



*Implementación de un Filtro de Kalman para la  
estimación del Dinero Lavado a través de  
Remesas Familiares en Colombia*

Por

Jackelyne Gómez Restrepo

Asesor

Carlos Mario Vélez S.

Universidad EAFIT

Departamento de Ciencias Básicas

Medellín

3 de junio de 2013

Nota de aceptación

---

---

---

---

Presidente del jurado

---

Jurado

---

Jurado

---

Ciudad y fecha (día, mes y año)

## Resumen

La economía subterránea abarca la economía ilegal (producción y distribución de bienes y servicios prohibidos por la ley), economía no declarada (acciones que evaden las normas tributarias), economía no registrada y la economía informal; ésta corresponde al 3% del PIB, dato que resulta realmente perturbador y alarmante para la economía colombiana, ya que coincide con el porcentaje de correspondencia entre las remesas familiares y el PIB.

De acuerdo con este entorno económico, en este trabajo se pretende estimar la cantidad del dinero ilícito transado en las remesas familiares en Colombia por medio de la implementación en Matlab de un filtro de Kalman, junto con un algoritmo genético encargado de generar las condiciones iniciales de los parámetros que inicializarán el filtro. Dicho filtro estimará al mismo tiempo la serie de dinero lavado por remesas familiares y los parámetros del modelo. Así, se genera una contribución al mundo académico de la Economía y la Computación, y a entidades financieras y del sector real e instituciones del gobierno que busquen entender la dinámica del dinero ilícito que ingresa al país por las remesas familiares, así como a medir la cuantía e identificar aquellas variables que influyen en dicha dinámica.

Los resultados obtenidos en el trabajo son validados de manera cuantitativa realizando predicciones de otras variables involucradas en el modelo, y cualitativamente a través de acontecimientos históricos del país relacionados con la migración, la ola de inseguridad, tasas de desempleo, variables relacionadas con el narcotráfico, entre otras que ayudan a sustentar la magnitud y signo de los coeficientes estimados. A través del volumen calculado es posible entonces validar la tesis de expertos que sugieren que en la actualidad las remesas familiares son poco usadas para el blanqueo de dinero.

**Palabras Clave:** *Remesas familiares, Dinero ilícito, Filtro de Kalman, Algoritmo Genético, Estimación.*

## **Agradecimientos**

Deseo expresar un agradecimiento especial a un gran maestro: el Doctor Carlos Mario Vélez Sánchez por su valiosa colaboración, asesoría y sugerencias en la elaboración de este proyecto. Su vasta experiencia en investigación fue de gran importancia ya que con sus indicaciones se realizó la elección adecuada del proyecto, metodologías, validaciones y forma del trabajo final.

# Tabla de Contenido

<b>Lista de Figuras</b>	<b>5</b>
<b>Lista de Tablas</b>	<b>7</b>
<b>1. Introducción</b>	<b>8</b>
1.1. Objetivo general . . . . .	8
1.2. Objetivos específicos . . . . .	8
1.3. Justificación . . . . .	9
1.4. Metodología . . . . .	11
1.5. Organización del trabajo . . . . .	12
<b>2. Marco de Referencia</b>	<b>13</b>
2.1. Glosario . . . . .	13
2.2. Antecedentes en el campo de la Economía . . . . .	14
2.3. Economía Subterránea . . . . .	18
2.3.1. Narcotráfico en Colombia . . . . .	19
2.4. Remesas familiares . . . . .	22
2.5. Métodos . . . . .	23
2.5.1. Filtros de Kalman y Representación Estado-Espacio . . . . .	23
2.5.2. Algoritmos Genéticos (AG) - <i>Genetic Algorithms (GA)</i> . . . . .	37
2.5.3. Dinámica del dinero ilícito . . . . .	43
2.5.4. Criterios de selección de modelos . . . . .	45
2.6. Resumen . . . . .	46
<b>3. Resultados y discusión</b>	<b>49</b>
3.1. Definición del Problema . . . . .	50

3.2. Modelos propuestos . . . . .	51
3.2.1. Modelo 1 . . . . .	51
3.2.2. Modelo 2 . . . . .	54
3.2.3. Modelo 3 . . . . .	63
3.2.4. Modelo 4 . . . . .	70
3.3. Análisis de Resultados . . . . .	79
3.3.1. Modelos no exitosos . . . . .	79
3.3.2. Modelo elegido . . . . .	80
3.4. Resumen . . . . .	86
<b>4. Conclusiones y Trabajo Futuro</b>	<b>89</b>
<b>Referencias</b>	<b>94</b>

# Lista de Figuras

2.1. Producción de hoja de coca en Colombia, Perú y Bolivia, 1990-2000. Tomada de [30] . . . . .	20
2.2. Noticia Pitufeo en Colombia. Tomada de [32] . . . . .	21
2.3. Remesas familiares por país de origen. Tomada de [3] . . . . .	22
2.4. % de remesas familiares con respecto al PIB. Tomada de [3] . . . . .	23
2.5. Filtro Kalman. Tomado de [13] . . . . .	27
2.6. Estimación Dual. Tomada de [50] . . . . .	35
2.7. Filtro de Kalman Dual - Estimación Simultánea. Tomada de: [56] . .	36
2.8. Algoritmo genético. Tomada de [59] . . . . .	40
2.9. Identificación de Sistemas con Algoritmos genéticos. Tomada de: [62]	42
3.1. Modelo 2 - Parámetro $\theta_8$ . . . . .	60
3.2. Modelo 2 - Dinero de remesas familiares en logaritmo natural. Elabo- ración propia . . . . .	60
3.3. Modelo 2 - Dinero ilícito de remesas familiares en logaritmo natural. Elaboración propia . . . . .	61
3.4. Modelo 2 - QQ-Plot . . . . .	62
3.5. Modelo 3 - Dinero lavado por remesas. Elaboración propia . . . . .	68
3.6. Modelo 3 - Dinero de remesas familiares. Elaboración propia . . . . .	68
3.7. Modelo 3 - Parámetro $\theta_7$ . . . . .	69
3.8. Dinero de remesas familiares en logaritmo natural. Elaboración propia	75
3.9. Dinero ilícito de remesas familiares en logaritmo natural. Elaboración propia . . . . .	75
3.10. Modelo 4 - Parámetro $\theta_7$ . . . . .	77
3.11. Modelo 4 - QQ-Plot . . . . .	79



3.12. Estimación Dinero Legal. Elaboración propia . . . . .	85
3.13. Estimación Dinero enviado a través de remesas familiares. Elaboración propia . . . . .	85

# Lista de Tablas

3.1. Parámetros del algoritmo genético para el Modelo 2 . . . . .	58
3.2. Condiciones Iniciales - Modelo 2 . . . . .	59
3.3. Criterios de evaluación- Modelo 2 . . . . .	62
3.4. Condiciones Iniciales - Modelo 3 . . . . .	67
3.5. Criterios de evaluación- Modelo 3 . . . . .	69
3.6. Parámetros del algoritmo genético . . . . .	73
3.7. Condiciones Iniciales - Modelo 4 . . . . .	74
3.8. Parámetros Estimados . . . . .	76
3.9. Criterios de evaluación- Modelo 4 . . . . .	78

# Capítulo 1

## Introducción

La introducción contiene la descripción de todos los aspectos que ubican el proyecto de investigación en un contexto, el detalle de la pregunta de investigación junto con los objetivos que han sido planteados para el desarrollo del proyecto y la justificación del mismo.

### 1.1. Objetivo general

Implementar un filtro de Kalman para determinar el volumen de dinero lavado a través de las remesas familiares enviadas a Colombia.

### 1.2. Objetivos específicos

- Describir el impacto que tienen las remesas familiares sobre la economía nacional.
- Identificar trabajos que se hayan realizado sobre el tema, con el fin de tener una amplia referenciación del uso del filtro de Kalman en problemas de la economía.
- Formular un modelo que describa el volumen de dinero ilícito transado a través de las remesas familiares en Colombia, basado en el conocimiento adquirido a través de las investigaciones previas.
- Implementar un filtro de Kalman que estime los parámetros y estados del modelo propuesto.

- Combinar las bondades del filtro de Kalman y de los algoritmos heurísticos para mejorar el ajuste del modelo.
- Evaluar los resultados obtenidos a través de técnicas de validación y verificación.

### 1.3. Justificación

Por su gran impacto en la sociedad los delitos relacionados con la economía subterránea han llamado la atención del Gobierno, autoridades, empresas y académicos quienes se han puesto en la tarea de desarrollar toda una normatividad, teoría e incluso modelos para entender, describir y combatir estos delitos; sin embargo, a pesar de los diferentes estudios que se han realizado, ha sido difícil medir el tamaño de la economía subterránea a través de métodos directos (encuestas, supervisiones, medios de cumplimiento) [1]. De esta forma muchas de las instituciones que luchan en contra de esta economía ilícita, han trabajado con el fin de identificar las tipologías más utilizadas para el lavado de activos entre las cuales han reconocido tipologías que involucran las remesas familiares. En [2] se denomina como estructuración o pitufeo a la acción de realizar varios depósitos de montos por debajo de los límites de reporte a través de diferentes medios, como pueden ser las remesas familiares que ingresan al país. Dadas las limitaciones de los datos observables y de las herramientas de modelado, y dada la importancia de este ingreso en Colombia, surgen preguntas como, ¿de qué manera se puede aproximar el volumen de dinero ilícito que ingresa al país por este medio?, o ¿cuál es la relación que existe entre ciertas variables y el dinero lavado?. La preocupación es conocer si las remesas familiares, más que un ingreso para las familias colombianas, es un medio de lavado de dinero. Además, por supuesto, se desea demostrar la eficiencia de utilizar este tipo de herramientas de modelado para la estimación óptima de estados y parámetros no observados.

En Colombia, el crecimiento en la recepción de remesas se debe a la última ola migratoria ocurrida desde 1996. A partir de 1999 los ingresos por este medio registran altas tasas de crecimiento, adquiriendo una importancia significativa, ya que pasaron de representar cerca del 1.9 % del PIB en el 2000 a un máximo de 3.9 % en el 2003 y 2.9 % del PIB para el 2006 [3]. Las remesas tienen carácter de ingreso “cuasipermanente”, razón por la cual han sido incorporadas en los hábitos de consumo

y gasto de los hogares. De esta manera, el desarrollo de este trabajo podrá ser una base no solo informativa, sino también para la toma de decisiones en instituciones privadas y públicas que se encuentren en la necesidad de estudiar, conocer, detectar y/o prevenir el lavado de activos a través de las remesas familiares, lo que genera que este proyecto cobre un valor lo suficientemente alto frente a temas de riesgos de reputación para empresas privadas (principalmente aquellas que reciben y pagan remesas) y de desarrollo de la economía para la Nación.

A nivel académico, resulta ser una propuesta que intenta resolver un problema económico a través de metodologías que han tenido mayor desarrollo e impacto en los problemas ingenieriles, pero que pueden resultar bastante útiles siempre que se garantice su buen uso. El principal objetivo va orientado a obtener un modelo híbrido que tome las bondades de dos metodologías robustas para la optimización y estimación de sistemas, con el fin de hallar los parámetros óptimos que permitan estimar la serie (dinero lavado por remesas familiares).

El tema sobre el cual se desarrollará el proyecto (dinero que ingresa a Colombia a través de remesas familiares) corresponde al campo de la economía, más propiamente la macroeconomía colombiana; sin embargo, este desarrollo se vale de metodologías como los algoritmos heurísticos y el filtro de Kalman, desarrollados en Matlab, los cuales corresponden a áreas ingenieriles como la Computación Óptima.

Básicamente, se plantea un proyecto de investigación aplicada de cobertura nacional y que puede ser útil tanto para entidades financieras privadas como para instituciones del gobierno que busquen entender la dinámica del dinero ilícito que ingresa al país por las remesas familiares, así como medir su cuantía e identificar aquellas variables que influyen en su crecimiento. Se pretende estimar la cantidad del dinero ilícito transado en las remesas familiares con base en un conjunto de variables que se observan y se miden a diario como indicadores de la economía colombiana por su relación con el mercado laboral, el entorno fiscal y el entorno monetario; de esta forma, es posible proponer un modelo capaz de describir en el tiempo, el comportamiento del dinero producto del lavado.

Este es un tema estrechamente relacionado con la economía subterránea, lo que

conlleva a que existan pocas mediciones exactas y muchas menos propuestas de modelado. Debido a que no se consiguen fácilmente series completas y confiables sobre el dinero lavado en el tiempo, no es posible estimar un modelo de series de tiempo con las metodologías tradicionales, lo que genera la necesidad de buscar otras alternativas para realizar la estimación a través de una medición indirecta de esta variable; es allí donde el filtro de Kalman cobra relevancia y se vuelve un factor importante para el completo desarrollo de esta propuesta.

De esta manera, se propone realizar una estimación indirecta del dinero lavado a través de remesas familiares. La estrecha interacción entre el dinero lavado y la demanda de efectivo por remesas se basa en el supuesto de que la economía subterránea, que usa como fuente dichos ingresos, efectúa la totalidad de las transacciones en efectivo, de manera que se puede suponer que un agente de la economía subterránea realiza transacciones en efectivo, las cuales aumentan el dinero ilegal transado a través de las remesas familiares (variable de estado para el modelo de estado-espacio), hecho que de manera directa, además de otras variables como la migración del país y las tasas de desempleo de los países a los que migran los colombianos, contribuye al incremento del volumen de remesas que ingresan a Colombia.

## 1.4. Metodología

La información nacional utilizada procede de bases libres que almacenan las principales variables de la macroeconomía colombiana como lo es la página del DANE y el Banco de la República. La información de la cantidad de servidores públicos fue solicitada al DAFP (Departamento Administrativo de Funcionarios Públicos), mientras que la información de otros países se consultó en la página del World Bank Group. Esta información fue transformada aplicando logaritmo natural a las variables para tenerlas en una misma escala y así reducir la variabilidad en los resultados.

A partir del conocimiento y tratamiento de los datos se construye un algoritmo genético en Matlab que se encarga de hallar de manera óptima las condiciones iniciales que serán empleadas en el filtro de Kalman, también implementado en Matlab, para la estimación de la serie de dinero lavado por remesas familiares y de los paráme-

tos del modelo. El modelo final corresponde a un híbrido entre dos metodologías que garantizan la obtención de valores óptimos (la teoría al respecto se presentará en el capítulo 2).

## **1.5. Organización del trabajo**

La estructura del documento se encuentra distribuida en cuatro capítulos: en el capítulo 2 se encuentra el marco de referencia, el cual inicia con la presentación de antecedentes, definiciones y contextualización de la economía ilegal y su conexión con las remesas familiares; posteriormente se realiza un desarrollo de la teoría del filtro de Kalman, los algoritmos genéticos y un modelo teórico de la dinámica del dinero ilícito (principales herramientas matemáticas utilizadas para el desarrollo del proyecto).

El capítulo 3 tiene los resultados teóricos y prácticos obtenidos, comenzando con la propuesta del modelo de estimación de dinero lavado a través de remesas familiares, la descripción del problema y el planteamiento para resolverlo con las metodologías antes mencionadas. Los resultados, gráficas, validación y discusión podrán ser encontrados allí mismo. Finalmente se presenta el capítulo de conclusiones y trabajos futuros.

# Capítulo 2

## Marco de Referencia

En esta sección se encontrarán algunos antecedentes sobre aplicaciones del filtro de Kalman en la economía, junto con otras propuestas de estimación del dinero producto del blanqueo. Se definirá para este contexto, la economía subterránea, el dinero producto del lavado de activos y el papel de las remesas familiares tanto en la economía colombiana como en el mundo del narcotráfico.

Luego de la contextualización se desarrollará la teoría para la estimación de estados y parámetros con el filtro de Kalman, así como algunas variaciones que se han hecho del mismo para resolver diferentes tipos de problemas. También se presenta una revisión de los algoritmos genéticos, historia, algoritmo y terminología, orientados a la búsqueda de los valores que serán las condiciones iniciales para el filtro de Kalman. Para terminar se presenta el desarrollo hecho por Loayza en [4] para proponer un modelo que represente la dinámica del dinero ilícito y que posteriormente se convertirá en la ecuación de transición.

### 2.1. Glosario

**ARIMA:** Modelos Autorregresivos Integrados de Medias Móviles, definido así debido a que la variable endógena en el periodo  $t$  es explicada por las observaciones de periodos anteriores más una perturbación .

**BID:** Banco Interamericano de Desarrollo .



**Blanqueo de dinero:** Sinónimo de lavado de activos, lavado de dinero, lavado de capitales.

**Caminata Aleatoria:** Trayectoria que resulta de hacer sucesivos pasos aleatorios.

**Entorno Fiscal:** Relacionado con los impuestos, gasto público y presupuesto del estado.

**Entorno Monetario:** Referente al mercado de dinero con el objetivo de mantener y controlar la estabilidad en la economía.

**Estacionalidad:** Variación periódica y predecible de una serie de tiempo.

**GAFI:** Grupo de Acción Financiera Internacional.

**GARCH:** Generalized AutoRegressive Conditional Heteroscedasticity. Son modelos que permiten modelar varianzas variables en las diferentes series de tiempo.

**Heurístico:** Procedimiento utilizado para la solución de problemas; el enfoque dado en este proyecto se orienta a problemas de optimización.

**Intermediario Bancario:** Instituciones encargadas de mediar entre prestamistas y prestatarios.

**Máxima Verosimilitud:** Método utilizado usualmente para ajustar e identificar modelos estadísticos.

**Macroeconomía:** Parte de la economía que estudia en general la economía a través de variables agregadas.

**PIB:** Producto Interno Bruto.

**Proceso Estocástico:** Sucesión de variables aleatorias que evolucionan en función de una variable.

**Tendencia:** Corresponde a un patrón de comportamientos que indica la dirección o rumbo de la serie.

**UIAF:** Unidad de Información y Análisis Financiero.

## 2.2. Antecedentes en el campo de la Economía

Como se afirma en [6], el filtro de Kalman puede no ser muy interesante debido a que su interpretación no es tan natural. A su vez, autores como Shuster en

[7] recomiendan no “enamorarse” del filtro de Kalman, el cual puede ser admirable por su funcionamiento computacional e interesante teoría, pero quizás no es el más adecuado en el análisis teórico ya que no todos los problemas necesitan de una representación compleja como lo puede llegar a ser un filtro de Kalman; sin embargo, estas percepciones pueden deberse a una mala interpretación en la modelación con un filtro de Kalman, ya que una formulación estado-espacio (necesaria en la implementación de un filtro de Kalman) puede generar una estructura preferible a la que se propone con un modelo ARIMA [6].

Los modelos ARIMA son muy comunes para describir procesos o series económicas, ya que pueden ser escritos por medio de una representación estado-espacio y así utilizar el filtro de Kalman para hallar sus parámetros. En [8] el autor sugiere usar el filtro de Kalman para estimar los parámetros variantes en el tiempo de una regresión lineal y obtener la máxima verosimilitud del modelo estado-espacio, lo que sería equivalente a encontrar los parámetros de cualquier modelo ARIMA.

Las series de tiempo (tradicionalmente modeladas a través de modelos ARMA) consisten en observaciones que tienen una tendencia y estacionalidad dentro del periodo de tiempo analizado. Cuando se analiza una serie existe gran interés por examinar la tendencia y así identificar los principales movimientos de la variable. Un modelo estructural de series de tiempo es un modelo que puede ser interpretado de forma directa; sin embargo, debe garantizarse que al ser estimado sus componentes sean estocásticas. Dentro de las metodologías utilizadas para la estimación de modelos que reflejen el comportamiento de series de tiempo, el filtro de Kalman ha empezado a cobrar gran aceptación dentro del mundo econométrico; la representación estado-espacio abre la posibilidad de proponer modelos que además de representar el comportamiento de las series pueden ser interpretados directamente. De esta forma, luego de ser estimado, este puede pasar por los típicos tests de residuales o los test para validar si las componentes son consistentes de acuerdo con el conocimiento del experto [9].

El análisis de series de tiempo está basado en la teoría de procesos estocásticos, los cuales, en su mayoría, pueden aproximarse a un modelo autoregresivo de medias móviles (ARMA), o a un ARIMA (diferenciando un ARMA). A través del tiempo

se han incluido importantes criterios para obtener un buen modelo: parsimonioso (obtener el modelo más simple posible), coherencia de los datos (consistencia con la información), consistencia con el conocimiento a priori, admisibilidad de los datos, globalidad y estabilidad estructural (buen modelo con datos de entrenamiento, validación y prueba).

Desarrollado inicialmente para la navegación espacial, el filtro de Kalman se ha convertido en una herramienta de gran utilidad para muchas aplicaciones, algunas de ellas encontradas en [10], son: graficación por computador en tiempo real, expansión de escenarios, simulación de instrumentos virtuales, extracción del movimiento de los labios a partir de un video, ajuste de una superficie a un conjunto de puntos, modelos demográficos, monitoreo ambiental, predicción de indicadores económicos, navegación, entrenamiento de redes neuronales, identificación de parámetros en sistemas, procesamiento digital de señales, entre otros. El filtro de Kalman es utilizado principalmente para estimar estados de sistemas que solo pueden ser observados indirectamente [11], ventaja que ha sido aprovechada para la solución de problemas econométricos a nivel nacional e internacional, tal y como se mencionará posteriormente.

A nivel nacional, en [12] proponen estimar el volumen de la economía subterránea con un filtro de Kalman. A partir de una función de demanda de dinero en efectivo se estima la dinámica y tamaño de la economía subterránea colombiana en el periodo 1976-2003. De acuerdo con la descripción del artículo, los autores buscan estimar la demanda por efectivo la cual involucra dinero ilícito y lícito, relacionando variables como los impuestos, las restricciones laborales y financieras, la importación de cajeros automáticos, el costo de los servicios financieros, entre otros regresores que conllevan sólo a capturar actividades intensivas en efectivo las cuales pueden ser la mayoría pero no el total de las transacciones en la economía subterránea.

Misas y Vásquez en [13] se concentran en estimar los parámetros involucrados en la dinámica de la tasa de interés real y la inflación esperada. Con este propósito, plantean una representación estado-espacio que supone perturbaciones estocásticas independientes, con la cual estiman los parámetros estructurales y la variable no observada de tasa de inflación esperada; con su estimación, validan los resultados

obtenidos por otros autores que en su momento emplearon otras metodologías.

El autor de [14] emplea un filtro de Kalman para estimar un sistema de ecuaciones para la tasa natural de desempleo, el desajuste del mercado laboral, los salarios y los precios; con los resultados obtenidos logra identificar varios de los determinantes de la tasa natural de desempleo, así como aquellas variables que poco influyen. Dentro de sus conclusiones menciona un gran beneficio al usar la metodología: El filtro de Kalman permite estimar simultáneamente parámetros, hiper-parámetros y la variable no observada del sistema.

A nivel internacional, [1] es un artículo que hace referencia a la estimación de la economía subterránea en Perú. Este autor se concentra en obtener la serie de tiempo que almacene el volumen del dinero ilícito peruano entre 1980 y 2005, para posteriormente compararla con la información existente en el medio y así proponer políticas económicas para combatir el narcotráfico.

En el artículo [15], Giles utiliza técnicas econométricas, como el modelo MIMIC (Multiple Indicators, Multiple Causes), para medir el tamaño de la economía subterránea, debido a que ésta hace parte de la construcción de modelos útiles para la generación de políticas. Al igual que en [15], el mismo autor en [16] propone un MIMIC para desarrollar y estimar un modelo para la economía subterránea en Nueva Zelanda, con el fin de identificar el impacto de los movimientos de los impuestos sobre las actividades ilícitas. Para ambos estudios el autor aprovechó las bases teóricas del modelo, donde la dinámica de indicadores (efectivo, empleo informal, tasa de participación, consumo de electricidad) puede ser modelada en función de una variable no observable, la cual a su vez depende de otras variables. Sin embargo, Watson y Engle en [17] logran demostrar que el modelo MIMIC corresponde a un caso particular de una familia de modelos con un componente no observado que pueden ser escritos a través de su representación estado-espacio y por ende estimados con un filtro de Kalman.

