
UNIVERSIDAD NACIONAL DE CÓRDOBA



TÍTULO DE TESIS:

**PREDICCIÓN DE PRECIOS MEDIANTE MODELIZACIÓN
MULTIVARIADA DE SERIES DE TIEMPO. UNA APLICACIÓN AL
SECTOR LÁCTEO ARGENTINO**

PARA OPTAR EL GRADO DE:

MAGISTER EN ESTADÍSTICA APLICADA

AUTOR:

**LIC. EN ECONOMÍA
JIMENA VICENTIN MASARO**

Año 2016



PREDICCIÓN DE PRECIOS MEDIANTE MODELIZACIÓN MULTIVARIADA DE SERIES DE TIEMPO.
UNA APLICACIÓN AL SECTOR LÁCTEO ARGENTINO por JIMENA VICENTIN MASARO se distribuye
bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0
Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Comisión Asesora de Tesis

Director:

Dr. Rodrigo García Arancibia
Universidad Nacional del Litoral

Miembros:

Dr. Carlos Walter Robledo
Universidad Nacional de Córdoba

Fecha de Defensa y Aprobación de Tesis: 21 de abril de 2016.

Agradecimientos

En primer lugar, a la Ph.D Edith Depetris Guiguet y al Ph.D Gustavo Rossini quienes me han dado la oportunidad de trabajar con ellos y que, con su confianza, ayuda y paciencia me permitieron alcanzar no sólo un crecimiento profesional sino también personal.

Al director de esta tesis y compañero de trabajo Dr. Rodrigo García Arancibia. Es un referente profesional que, con su actitud proactiva, siempre me ha incentivado a ir más lejos.

Un agradecimiento muy especial a todo el personal docente y no docente de la maestría, que hacen que las personas foránea a la UNC nos sintamos como en casa.

Quiero agradecer enfáticamente a las siguientes personas que, tanto directa como indirectamente, han marcado mi evolución a lo largo de la Maestría:

A mi mamá, Sonia Marcela María, por apoyarme siempre en mis gustos y excentricidades a lo largo de toda mi vida; por permitirme proyectarme tan lejos como quisiese y enseñarme lo que es ser '*Feliz*'. Por estar al lado mío para sostenerme, cualquiera fuese mi decisión.

A mi hermanita, Rocio, compañera de aventuras desde que tengo uso de razón. Cómplice de vida, que me acompaña y alienta en todas mis ideas; con quien soñamos metas y nos apoyamos mutuamente... Muchas gracias por creer '*Siempre*' en mí, incluso más de lo que yo creo en mí misma.

A mis abuelos, Nelcy y Juan, quienes siempre contextualizaron mi existencia y que, conjuntamente con mi mamá, me inculcaron los valores del esfuerzo, dedicación, pasión y trabajo; de que la vida muchas veces no se plantea como uno espera, pero siempre hay que seguir para adelante...

A mi tía Fany y mis primos, Juan, Bauti y Balta. Con su amor e inocencia, llenan mi mundo de ternura. En ellos encuentro la incondicionalidad de la familia.

A mi amiga Laura Zingaretti que con sus conocimientos y experiencia, ha logrado guiarme a lo largo de la construcción de la tesis; sus auxilios han permitido enriquecerme con sus posturas críticas. Agradezco que en algún punto de la historia nos hayamos permitido fomentar la amistad.

A mis amigos Debora Arcucci, Roberto Ibareuren y Pablo Rossi con quienes iniciamos este proceso y, a partir del mismo, supimos construir una amistad. Hemos atravesado conjuntamente cada uno de los obstáculos de la maestría, y en ese proceso, nos hemos apoyado mutuamente para no claudicar en nuestros esfuerzos. Espero que pronto, ellos también puedan finalizarlo.

Y por último, a mi amiga Daniela. La amo en las diferencias y, aunque ella no lo sepa, ha crecido conmigo en la construcción de mi profesión.

Resumen

Palabras Clave: * Predicción * Multivariados * Cointegración * ANN *Sector Lácteo*

Realizar pronósticos sobre el comportamiento de precios es de suma relevancia en un sector productivo, ya que permite a los agentes tomar decisiones con menor incertidumbre y realizar una planificación más seria sobre sus actividades.

Sin embargo, no siempre es fácil contar con una metodología a tal fin, sobre todo en el sector agrícola ganadero por estar a sujeto múltiples imprevistos, tanto de tipo climáticos como políticos, que hacen difícil su proyección.

El sector lácteo argentino está compuesto por distintos niveles, a saber, primario, industrial y de comercialización. Estos niveles están coordinados por un sistema de precios, los cuales son objeto de constantes conflictos debido a los niveles los precios recibidos por cada uno ellos. En particular, los productores primarios son el eslabón con menor poder de negociación y con problemas para obtener información sobre dichos precios, motivos por los cuales, la presente tesis está centrada en comparar metodologías para predicciones confiables de los precios al productor primario lácteo.

Con tal objetivo, se aborda el problema de la modelización multivariada de series de tiempo para la predicción, con el propósito de comparar la capacidad predictiva dos metodologías distintas, utilizando para ello información del sector. Por un lado, una más tradicional y generalmente usada en la el área de economía, los VECM; y por otro lado, una más innovadora en esta área, como son las ANN.

En primer lugar, se modeliza con cada metodología, buscando para ello un diseño parsimonioso, y que estuviese fundamentado dentro de cada uno de los métodos. En el proceso de búsqueda del VECM se sigue la metodología de Johansen; y en el de la ANN, un diseño a priori sencillo y coherente con la teoría económica.

Luego, se realiza predicciones sobre el último año disponible a partir de dos procedimientos distintos, lográndose proyecciones de tipo estáticas y dinámicas. Finalmente se

procede a comparar la capacidad predictiva de ambos modelos (para cada tipo de pronóstico) para predecir los precios pagados a los productores primarios de leche cruda, y poder seleccionar el modelo es el más adecuado a dicho propósito.

Prices Prediction by Multivariate Time Series Modeling. An Application To Argentine's Dairy Sector.

Summary

Key Words: * Predictions * Multivariates * Cointegration* ANN * Dairy Sector *

Making predictions about prices' behavior is extremely important in a particular productive sector. This allow agents take decisions with less uncertainty and, doing a more serious planning about their activities.

However, it is not always easy found a methodology for this purpose, especially in the livestock farming sector, because is subject to multiple incidentals, both type weather and politicians ones, who make them difficult to forecast.

Argentine's dairy sector is composed of different levels, namely primary, manufacturing and marketing. These levels are coordinated by a price system, which are subject to constant conflicts due to price levels received by each of them. In particular, primary producers are the ones with less bargaining power and problems obtaining information on those prices. This are the reasons why, this thesis is focused on find out methodologies for reliable predictions of prices of dairy primary producer.

To that end, in this thesis is treated the problem of multivariate time series modeling to do predictions, in order to compare the predictive capability two different methodologies, using the sector information. That methodologies are, first one more traditional and, generally used in the area of economy, named VECM; and secondly, a more innovative in this area, such as ANN.

Firstly, it is modeled with each methodology, looking for a parsimonious design and, as the correct steps of each of the methods. In the search process of VECM is followed the Johansen's methodology; and to the ANN, an easy design and consistent with economic theory.

Then, it is made predictions about the last year available by two different procedures, called static and dynamic forecasting. Finally, is proceed to compare the predictive ability of both models to predict the prices paid to primary producers of raw milk, and to select the model is best suited to that purpose. This comparative process is made for each type of forecast.

Tabla de Contenidos

Agradecimientos	iii
Resumen.....	1
Summary	3
Lista de Tablas	7
Lista de Figuras	8
Capítulo 1	10
Introducción.....	10
1.1. Problema de Investigación	10
1.2. Objetivos	13
1.3. Hipótesis.....	14
1.4. Organización de la Tesis	14
PARTE I. Marco Teórico y Metodológico	16
Capítulo 2	17
Marco Teórico y Antecedentes	17
2.1. Introducción	17
2.2. Economía del Sector Lácteo.....	17
2.3. Modelización y Predicción de Precios en la Cadena	22
2.4. Antecedentes	25
Capítulo 3	29
Metodología.....	29
3.1. Introducción	29
3.2. Predicción Mediante el VECM	30
3.3. Predicción con ANN	47
3.4. Comparación de los Modelos.....	64
Capítulo 4	66
Datos, Variables y Procedimientos.....	66
4.1. Introducción	66
4.2. Variables	66

4.3. Fuentes de los Datos.....	69
4.4. Descripción de las Variables.....	70
4.5. Implementación y Software	72
PARTE II. Resultados.....	75
Capítulo 5	76
Predicción del Precio al Productor Primario	76
5.1. Introducción	76
5.2. Estimación del VECM	77
5.3. Estimación del ANN	83
5.4. Comparación de los Modelos.....	89
5.5. Síntesis y Conclusiones.....	92
Capítulo 6	93
Discusión, Conclusiones y Recomendaciones.....	93
6.1. Discusión.....	93
6.2. Conclusiones	94
6.3. Recomendaciones y Futuras Líneas de Investigación.....	97
Bibliografía	99
Anexo	109

Lista de Tablas

Tabla 5.1. Prueba de Estacionariedad ADF.	78
Tabla 5.2. Pruebas de Estacionariedad KPSS	78
Tabla 5.3. Tests de Cointegración de Johansen, traza y max.....	81
Tabla 5.4. Estimaciones del VECM.	82
Tabla 5.5. Pesos Estimados de la ANN.....	88
Tabla 5.6. Valores Observados y Predichos de los Modelos. Período Mensual 2012. ..	89
Tabla 5.7. Medidas de Performance Predictiva de los Modelos Estáticos.....	90
Tabla 5.8. Medidas de Performance Predictiva de los Modelos Dinámicos.....	92

Lista de Figuras

Figura 3.1. Esquema General de una ANN.....	50
Figura 4.1. Estacionalidad del Precio Pagado al Productor.	68
Figura 4.2. Precios Pagados al Productor con Fuente Nacional y Provincia de Santa Fe.	70
Figura 4.3. Precio Pagado al Productor. Período 2000-2012.....	71
Figura 4.4. Precio (índice) Mayorista y Minorista. Período 2000-2012.....	71
Figura 4.5. Cantidad (mill. Its.) de Leche Cruda Producida. Período 2000-2012.....	72
Figura 4.6. Precios (Índice) de Exportación de un Mix Lácteo. Período 2000-2012. ...	72
Figura 5.1. Correlaciones entre las Variables.	84
Figura 5.2. Errores del Proceso de Estimación de los Parámetros.....	86
Figura 5.3. Valores Ajustados versus Valores Predichos.	87
Figura 5.4. Estructura de la Red para Predecir el lp.prod.	88
Figura 5.5. Valores Observados y Predicciones Realizadas para el Precio al Productor Lechero con los Modelos Estáticos. Año 2012.	90
Figura 5.6. Valores Observados y Predicciones Realizadas por los Modelos Dinámicos para el Período 2012.	91

Capítulo 1

Introducción

1.1. Problema de Investigación

La predicción de los precios es una de las cuestiones claves en el análisis sectorial. Sin embargo, tanto el nivel de producción como los precios de los sectores agropecuarios son características altamente variables con fuertes dependencias de eventualidades, ya que están sujetos a shocks climáticos y políticos, complejizando la modelización de su comportamiento y, por consiguiente, la tarea de predecir o pronosticar su evolución futura (Jha & Sinha, 2013).

El complejo lácteo argentino no escapa a dicha complejidad y problemas de variabilidad a lo largo de su estructura. Es un sector que está integrado por niveles, a saber, la producción primaria por un lado; la industrialización por otro; la comercialización mayorista interna y externa; y por último, el consumo final (Castellano, Issaly, Iturrioz, Mateos, & Terán, 2009; Depetris, Rossini, García Arancibia, & Vicentin Masaro, 2011; Terán, 2008). El flujo de productos e insumos a lo largo de toda la cadena está básicamente coordinado por el sistema de precios, que además de ser clave para la distribución de ingresos sectoriales, constituye uno de los elementos más críticos de los incentivos en la misma, y en la determinación de su competitividad.

Sin embargo, existen constantes conflictos entre los agentes que la constituyen, fundamentalmente por los niveles de precios pagados a cada eslabón (Gutman, Guiguet, & Rebolini, 2003). Éstos devienen, principalmente, de la falta de previsibilidad y transparencia en la formación de precios del sector, generando descontentos en todos los niveles, sobre todo en aquellos agentes con menor poder de negociación. Frecuentemente ésta es la situación de los productores primarios, quienes producen un bien altamente perecedero, que necesita ser vendido con urgencia y no puede ser almacenado, y por tanto, debe aceptar condiciones de compra que no son siempre las más favorables.

Una posible solución a estas constantes tensiones podría ser la decisión política de establecer las reglas de juego claras en el sector y, a su vez, transparentar el proceso de formación de precios, donde dichas pautas surjan de la interacción y acuerdo entre los agentes de los distintos niveles. La ausencia de conflictos y el establecimiento de acuerdos tácitos-explícitos sobre la forma en que funciona la cadena, permitiría reducir la variabilidad e imprevisibilidad del comportamiento de las variables claves en la toma de decisiones, a saber, los precios.

Contar con estimaciones sobre la evolución futura de precios del sector, ayudaría en la toma de decisiones tanto de tipo productivas como de inversiones, ya que las mismas se realizarían sobre escenarios más probables. Por lo tanto, disponer de herramientas de predicción para dichos precios en Argentina, que a su vez sean confiables, sería de suma utilidad para todos los agentes de la cadena, a saber, tanto productores como industriales y comerciantes; y hasta los gobernantes obtendrían beneficios de esta información, ya que las decisiones de política estarían basadas sobre fundamentos reales y no en meras suposiciones.

En términos teóricos, no existe una sola herramienta de predicción. Dentro de las metodologías estadísticas y econométricas más tradicionales de predicción multivariada de series de tiempo, se pueden mencionar dos grandes grupos: las basadas en modelos estructurales y las específicas de series de tiempo (Claveria, Monte, & Torra, 2013; Claveria, & Torra, 2013; Fouladgar, Yazdani, Khazaei, Zavadskas, & Fouladgar, 2013; Jha, & Sinha, 2013). Los primeros enfoques son demasiado onerosos desde lo computacional y de disponibilidad de información; en cambio, los segundos son menos costosos. Dentro de estos últimos, hay una extensa lista de posibilidades, pero los más usados son Cointegración y Modelos de Corrección del Error (VECM) y Vectores Autoregresivos (Barrientos, Rodas, Velilla, Lopera, & Villada, 2012; Claveria, & Torra, 2013; McGuire, 2012). Los VECM son los más adecuados para la realización de predicciones de variables económicas ya que, generalmente las mismas no son estacionarias y esta metodología tiene en cuenta esto de modo de no generar resultados espurios (Ahumada, & Cornejo, 2014; Arias, & Torres, 2004; Basso, 2009; Engle, & Granger, 1987; Granger, 1981; Lütkepohl, 2007; Otero, & Trujillo, 1993).

Por otra parte, existe una utilización creciente de las Redes Neuronales Artificiales (ANN¹) en la modelización de series de tiempo con objetivos predictivos, dada su flexibilidad para reconocer tanto patrones lineales como no lineales (Binner, Gazaly, Chen, & Chie, 2004; Huarng, & Yu, 2006; Joekes, 2002; Maciel, & Ballini, 2010; Swanson, & White, 1997; entre otros). La ventaja de este método sobre los tradicionales es que descansa sobre hipótesis más débiles de las distribuciones subyacentes; además de poseer una capacidad de adaptación a entornos cambiantes, dándole mayor robustez a sus resultados (Binner et al., 2004; Lippmann, 1987).

Se ha encontrado diversa literatura donde se compara la performance predictiva de los modelos de serie de tiempo tradicionales con los de ANN en aplicaciones de diferentes áreas. Dependiendo de los datos utilizados, algunos son univariados (Abudu, King, & Sheng, 2012; Claveria, & Torra, 2013; Jha, & Sinha, 2013; entre otros) y otros involucran múltiples variables (Binner et al., 2004; Fouladgar et al., 2013; Maciel, & Ballini, 2010; Otero, & Trujillo, 1993; Swanson, & White, 1997; entre otros). Los resultados obtenidos por dichos estudios fueron múltiples, pero en general destacaron la mejor performance de las ANN.

Con respecto a la predicción de precios agropecuarios específicamente, también se han encontrado antecedentes de estudios comparativos de metodologías. Sin embargo, todos ellos aplican técnicas de series de tiempo univariada versus ANN (Chakraborty, Mehrotra, Mohan, & Ranka, 1992; Hamm, & Brorsen, 1997; Kohzadi, Boyd, Kermanshashi, & Kaastra, 1996; Zou, Xia, Yang, & Wang, 2007; entre otros). En estos casos univariados también se destaca la superior performance predictiva de las ANN en comparación con los modelos tradicionales.

Por lo tanto, dentro de la literatura económica existe un vacío en la comparación de metodologías multivariadas de predicción de series de tiempo. Más precisamente, no se han observado trabajos que enfrenten el desempeño de un VECM con el de una ANN en esta área; incluso siendo el VECM un método ampliamente usado en economía y a la vez que las ANN son cada vez más utilizadas por su buen desempeño en predicciones.

Por otro lado, con respecto a predicciones de precios del sector lácteo en particular sucede algo similar, con el agregado de que tampoco se observan aplicaciones de ANN,

¹ Por sus siglas en inglés *Artificial Neural Networks*

pero sí metodologías clásicas de series de tiempo, como por ejemplo, el trabajo de Fernández-Amador, Baumgartner y Crespo-Cuaresma (2010) o, el de Haden y VanTassel (1988), entre otros (Hansen, & Li, 2015; Keough, 1991; Mosheim, 2012).

La presente tesis intenta llenar estos vacíos en la literatura o al menos, realizar un aporte en dicho sentido. En particular, se propone evaluar la capacidad predictiva de modelos multivariados alternativos de series de tiempo para pronósticos de precios del sector lácteo, usando información de distintas fuentes de esta actividad agroindustrial. A parte de la contribución metodológica en el área económica, se espera aportar a la búsqueda de predicciones confiables en el sector lácteo, bajo la hipótesis central de que las ANN son competitivas para predecir la serie económica de interés, generando menores errores y mayor parsimonia que la metodología tradicional.

Además, dado que la presente tesis se realiza en el marco de la Maestría en Estadística Aplicada, se busca desarrollar un trabajo que aplique los conocimientos adquiridos en los diferentes cursos, como así también extender el conocimiento en otros tópicos no vistos en el cursado regular, como ser análisis multivariado de series de tiempo y ANN, considerando la importancia que los mismos tienen en la modelización económica actual.

1.2. Objetivos

El objetivo general del presente trabajo es evaluar la capacidad predictiva de modelos de serie de tiempo multivariados alternativos para pronósticos de precios en el sector lácteo argentino

Los objetivos específicos son:

1. Obtener predicciones de precios al productor primario lácteo argentino, a partir de la especificación de modelos multivariados tradicionales de series de tiempo.
2. Obtener la predicción de los mismos precios, pero mediante la especificación de modelos de redes neuronales.
3. Analizar el ajuste en las predicciones obtenidas por cada una de las metodologías utilizadas y, comparar dichos ajustes.

4. Realizar conclusiones sobre la capacidad predictiva de los modelos empleados que permitan guiar futuros estudios sobre predicción multivariante precios.

Para cumplimentar el Objetivo Específico 1, se modelan las series de tiempo del sector lácteo argentino bajo la metodología tradicional de VECM para obtener predicciones de los precios al productor primario lácteo. Se busca el modelo de mejor ajuste, probando la cantidad de rezagos, la existencia de cointegración y la inclusión de factores determinísticos en la relación; así como también la adecuación del modelo a los supuestos de partida, a saber: normalidad, homocedasticidad y no autocorrelación de los residuos, así como estacionariedad de los mismos.

Para el Objetivo Específico 2, se modelan las mismas series usando la metodología de ANN, para obtener predicciones de los mismos precios del Objetivo Específico 1. Se busca un modelo parsimonioso, eligiendo para ello una arquitectura adecuada, a saber, los nodos de *inputs*, *hidden* y *output*, siendo este último los precios a predecir.

Para el cumplimiento del Objetivo Específico 3, se analizan los resultados de las predicciones obtenidas por cada uno de los modelos y se comparan dichas predicciones. Se realizan dos tipos de predicciones para el período de un año, una estática y otra dinámica.

Por último, se comparan medidas de los errores cometidos en las predicciones, para concluir sobre el modelo más adecuado para predecir los precios de interés. Al realizar esta actividad se cumple con el Objetivo Específico 4.

1.3. Hipótesis

La hipótesis general del trabajo es que a pesar de que los modelos de series de tiempo, tal como el VECM, están más difundidos en el ámbito económico para predicción de precios, las ANN pueden tener ventajas en términos de poder predictivo con modelos más parsimoniosos.

1.4. Organización de la Tesis

La tesis está organizada en tres partes. La PARTE I incluye el marco teórico y metodológico que se utiliza. Está compuesta por tres capítulos: en el primero de ellos, se hace un recorrido sobre el funcionamiento de la cadena láctea argentina, así

como los fundamentos económicos y teóricos que influyen sobre las relaciones entre los precios a lo largo de una cadena productiva agrícola-ganadera. Luego, se analiza y describen los antecedentes sobre las diferentes metodologías de predicción de precios. El Capítulo 3, corresponde a la presentación de los procedimientos metodológicos que se aplican para predicción de los precios en la tesis; como también la forma de evaluación de dichos métodos. Y por último, en el Capítulo 4 se explica el tratamiento de la base de datos utilizada, las variables con las que se trabaja y los procedimientos computacionales llevados a cabo.

La PARTE II está compuesta por el Capítulo 5 y contiene los resultados empíricos de la tesis. Se expone el procedimiento utilizado para obtener los parámetros del VECM por un lado, y los de la ANN por otro. Luego se muestran las predicciones tanto estáticas como dinámicas realizadas por cada modelo, de modo de compararlas con los valores reales observados de la serie y, poder decidir cuál modelo es el de menores errores.

Por último, en el Capítulo 6 se realiza una discusión, confrontando los resultados obtenidos con otros correspondientes a antecedentes más similares. Además se muestran las conclusiones y recomendaciones arribadas.

PARTE I. MARCO TEÓRICO

Y METODOLÓGICO

Capítulo 2

Marco Teórico y Antecedentes

2.1. Introducción

Conocer la estructura de un sector productivo, así como los agentes que interactúan en la misma se convierte en una parte fundamental del proceso de análisis y modelización de un sector. Más aún si el propósito es tratar de estudiar las relaciones económicas entre ellas, y se desea predecir algún tipo de comportamiento. En este capítulo, en una primera etapa se analiza la estructura del sector lácteo y los tipos de relaciones que existen a lo largo de la cadena.

Luego, se hace una revisión de los diferentes caminos factibles de ser utilizados en la modelización de series de tiempo que permitan arribar a predicciones de las variables.

Y por último, se expone una revisión de los antecedentes encontrados que se relacionen tanto con la metodología elegida como también con el área de aplicación. Este estado del arte permite reconocer cuáles son los vacíos desde la bibliografía y qué aportes se realizan por medio de la presente tesis.

2.2. Economía del Sector Lácteo

En el entramado productivo de un país existe una gran variedad de cadenas productivas, y cada una de ellas contiene el trabajo conjunto y concomitante de agentes que interactúan obteniendo el producto final disponible al consumidor. Particularmente, los sectores productivos agrícolas-ganaderos de alimentos están compuestos por tres niveles interconectados, a saber, el nivel primario, el nivel procesador/industrial y los de distribución, ya sea mayorista o minorista (Bukeviciute, Dierx, Ilzkovitz, & Roty, 2009; McGuire, 2012).

Las relaciones entre los agentes, los flujos de información y recursos a lo largo de una cadena productiva describen el funcionamiento de la misma, afectando la forma en que

se determinan y modifican los precios que la componen (Heien, 1977; LaFrance, & de Goter, 1985; Mosheim, 2012). Por tanto, la descripción de cómo funciona una cadena ayuda a entender el proceso de formación de los precios a lo largo de la misma, la manera en la que intervienen los costos, cuáles son las interacciones entre los agentes y dónde pueden impactar las regulaciones y políticas gubernamentales (McGuire, 2012).

Particularmente, el complejo lácteo argentino está compuesto por la producción primaria, que aporta el principal insumo de la cadena; la industrialización, que elabora los diferentes tipos de productos lácteos; y la comercialización, que incluye la venta interna y externa mayorista como así también, al consumo final (Castellano et. al, 2009). Todos estos interactuando concomitantemente. Depetris, Rossini, García Arancibia y Vicentin Masaro (2011) describen al complejo lácteo argentino como un conjunto de eslabones verticales donde el flujo de productos e insumos a lo largo del mismo está básicamente coordinado por el sistema de precios. Sin embargo, cada eslabón o nivel representa, por sí mismo, una interacción de oferta y demanda.

En general, existen distintos factores determinantes de la oferta y demanda, los cuales a su vez dependen del mercado en particular. Siguiendo al *mainstream*, se pueden mencionar como determinantes generales de la demanda al precio del bien en particular y el de sus sustitutos, los gustos y preferencias, el ingreso disponible, el tamaño y distribución de la población, expectativas o demanda rezagada, demanda derivada, entre otros. Por otra parte, los determinantes de la oferta son: el precio esperado del bien, el precio de los factores, la tecnología disponible, factores aleatorios como los climáticos, factores institucionales y políticos, entre otros.

La teoría neoclásica supone el vaciamiento de los mercados, es decir, que las cantidades demandadas y ofrecidas deben estar equilibradas en el largo plazo y por lo tanto, no deben haber excedentes. Sin embargo, en el corto plazo esta condición no se da estrictamente, sobre todo en bienes atados a ciclos biológicos. En general, la oferta no es lo suficientemente elástica y los bienes deben venderse, incluso afectándose los precios, actuando éstos como instrumentos para racionalizar la demanda (Nicholson, 2005).

Utilizando la explicación de Heien (1977) sobre cómo se determinan los precios en los sectores agrícolas-ganaderos, y trasladándola al sector lácteo específicamente, se puede describir el funcionamiento de éste teniendo en cuenta la estructura de cadena así como

los procesos de corto y largo plazo. En el equilibrio, Heien afirma que la cantidad ofrecida y demandada en cada nivel del sector se igualan, indicando los factores que afectan tanto a oferta como demanda para cada nivel. Es así como, en el nivel de producción primaria, los principales factores que afectan a la oferta de leche cruda son: la cantidad de vacas en ordeño y su productividad. El primero, a su vez, depende de precios esperados de dicha leche cruda ($p_{\text{prod}_t}^e$) y de los costos (c_{prim_t}); y el segundo, de la tecnología disponible ($\text{tec}_{\text{prim}_t}$) así como también de la estación climática del año (clima_t). La demanda de leche cruda, por otro lado, está determinada por el precio esperado de dicho bien, así como también de los precios esperados de los bienes producidos con dicha materia prima, tanto los destinados al mercado interno ($p_{\text{may}_t}^e$) como externo ($p_{\text{exp}_t}^e$). En este caso la industria procesadora actúa como demandante. Por tanto, en el nivel primario las cantidades ofrecidas y demandadas que se equilibran en el largo plazo pueden ser resumidas de la siguiente forma,

$$Q_{\text{prim}_t}^o = f_{op} \left(p_{\text{prod}_t}^e, c_{\text{prim}_t}, \text{tec}_{\text{prim}_t}, \text{clima}_t, z_t^{op} \right), \quad (2.1)$$

$$Q_{\text{prim}_t}^d = f_{dp} \left(p_{\text{prod}_t}^e, p_{\text{may}_t}^e, p_{\text{exp}_t}^e, z_t^{dp} \right), \quad (2.2)$$

donde z_t^{op} y z_t^{dp} son otros factores determinantes, para la oferta y la demanda, respectivamente.

Por otro lado, Heien indica que la oferta de bienes industriales depende de la tecnología empleada ($\text{tec}_{\text{ind}_t}$), de los precios de los bienes producidos, ya sea en el mercado interno (mayorista-minorista) o externo; así como de los costos de producción (c_{ind_t}). Aquí los industriales asumen el rol de oferentes de productos. Por el lado de la demanda, están tanto los demandantes externos como los mayoristas internos; y las variables que afectan a esta demanda son: el precio de los productos mayoristas y el precio de importación. Resumiendo, en el nivel industrial, la oferta y la demanda en el largo plazo son:

$$Q_{\text{ind}_t}^o = f_{oi} \left(p_{\text{may}_t}^e, p_{\text{exp}_t}^e, \text{tec}_{\text{ind}_t}, c_{\text{ind}_t}, z_t^{oi} \right), \quad (2.3)$$

$$Q_{\text{ind}_t}^d = f_{di} \left(p_{\text{may}_t}^e, p_{\text{bs_imp}_t}^e, z_t^{di} \right), \quad (2.4)$$

con z_t^{oi} y z_t^{di} como otros factores de esta oferta y demanda, respectivamente.

