



Vigilada Mineducación

Predicción de tipo funcional para la demanda de productos de maquinaria agrícola a partir de suavizados robustos Nadaraya-Watson

Functional prediction for the demand of agricultural machinery products from robust smoothing Nadaraya-Watson

MARIBEL MEDINA ARBOLEDA
mmedinaa@eafit.edu.co

Trabajo de Grado

Director
Santiago Ortiz
sortiza2@eafit.edu.co

UNIVERSIDAD EAFIT
Escuela de Ciencias Aplicadas e Ingeniería
Maestría en Ciencias de los Datos y Analítica
Medellín, Colombia
2022

Resumen

En la cadena de suministro, la previsión es un factor importante para determinar la eficiencia de la operación. Un nivel de servicio óptimo al cliente o un manejo de capital de trabajo adecuado son conceptos en las empresas que se ven afectados positivamente en la forma en que se logra tener una predicción adecuada de la demanda de un producto. Este trabajo expone dos métodos para abordar el pronóstico de la demanda de productos de maquinaria agrícola, el primero se basa en utilizar técnicas de regresión no paramétrica con datos explicativos funcionales y el segundo, un modelo semi-funcional lineal parcial. En este enfoque funcional de predicción de series temporales, se propone utilizar una trayectoria continua de las ventas en el pasado para predecir los valores futuros y un modelo lineal parcial que considera variables adicionales exógenas que permitan mejorar el pronóstico. Asimismo, el uso de técnicas robustas del estimador de Nadaraya-Watson basada en una función de peso que penaliza los valores más alejados de las observaciones multivariantes utilizando una medida de profundidad de Mahalanobis, que en consecuencia, presentan propiedades de robustez y mejor precisión en la previsión de la demanda de productos de maquinaria agrícola la cual, tiene como principal característica, la variabilidad y la presencia de valores atípicos.

Palabras clave: Datos funcionales, estadística no paramétrica, regresión lineal parcial, estimador Nadaraya-Watson.

1 Introducción

La predicción de la demanda es la base de muchas decisiones de gestión en la cadena de suministro, como la planificación y el cumplimiento de los pedidos (Narayanan et al., 2019), el apoyo a la planificación de producción, de servicio posventa, la distribución de los productos (Acar and Gardner, 2012) y del control de los inventarios (Abolghasemi et al., 2020), de igual manera, un pronóstico preciso permite desarrollar una estrategia de precios adaptable para mejorar la gestión de los ingresos (Seyedan et al., 2022). Las empresas comercializadoras de productos de maquinaria agrícola tienen, por ende, que gestionar adecuadamente los procesos de planificación de la demanda y las fluctuaciones de estas, lo cual facilita consecuentemente, un abastecimiento óptimo de productos, una reducción en los costos de operación y una mayor satisfacción del cliente.(Nguyen et al., 2018; Wang et al., 2016; Fildes et al., 2019).

Abolghasemi et al. (2020) define la predicción como una estimación de las ventas de un producto durante un determinado periodo de tiempo, mientras que Kirkpatrick and Gaynor (1994) definen la predicción como “el esfuerzo de prever eventos futuros examinando el pasado”. Sin embargo, en la predicción existe un elemento importante llamado volatilidad de la demanda, la cual se hace compleja debido a muchos factores endógenos y exógenos, como: las variaciones de los mercados, nuevos competidores, el clima, variaciones en la tasa representativa del mercado, las importaciones de los productos, el impacto de la promoción, las ventas de otros productos, entre otros (Gilliland, 2010).

Una gran variedad de métodos estadísticos se han desarrollado para generar pronósticos, el mayor inconveniente es que estos métodos están basados en supuestos sobre la tendencia de los datos, teniendo que utilizar, un modelo distinto para diferentes tendencias de las ventas (Thiesing and Vornberger, 1997). Pronósticos de series temporales, agrupación, redes neuronales (NN, por sus siglas en inglés), K-Vecinos más cercanos (KNN, por sus siglas en inglés), regresión lineal (MLR, por sus siglas en inglés), árboles de decisión (DT, por sus siglas en inglés), máquinas de vectores de soporte (SVM, por sus siglas en inglés) y regresión de vectores de soporte (SVR, por sus siglas en inglés), son métodos comunes que se han utilizado para la predicción de la demanda (Seyedan and Mafakheri, 2020; Stepnicka et al., 2011; Bozkir and Sezer, 2011). Asimismo, En los últimos años, se pueden encontrar algunos estudios que hacen frente al problema del pronóstico de la demanda desde una perspectiva funcional, esto es, definir las curvas diarias de demanda mediante alguna forma funcional, conocido como análisis de datos funcionales. Literatura sobre este tema puede encontrarse en Ramsay and Silverman (2005), Ferraty and Vieu (2006) y Kokoszka and Horvath (2012).

En la predicción de la demanda se consideran dos tipos de previsiones: las previsiones a corto plazo, es decir, las previsiones horarias con un día de antelación, y las previsiones a medio plazo, es decir, las previsiones con varios días de antelación para los datos diarios. En la comercialización de productos de maquinaria agrícola las series de demanda presentan las siguientes características: tienen alta variabilidad, en algunos productos se superponen niveles de estacionalidad, en particular estacionalidad mensual, presentan dinámica de corto plazo y semanas especiales (efecto calendario, donde se presenta una menor demanda en la primera mitad del mes comparado con la segunda mitad). Estas series incorporan efectos no lineales de variables como la tasa representativa del mercado, índice de precios al consumidor y el clima en especial. Los valores atípicos también están presentes. Las características anteriores implican que, encontrar un ajuste adecuado para estas series, sea un problema complejo.

Este documento se centra en, primero, la aplicación de algoritmos basados en modelos funcionales no paramétricos para la predicción de la demanda considerando la variable respuesta de tipo funcional y variables explicativas funcionales, segundo, la aplicación de modelos de regresión semi-funcional lineal parcial considerando variables exógenas que permitan mejorar el modelo de predicción, estos modelos de predicción han sido estudiados en Vilar et al. (2012), Aneiros et al. (2016) y Vilar et al. (2018). Procedimientos bootstrap para evitar supuestos sobre la distribución de los datos. Lo propuesto en este documento, a diferencia de lo expuesto en la literatura que tratan la predicción a partir de datos funcionales, es la aplicación de un modelo de regresión no paramétrico robusto del estimador de Nadaraya-Watson el cual es resistente a valores atípicos.

Este documento está organizado de la siguiente manera. La Sección 2, los datos reales de demanda de maquinaria agrícola en una empresa comercializadora de productos en el sector de la agricultura desde el año 2018 al año 2021 en Colombia, covariables funcionales exógenas como la tasa representativa del mercado y el índice de precios al consumidor y; datos de los descuentos aplicados a las ventas de los productos. En la Sección 3, los

métodos de predicción implicados en el estudio, funcional no paramétrico y el modelo semi-funcional lineal parcial, así como los intervalos de predicción. En la Sección 4, se presenta los resultados de los modelos aplicados. Finalmente, en la Sección 5, se muestran las conclusiones derivadas de este trabajo.

2 Datos

Esta base de datos consiste en la demanda de productos de maquinaria agrícola para el sector de la agricultura en el mercado colombiano de una empresa comercializadora, de igual forma, se presentan algunas covariables, que pueden afectar o no, la demanda de estos productos. La principal característica del comportamiento de las ventas de estos productos es su variabilidad y la presencia de datos atípicos, es decir, es un comportamiento no estacionario. El comportamiento de la demanda al igual que las variables exógenas se presentan en el periodo de tiempo: semanal, esto es, para la predicción de la demanda de las 52 semanas del año siguiente, con base en las semanas del año anterior.

Las observaciones de las variables exógenas de la tasa representativa del mercado (TRM) y el índice de precios al consumidor (IPC) están disponibles en los sitios oficiales del Banco de la República de Colombia (2022) y el Departamento Administrativo Nacional de Estadística (2022).

La Figura 1 muestra la demanda histórica, en unidades, para los años 2018 al 2021 en las 52 semanas del año. Las principales características de la demanda se pueden resumir en la variabilidad, no se observa tendencia ni estacionalidad. Tanto la demanda como la TRM y los descuentos aplicados a las ventas presentan datos atípicos, para ello no se necesitará herramientas para detectar y manejar valores atípicos en series temporales funcionales. Los métodos presentados en Osorio et al. (2021), los cuales se exponen en la Sección 3.1.1, se aplicaron a los datos objeto de análisis, en los cuales, el método propuesto es resistente ante la presencia de valores atípicos. Se utilizan técnicas de suavizado para convertir los datos de 52 semanas en una observación funcional. Además, existen variables exógenas que podrían mejorar las predicciones de la demanda de productos de maquinaria agrícola y podrían incorporarse al modelo de predicción incluido en la Sección 3.2. Se introdujo las variables TRM y el IPC al modelo, estas dos covariables podrían influir positiva o negativamente en la demanda de productos en un país.