El artículo [8] presenta tres aplicaciones bastante interesantes. La primera enfocada en un modelo de demanda de reservas internacionales que básicamente se concentra en estimar los parámetros variantes en el tiempo de una regresión lineal; el modelo

supone que los parámetros son caminatas aleatorias que hacen que estos parámetros sean inestables lo que dificulta el uso de técnicas econométricas convencionales y da lugar a la implementación de un filtro de Kalman. Una segunda aplicación es de un modelo de primas de riesgo del tipo de cambio, lo que implica la implementación de un filtro de Kalman con términos de error correlacionados, lo que hace referencia a los casos en los que el impacto o *shock* causa movimientos tanto en las variables observadas como en las no observadas. Finalmente, el autor propone el uso del filtro de Kalman para estimar los parámetros de modelos financieros como los modelos de volatilidad estocástica, los cuales generalmente se estiman con modelos de la familia GARCH.

Finalmente, uno de los autores más mencionado en los artículos anteriores es Loayza, quien en [4] realiza una investigación donde consigue evidencias cuantitativas que relacionan la economía informal con la existencia de regulaciones y altas tasas impuestas por los gobiernos latinoamericanos. El autor estudia la economía subterránea a partir de un modelo de crecimiento y sus resultados sugieren que el tamaño del sector informal depende de manera directa de las variables asociadas a los impuestos y regulaciones, y depende de manera inversa de la calidad y eficacia de las instituciones del gobierno.

### **2.3. Economía Subterránea**

Los autores de [18] proponen describir la economía subterránea como aquella en la que se abarca la economía ilegal (producción y distribución de bienes y servicios prohibidos por la ley), economía no declarada (acciones que evaden las normas tributarias), economía no registrada y la economía informal. De acuerdo con esta definición la Economía Subterránea en Colombia comprende delitos fuente de lavado de activos como el enriquecimiento ilícito, narcotráfico, extorsión, rebelión y secuestro [19], tráfico de migrantes y trata de personas [20], tráfico de niñas, niños y adolescentes [21], la financiación del terrorismo y la administración de recursos relacionados con actividades terroristas [22], delitos contra la administración pública, tráfico de armas, concierto para delinquir y delitos contra el sistema financiero [23]. Con la definición de los delitos se pueden delimitar las acciones relacionadas con

el blanqueo de dinero, el cual es definido por [24] como el proceso de hacer que el dinero sucio parezca limpio, haciendo que las organizaciones criminales o delincuentes obtengan beneficios de las ganancias de sus acciones sin poner en peligro su uso en actividades posteriores ilícitas o lícitas. Según estudios del GAFI presentados en [25], para materializarse el lavado de dinero el proceso debe pasar por tres etapas básicas:

- Colocación: Recepción física de bienes o de dinero. Suele ser el paso más difícil ya que el dinero ilícito debe ser incorporado rápidamente en el sistema financiero.
- Ocultamiento: Busca realizar repetidas operaciones con el dinero ilícito en el sistema financiero con el fin de eliminar su trazabilidad; de esta manera se crea una serie de estratos que dificultan la tarea de determinar la manera en que los fondos ingresaron a la economía legal.
- Integración: Corresponde a la etapa final en la que el dinero ya se considera legal y puede ser utilizado para adquirir bienes e inmuebles.

De acuerdo con el último informe de la UIAF en [24], el dinero lavado en Colombia por narcotráfico corresponde al 3 % del PIB.

### **2.3.1. Narcotráfico en Colombia**

Dado el alcance de esta investigación y su directa relación con las remesas familiares, solo se tomarán en cuenta los acontecimientos, delitos fuente y variables relacionadas con el narcotráfico (esto indica que no se tendrá en cuenta el dinero lavado por informalidad o contrabando que también corresponden a la definición de la economía subterránea).

A inicio de los años 70's se da la aparición de los primeros cultivos de marihuana en Colombia; de acuerdo con información publicada por la Procuraduría General de la Nación en 1978 existían alrededor de 70.000 hectáreas de marihuana sembradas, aunque para la misma fecha en [26] estiman tan solo 40.000 hectáreas; más adelante, a través de imágenes satelitales, en [27] estiman 3.000 hectáreas en 1991.

Cerca a los años 90, la coca se producía principalmente en Bolivia y Perú, pero

luego este proceso pasó a Colombia. Este país era el encargado de movilizar la coca, importándola desde Bolivia y Perú y enviándola a Estados Unidos; en 1992 el gobierno peruano puso en pie un plan de erradicación de cultivos y captura de vehículos y aviones que presuntamente transportaban coca [28], y en 1993 las autoridades ejecutan la interrupción del puente aéreo. Sin embargo, la militarización de la guerra de la droga y la interrupción del puente aéreo no reducen el cultivo, la producción y el abastecimiento de coca [29], ya que Colombia se “adueña ” de todo el proceso (Ver Fig. 2.1, tomada de [30]), desde el cultivo de la hoja, la comercialización, el procesamiento y el transporte, lo cual resultó en la facilitación de los mecanismos y procesos necesarios para el lavado de activos [31].

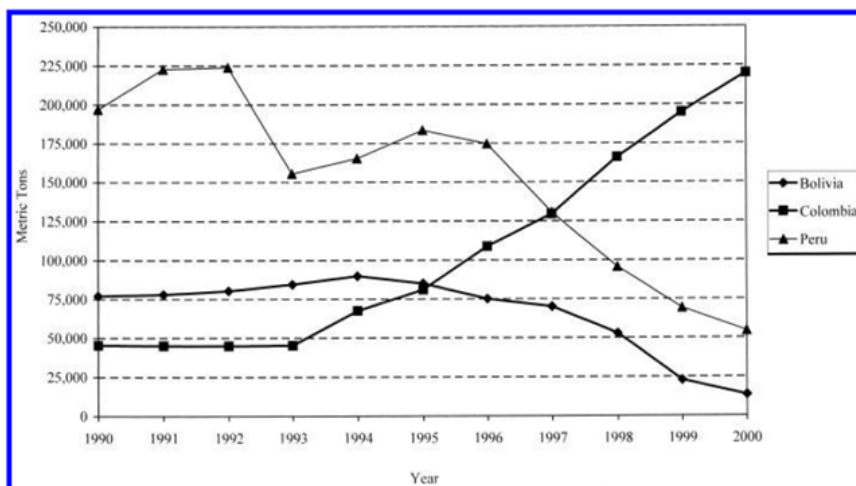


Figura 2.1: Producción de hoja de coca en Colombia, Perú y Bolivia, 1990-2000. Tomada de [30]

Posteriormente se comenzaron a encontrar cultivos de amapola, la cual es la base para la heroína. Para el 2003 departamentos como Huila, Tolima, Cesar, Caquetá, Cauca, Cundinamarca y Boyacá conformaban en conjunto 4.026 hectáreas sembradas con amapola [12].

La expansión del narcotráfico, el cual muchas veces se ha descrito como el combustible de la explosiva situación social y política del país, estuvo acompañada por el secuestro, el tráfico creciente de armas y de insumos químicos, actividades que generaron los recursos para la continuidad de los cultivos ilícitos [12].

## Pitufeo

En el estudio del lavado de dinero los entes de control relacionados con la prevención del lavado de activos y financiación del terrorismo han caracterizado los comportamientos y estrategias utilizadas en tipologías que describen de forma general los comportamientos típicos que se han usado para lavar dinero. En [24] definen las tipologías como los estudios que se concentran en analizar sectores, tendencias o modalidades en las operaciones de lavado; éstas suelen ser descriptivas y se construyen con el fin de proporcionar herramientas que le permitan a los sectores reportantes y entes de control contar con información para mejorar los mecanismos de control, detección y prevención.

2011-07-07 09:00:00

### Por 'Pitufeo' 10 personas fueron enviadas a prisión



Por 'Pitufeo' 10 personas fueron enviadas a prisión

Tweet 0 +1 0 Blogger Me gusta 0 Enviar ShareThis 4

*En una redada hecha por la Policía Nacional entre martes y miércoles, fueron capturados un total de 10 personas en Pereira y Dosquebradas, los cuales tenían deudas pendientes con la justicia, al ser señalados dentro de una investigación hecha por la Fiscalía General de la Nación de cobrar giros del exterior provenientes del narcotráfico, actividad conocida como pitufeo.*

Figura 2.2: Noticia Pitufeo en Colombia. Tomada de [32]

Dentro de las tipologías más conocidas está el pitufeo o fraccionamiento de dinero por medio de giros internacionales. La conectividad que ha generado la globalización ha facilitado el envío de dinero de personas que se encuentran en el exterior a sus familiares, pero también ha abierto un canal para que se ingresen dineros ilícitos por este mismo medio.

En el pitufeo la operación se basa en fraccionar grandes sumas de dinero en varios envíos que realizan y reciben diferentes pitufos con la finalidad de evadir los



controles existentes tanto en el país de origen como en el destinatario. El dinero viaja a través del intermediario cambiario y es cobrado por un beneficiario falso que presta su nombre para reclamarlo pero que finalmente se lo entrega a un tercero quien le paga un pequeño porcentaje. De esta forma las remesas se han convertido en un medio atractivo por su facilidad para el envío de dinero y evasión de controles, lo que la convierte en una tipología de uso común (Ver Fig. 2.2).

## 2.4. Remesas familiares

Las remesas familiares son el medio más utilizado por los emigrantes para enviar dinero a sus familiares radicados en su país de origen, lo que las ha convertido en una de las principales fuentes de ingresos para los países tercermundistas, los cuales han recibido hasta 30 millones de dólares al año desde Estados Unidos de acuerdo a un informe del BID realizado 9 años atrás [33]. Para el 2004 un promedio del 61 % de inmigrantes latinoamericanos enviaban dinero a sus familias manejando montos de aproximadamente US235 por remesa enviada[34].

<b>Remesas familiares por país de origen y por agente pagador en Colombia - año 2003</b>				
<b>Participación porcentual relativa</b>				
	<b>Pagador 1/</b>			<b>Total</b>
	<b>Grande</b>	<b>Mediano</b>	<b>Pequeño</b>	
Estados Unidos	47,4	51,3	62,4	49,2
España	40,2	17,3	21,0	35,6
Gran Bretaña	4,1	3,0	8,0	4,3
Caribe	3,7	1,6	1,9	3,3
Ecuador	0,6	0,1	5,0	0,9
Italia	0,6	7,5	0,5	1,5
Venezuela	0,1	4,5	0,1	0,7
Resto Europa	1,9	0,0	0,9	1,6
Resto América Latina	0,5	0,0	0,1	0,4
Resto del mundo	0,9	14,7	0,1	2,5
<b>Total</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>

1/ Pagador grande: paga más de US\$10 m al mes.  
Pagador mediano: paga entre US\$5 m y US\$10 m al mes.  
Pagador pequeño: paga menos de US\$5 m al mes.  
Fuente: Encuesta de costos de transacción de remesas de trabajadores, elaborada por el Banco de la República.

Figura 2.3: Remesas familiares por país de origen. Tomada de [3]

Colombia no ha estado exenta del fenómeno de la migración internacional, al

punto en que de cada cien colombianos cerca de 8 a 10 reside de manera permanente en el exterior [35]. En efecto el uso de las remesas internacionales para la adquisición de bienes básicos (mercado y servicios públicos, salud, educación, arriendo o pago de cuota de vivienda propia) ilustra su grado de importancia para la manutención y supervivencia de los hogares nacionales.

Según [36], para el 2003 aproximadamente el 50 % de los ingresos por remesas en Colombia entraron por envíos realizados desde España (Ver Fig. 2.3). De acuerdo con la Fig. 2.4, tomada de [3], las remesas pueden corresponder al 3 % del PIB nacional.

*Millones de USD, porcentajes y número de veces*

	2000	2001	2002	2003	2004	2005p	2006p
Remesas de Trabajadores: Ingresos ( Millones de USD)	1.578	2.021	2.454	3.060	3.170	3.314	3.890
Como % del PIB	1,9%	2,5%	3,0%	3,8%	3,2%	2,7%	2,9%
Como % de los ingresos corrientes - Balanza de Pagos	8,4%	10,9%	13,7%	15,4%	13,1%	11,1%	11,1%
Como % de las exportaciones de bienes	11,5%	15,7%	19,9%	22,1%	18,4%	15,2%	15,4%
Como proporción de:							
Ingresos por Inversión Extranjera Directa (%)	65,9%	80,1%	116,0%	169,9%	101,7%	32,3%	61,8%

Figura 2.4: % de remesas familiares con respecto al PIB. Tomada de [3]

## 2.5. Métodos

Esta segunda parte del marco de referencia contiene la teoría de los métodos matemáticos que se utilizan para la modelación.

### 2.5.1. Filtros de Kalman y Representación Estado-Espacio

Dado un sistema el modelador debe desarrollar un modelo matemático que pueda representar de manera adecuada la mayor parte de los comportamientos del mismo.

En la estimación se construyen las relaciones entre las variables de interés, las entradas y salidas del sistema [37]. Esta implementación se realiza con el fin de estimar los parámetros y estados de un sistema; sin embargo, para llevar a cabo este proceso, es importante resaltar que detrás existe todo un marco matemático para sustentar la teoría de la estimación, que de acuerdo con [38] provee una serie de herramientas necesarias para entender y abordar una amplia gama de problemas. La teoría de la estimación puede considerarse como un fundamento matemático natural para la asimilación de los datos ya que suministra un lenguaje fácil de entender y un número extenso de algoritmos para generar estimaciones; dentro de esta teoría, el filtro de Kalman es considerado como uno de los algoritmos de mayor importancia en las estimaciones, ya que éste obtiene una estimación óptima de las variables.

## Representación Estado-Espacio

El principio básico es modelar el sistema como un sistema dinámico lineal excitado con una perturbación aleatoria [39]. Dicho de otra forma en [40], los modelos en su representación estado-espacio son una conveniencia notacional para problemas de estimación y/o control desarrollados para facilitar el manejo del sistema en la modelación.

El estado hace referencia al vector más pequeño que resume el pasado del sistema [41], conocerlo permite la predicción teórica del estado futuro y por ende la dinámica completa del sistema.

La representación estado-espacio aplicada a una serie de tiempo se describe a través de una ecuación llamada ecuación de medida:

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t \quad (2.1)$$

$$Var(\varepsilon_t) = H_t \quad (2.2)$$

Donde:

$y_t$  : Es una serie de tiempo con  $N$  elementos.

$\alpha_t$  : Representa las variables observables de tamaño  $m \times 1$ . Se conoce como vector de estado.

$Z_t$  : Es una matriz de tamaño  $N \times m$ .

$d_t$  : Es un vector de tamaño  $N \times 1$ .

$\varepsilon_t$  : Es un vector de ruidos con media 0, de tamaño  $N \times 1$ .

Como los elementos del vector de estados son desconocidos [9], se deben modelar con un proceso de Markov de primer orden para darle relevancia solo a lo que ocurra en el momento actual:

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t \quad (2.3)$$

$$Var(\eta_t) = Q_t \quad (2.4)$$

Donde:

$T_t$  : Es una matriz cuadrada de tamaño  $m \times m$ .

$c_t$  : Es un vector de tamaño  $m \times 1$ .

$R_t$  : Es una matriz de tamaño  $m \times g$ .

$\eta_t$  : Es un vector de perturbaciones con media 0, de tamaño  $g \times 1$

El objetivo de esta representación es hallar los valores de  $\alpha_t$  de tal forma que éstos recojan toda la información posible del sistema en el tiempo  $t$ . Otra de las grandes bondades de esta metodología es la de poder estimar los parámetros; es decir, si se tiene que  $c_t$  y  $d_t$  son vectores de parámetros desconocidos, estos podrían ser estimados y pasar a ser tratados como variables de estado.

### **Filtro de Kalman clásico**

En términos matemáticos el filtro de Kalman es una herramienta que estima los estados de un sistema lineal [11] (utilizando la representación estado-espacio) sujeto a una perturbación aleatoria durante su evolución, siguiendo una ecuación de movimiento [42]. De acuerdo con [9] el filtro de Kalman es un procedimiento recursivo para computar el estimador óptimo del vector de estado y de los parámetros, basado en la información disponible en el periodo  $t$ .

El filtro de Kalman es óptimo en el sentido en que incorpora toda la información disponible [37]; éste procesa todas las mediciones sin tener en cuenta su precisión para estimar el valor actual de las variables, a través del conocimiento del sistema y la medición dinámica, los ruidos del sistema, los errores de medición y la información disponible en las condiciones iniciales; además minimiza el error en las estimaciones [41]. Es recursivo porque no requiere almacenar toda la información anterior y reprocesarla y en lugar de esto toma las nuevas predicciones como valores iniciales para la siguiente iteración [43].

**Historia del FK clásico** De acuerdo con [8] el filtro de Kalman fue desarrollado y publicado por Rudolph Kalman, aunque se conoce que en 1958 Peter Swerling había desarrollado un algoritmo similar. En algunas ocasiones lo llaman filtro de Kalman-Bucy ya que el trabajo previo de Richard Bucy condujo a Kalman a obtener dicho algoritmo.

Después de su publicación, Kalman continuó trabajando en un segundo paper para el caso continuo del filtro; sin embargo, en el primer intento de publicación fue rechazado, pero gracias a su persistencia luego de un tiempo fue aceptado, convirtiéndose en la base de diferentes temas de investigación en muchas universidades.

Las raíces del algoritmo van atrás hasta el método de Karl Gauss de los mínimos cuadrados, tal y como lo menciona Sorenson en [44] donde afirma que “el concepto gaussiano de estimación a través de mínimos cuadrados originalmente simulado para estudios astronómicos, ha construido las bases para un gran número de teorías y técnicas de estimación ”.

En 1960 en la NASA Stanley F. Schmidt le vio gran potencial para la aplicación a uno de los problemas en los que venía trabajando (estimación de la trayectoria y control para el proyecto Apolo la cual era una misión para un viaje tripulado de ida y regreso a la Luna); en ese momento el filtro de Kalman tuvo su más grande implementación.

**Algoritmo del FK clásico** La Fig. 2.5 presenta el algoritmo del filtro de Kalman, donde  $\alpha_0$  y  $P_0$  son los valores iniciales para el vector de estado o parámetros a estimar y la matriz de covarianzas para el error, respectivamente.

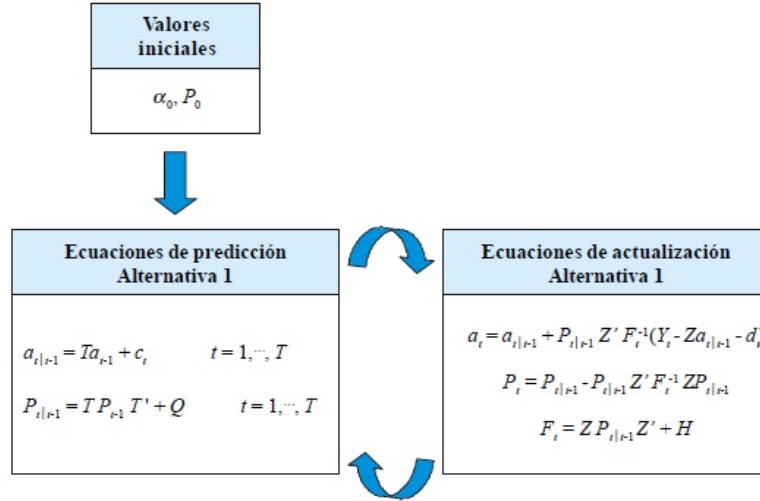


Figura 2.5: Filtro Kalman. Tomado de [13]

Este procedimiento requiere de dos ecuaciones básicas para realizar dicha estimación; las ecuaciones de predicción, utilizadas para extrapolar la solución de la iteración anterior hacia la actual [43]:

$$a_{t|t-1} = T_t a_{t-1} + c_t \quad (2.5)$$

$$P_{t|t-1} = T_t P_{t-1} T_t' + R_t Q_t R_t', t = 1, 2 \dots T \quad (2.6)$$

Una vez que la nueva observación  $y_t$  se encuentra disponible, el estimador se puede actualizar a través de las ecuaciones de actualización [9] que se encargan de corregir la predicción de la información medida [43]:

$$a_t = a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} (y_t - Z_t a_{t|t-1} - d_t) \quad (2.7)$$

$$P_t = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1} \quad (2.8)$$

$$F_t = Z_t P_{t|t-1} Z_t' + H_t \quad (2.9)$$

Esto implica tener los valores iniciales de  $a$  y  $P$  para  $t = 0$ , además de definir  $a_t$  como el estimador óptimo de  $\alpha_t$ . De acuerdo con [45] este filtro fue desarrollado en el contexto de una asimilación de datos en la cual cada estado inicial es provisto de una predicción que se inicia desde el análisis previo, de esta forma los valores iniciales están dados por la media y matriz de covarianzas de la distribución no condicionada del vector de estado.

Al implementar las ecuaciones se espera obtener buenas estimaciones; es decir, que el valor esperado de la estimación sea igual al valor esperado del estado y que el estado estimado varíe muy poco con respecto al estado real (poco error en la variación) [11].

En resumen, la información utilizada en un filtro corresponde al conocimiento del sistema y de la dinámica de los dispositivos de medición, la descripción estadística de los ruidos del sistema, errores de medición e incertidumbres de los modelos dinámicos y cualquier otra información existente en las condiciones iniciales [10]. En [46] recuerdan que el filtro de Kalman es solo una herramienta creada para modelos matemáticos, es un programa computacional ya que usa una representación finita del problema de estimación, es una caracterización estadística de un problema de estimación y es un método de aprendizaje en un contexto limitado, así que no basta solo con implementarla sino que se necesita de un modelador que tenga las suficientes herramientas para elegirla, analizarla, adecuarla y evaluarla.

**Deducción del FK clásico** Considere un sistema lineal y un vector inicial que tiene una distribución normal multinomial con media  $a_0$  y matriz de covarianza  $P_0$ . Las perturbaciones  $\eta_t$  y  $\varepsilon_t$  se distribuyen normal multivariado independiente entre sí. Tenga presente que las dinámicas no modeladas, que hacen que las anteriores

condiciones del ruido no se cumplan, pueden degradar seriamente el desempeño del filtro [10].

El vector de estado para el momento  $t = 1$ , se da en la ecuación (2.10):

$$\alpha_1 = T_1\alpha_0 + c_1 + R_1\eta_1 \quad (2.10)$$

De esta forma  $\alpha_1$  se puede escribir como una combinación lineal de dos vectores de variables aleatorias con distribución normal multivariada y un vector de constantes, lo que indica que  $\alpha_1$  es por sí mismo normal multivariado con media:

$$a_{1|0} = T_1a_0 + c_1 \quad (2.11)$$

Y matriz de covarianza:

$$P_{1|0} = T_1P_0T_1' + R_1Q_1R_1' \quad (2.12)$$

La notación  $a_{1|0}$ , indica la media de la distribución de  $\alpha_1$  condicionada a la información que se tiene en  $t = 0$ . Para obtener la distribución de  $\alpha_1$  condicionada a  $y_1$ , se pueden usar las ecuaciones (2.13) y (2.14):

$$\alpha_1 = a_{1|0} + (\alpha_1 - a_{1|0}) \quad (2.13)$$

$$y_1 = Z_1a_{1|0} + d_1 + Z_1(\alpha_1 - a_{1|0}) + \varepsilon_1 \quad (2.14)$$

De acuerdo con las dos ecuaciones anteriores se puede deducir que el vector  $\begin{bmatrix} \alpha_1' & y_1' \end{bmatrix}'$ ,



tiene una distribución normal multivariada con media  $\left[ a'_{1|0} \quad (Z_1 a_{1|0} + d_1)' \right]'$  y matriz de covarianzas: 
$$\begin{bmatrix} P_{1|0} & P_{1|0} Z_1' \\ Z_1 P_{1|0} & Z_1 P_{1|0} Z_1' + H_1 \end{bmatrix}$$

Finalmente aplicando el lema: *Sean  $x$  y  $y$  dos vectores que distribuyen normal multivariado, tal que su media y matriz de covarianzas están dadas por:*

$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

*Entonces, la distribución de  $x$  dada  $y$  es también normal multivariada con media  $\mu_{x|y} = \mu_x + \Sigma_{xy} \Sigma_{yy}^{-1} (y - \mu_y)$  y matriz de covarianzas:  $\Sigma_{xx|y} = \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx}$*   
Así, se obtiene que la distribución de  $\alpha_1$  condicionada a un valor particular de  $y_1$ , es normal multivariada con media:

$$a_1 = a_{1|0} + P_{1|0} Z_1' F_1^{-1} (y_1 - Z_1 a_{1|0} - d_1) \quad (2.15)$$

Matriz de varianzas y covarianzas:

$$P_1 = P_{1|0} - P_{1|0} Z_1' F_1^{-1} Z_1 P_{1|0} \quad (2.16)$$

$$F_1 = Z_1 P_{1|0} Z_1' + H_1 \quad (2.17)$$

Si lo anterior se utiliza para todos los valores de  $t$ , se obtiene un conjunto de ecuaciones que tienen exactamente la misma forma de las ecuaciones recursivas del filtro de Kalman. En general para cualquier  $t$ , usando (2.11), (2.12), (2.15), (2.16) y (2.17), se tiene las mismas ecuaciones de (2.5)- (2.9).