En el proceso de comercialización se puede distinguir dos niveles, el interno y el externo. En el interno, la oferta está determinada principalmente por los precios minoristas esperados ($p_{\min_t}^e$), la oferta rezagada ($Q_{\text{com_int}_{t-1}}^o$), los costos de comercialización ($c_{\text{com_int}_t}$), y otros factores (z_t^{oci}). La demanda, por los precios minoristas, los ingresos disponibles (YD_t), el precio de bienes sustitutos (p_{BS_t}), gustos y preferencias (GP_t), y otros factores (z_t^{dci}). En el nivel externo, la oferta depende de los precios internacionales recibidos por los productos (p_{exp_t}), los aranceles de exportación (tax_{ext_t}), la tasa de cambio (tc_t) y otros (z_t^{oce}). La demanda, por otro lado, del precio de exportación (p_{exp_t}), la variación del PBI del país (ΔY_t), el precio de exportación de los competidores ($p_{\text{exp}_t}^{\text{otros}}$), el tipo de cambio y otros (z_t^{dce}). Resumiendo, las cantidades ofrecidas y demandadas del nivel de comercialización, interna y externa, determinadas conjuntamente son:

$$Q_{\text{com_int}_t}^o = f_{oci}(p_{\min_t}^e, Q_{\text{com_int}_{t-1}}^o, c_{\text{com_int}_t}, z_t^{oci}), \quad (2.5)$$

$$Q_{\text{com_int}_t}^d = f_{dci}(p_{\min_t}, YD_t, p_{BS_t}, GP_t, z_t^{dci}), \quad (2.6)$$

$$Q_{\text{com_ext}_t}^o = f_{oce}(p_{\text{exp}_t}, tax_{\text{ext}_t}, tc_t, z_t^{oce}), \quad (2.7)$$

$$Q_{\text{com_ext}_t}^d = f_{dce}(p_{\text{exp}_t}, \Delta Y_t, p_{\text{exp}_t}^{\text{otros}}, tc_t, z_t^{dce}). \quad (2.8)$$

Todos estos niveles están interactuando de manera concomitante. Ahora, se expresan las funciones de oferta y demanda de los distintos niveles de manera inversa en función a los distintos precios de los niveles de la cadena, y considerando al resto de las variables como dadas (o fijas)². Entonces se puede observar que la determinación de precios es endógena dentro del sistema.

² Sea f invertible, entonces si $f : \mathbb{R}^n \rightarrow \mathbb{R} \Rightarrow f^{-1} : \mathbb{R} \rightarrow \mathbb{R}^n$. Para que, siendo $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sea $f^{-1} : \mathbb{R} \rightarrow \mathbb{R}$ las $n-1$ variables restantes deben considerarse fijas.

En el nivel primario, las funciones (2.1) y (2.2) expresadas de manera inversa y dados los niveles de todas las variables excepto las relacionadas con cantidad son:

$$p_{\text{prod}_t} = f_{op}^{-1} \left(Q_{\text{prim}_t}^o, c_{\text{prim}_t}, \text{tec}_{\text{prim}_t}, \text{clima}_t, z_t^{op} \right), \quad (2.1')$$

$$p_{\text{prod}_t} = f_{dp}^{-1} \left(Q_{\text{prim}_t}^d, p_{\text{may}_t}^e, p_{\text{exp}_t}^e, z_t^{dp} \right). \quad (2.2')$$

En el nivel industrial: teniendo en cuenta que los precios de externos no son una variable de decisión interna, y fijando el resto de las variables (excepto cantidades) las funciones de demanda y oferta inversas son:

$$p_{\text{may}_t} = f_{oi}^{-1} \left(Q_{\text{ind}_t}^o, p_{\text{exp}_t}^e, \text{tec}_{\text{ind}_t}, c_{\text{ind}_t}, z_t^{oi} \right), \quad (2.3')$$

$$p_{\text{may}_t} = f_{di}^{-1} \left(Q_{\text{ind}_t}^d, p_{\text{bs_imp}_t}^e, z_t^{di} \right). \quad (2.4')$$

Y por último, del nivel de comercialización, y considerando que los agentes de un país pequeño no influyen sobre los precios externos (los precios externos son exógenos y constantes), las funciones de oferta y demanda inversas (dados los niveles de todas las variables restantes, menos las cantidades de oferta y demanda) correspondientes a las ecuaciones (2.5) y (2.6):

$$p_{\text{min}_t} = f_{oci}^{-1} \left(Q_{\text{com_int}_t}^o, Q_{\text{com_int}_{t-1}}^o, c_{\text{com_int}_t}, z_t^{oci} \right), \quad (2.5')$$

$$p_{\text{min}_t} = f_{dci}^{-1} \left(Q_{\text{com_int}_t}^d, YD_t, p_{\text{BS}_t}, GP_t, z_t^{dci} \right). \quad (2.6')$$

Heien también aclara que los efectos descritos corresponden a situaciones de equilibrio o relaciones de largo plazo. En el corto plazo, puede suceder que la demanda minorista de lácteos se incremente por un aumento en el ingreso de los consumidores, lo que se traslada a la industria que, a su vez, demandará mayor cantidad de leche al sector primario para responder a dicho desequilibrio. Como la producción de leche cruda es inelástica en el corto plazo y no puede responder en forma inmediata, pasará un cierto tiempo hasta que la oferta pueda aumentar su producción y adecuarse a la nueva demanda. Mientras tanto, la industria habrá reducido/agotado los stocks, y ante la escasez relativa de leche, los precios al productor aumentarán. Ese aumento será trasladado al minorista, lo que enfriará la demanda, y ese aletargamiento o disminución, recomenzará un nuevo

ciclo en la cadena. Por lo tanto, en el corto plazo se producirán desequilibrios entre la oferta y demanda, produciendo fluctuaciones en los precios resultantes, mientras que en el mediano/largo plazo deberán equilibrarse. En el corto plazo existen efectos de las cantidades sobre los precios a lo largo de la cadena, lo que Heien llama «Noción de la Dependencia del Precio».

Deaton y Laroque (1992; 2003) se han pronunciado sobre las relaciones o efectos de corto y largo plazo en precios. Según estos autores, la dinámica de corto plazo de los precios está fuertemente guiada por los niveles de inventarios mantenidos; en cambio, en el largo plazo, están focalizados en las relaciones de equilibrio entre la oferta y la demanda, y por consiguiente, por los factores que los determinan. Por lo tanto, afirman que los precios de estos tipos de bienes están simultáneamente determinados por los factores de la oferta y la demanda en un contexto de variaciones de los stocks.

En un estudio realizado por la Oficina Gubernamental Contable de los Estados Unidos (GAO) en 2005, se describen los factores que empíricamente afectan a los precios en los diferentes niveles del sistema productivo lácteo estadounidense. Bajo un sistema de producción con fuerte presencia de cooperativas procesadoras, afirman que los precios pagados a los productores están afectados fuertemente por la cantidad ofrecida de leche cruda, como así también por la demanda de productos lácteos. En el nivel cooperativo, los precios están influidos por los costos de los servicios que brindan y el poder de negociación que tienen con respecto a los procesadores de leche. Por último, en los niveles de comercialización mayoristas y minoristas, los precios están afectados por los costos internos, laborales y de energía, y por la consolidación que tengan las empresas en el mercado (GAO, 2005). Además afirman que los precios están correlacionados a lo largo de los distintos niveles del sector, con mayores correlaciones entre precios adyacentes.

2.3. Modelización y Predicción de Precios en la Cadena

Predecir precios provenientes de un sector agrícola-ganadero no es lo mismo que hacerlo sobre cualquier otro bien, dado que tienen características especiales, como la presencia de estacionalidades, la naturaleza derivada de la demanda, estructura productiva que devienen de procesos biológicos, entre otros.

La cantidad de metodologías disponibles para modelar estos tipos de datos es extensa. Fouladgar et al. (2013) afirman que hay tres tipos de métodos para la predicción: (a)

Subjetivos, es decir, basado en intuición; (b) Univariados, cuando la predicción del precio depende de su valor presente y de los pasados; y (c) Multivariados, cuando se desea predecir en función de otras variables, las cuales pueden estar en valor presente o pasado. Los últimos dos métodos tienen sus aproximaciones mediante modelos de series de tiempo, siendo los Vectores Autoregresivos (VAR) y los VECM los más usados en series de tiempo para predicciones multivariadas.

Jha y Sinha (2013) destacan dos enfoques para predicciones de los precios específicamente, a saber, los modelos estructurales y los de series de tiempo. La utilización de uno u otro depende de distintas cuestiones, pero principalmente del objetivo del trabajo y de la disponibilidad de información.

Los modelos estructurales se sustentan en los principios de la teoría del consumidor y productor para identificar la oferta y la demanda, así como los factores que las determinan y por tanto, los precios por los cuales se equilibran dichas cantidades. Este tipo de estudios empíricos permite conocer el funcionamiento de la cadena, probar la implementación de políticas o estudiar las relaciones existentes a lo largo de la misma (e.g. Liu, Kaiser, Mount, & Forker, 1991; Mosheim, 2012; Novakovic, & Thompson, 1977). Sin embargo, se requiere una alta cantidad de información debido a su especificación más realista, la cual no siempre está disponible. Una forma de superar este problema es la realización de estimaciones por niveles de la cadena, obteniendo así resultados parciales.

Se pueden usar técnicas econométricas de series de tiempo que no descansan sobre diseños y supuestos tan estrictos como lo tienen los modelos estructurales, pero que tienen una buena performance hacia la predicción y demandan una menor cantidad de datos (Jha, & Sinha, 2013). Claveria y Torra (2013) señalan que los modelos de series de tiempo multivariados más usados en la literatura, dejando de lado los estructurales, son Cointegración-VECM y VAR. Con series de tiempo no estacionarias, la modelización para predicción mediante VECM resulta más adecuada para no tener problemas de estimación de relaciones espurias (Engle, & Granger, 1987; Granger, 1981; entre otros).

En la actualidad, los modelos de ANN han tomado creciente importancia, dada su flexibilidad para reconocer tanto patrones lineales como no lineales en las relaciones entre variables (e.g. Claveria, & Torra, 2013; Maciel, & Ballini, 2010). Su funcionalidad es

tanto la clasificación como la predicción, motivo por el cual es considerada una forma de modelización factible en el presente trabajo. Lippmann (1987) y otros autores (e.g. Binner et al., 2004) sugieren que las ventajas de este método sobre los tradicionales surgen del hecho de tratar de forma más robusta las distribuciones generadas por procesos no lineales y no gaussianos, ya que descansan sobre hipótesis más débiles de las distribuciones subyacentes; además de poseer una capacidad de adaptación en entornos cambiantes, lo que le da mayor robustez a sus resultados.

Las ANN se han aplicado en diferentes campos, como la biología, la ingeniería y la física, entre otros (Kuan, & White, 1994). Debido al aporte de White (1988) también se ha incorporado a la economía y finanzas. Sin embargo, para las ANN no es fácil introducirse en esta área. Según Kuan y Liu (1995) se debe a que es difícil para los economistas aplicados entender cómo trabajan las ANN y cómo implementarlas apropiadamente. Según Arangón-Torres, Calzada-Arroyo, García-Güemes y Pacheco-Bonrostro (1997), el principal uso dentro de la economía es la predicción.

Los modelos multivariados aplicables a predicción de precios que se han seleccionado en la presente tesis son: los VECM y las ANN. El primero de ellos por ser el más utilizado en modelización multivariada clásica para series de tiempo que no son estacionarias; el segundo, por ser un modelo que se está incorporando en la modelización económica, y más particularmente, en la predicción de precios, pero que no se han encontrado trabajos que utilicen precios del sector lácteo, ni a nivel internacional ni en Argentina.

La principal diferencia entre una u otra metodología radica en el supuesto de linealidad sobre la relación entre valor presente de una o más variables respuestas y los valores pasados (Fouladgar et al., 2013). Los VECM parten de dicho supuesto de linealidad, en cambio las ANN, en esencia, permiten la no linealidad de las relaciones. La característica de permitir relaciones no lineales resulta una ventaja del modelo ANN en aplicaciones económicas, pues en el mundo real, las variables no exhiben regularidades simples (Chakraborty et al., 1992). Por otro lado, y dependiendo del diseño de la red neuronal, suelen tener un buen ajuste y desempeño, con estructuras bastante sencillas (Kohzadi et al., 1996).

2.4. Antecedentes

En predicciones de precios agrícolas-ganaderos se han encontrado trabajos que utilizan diferentes tipos de modelizaciones. Aunque la más común corresponde a métodos de series de tiempo univariados, también se han observado trabajos donde utilizan ecuaciones estructurales bajo diferentes métodos de estimación; así como modelos de series de tiempo multivariados, como ser, VAR y VECM.

En general, los modelos de equilibrio parcial con ecuaciones estructurales son ampliamente usados para el análisis de sectores productivos agrícolas-ganaderos (Heien, 1977). En particular, para el sector lácteo (Liu, Sun, & Kaiser, 1995; Novakovic, & Thompson, 1977; Schmit, & Kaiser, 2006; entre otros). Estos trabajos intentan imitar la estructura de dicho sector, incorporando las distintas formas de comercialización. Algunos obtienen resultados modelando para el conjunto de productos lácteos (LaFrance, & de Gorter, 1985); otros separan por grupos de productos: Kaiser, Streeter y Liu (1988) lo hacen para productos fluidos por una parte, y productos manufacturados, por otra; mientras que Novakovic y Thompson (1977) utilizaron sub-modelos para la leche fluida, quesos, manteca, leche en polvo descremada y postres helados. Este tipo de modelos son usados principalmente por organismos encargados de predecir la evolución sectorial, con el objetivo de evaluar los efectos de modificaciones en el sistema impositivo o de comercio exterior. Así, fue incorporado en los actuales modelos de pronósticos de largo plazo que ofrece el FAPRI³ y más recientemente los modelos de Bailey de Penn State University (2003; 2004; 2006; 2009).

Mosheim (2012), investigador del *Economic Research Service* del Departamento de Agricultura de Estados Unidos (USDA⁴), analiza la capacidad predictiva de diferentes modelos aplicados al sector lácteo de Estados Unidos. Mosheim utiliza seis modelos para predicción de distintas variables del sector. Cuatro de ellos son ecuaciones estructurales utilizando las siguientes metodologías de estimación: Mínimos Cuadrados Ordinarios (OLS⁵), Mínimos Cuadrados en Dos Etapas (2SLS⁶), Mínimos Cuadrados en Tres Etapas (3SLS), y Regresión con Ecuaciones Aparentemente No Relacionadas

³ Por sus siglas en inglés *Food and Agricultural Policy Research Institute*.

⁴ Por sus siglas en inglés *United States Department of Agricultural*

⁵ Por sus siglas en inglés *Ordinary Least Squares*

⁶ Por sus siglas en inglés *Two Step Least Squares*

(SURE⁷). El resto de los modelos, corresponden a sistemas de ecuaciones con variables endógenas autoregresivas, restringidas y no restringidas, ambas con variables exógenas, a saber, SVARX y VARX, respectivamente. Usando variables trimestrales referentes a cada nivel del sector, desde el cuarto trimestre de 1998 al primero de 2009, realiza predicciones sobre variables del sector lácteo, incluyendo precios de la leche cruda. Esta variable ha sido pronosticada *ex-post*⁸ hasta el primer trimestre de 2010. Mosheim concluye que el modelo de VARX tiende a predecir mejor que los otros modelos tanto los precios como el número de vacas en ordeño; sin embargo, no es mejor para predecir la cantidad producida de leche cruda ni stocks del sector.

Por otra parte, la utilización de VECM también es usada para predecir variables en el área de economía, incluso por organismos oficiales públicos (Juseluis, 2008; Maddala, & Kim, 1998). Esta metodología permite analizar la relación entre variables bajo la simultaneidad de las mismas, con la característica distintiva de que las mismas no sean estacionarias (Arias, & Torres, 2004). Particularmente, hay estudios empíricos focalizados en predecir precios de bienes agrícolas-ganaderos teniendo en cuenta el movimiento correlacionado con otras variables como ser, otros precios o índices de precios, como también cantidades, bajo la hipótesis de que estas incorporaciones mejoran el poder predictivo (Ahumada, & Cornejo, 2014; Fanchon, & Wendell, 1992; Zapata, & García, 1990; entre otros). Por ejemplo, con el objetivo de predecir los precios de tres diferentes *commodities* agrícolas (maíz, soja y trigo), Ahumada y Cornejo (2014) han modelado dichos precios de dos formas diferentes. Por un lado, diseñaron un modelo individualizado para cada precio de las *commodities*, y por otro, un VECM, donde los tres precios son considerados endógenos. Tuvieron en cuenta otras variables como la cantidad producida y los *stocks* de dichos bienes, algunas variables macroeconómicas, y componentes determinísticos. Las variables fueron trabajadas en logaritmo e hicieron predicciones para datos dejados fuera de la modelización (*ex-post*) a partir de una muestra fija por un lado (*predicciones estáticas*), y por otro, con la incorporación recursiva de datos (*predicciones dinámicas*). Evaluaron la performance predictiva de los distintos modelos y concluyeron que el más adecuado es el que tiene en cuenta las interacciones de todos los precios, es decir, el VECM.

⁷ Por sus siglas en inglés *Seemingly Unrelated Regressions Equations*

⁸ Es decir, se han dejado fuera de la estimación un conjunto de datos y luego, esos mismos, han sido sujeto de predicción.

Del mismo modo, para precios ganaderos se han encontrado dos trabajos que comparan metodologías de predicción bajo la presencia de no estacionariedad en las series, a saber, el de Zapata y Garcia (1990) y, el de Fanchon y Wendel (1992). En ambos, se utilizaron tanto modelos univariados como multivariados, llegando a la conclusión de que los segundos producen resultados superiores en la predicción *ex-post*. De los modelos multivariados (VAR y VECM), el VECM mostró mejores resultados predictivos de los precios.

Por otro lado, las ANN constituyen una metodología relativamente nueva dentro de la modelización económica, por lo que se han hecho esfuerzos para comenzar a tenerla presente para realizar predicciones (Jha, & Sinha, 2013; Kuan, & White, 1994). Casi todos los trabajos que se han encontrado para predicciones de precios agrícola-ganaderos han evaluado la capacidad predictiva de las ANN univariadas en contraposición a modelos univariados de series de tiempo, a saber, Autorregresivos Integrado de Medias Móviles (ARIMA) o Autorregresivos de Medias Móviles (ARMA). No se han encontrado trabajos en esta área de aplicación que compare ANN con modelos multivariados de series de tiempo para predicción. En general, las conclusiones obtenidas fueron variadas, pero destacan la performance de las ANN para realizar predicciones, resaltando que con modelos simples se consiguen resultados bastante robustos en términos de predicciones.

Por ejemplo, Li, Xu y Li (2010) evaluaron la capacidad predictiva de una ANN versus las predicciones realizadas con un ARIMA de los precios del tomate en China a corto plazo. Observaron que las predicciones realizadas con el ANN resultaron ser las más cercanas a los valores reales de los precios. Del mismo modo, en otro trabajo se evalúan los mismos modelos, pero aplicados a la predicción de precios mensuales del trigo en el mercado Chino (Zou et al., 2007); llegando a la misma conclusión de que el mejor modelo para predecir estos precios es el ANN estimado.

Hay otros trabajos que combinan ANN y ARIMA univariadas llegando a un modelo híbrido y, comparan la capacidad predictiva de éste con respecto a la de los modelos individuales (Jha, & Sinha, 2013; Shahwan, & Odening, 2007). Éstos concluyen que la metodología híbrida es la más conveniente para predecir precios agrícolas, dado que estos precios se comportan de manera muy compleja en la realidad, y los modelos individuales no alcanzan a capturar distintos patrones que se suceden en la misma.

Se ha encontrado sólo un trabajo de predicción de precios de un producto agroindustrial con un diseño de ANN multivariada (Chakraborty et al., 1992). Sin embargo, su capacidad predictiva es también comparada con la de un modelo univariado de serie de tiempo, a saber, un ARMA. El trabajo está aplicado a la predicción de precios de la harina en tres ciudades de Estados Unidos distintas, a saber, Buffalo, Minneapolis y Kansas. Los resultados obtenidos indican que la ANN predice mejor que el modelo ARMA; aunque proponen como futuras líneas explorar la posibilidad de combinar ambas metodologías.

Otra área de aplicación destacable corresponde a la de precios de productos energéticos. En la predicción de éstos se han encontrado trabajos que comparan los pronósticos de ANN con modelos de series de tiempo multivariados, más precisamente VECM (Barrientos et al., 2012; Fouladgar et al., 2013). Fouladgar et al. (2013) modela los precios del barril de petróleo (*WTI Oil*) llegando a la conclusión de que la ANN resulta la metodología más adecuada para predecirlos, ya que ante la evaluación *ex-post*, produce los menores errores en comparación con el VAR y el VECM. En cambio, Barrientos et al. (2012) no tuvieron una postura concluyente sobre qué metodología es mejor para predecir los precios de la energía eléctrica en Colombia, ya que utilizaron diferentes supuestos en los VECM y ANN, y sus conclusiones dependían de los mismos.

Por lo tanto, ante este estado de arte, la presente tesis constituye una contribución en el área de predicción de precios en general, a pesar de estar focalizado en el sector lácteo en particular.

Capítulo 3

Metodología

3.1. Introducción

Como el objetivo es evaluar predicciones de serie de tiempo con diferentes metodologías, a continuación se detalla la formulación teórica de cada una de las mismas, a saber, del VECM - Cointegración y ANN. Tal como se ha visto en la Sección 2.4, ambas son utilizadas para realizar predicciones, la primera de ellas es más comúnmente usada en el área de economía, mientras que la segunda, en áreas de ingeniería.

Aún cuando el interés predictivo sea una sola variable, como dichas metodologías permiten la utilización conjunta de distintas series de tiempo, ambos pertenecen a modelos multivariados. La primera de ellas es generalmente utilizada en el contexto de datos de series de tiempo o de panel; no así la segunda, ya que es factible de aplicarla en otros tipos de datos. Sin embargo, como se menciona anteriormente, se modela solamente datos de serie de tiempo en el presente caso.

La característica primordial del VECM radica en la modelización conjunta de las variables dependientes involucradas, bajo el supuesto de que las mismas son endógenas y que, a su vez, guardan una relación lineal con los valores pasados de las mismas. Dependiendo del diseño, los VECM no restringidos estiman un gran número de parámetros, lo que puede ser problemático con relación a la disponibilidad de datos de series de tiempo.

Por otro lado, las ANN no realizan suposiciones sobre relaciones endógenas. El diseño de la red neuronal depende del conocimiento previo que se tenga del funcionamiento de lo que se está modelando. En general, las ANN con forma de embudo suelen llamarse parsimoniosas; la cantidad de variables de ingreso (o independientes) suelen ser mayores que la cantidad de variables de salida; y los nodos ocultos, que permiten la especificación no lineal, se acoplan a dicho formato parsimonioso. La cantidad de parámetros a

estimar en estos modelos suele crecer rápidamente a medida que se le incorporan elementos al diseño, sobre todo en las capas ocultas. En general, el ajuste intra-modelo suele ser bueno; pero si existe una sobreparametrización, las predicciones por fuera de la muestra no siempre lo son.

Dado que se quiere evaluar cuál metodología es la más adecuada para realizar predicciones de precios en el sector lácteo, más precisamente en los precios primarios, se comparan dichas metodologías a partir de medidas de error sobre las predicciones de cada modelo. Luego de describir cada una de las metodologías y cómo se predice a partir de ellas, se muestran las medidas de performance que se utilizan para compararlas.

Por tanto, el presente capítulo se organiza en tres bloques. Un primer bloque donde se explica el procedimiento teórico a tener en cuenta para lograr predicciones por medio de un VECM. El segundo, el correspondiente para realizar predicciones por medio de la metodología de ANN; y el tercero, se presentan las medidas de performance tenidas en cuenta para comparar las predicciones realizadas con los distintos modelos, y poder distinguir cuál tiene mejor performance en términos del poder de predicción.

3.2. Predicción mediante el VECM

Los objetivos de modelar múltiples series de tiempo pueden ser varios, el más importante es realizar predicciones con el objetivo de anticipar el futuro. El hecho de tener en cuenta más de una variable de serie de tiempo indica que, en principio, hay interés en conocer las posibles interacciones que existen entre dichas variables, suponiéndolas endógenas.

El primer paso lógico en la modelación de múltiples series de tiempo es la utilización de la metodología VAR, que es la versión multivariada de los modelos autorregresivos (AR). Pero los VAR generan resultados consistentes y útiles, siempre y cuando las variables involucradas sean estacionarias. Sin embargo, la mayoría de las series económicas en la práctica no lo son. En general, están dominadas por una tendencia que son la principal causa de sus valores futuros. En algunos casos, un grupo de series de tiempo no estacionarias obedece a una o más relaciones de largo plazo y sus desviaciones de esa/s relación/es pueden ser estacionarias. Entonces en esos casos, se dice que las variables están cointegradas y se utiliza el VECM para ajustar dichas relaciones, y poder predecir los valores de las variables involucradas.

Por tanto, para predecir a partir de un VECM, en un primer lugar se constata el orden de integración de las series de tiempo involucradas. Existen varios test que permiten responder sobre el orden de integración de las series univariadas; sin embargo, en el presente análisis se utilizan dos de ellos, el Dickey Fuller Aumentado (Fuller, 1976; Dickey, & Fuller, 1979, 1981) y el *KPSS* (Kwiatkowski, Phillips, Schmidt, & Shin, 1992).

Una vez que se corrobora la presencia de raíces unitarias, se debe probar la existencia de cointegración entre las variables. Enders (1995) expone dos técnicas para realizar pruebas de cointegración, a saber, la metodología empleada por Engle y Granger (1987) por un lado, y la metodología de Johansen (1988) y Stock-Watson (1988), por otro. Basso (2009) resume estos trabajos diciendo que la primera de ellas asume la existencia de cointegración entre variables si, estimada la relación de largo plazo, los residuos generados son estacionarios; en cambio, la metodología de Johansen analiza el rango de la matriz que contiene los coeficientes de la relación de largo plazo. En el presente trabajo se utiliza esta última postura, dado que resulta la más apropiada cuando se analiza la relación entre más de dos variables (Juseluis, 2008).

Sin embargo, para emplear esta metodología se debe tener conocimiento de la forma que tiene la relación entre las variables, es decir, la cantidad de rezagos a incluir en el modelo y la presencia (o no) de componentes determinísticos en la relación de largo plazo (tendencia lineal y/o cuadrática). Estas cuestiones se definen en un primer paso antes de realizar el test de cointegración de Johansen.

Por último, probada la presencia de cointegración, se procede a estimar efectivamente el VECM, y su equivalente VAR para realizar predicciones de la variable de interés.

3.2.1 Orden de Integración de las Series

Test de Dickey Fuller Aumentado

Este test permite constatar si una serie de tiempo univariada (x_t) es estacionaria. Se asume que dicha serie responde a un proceso autorregresivo de orden k ⁹. Dependiendo de los componentes determinísticos que se incluyan en la serie univariada, la diferencia de primer orden puede ser especificada de las siguientes maneras:

⁹ En realidad la prueba fue desarrollada y demostrada para una serie autorregresiva de orden uno, y luego extendida a procesos de órdenes superiores.

$$\Delta x_t = \beta_1 + \beta_2 t + \pi x_{t-1} + \sum_{j=1}^k \gamma_j \Delta x_{t-j} + u_{1t}, \quad (3.1)$$

$$\Delta x_t = \beta_1 + \pi x_{t-1} + \sum_{j=1}^k \gamma_j \Delta x_{t-j} + u_{2t}, \quad (3.2)$$

$$\Delta x_t = \pi x_{t-1} + \sum_{j=1}^k \gamma_j \Delta x_{t-j} + u_{3t}, \quad (3.3)$$

donde u_{it} corresponde al término de perturbación del modelo en diferencia, el cual se supone $N(0, \sigma^2)$; Δ indica la primera diferencia, a saber $\Delta x_t = x_t - x_{t-1}$; y $\pi = \rho_1 - 1$, donde ρ_1 es el coeficiente del término x_{t-1} en el proceso autorregresivo que explica x_t .

Las ecuaciones planteadas en (3.1), (3.2) y (3.3) están basadas en tres especificaciones distintas del comportamiento de la variable, donde en todos los casos se tienen en cuenta rezagos en diferencias (que devienen de los k -ésimos órdenes de la autorregresión). La ecuación (3.1) tiene en cuenta tanto una constante como un tendencia en la relación; en la (3.2) sólo se incluye una constante; y por último, la (3.3) no tiene componentes determinísticos.

La prueba consiste en corroborar las siguientes hipótesis, H_0 : existe raíz unitaria o lo que es lo mismo, la serie no es estacionaria, condición que se cumple si $\pi = 0$ o $\rho_1 = 1$; versus H_1 : existe raíz unitaria o lo que es lo mismo, la serie es estacionaria.

El estadístico utilizado para comparar las hipótesis (τ), es construido de manera homóloga al caso de t -Student, sin embargo no sigue tal distribución. A su vez, es distinto en función a la especificación del modelo, es decir, dependiendo de la presencia de componentes determinísticos.

