matriz U

$$[U] = [L]^T$$

Y la Matriz LU es:

Ecuación 39: Matriz LU

$$Lu = \begin{bmatrix} 0.0470 & 0.5605 & 0.0708 & -0.1102 & 0.0949 \\ -0.0313 & 0.1274 & 0.1615 & 0.2142 & 0.2247 \\ -0.0432 & -0.1242 & 0.0540 & -0.1735 & 0.0801 \\ -0.0052 & -0.0173 & -0.0027 & -0.0306 & -0.0547 \\ -0.0772 & 0.1719 & 0.1785 & 0.2449 & 0.0959 \end{bmatrix}$$

Como punto de partida del modelo, se establece el valor de la variable al año 2020 y los datos aleatorios son obtenidos con la semilla “171239” utilizando la librería numpy.random. Esto se ve con mayor claridad en el anexo 4.

Es de mencionar que la simulación funciona como un proceso de Márkov donde cada año de la simulación es representado como una matriz y basando sus probabilidades de variación en la matriz Lu previamente obtenida.

Así, se genera una matriz Z basada en la simulación de números aleatorios donde se obtiene un multiplicador de desviación estándar de cada uno de los años por cada una de las variables.

Posteriormente se obtienen los valores anuales, usando el producto interno ordinario de vectores para las matrices Lu y una matriz obtenida de: $(Volatilidad + Desviación\ estándar * Z)$

Este proceso se repite un número determinado de veces a fin de obtener un valor medio acumulativo para las simulaciones.

Para el análisis de los datos se realizan simulaciones con 10.000, 20.000, 50.000 y 100.000 intentos buscando el punto de estabilidad de las cadenas de Márkov utilizadas en el modelo.

Ilustración 11 – 11: Tendencia de patentes 2030 (10.000, 20.000, 50.000 y 100.000 simulaciones)

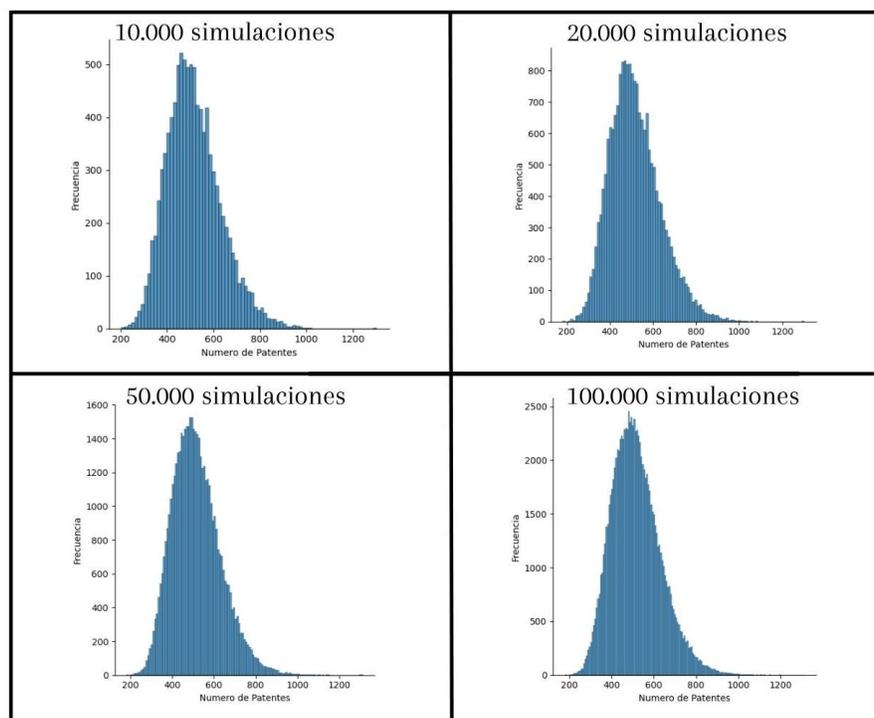


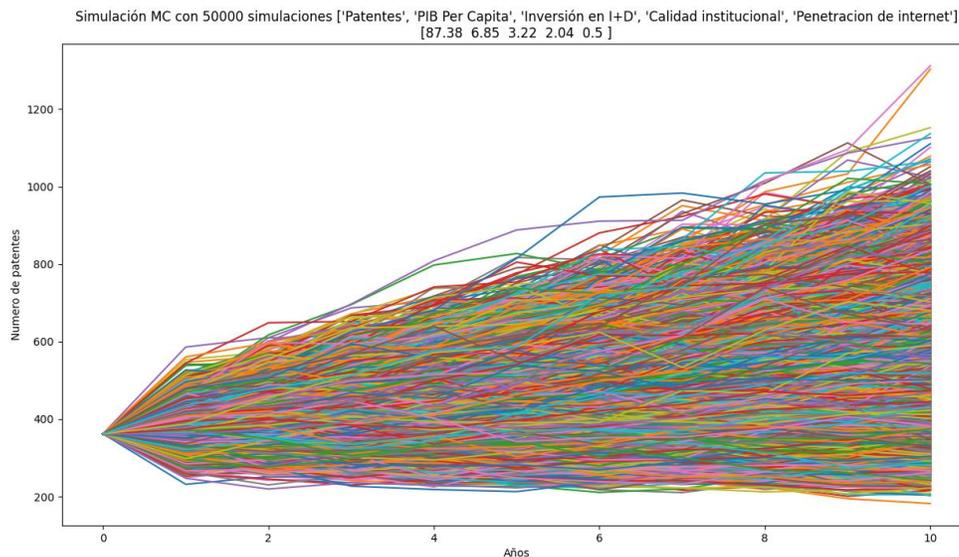
Tabla 4 – 4: Estadísticas descriptivas de las simulaciones

Numero de simulaciones	10.000	20.000	50.000	100.000
Media	440,95	441,13	440,7	440,93
Desviación estándar	90,39	90,9	90,45	90,53
Max	1303,06	1303,06	1311,86	1311,86
Min	200,81	182,06	182,06	182,06
1%	297,21	297,56	297,36	297,10
5%	333,47	333,45	333,01	333,18
95%	614,70	616,28	614,87	615,22
99%	727,89	728,41	724,28	725,14
Cuartil 1	369,56	369,2	369,05	369,30
Cuartil 2	422,68	422,32	422,07	422,34
Cuartil 3	487,81	488,49	488,01	488,33