**Propiedades del FK clásico** Ciertas propiedades para los modelos expresados con la ecuaciones (2.1) y (2.3) pueden caracterizarse con propiedades de sistemas análogos de la ingeniería de control:

$$y_t = Z\alpha_t \quad (2.18)$$

$$\alpha_t = T\alpha_{t-1} + Gu_t \quad (2.19)$$

Donde  $u_t$  es el vector de variables de control de tamaño  $n \times 1$  y  $G$  es una matriz fija de tamaño  $m \times n$ .

A continuación se definen algunas de las propiedades:

- Controlabilidad: El sistema (2.18) - (2.19) es controlable si se cumple que  $Rango \left[ GTG \dots T^{m-1}G \right] = m$ . La controlabilidad implica estabilidad.

- Observabilidad: El sistema (2.18) - (2.19) es observable si:  $Rango \left[ Z'T'Z' \dots (T')^{m-1}Z' \right] = m$ . La controlabilidad implica detectabilidad.

- Estabilizabilidad: El sistema (2.18) - (2.19) es estabilizable si existe una matriz  $m \times n$  tal que:  $|\lambda_i(T + GS')| < 1$ ; donde  $\lambda_i$  representa la  $i$ -ésima raíz característica.

- Detectabilidad: El sistema (2.18) - (2.19) es detectable si existe una matriz  $D$  de tamaño  $m \times N$  tal que:  $|\lambda_i(T + DZ)| < 1$ .

Las propiedades anteriores fueron definidas para el sistema (2.18) - (2.19) pero pueden aplicarse a modelos estado-espacio estadísticos a través de una modificación en la ecuación de transición: Sea  $R^*$  una matriz de  $n \times n$ , tal que  $R^*R^{*'} = Q$ , obteniendo:

$$\alpha_t = T\alpha_{t-1} + c_t + G\eta_t^* \quad (2.20)$$

El filtro de Kalman aplicado al modelo (2.18) y (2.19), está en estado estacionario si la matriz de covarianzas del error es invariante en el tiempo:

$$P_{t+1|t} = \bar{P} \quad (2.21)$$

Por lo cual la recursión para la matriz de covarianzas del error es redundante en estado estacionario, mientras que la recursión para el estado se vuelve:

$$a_{t+1|t} = \bar{T}a_{t|t-1} + \bar{K}y_t + (c_{t+1} - \bar{K}d_t) \quad (2.22)$$

Donde las matrices de transición y ganancia se definen como:

$$\bar{T} = T - \bar{K}Z \quad (2.23)$$

$$\bar{K} = T\bar{P}Z'(Z\bar{P}Z' + H)^{-1} \quad (2.24)$$

El filtro en estado estacionario es estable si las raíces de  $T$  en valor absoluto son menores que 1.

### **Filtro de Kalman Extendido - FKE (*EKF - Extended Kalman Filter*)**

Para modelos dinámicos el filtro de Kalman es el estimador con la mínima varianza si las condiciones iniciales de la matriz de covarianzas del error y las perturbaciones  $\eta_t$  y  $\varepsilon_t$  son gaussianas [47]. De esta forma, para los casos en que dichas perturbaciones no son gaussianas el modelo se considerará no lineal y por tanto se empleará el *EKF* para la estimación.

El *EKF* es una variación para la estimación de los modelos no lineales; éste linealiza las ecuaciones de estado en cada instante del tiempo [48] alrededor de la estimación de estado más reciente, de esta forma se obtiene un modelo lineal al cual se le puede aplicar el filtro de Kalman clásico.

**Algoritmo del FKE** Considere el modelo estado-espacio (2.25) - (2.26) donde  $\varepsilon_t$  y  $\eta_t$  son procesos gaussianos independientes con media cero y matrices de covarianzas  $Q_t$  y  $R_t$  respectivamente.

$$\alpha_{t+1} = t_t(\alpha_{t-1}) + R_t(\alpha_{t-1}) + \eta_t \quad (2.25)$$

$$y_t = z_{\alpha_t} + \varepsilon_t \quad (2.26)$$

Defina:

$$\hat{Z}_t = \frac{\partial z_t(\alpha_t)}{\partial \alpha'_{t-1}} \Big|_{\alpha_t = \hat{a}_{t|t-1}} \quad (2.27)$$

$$\hat{T}_t = \frac{\partial t_t(\alpha_{t-1})}{\partial \alpha'_{t-1}} \Big|_{\alpha_{t-1} = \hat{a}_{t-1}} \quad (2.28)$$

**Inicializar:**  $t = 0$ ,  $\hat{a}_0 = E[\alpha_0]$ ,  $P_0 = E[(\alpha_0 - E[\alpha_0])(\alpha_0 - E[\alpha_0])^T]$ ,  $n = T$

**Para**  $t = 1$  hasta  $n$  haga:

$$\begin{aligned} \hat{a}_{t|t-1} &= t_t(\hat{a}_{t-1}) \\ P_{t|t-1} &= \hat{Z}_t P_{t-1} \hat{Z}_t^T + Q_{t-1} \\ K_t &= P_{t|t-1} \hat{T}_t^T [\hat{T}_t P_{t|t-1} \hat{T}_t^T + R_t]^{-1} \\ \hat{a}_t &= \hat{a}_{t|t-1} + K_t y_t - \hat{T}_t(\hat{a}_{t|t-1}) \\ P_t &= (I - K_t \hat{T}_t) P_{t|t-1} \end{aligned}$$

**Fin**

**Deducción del FKE** Considere el modelo no lineal:

$$y_t = z_t(\alpha_t) + \varepsilon_t \quad (2.29)$$

$$\alpha_t = t_t(\alpha_{t-1}) + R_t(\alpha_{t-1}) + \eta_t \quad (2.30)$$

Donde:

$z_t(\alpha_t)$  es un vector de tamaño  $N \times 1$

$t_t(\alpha_{t-1})$  es un vector de tamaño  $m \times 1$

$R_t(\alpha_{t-1})$  es una matriz de tamaño  $m \times g$

Además los elementos de  $z_t(\alpha_t)$  y  $t_t(\alpha_{t-1})$  no son necesariamente funciones lineales de los elementos del vector de estado, aunque  $R_t(\alpha_{t-1})$  sí depende de éste.

Debido a que es un problema no lineal se puede aproximar la solución a través de la linealización del modelo y luego aplicar una modificación del filtro de Kalman clásico. Luego, si las funciones no lineales:  $z_t(\alpha_t)$ ,  $t_t(\alpha_{t-1})$  y  $R_t(\alpha_{t-1})$  son lo suficien-

temente suaves se pueden llevar a series de Taylor alrededor de las medias  $\hat{a}_{t|t-1}$  y  $\hat{a}_{t-1}$ . Así:

$$z_t(\alpha_t) \simeq z_t(\hat{a}_{t|t-1} + \hat{Z}_t(\alpha_t - \hat{a}_{t|t-1})) \quad (2.31)$$

$$t_t(\alpha_{t-1}) \simeq t_t(\hat{a}_{t-1}) + \hat{T}_t(\alpha_{t-1} - \hat{a}_{t-1}) \quad (2.32)$$

$$R_t \simeq \hat{R}_t \quad (2.33)$$

Donde:

$$\hat{Z}_t = \frac{\partial z_t(\alpha_t)}{\partial \alpha'_{t-1}} \Big|_{\alpha_t = \hat{a}_{t|t-1}} \quad (2.34)$$

$$\hat{T}_t = \frac{\partial t_t(\alpha_{t-1})}{\partial \alpha'_{t-1}} \Big|_{\alpha_{t-1} = \hat{a}_{t-1}} \quad (2.35)$$

$$\hat{R}_t = R_t(\hat{a}_{t-1}) \quad (2.36)$$

Sustituyendo las ecuaciones (2.31) - (2.36) en (2.7) - (2.9) y asumiendo que se conocen  $\hat{a}_{t|t-1}$  y  $\hat{a}_{t-1}$  el modelo no lineal se puede aproximar a:

$$y_t \simeq \hat{Z}_t \alpha_t + \hat{d}_t + \varepsilon_t \quad (2.37)$$

$$\alpha_t \simeq \hat{T}_t \alpha_{t-1} + \hat{c}_t + \hat{R}_t \eta_t \quad (2.38)$$

Donde  $\hat{d}_t$  y  $\hat{c}_t$  se definen como:

$$\hat{d}_t = z_t(\hat{a}_{t|t-1}) - \hat{Z}_t \hat{a}_{t|t-1} \quad (2.39)$$

$$\hat{c}_t = t_t(\hat{a}_{t-1}) - \hat{T}_t \hat{a}_{t-1} \quad (2.40)$$

Teniendo en cuenta que la ecuación de predicción de estado es  $\hat{a}_{t|t-1} = t_t(\hat{a}_{t-1})$ , y que la ecuación de actualización de estado es  $\hat{a}_t = \hat{a}_{t|t-1} + P_{t|t-1} Z'_t F_t^{-1} [y_t - z_t(\hat{a}_{t|t-1})]$ , se pueden calcular  $\hat{a}_t$  y  $\hat{a}_{t|t-1}$  aplicando el filtro de Kalman al sistema descrito por (2.29)- (2.36).

### Filtro de Kalman Dual - FKD (*DKF - Dual Kalman Filter*)

De acuerdo con [50] gran cantidad de métodos se han propuesto para realizar estimaciones duales, algunos iterativos que se basan en realizar las estimaciones usando el modelo actual y toda la información disponible para posteriormente estimar el modelo a través de las estimaciones previas (ver imagen de la izquierda en la Fig. 2.6). Existen también propuestas secuenciales para las estimaciones duales, estos algoritmos suelen ser recursivos y trabajan necesariamente en el dominio del tiempo (ver imagen de la derecha de la Fig. 2.6).

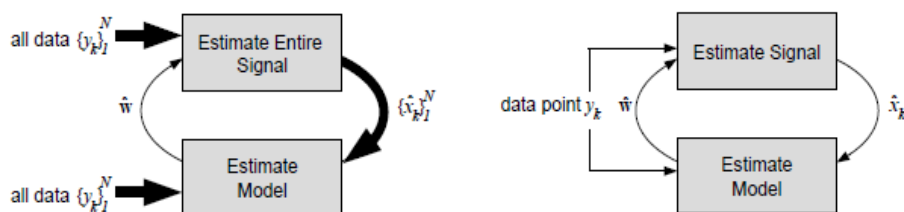


Figura 2.6: Estimación Dual. Tomada de [50]

El FKD corresponde a un algoritmo secuencial con una variación en el algoritmo del filtro clásico para estimar de manera simultánea el estado y los parámetros del modelo manteniendo los supuestos de no correlación entre las perturbaciones las cuales siguen un comportamiento gaussiano. El filtro está basado en una forma canónica para las ecuaciones estado-espacio de un sistema lineal, lo cual permite estimar a la vez los estados y parámetros usando dos estimadores lineales [51]; sin embargo, de acuerdo con [52] es posible que para algunos casos el filtro tenga problemas de convergencia. El deseo de predecir simultáneamente es bastante interesante ya que se convierte en un problema de estimar y modelar los parámetros y la variable de estado de interés.

**Historia del FKD** Entre 1963 y 1964 Kopp y Orford en [53] y Cox en [54] proponen incluir de forma conjunta como vector de estado el rezago de la señal y los parámetros desconocidos, formando así una representación estado-espacio no lineal de forma que el *EKF* entra en juego para la solución de este tipo de problemas.

Sin embargo, aunque el filtro de Kalman extendido era utilizado para las soluciones presentaba problemas de convergencia, lo que motivó a Nelson [51] a buscar un algoritmo que estimara de manera paralela el vector de estados que incluía variables y parámetros desconocidos. Ante esta necesidad surge la propuesta de utilizar dos filtros de Kalman anidados como solución alternativa al problema de estimación dual; bajo el nuevo enfoque de estimación dual uno de los filtros genera las estimaciones de la señal asumiendo conocido el modelo, mientras que el otro filtro estima los parámetros asumiendo que la señal es conocida. Posteriormente en 1983 Ljung y Söderström en [55] denominan el *DKF* como un algoritmo recursivo de identificación.

**Algoritmo del FKD** Básicamente contiene dos filtros de Kalman que corren en paralelo (ver Fig. 2.7) pero que necesitan de las estimaciones previas de parámetros y estados. Esto quiere decir que para obtener el *DKF* basta con anidar dos filtros de Kalman clásicos (para sistemas lineales) o extendidos (para sistemas no lineales), de acuerdo con el algoritmo en cada iteración el filtro estima el estado y a su vez los parámetros, usando las estimaciones previas de ambos [49].

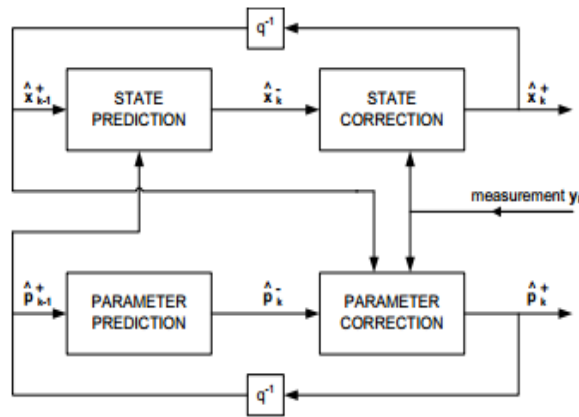


Figura 2.7: Filtro de Kalman Dual - Estimación Simultánea. Tomada de: [56]

**Inicializar:**  $t = 0$ ,  $\hat{x}_0 = Ex$ ,  $P_{x,0} = E[(x - \hat{x}_0)^T x - \hat{x}_0]$ ,  $\hat{a}_0 = E(a)$ ,  $P_{a,0} = E[(a - \hat{a}_0)^T a - \hat{a}_0]$ ,  $n = T$

**Para**  $t = 1$  hasta  $n$ , haga:

$$\begin{aligned}
a_{t|\hat{t}-1} &= a_{\hat{t}-1} \\
P_{a,t|\hat{t}-1} &= P_{a,t-1} + R_r \\
x_{t|\hat{t}-1} &= A_{t-1}x_{t-1} \\
P_{x,t|\hat{t}-1} &= A_{t-1}P_{x,t-1}A_{t-1}^T + R_v \\
K_{x,t} &= P_{x,t|\hat{t}-1}C^T(CP_{x,t|\hat{t}-1}C^T + r_n)^{-1} \\
\hat{x}_t &= x_{t|\hat{t}-1} + K_{x,t}(y_t - Cx_{t|\hat{t}-1}) \\
P_{x,t} &= (I - K_{x,t}C)P_{x,t|\hat{t}-1} \\
K_{a,t} &= P_{a,t|\hat{t}-1}C_a(C_aP_{a,t|\hat{t}-1}C_a + r_c)^{-1} \\
\hat{a}_t &= a_{t|\hat{t}-1} + K_{a,t}(y_t - Cx_{t|\hat{t}-1}) \\
P_{a,t} &= (I - K_{a,t}C_a^T)P_{a,t|\hat{t}-1}
\end{aligned}$$

**Fin**

## 2.5.2. Algoritmos Genéticos (AG) - *Genetic Algorithms (GA)*

Los AG son un heurístico adaptativo que operan sobre una población de puntos, de forma que en cada iteración un nuevo grupo de puntos se crea a partir de los cruces de los puntos anteriores, lo que significa que los AG combinan el intercambio de la información entre los individuos para hallar entre ellos al más fuerte. Desde su aparición han sido utilizados para resolver problemas de combinatorias en campos como la ingeniería, ciencia y administración.

### Historia del AG

Los algoritmos genéticos fueron inventados por John Holland y algunos de sus colegas a comienzos de 1970. Holland incorporó características naturales de la evolución para proponer una técnica robusta, computacionalmente simple y lo suficientemente poderosa para resolver problemas de optimización [57].

Básicamente, Holland quería simular el proceso de reproducción natural combinado con el principio de supervivencia biológica; de esta forma con el paso de las generaciones se encontrarán individuos (soluciones) cada vez mejores.



## **Características del AG**

Trabajan con parámetros codificados: El conjunto de parámetros debe estar en “código” por lo que es necesario encontrar la manera más eficiente para representar los cromosomas.

Buscan a partir de un set de puntos: Trabaja a partir de una gran colección de soluciones, lo que le permite aumentar la probabilidad de hallar el óptimo global.

Solo requieren los valores de la función objetivo: No están limitados por supuestos sobre el espacio de búsqueda.

Son no determinísticos: Usan reglas de transición estocásticas.

Son ciegos: No saben cuando encuentran el óptimo, por esta razón necesitan una regla de parada.

## **Terminología del AG**

- Cromosomas: Estructura que almacena la forma en que el organismo es construido.
- Ajuste: Este valor es propio del individuo y corresponde a un número que mide qué tan bueno es el individuo.
- Población Inicial: Es el número predefinido de soluciones con el que comenzará a correr el algoritmo. De acuerdo con [58] la calidad de la solución depende del tamaño de la población y de la forma en que se construyó la población inicial.
- Padres: Son los dos individuos elegidos para cruzar.
- Hijos o descendientes: Corresponde al resultado de cruzar los padres.
- Elección de Padres: Suele ser un proceso probabilístico, aunque puede estar influenciado bajo el supuesto biológico de que los organismos superiores tienden a atraerse.
- Cruce: Es el principal operador genético ya que hace referencia a la forma en que los hijos herederán las características de sus padres. Es un operador que funciona sobre 2 individuos (padres).
- Mutación: Este fenómeno produce cambios aleatorios sobre los hijos modificando algunas características de los mismos.

- Generación: Iteración, en este paso se seleccionan los individuos para cruzar.

## Algoritmo del AG

Para implementar un AG es importante codificar la solución como una cadena de símbolos que representarán a un cromosoma. El algoritmo operará sobre un conjunto de soluciones y no sobre una única posible solución, por esta razón es necesario tener un proceso constructor de la población inicial que genere una cantidad predefinida de posibles buenas soluciones. Además de este mecanismo se necesita uno adicional para generar los hijos durante el cruce de los padres seleccionados a través de su función de ajuste, una buena definición del operador de cruce generará hijos con las mejores características de cada uno de sus padres.

Con la intención de generar nuevas características es posible aplicar un operador de mutación que operará de forma aleatoria. De esta manera, luego de obtener los hijos, de esta nueva gran población se elegirán aquellos individuos con mejor función de ajuste para conformar la siguiente generación la cual pasará por este mismo proceso, garantizando que con el paso de las generaciones se obtengan individuos cada vez mejores.

El algoritmo se encuentra resumido en la Fig. 2.8, tomada de: [59])

**Variaciones en los operadores** De acuerdo con la Fig. 2.8 los operadores pueden ejecutarse de diferentes maneras, allí por cada operador mencionan tres posibles variaciones; sin embargo, existe una gran cantidad de funciones para cada uno ellos.

Cruce: Debido a que este proceso es de gran importancia es necesario realizar una buena elección para generar buenas poblaciones de hijos. Dentro de los cruces más comunes se encuentran:

- Básico: Se selecciona un punto al azar de la cadena. La parte anterior del punto es copiada de la información del padre y la posterior de la de la madre.
- Cruce multipunto: Es igual que el cruce básico solo que estableciendo más de un punto de cruce.

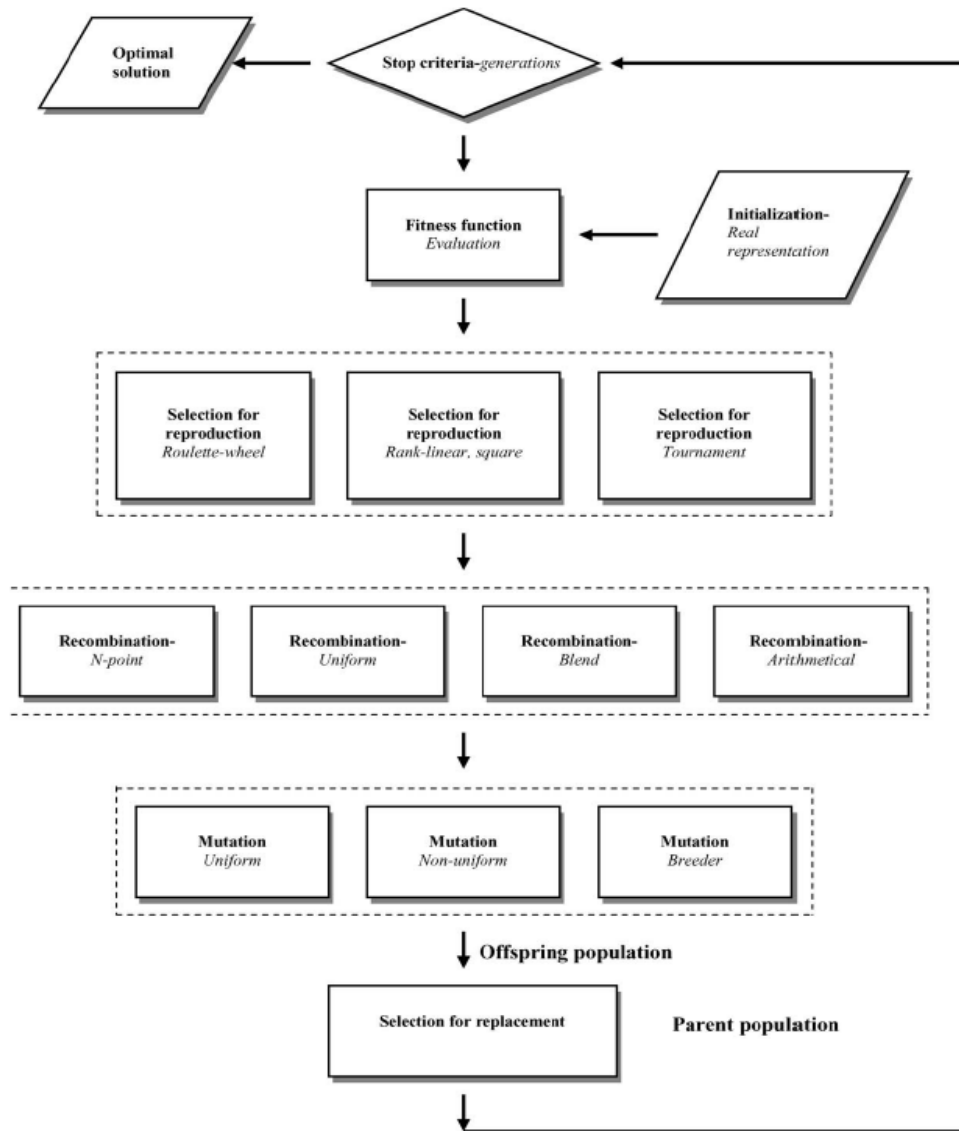


Figura 2.8: Algoritmo genético. Tomada de [59]

- Cruce uniforme: Para cada gen de la cadena del descendiente existe una probabilidad de que el gen pertenezca al padre y otra de que pertenezca a la madre.
- Cruce de mapeamiento parcial: Toma una subsecuencia del padre y procura preservar el orden absoluto de los fenotipos del resto del genoma lo más parecido posible de la madre.
- Cruce de orden: Toma una subsecuencia del genoma del padre y procura preservar

el orden relativo de los fenotipos del resto del genoma lo más parecido posible de la madre.

- Cruce de ciclo: Toma el primer gen del genoma del padre, poniéndolo en la primera posición del hijo y el primer gen del genoma de la madre, poniéndolo dentro del genoma del hijo en la posición que ocupe en el genoma del padre. El fenotipo que está en la posición que ocupa el gen del genoma del padre igual al primer gen del genoma de la madre se va a colocar en la posición que ocupe en el genoma del padre, y así hasta rellenar el genoma del hijo.

Selección Como se vió en el algoritmo el proceso se realiza sobre un grupo de la población, la forma de elegir a los individuos puede variar de acuerdo con las funciones elegidas para la selección:

- Selección directa: Toma elementos de acuerdo a un criterio objetivo (los  $n$  mejores, los  $n$  peores, etc).

- Selección equiprobable: Todos tienen la misma probabilidad de ser escogidos.