Dickey y Fuller (1979, 1981) han analizado la distribución asintótica de los diferentes estadísticos mediante simulaciones, obteniendo la distribución aproximada de los mismos para una variable que responde a un proceso autorregresivo de orden 1. A partir de allí se han tabulado valores críticos de los estadísticos para constatar las hipótesis plan-

teadas (raíz unitaria) según su forma funcional, a saber, τ_t, τ_μ, τ , en función de si posee constante y tendencia determinística, tiene sólo constante determinística o ninguna de las dos, respectivamente. Estos autores demostraron que las distribuciones encontradas son extensibles a los casos de procesos autorregresivos de órdenes superiores, como los expuestos en las ecuaciones (3.1) a (3.3). Además de calcular los estadísticos para probar raíces unitarias y sus correspondientes distribuciones asintóticas, también hicieron lo mismo para probar la significancia de los componentes determinísticos.

Siguiendo el procedimiento de Dickey y Fuller (1979, 1981), el paquete *urca* (Pfaff, & Stiggler, 2013a) del programa R incorpora las rutinas correspondientes para hacer estas pruebas, donde en la especificación de la prueba se incluye la selección de rezagos mediante distintos criterios de información. En particular, en el presente caso se seleccionan los rezagos en función al criterio de información de Akaike (*AIC*). Estas rutinas permiten constatar el orden de integración para procesos AR de cualquier orden, donde la selección óptima de rezagos se realiza de manera iterativa: comienza con un número máximo de rezagos, y va sacando uno a uno hasta quedarse con aquella cantidad de rezagos que haga mínimo el *AIC*.

A priori, no se conoce la forma funcional del proceso autorregresivo de las variables. Por tanto, el test consiste en seguir un proceso progresivo de mayor a menor, contrastando en cada paso que los componentes determinísticos sean significativos, tomando para ello los estadísticos adecuados expuestos por Dickey y Fuller (1981). A saber, se comienza estimando el modelo con constante y tendencia determinística, y se evalúa la significancia de estos componentes. En caso de que las mismas no fuesen significativas, se ajusta un modelo con constante determinística solamente, y si el coeficiente correspondiente no es significativo, se finaliza en un modelo sin componentes determinísticos.

Una vez seleccionado el modelo mediante esta metodología, se prueba si los residuos resultantes están correlacionados, lo cual indicaría que el modelo no está bien especificado. La prueba utilizada se debe a Ljung y Box (1978): la hipótesis nula es la ausencia de correlación serial o que los datos se distribuyen de manera independiente, versus la alternativa que sostiene la dependencia de los residuos. Se calcula el estadístico Q , definido como

$$Q = T(T + 2) \sum_{i=1}^{\nu} \frac{\delta_i^2}{T-i}, \quad (3.4)$$

donde T es el número de observaciones, δ_i es el coeficiente de autocorrelación de orden i de los residuos, ν es el orden de la autocorrelación que se está probando (en el presente trabajo se utiliza $\nu=10$). Bajo hipótesis nula, el estadístico Q definido en (3.4) sigue una distribución teórica asintótica χ^2_{ν} .

Por lo tanto, para un nivel de significancia α , si el estadístico $Q > \chi^2_{1-\alpha, \nu}$, entonces se rechaza hipótesis nula, concluyéndose que los residuos presentan autocorrelación, y por ello la especificación del modelo debe modificarse.

En la presente tesis, se implementan estas pruebas para conocer si cada una de las variables es estacionaria, buscando el modelo que cumpla con las especificaciones mencionadas. En caso de que alguna no fuese estacionaria, se realiza la prueba para la variable en diferencia, con el objeto de conocer el orden de integración de ésta.

Prueba Kwiatkowski, Phillips, Schmidt y Shin

Contrariamente a la prueba de Dickey Fuller Aumentada (*ADF*), la hipótesis nula de la prueba Kwiatkowski, Phillips, Schmidt y Shin (*KPSS*) afirma estacionariedad de la serie ($H_0: x_t \sim I(0)$), versus la alternativa de que no lo es ($H_1: x_t \sim I(1)$). Llamada así debido a quienes la propusieron, a saber, Kwiatkowski et al. (1992), es una alternativa al *ADF* para constatar estacionariedad de una serie de tiempo.

Siguiendo al planteamiento de Lütkepohl (2004a), la prueba supone que x_t proviene de un proceso generador de datos (sin componente de tendencia) como el siguiente,

$$x_t = e_t + z_t, \quad (3.5)$$

siendo e_t una caminata aleatoria dada por $e_t = e_{t-1} + v_t$, con $v_t \sim iid(0, \sigma_v^2)$ y, z_t un proceso estacionario. De esta forma, las hipótesis planteadas son equivalentes a: $H_0: \sigma_v^2 = 0$ versus $H_1: \sigma_v^2 > 0$, dado que si $\sigma_v^2 = 0$, entonces e_t es sólo una constante en el proceso estacionario y, por tanto, x_t también lo es.

Kwiatkowski et al. (1992) propusieron el siguiente estadístico,

$$KPSS = \frac{1}{T} \sum_{i=1}^T \frac{S_i^2}{\hat{\sigma}_\infty^2}, \quad (3.6)$$

con $S_t = \sum_{j=1}^t \hat{\sigma}_j$, $\hat{\sigma}_j = x_t - \bar{x}$, y $\hat{\sigma}_\infty^2 = \lim_{T \rightarrow \infty} T^{-1} Var \left(\sum_{t=1}^T z_t \right)$. Este último es un estimador de la varianza de largo plazo del proceso estacionario z_t . Entonces, si x_t es un proceso estacionario, S_t^2 está acotado (ya que es un estimador de su varianza), y entonces el $KPSS$ no será muy grande. Pero si el proceso no es estacionario, entonces S_t^2 crece indefinidamente, y entonces el $KPSS$ es muy grande.

Para evitar que la distribución del estadístico dependa de parámetros irrelevantes, Kwiatkowski et al. (1992) han propuesto el siguiente estimador no paramétrico de $\hat{\sigma}_\infty^2$:

$$\hat{\sigma}_\infty^2 = \frac{1}{T} \sum_{i=1}^T \hat{\sigma}_i^2 + 2 \sum_{j=1}^{l_q} \omega_j \left(\frac{1}{T} \sum_{t=j+1}^T \hat{\sigma}_t \hat{\sigma}_{t-j} \right), \quad (3.7)$$

donde $\omega_j = 1 - \frac{j}{l_q + 1}$, y $l_q = q \left(\frac{T}{100} \right)^{1/4}$ es un parámetro de truncamiento de rezagos.

Uno de los problemas del presente test es la selección de este parámetro de truncamiento, aunque habitualmente se toma $q=4$ ó $q=13$.

Si se sospecha la presencia de una tendencia determinística en el proceso generador de datos, la ecuación (3.5) se ve modificada de la siguiente manera

$$x_t = \mu_1 t + e_t + z_t, \quad (3.8)$$

y el estadístico es calculado de la misma forma que en la ecuación (3.6); sin embargo, la distribución asintótica bajo H_0 es diferente al caso sin tendencia – ecuación (3.5). Los valores críticos de dicha distribución son expuestos por Kwiatkowski et al. (1992).

El presente test también es implementado por medio del paquete *urca* (Pfaff, & Stigler, 2013a). Al no conocerse a priori qué forma funcional es la más adecuada, se realizan ambos test $KPSS$, tanto con presencia y ausencia del término de tendencia; como tam-

bién con ambos parámetros de truncamiento de rezagos, es decir, $q=4$, 13. Por lo tanto, para cada variable se realizan cuatro test, en función a la combinación de las características mencionadas.

Aún con hipótesis distintas, tanto el *KPSS* como el *ADF* tienen el mismo objetivo. Sin embargo, no necesariamente resultan en iguales conclusiones, por lo que se realizan ambos para obtener resultados más robustos.

3.2.2 Definición del VECM

Cointegración y Vector de Corrección de Errores

Engle y Granger (1987), Juselius (2008) y Lütkepohl (2007) explican que, en algunas situaciones, se espera que ciertas series de tiempos se muevan conjuntamente sin distanciarse mucho una de otra. Esto da idea de que existen ciertas relaciones de equilibrio entre ellas, constituyendo puntos estacionarios a largo plazo.

Destacan que una serie de tiempo estacionaria sin componentes determinísticos puede ser representada como un promedio móvil infinito que es generalmente aproximado por un ARMA. Sin embargo, comúnmente, las series económicas no son estacionarias, y se debe obtener la primera diferencia a fin de asumir la estacionariedad. Estas ideas motivaron a la definición de *Integración*.

Definición 3.1. Un proceso estocástico univariado x_t se dice integrado de orden d si, para $d > 0$, después de haber sido diferenciado d veces, es estacionario (denotado por $x_t \sim I(d)$). Es decir, si $\Delta^d x_t = (1-L)^d x_t$ es estable; donde L es el operador rezago (*lag*). (Lütkepohl, 2007, p. 242).

A partir de la **Definición 3.1** y extendiendo su uso al ámbito multivariado, se define Cointegración siguiendo a Enders (1995).

Definición 3.2. «Los componentes del vector $\mathbf{x}_t = (x_{1t}, x_{2t}, \dots, x_{nt})^T$ se dicen que están cointegrados de orden d, b , denotado por $\mathbf{x}_t \sim CI(d, b)$, si:

1. todos los componentes de \mathbf{x}_t son integrados de orden d ;

2. existe un vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_n)$ tal que la combinación lineal $\boldsymbol{\beta}^T \mathbf{x}_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt}$ es integrada de orden $(d-b)$, donde $b > 0$ ».
(Enders, 1995, p. 358).

La reducción del orden de integración genera un especial tipo de relaciones entre variables, con cierta interpretación y consecuencias contrastables. La existencia de cointegración significa que las variables involucradas tienen tendencias estocásticas (y/o determinísticas) comunes, y tienden a moverse juntas en el largo plazo (Juselius, 2008). Estas relaciones pueden interpretarse como relaciones estables en el largo plazo, de allí el interés en estudiarlas si las variables son económicas.

Siguiendo con la notación de la **Definición 3.2**, Engle y Granger (1987) en su análisis formal, comienzan diciendo que si n variables (\mathbf{x}_t) no estacionarias $I(1)$ están en un equilibrio de largo plazo, entonces se debe dar que $\boldsymbol{\beta}^T \mathbf{x}_t = \mathbf{0}$; las desviaciones del largo plazo son llamadas error del equilibrio (ζ_t) :

$$\boldsymbol{\beta}^T \mathbf{x}_t = \zeta_t, \quad (3.9)$$

y por tanto, si la relación a largo plazo es significativa, entonces el error de equilibrio debería ser estacionario, es decir, $\zeta_t \sim I(0)$.

Un VECM parte de que, bajo la existencia de un equilibrio estable a largo plazo entre las variables no estacionarias, a corto plazo las mismas se pueden enfrentar a desviaciones que se van corrigiendo hasta llegar nuevamente a dicho equilibrio. Por lo tanto, a partir del VECM es posible conocer la dinámica de corto plazo, siempre que las variables tengan una relación estable en el largo plazo, es decir, que estén cointegradas.

Entonces, para definir el modelo en términos de VECM, sea \mathbf{x}_t una matriz $n \times t$ de variables económicas en serie de tiempo $I(1)$ y cointegradas, entonces, en nivel conforman un sistema VAR de orden p no restringido dado por:

$$\mathbf{x}_t = \boldsymbol{\mu} + \sum_{i=1}^p \Phi_i \mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t, \quad (3.10)$$

donde $\boldsymbol{\mu}_{n \times 1}$ corresponde a un vector de constantes, Φ_i es un vector de parámetros $n \times 1$ correspondiente a los efectos de los i -ésimos rezagos, y $\boldsymbol{\varepsilon}_t$ es un vector de error de $n \times 1$.

Si se sustrae \mathbf{x}_{t-1} de ambos lados de la ecuación (3.10) y, se suma y resta $\sum_{i=1}^{p-1} \Phi_i \mathbf{x}_{t-1}$ del lado derecho de la mismas, el sistema (3.10) puede ser reescrito como un VECM,

$$\mathbf{x}_t - \mathbf{x}_{t-1} = \boldsymbol{\mu} + \sum_{i=1}^p \Phi_i \mathbf{x}_{t-i} - \mathbf{x}_{t-1} - \sum_{i=1}^{p-1} \Phi_i \mathbf{x}_{t-1} + \sum_{i=1}^{p-1} \Phi_i \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t,$$

$$\Delta \mathbf{x}_t = \boldsymbol{\mu} + \left(\sum_{i=1}^p \Phi_i - 1 \right) \mathbf{x}_{t-1} + \sum_{i=1}^{p-1} \Phi_i \mathbf{x}_{t-i} - \sum_{i=1}^{p-1} \Phi_i \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t,$$

$$\Delta \mathbf{x}_t = \boldsymbol{\mu} + \left(\sum_{i=1}^p \Phi_i - 1 \right) \mathbf{x}_{t-1} - \sum_{i=1}^{p-1} \Phi_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t,$$

$$\Delta \mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\Pi} \mathbf{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t, \quad (3.11)$$

donde $\boldsymbol{\Pi}_{n \times n} = \sum_{i=1}^p \Phi_i - 1$, $\boldsymbol{\Gamma}_i = -\sum_{i=1}^{p-1} \Phi_i$ es $n \times 1$ y, $\boldsymbol{\varepsilon}_t$ es iid: $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Omega}_{n \times n})$ y corresponde al vector de errores $n \times 1$ o término de perturbación.

El modelo VAR cointegrado expuesto en (3.11) debe ser estacionario, es decir $\Delta \mathbf{x}_t \sim I(0)$, y por tanto sus errores tienen que serlo.

Engle y Granger (1987) mostraron que $\boldsymbol{\Pi}_{n \times n}$ tiene rango r ($0 \leq r \leq n$), el cual indica la cantidad de relaciones de cointegración linealmente independiente entre las variables que componen el vector \mathbf{x}_t . Dependiendo de cuál sea el rango de dicha matriz, pueden suceder tres situaciones:

1. Cuando el $\text{rango}(\boldsymbol{\Pi}) = 0$, entonces todos los elementos de dicha matriz son cero y, el VECM es un modelo VAR en primeras diferencias.

2. Si el $\text{rango}(\mathbf{\Pi}) = n$, entonces es de rango completo. Esto sucede cuando las variables involucradas en \mathbf{x}_t son estacionarias.

3. Si el $0 < \text{rango}(\mathbf{\Pi}) = r < n$, entonces es el caso donde las variables no son estacionarias, y hay r vectores de cointegración entre ellas. Este es el único caso en que el análisis mediante VECM es correcto.

Lütkepohl (2007) concluye que ante la presencia de r vectores de cointegración entre las variables \mathbf{x}_t ($\text{rango}(\mathbf{\Pi}) = r$), se puede descomponer la matriz $\mathbf{\Pi}_{n \times n}$ de la siguiente manera: $\mathbf{\Pi} = \mathbf{\alpha}\mathbf{\beta}^T$, con $\mathbf{\alpha}$ y $\mathbf{\beta}$ matrices $n \times r$, ambas con rango r , donde $r < n$ y n , la cantidad de variables endógenas involucradas.

A su vez, al modelo (3.11) se pueden incorporar componentes determinísticos, a saber, una tendencia determinística (δt). Por tanto, el modelo (3.11) con estos componentes queda de la siguiente manera

$$\Delta \mathbf{x}_t = \boldsymbol{\mu} + \delta t + \mathbf{\alpha}\mathbf{\beta}^T \mathbf{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t. \quad (3.12)$$

Una constante en la ecuación de primera diferencia de una variable, representa una tendencia lineal cuando la variable está en nivel. Del mismo modo, una tendencia lineal en la ecuación en diferencia, representa una tendencia cuadrática en la variable en nivel. Por tanto, los componentes determinísticos $\boldsymbol{\mu} + \delta t$ de la ecuación (3.12) representan una tendencia lineal y cuadrática, respectivamente, de las variables en niveles (\mathbf{x}_t).

Los componentes determinísticos pueden ser descompuestos en tendencia de \mathbf{x}_t , los que a su vez se pueden separar de la tendencia de la ecuación de cointegración. Siguiendo a Juselius (2008), si se reescribe la constante y la tendencia del VECM de la siguiente forma

$$\boldsymbol{\mu} = \mathbf{\alpha}\boldsymbol{\nu} + \boldsymbol{\gamma}, \quad (3.13)$$

$$\delta t = \mathbf{\alpha}\boldsymbol{\rho}t + \boldsymbol{\tau}t, \quad (3.14)$$

si \mathbf{v} y $\boldsymbol{\rho}$ son vectores $r \times 1$, $\boldsymbol{\gamma}$ y $\boldsymbol{\tau}$ son de dimensión $n \times 1$, $\boldsymbol{\gamma}$ ortogonal a $\boldsymbol{\alpha}\mathbf{v}$, y $\boldsymbol{\tau}$ lo es con respecto a $\boldsymbol{\alpha}\boldsymbol{\rho}$, entonces el modelo (3.12) puede ser re-escrito de la siguiente forma

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\tau}t + \boldsymbol{\alpha}(\boldsymbol{\beta}^T \mathbf{x}_{t-1} + \mathbf{v} + \boldsymbol{\rho}t) + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t. \quad (3.15)$$

Luego, las ecuaciones de cointegración representan las relaciones de equilibrio de largo plazo entre las variables en niveles, que a su vez no son estacionarias. Los términos determinísticos en las relaciones de cointegración, $\mathbf{v} + \boldsymbol{\rho}t$, son la media y la tendencia lineal de dichas relaciones de largo plazo.

Esta forma de configuración es bastante flexible, y da lugar a cinco tipos de modelos, según la configuración de la tendencia que se seleccione. Pueden darse los siguientes casos:

- a) *Tendencia no restringida*, es cuando los cuatro parámetros son estimados, a saber, $\boldsymbol{\gamma}$, \mathbf{v} , $\boldsymbol{\tau}$ y $\boldsymbol{\rho}$. Se asume que existe una tendencia cuadrática en la variable a nivel \mathbf{x}_t , y por lo tanto, la/s ecuación/es de cointegración es/son estacionaria/s en tendencia.
- b) *Tendencia Restringida*, cuando $\boldsymbol{\tau} = 0$, entonces \mathbf{x}_t incluye una tendencia lineal, pero no cuadrática. Al igual que el caso a), la/s ecuación/es de cointegración es/son estacionaria/s en tendencia.
- c) *Constante no restringida*, cuando $\boldsymbol{\tau} = 0$ y $\boldsymbol{\rho} = 0$. Las variables observadas siguen una tendencia lineal, pero la/s ecuación/es de cointegración es/son estacionaria/s alrededor de una media constante.
- d) *Constante restringida*, cuando $\boldsymbol{\tau} = 0$, $\boldsymbol{\rho} = 0$, y $\boldsymbol{\gamma} = 0$. No hay tendencias en \mathbf{x}_t , pero la tendencia está restringida a estar en la relación de cointegración. La/s ecuación/es de cointegración es/son estacionaria/s alrededor de una media constante.
- e) *Sin tendencias*, cuando $\boldsymbol{\tau} = 0$, $\boldsymbol{\rho} = 0$, $\boldsymbol{\gamma} = 0$ y $\mathbf{v} = 0$, es decir, no hay medias ni tendencias distintas de cero.

El modelo representado en (3.15), así como todas sus versiones posibles, puede contener variables exógenas, continuas o discretas. La ecuación (3.15) y su equivalente en nivel, con estos componentes adicionales quedan, entonces, especificados de la siguiente manera,

$$\mathbf{x}_t = \boldsymbol{\alpha}\mathbf{v}t + \boldsymbol{\gamma}t + \boldsymbol{\alpha}\boldsymbol{\rho}t^2 + \boldsymbol{\tau}t^2 + \sum_{i=1}^p \Phi_i \mathbf{x}_{t-i} + \boldsymbol{\theta}\mathbf{y}_t + \boldsymbol{\eta}\mathbf{D}_t + \boldsymbol{\varepsilon}_t, \quad (3.16)$$

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\tau}t + \boldsymbol{\alpha}(\boldsymbol{\beta}^T \mathbf{y}_{t-1} + \mathbf{v} + \boldsymbol{\rho}t) + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\theta} \Delta \mathbf{y}_t + \boldsymbol{\eta} \mathbf{D}_t + \boldsymbol{\varepsilon}_t, \quad (3.17)$$

donde $\boldsymbol{\theta}$ y $\boldsymbol{\eta}$ corresponden a los vectores de coeficientes de las variables predictoras exógenas continuas y de las dummies, respectivamente; a su vez, \mathbf{x}_t y \mathbf{D}_t son los vectores que contienen dichas variables, respectivamente.

Los componentes de la descomposición de la matriz $\boldsymbol{\Pi}$, tienen importantes interpretaciones en el área económica. Los coeficientes de la matriz $\boldsymbol{\beta}$ son interpretados como las relaciones estables de largo plazo entre las variables, y $\boldsymbol{\alpha}$ son los coeficientes de ajustes en el corto plazo hacia el equilibrio de largo plazo.

Prueba de Exogeneidad Débil

Hasta ahora se ha asumido que todos los componentes de \mathbf{x}_t son endógenos. Sin embargo, puede sospecharse que alguno de ellos no lo es; y en caso de confirmarse la exogeneidad, la variable se debe incorporar al modelo como tal. En primer lugar, y siguiendo Engle, Hendry y Richard (1983) se define exogeneidad débil.

Considérese dos variables de series de tiempo, y_t y z_t , normalmente distribuidas e independientes:

$$[y_t \quad z_t] \sim N(\boldsymbol{\mu}, \boldsymbol{\Omega}) \quad t = 1, \dots, T. \quad (3.18)^{10}$$

Si \mathbf{x}_t es $(y_t, z_t)^T$ y se define a $\boldsymbol{\varepsilon}_t = \mathbf{x}_t - E(\mathbf{x}_t) = \mathbf{x}_t - \boldsymbol{\mu}$ como el error, entonces

¹⁰ Los valores en negrita denotan vectores o matrices.

$$\mathbf{x}_t = \boldsymbol{\mu} + \varepsilon_t \quad \varepsilon_t \sim N(\mathbf{0}, \boldsymbol{\Omega}), \quad (3.19)$$

Sin pérdida de generalidad, la función de densidad de \mathbf{x}_t ($F_{\mathbf{x}}(\mathbf{x}_t; \boldsymbol{\theta})$) puede ser factorizada en la densidad condicional de y_t dada z_t ($F_{y|z}(y_t | z_t; \boldsymbol{\gamma}_1)$) y la densidad marginal de z_t ($F_z(z_t; \boldsymbol{\gamma}_2)$) de la siguiente forma:

$$F_{\mathbf{x}}(\mathbf{x}_t; \boldsymbol{\theta}) = F_{y|z}(y_t | z_t; \boldsymbol{\gamma}_1) \cdot F_z(z_t; \boldsymbol{\gamma}_2), \quad (3.20)$$

donde $\boldsymbol{\theta}$ es el set de parámetros completo del proceso conjunto; $\boldsymbol{\gamma}_1$ y $\boldsymbol{\gamma}_2$ son los parámetros del proceso condicional y marginal, respectivamente, cuyos espacios de parámetros son Θ , Λ_1 y Λ_2 . Si se define $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T; \boldsymbol{\gamma}_2^T)^T$ y se denota el espacio de parámetros como Λ .

Entonces, sea $\boldsymbol{\psi}$ es el vector de parámetros de interés:

Definición 3.3: «La variable z_t es débilmente exógena sobre el período de muestra respecto de un conjunto de parámetros de interés $\boldsymbol{\psi}$ si y sólo si existe una reparametrización $\boldsymbol{\theta}$ como $\boldsymbol{\gamma}$, con $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T; \boldsymbol{\gamma}_2^T)^T$ tal que:

(i) $\boldsymbol{\psi}$ es una función de $\boldsymbol{\gamma}_1$ solamente, y

(ii) la factorización de la ecuación (3.20) opera como un corte secuencial, i.e.,

$$F_{\mathbf{x}}(\mathbf{x}_t; \boldsymbol{\theta}) = F_{y|z}(y_t | z_t; \boldsymbol{\gamma}_1) \cdot F_z(z_t; \boldsymbol{\gamma}_2),$$

con $\boldsymbol{\gamma} \in \Lambda_1 \times \Lambda_2$.» (Engle, Hendry, & Richard, 1983, p. 282)

Si existe exogeneidad débil, entonces se puede realizar estimaciones eficientes y realizar pruebas a través de modelos condicionales, ignorando la información del proceso marginal.

El test de exogeneidad débil ha sido debidamente instrumentado por Pffaf (2008), quien se ha basado en los trabajos de Johansen y Joselius (1990), y Johansen (1991, 1995). El mismo consiste en probar a partir de un test de razones de verosimilitudes si el modelo

VEC restringido (donde se supone exógena/s la/s variables) es el más adecuado, versus el modelo con todas las variables endógenas.

El test consiste en imponer restricciones en la matriz de carga α de (3.11), (3.12), (3.15) o (3.17), según corresponda. Las hipótesis son $H_0: \alpha = A\alpha$ versus $H_1: \alpha = I\alpha$, donde $A_{g \times n}$ es la matriz que contiene las restricciones y g es el número de variables no restringidas, por tanto, se tiene $c = n - g$ restricciones (las mismas buscan constatar si el coeficiente de la variable posiblemente exógena igual a cero, por ejemplo, $H_0: \alpha_3 = 0$). A partir de los autovalores (λ) de la matriz de las varianzas y covarianzas obtenidos del modelo VECM restringido estimado por máxima verosimilitud, y los del modelo completo, el estadístico de la razón de verosimilitud utilizado es definido por Johansen (1991) de la siguiente manera,

$$-2 \ln(Q: H_0 | H_1) = T \sum_{i=1}^r \left\{ \frac{(1 - \lambda_{H_0})}{(1 - \lambda_{H_1})} \right\}, \quad (3.21)$$

el cual es asintóticamente distribuido como $\chi_{r \times c}^2$. Si la hipótesis nula no puede ser rechazada, entonces el modelo VECM a estimar debe ser reducido, a uno g -dimensional.

Selección del Número de Rezagos del VECM

Otra cuestión a resolver en la definición del modelo es el número de rezagos a incorporar en el VECM. Este proceso de selección consiste en estimar modelos VAR con las variables endógenas teniendo en cuenta diferentes rezagos. Éstos no sólo tienen en cuenta las variables endógenas sino también aquellas exógenas y componentes determinísticos. No existe un número de rezagos máximo a calcular, eso dependerá del conocimiento del problema y de los datos disponibles.

Una vez que fueron estimados los modelos alternativos, se debe seleccionar cuál de ellos es el más apropiado. Enders (1995) y Lütkepohl (2007) señalan que hay varios procedimientos factibles de ser utilizados en dicha selección. Uno de ellos consiste en realizar un proceso de truncamiento de rezagos partiendo de un número alto de los mismos, e ir eliminándolos siempre que sus coeficientes no sean estadísticamente diferentes de cero. El otro procedimiento se basa en seleccionar el número de rezagos que optimi-

ce diferentes criterios de información. En el presente trabajo se utiliza esta última metodología de selección.

Los criterios de información factibles de ser obtenidos en cada modelo estimado son: Akaike (*AIC*), Hannan-Quinn (*HQ*) y Schwartz-Bayes (*SB*). Sus versiones multivariantes se definen a continuación:

$$AIC(p) = \left[\ln |\Sigma_u(p)| \right] + \frac{2}{T} pn^2, \quad (3.22)$$

$$HQ(p) = \left[\ln |\Sigma_u(p)| \right] + \frac{2 \ln(\ln(T))}{T} pn^2, \quad (3.23)$$

$$SB(p) = \left[\ln |\Sigma_u(p)| \right] + \frac{\ln(T)}{T} pn^2, \quad (3.24)$$

donde $|\Sigma_u(p)|$ es el determinante de la matriz $\Sigma_u(p) = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t^T$, la cual es un estimador de la matriz de varianzas y covarianzas para un modelo con p rezagos, T es el tamaño de la muestra o cantidad total de períodos, y n es el número de variables endógenas del modelo. Se asume que el tamaño de la muestra es constante para todos los órdenes de rezagos, por lo que se establece un número máximo de rezagos.

Si bien se tienen en cuenta los tres criterios, no siempre los mismos son idénticos en la decisión final del número de rezagos óptimo. Por tanto, siguiendo a Pfaff (2008), se realiza un análisis de diagnóstico de los VAR con los rezagos óptimos indicados por los criterios de información, para decidir por aquel modelo cuyos rezagos genere las condiciones más adecuadas en torno al comportamiento de los residuos producidos. Se hacen pruebas de autocorrelación serial (test de Portmanteau), normalidad (test de Jarque-Bera) y proceso autoregresivo heteroscedástico (test ARCH¹¹) sobre los errores obtenidos.

Test de Cointegración de Johansen

Una vez seleccionado el número de rezagos y la forma del modelo (componentes exógenos y dummies), se procede a efectuar las pruebas de cointegración. En cuanto a la

¹¹ Por sus siglas en inglés *Autoregressive Conditional Heteroskedasticity*.

introducción de componentes determinísticos, el proceso es de prueba y error, y en definitiva se selecciona el modelo que tenga mayor capacidad predictiva.