3 Modelos funcionales

El análisis de datos funcionales es una rama de la estadística que analiza datos que proporcionan información sobre curvas, superficies o cualquier otro objeto matemático que varíe sobre un continuo. El continuo suele ser el tiempo, por ejemplo. Estas curvas se definen mediante alguna forma funcional. Los libros publicados por Ramsay and Silverman (2002) y Ramsay and Silverman (2005) son buenas referencias generales para el análisis de datos funcionales utilizando una visión lineal. El libro de Ferraty and Vieu (2006), analiza

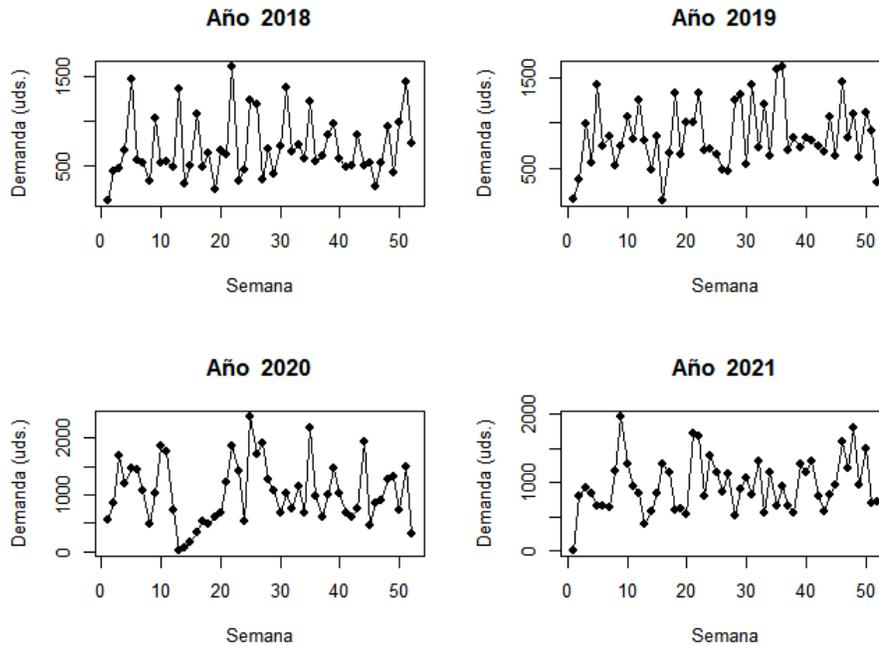


Figura 1: Comportamiento de la demanda de productos de maquinaria agrícola de una empresa comercializadora Colombiana.

los datos funcionales desde el punto de vista estadístico no paramétrico, Kokoszka and Horvath (2012), centrado en la inferencia y, más recientemente, Eubank and Hsing (2015), el cual presenta un compendio de los conceptos y resultados matemáticos relevantes para el desarrollo teórico del análisis de datos funcionales. Debido a la naturaleza de los datos, los métodos de suavización no paramétricos también son herramientas útiles para análisis funcional, enfoque seguido en este documento. En las siguientes subsecciones se presenta el modelo funcional no paramétrico y el modelo semi-funcional lineal parcial utilizados para la previsión de la demanda.

3.1 Modelo funcional no paramétrico

Los modelos de regresión buscan estimar la relación entre una variable dependiente o respuesta y una o más variables independientes o predictoras. Dentro de las metodologías de regresión existentes, el método de mínimos cuadrados ordinarios (MCO) se ha adoptado generalmente por su tradición y facilidad de cálculo, sin embargo, se ve afectada por la presencia de valores atípicos, la no normalidad, la multicolinealidad y los datos faltantes (Ho and Naugher, 2000). Los valores atípicos son observaciones que parecen incoherentes con el resto de los datos, a menudo, estos puntos influyentes, además de permanecer ocultos para el usuario, hacen que los modelos no presenten resultados fiables (Alma, 2011; Rousseeuw, 1984).

Para resolver este problema, se ha desarrollado nuevas técnicas estadísticas que no se

ven afectadas ante la presencia de valores atípicos. Algunas alternativas son: los mínimos cuadrados recortados (LTS, por sus siglas en inglés), la estimación M de Huber, la estimación MM, Regresiones robustas a partir de M-estimadores, S-estimadores (Birkes and Dodge, 1993; Schumacker et al., 2002; Wilcox, 2004; Staudte and Sheather, 1990) y, Velasco et al. (2020) proponen un estimador basado en la regresión C3S. Otros autores también adoptaron métodos robustos para reforzar los estimadores muestrales como para detectar observaciones atípicas, ver por ejemplo Peña and Prieto (2001), Maronna and Zamar (2002) y Olive (2004). El objetivo principal de la regresión robusta es proporcionar resultados resistentes en presencia de valores atípicos.

El análisis de regresión puede ampliarse para tratar datos funcionales, lo que da lugar a la regresión funcional. Probablemente, el modelo de regresión más utilizado es el modelo lineal clásico en el que las variables son escalares. Si se introducen curvas en lugar de valores escalares, se pueden obtener tres modelos de regresión diferentes:

- Modelo totalmente funcional: tanto la respuesta como los regresores son curvas.
- Modelo de respuesta escalar: respuesta escalar y regresores funcionales.
- Modelo de respuesta funcional: respuesta funcional y regresores escalares

Esta clasificación dada para el modelo de regresión lineal puede aplicarse a otro tipo de modelos, como los modelos de regresión lineal-parcial o no paramétrica (Raña, 2016).

El primer método de predicción abordado en el presente documento, es el modelo funcional no paramétrico (FNP), en particular para la previsión de la demanda. Un enfoque basado en la regresión no paramétrica con funcional tanto de la variable respuesta como la variable predictora, fue propuesto en Ferraty et al. (2011) en un entorno de curvas independientes. En Aneiros et al. (2013) se utilizó dicho modelo para predecir las curvas de demanda residual. También en Vilar et al. (2012) se empleó este modelo considerando la variable respuesta como escalar. En el presente trabajo, se considerará la variable respuesta funcional, para predecir la demanda del siguiente año. Es decir, al considerar la respuesta funcional, predice directamente la curva semanal mientras que, con la respuesta escalar, se predice solo un día por separado. Este modelo es posible extrapolarlo a otros periodos de tiempo, incluso a otras jerarquías presentes en la información de las ventas de productos.

La serie temporal de demanda se considerará realizaciones de tiempo discreto de un proceso estocástico de tiempo continuo, $\{\zeta(t)\}_{t \in R}$, las unidades para t son semanas del año, es decir, $\tau = 52$ para la previsión de la demanda del siguiente año. Considerando que dicho proceso se observa en el intervalo $t \in [a, b)$, con $b = a + N\tau$, donde $N = 52$, es la longitud del ancho de banda, las curvas observadas semanales de la demanda puede escribirse como $\{\zeta_i\}_{i=1}^N$, donde $\zeta_i(t) = \zeta(a + (i - 1)\tau + t)$, con $t \in (0, \tau]$. Se obtiene la predicción para cada semana, con una semana de antelación de la demanda de productos de maquinaria agrícola, $\zeta_{N+1}(t)$, con $t \in \{1, 2, \dots, 52\}$, en el año 2019 a partir de la información dada por las 52 semanas anteriores.

Se desea predecir la variable $\zeta_{N+1}(t)$ donde $t \in \{1, \dots, 52\}$ usando información dada de $\{\zeta_i(t) : i = N - 52, \dots, N\}$, el modelo no paramétrico (FNP) puede escribirse como:

$$\zeta_{i+1}(t) = m_t(\zeta_i) + \xi_{t,i+1}, i = 1, \dots, n, \quad (1)$$

donde $m_t(\cdot)$ es una función no paramétrica desconocida de la serie y $\xi_{t,i+1}$ es un término de error, con media cero. Así pues, una predicción $\widehat{\zeta}_{N+1}(t)$ para la variable $\zeta_{N+1}(t)$ puede ser obtenida estimando $\widehat{m}_t(\zeta_N)$ en (1), es decir, $\widehat{\zeta}_{N+1}(t) = \widehat{m}_t(\zeta_N)$. En el contexto actual, este enfoque consiste en estimar la autorregresión funcional, m_t , utilizando los valores anteriores de la demanda del año completo y aplicar esta funcional estimada al último año observado.