- Selección por sorteo: Cada individuo de la población tiene asignado un rango proporcional a su adaptación. Se escoge un número aleatorio dentro del rango global y el escogido es aquel que tenga dicho número dentro de su rango. La probabilidad de ser escogido es proporcional al grado de adaptación del individuo.

- Por ruleta: Se define un rango con las características de la selección por sorteo. El número al azar será un número aleatorio forzosamente menor que el tamaño del rango. El elemento escogido será aquel en cuyo rango esté el número resultante de sumar el número aleatorio con el resultado total que sirvió para escoger el elemento anterior. Tiene la ventaja de que no es posible escoger dos veces consecutivas el mismo elemento.

- Por torneo: Escoge un subconjunto de individuos de acuerdo con una de las técnicas anteriores y de entre ellos selecciona el más adecuado por otra técnica. Permite un cierto grado de elitismo.

## **AG para la identificación de sistemas**

De acuerdo con [60] la identificación paramétrica parte de un conjunto de ecuaciones dinámicas con parámetros desconocidos. La identificación de sistemas puede

dividirse en dos, la primera enfocada en la identificación estructural de las ecuaciones y la segunda en la estimación de los parámetros del modelo [61]. La orientación que se tratará será la segunda, con el fin de ajustar un modelo a datos experimentales para obtener los parámetros iniciales para el filtro de Kalman.

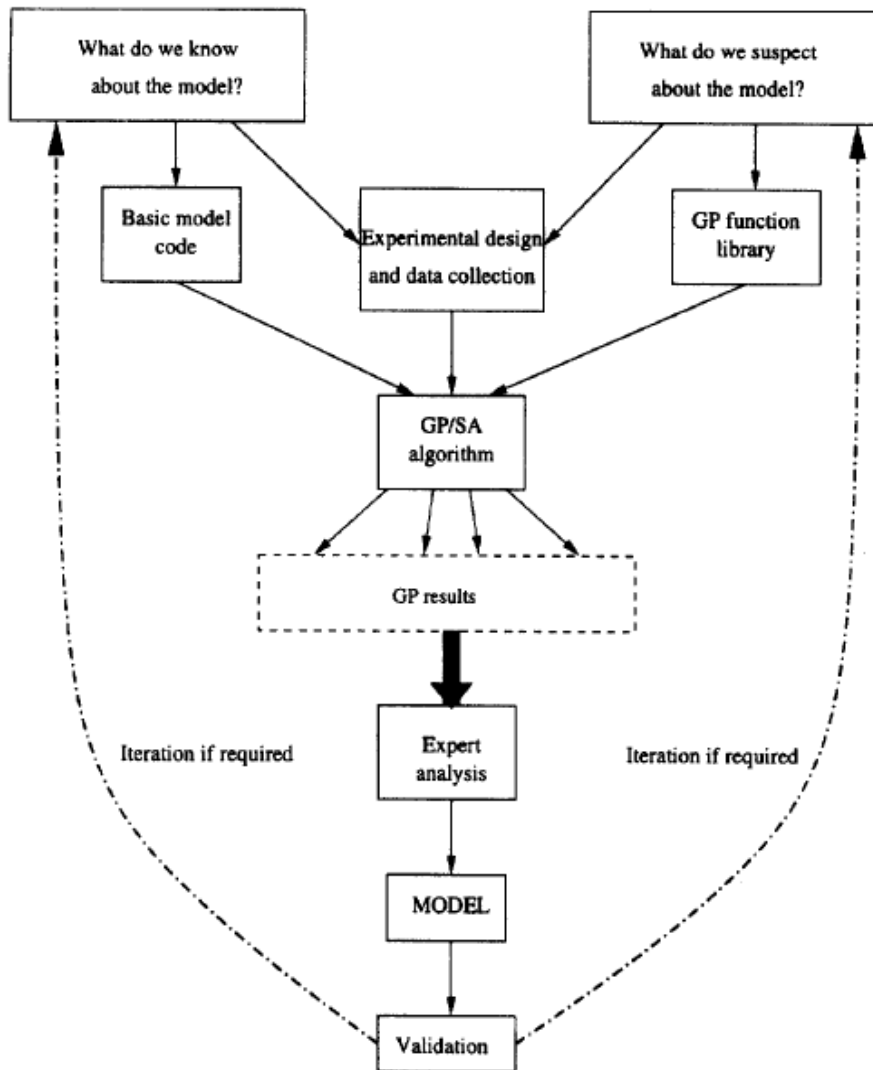


Figura 2.9: Identificación de Sistemas con Algoritmos genéticos. Tomada de: [62]

En [62] se describe la idea general de la identificación con algoritmos genéticos aplicado a modelos clásicos de la física; básicamente el autor propone una estimación guiada a través de diferentes etapas (Ver Fig. 2.9), su propuesta radica en mezclar

el conocimiento previo del sistema y del experto, el cual corresponde a un proceso importante que permitirá elegir de forma correcta los métodos a emplear para cada operador y por ende construir un AG que arrojará un resultado óptimo (correspondiente a la mejor solución de parámetros) que posteriormente será validado y de ser necesario retroalimentado.

Para el caso de la identificación de parámetros se parte de un conjunto de ecuaciones dinámicas con parámetros desconocidos, estos parámetros son estimados con el AG que se encarga de variar los valores en el sistema real, medir las respectivas respuestas y fijar los valores de los parámetros en el punto donde se logre mayor similitud entre ambas salidas (menor error).

### 2.5.3. Dinámica del dinero ilícito

De acuerdo con [4], ya mencionado con anterioridad, los agentes pueden elegir entre operar formal o informalmente, en cuyo caso sus ingresos se verán penalizados por una tasa efectiva de penalización  $\pi$ , además del poco acceso que podrán tener a los servicios públicos ( $\delta$ ); mientras que quien elige el sector formal estará afectado simplemente por una tasa impositiva  $\tau$  (impuestos y sobrecostos del Gobierno). Las ecuaciones a continuación representan los ingresos para cada sector:

$$y_t^F = (1 - \tau)A\left(\frac{G}{Y}\right)^\alpha k_j, \quad 0 < \tau < 1 \quad (2.41)$$

$$y_t^I = (1 - \pi)A\left(\frac{\delta G}{Y}\right)^\alpha k_j, \quad 0 < \pi < 1 \quad (2.42)$$

En las ecuaciones (2.41) - (2.42)  $j$  representa al agente,  $y_t^F$  corresponde a los ingresos para el sector formal,  $y_t^I$  son los ingresos para el sector informal,  $A$  es un parámetro de productividad,  $G$  es el flujo de los servicios públicos,  $Y$  es la producción total en la economía y  $\alpha$  mide la elasticidad de la salida con respecto a  $\frac{G}{Y}$ .

La ecuación (2.43) define el tamaño relativo de la economía informal:

$$I = \frac{Y^I}{Y} \quad (2.43)$$

Los servicios públicos son financiados por los impuestos o ingresos fiscales (que claramente provienen del sector formal), de acuerdo con:

$$G = \eta(q, \lambda)(\tau Y^F), \quad 0 < \eta(\cdot) \leq 1 \quad (2.44)$$

Donde  $\eta(\cdot)$  está en función de  $q$  la cual es una variable proxy que mide la calidad de las instituciones gubernamentales y  $\lambda$  que mide la eficacia de los controles a la evasión.

En el equilibrio, las ecuaciones (2.41) y (2.42) se igualan ya que los retornos relativos al capital de cada sector deberían ser iguales. De esta forma, y dado que  $\pi$  se puede escribir como el producto entre  $\lambda$  e  $I$ , es posible expresar la ecuación que mide el tamaño de la economía informal para un mercado en equilibrio como:

$$I = \frac{\delta^\alpha + \tau - 1}{\lambda \delta^\alpha} \quad (2.45)$$

De acuerdo con este modelo se puede deducir que el tamaño de la economía subterránea se puede medir a través de la existencia de restricciones regulatorias, eficacia del estado para hacerlas cumplir y/o ineficiencias en el sistema. De esta forma, para medir el tamaño relativo del dinero ilícito proveniente de las remesas familiares se propone:

$$\frac{Y_{It}}{Y_{Lt}} = \left[ \frac{Y_{It-1}}{Y_{Lt-1}} \right]^{\beta_0} \prod_{k=1}^n x_k^{\beta_k} e^{\varepsilon_{2,t}} \quad (2.46)$$

De la ecuación (2.46), la razón entre ambas economías está en función de sus niveles previos, variables que miden el crecimiento del narcotráfico, la eficacia de los controles, los impuestos, entre otros. Tomando logaritmos a (2.46) y asumiendo que  $Y_{Lt} = \lambda_1 Y_{Lt-1} + \varepsilon_{3,t}$ , se tiene:

$$\ln(Y_{It}) = \beta_0 \ln(Y_{It-1}) + \sum_{k=1}^n \beta_k x_k + \varepsilon_{2,t} \quad (2.47)$$

La ecuación (2.47) representa la forma general de la dinámica del dinero ilícito, basta con indicar los  $k$  regresores ( $x_k$ ) que describen dicho comportamiento. Dentro de las variables posibles se encuentran todas aquellas relacionadas con el crecimiento del narcotráfico (cultivos ilícitos, seguridad nacional, entre otros) y con los controles (cantidad de empleados públicos, leyes y penalización, impuestos).

#### 2.5.4. Criterios de selección de modelos

Elegir el modelo correcto para describir los datos que se tienen no es tarea sencilla, ya que debe existir una relación importante entre menor error, parsimonia y qué tan explicativo puede ser el modelo, por esta razón surgen algunos criterios que serán de gran utilidad a la hora de comparar los resultados de los modelos propuestos.

Cuando se tiene una serie de modelos con una cantidad de parámetros estimados, la función de máxima verosimilitud puede ser empleada para elegir aquel que mejor ajuste tiene; sin embargo, ésta no penaliza la complejidad del modelo como lo hacen el criterio de Akaike (AIC) y de bayes (BIC) [5]. De igual forma, es necesario medir qué tan bueno es el modelo para explicar los datos, para lo cual es utilizado el  $R^2$ .

- Criterio de información de Akaike: Combina la teoría de la máxima verosimilitud, la información teórica y la entropía de la información, teniendo en cuenta los cambios en la bondad de ajuste y las diferencias en la cantidad de parámetros en los modelos. El mejor modelo será aquel que tenga un menor valor AIC. El cálculo de este criterio es realizado con la siguiente ecuación:



$$AIC = n \ln \left( \frac{\sum_{i=1}^n (\hat{x}_i - x_i)^2}{n} \right) + 2k \quad (2.48)$$

- Criterio de información bayesiano: Al igual que el AIC tiene en cuenta la bondad del ajuste y el número de parámetros, junto con la cantidad total de datos. Un menor BIC indica un mejor modelo. Se calcula utilizando la fórmula en la (2.49):

$$BIC = n \ln \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \right) + k \ln(n) \quad (2.49)$$

- Coeficiente de determinación ( $R^2$ ): Mide el porcentaje de la variación total explicada por los modelos, dando de esta manera el porcentaje de información del modelo que es creíble. Es calculado a partir de la suma de cuadrados del error y la suma de cuadrados total.

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.50)$$

## 2.6. Resumen

A nivel nacional se encontró tan solo un trabajo orientado a la estimación de la economía subterránea, adicional a éste no existe mayor referenciación. Luego de la búsqueda del uso del FK en temas económicos se evidencia la acogida de la metodología por parte de algunos economistas nacionales, a nivel internacional existen trabajos orientados a analizar las variables relacionadas con el lavado de activos pero con metodologías diferentes.

El narcotráfico corresponde a uno de los delitos fuente de la economía subterránea, por medio de él se intenta blanquear el dinero obtenido a través de actividades relacionadas con los cultivos ilícitos. Para este fin se han encontrado diferentes tipologías

orientadas a camuflar el proceso de lavado, entre las más conocidas se encuentra el Pitufeo la cual involucra un gran número de personas en el exterior que envían pequeñas cuantías de dinero ilícito a través de remesas familiares; no obstante, el dinero es recibido por otras personas que se lo entregan a un “lavador” quien les da una pequeña comisión. Dada la importancia que han tenido las remesas familiares en la economía colombiana el impacto del dinero ilícito que ingresa a través de este medio tiene que ser medido de manera indirecta debido a que es una cuantía difícil de observar directamente en las transacciones.

El filtro de Kalman es una metodología creada por Rudolph Kalman la cual tuvo su primera aplicación en un proyecto de la NASA, básicamente estima de manera óptima los estados o parámetros no observados de un sistema lineal utilizando toda la información disponible. Sin embargo, no todos los sistemas pueden tener una representación lineal y surge entonces la necesidad de representarlos a través de modelos no lineales, de allí nace el filtro de Kalman Extendido que corresponde a una modificación en el algoritmo del FK, pero que conserva sus características de optimización. Los dos algoritmos antes mencionados tuvieron sus aplicaciones inicialmente para estimar estados, pero posteriormente se presentaron sistemas que además de tener estados no observados tenían parámetros no conocidos, lo que condujo a la creación del filtro de Kalman Dual que se encarga de anidar en un solo algoritmo las cinco ecuaciones del FK y de esta forma obtener 10 ecuaciones (5 utilizadas para estimar los estados y 5 para estimar los parámetros). Así, con estas 3 variaciones se podrían estimar estados de sistemas lineales (con el FK clásico), estados de sistemas no lineales (con el FKE) y estados junto con los parámetros del sistema (con el FKD).

Los AG son un heurístico adaptativo que combinan el intercambio de la información entre los individuos para hallar entre ellos al individuo más fuerte. Son utilizados para resolver problemas de optimización debido a que, de acuerdo con su configuración de parámetros, pueden sacar lo mejor de cada individuo o posible solución y hallar la mejor solución de las poblaciones con las que trabajan.

El modelo de Loayza es de gran utilidad para la definición general de la ecuación de transición la cual describe la dinámica del dinero ilícito, a través de dicha representación se deduce una de las hipótesis claves para la elección de las variables: el agente

que decide transar de manera informal se verá afectado por los controles del Estado, los impuestos, las penalizaciones y las variables relacionadas con el narcotráfico.

## Capítulo 3

# Resultados y discusión

El gran resultado de este capítulo se puede dividir en dos: un resultado teórico orientado a la propuesta de los modelos y un resultado práctico orientado a presentar las gráficas y ajustes.

En esta sección se presenta el problema a resolver de manera muy general, junto con los argumentos utilizados para la elección de las variables involucradas en los modelos propuestos para describir el volumen de dinero lavado por medio de las remesas familiares; posteriormente se generan algunos modelos específicos que se escriben en su forma estado-espacio para luego calcular las condiciones iniciales que se utilizan para inicializar el FKD.

Se realiza la propuesta de cuatro modelos que involucran las variables que se consideran tienen mayor impacto sobre las remesas familiares, junto con aquellas variables que interactúan en la dinámica del dinero lavado (variables de control por parte del Estado y de narcotráfico); dichas variables son utilizadas en diferentes combinaciones entre ellas para estimar los modelos que se escribirán en su forma estado-espacio, la cual es una representación necesaria para proceder a implementar cualquier FK.

El FKD necesita tener las condiciones iniciales de las variables, éstas serán calculadas con un algoritmo genético; en esta sección podrán encontrarse las rutinas y parámetros utilizados en la corrida del heurístico para la generaciones de dichos valores.

Se presentan también las soluciones halladas para los diferentes modelos híbridos presentando los parámetros encontrados y la serie de tiempo estimada (variable de estado) que corresponde al volumen de dinero ilícito no observado. A partir de estos resultados se realiza la elección del modelo seleccionando, el cual será aquel que tenga un menor error con respecto al valor observado.

Con respecto a estos resultados se espera poder identificar las variables que tienen mayor impacto no solo en el comportamiento de las remesas familiares en Colombia sino también en el dinero ilícito transado a través de ellas, usando como factor decisivo el mejor ajuste de las series.

Para el desarrollo de este capítulo se tienen en cuenta los ajustes obtenidos y las variables de cada uno de los modelos propuestos, se presentan los argumentos que descartan los modelos con el peor ajuste (modelos no exitosos) y que ayudan a sustentar la elección del mejor modelo.

### 3.1. Definición del Problema

A través de este proyecto de investigación se propone estimar el volumen del dinero ilegal movido por medio de las remesas familiares con base en variables medidas en la economía colombiana; así, se propone un modelo capaz de describir el comportamiento de este dinero.

Se propone realizar una estimación indirecta del dinero lavado a través de remesas familiares (variable no observable) usando variables observables. De acuerdo con esto se propone, de manera general, la estimación del modelo (3.1) y (3.2).

$$DR_t = \theta_1 DLR_t + \sum_{i=2}^n \theta_i VAR_i \quad (3.1)$$

$$DLR_t = \theta_{n+1} DLR_{t-1} + \sum_{i=n+2}^N \theta_i VAR_i \quad (3.2)$$

Donde:

*DR*: Dinero enviado a través de remesas familiares

*DLR*: Dinero Lavado a través del envío de remesas familiares

*VAR*: Variables definidas para describir los comportamientos de (3.1) y (3.2)

## 3.2. Modelos propuestos

### 3.2.1. Modelo 1

El primer modelo propuesto para la estimación del dinero que ingresa a Colombia por remesas familiares involucra variables como dinero legal e ilegal, tasas de desempleo de España, Reino Unido y Estados Unidos, tasa de emigración y tasa de cambio; con estos regresores se tiene el modelo(3.3). La ecuación de transición depende de los cultivos ilícitos, la cantidad de empleados públicos y la penalización por delitos del narcotráfico.

La ecuación de medida se presenta a continuación:

$$\begin{aligned} \ln(DR_t) = & \alpha_0 + \alpha_1 \ln(Y_{Lt}) + \alpha_2 \ln(Y_{It}) + \alpha_3 \ln(TE_t) \\ & + \alpha_4 TDS_t + \alpha_5 TDUK_t + \alpha_6 TDUS_t + \alpha_7 TC_t + \varepsilon_{1,t} \end{aligned} \quad (3.3)$$

Donde el dinero anual enviado a Colombia a través de remesas (*DR*) depende de factores como la tasa de emigración en el periodo (*TE*), la tasa de desempleo de España (*TDS*), Reino Unido (*TDUK*) y Estados Unidos (*TDUS*) los cuales son los principales países a los que migran los colombianos y la tasa de cambio dólar peso (por lo general los grandes pagadores de remesas monetizan en dólares). A estas variables se adiciona una variable que mida el volumen del dinero de las remesas descompuesta a través del dinero lícito y el ilícito, como variable que mide el dinero lícito se utiliza el PIB nacional del año correspondiente, ya que en él se representan los dineros que ingresan en la economía a través de actividades con un alto cumplimiento de las restricciones, el dinero ilegal de las remesas recoge aquel dinero que no es legal y estará asociado a aquellas variables que faciliten la evasión de los controles.

## Representación Estado-Espacio

Siguiendo las ecuaciones (2.1) - (2.4) y a partir de la ecuación (3.3), la ecuación de medida en su representación estado-espacio es:

$$\left[ \ln(DR_t) \right] = \left[ \theta_0 \right] \left[ \ln(Y_{It}) \right] + \left[ \theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4 \quad \theta_5 \quad \theta_6 \quad \theta_7 \right] \begin{bmatrix} c \\ \ln(Y_{Lt}) \\ \ln(TE_t) \\ TDS_t \\ TDUK_t \\ TDUS_t \\ TC_t \end{bmatrix} + \varepsilon_t \quad (3.4)$$

$$H = VAR(\varepsilon_t) = \sigma_\varepsilon^2 \quad (3.5)$$

La ecuación de transición de acuerdo con (3.2), encargada de representar la dinámica del dinero ilícito ingresado a través de las remesas, corresponde a:

$$\left[ \ln(Y_{It}) \right] = \left[ \theta_8 \right] \left[ \ln(Y_{It-1}) \right] + \left[ \theta_9 \quad \theta_{10} \quad \theta_{11} \quad \theta_{12} \right] \begin{bmatrix} \ln(Y_{Lt-1}) \\ \ln(CL_t) \\ \ln(LP_t) \\ \ln(P_t) \end{bmatrix} + \eta_t \quad (3.6)$$

$$Q = VAR(\eta_t) = \sigma_\eta^2 \quad (3.7)$$

Esta representación estado-espacio indica que 13 parámetros deben ser estimados.

## Cálculo de las condiciones iniciales

De acuerdo con el algoritmo del filtro de Kalman se necesitan las condiciones iniciales para la variable  $\ln(Y_{It})$ , los 13 parámetros  $\theta'$ s y la matriz de covarianzas  $P_0$ . Para obtener estos valores iniciales es necesario implementar un algoritmo genético que estime los valores de los parámetros del sistema, el AG se encarga de hallar estas variables para las ecuaciones (3.4) y (3.6) en el momento  $t = 1$ :

$$\left[ \ln(DR_1) \right] = \left[ \theta_0 \right] \left[ \ln(Y_{I1}) \right] + \left[ \theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4 \quad \theta_5 \quad \theta_6 \quad \theta_7 \right] \begin{bmatrix} c \\ \ln(Y_{L1}) \\ \ln(TE_1) \\ TDS_1 \\ TDUK_1 \\ TDUS_1 \\ TC_1 \end{bmatrix} + \varepsilon_1 \quad (3.8)$$

$$\left[ \ln(Y_{I1}) \right] = \left[ \theta_8 \right] \left[ \ln(Y_{I0}) \right] + \left[ \theta_9 \quad \theta_{10} \quad \theta_{11} \quad \theta_{12} \right] \begin{bmatrix} \ln(Y_{L0}) \\ \ln(CL_1) \\ \ln(LP_1) \\ \ln(P_1) \end{bmatrix} + \eta_1 \quad (3.9)$$

## Resultados

El primer modelo involucra variables relacionadas con el control y el narcotráfico como la cantidad de empleados públicos (al aumentar la cantidad de empleados públicos, y bajo el supuesto de que no son corruptos, disminuye el volumen de dinero lavado), área de cultivos ilícitos (está directamente relacionada con el crecimiento del dinero ilícito) y penalizaciones de delitos fuente de lavado de activos (esta variable esta relacionada de manera inversa, al aumentar los años en prisión o fianzas las personas evitarán cometer delitos de lavado de dinero). Mientras que la ecuación de medida relaciona variables de gran impacto en el crecimiento del envío de remesas como la tasa de emigración, el dinero legal e ilegal y tasa de cambio USDCOP, y variables de decrecimiento de remesas como tasas de desempleo de los países a los que migran los colombianos.

Siguiendo la lógica del AG (descrita en el capítulo 2) para la generación de los cromosomas se deben plantear algunas condiciones en los parámetros a estimar; sin embargo, se debe garantizar la existencia de información. Para modelar cualquier sistema es imprescindible contar con la mayor cantidad de información disponible a priori, ya sean mediciones, conocimiento para construir funciones que expliquen los comportamientos entre las variables o simple intuición o experiencia que puedan ser generalizadas a través de restricciones o expresiones matemáticas; de esta forma



tener información disponible se vuelve un requisito antes de modelar. Para este caso la dificultad se presentó al momento de buscar los datos de la variable penalizaciones ya que no se encontró un rango de valores; es decir, para un delito fuente es posible encontrar diversas penalizaciones debido a que la severidad depende de las negociaciones hechas con la fiscalía o al impacto que el delito tuvo en la sociedad. De esta manera se refleja la gran diversidad en los delitos asociados al lavado cuyas penalizaciones están asociadas al grado de participación, de conocimiento, de impacto, de volumen y de una gran cantidad de factores que hacen que cada caso sea diferente y por lo tanto se analice y castigue de manera particular.

### 3.2.2. Modelo 2

Esta segunda propuesta elimina la variable  $P_t$  (penalizaciones) e involucra variables como dinero legal e ilegal, tasas de desempleo de España, Reino Unido y Estados Unidos, tasa de emigración y tasa de cambio; con estos regresores se tiene la ecuación de medida:

$$\begin{aligned} \ln(DR_t) = & \alpha_0 + \alpha_1 \ln(Y_{Lt}) + \alpha_2 \ln(Y_{It}) + \alpha_3 \ln(TE_t) \\ & + \alpha_4 TDS_t + \alpha_5 TDUK_t + \alpha_6 TDUS_t \\ & + \alpha_7 TC_t + \varepsilon_{1,t} \end{aligned} \quad (3.10)$$

De acuerdo con la ecuación anterior el dinero anual enviado a Colombia a través de remesas ( $DR$ ) depende de factores como la tasa de emigración en el periodo ( $TE$ ), la tasa de desempleo de España ( $TDS$ ), Reino Unido ( $TDUK$ ) y Estados Unidos ( $TDUS$ ), los cuales son los principales países a los que migran los colombianos y la tasa de cambio dólar peso (por lo general los grandes pagadores de remesas monetizan en dólares). A estas variables se adiciona una variable que mida el volumen del dinero de las remesas descompuesta en el dinero lícito y el ilícito, como variable que mide el dinero lícito se utiliza el PIB nacional del año correspondiente, el dinero ilegal de las remesas recoge aquel dinero que no es legal y estará medido con variables asociadas a la efectividad de controles.