Definida la forma del VECM, con el test de cointegración de Johansen se prueba la existencia de un equilibrio estable a largo plazo entre las variables involucradas. Basso (2009) demuestra que dicho test es una versión multivariante de la prueba de *ADF* (Dickey, & Fuller, 1979,1981).

El test de cointegración de Johansen consiste en obtener los valores propios de la matriz $\mathbf{\Pi} = \mathbf{\alpha}\mathbf{\beta}^T$, definida en las distintas versiones del VECM, en las ecuaciones (3.11), (3.12), (3.15) y (3.17).

Siguiendo a Enders (1995), sean λ_i $i=1,2,\dots,r$ los valores propios de la matriz $\mathbf{\Pi}$, pueden suceder las siguientes situaciones:

- Si las variables involucradas en \mathbf{x}_t no están cointegradas, todos los valores propios son estadísticamente iguales a cero. Por tanto, la matriz $\mathbf{\Pi}$ es nula, y su rango también lo es.
- Si las variables están cointegradas, existe/n valor/es propio/s distinto/s de cero y, por tanto, el $\text{rango}(\mathbf{\Pi}) = r$ será del tamaño de la cantidad de valores propios distintos de cero (r).

Pero como los verdaderos valores propios de la matriz $\mathbf{\Pi}$ no se conocen, y entonces son estimados ($\hat{\lambda}_i$), se realiza una prueba estadística sobre λ_i . Sea T el total de observaciones (o cantidad total de períodos) disponibles en la prueba, Johansen (1998) define dos pruebas basadas en dos estadísticos distintos, a saber,

$$\lambda_{\text{traza}}(r) = -T \sum_{i=r+1}^n \log(1 - \hat{\lambda}_i), \quad (3.25)$$

$$\lambda_{\text{max}}(r, r+1) = -T \log(1 - \hat{\lambda}_{r+1}), \quad (3.26)$$

donde $0 \leq r \leq n$.

Se contrastan las siguientes hipótesis con el estadístico λ_{traza} , $H_0: \text{rango}(\mathbf{\Pi}) \leq \lambda_{\text{traza}}$ versus $H_1: \text{rango}(\mathbf{\Pi}) > \lambda_{\text{traza}}$; y con el estadístico λ_{max} , $H_0: \text{rango}(\mathbf{\Pi}) = r$ versus $H_1: \text{rango}(\mathbf{\Pi}) = r + 1$.

La distribución de los estadísticos depende del número de componentes no estacionarios involucrados bajo la hipótesis nula (i.e. $n - r$) y de la inclusión de elementos determinísticos en la relación.

En la presente tesis, el modelo con constante y tendencia determinística en la relación de cointegración tiene el mejor desempeño predictivo. Por tanto, se utilizan los estadísticos de la prueba de Johansen publicados por Osterwald-Lenum (1992), ya que éste ha estudiado las distribuciones de los mismos en modelos con constante y tendencia determinística en el largo plazo, y por tanto, sus correspondientes valores críticos.

3.2.3 Método de Estimación del VECM y Diagnóstico-Verificación

Definida la forma del modelo, constatada la existencia de cointegración así como el rango de la matriz de cointegración, se procede a estimar los parámetros. Se utiliza el método de mínimos cuadrados ordinarios (MCO). Dicho método es implementado a partir del paquete de *R*, llamado ‘*urca*’ (Pffaf, & Stiggler, 2013a). La ventaja del mismo es la posibilidad de encontrar, de manera relativamente fácil, el modelo VAR equivalente para el VECM estimado. Por tanto, a partir del VECM estimado mediante MCO se obtiene el VAR equivalente que permite efectuar las predicciones sobre la variable en nivel.

Una vez estimado el VECM, se procede a probar si es el adecuado o, por el contrario, existen irregularidades que deben revisarse. Para ello, se realizan verificaciones sobre los residuos resultantes. En primer lugar, se prueba la estacionariedad de los residuos mediante las pruebas descriptas, a saber, *ADF* y *KPSS*.

También se prueba: a) la presencia de proceso autoregresivo heteroscedástico de los residuos a partir de test de ARCH; b) Normalidad de los residuos, por medio del test Jarque-Bera, y c) Autocorrelación serial de los residuos, a partir del test de Portmanteau.

3.2.4 Predicción con VECM

En el presente trabajo se realiza la estimación del VECM y su equivalente VAR para poder predecir la variable de interés.

Se utilizan dos procedimientos para realizar las predicciones a partir del VECM. El primero, llamado *Estático*, que predice 12 meses futuros con parámetros estáticos. Este procedimiento funciona de la siguiente manera. A partir de la información de los últimos períodos utilizados en el proceso de estimación de cada una de las series, y de las estimaciones de los parámetros del modelo, genera proyecciones de un período hacia adelante de cada una de esas variables endógenas. Dichas predicciones son tomadas como datos para predecir el período subsiguiente. Así, este proceso es sucesivo hasta completar los 12 períodos de interés. Por lo tanto, cada predicción de un período determinado, pasar a ser dato de entrada para el siguiente período a predecir, pero los parámetros del modelo se mantienen intactos, de allí el nombre de modelo estático de predicción.

La otra forma de predicción se denomina en la presente tesis *Dinámica*, y consiste en predecir los 12 meses de interés de a uno por vez, incorporando a la matriz de estimación los datos reales de a uno y, reestimando los parámetros en cada paso. Por ejemplo, para predecir a las variables en el período t , se incorpora información relevante del período $t-1$ y se reestiman los parámetros del modelo.

El primero de los procesos sirve para predecir períodos más prolongados de tiempo, bajo el supuesto de que no se conocen los valores reales de las variables y se quiere estimar a más largo plazo; el segundo, se trata de predicciones de corto plazo actualizándolo mes a mes.

En la presente tesis, se muestran solo las estimaciones de los parámetros del primer formato de predicción, ya que del segundo correspondería mostrar las 12 reestimaciones de los modelos y no se ajusta al objetivo primordial del presente trabajo. Sin embargo, se muestran los resultados predictivos de ambos procesos de estimación.

3.3. Predicción con ANN

En el intento de imitar la forma que el cerebro procesa información, se han desarrollado modelos matemáticos llamados ANN o modelos de computación conexionistas (Haykin,

2001). A partir del conocimiento científico de los rezagos y características de los elementos conductores de la información dentro del sistema nervioso y sus diferentes conexiones, se han diseñado modelos ideales, que a pesar de ser limitados (si los comparamos con el funcionamiento real) se han vuelto relevantes para la investigación científica práctica de otras áreas distintas de las neurociencias.

Las ANN se presentan como un modelo matemático computacional no lineal que, a partir de un proceso de aprendizaje, pueden detectar relaciones funcionales entre variables por medio de conjuntos de datos concretos, y reconocer patrones de comportamiento que permitan clasificar, predecir y controlar. Son relativamente flexibles, ya que no necesitan de un conocimiento previo sobre las relaciones entre las variables de entradas y las salidas, y permite realizar generalizaciones e inferir fuera de la muestra.

Sus ventajas se centran en: (i) la no linealidad; (ii) una estructura de entrada y salida y, (iii) que es adaptable. La no linealidad permite que las interconexiones entre las neuronas no obedezcan a patrones lineales. La red 'aprende' de los datos mediante la construcción de un mapeo de entrada a salida, ajustando sus ponderaciones en la medida de que lo aprendido se aleje de la realidad. Es adaptable, pues tiene la capacidad para 'ajustarse' a cambios del medio.

Las ANN pueden ser de distintas formas y son aplicables a distintos tipos de datos, particularmente a las series de tiempo. Predecir estos tipos de datos mediante ANN implica diseñar el procesamiento de patrones que se extienden en el tiempo. Un valor particular de una variable en el tiempo t puede depender no sólo de los valores pasados de la misma, sino también del valor en t de otras variables relacionadas (y sus respectivos valores anteriores). El tipo de ANN que contempla esta arquitectura, son aquellas inventadas por Rosenblatt (1958) llamadas perceptrón multicapa, ya que contienen varias capas con estructura no lineal.

Para estimar una ANN primero se necesitan definir sus diferentes componentes, las relaciones entre ellos, así como los procedimientos y parámetros de estimación, a saber: (a) la arquitectura e interconexiones; (b) la/s función/es de activación o de transferencia; (c) la función de costo, que evalúa la/s salidas de la red con sus valores originales; (d) el algoritmo de entrenamiento, que cambia de manera iterativa las ponderaciones para minimizar la función de costo.

En general, el procedimiento de diseño y estimación de una red no es único. Muchos autores definen el diseño final del modelo a partir de un procedimiento de prueba y error. Algunos han teniendo en cuenta estas dificultades, y propusieron una serie de pasos estimativos para la implementación y estimación de una ANN (Kaastra, & Boyd, 1996; Zhang, 2004). Por ejemplo, Kaastra y Boyd (1996), enumeran en ocho pasos las posibles actividades tendientes a utilizar en la estimación ANN. Las mismas son:

- a. Selección de Variables.
- b. Recopilación de datos.
- c. Procesamiento de los datos.
- d. División de la base de datos en: conjunto de entrenamiento, de prueba y validación.
- e. Definición de: (a) número de capas ocultas, (b) número de neuronas ocultas, (c) número de neuronas de salida y, (d) funciones de transferencia.
- f. Selección de criterios de evaluación.
- g. Entrenamiento de la ANN, lo que implica: a) seleccionar el número de iteraciones y, b) elegir la tasa de aprendizaje.
- h. Implementación.

Dicha estructuración es tomada en cuenta en el entrenamiento y evaluación de la ANN que se estiman en la presente tesis.

3.3.1 Características Generales de las ANN

Las ANN son modelos matemáticos no lineales formados por elementos que están interconectados. Éstos están dispuestos en capas y, con cada par de elementos conectados, hay asociado un valor llamado ponderación o peso. Por lo tanto, una ANN típica está formada por los siguientes elementos de forma:

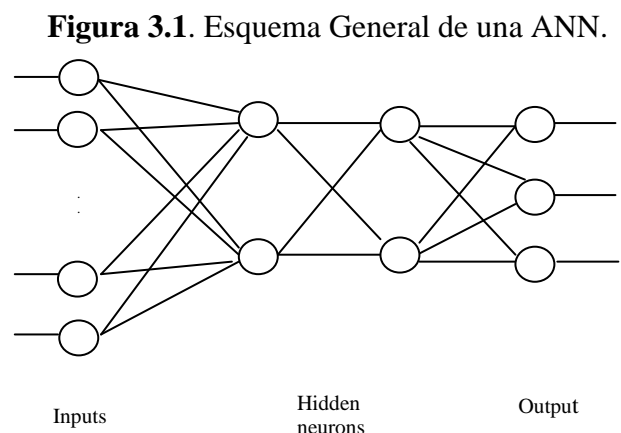
- a. La *capa de entrada* o *input layer*, por medio de la cual se suministra de información a la red. Los elementos que componen esta capa tienen la característica de ser datos numéricos.

b. Una o varias *capas ocultas* o *hidden layer/s*, sobre la/s que se proyectan los datos desde los elementos de la capa de entrada. En cada uno de los elementos de la capa oculta, se calcula una suma ponderada de los datos de entrada (o de la capa inmediatamente anterior en el caso de que haya más de una capa oculta). A dicha suma ponderada se le aplica una función de transferencia o de activación (en general, para problemas económicos se utiliza la “sigmoide”), siendo el valor resultante, la salida de ese elemento. Una red puede tener más de una capa oculta, aunque en general suelen funcionar bastante bien con un solo *hidden layer*.

c. La *capa de salida* u *output*. Punto final de la información proveniente desde la capa input o de las intermedias. Cada elemento de esta capa se obtiene a partir de las sumas ponderadas de las salidas de las capas ocultas que, a su vez se transforman mediante otra función de transferencia o activación.

En la figura 3.1 se puede observar el esquema de una ANN perceptrón multicapa, con indefinidos nodos en la capa de entrada, dos capas ocultas, cada una con dos nodos, y tres nodos en la salida.

La arquitectura de una ANN se define por su forma, el tipo de conexiones entre sus nodos y por la función de activación o transferencia. Existen muchas formas factibles, dependiendo de la naturaleza del problema. Pueden haber unas ANN sin capa oculta, y son llamadas redes lineales (Joekes, 2002). Sin embargo, la fuerza de las ANN es la no linealidad obtenida mediante la inserción de una o más capas ocultas (Weigend, & Gershenfeld, 1994).



Fuente: Elaboración propia a partir de bibliografía variada.

El flujo de información puede recorrer a lo largo de la red de dos maneras: a) *proalimentada* o *feedforward*, cuando fluye en una única dirección o sentido; o b) *retroalimentada*, *feedback* ó *recurrentes*, cuando posee información que fluye en ambas direcciones. El tipo de conexiones posibles no es el mismo en ambos casos. En una red proalimentada la información fluye desde la capa de entrada a la salida; en cambio en una ANN retroalimentada, la información fluye desde la entrada a la salida, pero puede haber nodos que transmitan su información a nodos anteriores, generando un proceso dinámico.

Diseñada la red, en el proceso de modelización se pueden distinguir dos fases, la de aprendizaje o entrenamiento y la de prueba. En la primera, se ajustan las ponderaciones, y en la segunda, se analiza la idoneidad de la red diseñada. Del conjunto de información disponible, se subdivide en dos partes, una para realizar el proceso de entrenamiento, y la otra, para la fase de prueba.

El proceso de entrenamiento es iterativo, y consiste en la adecuación de los pesos o ponderaciones de la red, en respuesta al estímulo generado por las entradas y de acuerdo con el tipo de función de transferencia o activación. Los métodos de aprendizaje se pueden clasificar en *supervisado* o *no supervisado*. El primero, es un proceso que incorpora un tutor externo y/o información global. De modo que, con cada conjunto de datos de entrada que se suministra a la red, la salida esperada obtenida se compara con los valores reales, con el objetivo de evaluar el error cometido y poder minimizarlo en el proceso de estimación. El segundo método, es un proceso auto-organizativo que no tiene en cuenta un tutor externo, sólo utiliza información local. Este proceso auto-organiza los datos presentados y descubre las propiedades internas más sobresalientes. De los dos métodos, el primero es el más utilizado, ya que genera una ANN con un método de evaluación general de su idoneidad, contando con una función de costo que se desea minimizar en el proceso iterativo de estimación de los pesos.

Por otra parte, el algoritmo con mayor difusión en el área económica es el de retro propagación o *backpropagation* desarrollado por Rumelhart, Hinton y Williams (1986). Sin embargo existen otros algoritmos posibles de ser utilizados en el proceso de estimación, como el *Adalain*, el *Madalain* y la *Función Básica Radial (RBF)*, entre otros.

A continuación se muestra el proceso en el cual se define la arquitectura y demás especificidades de la estimación.

3.3.2 Forma Teórica de una ANN

Una ANN es un modelo matemático. Si se analiza a la red desde atrás hacia adelante, se puede describir a las salidas o *outputs* como la transformación de la suma ponderada de los elementos precedentes; los cuales a su vez, son el resultado de una transformación de los nodos que los antecede, y así sucesivamente. Los únicos nodos que contienen información genuina son los iniciales, es decir, los correspondientes a la capa de entrada; el resto son el resultado de la transformación de las sumas ponderadas. Puede haber más de una función de activación o transformación, dependiendo de la cantidad de capas a transformar (a saber, la/s oculta/s y de salida), donde dicha/s función/es no es/son necesariamente lineal/es.

Una formulación general de una ANN proalimentada (o *feedforward*) de tres capas con posibilidades de tener conexiones directas entre *inputs* y *outputs* se escribe de la siguiente manera:

$$y_u = f_o \left(\sum_{u=1}^m w_{iu} x_i + \sum_{j=1}^s w_j f_h \left(\sum_{i=1}^n w_{ij} x_i \right) \right), \quad (3.27)$$

donde n es el total de nodos en la capa de entrada; s la cantidad total de nodos en la capa oculta y, m la cantidad de nodos en la capa salida. Por otro lado, x_i es el i -ésimo *input*, w corresponde a las ponderaciones, las cuales pueden ser: (a) interconexiones *inputs-hidden layer* (w_{ij}), (b) *hidden layer-output* (w_j), o (c) *input-output* (w_{iu}); f_h y f_o son las funciones de transferencias (o activación) de las capas ocultas y del *output*, respectivamente; y por último, y_u es el u -ésimo *output*.

En el presente trabajo se estima una ANN como la descrita en (3.27), pero sin conexiones directas entre *inputs* y *outputs*, y con la posibilidad de incluirse una segunda capa oculta con un total de e nodos, a saber

$$y_u = f_o \left(\sum_{d=1}^e w_d f_{h_2} \left(\sum_{j=1}^s w_j f_{h_1} \left(\sum_{i=1}^n w_{ij} x_i \right) \right) \right) \quad (3.28)$$

con f_{h_1} y f_{h_2} son las funciones de activación de la primera y la segunda capa oculta, respectivamente.

Las funciones (3.27) y (3.28) prevén la inclusión de un nodo en cada capa de la arquitectura de la red (exceptuando la capa de salida) que solamente está conectado hacia adelante con la capa siguiente; éste corresponde al nodo de sesgo o regulador externo de ajuste (*offset*). En general, dicho/s nodo/s de sesgo no son mencionados, pero en la estimación son tenidos en cuenta.

Las funciones de transferencias o activación pueden ser de distintos tipos, a saber, lineales, logísticas o sigmoideas, hiperbólicas tangentes, etc. Incluso puede haber distintos tipos de funciones de transferencias en una misma ANN (Jha, & Sinha, 2013; Joekes, 2002; Otero, & Trujillo, 1993).

Por ejemplo, suponiendo una ANN proalimentada con dos tipos de funciones de transferencia: a) n nodos en la capa de entradas, b) las entradas están completamente interconectadas a la capa oculta logísticamente, la cual tiene un total de s nodos, c) los j -ésimos nodos de la capa oculta están conectados a una única salida lineal, d) tanto la capa de entrada como la oculta tienen un regulador externo de ajuste o sesgo, donde el peso del sesgo de entrada en el j -ésimo nodo oculto es $w_{(n+1)j}$ y, el peso del sesgo oculto en la salida es $w_{(s+1)j}$; y, e) las ponderaciones w_{ij} pueden tener cualquier signo, incluso ser iguales a cero. Entonces, la entrada del j -ésimo nodo oculto (v_j), es la suma ponderada de los nodos de entrada más el sesgo $w_{(n+1)j}$,

$$v_j = \left(\sum_{i=1}^n w_{ij} x_i \right) + w_{(n+1)j}. \quad (3.29)$$

Ahora, la salida de dicho nodo se determina por medio de la transformación de v_j , a través de la función de activación logística de la siguiente manera:

$$\varphi(v_j) = \frac{1}{1 + \exp(-\varphi v_j)}, \quad (3.30)$$

donde ϕ es un parámetro desconocido. Luego, cada una de las salidas de los nodos ocultos es linealmente combinada con los pesos correspondientes a las conexiones de la capa oculta con la de salida, para formar la entrada del nodo de salida. La salida del *output* es exactamente igual a su entrada, ya que la función de transformación supuesta es la lineal.

En el presente trabajo, se define una ANN multicapa proalimentada, donde se utilizan dos tipos de funciones de activación, a saber, la logística para los nodos de la/s capas ocultas, y la lineal para la salida. Cuando el interés es hacer predicciones del *output*, la función lineal es el tipo de función de activación en la capa de salida a utilizar (Barrientos et al., 2012).

3.3.3 Diseño o Arquitectura de la ANN

Diseñar una red no es una tarea fácil, y muchas veces existe incertidumbre sobre cómo iniciar dicho proceso de diseño. Se ha puesto mucho esfuerzo en el desarrollo de redes y algoritmos de aprendizaje pero, sin embargo, el diseño de una arquitectura óptima y parsimoniosa de una ANN ha permanecido como un problema no resuelto (Barrientos et al., 2012; Fouladgar et. al, 2014; Joekes, 2002). Optimizar la arquitectura implica encontrar el diseño con el menor tamaño, que produzca errores mínimos tanto para los datos de entrenamiento como para los de test.

Hay un número infinito de formas de construir una red neuronal. Morgan y Bourland (1990) aseguran que, en general, las ANN son bastantes robustas respecto a una arquitectura específica, pero a menudo sus resultados dependen del modelo seleccionado.

Se ha establecido cierto orden en el proceso de diseño de una red. Específicamente,

a) Selección de Variables de Entrada como primer paso. Kaastra y Boyd (1996) señalan que la clave para tener éxito en este objetivo está en el conocimiento del problema en particular. Es una cuestión crítica conocer qué variables son relevantes dentro del problema. Sin embargo, esta tarea no es fácil y se requiere del auxilio de la teoría específica para encontrar dichas variables.

En aplicaciones económicas, para la construcción de ANN con el objetivo de predecir determinada/s variable/s de un mercado, se pueden usar como nodos de entrada varia-

bles económicas tanto *técnicas* como *fundamentales*. Las variables técnicas son las mismas variables de salida pero rezagadas; y las fundamentales corresponden a otras variables económicas que pueden influenciar el funcionamiento del mercado (Kaastra, & Boyd, 1996, p. 219).

b) El Número de Capas Ocultas. Éstas proveen a la red la habilidad de generalizar. En la práctica, son muy utilizadas las redes con una o dos capas ocultas, ya que funciona bastante bien. Incrementar el número de capas ocultas aumenta el tiempo computacional y los parámetros del modelo, y a su vez genera una alta probabilidad de caer en sobreajustes (Baum, & Haussler, 1989).

c) El Número de Nodos Ocultos. No hay una fórmula mágica para seleccionar la cantidad óptima de nodos en cada capa oculta. En general se trata de un experimento de prueba y error, teniendo presente que se recomienda que la cantidad de nodos disminuya a medida que avanza la estructura de la red en sentido de la capa de salida.

d) El Número de Neuronas de Salida. Decidir el número de nodos en la capa de salida depende del problema en particular que se está tratando. Generalmente, una ANN con múltiple nodos en la capa de salida produce resultados inferiores comparándolo con aquellas en que se tiene un solo nodo de salida (Masters, 1993).

e) Función de transferencia o Activación. Haykin (2001) señala que la función de transferencia más usada es la sigmoïdal o logística. Sin embargo, pueden usarse otras como la tangente hiperbólica, la arco-tangente y la lineal. La función lineal no es útil para propósitos de clasificación y mapeo no lineal, pero es muy utilizada en la capa de salida cuando el propósito es el pronóstico. Se ha demostrado que, para relaciones no lineales (como los mercados financieros), las funciones de transferencias no lineales son mejores (Kao, & Ma, 1992; Levich, & Thomas, 1993). En series de tiempo, la función sigmoïdal en las capas ocultas es la más utilizada (Maciel, & Ballini, 2010)

Existen ciertos mecanismos que sirven de guía para optimizar la estructura de una ANN. Park, Murray y Chen (1996) mencionan que la selección final de la arquitectura puede ser definida por medio de diferentes aproximaciones, por ejemplo, aproximación ad-hoc, dinámica, por la distribución y el pruning o poda, entre otras.

La primera de ellas se basa en el conocimiento que tiene el investigador sobre el problema, cuya selección y diseño de la arquitectura se realiza de acuerdo a su experiencia. Esta metodología de diseño se justifica en la característica de robustez de las ANN (Tang, de Almeida, & Fishwick, 1991), y su formato depende del conocimiento que se tenga del problema.

La aproximación dinámica, automatiza el procedimiento de selección de la arquitectura. A partir de una matriz de entrada y salida, varía el número de parámetros de la red (o conexiones) durante el entrenamiento sobre la base de los errores del sistema. Los métodos más conocidos son el de correlación en cascada (Fahlman, & Lebiere, 1989), el algoritmo de llegada (Frean, 1990) y, el método de eliminación ponderada (Weigend, Rumelhart, & Huberman, 1991).

La aproximación por distribución consiste en encontrar la función de distribución de los estimadores, y formular una medida de la performance de la red para luego compararlas. Hay diferentes formas de cumplimentar esta metodología, por ejemplo, Fogel (1990) propone un estadístico de información final (FIS¹²) bajo el supuesto de que la entrada de los nodos tiene una distribución normal. Dicho estadístico es una derivación del *AIC* y sirve para seleccionar la arquitectura del ANN.

Por último, las técnicas de diseño de la arquitectura por poda o *pruning* consisten en un proceso que identifica las unidades de una red que son innecesarias en la solución, y las elimina de la red. Dentro de este tipo, se tiene el método de selección de variable de entrada llamado Recorte Basado en la Sensibilidad (*Sensitivity-Based Pruning*), o aquellos que eliminan las conexiones con menos relevancia como, por ejemplo, el Recorte Basado en las Magnitudes (*Magnitude Based Pruning*), el *Optimal Brain Damage* y el *Optimal Brain Surgeon*, entre otros.

La presente tesis implementa el diseño de una ANN con el objetivo de comparar la capacidad predictiva de la misma para proyectar una variable versus la realizada a partir de una VECM. Por lo tanto, se utiliza el procedimiento más simple en el diseño de dichas ANN, a saber, un diseño ad-hoc. En el mismo, las variables de entrada surgen de la teoría económica y, del conocimiento y la disponibilidad de información del sector económico de aplicación para reconocer los factores que pueden estar relacionados a lo

¹² Por sus siglas en ingles *Final Information Statistic*

largo de la cadena productiva láctea argentina. Según la definición de Kaastra y Boyd (1996) se usan tanto variables técnicas como fundamentales, y se incorporan como *input* aquellas variables que no tengan altas correlaciones entre ellas. Para ello se realiza previamente un análisis de las correlaciones entre variables.

En cuanto a la selección del número de neuronas de salida se decide utilizar sólo la variable de interés predictivo, con la intención de obtener un modelo parsimonioso y con la mayor cantidad de grados de libertad en las estimaciones que permitan arribar al objetivo.

Por otra parte, la cantidad de capas ocultas y sus correspondientes nodos, se seleccionan a partir de un procedimiento de prueba y error, quedando el modelo que más capacidad predictiva tenga, resguardando los requisitos de forma de las ANN.

3.3.4 Método de Aprendizaje

Una vez diseñada la ANN, el proceso de estimación de los parámetros consiste en encontrar las ponderaciones que interconectan los componentes de la misma. A este proceso se lo llama proceso de aprendizaje, y es iterativo. Durante dicho proceso se va ajustando las ponderaciones de cada uno de los elementos, que a su vez, son la respuesta al estímulo generado por sus correspondientes entradas y teniendo en cuenta la función de transferencia estipulada.

Tal como se menciona en el apartado de las características generales, no hay un único método de aprendizaje en las ANN. Se puede distinguir dos:

a) *Supervisado*: proceso que incorpora un tutor externo y/o información global de control, de modo que, con cada conjunto de *inputs*, se suministre a la red el *output* esperado y, permita evaluar el error producido. En economía, este tipo es el más utilizado.

b) *No supervisado*: proceso que no tiene un tutor externo y cuenta sólo con información local y de control interno.

En el presente trabajo se utiliza el método de aprendizaje Supervisado. Esta técnica permite deducir patrones a partir de datos de entrenamiento, para luego, controlar el desempeño de los resultados estimados a partir de datos control, que en este caso son los

valores observados de la variable de salida. Por tanto, para poder llevar a cabo las estimaciones utilizando dicho método de aprendizaje se divide la base de datos con la que se cuenta como se explica a continuación.

División de Base de Datos

Hay ciertos temas centrales en la modelación de una ANN. Uno de ellos es el tamaño de la muestra usada para construirlas. Mientras que no existen reglas específicas sobre este tema, siempre es preferible contar con una muestra grande, ya que no sólo permite la introducción de un mayor número de parámetros, sino que además, se cuenta con mayores datos en los subconjuntos de entrenamiento y prueba.

En general, el conjunto total de datos disponibles se divide, generando dos fases en el proceso. Durante la primera, se realiza el ajuste de las ponderaciones de la red y selección del modelo. Se suministra información a la red a fin de que se produzca el proceso de adaptación de las ponderaciones, donde las ponderaciones iniciales se determinan de manera aleatoria. A su vez, este conjunto de datos, es subdividido en entrenamiento y de validación, donde este último permite la selección del modelo propiamente dicha.

La segunda, es la fase de prueba. Durante ésta, la red procesa una información de entrada ajena a la utilizada en la fase anterior, y dadas las ponderaciones estimadas allí, crea una salida u *output*, que constituye el valor predicho de la red. Éste corresponde a las predicciones *ex-post* del modelo correspondiente. El valor real observado es llamado *target* en el contexto de las ANN.

Concluyendo, en total se divide al conjunto inicial de datos en tres grupos, conjunto de entrenamiento, validación y prueba.