En la ilustración 4-11 se presentan los resultados en histogramas, es decir, representaciones gráficas de las variables en forma de barras que en este caso representan la frecuencia de aparición en función del número de patentes obtenidas. Estos histogramas son construidos con base en los resultados de las simulaciones del número de patentes a 10 años. Se establecen puntos de muestreo en 10.000, 20.000, 50.000 y 100.000 simulaciones buscando la cantidad donde se encuentre un punto donde no haya fuertes variaciones entre los resultados de las simulaciones. Las cantidades obtenidas de la simulación y representadas en la ilustración 4-11 se presentan en la tabla 4 – 4. No se observan cambios significativos en los valores, solo de algunas pocas cifras decimales, dependiendo del número de simulaciones. Esto indica que, el modelo muy estable, y que no varía de forma significativa ante variaciones en la cantidad de simulaciones realizadas. Sin embargo, se pueden evidenciar ligeros cambios en los mínimos y máximos de las simulaciones entre el grupo de 10.000/20.000 simulaciones y el grupo de 50.000/100.000 simulaciones. Por lo anterior, se utiliza el modelo obtenido con 50.000 simulaciones, ya que se considera que a partir de este punto el modelo alcanza la estabilidad y es eficiente en uso de recursos computacionales. Con esta misma base se simularán posteriormente las variables independientes.

Así, se obtienen los resultados que se presentan a continuación en la ilustración 4-12 utilizando 50.000 simulaciones.

Ilustración 12 – 12: Simulaciones gráficas a 10 años

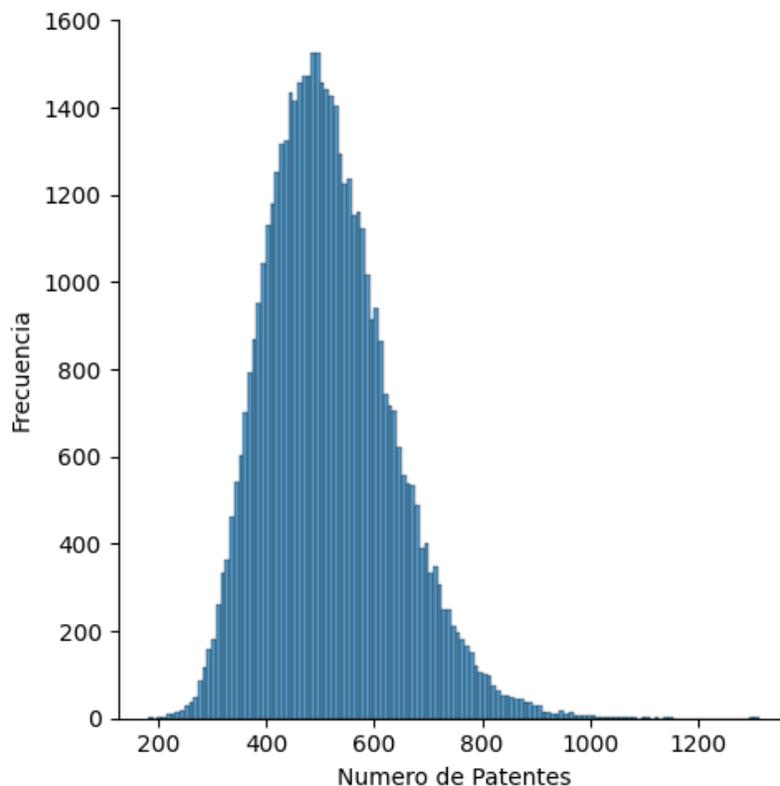


En la ilustración 4-12 se representa el número de patentes en función del tiempo en años. Cada una de las líneas de color representa una simulación, estas están dadas por los valores aleatorios que toman las variables en la matriz de distribución aplicando el modelo descrito en la ecuación 27.

Considerando la media de aproximadamente 440 casos que arroja el modelo se puede predecir un comportamiento desacelerado en el registro de patentes en Colombia. El incremento de números mínimos y máximos al incrementar las simulaciones se debe a el mayor número de escenarios que el modelo predice.

En la Ilustración 4-13 se puede observar la gráfica de distribución donde es evidencia de mejor forma el comportamiento de los datos de la simulación.

Ilustración 13 – 13: Grafica de distribución a 10 años



La prueba de normalidad de Kormogolov – Smirnov usando la librería Scipy.Stats arrojan los siguientes resultados:

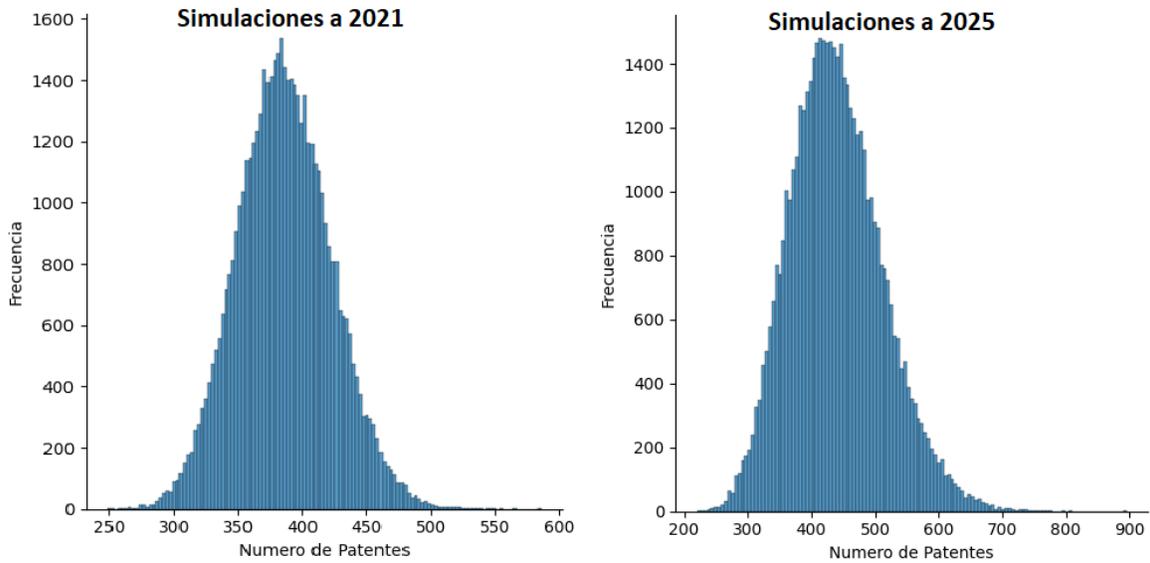
Statistics = 0.049 p-valor = 2.0223117796109084e-228

Así, el tener un p-valor tan cercano a cero indica que la distribución de la muestra no es normal o gaussiana. Esto puede deberse a la amplitud de los datos. Sin embargo, es importante observar como la mayor cantidad de datos tiende a la media de 440 patentes.

gráfica abarca un amplio conjunto de datos, desde 200 hasta 1303 patentes registradas para este periodo, sin embargo, para el análisis se debe observar cómo los valores a partir del 25% y del 75% son muy dispersos y casi se pueden considerar ruido. Por lo cual se establece que el valor de las patentes para el año 2030 se debería encontrar entre 422 y 488 patentes registradas en ese año.