## Representación Estado-Espacio

De acuerdo con las ecuaciones (2.1) - (2.4), la representación estado-espacio de la ecuación (3.10) es:

$$\begin{bmatrix} \ln(DR_t) \end{bmatrix} = \begin{bmatrix} \theta_0 \end{bmatrix} \begin{bmatrix} \ln(Y_{It}) \end{bmatrix} + \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 & \theta_6 & \theta_7 \end{bmatrix} \begin{bmatrix} c \\ \ln(Y_{Lt}) \\ \ln(TE_t) \\ TDS_t \\ TDUK_t \\ TDUS_t \\ TC_t \end{bmatrix} + \varepsilon_t \quad (3.11)$$

$$H = VAR(\varepsilon_t) = \sigma_\varepsilon^2 \quad (3.12)$$

$$(3.13)$$

La ecuación de transición corresponde a la expresión en (3.14):

$$\begin{bmatrix} \ln(Y_{It}) \end{bmatrix} = \begin{bmatrix} \theta_8 \end{bmatrix} \begin{bmatrix} \ln(Y_{It-1}) \end{bmatrix} + \begin{bmatrix} \theta_9 & \theta_{10} & \theta_{11} \end{bmatrix} \begin{bmatrix} \ln(Y_{Lt-1}) \\ \ln(CL_t) \\ \ln(LP_t) \end{bmatrix} + \eta_t \quad (3.14)$$

$$Q = VAR(\eta_t) = \sigma_\eta^2 \quad (3.15)$$

De acuerdo con la representación estado-espacio (3.11) - (3.15) se deben estimar 12 parámetros.

## Cálculo de las condiciones iniciales

Como se ha mencionado el FK necesita las condiciones iniciales para ser implementado, dichos valores iniciales serán proporcionados con un AG que generará valores óptimos iniciales para:  $\ln(Y_{It})$ , los 12 parámetros  $\theta$ 's y la matriz de covarianzas  $P_0$ :

$$\left[ \ln(DR_1) \right] = \left[ \theta_0 \right] \left[ \ln(Y_{I1}) \right] + \left[ \theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4 \quad \theta_5 \quad \theta_6 \quad \theta_7 \right] \begin{bmatrix} c \\ \ln(Y_{L1}) \\ \ln(TE_1) \\ TDS_1 \\ TDUK_1 \\ TDUS_1 \\ TC_1 \end{bmatrix} + \varepsilon_1 \quad (3.16)$$

$$\left[ \ln(Y_{I1}) \right] = \left[ \theta_8 \right] \left[ \ln(Y_{I0}) \right] + \left[ \theta_9 \quad \theta_{10} \quad \theta_{11} \right] \begin{bmatrix} \ln(Y_{L0}) \\ \ln(CL_1) \\ \ln(LP_1) \end{bmatrix} + \eta_1 \quad (3.17)$$

Para la generación de las posibles soluciones se suponen algunas condiciones para los parámetros:  $\theta_0$  debe ser positivo ya que está adicionando el dinero que ingresa a las remesas,  $\theta_1$  está relacionado con el término independiente así que no se tiene ningún supuesto sobre su signo pero sí sobre su proporción debido a que los datos están en escala logarítmica se espera que el parámetro esté en una escala pequeña,  $\theta_2$  debe ser positivo ya que corresponde al dinero legal,  $\theta_3$  debe ser positivo ya que a mayor migración se espera más dinero entrando al país por remesas familiares; mientras que  $\theta_4, \theta_5$  y  $\theta_6$  deberán tener signo negativo debido a que si la tasa de desempleo de los países a los que migran los colombianos es cada vez mayor los ordenantes enviarán menos dinero, por su parte se espera que  $\theta_7$  sea positivo ya que a mayor tasa de cambio (USDCOP) más dinero enviarán dado que sus beneficiarios percibirán más pesos colombianos. Con respecto a los parámetros relacionados con la ecuación de transición,  $\theta_8$  debe ser positivo y menor que 1 (para garantizar que el proceso de la ecuación de transición sea estacionaria y por ende obtener un sistema estable), para  $\theta_9$  no se tiene supuestos para su signo,  $\theta_{10}$  debe ser positiva puesto que a mayor cantidad de cultivos ilícitos se espera mayor dinero lavado, por su parte  $\theta_{11}$  será negativa ya que la existencia de más servidores públicos dificulta el proceso de lavado de dinero.

Utilizando los supuestos anteriores sobre los parámetros del sistema y diferentes combinaciones de parámetros del AG, se construye el algoritmo a continuación:

**Inicializar:**  $NG, p, k, tol, m$

**Generar Espacio Solución**

**Mientras**  $n \leq NG$  o  $e \leq tol$  haga:

Evaluar cada individuo del espacio solución usando (3.16) y (3.17)

Elegir los  $k$  mejores y asignarlos como padres

Calcular  $e = 0,5 \sum_{i=1}^m (ys_i - yso_i)^2 + (y_i - y0_i)^2$

**Para**  $i = 1$  hasta  $k/2$

Cruzar padres y obtener 2 hijos por cruce

**Fin Para**

**Para**  $j = 1$  hasta  $k$

Mutar los hijos con probabilidad  $p$

Evaluar los hijos mutados usando (3.16) y (3.17)

Calcular  $e = 0,5 \sum_{i=1}^m (ys_i - yso_i)^2 + (y_i - y0_i)^2$

Remover los  $k$  peores individuos del espacio solución

Incluir los  $k$  mejores hijos (entre mutados y no mutados)

**Fin Para**

$n = n + 1$

**Fin Mientras**

Evaluar los individuos del espacio solución final usando (3.16) y (3.17)

Calcular  $e = 0,5 \sum_{i=1}^m (ys_i - yso_i)^2 + (y_i - y0_i)^2$

Elegir el mejor individuo

**Fin**

Donde  $NG$  es el número de generaciones,  $p$  es la probabilidad de mutación,  $k$  es la cantidad de individuos por generación,  $tol$  es el error mínimo permitido,  $m$  es la cantidad de individuos por generación,  $e$  es el error medio cuadrático,  $y$  es el valor estimado en (3.16),  $y0$  es el valor real del dinero que ingresa por remesas en  $t = 1$ ,  $ys$  corresponde al valor estimado en (3.17) y  $yso$  es el valor real del dinero lavado a través de remesas en  $t = 1$ .

En el AG la asignación de padres se realiza de forma determinística para garantizar que son las mejores soluciones de la población, luego de seleccionarlos los  $k$  mejores padres se cruzan por corte, es decir se genera un valor uniforme que de-

Cuadro 3.1: Parámetros del algoritmo genético para el Modelo 2

Parámetro	Valor o método
Población Inicial	3500
Generación Población Inicial	Aleatoria
Cantidad de Generaciones	10
Selección de Padres	Directa
Cantidad de Padres	1000
Cruce	Por corte
Probabilidad de Mutación	0,6
Criterio de Parada	Iteraciones o Error

termine en qué posición será cortado el padre, así el primer hijo tendrá la primera parte del padre y la segunda de la madre y al segundo hijo le corresponderán las partes restantes. Con el fin de obtener variabilidad se calcula una probabilidad de mutación de los parámetros de cada hijo para proceder a incorporarlos en un nuevo espacio solución, este proceso se repite hasta que el número de generaciones se cumpla. Finalmente se obtiene un espacio solución con las mejores respuestas de todo el proceso, de allí se extrae el mejor. La tabla 3.1. resume los valores de los parámetros para el AG.

### Algoritmo del filtro de Kalman

Luego de tener las condiciones iniciales dadas en la corrida de los AG se procede con la estimación de los parámetros y la variable de estado de cada uno de los modelos planteados en las secciones anteriores de este mismo capítulo. A continuación se presenta el algoritmo utilizado:

**Inicializar:**  $\hat{a}_1, P_{a,1}, \hat{x}_1, P_{x,1}, n = T, Q, R, R_a$

**Para**  $t = 2$  hasta  $n$ , haga:

Predicción de los parámetros

$$\hat{a}_{t|t-1} = \hat{a}_{t-1}$$

$$P_{a,t|t-1} = P_{a,t-1}$$

Predicción del estado

$$\hat{x}_{t|t-1} = \hat{a}_{t|t-1,9}x_{t-1} + \hat{a}_{t|t-1,10}\ln(Y_{Lt-1}) + \hat{a}_{t|t-1,11}\ln(CL_t) + \hat{a}_{t|t-1,12}\ln(LP_t)$$

$$P_{x,t|t-1} = \hat{a}_{t|t-1,9}P_{x,t-1}\hat{a}_{t|t-1,9}^T + Q$$

Corrección del estado

$$K_{x,t} = P_{x,t|t-1} \hat{a}'_{t|t-1,1} ((\hat{a}_{t|t-1,1} P_{x,t|t-1} \hat{a}'_{t|t-1,1}) + R)^{-1}$$

Definir  $D_t$  como el segundo término en (3.11)

$$\hat{x}_t = x_{t|t-1} + K_{x,t} ((y_t - (\hat{a}_{t|t-1,1} x_{t|t-1})) + D_t)$$

$$P_{x,t} = (I - K_{x,t} \hat{a}_{t|t-1,1}) P_{x,t|t-1}$$

Corrección de parámetros

$$K_{a,t} = P_{a,t|t-1} / (P_{a,t|t-1} + R_a)$$

Definir  $C$  como la ecuación (3.14) sustituida en (3.11)

$$\hat{a}_t = a_{t|t-1} + K_{a,t} (y_t - C)$$

$$P_{a,t} = (1 - K_{a,t}) P_{a,t|t-1}$$

**Fin**

## Resultados

Cuadro 3.2: Condiciones Iniciales - Modelo 2

Parámetro	Valor
$\theta_0$ ,	0,2822
$\theta_1$	0,2637
$\theta_2$	0,4579
$\theta_3$	0,3310
$\theta_4$	-0,8249
$\theta_5$	-1,4305
$\theta_6$	-1,0247
$\theta_7$	0,1672
$\theta_8$	0,6530
$\theta_9$	0,0534
$\theta_{10}$	1,8346
$\theta_{11}$	-0,8141

Utilizando los valores iniciales calculados con el AG dados en la tabla 3.2 se implementa un *DKF* el cual estima los parámetros y la variable de estado del modelo dado por (3.11) - (3.14). Al realizar la validación sobre el comportamiento de  $\theta_9$  se obtiene un parámetro estable en el tiempo y que se mantiene en el intervalo  $[0, 1]$  (ver Fig. 3.1), hecho que garantiza la estacionaridad o controlabilidad del sistema; dicha propiedad es deseable ya que con ella es posible garantizar la detectabilidad (el modelo tendrá solución) y estabilizabilidad, es decir que el sistema será estable y por

lo tanto el comportamiento de la serie estimada será suave y no caótico. Sin embargo, este análisis corresponde tan solo a un criterio para identificar el comportamiento del sistema y debe ser complementando con los ajustes de las series.

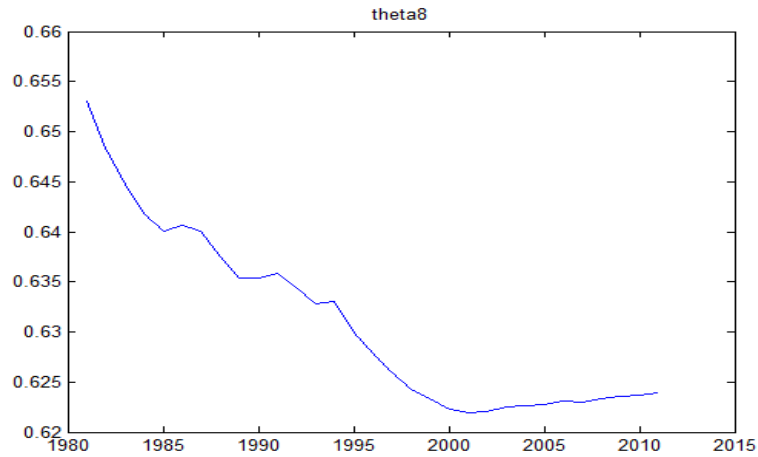


Figura 3.1: Modelo 2 - Parámetro  $\theta_8$

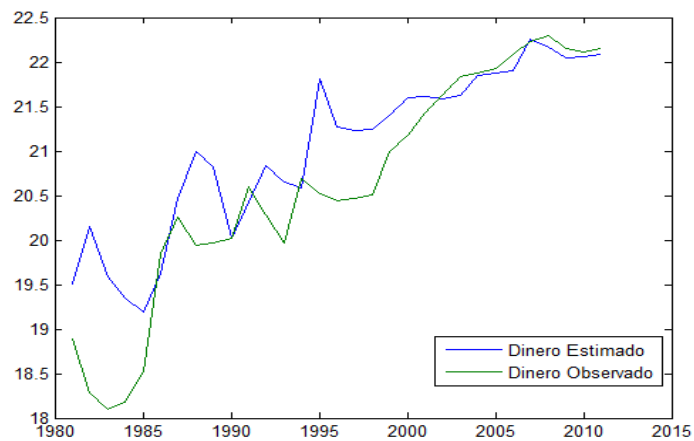


Figura 3.2: Modelo 2 - Dinero de remesas familiares en logaritmo natural. Elaboración propia

La estimación del dinero ingresado a Colombia por remesas familiares está representado en la Fig. 3.2 junto con la gráfica de la serie real (en escala logarítmica). De acuerdo con esta gráfica el ajuste que se tiene (luego de generar diferentes combinaciones de valores iniciales) contiene un error muy grande y poco tolerable lo que

no permite confiar en la estimación posterior del dinero lavado.

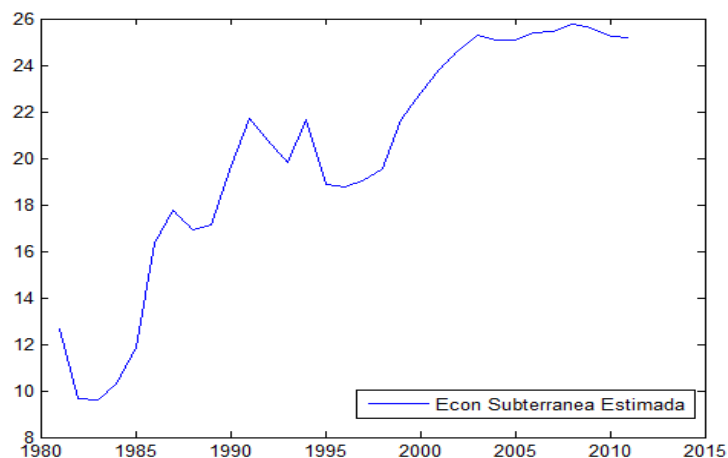


Figura 3.3: Modelo 2 - Dinero ilícito de remesas familiares en logaritmo natural. Elaboración propia

Por otro lado la Fig. 3.3 muestra el comportamiento del dinero ilegal que ha sido enviado a través de las remesas familiares durante el periodo de análisis (1980 - 2010). A medida que aumenta el tiempo el dinero lavado crece; no obstante, dicho crecimiento pasa por caídas y ascensos en diferentes periodos de tiempo quizás relacionados con las olas migratorias del país. Sin embargo, al analizar los valores y transformarlos a niveles se obtienen magnitudes muy grandes (del orden de  $10^{11}$ ), que corresponden incluso al 25 % del PIB para el 2010, dato que resulta bastante cuestionable debido a que para dicho periodo las remesas en Colombia no superaban el 3 % del PIB.

El objetivo en la estimación es obtener el menor error medio cuadrático ( $EMC$ ) posible entre la variable observada de dinero enviado a través de remesas y la variable medida, para este caso el error fue  $EMC = 7,3442$ ; dado que este error involucra propiedades como la varianza, (de acuerdo con el capítulo 2, se buscan estimaciones con variaciones pequeñas) si dicho valor es comparado con la variación anual de la variable observada (el cual fue de 1,2644) se puede concluir que el error obtenido es muy grande y que por ende deberían explorarse otros modelos. No obstante, para verificar esta conclusión se realizarán algunos test estadísticos sobre las estimaciones.



Los criterios más utilizados para comparar diferentes modelos que intentan representar el mismo sistema son los criterios de información AIC (*Akaike information criterion*) y BIC (*Bayesian information criterion*), los cuales miden la calidad de la estimación recompensando la calidad del ajuste (medida a través de la suma de los residuos al cuadrado) y penalizando la complejidad (el BIC es más exigente en este aspecto). De acuerdo con los valores calculados para cada modelo, aquel que tenga un menor valor AIC y/o BIC será el mejor, la tabla 3.3 presenta los criterios de información para el modelo 2, allí mismo el  $R^2$  informa el porcentaje de los datos que logran ser explicados con este modelo.

Cuadro 3.3: Criterios de evaluación- Modelo 2

<b>Criterio</b>	<b>Valor</b>
<i>EMC</i>	7,3442
<i>AIC</i>	15,1123
<i>BIC</i>	29,2283
$R^2$	70.37 %

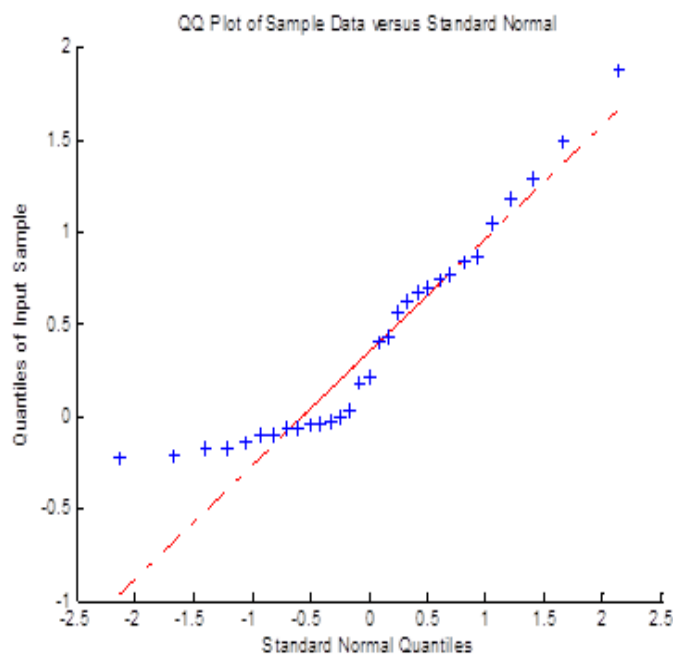


Figura 3.4: Modelo 2 - QQ-Plot

Finalmente se realiza un análisis del comportamiento de los residuos, de manera gráfica el QQ-plot (Fig. 3.4) muestra que los residuos (valores en forma de  $\times$ ) no logran seguir completamente el comportamiento de la recta la cual corresponde al comportamiento de teórico de unos datos que distribuyen normal; para verificar este resultado se ejecuta el test de Jarque-Bera obteniendo un  $p - value = 0,0740$  lo que sugiere rechazar la hipótesis, con un 95 % de certeza, de que los residuales distribuyen normal, supuesto que es necesario para la implementación del FK clásico.

### 3.2.3. Modelo 3

Esta propuesta se presenta en la ecuación (3.18):

$$\begin{aligned} \ln(DR_t) = & \alpha_0 + \alpha_1 \ln(Y_{Lt}) + \alpha_2 \ln(Y_{It}) + \alpha_3 \ln(TE_t) \\ & + \alpha_4 TDS_t + \alpha_5 TDUK_t + \alpha_6 TDUS_t + \varepsilon_{1,t} \end{aligned} \quad (3.18)$$

La relación entre las variables del modelo anterior definen que el dinero anual enviado a Colombia a través de remesas ( $DR$ ) depende de factores como la tasa de emigración en el periodo ( $TE$ ), la tasa de desempleo de España ( $TDS$ ), Reino Unido ( $TDUK$ ) y Estados Unidos ( $TDUS$ ). A estas variables se adiciona una variable que mida el volumen del dinero de las remesas descompuesta en dinero lícito e ilícito. Como variable que mide el dinero lícito se utiliza el PIB nacional del año correspondiente, ya que en él se representan los dineros que ingresan en la economía a través de actividades con un alto cumplimiento de las restricciones. El dinero ilegal de las remesas estará representado por aquellas variables que faciliten la evasión de los controles.

### Representación estado-espacio

De acuerdo con las ecuaciones (2.1) - (2.4) la ecuación de medida está dada por (3.19) y (3.20). Mientras que la ecuación de transición encargada de representar la dinámica del dinero ilícito ingresado a través de las remesas será (3.21) - (3.22).

De acuerdo con la representación estado-espacio de las ecuaciones (3.19) - (3.22) se deben estimar 12 parámetros.

$$\left[ \ln(DR_t) \right] = \left[ \theta_0 \right] \left[ \ln(Y_{It}) \right] + \left[ \theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4 \quad \theta_5 \quad \theta_6 \right] \begin{bmatrix} c \\ \ln(Y_{Lt}) \\ \ln(TE_t) \\ TDS_t \\ TDUK_t \\ TDUS_t \end{bmatrix} + \varepsilon_t \quad (3.19)$$

$$H = VAR(\varepsilon_t) = \sigma_\varepsilon^2 \quad (3.20)$$

$$\left[ \ln(Y_{It}) \right] = \left[ \theta_7 \right] \left[ \ln(Y_{It-1}) \right] + \left[ \theta_8 \quad \theta_9 \quad \theta_{10} \quad \theta_{11} \right] \begin{bmatrix} \ln(Y_{Lt-1}) \\ \ln(CL_t) \\ \ln(LP_t) \\ \ln(TDCol_t) \end{bmatrix} + \eta_t \quad (3.21)$$

$$Q = VAR(\eta_t) = \sigma_\eta^2 \quad (3.22)$$

### Cálculo de las condiciones iniciales

Para generar los valores iniciales de  $\ln(Y_{It})$ , los 12 parámetros  $\theta'$ s y la matriz de covarianzas  $P_0$  se implementó un algoritmo genético al que le corresponde estimar estas variables para las ecuaciones (3.19) y (3.21) en el momento  $t = 1$ :

$$\left[ \ln(DR_1) \right] = \left[ \theta_0 \right] \left[ \ln(Y_{I1}) \right] + \left[ \theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4 \quad \theta_5 \quad \theta_6 \right] \begin{bmatrix} c \\ \ln(Y_{L1}) \\ \ln(TE_1) \\ TDS_1 \\ TDUK_1 \\ TDUS_1 \end{bmatrix} + \varepsilon_1 \quad (3.23)$$

$$\left[ \ln(Y_{I1}) \right] = \left[ \theta_7 \right] \left[ \ln(Y_{I0}) \right] + \left[ \theta_8 \quad \theta_9 \quad \theta_{10} \quad \theta_{11} \right] \begin{bmatrix} \ln(Y_{L0}) \\ \ln(CL_1) \\ \ln(LP_1) \\ \ln(TDCol_1) \end{bmatrix} + \eta_1 \quad (3.24)$$

Para estos parámetros se espera que  $\theta_0$  sea positivo ya que está adicionando el dinero que ingresa a las remesas.  $\theta_2$  debe ser positivo ya que corresponde al dinero legal.  $\theta_3$  debe ser positivo ya que a mayor migración se espera más dinero entrando al país por remesas familiares; mientras que  $\theta_4, \theta_5$  y  $\theta_6$  deberán tener signo negativo debido a que si la tasa de desempleo de los países a los que migran los colombianos es cada vez mayor los ordenantes enviarán menos dinero. Con respecto a los parámetros relacionados en la ecuación de transición,  $\theta_7$  debe ser positivo y menor que 1 (para garantizar que el proceso de la ecuación de transición es estacionario y por ende obtener un sistema estable), para  $\theta_8$  no se tiene supuestos para su signo,  $\theta_9$  debe ser positiva puesto que a mayor cantidad de cultivos ilícitos se espera mayor dinero lavado, por su parte  $\theta_{10}$  será negativa ya que la existencia de más servidores públicos dificulta el proceso de lavado de dinero, finalmente se espera que  $\theta_{11}$  sea positivo debido a que a mayor tasa de desempleo nacional la posibilidad de incremento en actividades del narcotráfico será mayor.