Un estimador de tipo Nadaraya-Watson para m_t se define como:

$$\widehat{m}_{t,h}^{FNP}(\zeta_N) = \sum_{i=1}^n \omega_h(\zeta_N, \zeta_i) \zeta_{i+1}(t) \quad (2)$$

Los pesos $\omega_h(\cdot, \cdot)$ en (2) se construyen como:

$$\omega_h(\zeta_N, \zeta_i) = \frac{K(d(\zeta_N, \zeta_i)/h)}{\sum_{i=1}^n K(d(\zeta_N, \zeta_i)/h)}, \quad (3)$$

donde $K : [0, \infty) \rightarrow [0, \infty)$ es una función kernel, $h > 0$ es un parámetro de suavización y $d(\cdot, \cdot)$ es una semimétrica. Tanto el ancho de banda como la semimétrica son aspectos relevantes para el buen comportamiento asintótico y práctico del estimador. Los métodos de kernel son bien conocidos y utilizados intensamente ya que son una forma útil de hacer la ponderación local. La idea principal de la ponderación local entorno a ζ es atribuir a cada ζ_i un peso que tenga en cuenta la distancia entre ζ y ζ_i ; cuanto más distante esté ζ_i de ζ , menor será la ponderación. Existe una amplia literatura en este campo, para una presentación del estado del arte, véase Horová et al. (2002) y Ferraty and Vieu (2006). Para el desarrollo del presente trabajo se considerará el kernel Gaussiano, este se define como $k(\mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\mu^2}{2}\right\}$.

El estimador de Nadaraya-Watson, propuesto por Nadaraya (1964) y Watson (1964), es una media ponderada localmente que utiliza una función de kernel, como se muestra en (3). Sin embargo, la inclusión de este tipo de estimador puede traer problemas ante la presencia de observaciones atípicas. Como diferenciador de la metodología propuesta en Vilar et al. (2018), este trabajo, propone una variación de este estimador que incluye una función de peso basada en una versión robusta de la profundidad de Mahalanobis multivariante abordada en Osorio et al. (2021).

3.1.1 Modelo Nadaraya-Watson robusto como suavizador de observaciones

Dado un par de observaciones $(x_i, Y_i), \dots, (x_n, Y_n)$, donde la variable dependiente Y está relacionada con la variable independiente x . La distancia de Mahalanobis al cuadrado de

las observaciones $w_i = (x_i, y_i)$ mediante $PMD_R^2(w_i) = (w_i - \hat{\mu}_R)' \hat{S}_R^{-1} (w_i - \hat{\mu}_R)$, donde $\hat{\mu}_R$ y \hat{S}_R^{-1} son estimadores robustos multivariantes de localización y dispersión. El objetivo de la aplicación de esta distancia $PMD_R^2(w_i)$ es establecer una medida de profundidad de Mahalanobis de la forma $PM_h D(w_i) = (1 + PMD_R^2(w_i))^{-1}$, Osorio et al. (2021) propone considerar $PM_h D(w_i)$ para medir la periferia de los puntos, de ese modo, introducir una función de peso $d(w_i)$ que depende del rango de la medida de profundidad para cada w_i , esto es:

$$d(w_i) = PM_h D(w_i) \left(\sum_{j=1}^n PM_h D(w_j) \right)^{-1}, \quad (4)$$

donde $d(w_i) : \mathbb{R}^2 \rightarrow (0, 1)$ siguiendo el procedimiento introducido en Liu et al. (1999); de modo que las observaciones con poca profundidad, o gran perifericidad, obtienen pesos pequeños (Stahel (1981); Donoho and Gasko (1992)). $d(w_i)$ permite definir un nuevo peso para $\omega_h(\zeta_N, \zeta_i)$ basado en una estimación robusta de la profundidad de Mahalanobis. Por lo tanto, aplicando (4) en (3), la propuesta para el estimador robusto de Nadaraya-Watson se define como:

$$\omega_h^*(\zeta_N, \zeta_i) = \frac{d(w_i) K(d(\zeta_N, \zeta_i)/h)}{\sum_{i=1}^n d(w_i) K(d(\zeta_N, \zeta_i)/h)},$$

Se observa que este estimador penaliza los datos con poca profundidad. En la Sección 4 se presenta los principales resultados de este trabajo. Asimismo, dado que $d(w_i)$ está en función de estimadores robustos de localización y dispersión, se propone utilizar algunos estimadores conocidos de la literatura utilizada en Osorio et al. (2021) los estimadores son: la profundidad estándar de Mahalanobis MAH (Olive, 2004), Comedian (Falk, 1997), Fast-MCD (Rousseeuw and Van Driessen, 1999) MRD (Ro et al., 2015) y MRCD (Boudt et al., 2020).

3.1.2 Ancho de banda

El ancho de banda $h > 0$ juega un papel importante porque controla el número de términos en la media ponderada. En efecto, cuanto menor es h , menor es el número de ζ_{i+1} que se tienen en cuenta en la media. Por lo tanto, es más sensible $\hat{m}_{t,h}^{FNP}(\zeta_N)$ a las pequeñas variaciones de los ζ_{i+1} . En el caso contrario, cuanto mayor es h , mayor es el número de términos en la suma y menos sensible es $\hat{m}_{t,h}^{FNP}(\zeta_N)$ respecto a pequeñas variaciones de los ζ_{i+1} . En el contexto no paramétrico, h se denomina ancho de banda porque permite seleccionar el número de términos contenidos en la expresión del estimador (Ferraty and Vieu, 2006).

Diferentes métodos para la selección del ancho de banda se tuvieron en cuenta:

- Cálculo de ancho de banda de validación cruzada: encuentra un ancho de banda adecuado no paramétrico. Este método de selección de ancho de banda es basado en

datos, pueden ser muy exigentes desde el punto de vista computacional (Hayfield and Racine, 2008; Quintela and Vilar, 1991).

- Calculo de ancho de banda de función genérica densidad: calcula las estimaciones de la densidad del kernel. Su método por defecto lo hace con el kernel y el ancho de banda dados para la observación univariante funcionales (Sheather and Jones, 1991).

Por último, la base para elegir el ancho de banda óptimo es por medio de la minimización de la medida de error porcentual absoluto medio (MAPE, por sus siglas en inglés).

3.1.3 Índices de medición

En este trabajo se utilizan los siguientes índices de medición para validar el rendimiento de los modelos de predicción aplicados, también, para validar los métodos de selección del ancho de banda, estos índices son: el error absoluto medio (MAE, por sus siglas en inglés), la raíz del error cuadrático medio (RMSE, por sus siglas en inglés) y el MAPE para evaluar el rendimiento. Los valores relativamente bajos de estas medidas sugieren una mejor predicción.

- MAE: Utilizado con frecuencia para resolver problemas de regresión (Elizabeth Michael et al., 2022). Es el valor absoluto de la diferencia entre el valor previsto y el valor

observado:
$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n \left| \hat{\zeta}_i - \zeta_i \right|$$

- RMSE: Da mayor peso a los errores grandes. Cuando se analizan las previsiones en circunstancias en las que no se desean errores significativos, se utiliza:

$$\text{RMSE} = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n (\hat{\zeta}_i - \zeta_i)^2 \right)}$$

- MAPE: Es la media o promedio de los errores porcentuales absolutos de las previsiones:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{\zeta}_i - \zeta_i}{\zeta_i} \right| * 100$$

3.2 Modelo semi-funcional lineal parcial

En la Sección 2, se introdujo la existencia de variables exógenas que podrían mejorar las predicciones de las curvas semanales de demanda de productos en comparación con aquellos que solo consideran variables endógenas. Se considera una extensión de la Sección 3.1, propuesto por Vilar et al. (2018) aplicando el modelo, también, a respuesta funcional. Para una lectura más amplia se remite al lector a Aneiros et al. (2013).

Se introducen dos variables exógenas al modelo FNP como covariables funcionales, estas variables son la TRM y el IPC de Colombia, igualmente, se introduce la variable de descuentos aplicados en la venta. En este tipo de escenarios, parece natural generalizar el modelo FNP incorporando, en la función de regresión, una componente lineal con p

variables funcionales exógenas. Es decir, tomando como base la predicción $\zeta_{N+1}(t)$ donde t corresponde a las 52 semanas del año, el modelo semi-funcional lineal parcial (SFPL) se construye como:

$$\zeta_{i+1}(t) = X_{t+1}^T \beta_t + m_t(\zeta_i) + \xi_{t,i+1}, i = 1, \dots, n, \quad (5)$$

donde $X_{t+1}^T = (x_{i+1,1}, \dots, x_{i+1,p})$ es un vector de p variables exógenas funcionales y $\beta_t = (\beta_{t,1}, \dots, \beta_{t,p})^T$ es un vector de parámetros desconocidos a ser estimado y $\xi_{t,i+1}$ es el error funcional aleatorio con media cero. Se propone los estimadores para β_t y $m_t(\cdot)$ basados en los mínimos cuadrados ordinarios y en suavización por kernel, sus expresiones son:

$$\hat{\beta}_{t,h} = (\tilde{X}_h^T \tilde{X}_h)^{-1} \tilde{X}_h^T \tilde{\zeta}_{t,h}$$

y

$$\hat{m}_{t,h}^{SFPL}(\zeta) = \sum_{i=1}^n \omega_h(\zeta, \zeta_i) (\zeta_{i+1}(t) - X_{t+1}^T \hat{\beta}_{t,h}),$$

respectivamente. Tenga en cuenta que se denota $\tilde{X}_h = (I - W_h)X$ y $\tilde{\zeta}_{t,h} = (I - W_h)\zeta_t$, donde $W_h = (\omega_h(\zeta_i, \zeta_j))_{i+1,j+1}$, $X = (x_{i+1,j})_{i+1,1 \leq j \leq p}$ y $\zeta_t = (\zeta_{i+1}(t))_{i+1}$. Ahora, la previsión de $\zeta_{N+1}(t)$ se obtiene como:

$$\hat{\zeta}_{N+1}(t) = X_{N+1}^T \hat{\beta}_{t,h} + \hat{m}_{t,h}^{SFPL}(\zeta_N)$$

Si el objetivo es predecir un mes, Debe aplicarse procesos similares a los expuestos en la Sección 3.1 si, en lugar de predecir un valor semanal para el siguiente año, se realiza la predicción para el mes siguiente.