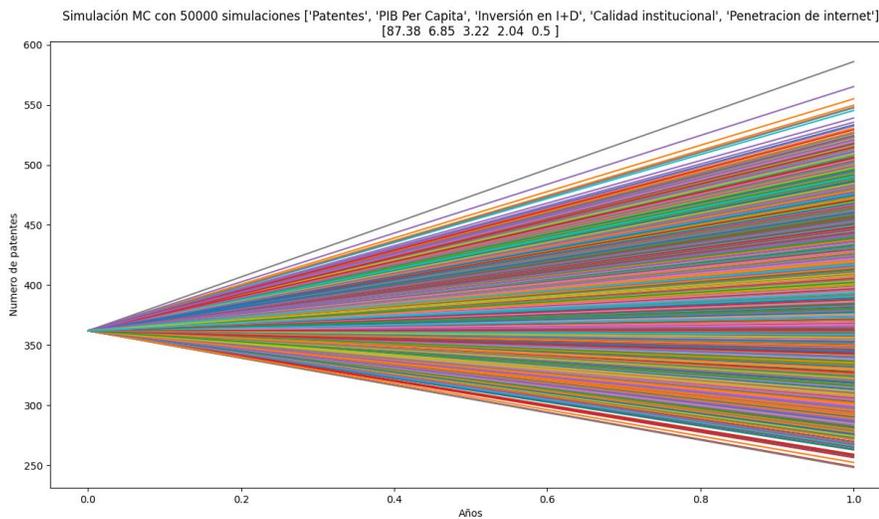
Posteriormente se simulan los datos al año 2021 y al año 2025 para establecer una tendencia, esto arroja las siguientes gráficas:

Ilustración 14 – 14: Simulaciones de patentes a 2021 y 2025



En las gráficas se muestra como a medida de que el tiempo se aleja del año inicial se incrementa en ancho de la distribución, aunque su media se mantiene muy cercana a medida que pasan los años, ya hay datos aislados que llegan a las 1.000 patentes, aunque estas son un muy pequeño porcentaje de las simulaciones realizadas.

Ilustración 15 – 15: Simulaciones de patentes por líneas 2021



Como se observa en la gráfica para el año 2021 el proceso de calculo es directo y unicamente calculado a partir del valor inicial de la simulación ya que no hay ninguna transición de una matriz a otra. Además en esta gráfica se muestra de mejor forma como es el proceso de transición entre un año n y un año $n+1$ a partir de un valor dado.

Tabla 5 – 5: Estadísticas descriptivas de patentes por años

	2021	2025	2030
Media	374,07	404,12	440,93
Desviación estándar	28,7	57,93	90,53
Max	585,97	894,89	1311,86
Min	248,22	220,72	182,06
1%	313,5	300,02	297,10
5%	339,54	330,76	333,18
95%	433,79	513,88	615,22
99%	464,47	579,39	725,14
Cuartil 1	362,0	362,0	369,30
Cuartil 2	362,0	392,18	422,34
Cuartil 3	385,11	436,91	488,33

La tendencia que muestran los datos es a incrementar, no obstante, se debe considerar que el registro de patentes en Colombia desde el año 2000 ha tenido una gran variabilidad y aunque ha tenido años estables, pero en algunos ha habido alzas o reducciones significativas, pero, el modelo contempla un amplio rango de posibilidades. Incluso únicamente simulando el año inmediatamente siguiente del año base, el modelo tiene un rango de entre 248 y 586 patentes registradas.

En las simulaciones se muestra una leve tendencia a incrementar a medida en que avanzan los años, no sólo basado por la tendencia de la variable dependiente, sino también la tendencia de las variables independientes.

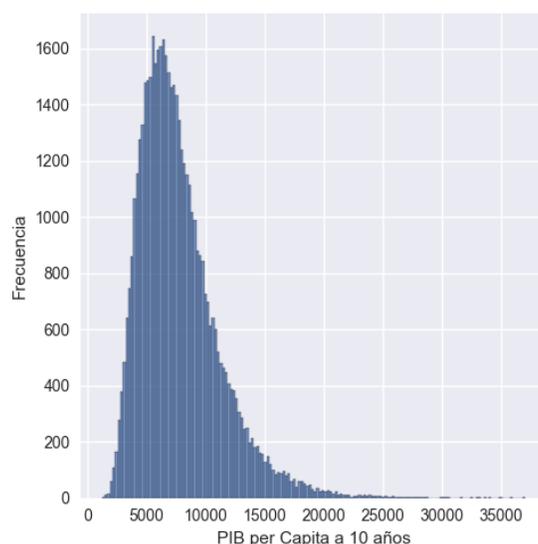
A medida que incrementan los años también se aumenta la desviación estándar y la dispersión de los datos, por lo que a su vez se ve una mayor amplitud en las gráficas y unos mínimos y máximos más lejanos a la media.

Así, se estima que en Colombia se registrarán para el año 2021 aproximadamente 374 patentes, para el año 2025 serán 404 y para 2030 serán cerca de 441. Reportando así la simulación un incremento sostenido en el registro de patentes por parte de residentes en Colombia y así de la innovación tecnológica.

4.7 Revisión de la evolución de las variables independientes por MCMC:

4.7.1 Distribución de simulaciones de PIB per cápita

Ilustración 16 – 16: Distribución de las simulaciones de PIB per cápita



En la ilustración 4 - 16 se observa como en la mayor parte de las simulaciones hay una tendencia creciente en los datos obtenidos. Los datos tienen una media de 6.488 y se espera que haya un crecimiento sostenido del PIB per cápita en los próximos 10 años. Esta gráfica se explica mejor en la siguiente gráfica.

Tabla 6 – 6: Estadísticas descriptivas del PIB en Colombia por años

	2021	2025	2030
Media	5.437	5.879	6.488,6
Desviación estándar	532	1.374	2.285,2
Max	9.577	20.331	37.063,5
Min	318	1.757	1.216,6
99%	7.204	10.541	14.650,5
1%	4.168	3.444	3.067,7
Cuartil 3	5334,55	5.161	5.178,6

Cuartil 2	5.495	5.459	5.873,3
Cuartil 1	5.495	6.467	7.358,6

Aunque los mínimos y máximos son valores bastante alejados de la tendencia, estos pueden ser considerados ruido, ya que se alejan demasiado del 99% de datos inferior y superior. Esto a su vez apoya la teoría de un crecimiento ligero pero sostenido que plantea la simulación del registro de patentes en Colombia realizada anteriormente.