Utilizando los supuestos anteriores sobre los parámetros del sistema y los valores de la tabla 3.6 se obtienen las condiciones iniciales utilizando el algoritmo a continuación:

**Inicializar:**  $NG, p, k, tol, m$

**Generar Espacio Solución**

**Mientras**  $n \leq NG$  o  $e \leq tol$  haga:

Evaluar cada individuo del espacio solución usando (3.23) y (3.24)

Elegir los  $k$  mejores y asignarlos como padres

Calcular  $e = 0,5 \sum_{i=1}^m (ys_i - yso_i)^2 + (y_i - y0_i)^2$

**Para**  $i = 1$  hasta  $k/2$

Cruzar padres y obtener 2 hijos por cruce

**Fin Para**

**Para**  $j = 1$  hasta  $k$

Mutar los hijos con probabilidad  $p$

Evaluar los hijos mutados usando (3.23) y (3.24)

Calcular  $e = 0,5 \sum_{i=1}^m (ys_i - yso_i)^2 + (y_i - y0_i)^2$

Remover los  $k$  peores individuos del espacio solución

Incluir los  $k$  mejores hijos (entre mutados y no mutados)

**Fin Para**

$$n = n + 1$$

**Fin Mientras**

Evaluar los individuos del espacio solución final usando (3.23) y (3.24)

$$\text{Calcular } e = 0,5 \sum_{i=1}^m (ys_i - yso_i)^2 + (y_i - y0_i)^2$$

Elegir el mejor individuo

**Fin**

En el algoritmo anterior  $NG$  corresponde al número de generaciones,  $p$  es la probabilidad de mutación,  $k$  es la cantidad de individuos por generación,  $tol$  es el error mínimo esperado,  $m$  es la cantidad de individuos por generación,  $e$  es el error medio cuadrático,  $y$  es el valor estimado en (3.16),  $yo$  es el valor real del dinero que ingresa por remesas en  $t = 1$ ,  $ys$  corresponde al valor estimado en (3.17) y  $yso$  es el valor real del dinero lavado a través de remesas en  $t = 1$ .

La asignación de padres se hizo de forma determinística para garantizar que éstos sean buenas soluciones, luego de seleccionar los  $k$  mejores padres estos se cruzan por corte. Con el fin de obtener variabilidad se calcula una probabilidad de mutación de los parámetros de cada hijo para proceder a incorporarlos en un nuevo espacio solución, este proceso se repite hasta que el número de generaciones se cumpla. Finalmente se tiene un espacio solución con las mejores respuestas de todo el proceso de donde se extrae el mejor. La tabla 3.1. resume los valores de los parámetros del AG.

**Algoritmo del filtro de Kalman**

Posterior al cálculo de las condiciones iniciales se realiza la estimación utilizando el algoritmo del FGD que se encuentra en el capítulo 2.

**Inicializar:**  $\hat{a}_1, P_{a,1}, \hat{x}_1, P_{x,1}, n = T, Q, R, R_a$

**Para**  $t = 2$  hasta  $n$ , haga:

Predicción de los parámetros

$$\hat{a}_{t|t-1} = \hat{a}_{t-1}$$

$$P_{a,t|t-1} = P_{a,t-1}$$

Predicción del estado

$$\hat{x}_{t|t-1} = \hat{a}_{t|t-1,9}x_{t-1} + \hat{a}_{t|t-1,10}\ln(Y_{L,t-1}) + \hat{a}_{t|t-1,11}\ln(CL_t) + \hat{a}_{t|t-1,12}\ln(LP_t)$$

$$P_{x,t|t-1} = \hat{a}_{t|t-1,9}P_{x,t-1}\hat{a}_{t|t-1,9}^T + Q$$

Corrección del estado

$$K_{x,t} = P_{x,t|t-1}\hat{a}'_{t|t-1,1}((\hat{a}_{t|t-1,1}P_{x,t|t-1}\hat{a}'_{t|t-1,1}) + R)^{-1}$$

Definir  $D_t$  como el segundo término en la ecuación (65)

$$\hat{x}_t = x_{t|t-1} + K_{x,t}((y_t - (\hat{a}_{t|t-1,1}x_{t|t-1})) + D_t)$$

$$P_{x,t} = (I - K_{x,t}\hat{a}_{t|t-1,1})P_{x,t|t-1}$$

Corrección de parámetros

$$K_{a,t} = P_{a,t|t-1}/(P_{a,t|t-1} + R_a)$$

Definir  $C$  como la ecuación (3.21) sustituida en (3.19)

$$\hat{a}_t = a_{t|t-1} + K_{a,t}(y_t - C)$$

$$P_{a,t} = (1 - K_{a,t})P_{a,t|t-1}$$

**Fin**

### Resultados Modelo 3

A partir de los resultados del algoritmo genético (ver tabla 3.4) se ejecuta el filtro de Kalman dual que estima de manera simultánea los parámetros y la variable de estado del modelo (3.19) - (3.22).

Cuadro 3.4: Condiciones Iniciales - Modelo 3

Parámetro	Valor
$\theta_0$	0,7715
$\theta_1$	1,6061
$\theta_2$	0,0634
$\theta_3$	0,2874
$\theta_4$	-0,8009
$\theta_5$	-0,0523
$\theta_6$	-0,6630
$\theta_7$	0,8325
$\theta_8$	0,7376
$\theta_9$	0,1770
$\theta_{10}$	-1,4584
$\theta_{11}$	0,4140

La serie estimada para la variable de estado (dinero ilícito) se presenta en la

Fig. 3.5. Evidentemente el volumen del dinero ilícito crece a medida que pasa el tiempo pasando por fluctuaciones que se pueden atribuir a variables externas. Contrario a los resultados del modelo anterior, las magnitudes se encuentran en una escala esperada (orden de  $10^8$ ); sin embargo, el error obtenido en el ajuste del dinero enviado por remesas (Fig. 3.6) genera inconformidad sobre los resultados del dinero ilegal transado por allí.

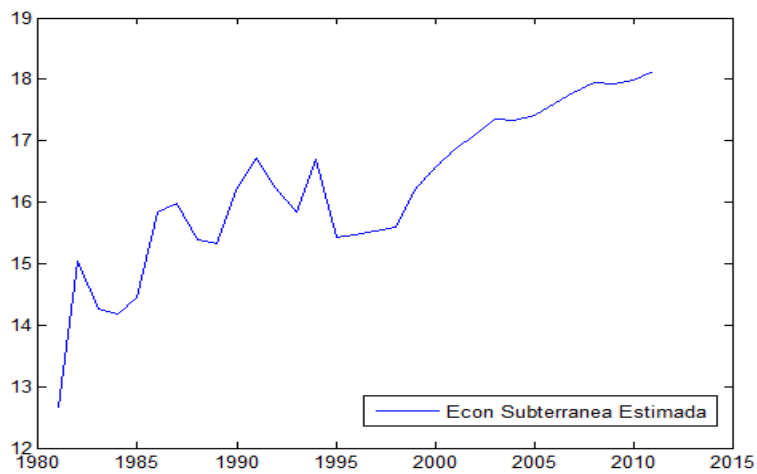


Figura 3.5: Modelo 3 - Dinero lavado por remesas. Elaboración propia

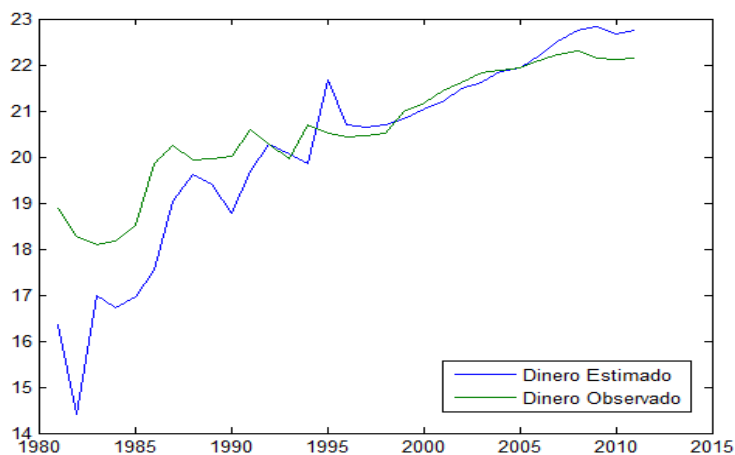


Figura 3.6: Modelo 3 - Dinero de remesas familiares. Elaboración propia

El dinero ingresado a Colombia por remesas familiares, observado y estimado, se

presenta en la Fig. 3.6; al inicio la serie ajustada intenta imitar los cambios de los datos reales aunque con mucho error, posteriormente cuando la brecha es menor la tendencia no es seguida por la serie estimada. Luego de calcular el *EMC* se obtuvo un error que supera en magnitud (ver Tabla 3.5) no solo la volatilidad de la serie sino también el error calculado con el modelo anterior; de acuerdo con este resultado este modelo es el que se ajusta en una menor medida y por ende genera mayor variabilidad entre los pronósticos y los datos reales.

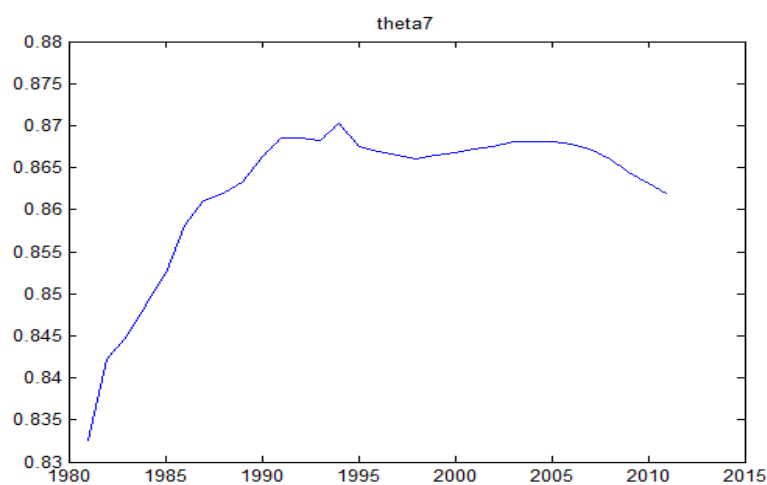


Figura 3.7: Modelo 3 - Parámetro  $\theta_7$

Al analizar el comportamiento usando la Fig. 3.7, el parámetro asociado al dinero ilícito en la ecuación de transición tiene un comportamiento muy estable que se mantiene entre 0 y 1. De esta forma se garantiza que la serie estimada del dinero ilícito sea controlable y estable.

Cuadro 3.5: Criterios de evaluación- Modelo 3

<b>Criterio</b>	<b>Valor</b>
<i>EMC</i>	20,2874
<i>AIC</i>	38,4255
<i>BIC</i>	52,5415
$R^2$	14.49%

La tabla 3.5 presenta los criterios de información para el modelo 3, de acuerdo



con el  $R^2$  este modelo no es el adecuado para estimar estas series, puesto que explica muy poca proporción de los datos.

Para este caso no se analiza el comportamiento de los residuales, ya que el modelo no es lo suficientemente bueno con las estimaciones.

### 3.2.4. Modelo 4

El último modelo implementado es (3.25):

$$\begin{aligned} \ln(DR_t) = & \alpha_0 + \alpha_1 \ln(Y_{Lt}) + \alpha_2 \ln(Y_{It}) + \alpha_3 \ln(TE_t) \\ & + \alpha_4 TDS_t + \alpha_5 TDUK_t + \alpha_6 TDUS_t + \varepsilon_{1,t} \end{aligned} \quad (3.25)$$

De (3.25) se tiene que el dinero anual que ingresa por remesas ( $DR$ ) depende de variables como la tasa de emigración anual ( $TE$ ), la tasa de desempleo de España, Reino Unido y Estados Unidos ( $TDS$ ,  $TDUK$  y  $TDUS$ , respectivamente). Adicionalmente se tiene en cuenta una variable que mide el dinero legal y el ilegal.

### Representación estado-espacio

Usando el sistema (2.1) - (2.4), se construye la ecuación de medida:

$$\left[ \ln(DR_t) \right] = \left[ \theta_0 \right] \left[ \ln(Y_{It}) \right] + \left[ \theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4 \quad \theta_5 \quad \theta_6 \right] \begin{bmatrix} c \\ \ln(Y_{Lt}) \\ \ln(TE_t) \\ TDS_t \\ TDUK_t \\ TDUS_t \end{bmatrix} + \varepsilon_t \quad (3.26)$$

$$H = VAR(\varepsilon_t) = \sigma_\varepsilon^2 \quad (3.27)$$

Por su parte la ecuación de transición está representada por:

$$\begin{bmatrix} \ln(Y_{It}) \end{bmatrix} = \begin{bmatrix} \theta_7 \end{bmatrix} \begin{bmatrix} \ln(Y_{It-1}) \end{bmatrix} + \begin{bmatrix} \theta_8 & \theta_9 & \theta_{10} \end{bmatrix} \begin{bmatrix} \ln(Y_{Lt-1}) \\ \ln(CL_t) \\ \ln(LP_t) \end{bmatrix} + \eta_t \quad (3.28)$$

$$Q = VAR(\eta_t) = \sigma_\eta^2 \quad (3.29)$$

De acuerdo con (3.27) y (3.28) se deben estimar 11 parámetros.

### Cálculo de las condiciones iniciales

De acuerdo con el algoritmo del filtro de Kalman se necesitan las condiciones iniciales para la variable  $\ln(Y_{It})$ , los 11 parámetros  $\theta$ 's y la matriz de covarianzas  $P_0$ . Para obtener dichos valores se implementó un algoritmo genético encargado de estimar los valores de los parámetros del sistema, de esta forma el AG se encarga de hallar estas variables para las ecuaciones (3.26) y (3.28) en el momento  $t = 1$ :

$$\begin{bmatrix} \ln(DR_1) \end{bmatrix} = \begin{bmatrix} \theta_0 \end{bmatrix} \begin{bmatrix} \ln(Y_{I1}) \end{bmatrix} + \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 & \theta_6 \end{bmatrix} \begin{bmatrix} c \\ \ln(Y_{L1}) \\ \ln(TE_1) \\ TDS_1 \\ TDUK_1 \\ TDUS_1 \end{bmatrix} + \varepsilon_1 \quad (3.30)$$

$$\begin{bmatrix} \ln(Y_{I1}) \end{bmatrix} = \begin{bmatrix} \theta_7 \end{bmatrix} \begin{bmatrix} \ln(Y_{I0}) \end{bmatrix} + \begin{bmatrix} \theta_8 & \theta_9 & \theta_{10} \end{bmatrix} \begin{bmatrix} \ln(Y_{L0}) \\ \ln(CL_1) \\ \ln(LP_1) \end{bmatrix} + \eta_1 \quad (3.31)$$

Para la generación de los cromosomas se tenían algunas condiciones para los parámetros a estimar, entre ellas:  $\theta_0$  debe ser positivo ya que está adicionando el dinero que ingresa a las remesas.  $\theta_1$  está relacionado con el término independiente así que no se tiene ningún supuesto sobre su signo pero sí sobre su proporción debido a que los datos están en escala logarítmica se espera que el parámetro tenga una escala pequeña.  $\theta_2$  debe ser positivo ya que corresponde al dinero legal.  $\theta_3$  debe ser positivo ya que a mayor migración se espera más dinero entrando al país por remesas

familiares; mientras que  $\theta_4, \theta_5$  y  $\theta_6$  deberán tener signo negativo debido a que si la tasa de desempleo de los países a los que migran los colombianos es cada vez mayor los ordenantes enviarán menos dinero. Con respecto a los parámetros relacionados con la ecuación de transición  $\theta_7$  debe ser positivo y menor que 1 (para garantizar que el proceso de la ecuación de transición es estacionario y por ende obtener un sistema estable), para  $\theta_8$  no se tiene supuestos para su signo,  $\theta_9$  debe ser positiva puesto que a mayor cantidad de cultivos ilícitos se espera mayor dinero lavado, por su parte  $\theta_{10}$  será negativa ya que la existencia de más servidores públicos dificulta el proceso de lavado de dinero.

Luego de correr el algoritmo y utilizando los supuestos anteriores sobre los parámetros del sistema se obtuvieron los valores iniciales para la estimación de los parámetros del modelo con el filtro de Kalman.

**Inicializar:**  $NG, p, k, tol, m$

**Generar Espacio Solución**

**Mientras**  $n \leq NG$  o  $e \leq tol$  haga:

    Evaluar cada individuo del espacio solución usando (3.30) y (3.31)

    Elegir los  $k$  mejores y asignarlos como padres

    Calcular  $e = 0,5 \sum_{i=1}^m (ys_i - yso_i)^2 + (y_i - y0_i)^2$

**Para**  $i = 1$  hasta  $k/2$

        Cruzar padres y obtener 2 hijos por cruce

**Fin Para**

**Para**  $j = 1$  hasta  $k$

        Mutar los hijos con probabilidad  $p$

        Evaluar los hijos mutados usando (3.30) y (3.31)

        Calcular  $e = 0,5 \sum_{i=1}^m (ys_i - yso_i)^2 + (y_i - y0_i)^2$

        Remover los  $k$  peores individuos del espacio solución

        Incluir los  $k$  mejores hijos (entre mutados y no mutados)

**Fin Para**

$n = n + 1$

**Fin Mientras**

Evaluar los individuos del espacio solución final usando (3.30) y (3.31)

Calcular  $e = 0,5 \sum_{i=1}^m (ys_i - yso_i)^2 + (y_i - y0_i)^2$

Elegir el mejor individuo

**Fin**

Cuadro 3.6: Parámetros del algoritmo genético

Parámetro	Valor o método
Población Inicial	3500
Generación Población Inicial	Aleatoria
Cantidad de Generaciones	10
Selección de Padres	Directa
Cantidad de Padres	1000
Cruce	Básico
Probabilidad de Mutación	0,6
Criterio de Parada	Iteraciones o Error

En el algoritmo, la variable  $NG$  se define como el número de generaciones,  $p$  como la probabilidad de mutación,  $k$  corresponde a la cantidad de individuos por generación,  $tol$  es el error mínimo aceptado,  $m$  es la cantidad de individuos por generación,  $e$  es el error medio cuadrático,  $y$  es el valor estimado en (3.16),  $y_0$  es el valor real del dinero que ingresa por remesas en  $t = 1$ ,  $ys$  corresponde al valor estimado en (3.17) y  $ys_0$  es el valor real del dinero lavado a través de remesas en  $t = 1$ .

La asignación de padres se hizo de forma determinística, los  $k$  mejores padres se cruzan por corte, es decir se genera un valor uniforme que determine en qué posición será cortado el padre, así el primer hijo tendrá la primera parte del padre y la segunda de la madre y al segundo hijo le corresponderán las partes sobrantes. Con el fin de obtener variabilidad se calcula una probabilidad de mutación de los parámetros de cada hijo para proceder a incorporarlos en un nuevo espacio solución, este proceso se repite hasta que el número de generaciones se cumpla. Finalmente se tiene un espacio solución con las mejores respuestas de todo el proceso que generará la mejor solución

### Algoritmo del filtro de Kalman

Luego de calcular las condiciones iniciales con el AG, se estima la serie de tiempo y los parámetros del modelo, utilizando el siguiente algoritmo:

**Inicializar:**  $\hat{a}_1, P_{a,1}, \hat{x}_1, P_{x,1}, n = T, Q, R, R_a$

**Para**  $t = 2$  hasta  $n$ , haga:

Predicción de los parámetros

$$\hat{a}_{t|t-1} = \hat{a}_{t-1}$$

$$P_{a,t|t-1} = P_{a,t-1}$$

Predicción del estado

$$\hat{x}_{t|t-1} = \hat{a}_{t|t-1,9}x_{t-1} + \hat{a}_{t|t-1,10}\ln(Y_{L,t-1}) + \hat{a}_{t|t-1,11}\ln(CL_t) + \hat{a}_{t|t-1,12}\ln(LP_t)$$

$$P_{x,t|t-1} = \hat{a}_{t|t-1,9}P_{x,t-1}\hat{a}_{t|t-1,9}^T + Q$$

Corrección del estado

$$K_{x,t} = P_{x,t|t-1}\hat{a}'_{t|t-1,1}((\hat{a}_{t|t-1,1}P_{x,t|t-1}\hat{a}'_{t|t-1,1}) + R)^{-1}$$

Definir  $D_t$  como el segundo término en (3.26)

$$\hat{x}_t = x_{t|t-1} + K_{x,t}((y_t - (\hat{a}_{t|t-1,1}x_{t|t-1})) + D_t)$$

$$P_{x,t} = (I - K_{x,t}\hat{a}_{t|t-1,1})P_{x,t|t-1}$$

Corrección de parámetros

$$K_{a,t} = P_{a,t|t-1}/(P_{a,t|t-1} + R_a)$$

Definir  $C$  como la ecuación (74) sustituida en (3.26)

$$\hat{a}_t = a_{t|t-1} + K_{a,t}(y_t - C)$$

$$P_{a,t} = (1 - K_{a,t})P_{a,t|t-1}$$

**Fin**

## Resultados

Cuadro 3.7: Condiciones Iniciales - Modelo 4

Parámetro	Valor
$\theta_0$	0,3985
$\theta_1$	0,5175
$\theta_2$	0,5538
$\theta_3$	0,0627
$\theta_4$	-1,4154
$\theta_5$	-0,7591
$\theta_6$	-2,5457
$\theta_7$	0,9390
$\theta_8$	-0,3027
$\theta_9$	0,9742
$\theta_{10}$	-0,1453

Utilizando las condiciones iniciales de la tabla 3.7 se implementa un  $DKF$  que estimará los parámetros y la variable de estado de forma que el error entre la variable observada y la estimada por el modelo sea cada vez más pequeño. La Fig. 3.8 muestra como la estimación del dinero que ingresa por remesas logra ajustarse en una buena proporción a la serie real.

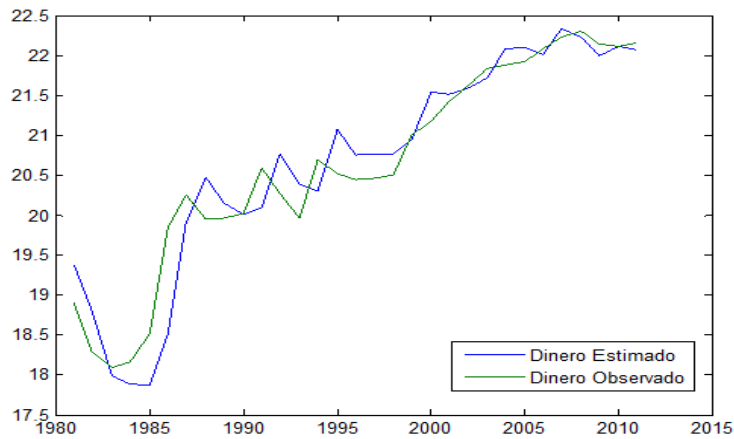


Figura 3.8: Dinero de remesas familiares en logaritmo natural. Elaboración propia

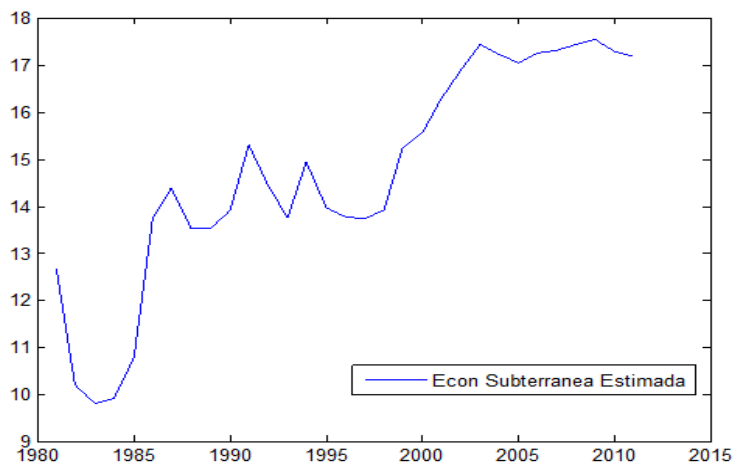


Figura 3.9: Dinero ilícito de remesas familiares en logaritmo natural. Elaboración propia

La Fig. 3.9 presenta el comportamiento del dinero ilícito enviado a través de las

remesas familiares durante el periodo de análisis. Con el pasar del tiempo el volumen ha pasado por ciclos de crecimiento y decrecimiento los cuales pueden deberse a las olas de migración de los colombianos (asociadas con los problemas de seguridad, crisis económicas nacionales y crecimiento de otros países, entre otros), o al auge en las economías de países industrializados (más adelante se realizará un análisis teniendo en cuenta estos factores).