No hay consensos en cómo dividir una muestra, sin embargo existe claridad en que deben ser más abundantes los datos de entrenamiento que los de prueba, y que, esa mayor disponibilidad debe ser usada para aprender los patrones presentes en los datos. La mayoría de los estudios prácticos aconsejan una proporción que va desde 70 al 90% para el conjunto de entrenamiento y validación, y por contraposición, entre 30 y 10% para el conjunto de prueba. Granger (1993) sugiere que para modelos no lineales se deje al menos el 20% de los datos en el conjunto de prueba; en cambio Hopcroft (1993) recomienda que al menos 10 datos integren dicho conjunto.

Así, en el presente trabajo la muestra se subdivide en los tres subconjuntos, donde dicha subdivisión se realiza en función a la cantidad de predicciones que se quieren realizar. Por lo tanto, como se quieren realizar 12 predicciones y evaluar las mismas, se dejan de prueba los últimos 12 períodos disponibles. Esto hace una proporción aproximada de 92% y 8% para el conjunto de entrenamiento-validación y el de prueba, respectivamente. Luego, el 92% es subdividido, dejando un 5% de los mismos como datos de validación (7 períodos en validación y 136 en el de entrenamiento). Esta división se realiza con el objetivo de que los resultados que se obtengan coincidan con los períodos predichos del VECM.

3.3.5 Entrenamiento de la ANN

Entrenar una red neuronal para conocer los patrones presentes en los datos, implica un proceso iterativo, en el que se busca minimizar una función de costo. Ésta utiliza información de la comparación entre la salida de la red y el/los verdadero/s valor/es de la variable/s respuesta/s ó *targets*. El objetivo es encontrar los valores de los pesos o ponderaciones que conectan las neuronas y que minimizan globalmente la función de error o costo. Si estos patrones encontrados por medio del proceso iterativo no provienen de un modelo sobre ajustado, entonces son una buena generalización y permiten obtener buenas predicciones por fuera de la muestra.

El método de aprendizaje más utilizado en el área económica, llamado retro propagación o *Backpropagation (BP)*, utiliza un algoritmo de entrenamiento iterativo. Éste ajusta los pesos de modo de generar los valores mínimos de la función de costo. Sin embargo, el proceso no garantiza encontrar dichos mínimos globales, ya que pueden existir mínimos locales en los cuales el proceso de estimación queda estancado.

A continuación se muestra el funcionamiento del *BP*, y los parámetros a tener en cuenta para disminuir las probabilidades de encontrar mínimos locales.

Algoritmo de Aprendizaje Backpropagation

Este algoritmo, desarrollado por Rumelhart, Hinton y Williams (1986), es el más utilizado en aplicaciones de ANN, particularmente en series de tiempo.

Siguiendo la exposición de Maciel y Ballini (2010), y de Ripley (2005), el proceso de errores de retro propagación consiste en la realización de dos fases que transcurren a

través de las diferentes capas de una ANN, a saber, la fase hacia adelante y la fase hacia atrás. En la primera, y a partir de un vector de datos de entrada, la información se propaga a lo largo de la red utilizando para ello unas ponderaciones fijas. Por medio de éstas, se produce la salida de la red como respuesta de la información de entrada propagada a lo largo de la red. Durante ese proceso los pesos están fijos. En la segunda fase, es decir, hacia atrás, los pesos son ajustados en concordancia con cierta regla de corrección del error, empezando desde la salida hacia la entrada de la ANN. Se ajustan los pesos de la última capa proporcionalmente al error; se pasa luego, a la capa anterior con una retro propagación del error, ajustando los pesos y continuando con este proceso hasta llegar a la primera capa. El procedimiento completo es llamado algoritmo de aprendizaje *stochastic gradient descent backpropagation*, que se explica su funcionamiento a continuación.

Suponiendo una ANN con n neuronas en la capa de entrada, m neuronas en la capa de salida, y s neuronas en una única capa oculta. Dados unos pesos iniciales aleatorios para las diferentes conexiones, y con los valores de las neuronas de entrada se procede a efectuar la primera fase: calcular la salida de cada capa, hasta llegar a la final/salida de la red. A saber,

a) La salida de la j -ésima neurona oculta es:

$$\varphi_j = f_h \left(\left(\sum_{i=1}^n w_{ij} x_i \right) + w_{(n+1)j} \right), \quad (3.31)$$

donde en total hay s neuronas ocultas.

b) La salida de la u -ésima neurona de salida es:

$$\hat{y}_u = f_o \left(\left(\sum_{j=1}^s w_{ju} \varphi_j \right) + w_{(s+1)u} \right), \quad (3.32)$$

de un total de m neuronas de salida.

Luego se comienza la segunda fase, calculando los términos de error para todas las neuronas y actualizando los pesos de la siguiente manera,

a) En el caso de los correspondientes a los de las neuronas de la capa de salida, el ajuste está dado por:

$$\Delta w_{ju} = \eta \delta_u \varphi_j, \quad (3.33)$$

con $\delta_u = (y_u - \hat{y}_u) \left(\frac{\partial f_o}{\partial w_{ju}} \right)$ para $j \in [1, \dots, s] \wedge u \in [1, \dots, m]$, y η es la tasa de aprendizaje.

b) En el caso de las neuronas ocultas, el ajuste es:

$$\Delta w_{ij} = \eta \delta_j x_i, \quad (3.34)$$

donde $\delta_j = \left(\frac{\partial f_h}{\partial w_{ij}} \right) \sum_{u=1}^m \delta_u w_{ju}$, es decir, que el error en las capas ocultas depende de todos los términos de error de la capa de salida. De aquí el nombre de propagación del error hacia atrás.

Una vez actualizados los pesos, se vuelve a la primera fase hacia adelante, continuando de manera iterativa este proceso hasta que el término de error total sea mínimo. Dicho error total, está definido por

$$Err(\mathbf{w}) = \sum_m \|\mathbf{y} - \hat{\mathbf{y}}\|^2. \quad (3.35)$$

Para poner en marcha el proceso de estimación de los parámetros, se deben definir, el número de iteraciones máximo y la tasa de aprendizaje.

Número de Iteraciones Máximo

Una de las cuestiones claves es el número máximo de iteraciones establecido en el proceso, ya que un punto de parada insuficiente puede llevar a un punto del proceso que no es mínimo.

Hay dos posturas con respecto a este tema. La primera, destaca el peligro de quedar atrapado en un mínimo local y la dificultad de encontrar un mínimo global (Masters, 1993). Se aconseja parar sólo cuando no hay disminuciones en la función del error, después de haber seleccionado los pesos de inicio de manera aleatoria. Ese punto es consi-

derado el de convergencia. La segunda postura, analiza la evolución de la capacidad de generalización dentro y fuera del conjunto de entrenamiento. Ese proceso debe ser repetido varias veces para ver si se ve modificada dicha evolución. Sin embargo, ninguna de las dos posturas asegura que el punto que se haya encontrado sea un mínimo global.

En el presente trabajo se sigue la segunda postura. En cada modelo se establecen diferentes números máximos de iteraciones y se analiza la evolución de los errores tanto en el conjunto de entrenamiento como en el de validación, con el objetivo de ver si los mismos tienden a un punto estacionario del error lo más cercano a cero.

Tasa de Aprendizaje

El objetivo del algoritmo de *BP* es encontrar las ponderaciones que minimicen el error total. En las ecuaciones (3.33) y (3.34) se involucra la tasa de aprendizaje. Ésta es una constante de proporcionalidad que determina el tamaño del cambio de las ponderaciones.

Cuando la tasa de aprendizaje es muy elevada, la función de error cambia abruptamente sin mostrar mejoras a lo largo de los pasos de la iteración. Por el contrario, cuando la tasa es demasiado pequeña, el proceso de entrenamiento se vuelve demasiado lento. Los valores posibles de dicha tasa están en el intervalo $(0,1)$.

En el presente trabajo se utiliza una tasa de aprendizaje de 0.2. La misma es un parámetro *default* en el programa utilizado para la estimación, por lo tanto se decide no modificarlo y dejarlo tal cual está por defecto, ya que no genera problemas en el proceso de estimación, y constituye un valor usado comúnmente en la bibliografía asociada.

3.3.6 Importancia Relativa de las Variables Inputs

Estimados los parámetros de la ANN, estos pesos/ponderaciones que conectan los nodos pueden ser usados para describir las relaciones entre las variables de entrada y salida/s.

El método descrito por Garson (1991) identifica la importancia relativa de las variables explicativas sobre las variables respuestas en una ANN, descomponiendo los pesos del modelo. El método consiste en identificar todas las conexiones ponderadas entre los nodos de interés, es decir, conocer la transferencia existente desde el ingreso de la in-

formación hasta la salida, pasando por los nodos ocultos. Este procedimiento se realiza para todas las variables de ingreso.

Técnicamente, el algoritmo de Garson, calcula la importancia relativa de la siguiente forma (Díaz-González, Hidalgo-Dávila, Santoyo, & Hermosillo-Valdéz, 2013):

$$IR_z = \frac{\sum_{j=1}^s \left(\frac{|w_{zj}| |u_j|}{\sum_{i=1}^n |w_{ij}|} \right)}{\sum_{i=1}^n \sum_{j=1}^s \left(\frac{|w_{zj}| |u_j|}{\sum_{i=1}^n |w_{ij}|} \right)}, \quad (3.36)$$

Donde el subíndice z corresponde a la variable de entrada de la cual se desea conocer su importancia relativa; s es el número total de neuronas ocultas; j es la j -ésima neurona oculta; n es el número total de entradas; $w_{..}$ y u , son los pesos desde la entrada a neuronas intermedias y desde las neuronas intermedias a las salidas, respectivamente.

Este algoritmo es usado para conocer la importancia relativa de las variables de entrada definidas en cada modelo.

3.3.7 Predicción con ANN

En este trabajo se diseña una ANN para predecir datos de series de tiempo, a saber, el precio pagado al productor lácteo a partir de información de entrada, compuesta por otras series de tiempo y de la variable de salida.

Con el objetivo de hacerlos comparables a los resultados predichos del VECM, se tienen en cuenta las siguientes cuestiones:

- a) Utilizar la variable de interés predictivo como salida de la ANN.
- b) Predecir con un formato similar a los procedimientos de predicción del VECM, es decir: (a) por un lado, usar los valores predichos en cada paso para predecir el siguiente hasta completar los 12 meses, es decir, realizar predicciones con Modelo Estático; (b) incorporar en cada paso de estimación, información pasada real, y re-estimar los

parámetros del modelo, actualizándolo con la información añadida; a saber, predicciones con Modelo Dinámico. Aquí también sólo se exponen los parámetros del primero.

c) Utilizar la misma división de la base de datos que en el VECM, y por tanto, predecir el mismo conjunto de períodos.

3.4. Comparación de los Modelos

Las medidas de ajuste de los modelos de predicción son relevantes cuando se necesita decidir qué metodología es la más adecuada para predecir los valores futuros de la variable de interés. En principio, un modelo es mejor predictor que otro si sus predicciones están más cercanas a los verdaderos valores de la variable predicha; situación que puede ser conocida si se realizan predicciones *ex-post*, es decir, cuando se deja de lado cierto conjunto de datos en la estimación de los parámetros para luego predecir con los resultados los valores dejados de lado. En la presente tesis se diseñan los modelos de modo de dejar cierta cantidad de datos afuera de la estimación para poder realizar dichas comparaciones, y por consiguiente evaluar la performance de las metodologías.

En general, la evaluación de la performance predictiva no es única. De hecho existen diferentes medidas que se pueden utilizar en dicha evaluación. La mayoría de ellas tiene en cuenta el error cometido entre la predicción y el verdadero valor. En la presente tesis se calculan tres medidas de performance de las predicciones de cada modelo:

1. La *Raíz del Error Cuadrático Medio* o *Root Mean Squared Error (RMSE)*, definida por la ecuación (3.37):

$$RMSE_i = \sqrt{\frac{\sum_{l \in L} (y_{il} - \hat{y}_{il})^2}{l}}. \quad (3.37)$$

2. La *Desviación Media Absoluta* o *Mean Absolut Deviation (MAD)*:

$$MAD_i = \frac{\sum_{l \in L} |y_{il} - \hat{y}_{il}|}{l}. \quad (3.38)$$

3. El *Máximo Porcentaje de Error* o *Maximum Percent Error (MPE)*:

$$MPE_i = \frac{100}{l} \sum_{l \in L} \frac{|y_{il} - \hat{y}_{il}|}{y_{il}}. \quad (3.39)$$

En todos los casos i hace referencia a la metodología, y l son las observaciones pertenecientes al conjunto de prueba L utilizado.

Por tanto, en la presente tesis se comparan el poder de predicción de los modelos, de modo de determinar qué modelo es el más adecuado para predecir dicho conjunto de datos.

Capítulo 4

Datos, Variables y Procedimientos

4.1. Introducción

Para completar la primera parte del marco teórico y metodológico, en el presente capítulo se presenta una descripción de los datos y variables que se utilizan en el análisis. Adicionalmente, se detallan los procedimientos que se llevan a cabo para aplicar las metodologías presentadas en el capítulo anterior, y su implementación computacional con el software correspondiente.

4.2. Variables

El precio pagado al productor lechero es la variable que interesa explicar y que se estima¹³ mediante los diferentes modelos. Dicho precio es determinado dentro de un sector económico, viéndose influenciado por la información que generan otros eslabones de la misma cadena. Por tanto, se utilizan otras variables del sector como información relevante para la modelización del precio pagado al productor. Estas variables son de tipo económicas y productivas, a saber: las que representan a los precios en el resto de los eslabones del sector, es decir, al eslabón mayorista, de exportación y minorista; las cantidades mensuales producida de leche cruda, introduciendo también información sobre la estacionalidad en los precios pagados al productor primario por la leche cruda.

Tanto del precio pagado al productor primario como la cantidad de leche cruda producida son variables que están disponibles. El resto, es decir, los otros precios del sector así como la variable de estacionalidad son construidas a fin de ser utilizadas en la presente tesis. A continuación se expone la forma en que fueron construidas.

Dada la naturaleza de lo producido por el sector, tanto los precios minoristas, mayoristas como de exportación no están disponibles como único precio, sino que son publica-

¹³ Entre otras variables para el caso del VECM.

dos por productos. Con el objetivo de solucionar esta problemática de multiplicidad de productos y obtener un resumen de la información de los distintos eslabones, se procede a construir un índice de precios de un mix ponderado para cada eslabón que posea dicha multiplicidad de productos. Se escribe una versión generalizada de los tres eslabones que tienen esta característica, ya que metodológicamente, los tres fueron calculados de igual manera. Entonces, el precio mix del eslabón J es:

$$P_J^{Mix} = \sum_{i \in J} b_{ij} \cdot P_{ij}, \quad (4.1)$$

donde b_{ij} es la ponderación de i -ésimo producto en J , para J que puede ser Minorista, Mayorista (o producción industrial) o Exportación; y P_{ij} es el precio de i -ésimo producto en J . A su vez, la ponderación del producto i en el precio mix J (b_{ij}) es,

$$b_{ij} = \frac{h_{ij}}{H_J}, \quad (4.2)$$

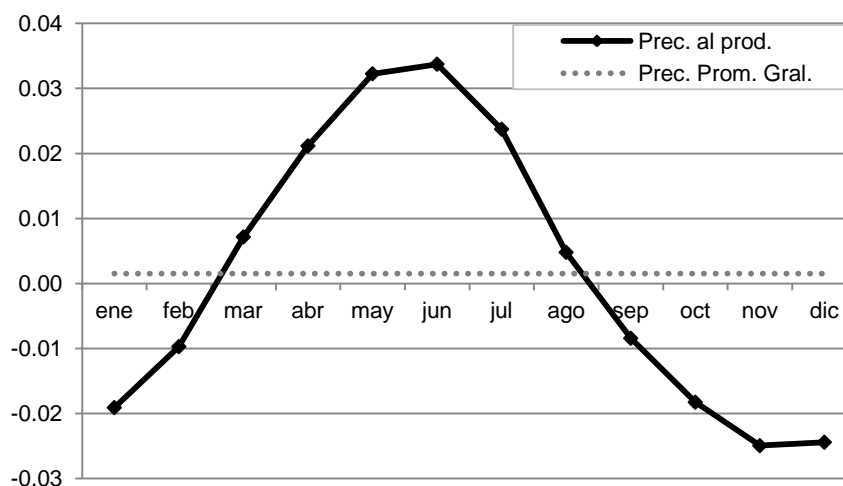
donde h_{ij} es la cantidad consumida, producida o exportada (según el eslabón J) del producto i ; H_J es la cantidad total consumida, producida o exportada (según el eslabón J), que corresponde a la suma total después de haber puesto todos los productos en una misma unidad de medida¹⁴.

Por otra parte, la variable de estacionalidad es construida a partir del análisis de estacionalidad de los precios al productor lácteo, una vez que a estos últimos se les saca tendencia. En la figura 4.1 se observa el comportamiento promedio mensual de los precios pagados al productor, y su relación con el precio promedio general para todo el período bajo análisis. Se puede ver que, en promedio, los precios mensuales se encuentran por encima del promedio general en el período marzo-agosto; y por debajo del promedio general en el período septiembre-febrero. Ello deviene principalmente de las características productivas del sector.

¹⁴ Dado que los índices de precios de los niveles mayorista, minorista y de exportación están medidos en \$/kg, el único producto que está medido en litros es la leche fluida. El coeficiente de conversión medido en lts/kg es de 1.036.

Con el objetivo de tener en cuenta dicha estacionalidad se decide crear una variable *dummy*. La misma indica estacionalidad alta en el precio pagado al productor tambero: la variable es igual a uno para los meses que componen el período marzo-agosto; e igual a cero para el resto de los meses (septiembre-febrero).

Figura 4.1. Estacionalidad del Precio Pagado al Productor.



Fuente: Elaboración propia a partir de datos de MAGyP.

Concluyendo, las variables que se utilizan en el presente trabajo son: a) los precios mensuales pagados al productor lácteo (P_{prod}), medidos en \$/lt.; b) los precios mensuales minorista mix lácteos ($P_{\text{min}}^{\text{mix}}$), medidos en \$/kg.; c) los precios mensuales mayorista mix lácteos ($P_{\text{may}}^{\text{mix}}$), medidos en \$/kg.; d) los precios de exportación mix lácteos ($P_{\text{exp}}^{\text{mix}}$), medidos en u\$/kg.; e) las cantidades (lts.) producidas mensuales de leche cruda (Q_{lts}); y f) una indicadora de la estacionalidad alta en los precios al productor (Alta.precios).

La periodicidad de todas las variables es mensual. Las mismas (excepto la de estacionalidad) son transformadas en logaritmo, estos son, $lp.prod$, $lp.min$, $lp.may$, $lp.exp$ y $lq.litros$, respectivamente. Los motivos por los que se hace dicha transformación, son varios. En primer lugar, es una transformación generalmente usada en economía, y permite analizar las estimaciones en términos de elasticidades. En segundo lugar, cuando se estima por medio del VECM utilizando las variables sin transformar, los residuos de los modelos no cumplen con los supuestos básicos para que los resultados obtenidos sean confiables. En tercer lugar, las variables así trabajadas no necesitan normalización para realizar las ANN. Por tanto, se trabaja con las variables transformadas por medio de la

función logarítmica, y las predicciones que se realizan son devueltas a su unidad de medida original, dado que el interés está centrado en la predicción de los valores reales de la variable de interés, así como evaluar y comparar la eficiencia de los modelos.

Por otra parte, en Argentina no es trivial el tema de la inflación. Se han buscado índices para corregir por inflación, sin embargo, debido a los problemas de credibilidad de los mismos se ha decidido trabajar las diferentes series de tiempo de manera original, sin deflactarlas, para no introducir mayor error en dichas series.

4.3. Fuentes de los Datos

Las bases provienen de distintas fuentes, donde la principal es el Ministerio de Agricultura, Ganadería y Pesca de la Nación Argentina, sección Lechería (MiniAgri). Desde la página oficial (MiniAgri, 2015) se obtuvo información sobre a) precios mayoristas por productos, con periodicidad mensual; b) cantidades destinadas al consumo interno de lácteos, mensuales por productos; así como c) la cantidad de leche cruda producida mensualmente en la Argentina.

En particular, los precios al productor primario lechero se obtienen tanto de la Secretaría de Lechería de la Provincia de Santa Fe y del MiniAgri. El motivo por el que se utilizan datos de la provincia de Santa Fe es que la serie de precios al productor primario de MiniAgri tienen una longitud menor. Recién a partir de noviembre de 2009 comenzaron a publicar información sobre pagos por litro de leche a nivel nacional.

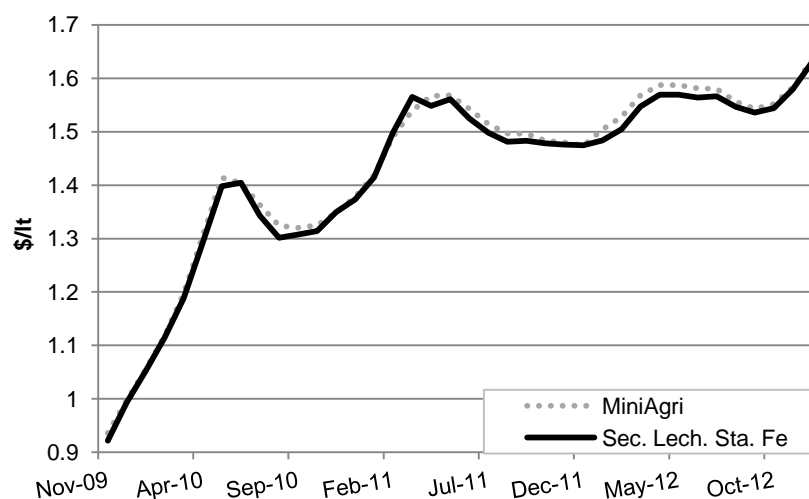
En la figura 4.2 se muestran los precios pagados al productor de ambas fuentes, de MiniAgri y de la provincia de Santa Fe, en el período coincidente. Se puede observar que los precios al productor de la provincia de Santa Fe no difieren significativamente de los correspondientes de la Nación y que tienen una evolución prácticamente idéntica.

El motivo radica en la importancia de la provincia de Santa Fe en la evolución y desempeño de la lechería nacional. Por lo tanto, se procede a hacer un empalme entre las series de estas dos fuentes a modo de obtener una serie de precios con mayor longitud.

Por otra parte, los precios minoristas mensuales por productos, son obtenidos del Instituto de Estadísticas y Censos de la Provincia de Santa Fe (IPEC). A nivel nacional no se dispone de datos sobre los precios minoristas, por lo tanto, teniendo en cuenta que la provincia mencionada es una de las más importantes en el consumo nacional, se asume

que los precios de la misma se comportan como lo hacen a nivel nacional. Claramente, ésta resulta una simplificación importante; sin embargo, ante la escasez de información, se propone como paliativo ante dicha situación.

Figura 4.2. Precios Pagados al Productor con Fuente Nacional y Provincia de Santa Fe.



Fuente: Elaboración propia con datos de MiniAgri y Secretaría de Lechería de Santa Fe.

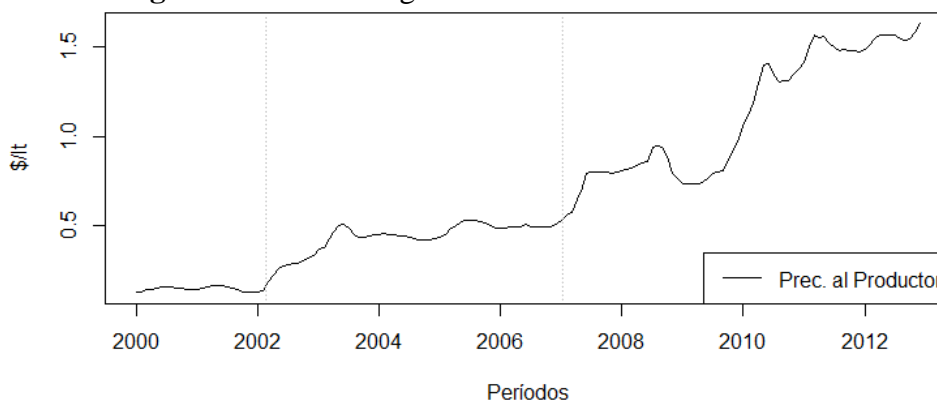
La series de precios de exportación por productos lácteos, así como las cantidades exportadas correspondientes, son obtenidos de bases de datos de Aduana (Penta-Transaction, 2014).

Dependiendo del origen de los datos, algunas de las series son más extensas que otras. De todos los datos disponibles se logra completar una base con una extensión desde enero del 2000 a diciembre del 2012. A partir de dicho período, no todas las series están completas. Sin embargo, la extensión final del período en cada modelo varía en función de la naturaleza del mismo, teniendo en cuenta la reducción por la incorporación de rezagos.

4.4. Descripción de las Variables.

El precio al productor primario (P_{prod}) es la variable a predecir en la presente tesis. En la figura 4.3 se muestra la evolución de la misma. A partir de la crisis del 2001-2002 los precios de la leche cruda han tenido una evolución positiva, que resulta más marcada a partir de 2007. A fines de 2009 y principios de 2010, los precios han tenido una escalada importante.

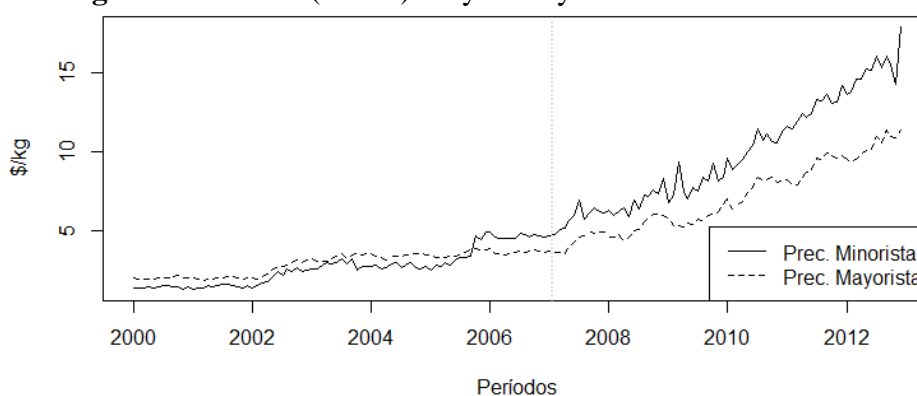
Figura 4.3. Precio Pagado al Productor. Período 2000-2012.



Fuente: Elaboración propia con datos de MiniAgri y Secretaría de Lechería de Santa Fe.

En cuanto a los precios mayoristas y minoristas del mix ponderado, la evolución también ha sido positiva y con similares comportamientos a los de la leche cruda (figura 4.4). La tasa de crecimiento de los precios minoristas ha sido más pronunciada que la de los mayoristas, sobre todo a partir de 2007. Sin embargo, antes de 2006, la evolución es muy similar entre ellos, con una pendiente similar y bastante estable.

Figura 4.4. Precio (Índice) Mayorista y Minorista. Período 2000-2012.

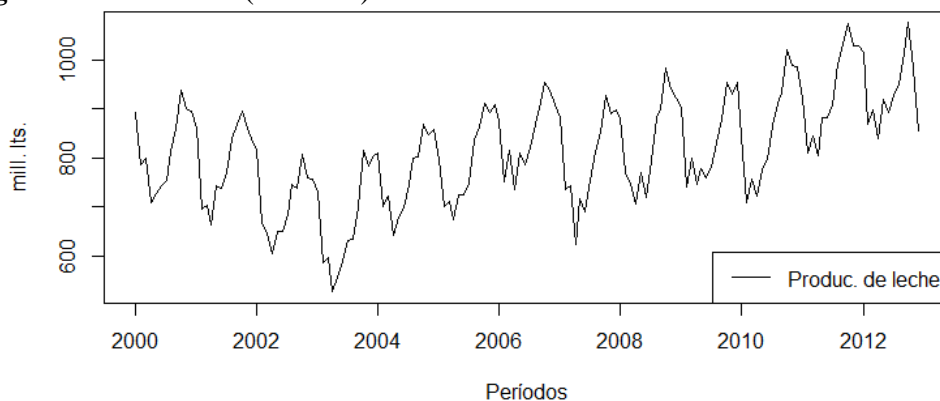


Fuente: Elaboración propia con datos de IPEC y MiniAgri.

La cantidad de litros de leche cruda producida es otra de las variables consideradas relevantes, ya que la escasez relativa en el corto plazo puede generar fricciones en los precios de los bienes lácteos.

En la figura 4.5 se muestra la evolución de la cantidad de leche producida en Argentina, en el período de análisis. Se puede observar un comportamiento marcadamente estacional, pero con una clara tendencia positiva a partir de 2003. Antes de 2003, la tendencia de las cantidades producidas de leche cruda era negativa.