4.7.2 Distribución de las simulaciones de la inversión en I+D:

La inversión en I+D ha presentado baja volatilidad en el transcurso de los años, con una desviación estándar de tan solo un 0,06. La inversión en I+D es la única variable considerada que puede ser establecida por el poder ejecutivo de Colombia. Más específicamente por el Ministerio de Hacienda y el Departamento Nacional de Planeación. Además, este es un recurso limitado que debe ser repartido con otros rubros en los cuales el gobierno debe invertir. No obstante, es de considerar la tendencia creciente que tiene éste debido a la alta importancia que ha tomado el desarrollo de Actividades de Ciencia e Innovación en el desarrollo económico de los países. Se debe tener en cuenta que según los estudios de Zhang L usando MCMC y modelando con BMA, la inversión en I+D es la variable que más impacto tiene en la innovación tecnológica en China.

Ilustración 17 – 17: Distribución de las simulaciones de I+D a 10 años

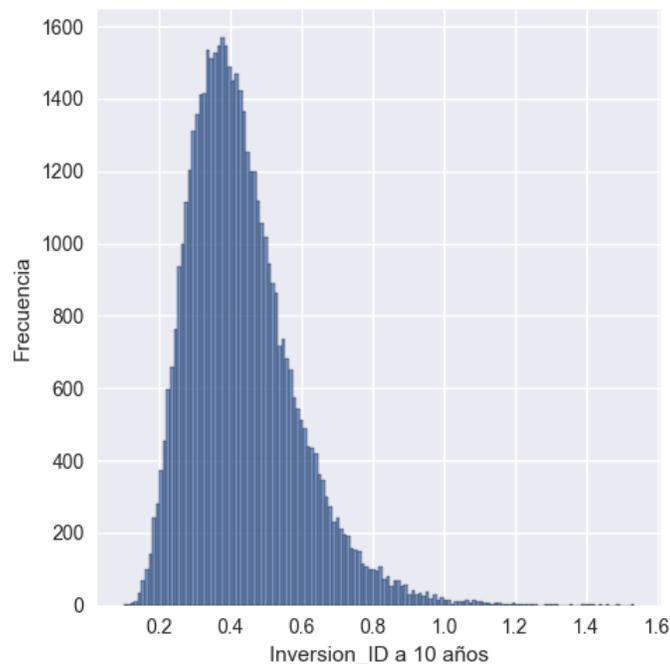


Tabla 7 – 7: Estadísticas descriptivas de la inversión en I+D en Colombia por años

	2021	2025	2030
Media	0,30	0,32	0,36
Desviación Estándar	0,02	0,06	0,10
Máximo	0,47	0,90	1,54
Mínimo	0,19	0,13	0,10
99%	0,37	0,53	0,72
1%	0,24	0,21	0,20
25%	0,29	0,29	0,29
50%	0,29	0,30	0,33
75%	0,30	0,35	0,40

Se observa un crecimiento sostenido de la variable, y aún en el cuartil 3 de las simulaciones, estas tienen un bajo incremento. Principalmente para el año 2021 se observa que los cuartiles 1 y 2 son iguales y con una décima de diferencia en el cuartil 3.

4.7.3 Distribución de las simulaciones de Calidad institucional

La calidad institucional es una variable porcentual, por lo que no puede superar los 100 puntos, según el Krauser (2021) Colombia se encuentra en el lugar 66 en el listado de

países organizado de mayor a menor calidad institucional. Por lo anterior, es de considerar que aún hay un amplio rango de mejora para este indicador. No obstante, los registros previos evidencian únicamente una lenta mejoría en el transcurso de los años.

Ilustración 18 – 18: Distribución de las simulaciones de Calidad Institucional

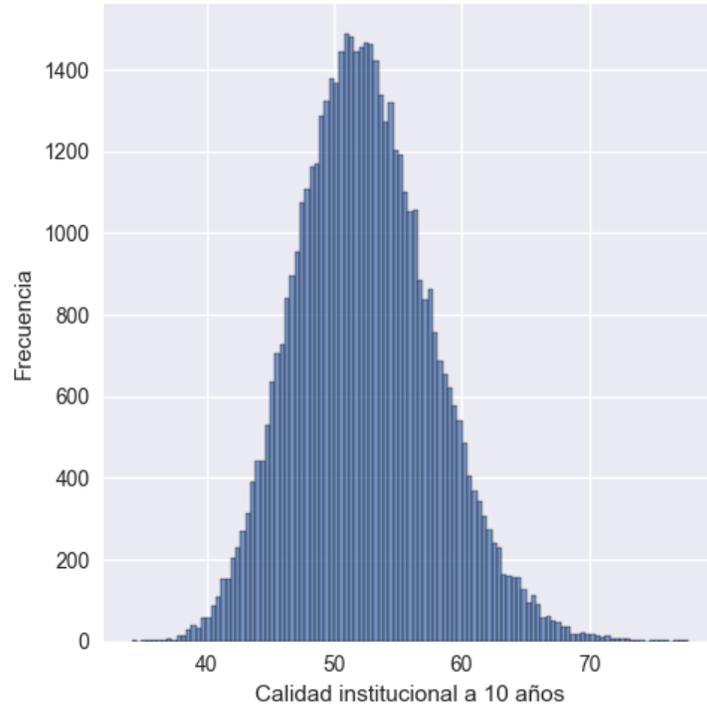


Tabla 8 – 8: Estadísticas descriptivas de la calidad institucional en Colombia por años

	2021	2025	2030
Media	46,12	47,41	49,08
Desviación Estándar	1,10	2,70	4,20
Máximo	53,09	65,27	77,81
Mínimo	40,67	36,10	33,84
99%	49,56	55,48	61,66
1%	43,42	42,12	41,71
25%	45,81	45,81	45,81
50%	45,81	46,78	48,21
75%	46,42	48,89	51,42

La calidad institucional muestra también una tendencia al alza significativa. Eso puede variar dependiendo de los gobiernos, pero es de observar como la calidad institucional de