3.3 Intervalos de predicción

Cuando se trata de pronósticos, es importante considerar también los intervalos de predicción (IP), ya que las previsiones puntuales no siempre son suficientes cuando se quiere obtener predicciones para datos que incluyen mucha variabilidad. Los IP, se consideraron tanto para el modelo FNP como para el modelo SFPL. Un IP es un intervalo asociado a una variable aleatoria aún no observada, con una probabilidad determinada de que la variable aleatoria se encuentre dentro del intervalo. Un intervalo de confianza (IC) es un intervalo asociado a un parámetro. Se supone que el parámetro es desconocido, y el IC se calcula a partir de los datos. Como los datos son aleatorios, el intervalo es aleatorio (Hall et al., 2004).

Para hallar el IC en el modelo FNP, donde $m_t(\cdot)$ es el operador de la regresión, que es el valor esperado de la respuesta condicionada a la variable predictora. Es decir, $m_t(\zeta_N) = E(\zeta_{N+1}(t)|\zeta_N)$, entonces, se puede obtener un IC del $(1 - \alpha)100\%$ para esta expectativa condicionada a partir de su estimador $\hat{m}_{t,h}^{FNP}(\zeta_N)$. Este intervalo puede ser construido como: $(\hat{m}_{t,h}^{FNP}(\zeta_N) - q_{t,1-\alpha/2}(\zeta_N), \hat{m}_{t,h}^{FNP}(\zeta_N) + q_{t,1-\alpha/2}(\zeta_N))$, donde $q_{t,1-\alpha/2}$ y $q_{t,1-\alpha/2}(\zeta_N)$ son los cuantiles de la distribución $m_t(\zeta_N) - \hat{m}_{t,h}^{FNP}(\zeta_N)$. Este IC se dedica a cubrir el verdadero

valor del operador de regresión, $m_t(\zeta_N)$, y se ocupa únicamente de la variabilidad debida a su estimación.

Ahora, en el caso del IP, el procedimiento es similar, estos, se dedican a cubrir la respuesta y no el operador de regresión, por lo que incluyen, no sólo la variabilidad debida a la estimación, sino también el error en el modelo. Es decir, en lugar de trabajar con la expectativa de $\zeta_{N+1}(t)$ condicionada a ζ_N se trata de la distribución $\zeta_{N+1}(t)$ condicionada a ζ_N . El punto de partida para construir un IP es buscar un intervalo (a, b) tal que, $P((\zeta_{N+1}(t)|\zeta_N) \in (a, b)) = 1 - \alpha$ (Vilar et al., 2018).

3.3.1 Intervalos de predicción para el modelo FNP

Para hallar los IP, Vilar et al. (2018) propone un procedimiento bootstrap, para aproximar las distribuciones de $m_t(\zeta_N) - \widehat{m}_{t,h}^{FNP}(\zeta_N)$ y $\xi_t|\zeta_N$ que, en la práctica, son desconocidas. Es importante tener en cuenta que el procedimiento bootstrap se aplica para un modelo que se considera homocedástico, en el que la componente de error aleatorio del modelo tiene una varianza constante. El procedimiento para construir el IP para el modelo FNP homocedástico propuesto por Vilar et al. (2018), es el siguiente:

Dada una curva x el cual es un valor fijo del predictor funcional y una muestra $S = \{(x_i, Y_i), \dots, (x_n, Y_n)\}$ el cual es asumido de (1), donde $S \in \{\zeta_i, \zeta_{i+1}(t)\}_{t \in R}$

Los intervalos bootstrap $(1 - \alpha)$ para $\zeta_{N+1}(t)|\zeta_N$ se construyeron como:

$$I_{\zeta_N, 1-\alpha}^* = (\widehat{m}_{t,h}^{FNP}(\zeta_N) + q_{t,\alpha/2}^*(\zeta_N), \widehat{m}_{t,h}^{FNP}(\zeta_N) + q_{t,1-\alpha/2}^*(\zeta_N)),$$

los cuantiles bootstrap $q_{t,p}^*(\zeta_N)$ son calculados mediante el siguiente algoritmo:

Paso 1: Calcular $\widehat{m}_{t,b}^{FNP}(\zeta_i), i = 1, \dots, n$ sobre el conjunto de datos S , usando un ancho de banda b .

Paso 2: Calcular los residuales $\widehat{\xi}_{i,i+1} = \zeta_{i+1}(t) - \widehat{m}_{t,b}^{FNP}(\zeta_i)$ donde $i = 1, \dots, n$.

Paso 3: Aplicar el procedimiento de bootstrap para obtener los errores: Crear n variables aleatorias i.i.d ξ_1^*, \dots, ξ_n^* a partir de la función de distribución empírica de $(\widehat{\xi}_{1,b} - \widehat{\xi}_b, \dots, \widehat{\xi}_{n,b} - \widehat{\xi}_b)$, donde $\widehat{\xi}_b = n^{-1} \sum_{i=1}^n \widehat{\xi}_{i,b}$.

Paso 4: Calcular $\zeta_{i+1}^*(t) = \widehat{m}_{t,b}^{FNP}(\zeta_i) + \xi_i^*, i = 1, \dots, n$ y $\widehat{m}_{t,hb}^{FNP}(\zeta_N) = \frac{\sum_{i=1}^n K(d(\zeta_N, \zeta_i)/h) \zeta_{i+1}^*(t)}{\sum_{i=1}^n K(d(\zeta_N, \zeta_i)/h)}$.

Paso 5: Repetir B veces los pasos 3-4, obteniendo B estimaciones $\left\{ \widehat{m}_{t,hb}^{*,r}(\zeta_N) + \tilde{\xi}^r \right\}_{r=1}^B$.

Paso 6: Crear B variables aleatorias i.i.d $\tilde{\xi}^1, \dots, \tilde{\xi}^B$ a partir de la función de distribución empírica de los residuos centrados en el Paso 2. $\tilde{\xi}$ se aproxima al error del modelo.

Paso 7: Calcular el conjunto de errores bootstrap:

$$E_{boot} = \left\{ \widehat{m}_{t,b}^{FNP}(\zeta_N) - \widehat{m}_{t,hb}^{*,r}(\zeta_N) + \tilde{\xi}^r \right\}_{r=1}^B$$

Paso 8: Calcular el cuantil bootstrap $q_{t,p}^*(\zeta_N)$ del cuantil de orden p de E_{boot} .

Cuando se trabaja con procedimientos bootstrap en la regresión no paramétrica, en estos algoritmos intervienen dos anchos de banda. El primer ancho de banda, b , se utiliza para construir los residuos que se van a remuestrear, y un segundo ancho de banda, h , se considera para suavizar la muestra bootstrap. La obtención de la validez asintótica de los procedimientos bootstrap requiere que b sea mayor que h (Raña et al., 2016).

3.3.2 Intervalos de predicción para el modelo SFPL

Para el modelo SFPL, el procedimiento es análogo. En este caso, la muestra será S , donde $S = \{(X_i, x_i, Y_i), \dots, (X_n, x_n, Y_n)\}$. Basado en la muestra para el modelo SFPL el cual es asumido de (5). La variable predictora para $\zeta_{N+1}(t) / \{X_{N+1}, \zeta_N\}$ es $\widehat{\zeta}_{N+1}(t) / \{X_{N+1}, \zeta_N\} = X_{N+1}^T \widehat{\beta}_{t,h} + \widehat{m}_{t,h}^{SFPL}(\zeta_N)$. Ahora la distribución está enfocada en $\zeta_{N+1}(t)$ condicionada por las covariables funcionales no paramétricas X_{N+1} y ζ_N . Se tiene la siguiente descomposición:

$$\begin{aligned} \zeta_{N+1}(t) / \{X_{N+1}, \zeta_N\} &= X_{N+1}^T \widehat{\beta}_{t,h} + X_{N+1}^T (\beta - \widehat{\beta}_{t,h}) + \widehat{m}_{t,h}^{SFPL}(\zeta_N) \\ &\quad + (m_t(\zeta_N) - \widehat{m}_{t,h}^{SFPL}(\zeta_N)) + (\xi_{t,N+1} | \{X_{N+1}, \zeta_N\}) \end{aligned}$$

Como los valores verdaderos de la función de regresión $m(t)$ y vector de parámetros β_t son desconocidos en la práctica, es necesario aproximar $(\beta_t - \widehat{\beta}_{t,h})$ y $(m_t(\zeta_N) - \widehat{m}_{t,h}^{SFPL}(\zeta_N))$ y el término del error $\xi_{t,N+1} | \{X_{N+1}, \zeta_N\}$. Es importante tener en cuenta que el procedimiento bootstrap se aplica para un modelo que se considera homocedástico.