Cuadro 3.8: Parámetros Estimados

<b>Parámetro</b>	<b>Estimación</b>
$\theta_0$	$0.399 \pm 0,003$
$\theta_1$	$0.518 \pm 0,003$
$\theta_2$	$0.555 \pm 0,001$
$\theta_3$	$0.0635 \pm 0,003$
$\theta_4$	$-1.42 \pm 0,001$
$\theta_5$	$-0.758 \pm 0,003$
$\theta_6$	$-2.54 \pm 0,002$
$\theta_7$	$0.940 \pm 0,003$
$\theta_8$	$-0.302 \pm 0,003$
$\theta_9$	$0.975 \pm 0,004$
$\theta_{10}$	$-0.144 \pm 0,002$

La tabla 3.8 presenta los resultados,  $\theta_0$  tiene el signo esperado y evidencia el peso que tiene el dinero ilícito sobre el dinero que ingresa a través de las remesas familiares, además de ser un valor menor a 1 lo que lleva a afirmar que el modelo de estimación de remesas es estable y detectable. El parámetro  $\theta_2$  resalta un peso superior por parte del dinero registrado, el hecho de que  $\theta_0$  sea menor que  $\theta_2$  implica que la proporción de dinero ilícito tiene menor participación en el total de las remesas familiares, dato que resulta ser convincente ya que es de esperarse que el dinero ilegal corresponda tan solo a una pequeña porción del total de las remesas transadas.

Los signos asociados a  $\theta_3, \theta_4, \theta_5$  y  $\theta_6$  son los esperados, lo que valida el supuesto de que a mayor migración ingresará una mayor cantidad de dinero (esta variable afecta positivamente a la variable observada), pero si las tasas de desempleo de países como España, Reino Unido y Estados Unidos aumentan el dinero enviado tenderá a ser menor (existe una relación inversa). De acuerdo con las magnitudes en valor absoluto la tasa de desempleo de Estados Unidos y de España tienen un

gran impacto (de acuerdo con la información de migración estos son los países con las tasas de emigración más altas y en gran parte del periodo analizado han tenido un buen crecimiento económico lo que facilita el envío de remesas por parte de trabajadores emigrantes)

Los parámetros asociados a la ecuación de transición tienen los signos esperados.  $\theta_7$  es menor que 1, hecho que garantiza que el proceso es estacionario lo que implica que el sistema es controlable y por ende estable. De acuerdo con el comportamiento de este parámetro en la gráfica se observa que oscila en un rango mucho menor en comparación con el mismo parámetro de los demás modelos, pero que se mantiene positivo.

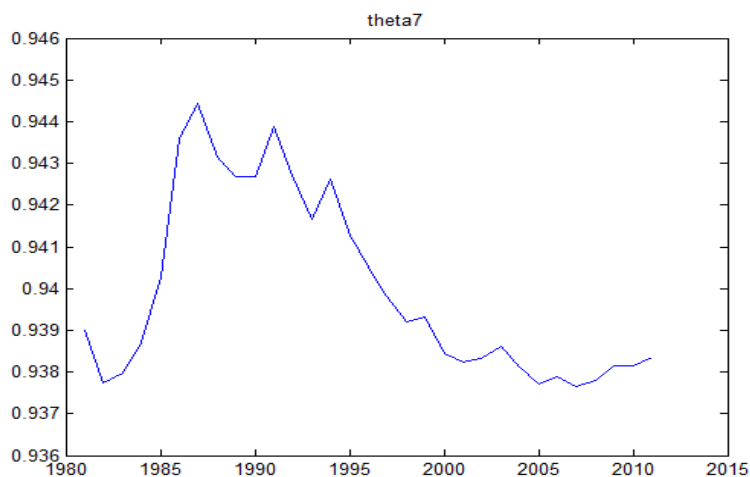


Figura 3.10: Modelo 4 - Parámetro  $\theta_7$

$\theta_9$  y  $\theta_{10}$  son parámetros que actúan de forma contraria; mientras que uno al subir (crecimiento de cultivos ilícitos) incrementa el volumen de dinero ilícito, el otro (aumento en la cantidad de servidores públicos) disminuye dicha cifra. A partir de esta ecuación es posible validar la hipótesis, planteada inicialmente en [4], que propone tener una respuesta estadísticamente significativa al establecer relaciones entre el dinero lavado y el crecimiento de la función pública y órganos de control; al analizar las magnitudes se encuentra que el crecimiento en los cultivos ilícitos es la variable que tiene mayor peso en la dinámica del dinero ilegal mientras que la cantidad de servidores públicos es la de menor impacto.



El ajuste del modelo es parcial aunque la estimación logra replicar la tendencia de la serie eal, este buen ajuste es corroborado por la medición del *EMC*, el cual es considerablemente menor al error obtenido con los modelos anteriores y relativamente cercano a la volatilidad de la serie observada (1,2644). De esta forma, al calcular esta diferencia que no es más que el promedio de las diferencias al cuadrado entre los valores estimados y reales, se garantiza que dichas estimaciones sean insesgadas (los valores sean similares) y eficientes (la variabilidad sea poca); es decir, que los valores que se están hallando para los parámetros y para la variable de estado son los apropiados para describir el comportamiento de las remesas familiares que ingresaron a Colombia durante 1980 - 2010.

No solo el *EMC* fue el menor, los criterios de información *AIC* y *BIC* son menores (ver tabla 3.3, tabla 3.5 y tabla 3.9) en comparación con los criterios de los modelos anteriores indicando que este es el mejor modelo. Por su parte el  $R^2$  indica que gran parte de la información es explicada por el modelo y que por ende sus resultados son confiables

Cuadro 3.9: Criterios de evaluación- Modelo 4

<b>Criterio</b>	<b>Valor</b>
<i>EMC</i>	2,4472
<i>AIC</i>	-8,5203
<i>BIC</i>	4,5047
$R^2$	88.91 %

Finalmente se realiza un análisis del comportamiento de los residuos, de manera gráfica el QQ-plot (Fig. 3.11) muestra que los residuos (valores en forma de  $\times$ ) siguen el comportamiento de la recta la cual corresponde al comportamiento de teórico de unos datos que distribuyen normal; para verificar este resultado se ejecuta el test de Jarque-Bera obteniendo un  $p - value = 0,0114$  lo que sugiere aceptar la hipótesis, con un 95 % de certeza, de que los residuales distribuyen normal.

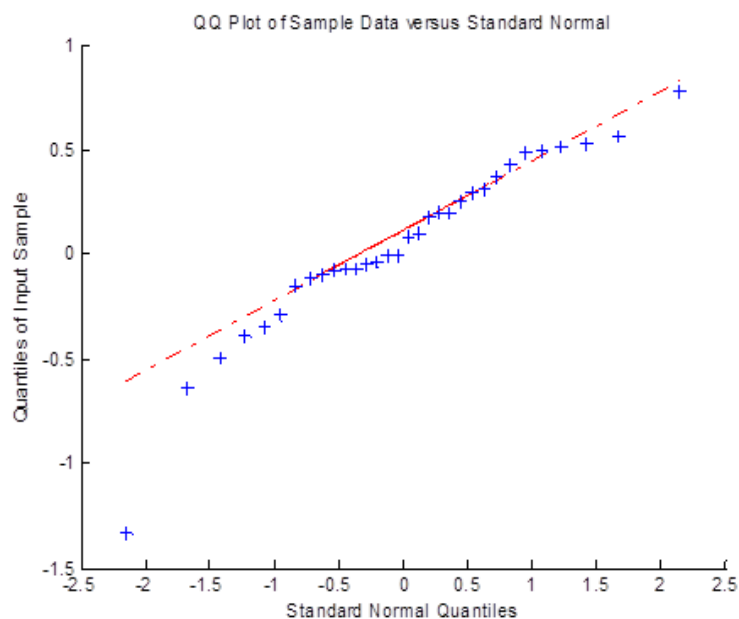


Figura 3.11: Modelo 4 - QQ-Plot

### 3.3. Análisis de Resultados

Al elegir el modelo que tiene menor error, menor criterio de información y mayor  $R^2$  deben buscarse argumentos adicionales para sustentar los resultados obtenidos. Luego de una vasta búsqueda sobre estimaciones del dinero lavado no se encontró una forma directa de validar las estimaciones debido a que es una serie difícil de observar y medir; ante esta dificultad se proponen dos formas para realizar la validación de los resultados, una cualitativa que utiliza los valores del modelo y estiman una variable de estado que es observada y una cuantitativa que relaciona los hechos históricos de Colombia relacionados con la inseguridad, la política, el narcotráfico y la migración para sustentar los incrementos y decrementos de la serie.

#### 3.3.1. Modelos no exitosos

De acuerdo con los resultados obtenidos, el modelo 1 fue descartado desde antes de realizar las estimaciones debido a la ausencia de información. Este modelo a pesar de tener una perspectiva interesante, ya que involucraba variables adicionales de control (creadas por el Estado), no pudo ser analizado puesto que para cualquier

modelo es indispensable no solo tener clara la pregunta de investigación a desarrollar sino también garantizar la existencia de información para las implementaciones pertinentes.

El segundo modelo propuesto agrega al comportamiento de las remesas familiares la posible relación con la tasa de cambio USDCOP; teniendo en cuenta que la gran mayoría de los intermediarios bancarios que pagan remesas en Colombia monetizan en dólares se adicionó solo esta tasa de cambio. Sin embargo, según los resultados de las estimaciones, los criterios *AIC*, *BIC* y el *EMC* los ajustes no fueron los esperados, se obtuvieron errores que superaban de manera significativa la variación anual de la serie, aun garantizando sistemas controlables, estables y detectables. Esto se puede sustentar en el hecho de que las remesas se han convertido en un ingreso necesario en los hogares colombianos, razón por la cual, independiente del nivel de la tasa de cambio, las personas en el exterior seguirán enviando la mesada mensual a sus familiares radicados en su país de origen.

El modelo 3 basado en los resultados de los modelos anteriores descarta el uso de la tasa de cambio y de las penalizaciones y adiciona la tasa de desempleo nacional basado en el supuesto de que a mayor tasa de desempleo en Colombia mayor envío de remesas. No obstante, luego de las estimaciones y el cálculo de los errores los ajustes tuvieron los errores más grandes con respecto a los demás modelos y el menor  $R^2$ , de esta forma se presenta la discusión sobre si dicha variable está indirectamente representada por la tasa de emigración (gran parte de los colombianos migran a otros países por la situación de seguridad o desempleo nacional), así las cosas se propone eliminar esta variable del modelo y analizar los resultados que se obtienen sin involucrarla.

### **3.3.2. Modelo elegido**

El modelo con el mejor ajuste corresponde al que involucra tasa de emigración, tasa de desempleo de los países a los que migran los colombianos, cultivos ilícitos y cantidad de empleados públicos. La validación de los resultados se ejecutó teniendo en cuenta un análisis cuantitativo y uno cualitativo:

## Validación Cualitativa

1980 es el primer año para el cual se realiza la estimación, para esta época Colombia era líder en el narcotráfico y además no tenía una buena situación de empleo; se presume que esto favorecía las condiciones para el lavado de activos y la “facilidad” de ingresar este dinero a través de las remesas familiares (fácilmente justificable por la primera oleada de emigración de 1965-1970).

Para 1981 y 1982 el PIB del país no mejoraba y se estimaba que los ingresos del narcotráfico estaban alrededor del 3.3% del PIB; sin embargo, dicha cantidad de dinero es difícil de identificar en la economía ya que Balisario Betancurt (ex presidente de Colombia) adelantó con los narcotraficantes estrechas relaciones hasta el punto de generar, en 1983, la amnistía tributaria que favoreció el ingreso del dinero del narcotráfico al sistema financiero; de esta forma lavar dinero era tan sencillo que las tipologías como el pitufeo quizás fueron desechadas en su momento.

A finales de 1984 comienza la reactivación financiera junto con el fin de la ola migratoria en 1985. A partir de este momento el flujo de emigración logra una estabilidad durante casi 10 años. Para la época es bien sabido que Medellín es una de las ciudades que guardan mayor relación con el narcotráfico; sin embargo, a pesar del crecimiento que se estaba presentando en el PIB el incremento en los depósitos al sistema financiero desde esta zona pueden parecer alarmantes y altamente relacionados con dineros ilícitos.

Finales de los años 80's es una época dura para Colombia ya que se enfrenta a los atentados del narcotráfico generando una guerra interna entre el gobierno, los grupos al margen de la ley y los carteles del narcotráfico (atentado contra el DAS, atentado contra Avianca, asesinato de Luis Carlos Galán, entre otros), además de enfrentar una crisis económica que genera una salida importante de colombianos al exterior.

César Gaviria sube a la presidencia para 1990; él debe enfrentarse directamente con las grandes cabezas del narcotráfico y vivir bajo su mandato la extrema violencia en cabeza de Pablo Escobar, la cual lleva al país a la inseguridad e inestabilidad .

Las amenazas del narcotráfico generan una disminución en la cantidad de servidores públicos logrando que de esta forma se facilite el lavado de dinero en el sistema financiero. A su vez, dicha situación intensifica la emigración de colombianos a países industrializados.

1991 es un año importante para Colombia ya que allí se firma la Nueva Constitución Política, seguido de la creación del Bloque de Búsqueda en 1992 y la interrupción del proceso de paz. Estos cambios trajeron consigo la generación de propuestas para la creación de medidas de protección de colombianos en el exterior, hecho que fácilmente pudo haber aumentado las condiciones de vida de estos y por ende el aumento en el envío de remesas. Para 1993 Pablo Escobar estaba acorralado, hasta que fue asesinado y con ello se presenta la recuperación política y económica.

En 1994 se genera una gran alarma ya que se calcula que el área sembrada de coca se ha multiplicado por 4, lo que está relacionado con el gran flujo de dinero que está produciendo y a su vez patrocinando el narcotráfico; este suceso puede sustentar el pico en la entrada de dinero lavado a través de remesas de la época.

1995 presenta una nueva ola de emigración pero ya con mayor seguridad para quienes salen del país, ya que existe asistencia para ellos en el exterior. Esta nueva etapa incrementa significativamente la entrada de dinero ilícito a Colombia lo que permite aumentar el envío de dinero sucio por remesas. Para este periodo, a pesar de que había mejorado la seguridad y empleo en el país, el narcotráfico seguía siendo representativo de Colombia, razón por la cual Estados Unidos retira al país de la lista de países que luchan contra la droga; este hecho se puede ligar directamente al aumento del narcotráfico y por ende del dinero lavado (ver periodo en la Fig. 3.8)

En 1998 Colombia se enfrenta a una nueva crisis financiera (que genera disminución en el PIB), para la cual los colombianos reaccionan migrando a países como España y Estados Unidos (que para la época contaban con buenas condiciones de empleo) haciendo que éste sea el segundo país latinoamericano con mayor flujo de emigrantes lo que posteriormente llevó al incremento en las remesas. 1999 no es un mejor año, Andrés Pastrana adelanta los diálogos de paz con las Farc y ELN y acuerda la entrega de una zona de distensión que queda al mando de los grupos al

margen de la ley (quienes posiblemente la usaban para la construcción de laboratorios de droga).

En el año 2000 la migración al exterior fue extremadamente grande debido al conflicto interno en el país; este flujo generó un aumento significativo en la entrada de remesas. Con la entrada del nuevo gobierno, en el 2002, comenzó la estabilidad financiera y crecimiento del PIB.

En el 2005 las remesas representan un importante porcentaje del PIB, por esta razón se busca implementar medidas que disminuyan el costo de las mismas. Se tiene información de que el 50 % de éstas provienen de Estados Unidos y casi el 30 % de España.

Para el 2008 los principales líderes guerrilleros y cabezas del narcotráfico han sido asesinados o capturados, hechos que conversan con una disminución en la entrada de dinero ilícito a través de remesas. En 2009 la protección de los colombianos en el exterior disminuye y la economía de los principales países a los que emigran sufre impactos negativos que disparan una caída en las remesas enviadas; sin embargo, a pesar de esta situación dicha disminución no es tan grande como lo esperaban los expertos en el tema, lo que genera sospechas acerca de la procedencia de ese dinero.

Para el 2011 las remesas a pesar de haber disminuido por la caída en la emigración, el flujo que ingresa de éstas sigue siendo representativo; hecho que se debe analizar como un posible lavado de dinero. De acuerdo con los comportamientos naturales, las remesas familiares tienen una estrecha relación con la migración de los colombianos hacia otros países; de esta forma y de manera indirecta el crecimiento en las tasas de desempleos de dichos países afectan de forma importante la cantidad de dinero que los emigrantes envían a su país de origen. El dinero que ingresa a través de las remesas familiares puede ser no solo consecuencia de la emigración, sino también producto del lavado de dinero.

Con el modelo se estima el comportamiento del dinero que ingresa a Colombia por medio de las remesas familiares con el fin de hacer una estimación indirecta del dinero que se lava a través de ellas, de esta forma será posible determinar la proporción

de remesas que corresponde a dinero ilícito, así como estimar la real relación entre las remesas, la emigración, el desempleo y el lavado. Los resultados del modelo indican que menos del 0.2 % de las remesas familiares (alrededor del 0.006 % del PIB) es dinero producto de actividades del narcotráfico, este valor no es lo suficientemente significativo en comparación con la proporción del PIB que pertenece al dinero lavado en la economía colombiana; sin embargo, es interesante analizarlo para generar políticas orientadas a disminuir dicha proporción hacia niveles más cercanos al 0 %.

A su vez, y con la intención de priorizar riesgos, estos resultados generan una visión global sobre el real comportamiento del dinero lavado por este medio, lo que lleva a concluir a que, en comparación con otras tipologías o medios de transacción de efectivo el monitoreo de remesas familiares puede tener una priorización menor o una menor asignación de recursos para mitigar el lavado de activos por medio de las remesas familiares.

### Validación Cuantitativa

Una buena validación puede realizarse utilizando el modelo y las variables estimadas para hallar una variable de estado que pueda relacionarse con el modelo, para este proceso basta con elegir una variable que se encuentre en la ecuación (3.26) y que puede asociarse a una ecuación dinámica (o de transición) para estimar dicha variable de estado. Para la estimación se conocen todos los  $\theta's$  y  $\ln(Y_{It})$  (ya que fueron estimados previamente utilizando el Modelo 4), con esto se propone estimar un modelo sencillo para hallar  $\ln(Y_{Lt})$  (economía legal):

$$\ln(Y_{Lt}) = \theta_7 + \theta_8 \ln(Y_{Lt-1}) \quad (3.32)$$

Para estimar este estado se utiliza un FK clásico que utiliza los parámetros previamente calculados para estimar la nueva variable de estado: *dinero legal*. Con este modelo se obtiene una aproximación satisfactoria del dinero legal, el cual es medible y por ende comparable con una serie real, tal y como lo muestra la Fig. 3.12.

A partir de la estimación presentada en Fig. 3.12 se obtiene el dinero lavado en-

viado a través de las remesas familiares (Ver Fig. 3.13). Este resultado basado en el ajuste del dinero enviado a través de remesas (el cual tuvo una buena aproximación con el modelo) muestra ciclos que se relacionan con las variables explicativas utilizadas.

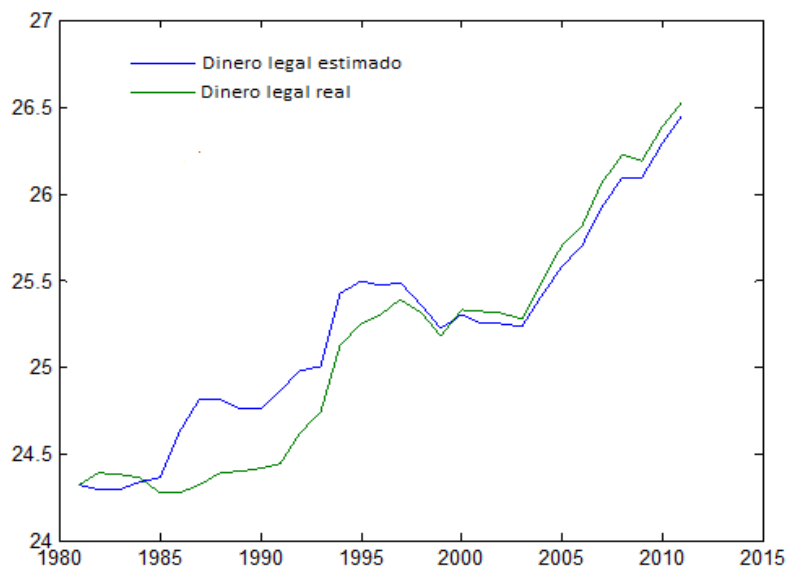


Figura 3.12: Estimación Dinero Legal. Elaboración propia

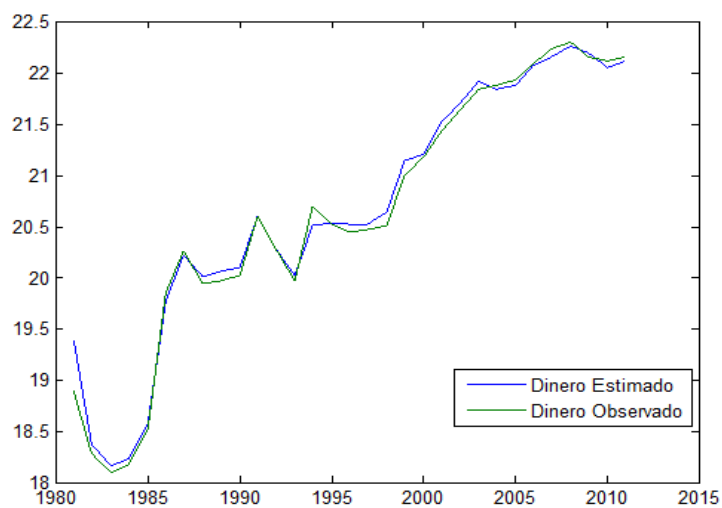


Figura 3.13: Estimación Dinero enviado a través de remesas familiares. Elaboración propia



### 3.4. Resumen

En el capítulo se presentó la forma general del modelo de estimación del dinero lavado por medio de las remesas familiares, teniendo en cuenta variables que aportan al crecimiento del dinero ilegal (como dimensión de cultivos ilícitos, cantidad de trabajadores públicos y desempleo en Colombia) y variables que están involucradas con el envío de remesas al país (tasa de emigración, tasa de desempleo de países como España, Reino Unido y Estados Unidos, dinero legal e ilegal que se transa por allí y tasa de cambio USDCOP).

Diferentes combinaciones de las variables antes mencionadas son ejecutadas para así obtener diversas propuestas de modelos, algunos de los cuales debieron ser descartados incluso antes de la implementación por falta de información, los demás fueron presentados en su forma estado-espacio y posteriormente las condiciones iniciales de las variables fueron calculadas usando un AG. De acuerdo con los algoritmos presentados en la sección el heurístico entregó buenos resultados en su forma más básica (elección determinística de padres, cruce por corte y mutación aleatoria), por lo que modificaciones en las funciones de cruce o de generación de soluciones por funciones más complejas no fueron necesarias, y por lo contrario en todos los modelos la estructura se mantuvo junto con los valores de los parámetros del algoritmo.

En la estimación de la variable de estado y de los parámetros el algoritmo del FGD fue modificado de acuerdo a la cantidad de variables a estimar en el modelo; sin embargo, en los tres casos éste convergió de manera satisfactoria. Luego de la implementación del FGD para los diferentes modelos (producto de diversas combinaciones de variables) se obtuvieron resultados que en algunos casos no se esperaban. Para el modelo 1 (el cual involucra la variable penalización  $P_t$ ) no fue posible implementar el FK debido a la falta de información.

El modelo 2 generó un ajuste del dinero de remesas con mucho error, lo que posiblemente llevó a sobredimensionar el dinero lavado a través de este medio hasta el punto de que este valor sobrepasara el porcentaje real de correspondencia de las remesas familiares con respecto al PIB colombiano.