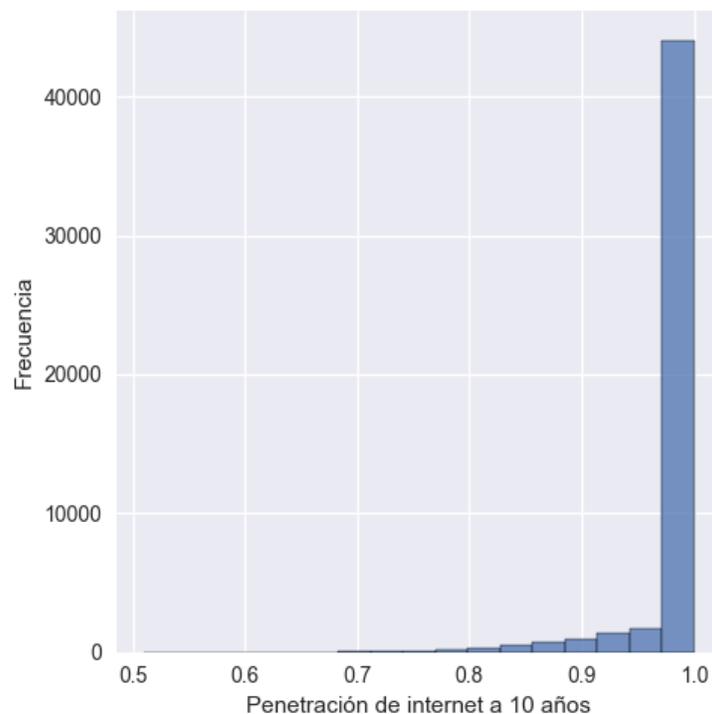
Colombia ha mejorado significativamente. La calidad institucional a su vez depende de múltiples variables explicadas por Barbier & Burgess (2021).

4.7.4 Distribución de las simulaciones de penetración de internet

Las simulaciones de penetración de internet tienen un peso muy bajo en el modelo, sin embargo, al tener las simulaciones una proyección a tan largo plazo se considera que el impacto en el transcurso de 10 años no es descartable.

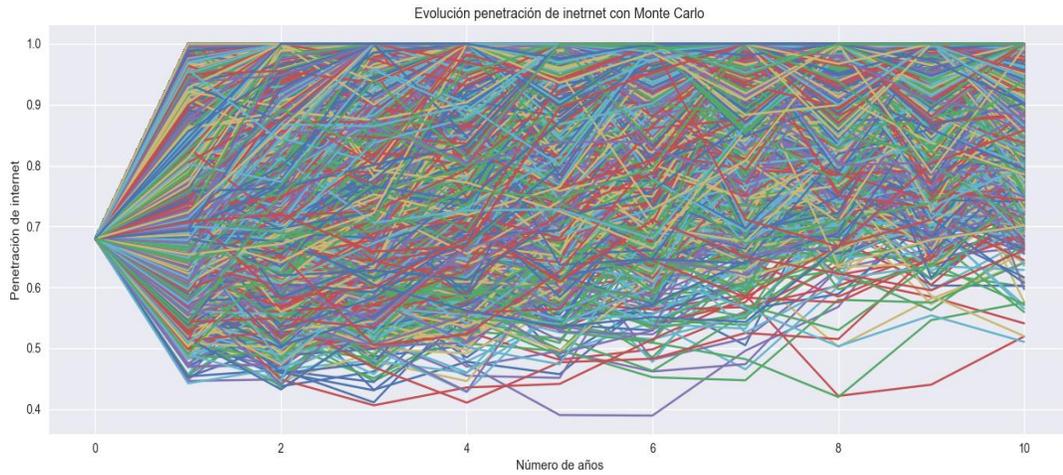
En este parámetro se debe considerar que es una variable porcentual y que su tope está en un 100% y con la tendencia que lleva la tendencia se espera que la penetración de internet en Colombia al año 2030 sea un valor cercano al 100%, así, todas las simulaciones que se aproximen o superen el valor de 100% son redondeadas y disminuidas al 100%.

Ilustración 19 – 19: Distribución de las simulaciones de penetración de internet



Así, en la ilustración 4-19 se observa como la mayor parte de las simulaciones llegan a un valor del 100% debido a la naturaleza de la variable. Esto se puede observar de mejor manera en la gráfica de simulaciones independiente:

Ilustración 20 – 20: Camino de simulaciones de penetración de internet



En la ilustración 4-20 se observa como ninguna de las simulaciones propuestas supera el 1 y que a partir de ese punto siguen su recorrido dependiendo de los valores que adquieran la simulación.

Tabla 9 – 9: Estadísticas descriptivas de la penetración de internet en Colombia por años

	2021	2025	2030
Media	0,742	0,879	0,928
Desviación Estándar	0,098	0,136	0,117
Máximo	1,000	1,000	1,000
Mínimo	0,442	0,390	0,389
99%	1,000	1,000	1,000
1%	0,592	0,605	0,633
25%	0,680	0,736	0,885
50%	0,680	0,947	1,000
75%	0,800	1,000	1,000

Debido a la fuerte tendencia al incremento que tiene el modelo, se observa como el máximo en los 3 años se encuentra en el 100%, sin embargo, la media de las simulaciones

únicamente llega al 92,8% en el transcurso de 10 años. Esto último es más probable ya que hay que considerar que en Colombia hay limitaciones serias en infraestructura para llegar a un acceso de internet del 100% de los habitantes.

4.8 Validar el modelo

Se realiza una validación del modelo usando los datos recolectados hasta el año 2018 y a partir de ese punto se realiza la simulación para corroborar que los valores obtenidos en la simulación se acercan a los obtenidos en la realidad.

Además, se consideran los datos que ya se han publicado en la Super Intendencia de Industria y Comercio de registro de patentes para el año 2021. así, se utiliza nuevamente el modelo utilizando los años del 2000 al 2018 y se aplica para verificar si los resultados obtenidos corresponden a los faltantes en los años de 2019 a 2021.