El intervalo de predicción bootstrap $(1-\alpha)$ para $\zeta_{N+1}(t) | \{X_{N+1}, \zeta_N\}$ se construye como: $I_{\{X_{N+1}, \zeta_N\}, 1-\alpha}^* = (L_t, U_t)$ donde,

$$L_t = X_{N+1}^T \widehat{\beta}_{t,h} + \widehat{m}_{t,h}^{SFPL}(\zeta_N) + q_{t,\alpha/2}^*(X_{N+1}, \zeta_N)$$

y

$$U_t = X_{N+1}^T \widehat{\beta}_{t,h} + \widehat{m}_{t,h}^{SFPL}(\zeta_N) + q_{t,1-\alpha/2}^*(X_{N+1}, \zeta_N)$$

los cuantiles bootstrap $q_{t,p}^*(X_{N+1}, \zeta_N)$ son calculados mediante el siguiente algoritmo:

Paso 1: Calcular $\widehat{\beta}_{t,b}$ y $\widehat{m}_{t,b}^{SFPL}(\zeta_i)$, $i = 1, \dots, n$ sobre el conjunto de datos S

Paso 2: Calcular los residuales $\widehat{\xi}_{t,i+1} = \zeta_{i+1}(t) - X_i^T \widehat{\beta}_{t,b} - \widehat{m}_{t,b}^{SFPL}(\zeta_i)$ donde $i = 1, \dots, n$.

Paso 3: Aplicar el procedimiento de bootstrap para obtener los errores: Crear n variables aleatorias i.i.d ξ_1^*, \dots, ξ_n^* a partir de la función de distribución empírica de $(\widehat{\xi}_{1,b} - \widehat{\xi}_b, \dots, \widehat{\xi}_{n,b} - \widehat{\xi}_b)$, donde $\widehat{\xi}_b = n^{-1} \sum_{i=1}^n \widehat{\xi}_{i,b}$.

Paso 4: Calcular $\zeta_{i+1}^*(t) = X_{i+1}^T \hat{\beta}_{t,b} + \hat{m}_{t,b}^{FNP}(\zeta_i) + \xi_i^*$, $i = 1, \dots, n$ y los estimadores bootstrap

$$\hat{\beta}_{t,b}^* = (\tilde{X}_b^T \tilde{X}_b)^{-1} \tilde{X}_b^T \tilde{\zeta}_{t,b}^*$$

y

$$\hat{m}_{t,hb}^{SFP L^*}(\zeta_N) = \sum_{i=1}^n \omega_h(\zeta_i, \zeta_N) (\zeta_{i+1}^* - X_{i+1}^T \hat{\beta}_{t,b}^*)$$

Paso 5: Repetir B veces los pasos 3-4, obteniendo B estimaciones $\left\{ \hat{\beta}_{t,b}^{*,r} \right\}_{r=1}^B$ y $\left\{ \hat{m}_{t,hb}^{*,r}(\zeta_N) \right\}_{r=1}^B$.

Paso 6: Crear B variables aleatorias i.i.d $\tilde{\xi}^1, \dots, \tilde{\xi}^B$ a partir de la función de distribución empírica de los residuos centrados en el Paso 2. $\tilde{\xi}$ se aproxima al error del modelo.

Paso 7: Calcular el conjunto de errores bootstrap:

$$E_{boot} = \left\{ X^T (\hat{\beta}_{t,b} - \hat{\beta}_{t,b}^{*,r}) + (\hat{m}_{t,b}(\zeta_N) - \hat{m}_{t,hb}^{*,r}(\zeta_N)) + \tilde{\xi}_{t,r} \right\}_{r=1}^B$$

Paso 8: Calcular el cuantil bootstrap $q_{t,p}^*(X_{N+1}, \zeta_N)$ del cuantil de orden p de E_{boot} .

4 Resultados

En esta Sección se presentan los resultados relacionados con la previsión para el año siguiente de las curvas semanales de demanda de productos de maquinaria agrícola, utilizando los enfoques descritos en la Sección 3. Las curvas semanales de demanda de productos deben predecirse para cada semana del año 2019, utilizando la información de las 52 semanas anteriores. Métodos para la detección de valores atípicos no fueron aplicados, en su lugar, se aplicó una versión robusta del estimador de Nadaraya-Watson, también, se aplicó este método para las variables exógenas y para la variable de descuentos que se dan en la venta.

Es importante recordar el objetivo propuesto en este trabajo, el cual es implementar un modelo de regresión no paramétrica con un enfoque robusto para el alisado de los datos, con el fin de obtener estimaciones que no se alteren por la presencia de puntos atípicos y que a su vez contribuyan a una correcta descripción de la trayectoria de los datos. Estas mejoras son tan importantes como la predicción en sí. Como se ha mencionado en las secciones anteriores, el parámetro de suavizado h tiene que ser seleccionado adecuadamente para asegurar un buen rendimiento práctico. Además, también se sabe que la selección de este parámetro de suavizado es de gran interés, en la Figura 2, se presenta diferentes valores de h calculados con los métodos expuestos en la Sección 3.1.2.

Se observa que, entre más grande el valor de h más suave es el alisado, es decir, menos sensible es la curva a los diferentes valores de la variable respuesta. Asimismo, menos sensible es a valores atípicos presentados en la demanda de los productos. Los tres primeros valores de h fueron calculados basados en los datos, mientras que los dos últimos valores de h fueron calculados dado las valores univariante, es decir, dado los x_i . Para efectos del presente trabajo, el h seleccionado será el de mejor alisado de los datos, sin perder la forma

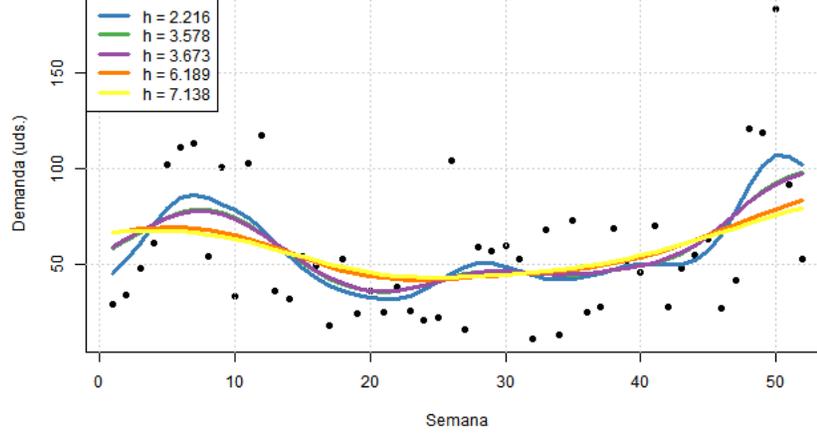


Figura 2: Comportamiento de la demanda de productos de la División Bombas y el alisado con base en el ancho de banda (h).

de estos, es decir, se continuará trabajando los modelos con $h = 7,138$.

En adelante, para comparar la precisión de cada modelo considerado y la previsión obtenida, ζ_{N+1} , a partir de los modelos de la Sección 3, se utilizaron los índices de medición de la Sección 3.1.3, sin embargo, el MAPE fue la medida comparativa de los resultados de las predicciones. Para explorar y validar el rendimiento de la robustez del estimador propuesto en 3.1, se aplicó en cada uno de los conjuntos de datos explicados en la Sección 2. A continuación se presenta cada una de las gráficas. Cada uno de los conjuntos de datos contienen valores atípicos en mayor o menor medida.

En la Figura 3, se observa el rendimiento del Nadaraya-Watson propuesto por Vilar et al. (2018) con el Nadaraya-Watson robusto propuesto en este documento. Este último con diferentes estimadores de localización y dispersión, se compara la profundidad clásica de Mahalanobis NW-MAH, NW-MCD (Fast-MCD), NW-COM (Comedian), NW-MRD y NW-MRCD. También se muestran los diferentes comportamientos de los modelos en cada conjunto de datos. En la Figura 3a se observa que los estimadores robustos obtienen resultados más precisos en presencia de valores atípicos, es decir, continúan la trayectoria según la forma de los datos y no se dejan influenciar por los valores atípicos, en comparación con el NW estándar. Como se presenta en la expresión 3.1.1, la inclusión de una función de peso basada en la profundidad trata de corregir la estimación de la regresión en los escenarios donde se presenta valores atípicos. Para la mayoría de los conjuntos de datos, el NW-MRCD obtiene las estimaciones más estables o precisas.

