



UNS
ESCUELA DE
POSGRADO

**“MODELOS DE DATOS PANEL PARA PRONÓSTICO DE LA
DESNUTRICIÓN INFANTIL EN CINCO REGIONES DEL NORTE
PERUANO, PERIODO 2010 – 2018”**

**Tesis para optar el grado de
Doctor en Estadística Matemática**

Autor:

Ms. Betsy Pamela Arrunátegui Huamán

Asesor:

Dr. Carlos Alberto Minchón Medina

**CHIMBOTE - PERÚ
2021**



UNS
ESCUELA DE
POSGRADO

CONSTANCIA DE ASESORAMIENTO DE LA TESIS DOCTORAL

Yo, Dr. Carlos Alberto Minchón Medina, mediante la presente certifico mi asesoramiento de la Tesis Doctoral titulada: "MODELOS DE DATOS PANEL PARA PRONÓSTICO DE LA DESNUTRICIÓN INFANTIL EN CINCO REGIONES DEL NORTE PERUANO, PERIODO 2010 - 2018", elaborada por la magister Betsy Pamela Arrunategui Huamán para obtener el Grado Académico de Doctor en Estadística Matemática en la Escuela de Posgrado de la Universidad Nacional del Santa.

Nuevo Chimbote, 18 de Agosto del 2021

Dr. Carlos Alberto Minchón Medina
Asesor



UNS
FACULTAD DE
POSGRADO

CONFORMIDAD DEL JURADO EVALUADOR

"MODELOS DE DATOS PANEL PARA PRONÓSTICO DE LA DESNUTRICIÓN
INFANTIL EN CINCO REGIONES DEL NORTE PERUANO, PERIODO 2010 – 2018"

TESIS PARA OPTAR EL GRADO DE DOCTOR en Estadística Matemática

Revisado y Aprobado por el Jurado Evaluador:

Dr. Alfonso Tesen Arroyo
PRESIDENTE

Dra. Yheni Farfán Machaca
SECRETARIA

Dr. Carlos Alberto Minchón Medina
VOCAL

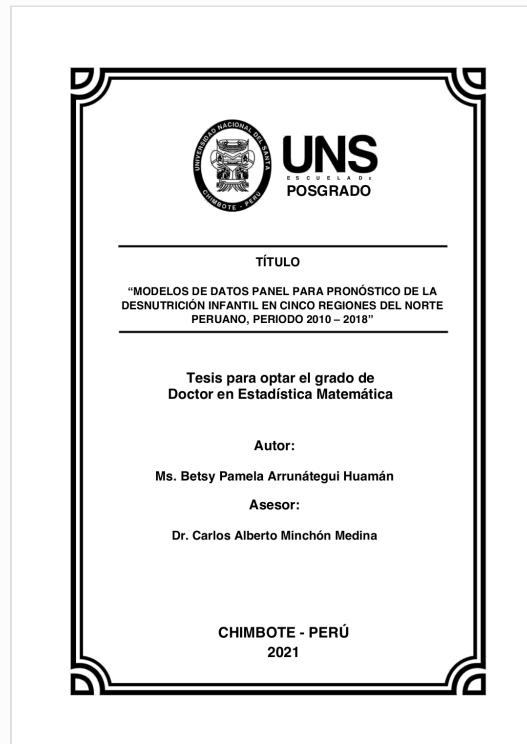


Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Betsy Arrunategui Huaman
Assignment title: PROYECTOS
Submission title: MODELOS DE DATOS PANEL PARA PRONÓSTICO DE LA DESN...
File name: Informe_Final-Betsy_Pamela_Arrunategui_Huam_n.docx
File size: 315.59K
Page count: 48
Word count: 12,091
Character count: 64,312
Submission date: 18-Aug-2021 03:48PM (UTC-0500)
Submission ID: 1632955312