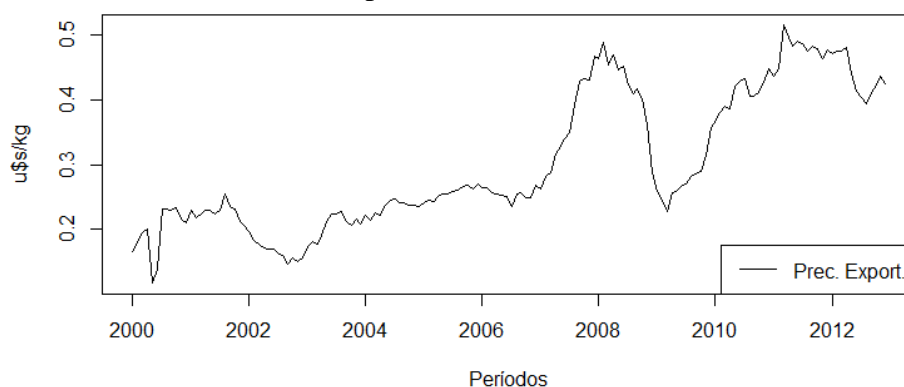
Figura 4.5. Cantidad (mill. lts.) de Leche Cruda Producida. Período 2000-2012.



Fuente: Elaboración propia a partir de datos de MiniAgri.

Por otra parte, los precios de exportación son determinados afuera del país, por lo que suele ser considerados una variable exógena en la determinación de los precios locales. Esto es probado mediante el test de exogeneidad débil.

Figura 4.6. Precios (Índice) de Exportación de un Mix Lácteo. Período 2000-2012.



Fuente: Elaboración propia a partir de datos de Aduana.

En la figura 4.6 se puede ver una evolución bastante parecida a la de los precios pagados al productor primario, excepto en el período 2007-2009 en donde se aplicaron precios de corte a los exportadores.

4.5. Implementación y Software

La tesis se desarrolla por completo con el programa R (R-Project, 2015), a saber, tanto el ordenamiento, exploración, estimaciones y verificaciones; sólo algunos gráficos se realizan en Excel. El primero permite la utilización sin licencia alguna, y además existen una gran cantidad de paquetes posibles para ser utilizado para un tema específico: no sólo se han encontrado varios paquetes para la modelación mediante VECM y para

ANN, sino que también variadas opciones para realizar el análisis descriptivo y de verificación/evaluación de modelos. Cada uno de esos paquetes tiene características particulares, permitiendo la comparación entre ellos y dando la libertad de seleccionar en función de los intereses particulares que se tenga.

Particularmente, los paquetes que se utilizan han sido elegidos por su flexibilidad y pertinencia sobre el problema presente; por el diseño e información brindada en la salida del programa después de cada estimación, y por la claridad del material de ayuda y de bibliografía citada.

Para la parte descriptiva se usa paquetes básicos de gráficos, por ejemplo, el *stats* (R Core Team, 2015) que es utilizado para graficar series de tiempo; el *graphics* (R Core Team, 2015) para hacer gráficos descriptivos.

De todos los paquetes disponibles para el análisis de series de tiempo y la estimación del VECM, se han elegido tres de ellos, estos son, *tseries* (Trapletti, & Hornik, 2015), *vars* (Pfaff, & Stigler, 2013b) y *urca* (Pfaff, & Stigler, 2013a). El primero contiene uno de los test de estacionariedad, a saber el *KPSS*. También contiene otros test y utilidades, sin embargo, ya están contenidos en los otros paquetes mencionados, y son utilizados por medio de ellos. Algunas de las utilidades del paquete *vars* son utilizados por el *urca* es decir, sus acciones son llamadas en este último paquete, pero no es el *urca* el que realiza la actividad.

Mediante estos últimos dos paquetes se realiza el análisis sobre la especificación de VECM, así como la estimación, pruebas y verificación. Una vez obtenido los resultados de la modelización, se realizan las predicciones, y a partir de ellos se calculan las medidas de performance de las estimaciones. Estas medidas no están contenidas en los paquetes, por lo que dichas funciones son construidas, comparando los valores observados y los predichos.

Por otro lado, también para la modelación de las ANN existen una gran variedad de paquetes utilizables. En particular, en el presente trabajo se utiliza principalmente el paquete *RSSNS* (Bergmeir, & Benitez, 2015). El principal motivo, es que permite realizar perceptrones multicapa explícitamente. Sin embargo, se incorporan ciertas utilidades de los paquetes *NeuralNetTools* (Beck, 2015) y *plyr* (Wickham, 2011), para realizar el test de importancia relativa de variables y para dividir el conjunto de datos, respectiva-

mente. El paquete *RSSNS* es actualmente el más utilizado en R para realizar modelaciones con ANN, ya que tiene una alta flexibilidad de modelación permitiendo definir parámetros, funciones de transferencia, función de aprendizaje, entre otras de las cuestiones fundamentales de la definición de una red. El presente paquete tampoco incluye medidas de performance. Las funciones también son construidas de la misma forma que en el VECM de modo que sean comparables.

PARTE II. **RESULTADOS**

Capítulo 5

Predicción del Precio al Productor Primario

5.1. Introducción

En el presente capítulo se muestra el proceso realizado para modelar las relaciones entre las variables económicas de series de tiempo bajo las diferentes metodologías según lo explicado en el Capítulo 3; así como la comparación de las predicciones obtenidas, por cada uno de ellos, del precio al productor primario de leche cruda argentina.

Para la modelación de series de tiempo multivariadas no estacionarias se utiliza un VECM tal como se detalla en la Sección 3.2. Primero se constata el orden de integración de las series mediante diferentes test, y luego, se constata la existencia de componentes determinísticos en las relaciones y la cantidad de rezagos, para así, poder estimar los parámetros que permitan realizar las predicciones.

Por otro lado, se muestra el proceso seguido en la modelación mediante ANN explicitada en la Sección 3.3: la definición de la arquitectura, lo que implica reconocer las variables de entrada y salida, la cantidad de capas ocultas y sus correspondientes nodos, así como la función no lineal que transforma las relaciones; y luego, los parámetros estimados.

Luego de realizar ambas modelizaciones, se procede a pronosticar la variable de interés para luego comparar cuál de ellas es la más correcta a partir de lo especificado en la Sección 3.4. Allí se prevén dos tipos de pronósticos. Uno que se denomina Estático, y corresponde a predicciones sobre el comportamiento de la variable en un año, es decir, donde se involucra más de un período, bajo el supuesto de que no se tiene ningún tipo de información sobre dicho período a predecir. El otro tipo de pronóstico, el Dinámico, donde mediante éste se predice de manera paulatina (es decir, mes a mes) el comporta-

miento de la variable durante los doce meses, pero realizando ajustes sobre los parámetros a partir de la incorporación de información mensual actualizada.

Para predicciones de mediano y largo plazo, el más adecuado es el formato *estático*. Si, en cambio, se quiere realizar predicciones de corto plazo, los pronósticos dinámicos son mejores. En la presente tesis se analiza primeramente el primer formato de predicción, mostrando los resultados de la modelización y estimación. En la segunda tipología de predicción, se muestran sólo los resultados de dicho proceso, pero no los resultados de los parámetros, ya que resultan engorrosos y no agregan información sobre el objetivo de la tesis.

Por tanto, el presente capítulo se organiza de la siguiente manera. En la próxima sección se muestra el proceso que se realiza en la estimación de los parámetros del VECM; luego, se muestra el correspondiente para la metodología ANN. En la Sección 5.4 se comparan las predicciones realizadas por cada uno de los modelos sobre el precio pagado al productor de leche cruda, tanto en su versión estática como dinámica. Y por último, se realiza una síntesis y se concluye sobre los resultados obtenidos.

5.2. Estimación del VECM

5.2.1 Pruebas de Estacionalidad de las Series

Prueba de Dickey Fuller Aumentada

En la tabla 5.1 se muestra tanto el número de rezagos óptimo que se incorporan en la prueba *ADF* de cada una de las variables (en nivel y diferencias) según el *AIC*, como también los estadísticos obtenidos y los valores críticos de los mismos.

Se observa que todas las variables medidas en nivel, excepto el logaritmo del precio al productor, poseen raíz unitaria a un nivel 5% y por ende, no son estacionarias. Sin embargo, el logaritmo del precio al productor no es estacionario al nivel 1%; por lo que se decide realizar la prueba también para dicha variable en diferencia.

A partir del estadístico *ADF* para las mismas variables en diferencia, se concluye que todas las variables consideradas poseen un orden de integración unitario, ya que al constatar presencia de raíces unitarias en las variables diferenciadas, se rechaza la hipótesis nula a un nivel del 5%.

Tabla 5.1. Prueba de Estacionariedad *ADF*.

Variables	Número de Rezagos Seleccionados	Tau observado	Tau crítico ¹⁵ (5%)	Tipo de serie
lp.prod	3	-2.4327	-1.95	Estacionaria *
Diff.lp.prod	2	-5.4892	-3.43	Estacionaria
lp.may	1	-0.4827	-2.88	No Estacionaria
Diff.lp.may	1	-10.3155	-3.43	Estacionaria
lp.min	7	0.485	-2.88	No Estacionaria
Diff.lp.min	6	-6.3633	-3.43	Estacionaria
lp.exp	2	-1.0986	-1.95	No Estacionaria
Diff.lp.exp	1	-5.8421	-3.43	Estacionaria
lq.litros	16	0.5465	-1.95	No estacionaria
Diff.lq.litros	15	-2.8627	-1.95	Estacionaria

Nota: * No estacionaria al 1% de nivel de significancia

Fuente: Elaboración propia.

Prueba KPSS

En la tabla 5.2 se muestran los resultados de los estadísticos observados así como la conclusión del test a un nivel del 5%, teniendo en cuenta un modelo con constante (Parte A) por un lado y, con constante y tendencia (Parte B), por otro.

Tabla 5.2. Pruebas de Estacionariedad *KPSS*.

Rezagos	Variables	Parte A: <i>CONSTANTE</i>		Parte B: <i>CONSTANTE Y TENDENCIA</i>	
		Tau observado	Tipo de serie (al 5%)*	Tau observado	Tipo de serie (al 5%)**
4	lp.prod	2.7125	No Estacionaria	0.2276	No Estacionaria
	Diff.lp.prod	0.063	Estacionaria	0.0408	Estacionaria
	lp.may	2.9138	No Estacionaria	0.0644	Estacionaria
	Diff.lp.may	0.0433	Estacionaria	0.0285	Estacionaria
	lp.min	2.7718	No Estacionaria	0.2763	No Estacionaria
	Diff.lp.min	0.1007	Estacionaria	0.0379	Estacionaria
	lp.exp	2.1821	No Estacionaria	0.1575	No Estacionaria
	Diff.lp.exp	0.0511	Estacionaria	0.0439	Estacionaria
	lq.litros	1.0064	No estacionaria	0.1336	Estacionaria
	Diff.lq.litros	0.0473	Estacionaria	0.0154	Estacionaria
13	lp.prod	1.0562	No Estacionaria	0.1111	Estacionaria
	Diff.lp.prod	0.0666	Estacionaria	0.0434	Estacionaria
	lp.may	1.1181	No Estacionaria	0.0414	Estacionaria
	Diff.lp.may	0.0632	Estacionaria	0.0421	Estacionaria
	lp.min	1.0805	No Estacionaria	0.1338	Estacionaria
	Diff.lp.min	0.1553	Estacionaria	0.0622	Estacionaria
	lp.exp	0.9069	No Estacionaria	0.0829	Estacionaria
	Diff.lp.exp	0.0424	Estacionaria	0.0367	Estacionaria
	lq.litros	0.662	No Estacionaria	0.1227	Estacionaria
	Diff.lq.litros	0.2343	Estacionaria	0.0845	Estacionaria

Notas: *El estadístico crítico es 0.463; ** El estadístico crítico es 0.146.

Fuente: Elaboración propia.

¹⁵ El valor crítico del *tau* de un modelo simple (sin tendencia y constante) y un nivel de confianza de 95% es de -1.95; para un modelo con constante es de -2.88; y el de un modelo con tendencia y constante es de -3.43.

Si los modelos correspondientes a cada una de las variables, en nivel, poseen una constante, sea con 4 o 13 rezagos, se concluye que las mismas no son estacionarias al nivel del 5%. Los estadísticos obtenidos para las mismas variables en diferencia, indican estacionariedad, por lo que se concluye que todas las variables $I(1)$.

Por otro lado, en el caso de que se incluya constante y tendencia en los modelos de cada una de las variables en nivel y, con cuatro rezagos, se puede concluir que casi todas las variables son estacionarias al nivel del 5%, excepto tres de ellas, a saber, el logaritmo del precio al productor, del precio minorista y el de exportación. Estas tres variables en diferencia son estacionarias, demostrando que son $I(1)$. Para el modelo con constante y tendencia, pero con 13 rezagos, el test concluye que todas las variables en nivel son estacionarias al nivel del 5%.

Los resultados de ambos test, a saber, el *ADF* y el *KPSS* (en sus diferentes versiones), no son coincidentes acerca de la estacionariedad de todas las series. La mayoría de las versiones del test *KPSS* concluye que todas las series son $I(1)$ a un nivel del 5%, excepto cuando se incluye en los modelos tendencia, constante y 13 rezagos. El *ADF* también arriba a similares conclusiones, excepto en la variable de interés predictivo, que sólo al nivel del 1% resulta $I(1)$. De todas formas, el hecho de que, en general la mayoría de las series sean no estacionarias justifica la modelización por medio de *VECM*.

5.2.2 Estructura del *VECM*

Variable Exógena - Prueba de Exogeneidad Débil.

Considerando que Argentina es un país tomador de precios con respecto a productos lácteos a nivel internacional, se tiene la hipótesis de que los precios de exportación no se relacionan con las otras variables que tienen que ver con el sector productivo interno de manera endógena. Para poder analizar dicha hipótesis se realiza el test de exogeneidad débil.

El resultado obtenido con dicho test indica que no se puede rechazar la hipótesis de exogeneidad débil de la variable correspondiente al precio de exportación de un mix de lácteos, ni siquiera a un nivel del 10%. Se obtuvo un p-valor de 0.7309 y un estadístico de 0.1183. Por lo tanto, tal como se esperaba y respondiendo a la hipótesis de forma, el

presente test confirma exogeneidad débil del logaritmo del precio de exportación, y entonces se la considera como variable exógena en el modelo.

Selección del Número de Rezagos del Modelo

Tal como se menciona en la metodología (Sección 3.2.2), se selecciona el número de rezagos teniendo en cuenta tres criterios de información, y en caso de no ser coincidentes, se realizan análisis de diagnóstico de los residuos de cada modelo, constatando normalidad, homocedasticidad y autocorrelación.

Tanto el *AIC* como el *HQ* sugieren la utilización de cuatro rezagos en el modelo VAR, mientras que el *SB*, tres. Por medio del análisis de diagnóstico de los residuos correspondientes a cada uno de modelos con los rezagos sugeridos por los criterios de información, se concluye que la estabilidad se logra con 4 rezagos sobre el VAR.

La selección de rezagos se realiza sobre el VAR y no sobre el VECM. Este último corresponde a una modelización de las variables en diferencia. Por lo tanto, los rezagos en el VECM es la cantidad de los del VAR menos uno. Así, la cantidad de rezagos seleccionada para el VECM es 3.

Prueba de Cointegración de Johansen

Seleccionada la cantidad de rezagos, se procede a efectuar la prueba de cointegración de Johansen. Siguiendo el procedimiento de la Sección 3.2.2 se concluye que se debe incorporar una tendencia en la relación de cointegración de las variables en diferencias, es decir, que tienen una tendencia cuadrática en niveles¹⁶.

Las pruebas de cointegración de Johansen se muestran en la tabla 5.3. Dichas pruebas fueron realizadas a partir de los contrastes λ_{traza} y λ_{max} .

Dado los resultados obtenidos, el test de cointegración de Johansen medido por el contraste de *traza* indica que, la hipótesis nula de que $rango(\Pi) = 0$, es rechazada al nivel del 5%. En cambio, la hipótesis nula de que la matriz de vectores de cointegración tiene un rango menor o igual a uno, no es posible rechazar al nivel del 5%. Por lo tanto, se concluye que existe un vector de cointegración entre las variables endógenas, donde el

¹⁶ Se debe agregar componentes determinísticos en la relación de cointegración dado que éstos resultan significativos y, además, su incorporación permite el cumplimiento de los supuestos sobre los errores del modelo, excepto el de normalidad.

logaritmo del precio de exportación es tomado como exógeno y se incluyen tres rezagos en el VECM.

Tabla 5.3. Test de Cointegración de Johansen, *traza* y *max*.

Prueba	H_0	H_1	$\hat{\lambda}_i$; observado	λ_i ; $\alpha=0.05$
λ_{traza}	$r=0$	$r>0$	74.60**	62.99
	$r\leq 1$	$r>1$	27.62*	42.44
λ_{max}	$r=0$	$r>0$	46.98**	31.46
	$r\leq 1$	$r>1$	20.45*	25.54

Notas: Al nivel del 5%, *No se rechaza H_0 ; ** Se Rechaza H_0 en favor de H_1
Fuente: Elaboración propia.

El test de cointegración medido por el contraste *max* arroja idénticos resultados que el de *traza*. Por lo tanto, existe suficiente evidencia a favor de la hipótesis de la existencia de un vector de cointegración entre las variables, y por tanto, una relación estable entre ellas a largo plazo.

5.2.3 Parámetros del VECM

El mejor VECM¹⁷ no restringido estimado posee una tendencia en la única relación de cointegración encontrada y, tiene tres rezagos en los ajustes de corto plazo. Las estimaciones de los parámetros se muestran en la tabla 5.4., y sirven para realizar las predicciones estáticas.

La cantidad de períodos disponibles para estimar los parámetros del VECM es de 144, es decir, desde enero del 2000 hasta diciembre del 2011. En cada uno de los modelos de las variables endógenas, se estiman 16 parámetros y se pierden 3 datos por rezagos, por tanto los grados de libertad en cada una de las ecuaciones es de 124.

Los coeficientes del vector de cointegración, una vez normalizados, indican la relación de largo plazo entre las variables endógenas. Así, sean $\hat{\beta}$ éstos coeficientes normalizados en el precio al productor lechero, se puede escribir dicha relación de largo plazo de la siguiente manera:

$$lp.prod_t = 0.33lp.may_t + 0.26lp.min_t - 3.36lp.litros_t + 0.011t \quad (5.1)$$

¹⁷ Los residuos del VECM son estacionarios, homocedásticos y no posee autocorrelación serial. Las pruebas se muestran en el Anexo. No se ha podido constatar normalidad de dichos residuos.

La ecuación (5.1) indica que en el largo plazo, ante un aumento del 10% en los precios mayoristas, los precios al productor primario se ven incrementados, en promedio, un 3.3% aproximadamente. A su vez, a largo plazo, un aumento del 10% en los precios minoristas se ve acompañado por un aumento promedio del 2.6% en los precios primarios, aproximadamente. Estos efectos resultaron significativos solamente al nivel del 10%.

Tabla 5.4. Estimaciones del VECM.

Vector de Cointegración				
	Coeficientes:			
	1			
lp.prod				
lp.may				-0.32995 (0.24242) •
lp.min				-0.26516 (0.24787) •
lq.litros				3.35858 (0.34725)***
Tendencia				-0.01107 (0.00403)***
Modelo de Corrección del Error (ECM)				
	Diff.lp.prod	Diff.lp.may	Diff.lp.min	Diff.lq.lts
ECT ¹	-0.04514 (0.01813)*	-0.07222 (0.01659) ***	0.044973 (0.03803)	-0.15953 (0.02781) ***
Constante	0.94583 (0.37859)*	1.51395 (0.34639)***	-0.92611 (0.79422)	3.27572 (0.58081)***
lp.exp	0.00839 (0.01201)	0.04412 (0.01098)***	-0.01579 (0.02518)	0.07639 (0.01842)***
Alta.precios	-0.03731 (0.01012)***	0.00211 (0.00926)	0.026 (0.02123)	0.05829 (0.01553)***
Diff.lp.prod.L1	0.462105 (0.08853)***	0.430427 (0.08099)***	0.37703 (0.18571)*	0.16613 (0.13581)
Diff.lp.may.L1	0.06933 (0.11284)	-0.43704 (0.10325)***	0.34182 (0.23672)	-0.41843 (0.17311)*
Diff.lp.min.L1	-0.05222 (0.04323)	0.00804 (0.03956)	-0.63079 (0.0907)***	0.15482 (0.06633)*
Diff.lq.lts.L1	-0.16042 (0.05912)**	0.31269 (0.05409)***	-0.222 (0.12401) •	0.32888 (0.09069)***
Diff.lp.prod.L2	0.20973 (0.09934)*	0.135491 (0.0909)	-0.030039 (0.20841)	0.05226 (0.1524)
Diff.lp.may.L2	-0.00341 (0.11396)	-0.36673 (0.10427)***	0.12069 (0.23906)	-0.60629 (0.17483)***
Diff.lp.min.L2	-0.05372 (0.05027)	0.05305 (0.046)	-0.23297 (0.10547)*	0.22965 (0.07713)**
Diff.lq.lts.L2	-0.15639 (0.0699)*	0.38232 (0.06396)***	-0.04826 (0.14665)	1.01556 (0.10724)***
Diff.lp.prod.L3	-0.11267 (0.09114)	0.02134 (0.08339)	0.26749 (0.19121)	-0.04026 (0.13983)
Diff.lp.may.L3	-0.05811 (0.10289)	0.05108 (0.09414)	0.34776 (0.21585)	-0.10682 (0.15785)
Diff.lp.min.L3	0.00925 (0.04352)	0.0638 (0.03982)	-0.01745 (0.09131)	0.1727 (0.06677)*
Diff.lq.lts.L3	0.06847 (0.06881)	0.04541 (0.06296)	-0.13976 (0.14435)	0.533453 (0.10556)
R ²	0.6479	0.5474	0.3987	0.6132
R ² Ajustado	0.6024	0.489	0.3212	0.5633
Estadístico F _{16,124}	14.26	9.374	5.14	12.29

Notas: Significativo al nivel del: • 10%, * 5%, ** 1%, *** 0.1%

¹Error Correction Term.

Fuente: Elaboración propia.

Por otra parte, la relación a largo plazo de los litros de leche cruda producida y los precios de ésta es muy importante. Los resultados indican que, a largo plazo, un aumento del 1% en las cantidades producidas a nivel nacional de leche cruda, va acompañado por una disminución del 3.4% en el precio que se le paga al productor primario de este producto.

Los términos de corrección del error (ECT) indican el ajuste de corto plazo que, conjuntamente con el modelo de corrección del error (ECM), permiten predecir la evolución de las variables endógenas.

El vector de ECT, particularmente, muestra cómo reacciona cada una de las variables ante una desviación de la relación estable que existe entre ellas en el largo plazo, donde dicha desviación puede provenir de cualquier shock que las afecte. Así, las estimaciones indican que tres de las cuatro variables endógenas son las que modifican su valor para volver al equilibrio de largo plazo cuando surge algún shock que afecta dicha estabilidad o equilibrio. Éstas son: el precio al productor primario, el precio mayorista y la cantidad de litros de leche cruda producida.

Las estimaciones de los parámetros de los 12 modelos que permiten realizar las predicciones dinámicas no se muestran por motivos expositivos. Sin embargo, tienen igual formato¹⁸ que el modelo estático presente, con parámetros diferentes en función a la información incorporada paulatinamente.

5.3. Estimación del ANN

5.3.1 Selección de las Variables Inputs

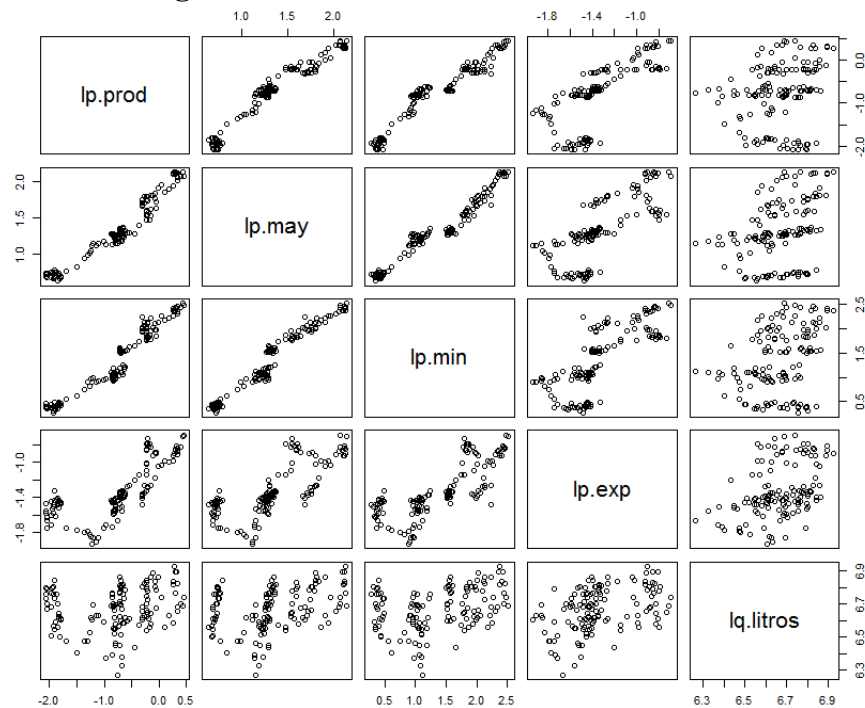
Existen numerosas formas de seleccionar las variables para ser usadas como *inputs* en una ANN; en la presente tesis se utilizan aquéllas relevantes desde la teoría económica y que están disponibles.

Se analizan las correlaciones de las variables utilizadas en el VECM, incluyendo en esta lista la exógena así como las endógenas utilizadas en esa oportunidad. En la figura 5.1 se muestran dichas correlaciones.

¹⁸ Es decir, los VECM tienen iguales variables endógenas, exógenas, cantidad de rezagos, y componentes determinísticos.

Tal como se observa en la figura 5.1, el logaritmo de precio al productor, el minorista y el mayorista están altamente correlacionados de manera positiva. Se puede observar que entre ellas la nube de puntos indica casi una línea recta sobre la diagonal. Utilizar una de estas tres variables como input resulta más que suficiente para capturar la información relevante; incorporarlas a todas generaría ineficiencias y produciría una red más compleja de lo necesario. Por lo tanto, de estas tres variables, sólo se utiliza el logaritmo del precio al productor, ya que es la de interés explicativo y, además se supone que existen efectos de inercia sobre este precio.

Figura 5.1. Correlaciones entre las Variables.



La correlación entre logaritmo del precio al productor y el logaritmo del precio de exportación es marcada, aunque las mismas no están tan correlacionadas como lo estaba el otro conjunto. Por tanto, el logaritmo del precio de exportación se considera también como variable input dentro del modelo.

A las tres variables que se seleccionan del análisis de correlación, se agrega la variable indicadora de los meses del año en que los precios suelen estar por encima del promedio, es decir, la variable Alta.precios.

5.3.2 Arquitectura de la ANN

Se selecciona una ANN con tres capas ya que, de todas las redes entrenadas, ésta es la que mejores resultados predictivos posee, con una estructura relativamente simple. Las tres capas son: - el *input layer* o capa de entrada; - *hidden layer* o capa oculta y, - *output layer* o capa de salida.

En la Sección 5.3.1 se indican cuáles son las variables de entrada tenidas en cuenta en el *input layer*. Dados los resultados del análisis de cointegración, se ha probado incorporar a las variables rezagadas en más de un período, pero los resultados predictivos no mejoran; por tanto se decide trabajar con sólo un rezago de las variables seleccionadas.

Por otro lado, en el proceso de diseño y búsqueda de la ANN final, se estimaron otras redes con múltiples nodos de salida, tantos como variable endógenas hayan sido utilizadas en el VECM. Sin embargo, no se producían resultados superiores en términos de las predicciones, ni sobre el conjunto de prueba ni sobre el de verificación. Por lo tanto, para el *output layer* se opta por un nodo de salida exclusivamente, ya que torna a la ANN de mayor parsimonia y menor complejidad.

La cantidad de capas ocultas así como los nodos correspondientes a éstas, se seleccionan en función de un proceso de prueba y error; quedando finalmente la cantidad de capas y nodos que mejor capacidad predictiva proporciona, y que a su vez, genere una estructura lo más sencilla posible. La ANN final seleccionada posee una capa oculta con dos nodos en ella.

Por otra parte, y en función a lo comentado en la Sección 3.3.3, se utilizan dos funciones de transferencias, a saber, logística para la capa oculta, y lineal en la de salida. Los parámetros son estimados por medio de un procedimiento supervisado y a través del algoritmo de backpropagation.

A continuación se muestran las estimaciones de la red entrenada.