El Cuadro 1 muestra el MAPE para cada uno de los modelos propuestos y el suavizado con el menor MAPE es el que se usará para el modelo SFPL propuesto en 3.2. De la tabla se concluye también que en datos con mayor variabilidad (Demanda y Descuentos) los

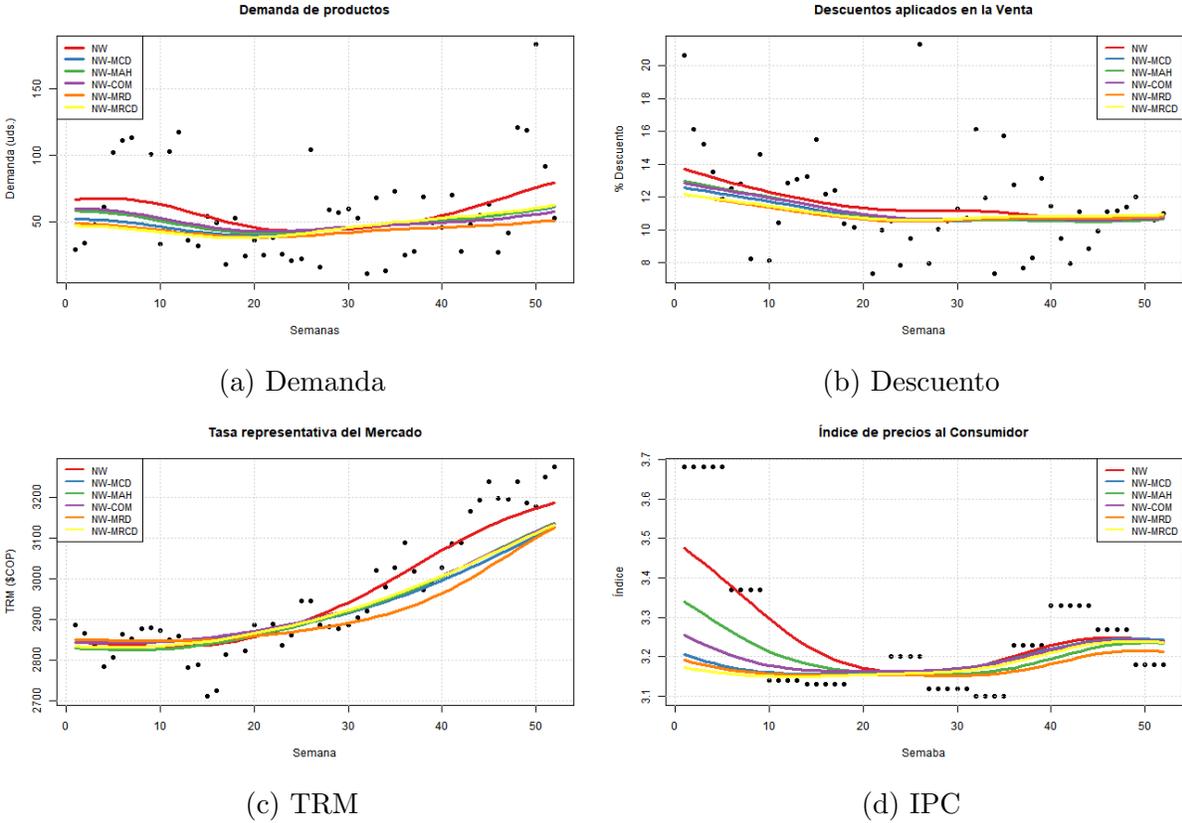


Figura 3: Modelos de Regresión robusta de Nadaraya-Watson aplicados a los datos presentados en la Sección 2

métodos robustos se comportan mejor que el Nadaraya-Watson estándar, mientras en la forma de los datos de la TRM y del IPC pasa todo lo contrario. Es importante aclarar que estos modelos se pueden extrapolar a otra unidad de tiempo, por ejemplo, mes, semana del mes y día. Más aún, se presentará los resultados de la predicción de la demanda para el año 2019 tomando como base la información del año 2018 que se acaba de procesar, involucrando variables exógenas que permitan mejorar el pronóstico. Se aplicará el modelo propuesto en la Sección 3.2 con los nuevos conjuntos de datos funcionales.

Es importante destacar el hecho de que las covariables, TRM e IPC, incluidas en los modelos SFPL no son observadas. Por lo tanto, para poner en práctica el procedimiento de SFPL, es necesario tener a mano buenas previsiones para el siguiente año de los valores de esas covariables. Sin embargo, es de conocimiento que existen sofisticados modelos para trazar el comportamiento de la TRM e IPC que realizan las entidades encargadas de pronosticar dichos comportamientos futuros. Aunque, ni esos modelos ni las previsiones son públicas, es posible simularlas como escenarios optimistas o pesimistas. Para no enmascarar el poder de predicción del modelo SFPL, se optó por incorporar las previsiones ideales dadas por los valores propios.

Modelos.FNP	Demanda			Descuentos		
	MAPE	MAE	RMSE	MAPE	MAE	RMSE
NW	62.01	25.21	31.23	17.05	1.90	2.69
NW-MCD	54.91	25.24	34.58	16.46	1.94	2.80
NW-MAH	56.28	25.05	33.82	16.38	1.91	2.76
NW-COM	56.76	25.17	34.14	16.48	1.91	2.77
NW-MRD	51.50	25.80	36.42	16.67	1.98	2.86
NW-MRCD	54.28	25.44	35.21	16.68	1.97	2.85
	TRM			IPC		
NW	1.44	42.71	51.19	2.22	0.07	0.10
NW-MCD	1.95	58.94	77.56	2.88	0.10	0.17
NW-MAH	1.89	57.06	72.18	2.61	0.09	0.13
NW-COM	1.88	56.58	72.96	2.78	0.10	0.16
NW-MRD	2.16	65.66	87.15	3.02	0.10	0.17
NW-MRCD	1.90	57.47	73.29	2.97	0.10	0.18

Cuadro 1: Índices de medición para las curvas de demanda de productos para cada modelo de regresión robusta de Nadaraya-Watson.

En la Figura 4 se presenta el pronóstico de la demanda, con base en las curvas de la TRM, IPC y los descuentos aplicados que se trabajaron con las pautas del modelo FNP (Sección 3.1). Se consideraron los nuevos predictores para el pronóstico de la demanda para el año 2019 obtenidos a partir de la aplicación del enfoque del modelo no paramétrico combinado con la información de las 52 semanas previas. El ajuste del modelo es mucho mejor, es importante aclarar que el propósito de la aplicación del modelo no paramétrico de Nadaya-Watson combinado con métodos robustos es el alisado de los datos, por ende, el juste puede no necesariamente ser el mejor u óptimo. Por ejemplo, en el Cuadro 1 se presentó el comparativo de los índices de medición en el que se observa el mejor MAPE para el pronóstico de la demanda el cual fue de 51,50, comparado con el MAPE presentado con el modelo SFPL de 3,60. Asimismo, cabe señalar que, desde el punto de vista de la bondad de ajuste, el rendimiento del modelo SFPL es mejor que el correspondiente al modelo FNP (el coeficiente de determinación del modelo SFPL ajustado es de 0,97 mientras que el del modelo FNP es de 0,25). Para solucionar este inconveniente el procedimiento de construcción de los IP con el modelo FNP es posible aplicarlo nuevamente pero utilizando un menor ancho de banda en los estimadores mejorando así la precisión.

Como se expuso anteriormente, es importante cuando se trabaja con predicciones calcular los IP. La construcción de los IP de la demanda de productos de maquinaria agrícola se hizo con base en los modelos homocedásticos desarrollados en la Sección 3.3. Esta elección de la homocedasticidad está avalada por algunos análisis estadístico, por ejemplo, se consideró la prueba de Breusch and Pagan (1979) (BP), aplicada a la demanda semanal de los productos en los cuatro años del conjunto de datos, asimismo, se verificó por métodos gráficos, como la gráfica de dispersión de los valores ajustados contra los residuos estandarizados determinando si la variabilidad es la misma entre esos periodos de tiempo. Con la prueba BP, la hipótesis nula es establecer homocedasticidad en los errores. Se obtiene que no se puede rechazar la hipótesis nula de ser homocedástica los datos a un nivel de confianza del

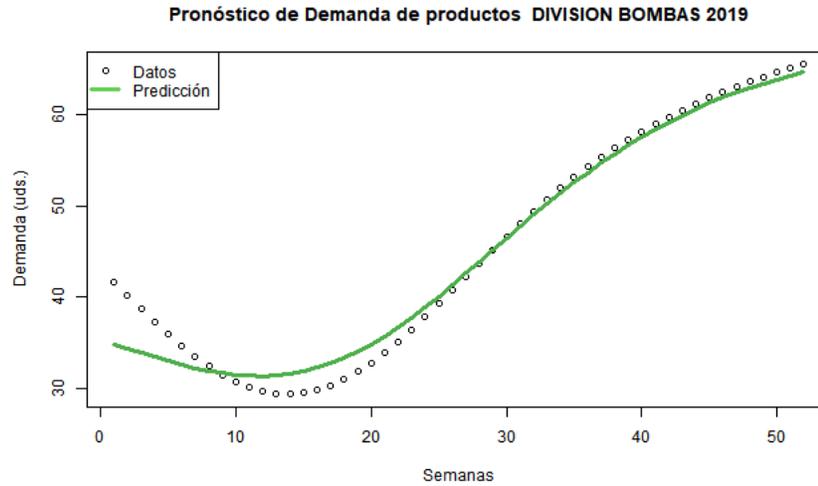


Figura 4: Predicción de la demanda de productos de la División Bombas para el año 2019 usando SFPL (línea verde) incluyendo variables exógenas.