Los datos reales de patentes de residentes registrados por la SIC son los siguientes:

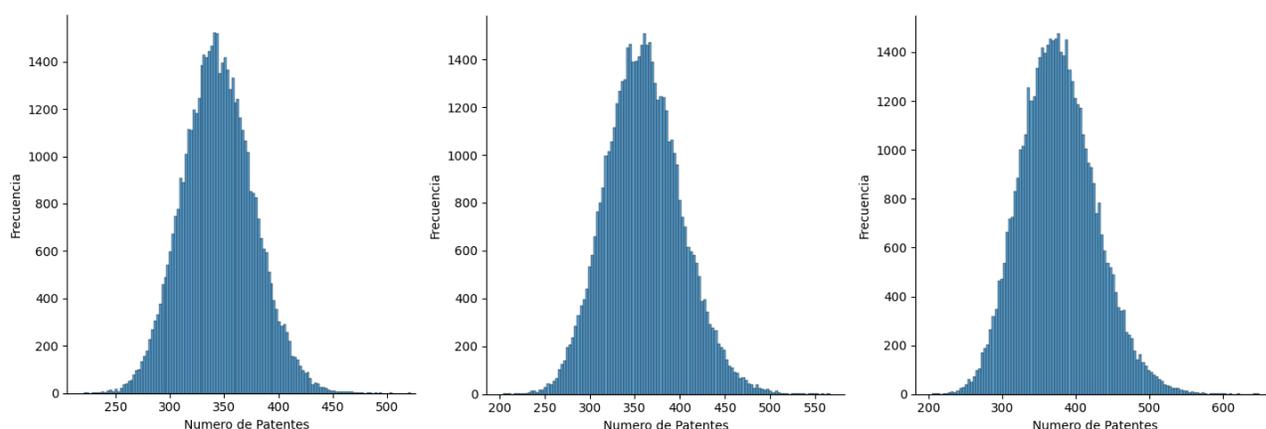
Tabla 10 – 10: Datos reales de registro de patentes por residentes (2019 – 2021)

Año	2019	2020	2021
Patentes	365	362	432

Datos recuperados de Superintendencia de Industria y Comercio (2021)

Aplicando la simulación se obtienen los siguientes datos con las siguientes distribuciones de tendencia:

Ilustración 21 – 21: Distribución de las simulaciones de patentes (2018 – 2021)



En las gráficas se observa como la media se sigue manteniendo en valores cercanos a las 320 patentes registradas al año, incluso después de transcurridos tres años.

Tabla 11 – 11: Estadísticas de patentes simuladas (2019 – 2021)

	2019	2020	2021
Media	330,25	340,4	349,32
Desviación Estándar	27,31	35,87	43,04
Máximo	521,62	566,58	648,92
Mínimo	220,69	203,86	204,86
99%	414,63	445,14	474,78
95%	387,27	409,66	431,09
1%	280,11	275,76	274,51
5%	303,31	300,03	299,43
Cuartil 1	316,0	316	316
Cuartil 2	316,0	326,04	339,33
Cuartil 3	343,88	362,80	375,87
Valores Reales	365	362	432
Diferencia de la media simulada y el registro real	9.58%	6%	19.21%

Se puede observar que todos los números se encuentran dentro de los valores que toma la distribución. No obstante, se debe considerar que las patentes incrementan de un modo mayor que el previsto por las medias que calculan las gráficas y aunque la dispersión de los datos cercanos a la media no es mucho. Los datos reales en los tres años evaluados son superiores o iguales al cuartil 3 de los datos simulados. Por lo que se debe estimar que en todos los años se presentará un incremento mayor al de la media.

En los años 2019 y 2020, el valor obtenido se encuentra por debajo del umbral del 95% de los datos, mientras que, en el año 2021, debido al fuerte incremento que se presentó en el registro de patentes de residentes en Colombia, el registro de patentes real se encuentra en el 5% superior de los datos simulados.

Así, aunque el modelo funciona para predecir el rango en el cual se encontrará el registro de patentes de residentes en Colombia, este valor en los años donde se ha validado el método y el modelo se suele ubicar por encima del cuartil 3 de los datos simulados. No obstante, se debe considerar que, en estos años, las patentes se mantienen estables o incrementan casi un 20% del año 2020 al año 2021, por lo que la fuerte volatilidad de estos

hace que sea improbable tener un modelo que prevea con total precisión su comportamiento.

Por otro lado, las simulaciones realizadas para el año 2021 tomando como año de referencia el año 2020 tienen una diferencia de 13,4% con respecto al valor real. Una diferencia que es menor a la obtenida a partir de las simulaciones que inician en el año 2018. Por lo que se evidencia que las simulaciones con un mayor lapso en el futuro tienden a incrementar de forma significativa su margen de error.

CAPITULO 5: CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones:

- 1.** Al aplicar el modelo de Análisis de Componentes Principales se puede observar como la tendencia del registro de patentes en Colombia y por lo tanto de la innovación tecnológica del país depende mayoritariamente de los registros de patentes en años anteriores y únicamente un 12,5% de las variables independientes consideradas para el estudio, (inversión en I+D, PIB per cápita, calidad institucional y penetración de internet) que pese a la gran correlación que presentan con la variable dependiente, el resultado del modelo parece arrojar que su comportamiento no afecta de forma tan directa el desarrollo de las patentes.
- 2.** La media de los modelos muestra una tendencia a un ligero incremento en el registro de patentes en Colombia, no obstante, luego de validados los datos de las simulaciones con registros reales de patentes por residentes en Colombia, se establece una tendencia a que los datos reales se encuentren en el cuartil superior de los datos simulados. No obstante, se debe tener en cuenta la alta volatilidad que tienen los datos reales de registro de patentes de residentes por la Super Intendencia de Industria y Comercio.
- 3.** El método Monte Carlo utilizando procesos de Weiner y Factorización de Cholesky logra hacer aproximaciones reales a las estadísticas del registro de patentes de Colombia, y puede servir como base para realizar cálculos reales de cómo se comportará la innovación tecnológica en Colombia en un mediano plazo con un margen de error de alrededor del 11% en los primeros años que se simulan y que parece tender a aumentar a medida en que incrementan la cantidad de años simulados.
- 4.** La innovación tecnológica en Colombia tiene una tendencia a incrementar, y se espera que para el año 2030 haya un mayor número de patentes registradas por parte de residentes, además de un incremento en el PIB per cápita, inversión en I+D, Calidad institucional y la penetración de internet que son las variables validadas en el modelo para la predicción de innovación tecnológica en Colombia.

5.2 Recomendaciones:

1. Para futuros estudios se recomienda experimentar con otro tipo de modelos al obtenido por medio de Análisis de Componentes Principales (PCA) ya que por este método el histórico del registro de patentes tiene un peso sobre la composición de las simulaciones mayor al 85%, y esto ocasiona que las variables independientes solo representen un 15% del peso total de la simulación, que aunque influye en los resultados de la misma, podría tener un mayor impacto con modelos como el Análisis por Máxima Verosimilitud.
2. En próximos estudios se recomienda aplicar el modelo en otros países no industrializados a fin de validar la funcionalidad del modelo y del método para predicciones estadísticas de la innovación tecnológica.
3. Ya que el registro de patentes por residentes en Colombia presenta una alta volatilidad, se recomienda para futuros estudios utilizar modelos tipo GARCH o ARCH donde se pueda describir el comportamiento de los cambios de las patentes en función del tiempo considerando una volatilidad altamente variable.