El tercer modelo incluye el desempleo de país; sin embargo, la estimación del dinero de las remesas familiares en el país no logra reflejar el comportamiento de la variable real (tal y como lo indicaba el bajo valor obtenido con el  $R^2$ ) salvo en su tendencia lo cual no es suficiente ya que se espera un menor error en los niveles y fluctuaciones de la serie.

El último modelo logra converger y ajustarse de manera apropiada arrojando parámetros que conversan con el conocimiento previo. El ajuste del modelo que describe el dinero que ingresa por remesas familiares al país, aunque logra reflejar la tendencia, presenta errores tolerables de estimación, los menores criterios de información y un  $R^2$  que explica una gran porporción de los datos, además de que los residuales son bien comportados. Por su parte el comportamiento de la estimación realizada pasa por ciclos de crecimiento y decrecimiento, los parámetros de la ecuación de medida y de transición tienen los signos esperados. Con respecto al parámetro relacionado con la variable de estado es importante resaltar que este tiene un valor menor a 1 en valor absoluto, lo que asegura que las estimaciones son estables en el tiempo. Obtener que  $\theta_0$  sea menor que  $\theta_2$  implica que la proporción de dinero ilícito tiene menor participación en el total de las remesas familiares, lo que lleva a deducir que el dinero que ingresa al país por este medio es en legal en un alto porcentaje.

Garantizar la existencia de información es vital para la implementación de cualquier modelo, para la primera opción este fue el factor que llevó a descartar un modelo que teóricamente tenía una combinación de variables interesantes y que se esperaba tuviera muy buen ajuste.

La elección del mejor modelo inicialmente se basó en escoger aquel que tuviera mejor ajuste utilizando el EMC (error que busca medir la variabilidad y que castiga severamente las equivocaciones grandes), menor coeficiente AIC y BIC y mayor  $R^2$  para garantizar que la representación fuera la correcta; sin embargo, luego de analizar los resultados y las variables involucradas se encontraron argumentos para sustentar la selección o descarte de un modelo en particular. Para el caso del modelo 2 la razón estuvo relacionada con el real peso que tienen las remesas sobre la economía de los hogares; mientras que para el tercer modelo el argumento estuvo en relacionar la variable con otra que tiene un peso mayor sobre el volumen de remesas en el país.

Para ambos modelos, es importante resaltar que las condiciones de controlabilidad, estabilización y detectabilidad de los sistemas fueran satisfechas, pero son tan solo condiciones necesarias más no suficientes para garantizar el buen comportamiento del sistema modelado.

Finalmente, el modelo elegido (aquel que tenía mejor ajuste, menor error y mayor porcentaje de información explicadas) fue validado para soportar los resultados. De acuerdo con la validación cuantitativa, al elegir otra variable con comportamiento dinámico, que se encuentra tanto en la ecuación de transición como en la medición y que además es posible observarla, y estimarla utilizando los parámetros y estados previamente hallados con el modelo original, se obtiene una aproximación satisfactoria del dinero legal. Por su parte a través de la validación cualitativa, basada en la identificación de hechos históricos relevantes durante la época, es posible sustentar los principales aumentos del dinero lavado a través de remesas, relacionados principalmente con el incremento del poder del narcotráfico, la inseguridad, la migración, el desempleo y las áreas de cultivos ilícitos.

## Capítulo 4

# Conclusiones y Trabajo Futuro

Temas como la medición del volumen de dinero lavado a través de las remesas familiares está ligado directamente con la economía subterránea; sin embargo, a pesar de su importancia e impacto existen pocas mediciones exactas y muchas menos propuestas de modelado. Debido a que no se consiguen fácilmente datos completos sobre el dinero lavado en el tiempo no es posible estimar estas series con las metodologías supervisadas, lo que genera la necesidad de buscar otras alternativas como el filtro de Kalman. De acuerdo con la revisión de antecedentes existen pocos estudios enfocados en modelar variables no observadas directamente, razón por la cual la implementación de modelos híbridos como lo es un filtro de Kalman que toma los *inputs* a partir de un algoritmo genético, se vuelve una propuesta robusta e interesante al garantizar resultados óptimos. La combinación de estas dos metodologías matemáticas permite modelar una variable no observada de gran impacto para la economía del país, obteniendo valores iniciales óptimos que serán usados para la estimación del problema dual con el filtro de Kalman dual, el cual también garantiza obtener el mejor resultado posible.

Varios estudios han estimado que el lavado de activos en Colombia por narcotráfico es del 3% del PIB, dato que no solo resulta perturbador y alarmante para la economía colombiana sino también para las empresas, las cuales pueden tener consecuencias legales y reputacionales. Debido a esto los entes de control relacionados con la prevención de lavado de activos y financiación del terrorismo han definido tipologías que describen de forma general los comportamientos típicos que se han

usado para lavar dinero; el pitufo ha sido una de las más antiguas y utilizadas para lavar dinero gracias a las facilidades que existen en el envío de remesas familiares y en el volumen que éstas representan en la economía colombiana.

No obstante, modelar esta serie resulta ser un trabajo extenuante si se desea utilizar las metodologías tradicionales en el campo de la econometría (que son métodos supervisados), ya que no existe una serie observada con la cual comparar las estimaciones; por esta razón se propone utilizar un filtro de Kalman dual con el que se estima de manera simultánea el estado y los parámetros para hallar el modelo que describa el volumen de dinero blanqueado por el envío de remesas familiares. Para este fin es necesario involucrar variables que tengan relación con las remesas y con el lavado. De acuerdo con investigaciones previas la economía informal depende de manera directa de variables relacionadas con los impuestos y regulaciones y depende de manera inversa de la calidad y eficacia de las instituciones del gobierno; mientras que las remesas familiares dependen de variables relacionadas con la migración, tasas de desempleo de los principales países a los que migran los colombianos, dinero legal e ilegal enviado a través de ellas, entre otras.

Luego de proponer varios modelos que combinaban las variables propuestas, llevarlos a su representación estado-espacio fue un proceso sencillo debido a las transformaciones que se realizaron para trabajar con modelos lineales. Dada dicha representación detectar la variable de estado correspondía a identificar aquella variable de interés, no medible directamente que se encontraba inmersa en la dinámica de otra variable que sí es medible pero que depende de un componente aleatorio en su comportamiento. La implementación de heurísticos adaptativos como los algoritmos genéticos que trabajan sobre poblaciones de posibles soluciones para hallar el mejor resultado dentro de la misma, en general han tenido muy buen desempeño, para este caso en particular la solución obtenida produjo una convergencia positiva en el FGD que utilizó dicha información. Una población de 3500 posibles soluciones con tan solo 10 iteraciones fue suficiente para encontrar entre los individuos al que menor error producía; sin embargo, para el caso fue necesario crear un parámetro aleatorio de mutación que le ingresaría un componente de variación el cual creaba características que posiblemente no se encontraban luego de los cruces de los padres. Adicional a los supuestos planteados sobre los parámetros de este algoritmo es de vital importancia

realizar algunas consideraciones sobre las posibles soluciones para las condiciones iniciales que se desean obtener, pues a partir de la definición de éstos se garantiza que el signo estimado posteriormente por el FK tenga un sentido económico o intuitivo. No obstante, para la implementación de este primer algoritmo es necesario contar con información disponible de las variables observadas, este factor tuvo un papel importante a la hora de descartar el primer modelo, dejando por fuera cualquier modelo que involucrara las penalizaciones por delitos relacionados con el narcotráfico.

A partir de los resultados arrojados por el heurístico y de la estructura estado-espacio, se realiza la estimación simultánea de parámetros y estados a través del FKD; dicha estimación debe realizarse en secuencia utilizando las estimaciones previas de ambos ya que no se conocen los parámetros (no pueden hallarse por métodos econométricos convencionales ya que no se cuentan con mediciones de la variable de interés) ni los valores de la serie de dinero lavado. Propiedades del filtro de Kalman clásico (heredadas por las variaciones mencionadas) aseguran la controlabilidad del sistema y por ende de los estados estimados, lo cual garantiza la estabilidad del mismo garantizando que los parámetros estimados no tomen valores “extremos”; es decir, se garantiza que el sistema no tenga un comportamiento caótico. Durante la implementación se evidenció una de las grandes fortalezas de este algoritmo: la recursividad, ya que se tomaban tan solo las mediciones anteriores y aun así los errores obtenidos no crecían (existencia de poca variación entre los valores observados y los estimados en la ecuación de medida); por su lado, el hecho de utilizar toda la información disponible (desde las condiciones iniciales, las observaciones anteriores, la medición de la dinámica en las ecuaciones de transición y medida) y de minimizar el error demuestra el proceso de optimización que gobierna en la metodología.

Los ajustes, criterios de información,  $R^2$  y  $EMC$  de cada uno de los modelos permitieron elegir aquel que sería el encargado de medir el dinero ilícito transado por las remesas familiares; en el caso del modelo 2 no solo el ajuste y el error sino también la incoherencia con respecto a los resultados (el dinero lavado por remesas era mayor al dinero enviado por remesas real) llevaron a deshacerse de dicho modelo. De acuerdo con estos criterios el mejor modelo era el 4, los resultados arrojados por él verifican el buen comportamiento de los residuales y los supuestos que se tienen sobre los

signos de las variables, lo que permite afirmar que el dinero que ingresa a Colombia a través de las remesas familiares se ve afectado directamente por el dinero legal e ilegal y las emigraciones, mientras que las tasas de desempleo afectan dicho volumen de manera negativa; de forma intuitiva la tasa de cambio podría considerarse una variable altamente relacionada pero no tuvo mucho efecto sobre las estimaciones, razón por la cual se eliminó del modelo final. A partir de la ecuación de transición (responsable de modelar la dinámica de la variable de estado) el dinero ilegal que ingresa al país como remesas familiares depende de las extensiones de áreas de coca cultivada, mientras que un crecimiento en la cantidad de servidores públicos dificulta la entrada o movimiento de dicho dinero. Por su parte, los coeficientes que acompañan al dinero legal e ilegal exponen de forma clara que el dinero legal no solo representa una mayor proporción de las remesas familiares sino que también circula con mayor velocidad; este suceso puede generar tranquilidad con respecto al producto pero también genera conclusiones como que a través de las remesas familiares no se lava una gran cantidad de dinero. La conclusión anterior es bastante importante ya que da pie a proponer este tipo de modelo como etapa previa para la toma de decisiones sobre qué negocios o sectores de la economía monitorear con mayor o menor prioridad. La situación del país, además de cifras, estadísticas y estimaciones son una clara evidencia de que en Colombia el dinero ilícito sigue entrando al sistema financiero pero probablemente por otro tipo de medios (quizás unos más complejos y difíciles de rastrear); de esta forma, luego de realizar un exhaustivo análisis a través de estas estimaciones se podrán tomar decisiones de ranqueo para la construcción de monitoreos, controles y políticas encargados de detectar y prevenir el lavado de activos.

Este tema tiene gran relevancia tanto en instituciones públicas como privadas debido a que ambas se exponen a riesgos legales, reputacionales y de contagio que pueden derivar en sanciones, condenas en la cárcel e incluso en la disolución de la organización. Dada esta situación las empresas están obligadas a implementar controles que permitan mitigar, vigilar, prevenir y detectar operaciones sospechosas y diferentes tipologías utilizadas para blanquear dinero. De esta forma se propone como trabajo futuro replicar el modelo para las empresas teniendo en cuenta las variables propias de manera que pueda tener una medición propia la cual le ayudará a priorizar los temas a monitorear. Adicionalmente es posible agregar otras variables que estén re-

lacionadas y permitan tener buenas estimaciones del dinero ilegal enviado a través de las remesas familiares.

La facilidad en la implementación y robustez teórica de este modelo pueden ser aprovechadas para realizar estimaciones de otras variables relacionadas principalmente con negocios que actualmente se encuentran más expuestos al lavado de activos y financiación del terrorismo (Divisas en general, negocios fiduciarios, negocios de valores, dinero de sectores más riesgosos, entre otros), mientras que su robustez puede tomarse como un beneficio para la obtención de resultados con poco error.

Como propuestas adicionales se propone realizar la inclusión de otros regresores que tengan información disponible pero tratando de mantener la parsimonia del modelo. También puede ser interesante construir un algoritmo genético un poco más robusto (funciones de cruce o elección de padres más complejas) o incluso otro tipo de heurísticos que generen otro tipo de soluciones óptimas para las condiciones iniciales, buscando con esto un mejor ajuste del modelo. Utilizar otro tipo de error para medir el ajuste, quizás uno que castigue menos los grandes errores (que son poco frecuentes) permite ser menos exigentes con la variabilidad esperada y quizás abra la posibilidad a tener modelos que involucren otras variables.



# Referencias

- [1] Escobar, J. Una Medición de la Economía Subterránea peruana: 1980-2005, pp 44-49 en *Economía y Sociedad 75*, CIES, 2010.
- [2] FELABAN. Señales de alerta y métodos para el lavado de activos/Métodos usados (en línea: Visitado el 14 de octubre de 2012 en: [http://www.felaban.com/lvdo/cap5\\_metodos\\_usados.html](http://www.felaban.com/lvdo/cap5_metodos_usados.html))
- [3] ... Remesas Internacionales en Colombia, pp 1-86 in *Programa de Remesas*, CEMLA, 2007
- [4] Loayza N. The Economics of the Informal Sector: A simple model and some empirical evidence from Latin America, pp 129-162 en *Carnegie-Rochester Conference Series on Public Policy, Vol. 45*, Estados Unidos, 1996.
- [5] Posada, S. Rosenero, R. Comparación de modelos matemáticos: Una aplicación en la evaluación de alimentos para animales, pp 141 - 148 en *Revista colombiana de ciencias pecuarias 20*, Colombia, 2007.
- [6] Bewley, T. Advances in Econometrics-Fifth World Congress, Volume I, pp 285 en *Econometric Society Monographs*, Estados Unidos, 1987.
- [7] Shuster, M. In My Estimation pp 273-297 en *The Journal of the Astronautical Sciences, Vol 54, Num 3 y 4*, 2006
- [8] Kaur,G. Kalman Filter and its Economic Applications en *MPRA Paper*, Alemania, 2006
- [9] Harvey A. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, pp 100-169. Reino Unido, 1989.

- [10] Vélez, C. Filtro de Kalman, Material de curso Modelación Experimental, Universidad EAFIT, Colombia, 2009
- [11] Simon, D. Kalman Filtering, pp 72-79 en *Embedded Systems Programming*, 2001
- [12] Arango C, Misas M y López E. Economía Subterránea en Colombia 1976-2003: una medición a partir de la demanda de efectivo, pp 154-211 en *Revista ESPE núm 50*, Colombia, 2006.
- [13] Misas M., Vásquez D. Expectativas de la inflación en Colombia: un ejercicio econométrico, pp 1-23 en *Subgerencia de Estudios Económicos del Banco de la República*, Colombia, 2002.
- [14] Tamayo J. La Tasa Natural de Desempleo en Colombia y sus determinantes, pp 1-31 en *Borradores de Economía No.491*, Colombia, 2008.
- [15] Giles D. Measuring the hidden economy: Implications for Econometric Modelling, pp 370-380 en *The Economic Journal, vol. 109*, Canadá, 1999.
- [16] Giles D. Modelling the hidden economy and the tax-gap in New Zeland, pp 621-640 en *Empirical Economics, No.24*, Canadá, 1999.
- [17] Watson, M. Engle, R. A Time Domain Approach to Dynamic Factor Analysis and Mimic Models, en *Les Cahiers de Seminaire d'Econometrie, Núm. 22*, 1980
- [18] Feige, Edgar L. Revised Estimates of the Underground Economy: Implications of U.S. currency held abroad, en *The Underground Economy: Global Evidence of its Size and Impact*, O. Lippert and M. Walker. Vancouver, Canadá, 1997.
- [19] Colombia. Ley 365 de 1997 artículos 247, 340, de 21 de febrero de 1997, del Código Penal. [http://www.secretariasenado.gov.co/senado/basedoc/ley/1997/ley\\_0365\\_1997.html](http://www.secretariasenado.gov.co/senado/basedoc/ley/1997/ley_0365_1997.html) (Visitado el 14 de octubre de 2012).
- [20] Colombia. Ley 747 de 2002 artículo 323, de julio 19 de 2002, del Código Penal. <http://www.dmsjuridica.com/CODIGOS/LEGISLACION/LEYES/L0747002.htm> (Visitado el 14 de octubre de 2012).

- [21] Colombia. Ley 1453 de 2011 artículo 323, del 24 de junio de 2011, del Código Penal. [http://www.secretariassenado.gov.co/senado/basedoc/ley/2011/ley\\_1453\\_2011.html](http://www.secretariassenado.gov.co/senado/basedoc/ley/2011/ley_1453_2011.html) (Visitado el 14 de octubre de 2012).
- [22] Colombia. Ley 1121 de 2006 artículo 323 del 30 de diciembre de 2006, del Código Penal. [http://www.secretariassenado.gov.co/senado/basedoc/ley/2006/ley\\_1121\\_2006.html](http://www.secretariassenado.gov.co/senado/basedoc/ley/2006/ley_1121_2006.html) (Visitado el 14 de octubre de 2012).
- [23] Colombia. Ley 599 de 2000 artículo 323 del 24 de julio de 2000, del Código Penal. <http://www.dmsjuridica.com/CODIGOS/LEGISLACION/LEYES/L0599000.htm> (Visitado el 14 de octubre de 2012).
- [24] UIAF. Cartilla: Lo que debe saber sobre prevención de lavado de activos y financiación del terrorismo. Colombia, 2012 [pdf: Tomado el 14 de octubre de 2012 de <https://www.uiaf.gov.co/?idcategoria34#>]
- [25] ABC del lavado de dinero. (pdf: Visitado el 14 de octubre de 2012 en <http://www.cuentasclarasdigital.org/wp-content/uploads/ABCDELLAVADODEDINERO.pdf>)
- [26] Junguito, R. Caballero, C. La Otra Economía, en *Coyuntura Económica*, Vol. 8, Núm 4, 1978
- [27] Gómez, H. Santamaría, M. La Economía Subterránea en Colombia, pp 313-320, en *Gran Enciclopedia Temática de Colombia*, Vol. 8, Colombia, 1994
- [28] Zirnite, P. The Militarization of the Drug War in Latin America, pp 166-173 en *Current History* 97, 618, 1998.
- [29] Angrist, J. Kugler, A. Rural Windfall or a new Resource Curse? Coca, Income and Civil Conflict in Colombia, pp 191-215, en *The Review of Economics and Statistics*, Vol 90, Núm 2, Estados Unidos, 2008
- [30] United Nations. Global Illicit Drug Trends 2001, en *New York: United Nations Office for Drug Control and Crime Prevention*, 2001.
- [31] Díaz, A. Sánchez, F. Geografía de los cultivos ilícitos y conflicto armado en Colombia, en *CEDE*, Núm 18, *Universidad de los Andes*, Colombia, 2004

- [32] La Tarde. Por “Pitufeo ” 10 personas fueron enviadas a prisión [en línea] Colombia, 2011 (Visitado el 17 de enero de 2012: <http://www.latarde.com/historico/44536-por-pitufeo-10-personas-fueron-enviadas-a-prision>)
- [33] Stefoni C. Migración, remesas y desarrollo: Estado del arte de la discusión y perspectivas, en Polis[en línea], 30, 2011 (Visitado el 17 de octubre de 2012: <http://polis.revues.org/2389> ; DOI : 10.4000/polis.2389)
- [34] FOMIN. Las remesas como instrumento de desarrollo: Informe sobre las remesas de Estados Unidos a América Latina. Lima, Perú, 2004 (en línea: visitado el 14 de octubre de 2012 en <http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=557131>)
- [35] Nieto L. Colombianos migran a España: quiénes y por qué, 2006 (en línea: visitado el 14 de octubre de 2012 en <http://historico.unperiodico.unal.edu.co/Ediciones/92/06.html>)
- [36] Banco de la República. Principales resultados de la encuesta de costos de transacción de remesas de trabajadores en Colombia, pp 1-14. Colombia, 2004.
- [37] Maybeck, P. Stochastic Models, Estimation and Control Volume 1, pp 1-23 en *Mathematics in Science and Engineering Vol 141*, Estados Unidos, 1979
- [38] Cohn, S. An Introduction to Estimation Theory en *Office Series on Global Modeling and Data Assimilation 97-01*, Estados Unidos, 1997
- [39] Kelly, A. Uncertainty 3: State Space Kalman Filters en *Introduction to Mobile Robots*, 1996
- [40] Welch, G. Bishop, G. An Introduction to the Kalman Filter Course 8 en *University of North Carolina at Chapel Hill. Department of Computer Science*, 2001.
- [41] Kleeman, L. Understanding and Applying Kalman Filtering, Monash University, Clayton
- [42] Fujii, K. Extended Kalman Filter pp 3-55, en *The ACFA-Sim-J Group*

- [43] Carroll, M. Overview of Kalman Filter Theory and Navigation Applications, 2003
- [44] Sorenson, H. Least-Squares estimation: from Gauss to Kalman, pp 63-68 en *IEEE Spectrum*, Vol 7, Estados Unidos, 1970
- [45] Malla, K. Asimilación de datos en Modelos de Dispersión Atmosférica en *Tesis de Ingeniería Matemática para obtener el título de Ingeniero Matemático en la Universidad de Concepción*, Chile
- [46] Grewal, M. Andrews, A. Kalman Filtering: Theory and Practice Using Matlab, Wiley-Interscience Publication, Segunda Edición, 2001
- [47] Nerger, L. Hiller, W. Shröter, J. A comparison of error subspace Kalman filters pp 715-735 en *Tellus 57A*, Alemania, 2005
- [48] Wenzel, T. Burnham, K. Blundell, M. Williams, R. Dual Extended Kalman Filter for vehicle state and parameter estimation, pp 153-171 en *Vehicle System Dynamics*, Vol. 44, No.2, Reino Unido, 2006
- [49] Hayken, S. Kalman Filtering and Neural Networks pp 1-20, Communications Research Laboratory. McMaster University, Hamilton, Ontario, Canada, 2001
- [50] Nelson, Nonlinear Estimation and Modeling of Noisy Time-Series by Dual Kalman Filtering Methods, en *Tesis para obtener Doctorado en Filosofía en Ingeniería Electrónica y de Computación*, Estados Unidos, 2000
- [51] Nelson, L. Stear, E. The simultaneous On-Line Estimation of Parameters and State in Linear Systems, pp 94-98 en *IEEE Transactions on Automatic Control*, 1976
- [52] Wan, E. Nelson, A. Dual Kalman Filtering Methods for Nonlinear Prediction, Smoothing and Estimation, en *In Advances in Neural Information Processing Systems 9*, 1997
- [53] Kopp, R. Orford, R. Linear Regression Applied to System Identification for Adaptive Control Systems, pp 2300-2306, en *American Institute of Aeronautics and Astronautics (AIAA)*, Vol. 1, Núm 10, 1963

- [54] Cox, H. On the Estimation of State Variables and Parameters for Noisy Dynamic Systems, pp 5-12, en *Automatic Control, IEEE Transactions on, Vol. 9, Núm 1*, Estados Unidos, 1964
- [55] Ljung, L. Söderström, T. Theory and Practice of Recursive Identification, en *MIT press*, Estados Unidos, 1983
- [56] Mateljak, P. Petrovic, V. Baotic, M. Dual Kalman Estimation of Wind Turbine States and Parameters, pp 14-17 en *18th International Conference on Process Control*, Eslovaquia, 2011
- [57] Holland, J. Adaptation in Natural and Artificial Systems, en *University of Michigan Press, Ann Arbor*, 1975
- [58] Sait, S. Youssef, H. Iterative Computer Algorithms with Applications in Engineering: Solving Combinatorial Optimization Problems, en *Wiley - IEEE Computer Society Press*, 2000
- [59] Shopova, E. Vaklieva, N. BASIC - A Genetic Algorithm for Engineering Problems Solution, pp 1293-1309, en *Computers and Chemical Engineering, Vol.30*, 2006.
- [60] Kenneth, W. Aircraft Parameter Estimation, en *Technical Report NASA TM-88281*, 1987.
- [61] Vladu, E. Using Genetic Algorithms in System Identification, en *6th International Symposium of Hungarian Researchers on computational Intelligence*, 2005
- [62] Gray, G. Murray, D. Li, Y. Sharman, K. Weinbrenner, T. Nonlinear Model Structure Identification Using Genetic Programming, pp. 1341-1352, en *Control Engineering Practice, Vol.6*, Reino Unido, 1998