95 %. Aplicando los procedimientos correspondientes, se obtienen los IP para la demanda semanal de productos de los modelos FNP y SFPL. Cada procedimiento bootstrap se generó con 1000 replicas.

En la Fifura 5 se presentan los IP del modelo FNP, en la gráfica se observa: valor esperado, es decir, las predicciones (línea azul), intervalos de predicción, es decir, rango para las predicciones sobre una observación individual (zona gris) y los valores reales observados de la demanda para el año 2019. Aunque como objeto de este documento no está el cálculo de los IC, es relevante expresar que estos son más estrechos, ya que solo se refieren a la incertidumbre en la estimación del modelo del valor esperado. Mientras que los IP son más amplios ya que se refieren tanto a la incertidumbre del modelo como a la incertidumbre del muestreo de cualquier observación, la última de las cuales contribuye a una varianza mucho mayor.

El aporte de los IP en los pronósticos de ventas es que presenta una posibilidad de inferir en los datos con base en el rango de valores que otorga el IP alrededor del valor de la predicción puntual.

La Figura 6 presenta el IP para el modelo SFPL, este IP tiene un menor amplitud porque existe menor variabilidad en los datos.

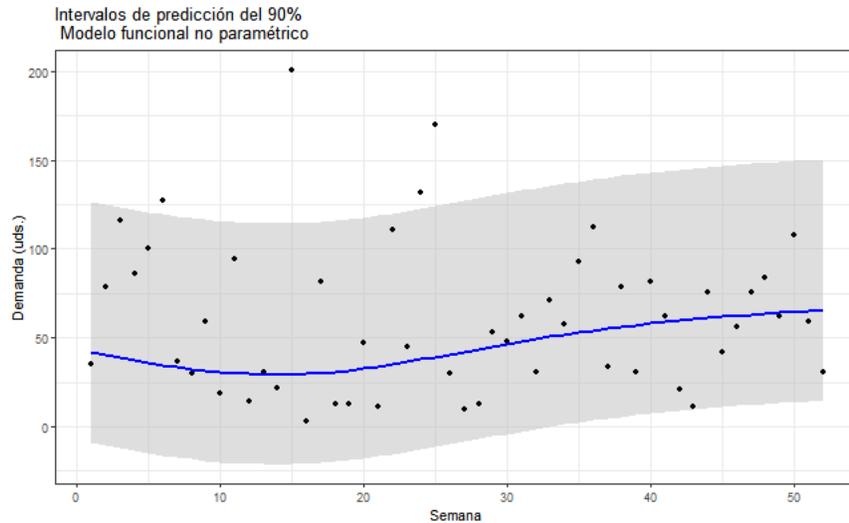


Figura 5: Intervalo de Predicción de la demanda de productos de la División Bombas sobre el modelo FNP para el año 2019.

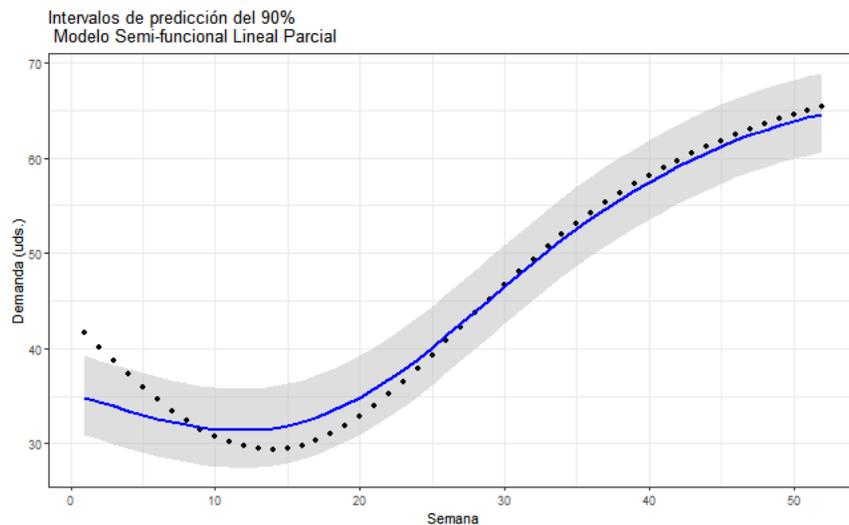


Figura 6: Intervalo de Predicción de la demanda de productos de la División Bombas para el año 2019 sobre el modelo SFPL incluyendo variables exógenas.

5 Conclusiones

Nuevos métodos para abordar la previsión de la demanda de productos de maquinaria agrícola se han utilizado con éxito, estos métodos no paramétricos funcionales permiten la previsión de la demanda a partir de datos funcionales arrojan buenos resultados ante la presencia de datos atípicos en las observaciones e involucrando variables exógenas, por otro lado, se ha presentado una versión robusta del estimador de Nadaraya-Watson, basada en la inclusión de una función de peso con enfoque en la profundidad, de acuerdo con los resultados obtenidos de las aplicaciones en los conjuntos de datos, es posible concluir

que este estimador presenta propiedades de robustez y mejor precisión en la previsión en presencia de valores atípicos.

Las versiones robustas presentadas en este trabajo tienen buen comportamiento respecto al alisado, ya que las curvas generadas por los métodos se ajustan muy bien a la forma general de los datos, es decir, no se muestran influenciadas ante la presencia de valores atípicos. Es importante considerar que tener presente variables exógenas permite mejorar la previsión de las ventas y es de gran ayuda para las compañías que comercializan este tipo de productos ya que la demanda se ve muy influenciada por la economía del país, el clima, los costos de transporte en las importaciones o el tiempo que demora un producto en estar disponible para venta, por ejemplo.

A pesar de su complejidad, el modelo SFPL puede aplicarse fácilmente a datos de alta frecuencia, por ejemplo, datos de 1 día o la semana inmediatamente anterior, en caso de la tasa representativa del mercado o el tiempo de llegada de los productos, cuando estén disponibles. Todas estas características hacen que este enfoque sea atractivo y con mucho potencial para mejorar.

Los métodos aplicados en el presente trabajo permitieron abordar una solución a una de las tareas importantes en la construcción del presupuesto comercial en las compañías, se trata de la limpieza o suavizado de las ventas extraordinarias o ventas perdidas por falta de stock en los históricos de ventas, los cuales, se convierten en observaciones atípicas. Se espera además, con el presente trabajo, mejorar el MAPE del presupuesto comercial actual, ya que éste no tiene en cuenta variables del mercado que pueden afectar el comportamiento de compra del consumidor. Asimismo, combinar los métodos aplicados en el modelo FNP da como resultados modelos que podrían aportar mucho a la mejora de la precisión de los pronósticos de la demanda cuando se presenta alta variabilidad en las ventas y la exploración de otros elementos como trabajo futuro, tales como medidas de profundidad, mayores dimensiones del espacio muestral o estructuras de dependencia multivariadas en caso de múltiples covariables.

Referencias

- Abolghasemi, M., Beh, E., Tarr, G., and Gerlach, R. (2020), “Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion,” *Comput. Ind. Eng.*, 142.
- Acar, Y. and Gardner, Jr, E. S. (2012), “Forecasting method selection in a global supply chain,” *Int. J. Forecast.*, 28, 842–848.
- Alma, Ö. G. (2011), “Comparison of Robust Regression Methods in Linear Regression,” *Int. J. Contemp. Math. Sciences*, 6, 409–421.
- Aneiros, G., Vilar, J., and Raña, P. (2016), “Short-term forecast of daily curves of electricity demand and price,” *Int. j. electr. power energy syst.*, 80, 96–108.