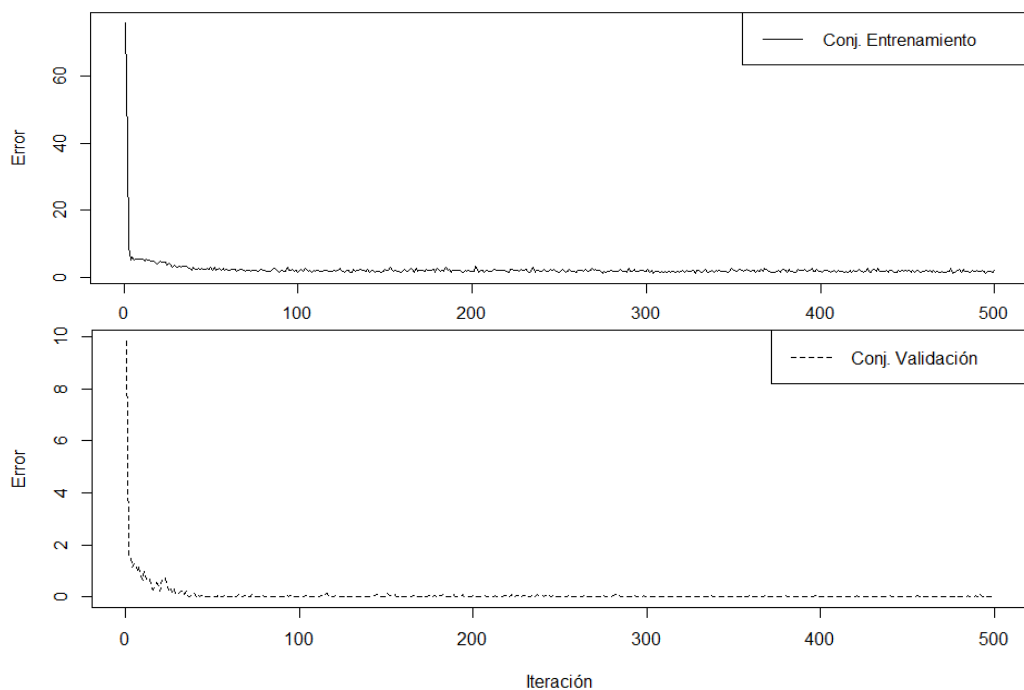
5.3.3 Entrenamiento y Verificación de la ANN

El conjunto de datos completo es dividido en función a lo mencionado en la Sección 3.3.4. El conjunto entrenamiento y el de verificación suman un total de 143 períodos, y el de prueba, 12.

Se realizaron 500 iteraciones en el proceso de estimación de los parámetros. En cada una de esas iteraciones se miden los errores sobre el conjunto de entrenamiento y sobre el de verificación, para ver si dichos errores tienen convergencia a cero (figura 5.2).

Los errores del proceso iterativo de la ANN entrenada tienen convergencia a valores muy pequeños. Antes de las 100 iteraciones ya se logran errores bajos cercanos a cero, tanto en el conjunto de entrenamiento como en el de validación; y a lo largo de los siguientes pasos iterativos, los errores no se disparan a valores muy alejados del cero, lo que muestra que se han logrado pesos que minimizan el error, aunque este procedimiento no garantiza que sea un mínimo global.

Figura 5.2. Errores del Proceso de Estimación de los Parámetros.



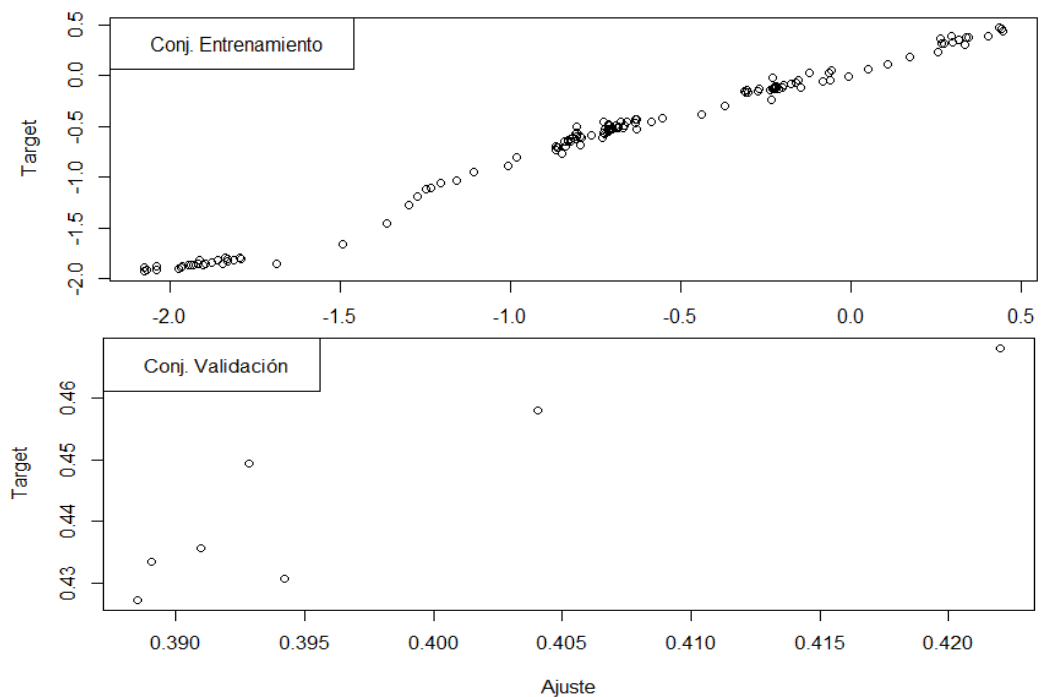
Fuente: Elaboración propia.

Con respecto a los valores ajustados del ANN sobre el conjunto de entrenamiento y el de validación, en la figura 5.3 se muestran los valores ajustados de cada conjunto versus los *targets* correspondientes.

Con respecto a los valores ajustados de la ANN sobre el conjunto de entrenamiento y el de validación, en la figura 5.3 se muestran los valores ajustados de cada conjunto versus los *targets* correspondientes.

Se puede observar una fuerte asociación entre los valores predichos de la red y los valores reales dentro del conjunto de entrenamiento. En el conjunto de validación, sólo hay siete observaciones, aún así se observa un buen ajuste de las estimaciones sobre estos datos.

Figura 5.3. Valores Ajustados versus Valores Predichos.



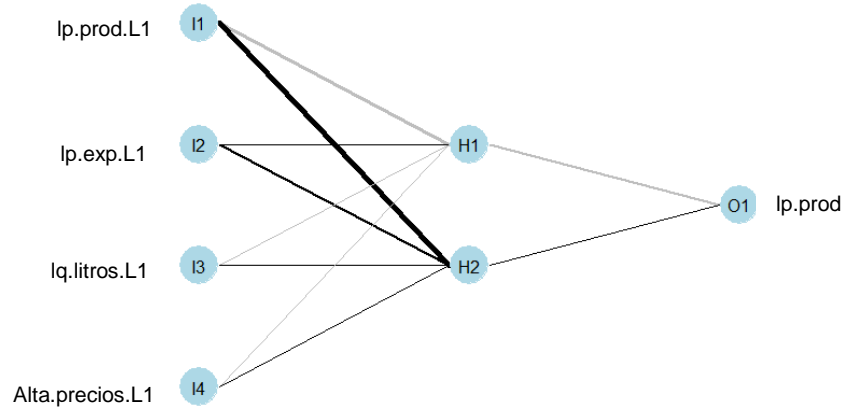
Fuente: Elaboración propia.

5.3.4 Parámetros Estimados de la ANN

En la figura 5.2. se puede ver la estructura de la red entrenada: cuatro nodos de entrada, dos en la capa oculta, y una en la salida. Los nodos de entrada corresponden a las variables *inputs*, a saber, por un lado, el precio al productor primario, el precio de exportación y la cantidad de litros producidos (todos transformados en logaritmo y rezagados en un período); y por otro, la variable dummy de estacionalidad. La capa de salida contiene sólo un nodo y corresponde al logaritmo del precio al productor tambero.

Los pesos/ponderaciones obtenidos en el proceso iterativo de estimación de cada una de las conexiones de la red, se muestran en la tabla 5.5. En total se estiman 13 parámetros, uno por cada conexión de la red (tres de ellos corresponden a reguladores de error que no se muestran en la tabla).

Figura 5.4. Estructura de la Red para Predecir el $lp.prod.$



Nota: Los nodos reguladores del error están contemplados en el modelo, aunque no son expuestos.
Fuente: Elaboración propia.

El algoritmo de Garson que indica que la variable *input* más relevante para explicar el *output* es el logaritmo del precio pagado al productor rezagado en un período ($IR_{lp.prod.L1} = 0.86$), es decir, que es para predecir con la ANN, el precio pagado al productor del mes anterior es la variable más importante; en segundo lugar se encuentra el logaritmo del precio de exportación ($IR_{lp.exp.L1} = 0.073$).

Tabla 5.5. Pesos Estimados de la ANN.

	<i>Hidden 1 (H1)</i>	<i>Hidden 2 (H2)</i>	<i>Output 1 (O1)</i>
<i>Input 1 (I1)</i>	-2.89657	4.01736	--
<i>Input 2 (I2)</i>	0.86308	1.69880	--
<i>Input 3 (I3)</i>	-0.17856	0.15560	--
<i>Input 4 (I4)</i>	-0.06759	0.08279	--
<i>Hidden 1 (H1)</i>	--	--	-2.03085
<i>Hidden 2 (H2)</i>	--	--	0.61281

Fuente: Elaboración propia.

Al igual que en el caso de los VECM dinámicos, las estimaciones de los parámetros de las ANN no se muestran por motivos expositivos; y se reservan iguales resguardos a los del VECM con respecto al formato.

5.4. Comparación de los Modelos

5.4.1 Predicciones del Precio al Productor Primario con Modelos

Estáticos

En las Secciones 5.2 y 5.3 se estiman los parámetros de ambos modelos para luego proceder a predecir la variable de interés, es decir, el precio al productor lácteo. Bajo el supuesto de que no se conocen los valores mensuales de esta variable durante el año 2012, se proyecta la variable para dicho período y se comparan los resultados obtenidos por cada metodología con los valores originales, de modo de ver cuál se ajusta más a ellos.

Tabla 5.6. Valores Observados y Predichos de los Modelos. Período Mensual 2012.

Período	Valor Observado	Pronóstico del VECM	Pronóstico del ANN	Observ.- Pron. VECM	Observ.- Pron. ANN
ene-12	1.48	1.50	1.54	-0.01	-0.05*
feb-12	1.50	1.55	1.54	-0.04	-0.03
mar-12	1.55	1.60	1.56	-0.06	-0.02
abr-12	1.57	1.64	1.59	-0.07	-0.02
may-12	1.57	1.66	1.60	-0.09	-0.03
jun-12	1.56	1.66	1.59	-0.10*	-0.02
jul-12	1.57	1.63	1.57	-0.06	0.00
ago-12	1.55	1.57	1.56	-0.02	-0.02
sep-12	1.54	1.53	1.53	0.01	0.00
oct-12	1.54	1.51	1.54	0.03	0.01
nov-12	1.58	1.52	1.55	0.06	0.03
dic-12	1.63	1.54	1.57	0.09	0.05*
Promedio:				-0.02	-0.01
Desvío:				0.06	0.03

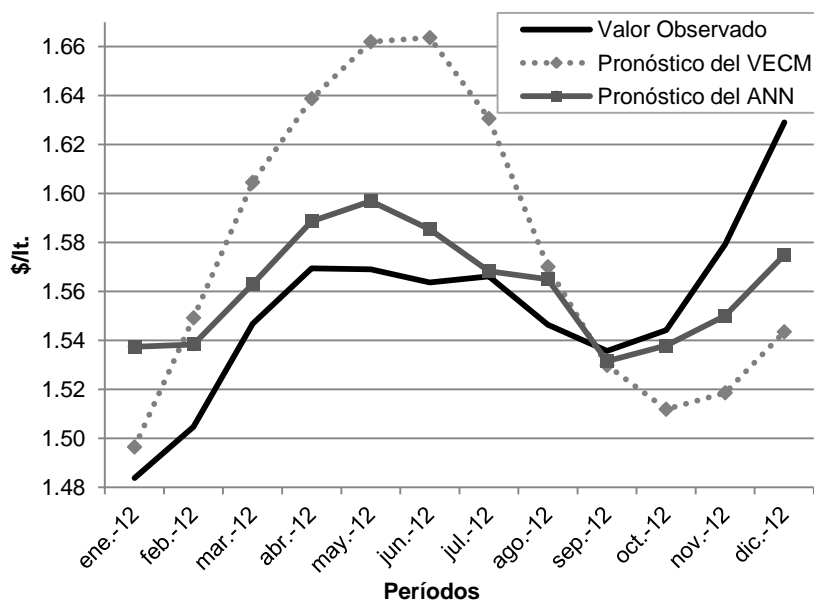
Nota: *Mayores diferencias observadas.

Fuente: Elaboración propia.

En la tabla 5.6. se pueden observar los valores predichos tanto del VECM como del ANN de los meses correspondientes al año 2012; así como también los correspondientes valores reales.

En promedio, las predicciones del VECM tienen errores superiores que las correspondientes a la de ANN. En la figura 5.5 se puede apreciar mejor la evolución de las predicciones de ambos modelos y los valores reales del precio pagado al productor argentino de leche cruda.

Figura 5.5. Valores Observados y Predicciones Realizadas para el Precio al Productor Lechero con los Modelos Estáticos. Año 2012.



Fuente: Elaboración propia.

Gráficamente puede observarse la mayor similitud de las predicciones de la ANN a los valores reales, así como también su evolución y comportamiento. Sin embargo, la evolución de las dos predicciones, en general, concuerda con la real, pero con una mayor dispersión de las del VECM.

Tabla 5.7. Medidas de Performance Predictiva de los Modelos Estáticos.

Medidas de Performance	Modelos:	
	VECM	ANN
RSME	0.0387675	0.01862861
MAD	0.03420706	0.01534307
MPE (%)	7.64504867	3.508897

Fuente: Elaboración propia.

Todas las medidas del error para analizar el ajuste de las predicciones *ex-post*, indican que el modelo ANN resulta ser el más adecuado (tabla 5.7), ya que dichas medidas son menos de la mitad a las obtenidas por medio de las predicciones del VECM.

5.4.2 Predicciones del Precio al Productor Primario con Modelos Dinámicos

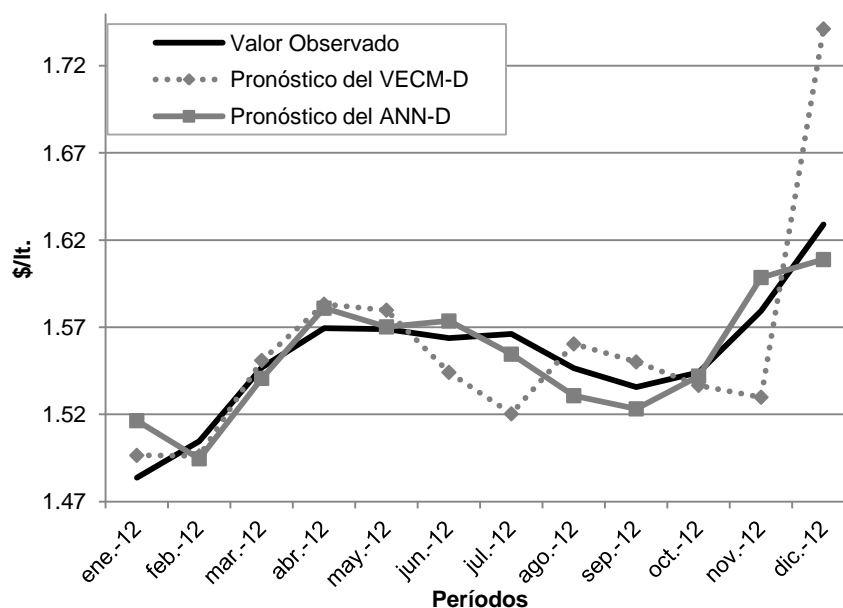
El otro formato de predicción implica parámetros de los modelos cambiantes. Es decir, a partir de parámetros re-estimados en cada período de predicción por medio de la in-

corporación paulatina de valores observados de las variables correspondientes al período inmediatamente anterior al de la predicción.

En la figura 5.6 se muestran los resultados obtenidos de estas predicciones dinámicas. Se puede observar que ambas metodologías proveen predicciones más parecidas a los valores reales de la variable en cada período, que los modelos estáticos. Ello se debe a que las predicciones se hacen para el mes inmediatamente siguiente; y luego, se va ajustando los parámetros de manera paulatina por medio de la incorporación de nuevos datos reales. Este procedimiento hace que se cometa menor error de predicción, ya que se tiene información más actualizada y se expone a menores riesgos sobre eventos futuros que distorsionen las predicciones.

Sin embargo, aún cuando las predicciones realizadas por ambas metodologías están más cerca de los valores reales, de la comparación entre ellas, se puede observar que las predicciones del ANN tienen un mejor desempeño.

Figura 5.6. Valores Observados y Predicciones Realizadas por los Modelos Dinámicos para el Período 2012.



Fuente: Elaboración propia.

Dicha superioridad en las predicciones no sólo se destacan con respecto a las medidas de error (tabla 5.8), sino que además, las predicciones del ANN guardan una evolución más similar a la real, que la que tienen las predicciones del VECM.

Tabla 5.8. Medidas de Performance Predictiva de los Modelos Dinámicos.

Medidas de Performance	Modelos:	
	VECM-D	ANN-D
RSME	0.02400339	0.009829479
MAD	0.01636527	0.008230446
MPE (%)	3.58597272	1.894586646

Nota: los números en **negrita** corresponden a los menores valores observados.
Fuente: Elaboración propia.

Más allá de que las medidas de performance sobre las predicciones realizadas indiquen que la ANN es mejor para predecir de modo dinámico, a su vez, muestran que los errores cometidos son la mitad de los realizados por el VECM.

5.5. Síntesis y Conclusiones

En el presente capítulo se muestra el procedimiento de modelización por medio de un VECM y una ANN, así como los resultados estimados de los parámetros, con el objetivo de predecir el precio pagado a los productores primarios de leche cruda argentina durante el año 2012.

Una vez estimados los modelos, se realizan las predicciones tanto estáticas como dinámicas sobre dicho período, para luego comparar los resultados obtenidos. La comparación no sólo se realiza a partir de las medidas definidas en la Sección 3.4, sino también por medio de una representación gráfica.

En las predicciones estáticas, se observa que ambas metodologías producen predicciones del precio al productor primario lácteo con una evolución similar a la observada. Sin embargo, el modelo ANN muestra predicciones más ajustadas a los valores reales, con menores errores de predicción, corroborado por todas las medidas de performance computadas.

Dichas conclusiones son similares para el caso de las predicciones dinámicas, con la diferencia de que, las correspondientes a la ANN tienen una evolución y comportamiento más similar a las reales que el VECM. Esta última en algunos períodos evoluciona de manera contraria a la real.

Por otro lado, comparando los resultados de las modelizaciones a partir de ambas metodologías, se concluye que se obtuvieron modelos más sencillos para las ANN que para los VECM. Se utilizaron estructuras más simples y parsimoniosas, con menores errores de predicción en todos los casos.

Capítulo 6

Discusión, Conclusiones y Recomendaciones

6.1. Discusión

Realizar predicciones de alguna variable en general, y de los precios en particular, resulta relevante en el análisis sectorial y la toma de decisiones de los agentes involucrados. Dentro del instrumental factible de utilizar para realizar predicciones de series de tiempo se pueden mencionar varias metodologías, todas ellas en general, requieren del cumplimiento de supuestos y se basan sobre distribuciones subyacentes de las variables involucradas. Tal como se menciona en las Secciones 2.3 y 2.4, las más usadas para modelización multivariada de series de tiempo económicas no estacionarias es la Cointegración y VECM (Ahumada, & Cornejo, 2014; Arias, & Torres, 2004; Basso, 2009; Engle, & Granger, 1987; Granger, 1981; Lütkepohl, 2004; Otero, & Trujillo, 1993).

Las proyecciones logradas en la presente tesis a partir del VECM han resultado tener una buena performance predictiva ya que, en todos los casos, las medidas sobre los errores cometidos obtenidas han sido relativamente bajas. Aquí no se han analizado otras metodologías multivariadas clásicas de series de tiempo, y por tanto, no se puede comparar los resultados obtenidos con esas otras posibilidades de modelización. Los motivos son: (a) no se modeliza mediante un VAR porque las series de tiempo utilizadas en el presente trabajo no son estacionarias, y por tanto no cumple uno de los supuestos de partida para aplicar dicho modelo; (b) para modelar con ecuaciones estructurales se necesita utilizar una mayor cantidad de variables para emular el funcionamiento del sector; por lo que, en principio, se decide realizar una primera aproximación con las series de tiempo disponibles.

Al igual que lo analizado en el estado de arte, el VECM ajustado se muestra como una buena metodología de aplicación cuando el interés es predecir variables económicas en

general, y de precios en particular, ya que genera niveles de errores de predicción bastante bajos, incluso cuando el horizonte de proyección no es tan a corto plazo.

Por otra parte, la modelización mediante de ANN ha tomado creciente importancia en la actualidad. Sin embargo, tal como Joekees (2002) afirma, nunca existe la seguridad de que el modelo encontrado ni que los resultados sean los mejores y, por tanto, se debe conformar con haber alcanzado una de las tantas posibilidades. Sin embargo, en la presente tesis se ha conseguido que las ANN (tanto estáticas como dinámicas) tuviesen una performance destacable sobre las predicciones *ex-post*. Con respecto a otros trabajos, los resultados obtenidos en la tesis muestran una performance predictiva muy competitiva, aún cuando dichos trabajos revisados en antecedentes hayan sido aplicados a la predicción de precios de otros productos agrícolas-ganaderos.

De los trabajos encontrados en el área de predicción de precios agrícolas que analizan comparativamente la performance predictiva de modelos tradicionales de serie de tiempo con los de ANN, no todos han concluido en una misma dirección. La mayoría ha destacado a las redes por encima de los otros modelos, sin embargo casi todos ellos se corresponden con problemáticas univariadas exclusivamente. Sin embargo, del presente trabajo, se incorpora como antecedente en la predicción de precios a partir de la modelización multivariada.

Cuando se observa lo ocurrido en la predicción de precios de otros bienes, por ejemplo, en productos energéticos como el petróleo (Fouladgar et al., 2013) o la electricidad (Barrientos et al., 2012), las conclusiones de la comparación entre VECM y ANN también demuestran un buen desempeño de la última por encima de la primera. En línea con estos trabajos, el presente estudio ha encontrado resultados similares, que permiten destacar la performance predictiva de las ANN, ya que no sólo presentan una mejora en la performance predictiva sino que a su vez, muestran una evolución más parecida a los valores reales, y a partir de una estructura más sencilla.

6.2. Conclusiones

Realizar pronósticos sobre el comportamiento de precios es de suma relevancia en el área económica en general, y en el sector lácteo en particular. Tener predicciones confiables que permitan tomar decisiones con menor incertidumbre es precisamente muy valorado por parte de los agentes que interactúan en el sector, dado que permitiría una

planificación más seria sobre sus actividades. Sin embargo, no siempre es fácil contar con una metodología que permita arribar a tal objetivo. El grado de fiabilidad de las predicciones muchas veces depende de la estabilidad del sector, y en particular el sector agrícola ganadero suele estar atado a imprevistos, tanto de tipo climáticos como políticos, que hacen difícil su proyección.

El sector lácteo argentino está compuesto por distintos niveles, a saber, primario, industrial y de comercialización. Todos estos niveles están coordinados por el sistema de precios y son éstos el objeto de constantes conflictos entre los distintos niveles del sector. En particular, los productores primarios son los que frecuentemente se encuentran en descontento con los precios pagados. Resultan ser el eslabón con menor poder de negociación y con problemas para obtener información sobre dichos precios recibidos. Por estos motivos, el interés predictivo de la presente tesis está centrado en los precios al productor primario lácteo.

Con tal objetivo, se abordó el problema de la modelización multivariada de series de tiempo para la predicción de dichos precios, con el propósito de comparar la capacidad predictiva de dos metodologías distintas. Por un lado, una más tradicional y generalmente usada en la modelización multivariada de series de tiempo dentro del área de economía en variables endógenas y no estacionarias (VECM); y por otro lado, una más innovadora en esta área, como son las ANN. Ésta última más flexible en los supuestos de partida y aplicables también a otros tipos de datos.

Se utilizaron distintas variables que representan al sector y que se creen relevantes a la hora de modelar los precios primarios. La mayoría corresponde a los precios de cada nivel, aunque también se incorporaron tanto cantidades producidas como indicadora de estacionalidad del nivel con interés predictivo. No todas estas variables se encuentran disponibles, por lo que debieron ser construidas algunas de ellas. El motivo es la multiplicidad de los productos elaborados a partir de la leche cruda, y que lleva a que los precios representativos de los niveles industriales y de comercialización sean múltiples también. En esos casos se construyeron índices de precios ponderados.

A partir de una variedad de fuentes, se logró formar una base de datos mensual completa que va desde enero del 2000 a diciembre del 2012. Todas las variables son trabajadas

en logaritmo (excepto la de estacionalidad). Algunas de ellas están disponibles para períodos más prolongados, sin embargo, no es la situación de todas ellas.

En primer lugar, se modelizó con cada metodología, buscando para ello un diseño parsimonioso, y que estuviese fundamentado dentro de cada uno de los métodos. Luego, se realizaron predicciones sobre el último año disponible a partir de dos procedimientos distintos, lográndose proyecciones de tipo estáticas y dinámicas. Las estáticas corresponden a las predicciones del precio al productor primario en el año 2012 realizadas (con cada metodología) a partir de la base de datos utilizada en la modelización. En cambio, las dinámicas corresponden a predicciones de igual período de tiempo, pero que provienen de una actualización mensual de la base de datos, incorporando de manera paulatina los valores observados en 2012 de las variables del modelo y reestimando los parámetros del mismo.

En el proceso de búsqueda del VECM se siguió la metodología de Johansen: luego de corroborar que las series no fuesen estacionarias, se procedió a constatar la presencia de cointegración, incluyendo la cantidad óptima de rezagos y elementos determinísticos. Encontrada la forma del modelo, se estimó por medio de MCO los valores de los parámetros para poder realizar las predicciones sobre los períodos de tiempo dejados fuera de la estimación. Manteniendo la misma forma del modelo, los parámetros fueron reestimados para realizar las predicciones dinámicas.

Por otro lado, se procedió a diseñar y estimar la ANN. Con respecto al diseño, partir de un análisis de correlaciones se seleccionaron las variables de entrada; en la capa de salida se incluyó a la única variable de interés predictivo; y por último, la selección de la cantidad de capas ocultas y sus correspondientes nodos se realizó por medio de prueba y error, buscando aquella arquitectura más parsimoniosa con buenos resultados. Se usó el método de aprendizaje supervisado, donde las estimaciones se realizaron por medio del algoritmo de *BP*. Luego, estimados los parámetros, se realizaron las doce predicciones de la variable de interés de manera estática, con igual aclaración que el VECM para las estimaciones dinámicas.

Finalmente se procedió a comparar la capacidad predictiva de los modelos. Dicha comparación no sólo se realizó de manera gráfica, sino que además se calcularon tres medidas distintas sobre el error de predicción. Tanto en las predicciones estáticas como las

dinámicas, los errores de predicción del modelo ANN fueron menores a las del VECM. A su vez, las predicciones del primero tuvieron una evolución más similar a la real que los correspondientes al VECM.

Para ambas metodologías, las medidas del error de predicción de las estimaciones dinámicas fueron menores a las estáticas, lo que indica que con información actualizada hay menos chances de alejarse demasiado de los valores reales. Sin embargo, esta situación es factible si el interés es predecir a corto plazo y se cuenta con información para actualizar la base de datos y las estimaciones de los parámetros.

Por tanto, en todos los casos, se concluye que en términos predictivos la utilización de la ANN permite cometer menores errores que el VECM con un modelo más parsimonioso y simple. Por consiguiente, para predecir los precios pagados a los productores primarios de leche cruda, dicho modelo es el más adecuado.

6.3. Recomendaciones y Futuras Líneas de Investigación

Dado que la presente tesis resulta una primera aproximación e introducción de las ANN en esta área de aplicación; como también, a lo largo de la preparación de la tesis han ido surgiendo determinadas inquietudes que, por cuestiones de focalización del tema, se dejan como propuestas para continuar en un futuro. Algunas propuestas o líneas de trabajo futuras son:

a) Complejizar las ANN aplicadas en el presente, incorporando por ejemplo, estructuras retroalimentadas, algoritmos de selección de variables de entrada, métodos de poda de la estructura, inclusión de otras variables en el nodo de salida, e incluso, la incorporación de distintos momentos de predicción en un mismo nodo de salida, analizar técnicas de robustez y utilización de otros algoritmos de entrenamiento, entre otros. Y, a partir de estas estructuras más complejas, volver a analizar la capacidad predictiva.

b) Analizar capacidad predictiva para períodos más largos de tiempo, incorporando para ello otras metodologías multivariadas como ser, ecuaciones estructurales.

c) Indagar sobre posibles metodologías que permitan obtener medidas de desvíos de los parámetros estimados de las ANN, para posibles evaluaciones sobre la significancia estadística de los mismos.

d) Intentar actualizar la base de datos, y con ella realizar estimaciones actualizadas para ver si las conclusiones se mantienen.

- e) Analizar la capacidad predictiva de VECM restringido, y compararlo con la ANN.
- f) Incorporar metodologías de evaluación de las diferencias entre las medidas de performance obtenidas entre modelos.
- g) Extender la comparación entre ANN y VECM para predicción de precios en otros complejos agroalimentarios, incluso para otros precios del mismo sector.