BIBLIOGRAFÍA:

1. Aguirre Ramirez J. (2010). *Modelo para medir y Evaluar capacidades de Innovacion*.
https://www.researchgate.net/publication/236159783_Modelo_para_medir_y_Evaluar_capacidades_de_Innovacion
2. Andersson, M., Kusetogullari, A., & Wernberg, J. (2021). Software development and innovation: Exploring the software shift in innovation in Swedish firms. *Technological Forecasting and Social Change*, 167.
<https://doi.org/10.1016/j.techfore.2021.120695>
3. Aponte Figueroa, G. M. (2016). Technological innovation management through the analysis of patent information. *Revista Científica Electrónica de Ciencias Gerenciales*, 33. <https://www.redalyc.org/pdf/782/78245566003.pdf>
4. Artz, K. W., Norman, P. M., Hatfield, D. E., & Cardinal, L. B. (2010). A longitudinal study of the impact of R&D, patents, and product innovation on firm performance. *Journal of Product Innovation Management*, 27(5), 725–740.
<https://doi.org/10.1111/j.1540-5885.2010.00747.x>
5. Azad, S. M., & Ghodsypour, S. H. (2018). Modeling the dynamics of technological innovation system in the oil and gas sector. *Kybernetes*, 47(4), 771–800. <https://doi.org/10.1108/K-03-2017-0083>
6. Banco Mundial. (2019). *Inversión en I+D*.
<https://datos.bancomundial.org/indicador/GB.XPD.RSDV.GD.ZS?locations=CO>
7. Banco Mundial. (2021). *PIB per cápita*.
<https://datos.bancomundial.org/indicador/NY.GDP.PCAP.CD?locations=CO>
8. Barbier, E. B., & Burgess, J. C. (2021). Institutional quality, governance and progress towards the SDGs. *Sustainability (Switzerland)*, 13(21).
<https://doi.org/10.3390/su132111798>
9. Barreto-Villanueva, A. (2012). El progreso de la estadística y su utilidad en la evaluación del desarrollo Papeles de Población. *Papeles de Población*, 18(73), 1–31.

10. Bolia, N., & Juneja, S. (2005). Monte Carlo methods for pricing financial options. *Sadhana - Academy Proceedings in Engineering Sciences*, 30(2–3), 347–385. <https://doi.org/10.1007/BF02706251>
11. Chang, S. H., Chang, H. Y., & Fan, C. Y. (2018). Structural model of patent quality applied to various countries. *International Journal of Innovation Science*, 10(3), 371–384. <https://doi.org/10.1108/IJIS-05-2017-0036>
12. Chi, M., Muhammad, S., Khan, Z., Ali, S., Yi, R., & Li, M. (2021). Technological Forecasting & Social Change Is centralization killing innovation, The success story of technological innovation in fiscally decentralized countries. *Technological Forecasting & Social Change*, 168(May 2020), 120731. <https://doi.org/10.1016/j.techfore.2021.120731>
13. Cigüenza, Noelia. (2018). La economía colombiana registró el peor crecimiento de los últimos ocho años. In *Diario La República*. La República. <https://www.larepublica.co/economia/la-economia-colombiana-registro-el-peor-crecimiento-de-los-ultimos-ocho-anos-2599653>
14. Departamento Nacional de Planeación, Consejo Nacional de Política Económica y Social, República de Colombia, “Conpes 4023,” (2021). <https://colaboracion.dnp.gov.co/CDT/Conpes/Economicos/4023.pdf>.
15. D’Ingiullo, D., & Evangelista, V. (2020). Institutional quality and innovation performance: evidence from Italy. *Regional Studies*, 54(12), 1724–1736. <https://doi.org/10.1080/00343404.2020.1765230>
16. Davison-Pilon, C. (2015). *Bayesian Methods For Hackers*. Addison-Wesley.
17. Di Sciorio, F., & Mattiozzi, S. (2020). Option pricing under multifractional brownian motion in a risk neutral framework. *Estudios de Economía Aplicada*, 38(3), 1–10. <https://doi.org/10.25115/EEA.V38I3.2902>
18. Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, 6(3), 274–290. <https://doi.org/10.1177/1745691611406920>
19. Ding, S., McDonald, F., & Wei, Y. (2021). Is Internationalization Beneficial to Innovation? Evidence from a Meta-analysis. In *Management International*

- Review* (Vol. 61, Issue 4). Springer Berlin Heidelberg.
<https://doi.org/10.1007/s11575-021-00451-0>
20. Dirección Nacional de Derechos de Autor. (2022). *Reportes DNDA*.
<http://derechodeautor.gov.co:8080/indicadores-de-gestion>
21. Departamento Nacional de Planeación. (n.d.). *Estadísticas: Indicadores de Ciencia, Tecnología e innovación*. Departamento Nacional de Planeación.
<https://www.dnp.gov.co/programas/desarrollo-empresarial/ciencia-tecnologia-e-innovacion/Paginas/estadisticas.aspx>
22. Expansión - Datos Macro. (2021). *Colombia - Exportaciones de Mercancía*.
<https://datosmacro.expansion.com/comercio/exportaciones/colombia>
23. Gortler, J., Spinner, T., Streeb, D., Weiskopf, D., & Deussen, O. (2020). Uncertainty-Aware Principal Component Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 822–831.
<https://doi.org/10.1109/TVCG.2019.2934812>
24. Gómez Hurtado, R. E. (2018). Tendencias de la innovación tecnológica en Colombia 1991-2013 a partir del análisis de patentes. *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información*, 32(77), 133.
<https://doi.org/10.22201/iibi.24488321xe.2018.77.57859>
25. IBM, “Prueba Kolmogorov-Smirnov para una muestra,” (2021).
<https://www.ibm.com/docs/es/spss-statistics/saas?topic=tests-one-sample-kolmogorov-smirnov-test>.
26. Igami, M., & Subrahmanyam, J. (2019). Patent Statistics as an Innovation Indicator? Evidence from the Hard Disk Drive Industry. *Japanese Economic Review*, 70(3), 308–330. <https://doi.org/10.1111/jere.12234>
27. Jiménez-Gómez, M., Prins, A., María, N., López, R., David, M., Miguel, L., Gómez, J., David, M., & López, R. (2017). *Medición de la innovación en Colombia*. 5.
28. Kim, S. I. (2022). ARMA–GARCH model with fractional generalized hyperbolic innovations. *Financial Innovation*, 8(1), 48.
<https://doi.org/10.1186/S40854-022-00349-2>