- Aneiros, G., Vilar, J. M., Cao, R., and Munoz San Roque, A. (2013), “Functional prediction for the residual demand in electricity spot markets,” *IEEE Trans. Power Syst.*, 28, 4201–4208.
- Banco de la República de Colombia (2022), URL <https://www.banrep.gov.co/es>. Accedido: 2022-6-30.
- Birkes, D. and Dodge, Y. (1993), *Alternative Methods of Regression*, Wiley Series in Probability and Statistics, Nashville, TN: John Wiley & Sons.
- Boudt, K., Rousseeuw, P. J., Vanduffel, S., and Verdonck, T. (2020), “The minimum regularized covariance determinant estimator,” *Stat. Comput.*, 30, 113–128.
- Bozkir, A. S. and Sezer, E. A. (2011), “Predicting food demand in food courts by decision tree approaches,” *Procedia Comput. Sci.*, 3, 759–763.
- Breusch, T. S. and Pagan, A. R. (1979), “A Simple Test for Heteroscedasticity and Random Coefficient Variation,” *Econometrica*, 47, 1287.
- Departamento Administrativo Nacional de Estadística (2022), URL <https://www.dane.gov.co/>. Accedido: 2022-6-30.
- Donoho, D. L. and Gasko, M. (1992), “Breakdown properties of location estimates based on halfspace depth and projected outlyingness,” *Ann. Stat.*, 20, 1803–1827.
- Elizabeth Michael, N., Mishra, M., Hasan, S., and Al-Durra, A. (2022), “Short-term solar power predicting model based on multi-step CNN stacked LSTM technique,” *Energies*, 15, 2150.
- Eubank, R. L. and Hsing, T. (2015), *Theoretical foundations of functional data analysis, with an introduction to linear operators*, Wiley Series in Probability and Statistics, Hoboken, NJ: Wiley-Blackwell.
- Falk, M. (1997), “On mad and comedians,” *Annals of the Institute of Statistical Mathematics*, 49, 615–644.
- Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2011), “Kernel regression with functional response,” *Electron. J. Stat.*, 5, 159–171.
- Ferraty, F. and Vieu, P. (2006), *Nonparametric functional data analysis*, Springer Series in Statistics, New York, NY: Springer, 2006 edition.
- Fildes, R., Ma, S., and Kolassa, S. (2019), “Retail forecasting: Research and practice,” *Int. J. Forecast.*
- Gilliland, M. (2010), *The business forecasting deal*, Wiley and SAS Business Series, Chichester, England: John Wiley & Sons.

- Hall, P., Hyndman, R. J., and Fan, Y. (2004), “Nonparametric confidence intervals for receiver operating characteristic curves,” *Biometrika*, 91, 743–750.
- Hayfield, T. and Racine, J. S. (2008), “Nonparametric Econometrics: ThenpPackage,” *J. Stat. Softw.*, 27.
- Ho, K. W. and Naugher, J. R. (2000), “Outlier Lies: An Illustrative Example of Identifying Outliers and Applying Robust Models,” *Multiple Linear Regression Viewpoints, 2000, Vol. 26(2)*, 4–8.
- Horová, I., Vieu, P., and Zelinka, J. (2002), “Optimal choice of nonparametric estimates of a density and of its derivatives,” *Stat. risk model.*, 20.
- Kirkpatrick, R. C. and Gaynor, P. E. (1994), *Introduction to time series modeling and forecasting for business and economics*, New York, NY: McGraw-Hill.
- Kokozka, P. and Horvath, L. (2012), *Inference for functional data with applications*, Springer series in statistics, New York, NY: Springer, 2012 edition.
- Liu, R. Y., Parelius, J. M., and Singh, K. (1999), “Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by Liu and Singh),” *Ann. Stat.*, 27, 783–858.
- Maronna, R. A. and Zamar, R. H. (2002), “Robust estimates of location and dispersion for high-dimensional datasets,” *Technometrics*, 44, 307–317.
- Nadaraya, E. A. (1964), “On Estimating Regression,” *Theory Probab. Appl.*, 9, 141–142.
- Narayanan, A., Sahin, F., and Robinson, E. P. (2019), “Demand and order-fulfillment planning: The impact of point-of-sale data, retailer orders and distribution center orders on forecast accuracy,” *J. Oper. Manage.*, 65, 468–486.
- Nguyen, T., Zhou, L., Spiegler, V., Ieromonachou, P., and Lin, Y. (2018), “Big data analytics in supply chain management: A state-of-the-art literature review,” *Comput. Oper. Res.*, 98, 254–264.
- Olive, D. J. (2004), “A resistant estimator of multivariate location and dispersion,” *Comput. Stat. Data Anal.*, 46, 93–102.
- Osorio, P., Ortiz, S., and Laniado, H. (2021), “Weighted Nadaraya-Watson Kernel Regression Based on a Robust Mahalanobis-Depth Estimation,” *Este artículo pertenece a las Memorias del Simposio Internacional de Estadística XXX Versión Evento-Virtual Septiembre 21 a 24 de 2021*, 272–277.
- Peña, D. and Prieto, F. J. (2001), “Multivariate outlier detection and robust covariance matrix estimation,” *Technometrics*, 43, 286–310.

- Quintela, A. and Vilar, J. (1991), “Técnicas de validación cruzada en la estimación de la densidad bajo condiciones de dependencia,” *Questió (Quaderns d’Estadística i Investigació Operativa)*; Vol.: 15 Núm.: 1, 15, 21–45.
- Ramsay, J. O. and Silverman, B. W. (2002), *Applied functional data analysis*, Springer Series in Statistics, New York, NY: Springer, 2002 edition.
- (2005), *Functional Data Analysis*, Springer Series in Statistics, New York, NY: Springer, 2 edition.
- Raña, P. (2016), *Pointwise forecast, con dence and prediction intervals in electricity demand and price*, Ph.D. thesis, Universidade da Coruña. Departamento de Matemáticas.
- Raña, P., Aneiros, G., Vilar, J., and Vieu, P. (2016), “Bootstrap confidence intervals in functional nonparametric regression under dependence,” *Electron. J. Stat.*, 10, 1973–1999.
- Ro, K., Zou, C., Wang, Z., and Yin, G. (2015), “Outlier detection for high-dimensional data,” *Biometrika*, 102, 589–599.
- Rousseeuw, P. J. (1984), “Least median of squares regression,” *J. Am. Stat. Assoc.*, 79, 871–880.
- Rousseeuw, P. J. and Van Driessen, K. (1999), “A fast algorithm for the minimum covariance determinant estimator,” *Technometrics*, 41, 212–223.
- Schumacker, R., Monahan, M. P., and Mount, R. E. (2002), “A Comparison of OLS and Robust Regression using S-PLUS,” *Multiple Linear Regression*, 10–13.
- Seyedan, M. and Mafakheri, F. (2020), “Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities,” *J. Big Data*, 7.
- Seyedan, M., Mafakheri, F., and Wang, C. (2022), “Cluster-based demand forecasting using Bayesian model averaging: An ensemble learning approach,” *Decision Analytics Journal*, 3, 100033.
- Sheather, S. J. and Jones, M. C. (1991), “A reliable data-based bandwidth selection method for kernel density estimation,” *J. R. Stat. Soc.*, 53, 683–690.
- Stahel, W. A. (1981), *Robuste Schätzungen: infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen*, Ph.D. thesis, Universität Zürich.
- Staudte, R. G. and Sheather, S. J. (1990), *Robust Estimation and Testing*, Wiley Series in Probability and Statistics, Nashville, TN: John Wiley & Sons.
- Stepnicka, M., Peralta Donate, J., Cortez, P., Vavrikova, L., and Gutierrez, G. (2011), “Forecasting seasonal time series with computational intelligence: contribution of a combination of distinct methods,” in *Proceedings of the 7th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-2011)*, Paris, France: Atlantis Press.

- Thiesing, F. M. and Vornberger, O. (1997), “Forecasting sales using neural networks,” in *Computational Intelligence Theory and Applications*, Lecture notes in computer science, Berlin, Heidelberg: Springer Berlin Heidelberg, 321–328.
- Velasco, H., Laniado, H., Toro, M., Leiva, V., and Lio, Y. (2020), “Robust three-step regression based on comedian and its performance in cell-wise and case-wise outliers,” *Mathematics*, 8, 1259.
- Vilar, J., Aneiros, G., and Raña, P. (2018), “Prediction intervals for electricity demand and price using functional data,” *Int. j. electr. power energy syst.*, 96, 457–472.
- Vilar, J. M., Cao, R., and Aneiros, G. (2012), “Forecasting next-day electricity demand and price using nonparametric functional methods,” *Int. j. electr. power energy syst.*, 39, 48–55.
- Wang, G., Gunasekaran, A., Ngai, E. W. T., and Papadopoulos, T. (2016), “Big data analytics in logistics and supply chain management: Certain investigations for research and applications,” *Int. J. Prod. Econ.*, 176, 98–110.
- Watson, G. S. (1964), “Smooth Regression Analysis,” *Sankhya, Ser. A*, 26, 359–372.
- Wilcox, R. R. (2004), *Introduction to robust estimation and hypothesis testing*, Statistical Modeling and Decision Science, San Diego, CA: Academic Press, 2 edition.