Bibliografía

- Abudu, S., King, J. P., & Sheng, Z. (2012). Comparison of the Performance of Statistical Models in Forecasting Monthly Total Dissolved Solids in the Rio Grande. *Journal of the American Water Resources Association*, 48(1), 10–24.
- Ahumada, H., & Cornejo, M. (2014). Explaining Commodity Prices by a Cointegrated Time Series-Cross Section Model. *Empirical Economics*, 48(4), 1667–1690.
- Ahumada, H., & Cornejo, M. (2014). How to Forecast Food Prices. *Anales del XLIX Reunión Anual de Economía Política*, Posadas, Argentina.
- Ansotegui, C., & Esteban, M. V. (1997). Relaciones entre Precios y Fundamentos en el Mercado de Capitales Español. *Revista de Economía Aplicada*, 5(14), 99–117.
- Aragón-Torres, A., Calzada-Arroyo, J. M., García-Güemes, A., & Pacheco-Bonrostro, J. (1997). Aplicaciones de Redes Neuronales en Economía. *Anales de las V Jornadas de la Asociación Española de Profesores Universitarios de Matemática para la Economía y la Empresa (ASEPUMA)*, Málaga, España.
- Arias, E., & Torres, C. (2004). *Modelos VAR y VECM para el Pronóstico de Corto Plazo de las Importaciones de Costa Rica* (Documento de Trabajo No. 22). Obtenido de la página del departamento de Investigaciones Económicas del Banco Central de Costa Rica: <http://www.bccr.fi.cr/investigacioneseconomicas/metodoscuantitativos/>.
- Bailey, K. W. (2003). *Impact of MPC Imports on 2002 U.S. Cheese Production*. (Staff Paper No. 362). Retrieved from Department of Agricultural Economics and Rural Sociology of The Pennsylvania State University website: <http://dairyoutlook.aers.psu.edu/reports/Pub2003/staffpaper362.pdf>.
- Bailey, K. W. (2004). A Multiple-Component Analysis of U.S. Dairy Trade. *Journal Dairy Science*, 87(6), 1934–1944.
- Bailey, K. W. (2009). *Documentation of a Dynamic and Simultaneous Econometric Model of the U.S. Dairy Industry* (Staff Paper No. 381). Retrieved from Department of Agricultural Economics and Rural Sociology of The Pennsylvania State University website: <http://purl.umn.edu/48653>.
- Bailey, K. W., Dunn, J., & Pajic, M. (2006). *Short Run Price Determination for the U.S. Dairy Industry* (Staff Paper No. 378). Department of Agricultural Economics and Rural Sociology of the Pennsylvania State University.
- Barrientos, J., Rodas, E., Velilla, E., Lopera, M., & Villada, F. (2012). A Model for Forecasting Electricity Prices in Colombia. *Lecturas de Economía*, 77(2), 91–127.

- Basso, M. A. (2009). *Variaciones en los Precios y Tipo de Cambio. La Paridad de Poder de Compra Argentina* (Tesis de Maestría). Universidad Nacional de Córdoba, Córdoba, Argentina.
- Baum, D., & Haussler, D. (1989). What Size Net Gives Valid Generalization?. *Neural Computation*, 1(1), 151–160.
- Beck, M. W. (2015). *NeuralNetTools*: Visualization and Analysis Tools for Neural Networks. R package versión 1.0.1. (<https://cran.r-project.org/web/packages/NeuralNetTools>).
- Bergmeir, C., & Benítez, J. M. (2015). *RSNNS*: Neural Networks in R using Stuttgart Neural Network Simulator (SNNS). R package 0.4-7. (Disponible en: <https://cran.r-project.org/web/packages/RSNNS>)
- Binner, J. M., Gazely, A. M., Chen, S. H., & Chie, B. T. (2004). Financial Innovation and Divisia Money in Taiwan: Comparative Evidence from Neural Network and Vector Error-Correction Forecasting Models. *Contemporary Economic Policy*, 22(2), 213–224.
- Bukeviciute, L., Dierx, A., Ilzkovitz, F., & Roty, G. (2009). *The Functioning of the Food Supply Chain and Its Effect on Food Prices in the European Union* (Occasional Papers No. 47). Retrieved from Directorate-General for Economic and Financial Affairs of European Commission, website: http://ec.europa.eu/economy_finance/publications/publication_summary15232_en.htm
- Castellano, A., Issaly, L., Iturrioz, G., Mateos, M., & Terán, J. C. (2009). *Análisis de la Cadena de la Leche en Argentina* (Documento de Trabajo No. 4). Obtenido de la página del Instituto Nacional de Tecnología Agropecuaria (INTA): <http://inta.gob.ar/documentos/analisis-de-la-cadena-de-la-leche-en-argentina-2>.
- Chakraborty, K., Mehrotra, K., Mohan, C. K., & Ranka, S. (1992). Forecasting the Behavior of Multivariate Time Series Using Neural Networks. *Neural Networks*, 5(6), 961–970.
- Chatterjee, A., Ayadi, O. F., & Boone, B. E. (2000). Artificial Neural Network and The Financial Markets: A survey. *Managerial Finance*, 26(12), 32–45.
- Chiarvetto Peralta, L. L., Rey Saravia, F. A., & Brignole, N. B. (2008). Aplicaciones de Redes Neuronales para la Predicción de Calidad del Aire. *Mecánica Computacional*, 27, 3607–3625.
- Claveria, O., & Torra, S. (2013). *Forecasting Business Surveys Indicators: Neural Networks vs. Time Series Models* (Working Paper No. 201320). Retrieved from Research Institute of Applied Economics of Univesitat de Barcelona website: http://www.ub.edu/irea/working_papers/2013/201320.pdf.
- Claveria, O., & Torra, S. (2013). Forecasting Tourism Demand to Catalonia: Neural Networks vs. Time Series Models. *Economic Modelling*, 36, 220–228.

- Claveria, O., Monte, E., & Torra, S. (2013). *Tourism Demand Forecasting with Different Neural Networks Models* (Working Paper No. 201321). Retrieved from Research Institute of Applied Economics of Univesitat de Barcelona website: <http://onlinelibrary.wiley.com/doi/10.1002/jtr.2016/epdf>
- Deaton, A., & Laroque, G. (1992). On the Behaviour of Commodity Prices. *The Review of Economic Studies*, 59(1), 1–23.
- Deaton, A., & Laroque, G. (2003). A Model of Commodity Prices After Sir Arthur Lewis. *Journal of Development Economics*, 71(2), 289–310.
- Depetris, E., & Rossini, G. (2005). Comportamiento de Precios de Leche Fluída al Productor y Minorista en las Ciudades de Rosario y Santa Fe. *Revista de Investigaciones de la Facultad de Ciencias Agrarias, UNR*, 5(8), 41–50.
- Depetris, E., & Rossini, G. (2006). Características de los Ajustes de los Márgenes en Quesos Frescos en el Mercado Santafesino. *Revista Argentina de Lactología*, 24, 9–19.
- Depetris, E., Rossini, G., García Arancibia, R., & Vicentin Masaro, J. (2011). *Competitividad del Complejo Lácteo Santafesino*. Santa Fe, Argentina: Ediciones UNL.
- Díaz-González, L., Hidalgo-Dávila, C. A., Santoyo, E., & Hermosillo-Valadez, J. (2013). Evaluación de Técnicas de Entrenamiento de Redes Neuronales para Estudios Geotermométricos de Sistemas Geotérmicos. *Revista Mexicana de Ingeniería Química*, 12(1), 105–120.
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366a), 427–431.
- Dickey, D. A., & Fuller, W. A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, 49(4), 1057–1072.
- Enders, W. (1995). *Applied Econometrics Times Series*. New York, United States: Wiley.
- Enders, W., & Granger, C. W. J. (1998). Unit-root Tests and Asymmetric Adjustment with an Example using the Term Structure of Interest Rates. *Journal of Business & Economic Statistics*, 16(3), 304–311.
- Engle, R. F., & Granger, C. W. J. (1987). Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica*, 55(2), 251–276.
- Engle, R. F., Hendry, D. F., & Richard, J. F. (1983). Exogeneity. *Econometrica: Journal of the Econometric Society*, 51(2), 277–304.
- Fahlman, S., & Lebiere, C. (1989). The Cascade-Correlation Learning Architecture. In D. Touretzky. (Ed.), *Advances in Neural Information Processing Systems*, (pp. 524–532). San Francisco, United States: Morgan Kaufmann.

- Fanchon, P., & Wendel, J. (1992). Estimating VAR Models Under Non-Stationarity and Cointegration: Alternative Approaches for Forecasting Cattle Prices. *Applied Economics*, 24(2), 207–217.
- Fernández-Amador, O., Baumgartner, J., & Crespo-Cuaresma, J. (2010). *Milking The Prices: The Role of Asymmetries in The Price Transmission Mechanism for Milk Products in Austria* (Working Paper No. 2010-21). Retrieved from University of Innsbruck website: <http://eeecon.uibk.ac.at/wopec2/repec/inn/wpaper/2010-21.pdf>
- Fogel, D. B. (1990). An Information Criterion for Optimal Neural Network Selection. *Transactions on Neural Networks*, 2(5), 490–497.
- Fouladgar, M. M., Yazdani, M., Khazaei, S., Zavadskas, E. K., & Fouladgar, V. (2013). Comparison of Vector Time Series and ANN Techniques for Forecasting of WTI Oil Price. *Economic Computation and Economic Cybernetics Studies and Research (ECECSR)*, 47(4), 19–35.
- Frean, M. (1990). The Upstart Algorithm: A Method for Constructing and Training Feedforward Neural Networks. *Neural computation*, 2(2), 198–209.
- Fuller, W. A. (1976). *Introduction to Statistical Time Series*. New York, United States: John Wiley and Sons.
- Garson, G. D. (1991). Interpreting Neural Network Connection Weights. *Artificial Intelligence Expert*, 6(4), 46–51.
- Government Accountability Office. (2005). *Dairy Industry: Information on Milk Prices, Factors Affecting Prices, and Dairy Policy Options* (GAO Publication No. 05-50). Washington, DC: United States Government Accountability Office. Retrieved from <http://www.gao.gov/assets/250/245016.pdf>
- Granger, C. W. (1993). Strategies for Modelling Nonlinear Time-Series Relationships. *Economic Record*, 69(3), 233–238.
- Granger, C. W. J. (1981). Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics*, 16(1), 121–130.
- Guiguet Depetris, E., & Rossini, G. (2005). Cointegración y Simetrías en Precios. El Caso del Dulce de Leche en Santa Fe y Rosario. *Ciencias Económicas*, 1(3), 11–21.
- Gutman, G., Guiguet, E., & Rebolini, J. M. (2003). *Los Ciclos en el Complejo Lácteo Argentino. Análisis de Políticas Lecheras en Países Seleccionados*. Buenos Aires, Argentina: Secretaria de Agricultura, Ganadería y Pesca y Alimentos. Obtenido el 10/07/2010 de http://www.minagri.gob.ar/site/ganaderia/lecheria/03-publicaciones/_archivos/Estudio_Lacteo.pdf
- Haden, K. L., & VanTassell, L. W. (1988). Application of Vector Autoregression to Dynamic Relationships Within the U.S. Dairy Sector. *North Central Journal of Agricultural Economics*, 10(2), 209–216.
- Hamm, L., & Brorsen, W. B. (1997). Forecasting Hog Prices with a Neural Network. *Journal of Agribusiness*, 15(1), 37–54.

- Hansen, B. G., & Li, Y. (2015). *Future World Market Prices of Milk and Feed Looking into the Crystal Ball* (Discussion Paper No. 2015/17). Retrieved from NHH Dept. of Business and Management Science website: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2592976.
- Haykin, S. (2001). *Neural Networks - A Comprehensive Foundation*. New York, United States: IEEE Press.
- Heien, D. (1977). Price Determination Processes for Agricultural Sector Models. *American Journal of Agriculture Economics*, 59(1), 126–132.
- Hernández-Roque, E., Rosales-Vázquez, S., & Chuairey-Rodríguez, A. (2004). *El Traspaso del Tipo de Cambio a los Precios: Una Aproximación al Caso de Cuba* (Estudio Económico No. 42). La Habana, Cuba: Facultad de Economía de la Universidad de la Habana. Obtenido el 20/01/2015 en: <http://www.bc.gob.cu/Anteriores/Otros/InvestPolMon/Traspaso%20Tipo%20de%20Cambio.pdf>
- Hoptroff, R. G. (1993). The Principles and Practice of Time Series Forecasting and Business Modelling Using Neural Nets. *Neural Computing & Applications*, 1(1), 59–66.
- Huarng, K., & Yu, T. H. K. (2006). The Application of Neural Networks to Forecast Fuzzy Time Series. *Physica A: Statistical Mechanics and its Applications*, 363(2), 481–491.
- Instituto Nacional de Estadísticas y Censos (2015) de la Nación Argentina. URL: www.indec.mecon.gov.ar/
- Jha, G. K., & Sinha, K. (2013). Agricultural Price Forecasting Using Neural Network Model: An Innovative Information Delivery System. *Agricultural Economic Research Review*, 26(2), 229–239.
- Joekes, S. (2002). *Aplicación de Redes Neuronales Robustas en Series de Tiempo: Un Estudio Comparativo* (Tesis de maestría). Universidad Nacional de Córdoba, Córdoba, Argentina.
- Johansen, S. (1988). Statical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, 12(2-3), 231–254.
- Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59(6), 1551–1580.
- Johansen, S. (1995). *Likelihood-based Inference in Cointegration Vector Autoregressive Models*. Oxford, United Kingdom: Oxford University Press.
- Johansen, S., & Juselius, K. (1990). Maximum Likelihood Estimation and Inference on Cointegration with Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics*, 52(2), 169–210
- Juselius, K. (2008). *The Cointegrated VAR Model: Methodology and Applications. Advanced Text in Econometrics*. New York, United States: Oxford University Press.

- Kaastra, I., & Boyd, M. (1996). Designing a Neural Network for Forecasting Financial and Economic Time Series. *Neurocomputing*, 10(3), 215–236.
- Kaiser, H. M. (1994). An Analysis of Policy Alternatives to the Dairy Price Support Program. *Agricultural and Resource Economics Review*, 23(2), 158–170.
- Kaiser, H. M., Streeter, D. H., & Liu, D. J. (1988). Welfare Comparisons of U.S Dairy Policies with and without Mandatory Supply Control. *American Agricultural Economics Association*, 70(4), 848–858.
- Kao, G. W., & Ma, C. K. (1992). Memories, Heteroscedasticity, and Price Limit in Currency Futures Markets. *Journal of Futures Markets*, 12(6), 679–692.
- Keough, G. (1991). *Evaluation of Time Series Model Forecasts for the Minnesota-Wisconsin Milk Price* (Research Report No. 91-07). Washington, United States: US Department of Agriculture. Retrieved 2/03/2015, from: <http://handle.nal.usda.gov/10113/43189>
- Kohzadi, N., Boyd, M. S., Kermanshahi, B., & Kaastra, I. (1996). A Comparison of Artificial Neural Network and Time Series Models for Forecasting Commodity Prices. *Neurocomputing*, 10(2), 169–181.
- Kuan, C. M., & Liu, T. (1995). Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks. *Journal of Applied Econometrics*, 10(4), 347–364.
- Kuan, C. M., & White, H. (1994). Artificial Neural Networks: An Econometric Perspective. *Econometric Reviews*, 13(1), 1–91.
- Kung, S. Y., & Hwang, J. N. (1988). An Algebraic Projection Analysis for Optimal Hidden Units Size and Learning Rates in Back-propagation Learning. International Conference on Neural Networks, Conference taking place in San Diego, United States.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure are We that Economic Time Series Have a Unit Root? *Journal of Econometrics*, 54(1), 159–178.
- LaFrance, J. T., & de Gorter, H. (1985). Regulation in a Dynamic Market: The US Dairy Industry. *American Journal of Agricultural Economics*, 67(4), 821–832.
- Lapedes, A., & Farber, R. (1987). Nonlinear Signal Processing Using Neural Networks: Prediction and System Modelling. International Conference on Neural Networks, Conference taking place in San Diego, United States.
- Levich, R. M., & Thomas, L. R. (1993). The Significance of Technical Trading-Rule Profits in the Foreign Exchange Market: A Bootstrap Approach. *Journal of International Money and Finance*, 12(5), 451–474.
- Li, G. Q., Xu, S. W., & Li, Z. M. (2010). Short-Term Price Forecasting for Agro-products Using Artificial Neural Networks. *Agriculture and Agricultural Science Procedia*, 1, 278–287.

- Lin, C. T., & Yeh, H. Y. (2009). Empirical of the Taiwan Stock Index Option Price Forecasting Model-Applied Artificial Neural Network. *Applied Economics*, 41(15), 1965–1972.
- Lippmann, R. P. (1987). An Introduction to Computing with Neural Nets. *ASSP Magazine*, 4(2), 4–22.
- Liu, D. J., Kaiser, H. M., Mount, T. D. & Forker, O. D. (1991). Modeling the U.S. Dairy Sector with Government Intervention. *Western Journal of Agricultural Economics*, 16(2), 360–373
- Liu, D. J., Sun, C. H., & Kaiser, H. M. (1995). Market Conduct Under Government Price Intervention in the U.S. Dairy Industry. *Journal of Agricultural and Resource Economics*, 20(2), 301–315.
- Ljung, G., & Box, G. (1978). On a Measure of Lack of Fit in Times Series Models. *Biometrik*, 65(2), 297–303
- Lütkepohl, H. (2004a). Univariate Time Series Analysis. In H. Lütkepohl y M. Kräzig. (Ed.), *Applied Time Series Econometrics*, Cambridge, United Kingdom: Cambridge University Press.
- Lütkepohl, H. (2004b). Vector Autorregressive and Vector Error Correction Models. In H. Lütkepohl & M. Kräzig. (Ed.), *Applied Time Series Econometrics*, Cambridge, United Kingdom: Cambridge University Press.
- Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. New York, United States: Springer.
- Maciel, L. S., & Ballini, R. (2008). Design a Neural Networks for Time Series Financial Forecasting: Accuracy and Robustness Analysis. Anales do 9º Encontro Brasileiro de Finanças, Sao Pablo, Brazil.
- Maciel, L. S., & Ballini, R. (2010). Neural Networks Applied to Stock Market Forecasting: An Empirical Analysis. *Journal of the Brazilian Neural Network Society*, 8(1), 3–22.
- Maddala, G. S., & Kim, I. M. (1998). *Units Roots, Cointegration, and Structural Change*. Cambridge, United Kingdom: Cambridge University Press.
- Marshall, A. (1890). *Principles of Political Economy*. New York, United States: Maxmillan.
- Mas-Collel, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic Theory*. New York, United States: Oxford University Press.
- Masters, T. (1993). *Practical Neural Network Recipes in C++*. San Francisco, United States: Morgan Kaufmann.
- McGuire, M. (2012). *Forecasting Class III Milk Prices in a Volatile Market* (Tesis Doctoral). Colorado State University, Colorado, United States.

- Ministerio de Agricultura, Ganadería, Pesca y Alimentos (SAGPyA) de la Nación Argentina. (2015). URL: http://www.minagri.gob.ar/site/_subsecretaria_de_lecheria/lecheria/index.php
- Morgan, N., & Bourland, H. (1990). Generalization and Parameter Estimation in Feedforward Nets: Some Experiments". In D. Touretzky and S. David. (Eds.), *Advances in Neural Information Processing Systems*. San Mateo, United States: Morgan Kaufmann.
- Mosheim, R. (2012). *A Quarterly Econometric Model for Short-Term Forecasting of the US Dairy Industry* (Technical Bulletin No. TB-1932). Washington, United States: US Department of Agriculture. Retrieved from <http://purl.umn.edu/184305>
- Nicholson, W. (2005). *Teoría Microeconómica. Principios Básicos y Aplicaciones*. Distrito Federal, México: Cengage Learning.
- Novakovic, A. M., & Thompson, R. L. (1977). The Impacts of Imports Manufactured Milk Products on the U.S Dairy Industry. *American Journal of Agriculture Economics*, 59(3), 507–519.
- Osterwald-Lenum, M. (1992). A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics. *Oxford Bulletin of Economics and Statistics*, 54(3), 461–472.
- Otero, J. M., & Trujillo, F. (1993). Predicción Multivariante y Multiperíodo de una Serie Temporal Económica Mediante una Red Neuronal. *Estadística Española*, 35(133), 345–375.
- Park, Y. R., Murray, T. J., & Chen, A. C. (1996). Predicting Sun Spots Using a Layered Perceptron Neural Network. *Transactions on Neural Networks*, 7(2), 501–505.
- Penta-Transaction Corporation (2015).
- Pfaff, B. (2008). *Analysis of Integrated and Cointegrated Time Series with R*. New York, United States: Springer.
- Pfaff, B., & Stigler, M. (2013a). *urca*: Unit Root and Cointegration Tests for Time Series Data. R package version 1.2. (Disponible en: <http://cran.r-project.org/web/packages/urca>).
- Pfaff, B., & Stigler, M. (2013b). *vars*: VAR Modelling. R package version 1.5-2. (Disponible en: <https://cran.r-project.org/web/packages/vars>).
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>.
- Ripley, B. D. (2005). *Pattern Recognition and Neural Networks*. United Kingdom: Cambridge University Press.
- Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6), 386–408.

- Rossini, G., Depetris, E., García Arancibia, R., Vicentin Masaro, J., & Coronel, M. (2014). Transmisiones de Precios en el Sector Lácteo: El Análisis del Comportamiento de los Precios de Exportación y el Recibido por los Productores. *Revista FAVE, Facultad de Ciencias Agrarias*, 12(1), 43–52.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning Internal Representations by Error Propagation*. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, United Kingdom: MIT Press.
- Schmit, T. M., & Kaiser, H. M. (2006). Forecasting Fluid Milk and Cheese Demands for the Next Decade. *Journal of Dairy Science*, 89(12), 4924–4936.
- Secretaría de Lechería de la Provincia de Santa Fe. (2015). URL: <https://www.santafe.gov.ar/index.php/web/content/view/full/112738/%28subtema%29/93794>
- Shahwan, T., & Odening, M. (2007). Forecasting Agricultural Commodity Prices Using Hybrid Neural Networks. In C. Shu-Heng, P. P. Wang & K. Tzu-Wen. (Eds.), *Computational Intelligence in Economics and Finance* (pp. 63–74). Berlin, Germany: Springer Berlin Heidelberg.
- Stock, J. H., & Watson, M. W. (1988). Testing for Common Trends. *Journal of the American Statistical Association*, 83(404), 1097–1107.
- Swanson, N. R., & White, H. (1997). Forecasting Economic Time Series Using Flexible versus Fixed Specification and Linear versus Nonlinear Econometric Models. *International Journal of Forecasting*, 13(4), 439–461.
- Tang, Z., de Almeida, C., & Fishwick, P. A. (1991). Time Series Forecasting Using Neural Networks vs. Box-Jenkins Methodology. *Simulation*, 57(5), 303–310.
- Terán, J. C. (2008). *Caracterización de la Cadena Agroalimentaria de la Leche en la Provincia de Santa Fe*. (Informe Anual 2008). Rafaela, Argentina: Instituto Nacional de Tecnología Agropecuaria. Obtenido de: <http://inta.gob.ar/documentos/caracterizacion-de-la-cadena-agroalimentaria-de-la-leche-en-la-provincia-de-santa-fe>
- Trapletti, A., & Hornik, K. (2015). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-33. (Disponible en: <https://cran.r-project.org/web/packages/tseries>)
- Vargas, H. (1995). *La Relación entre el Crédito y la Inflación* (Borradores Semanales de Economía No. 37). Bogotá, Colombia: Banco de la República de Colombia. Obtenido de: <http://www.banrep.gov.co/docum/ftp/borra069.pdf>
- Varian, H. R. (1994). *Microeconomía Intermedia. Un Enfoque Moderno*. Barcelona, España: Antoni Bosch Editor.
- Vicentin Masaro, J., Depetris, E., Rossini, G., & Coronel, M. (2012). Transmisión Vertical entre Precios de Exportación de Quesos y Precios Pagados al Productor Lechero Argentino. *Anales XLIII Reunión Anual de la AAEA*, San Juan, Argentina.

- Weigend, A. S., & Gershenfeld, N. A. (1994). *Time Series Prediction: Forecasting the Future and Understanding the Past*. Reading, United Kingdom: Addison-Wesley.
- Weigend, A. S., Huberman, B. A., & Rumelhart, D. E. (1990). Predicting the Future: A Connectionist Approach. *International Journal of Neural Systems*, 1(3), 193–209.
- Weigend, A. S., Rumelhart, D. E., & Huberman, B. A. (1991). Generalization by Weight-Elimination Applied to Currency Exchange Rate Prediction. International Joint Conference on Neural Networks (IJCNN), Conference taking place in Seattle, United States.
- Westcott, P. C. & Hoffman, L. A. (1999). *Price Determination for Corn and Wheat* (Technical Bulletin No. 1878). Retrieved from United States Department of Agriculture, Market and Trade Economics Division website: <http://www.ers.usda.gov/publications/tb-technical-bulletin/tb1878.aspx>.
- Westhoff, P., & Brown, D. S. (1999). The U.S. Dairy Sector without Price Supports. *Canadian Journal of Agricultural Economics/Revue Canadienne d'Agroeconomie*, 47(5), 19–27.
- White H. (1988). Economic Prediction Using Neural Networks: The Case of Ibm Daily Stock Returns. International Joint Conference on Neural Networks (IJCNN), Conference taking place in San Diego, United States.
- Wickham, H. (2011). *plyr*: Tools for Splitting, Applying and Comibing Data. R package, version 1.8.3. (Disponibile en: <https://cran.r-project.org/web/packages/plyr>)
- Wright, S. (1921). Correlation and Causation. *Journal of Agricultural Research*, 20(7), 557–585.
- Yao, J., Li, Y., & Tan, C. L. (2000). Option Price Forecasting Using Neural Networks. *Omega*, 28(4), 455–466.
- Zapata, H. O., & Garcia, P. (1990). Price Forecasting with Time-Series Methods and Nonstationary Data: An Application to Monthly US Cattle Prices. *Western Journal of Agricultural Economics*, 15(1), 123–132.
- Zhang, G. P. (2004). Business Forecasting with Artificial Neural Networks: An Overview. In Zhang G. (Ed.), *Neural Networks in Business Forecasting* (pp. 1–22). Hershey, United States: Idea Group Publishing.
- Zhang, G. P., Patuwo, B. E., & Hu, M. Y. (1998) Forecasting with Artificial Neural Networks: The State of the Art. *International Journal of Forecasting*, 14, 35–62.
- Zou, H. F., Xia, G. P., Yang, F. T., & Wang, H. Y. (2007). An Investigation and Comparison of Artificial Neural Network and Time Series Models for Chinese Food Grain Price Forecasting. *Neurocomputing*, 70(16), 2913–2923.

Anexo

A1. Diagnóstico del VECM

El análisis de diagnóstico de los residuos obtenidos del VECM consiste en comprobar la normalidad, la heteroscedasticidad y la no autocorrelación parcial, así como la estacionariedad de los mismos.

La prueba ARCH tiene como hipótesis nula la ausencia de autocorrelación heteroscedástica de los residuos multivariados del VECM. El estadístico obtenido es $\chi^2 = 529.7$, cuyo p-valor es 0.173. Por lo tanto, al nivel del 5% no se rechaza hipótesis nula sobre ausencia de autocorrelación heteroscedástica, es decir, el VECM estimado genera residuos que no están autocorrelacionados heteroscedásticamente.

Por otro lado, el test Portmanteau es una prueba asintótica que sirve para corroborar la presencia de autocorrelación serial en la matriz de residuos del VECM. Bajo la hipótesis de ausencia de autocorrelación serial en dicha matriz, se ha obtenido un estadístico $\chi^2 = 224.047$, cuyo p-valor es de 0.082, lo que indica que con un nivel del 5% no se rechaza la hipótesis nula no presencia de autocorrelación serial en los residuos.

No ha sido posible corroborar la normalidad de los residuos del VECM. Se realiza un tests Jarque-Bera. La hipótesis nula confirma la presencia de normalidad de los residuos. El estadístico obtenido es $\chi^2 = 736.92$, cuyo p-valor es 0.000. Al nivel del 5% se rechaza hipótesis nula.

Aún no siendo posible corroborar la normalidad de los residuos, se ha decidido quedarse con dicho modelo, ya que de todos los probados, ninguno ha satisfizo dicho supuesto.

Las pruebas sobre la estacionariedad de residuos son univariados. Por tanto se realizan los test *ADF* y *KPSS* para cada uno de los residuos correspondientes a las ecuaciones endógenas del VECM de la misma manera que se realiza en el cuerpo de la tesos. En

todos los casos los residuos han sido estacionarios. En las pruebas *ADF* se ha rechazado hipótesis nula sobre presencia de raíces unitarias incluso a un nivel del 1%. Por otro lado, en ninguna de las pruebas *KPSS* se ha podido rechazar hipótesis nula sobre estacionariedad de los residuos, incluso a un nivel del 10%.