29. Koh, P. S., & Reeb, D. M. (2015). Missing R and D. *Journal of Accounting and Economics*, 60(1), 73–94. <https://doi.org/10.1016/j.jacceco.2015.03.004>
30. Krause, M. (2021). Índice De Calidad Institucional 2021. *Relial* [ici-reli-al-2021-17mayo-publi-es-final.pdf](https://www.freiheit.org/ici-reli-al-2021-17mayo-publi-es-final.pdf) (freiheit.org).
31. Lebovits, J., Lebovits, J., Bender, C., & Lévy-véhel, J. (2014). *Stochastic calculus with respect to multi-fractional Brownian motion and applications to finance* To cite this version: *Thèse présentée par Stochastic Calculus With Respect to Multi-*.
32. Lux, T. (2021). Bayesian Estimation of Agent-Based Models via Adaptive Particle Markov Chain Monte Carlo. *Computational Economics*, 0. <https://doi.org/10.1007/s10614-021-10155-0>
33. Medina Hurtado, Santiago; Restrepo Morales, J. A. (2013). Estimación De La Utilidad En Riesgo De Una Empresa De Transmisión De Energía Eléctrica Considerando Variables Económicas. *Cuadernos de Economía*, 59, 103–137. <https://www.redalyc.org/pdf/2821/282126853006.pdf>
34. Molano-Rojas, A., Barrera, A., Garnica, L., Moncada, J. G., Moya, J., Penagos, A., Puentes, G., & Zarama, G. (2018). *Calidad Institucional Fundamento del desarrollo Calidad Institucional*.
35. Observatorio Colombiano de Ciencia y Tecnología. (2019). *Indicadores de ciencia y tecnología 2019*. <https://ocyt.org.co/Informeindicadores2019/indicadores-2019.pdf>
36. Observatorio Colombiano de Ciencia y Tecnología (2020). *Indicadores de Ciencia, Tecnología e Innovación*. Informe Indicadores CTeI 2020 v1.pdf (ocyt.org.co)
37. ECD (2015), Frascati Manual Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris. DOI: <http://dx.doi.org/10.1787/9789264239012-en>
38. Oh, G., & Ayoung, H. K. (2017). Analysis of technological innovation based on citation information. *Quality & Quantity*, 51(3), 1065–1091. <https://doi.org/10.1007/s11135-016-0460-9>

39. Pérez Padrón, S., Moreno Méndez, F., & Padrón Zardúa, T. (2018). La innovación tecnológica y la investigación de mercado en el sistema empresarial cubano. *Universidad y Sociedad*, 10(3), 134–141.
40. Popp, D. (2005). Lessons from patents: Using patents to measure technological change in environmental models. *Ecological Economics*, 54(2–3), 209–226.
<https://doi.org/10.1016/j.ecolecon.2005.01.001>
41. Salazar, E., & Alzate, W. (2018). Application of Monte Carlo simulation in the projection of the income statement. A case study. *Journal of Spaces*, 39(51).
<https://www.researchgate.net/publication/335001474>
42. Shirota, S., Omori, Y., Hedibert, F. L., & Piao, H. (2017). Cholesky realized stochastic volatility model. *Econometrics and Statistics*, 3, 34–59.
<https://doi.org/10.1016/j.ecosta.2016.08.003>
43. Sinha, A. K. (2021). The reliability of geometric Brownian motion forecasts of S&P500 index values. *Journal of Forecasting*, 40(8), 1444–1462.
<https://doi.org/10.1002/for.2775>
44. Solano, E., Alandete, N., & Vilorio, A. (2019). Análisis de componentes principales en la Competitividad en Colombia. *RISTI - Revista Ibérica de Sistemas e Tecnologías de Informação*, E24(11/2019), 260–271.
45. Sosa, J. (2021). *Métodos de Monte Carlo*. Rpubs by RStudio.
<https://rpubs.com/jcsosam/830745>
46. Superintendencia de Industria y Comercio. (n.d.). *¿Qué se debe hacer después de obtener el registro de las patentes de modelos de utilidad?*
<https://www.sic.gov.co/patente-de-modelos-de-utilidad>
47. Superintendencia de Industria y Comercio. (2021). *Estadísticas PI*.
<https://www.sic.gov.co/estadisticas-propiedad-industrial>
48. Tejada-Estrada, Gina-Coral, Cruz-Montero, Juana-María, Hernández-Uribe, & Cecilia-Yrene. (2019). *Innovación tecnológica: Reflexiones teóricas*. 24.
49. U.S. PATENT AND TRADEMARK OFFICE. (2019). *U.S. Patent Statistics Chart Calendar Years 1963 - 2019*.
https://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm

50. Vélez Ibarrola, R. (2021). Introducción al Movimiento Browniano. *Departamento de Estadística e I.O.*
http://portal.uned.es/pls/portal/docs/PAGE/UNED_MAIN/LAUNIVERSIDAD/UBICACIONES/01/DOCENTE/RICARDO_VELEZ_IBARROLA/CURSOS_COMPLEMENTARIOS/BM.PDF
51. Víctor Díaz. (2016). *Innovación tecnológica y comercio internacional: un análisis para los países de la Comunidad Andina de Naciones: Bolivia, Colombia y Perú.* [Universidad del Valle].
<http://bibliotecadigital.univalle.edu.co/xmlui/bitstream/handle/10893/10455/7380-0534435.pdf?sequence=1>
52. Wang, G., & Wang, Y. (2019). Evaluation research of science and technology innovation capability in ulanqab - On the basis of time sequence model. *Proceedings - 2019 International Conference on Economic Management and Model Engineering, ICEMME 2019*, 321–326.
<https://doi.org/10.1109/ICEMME49371.2019.00070>
53. Wang, Y., Li, C., & Khan, M. A. (2021). *Firm information disclosure environment and R & D investment: Evidence from Internet penetration.* 1–14.
<https://doi.org/https://doi.org/10.1371/journal.pone.0247549>
54. We Are Social. (2021). *Special Report Digital 2021.* We Are Social.
<https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/>
55. The World Bank. (2021). *Worldwide Governance Indicators.* The World Bank.
<http://info.worldbank.org/governance/wgi/Home/Reports>
56. World Intellectual Property Organization WIPO. (2021). *Global innovation Index 2021 (Colombia).* 1–9.
https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2021/co.pdf
57. Zhang, L., Zhou, D., Zhong, M., & Wang, Y. (2019). Improved closed-loop subspace identification based on principal component analysis and prior information. *Journal of Process Control*, 80, 235–246.
<https://doi.org/10.1016/j.jprocont.2019.06.001>

58. Zhang, M. (2018). *The Influencing Factors of Industrial Technological Innovation Ability--Using Bayesian Model Averaging*. *Emcs*, 216–222. DOI: 10.25236/emcs.2018.055