Índice General

RESUMEN.....	7
ABSTRACT.....	8
INTRODUCCIÓN.....	9
CAPÍTULO I.....	12
PROBLEMA DE INVESTIGACIÓN	12
1.1. Planteamiento y fundamentación del problema de investigación	12
1.2. Antecedentes de la investigación	14
1.3. Formulación del problema de investigación	17
1.4. Delimitación del estudio	17
1.5. Justificación e importancia de la investigación	18
1.6. Objetivos de la investigación:	19
CAPÍTULO II.....	20
MARCO TEÓRICO	20
2.1. Fundamentos teóricos de la investigación.....	20
Modelos Econométricos aplicados en el área de salud	20
Modelo de datos panel.....	23
Regresiones en datos de panel.....	25
Regresión agrupada (pooled ols)	26
Modelo de efectos fijos	27
Modelo de efectos aleatorios	28
Prueba de Hausman	29
Autocorrelación	30
Correlación contemporánea	31
Heterocedasticidad	31
Errores estándar corregidos para panel (PCSE)	32
Modelos presentes en la investigación presentada	33
2.2. Marco conceptual.....	34
CAPÍTULO III.....	35
MARCO METODOLÓGICO	35
3.1. Hipótesis central de la investigación	35
3.2. Variables e indicadores de la investigación	35
3.3. Métodos de la investigación	36
3.4. Diseño o esquema de la investigación	36
3.5. Población y muestra.....	37
3.6. Actividades del proceso investigativo	37
3.7. Técnicas e instrumentos de la investigación	38

3.8. Procedimiento para la recolección de datos	38
3.9. Técnicas de procesamiento y análisis de los datos.	39
CAPÍTULO IV	40
RESULTADOS Y DISCUSIÓN.....	40
Resultados.....	40
Discusión	46
CAPÍTULO V	49
CONCLUSIONES Y RECOMENDACIONES	49
5.1. Conclusiones	49
5.2. Recomendaciones	50
REFERENCIAS BIBLIOGRÁFICAS.....	51
ANEXOS.....	55
Anexo 1. Cuadro de operacionalización de variables.....	55
Anexo 2. Matriz de consistencia.....	556
Anexo 3. Datos analizados en logaritmo natural	557
Anexo 4. Modelo de datos agrupados.....	558
Anexo 5. Modelo de efectos aleatorios	58
Anexo 6. Prueba del Multiplicador de Lagrange para Efectos Aleatorios	59
Anexo 7. Modelo de efectos fijos	59
Anexo 8. Test de Hausman.....	60
Anexo 9. Test de Wooldridge.....	60
Anexo 10. Prueba de Breusch y Pagan	60
Anexo 11. Prueba modificada de Wald para Heterocedasticidad	61
Anexo 12. Modelo de efectos aleatorios con errores estándar corregidos	61

Lista de tablas

Tabla 1. Incidencia de la incidencia de la pobreza total, tasa de analfabetismo de la población femenina de 15 a más años de edad y desnutrición infantil en el norte peruano en el periodo 2010-2018.....	41
Tabla 2. Modelo de datos panel que pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018.....	42
Tabla 3. Prueba del Multiplicador de Lagrange para Efectos Aleatorios.....	43
Tabla 4. Prueba de Hausman para determinar qué modelo pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018.....	43
Tabla 5. Test de Wooldridge para analizar autocorrelación.....	44
Tabla 6. Prueba Modificada de Wald para analizar Heterocedasticidad.....	44
Tabla 7. Prueba de Breusch y Pagan para analizar correlación contemporánea.....	45
Tabla 8. Modelo de efectos aleatorios corregido a través de Errores Estándar Corregidos para Panel – PCSE, por problemas de correlación contemporánea.....	45

RESUMEN

El objetivo de esta investigación fue evaluar el modelo de datos panel que pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018, siendo esta de tipo longitudinal y espacial debido a que se obtuvieron datos de la misma población en distintos años y regiones con la finalidad de examinar sus variaciones en el tiempo; la muestra estuvo compuesta por las observaciones anuales de la desnutrición crónica infantil, incidencia de la pobreza total y tasa de analfabetismo para la población femenina de 15 años a más durante los años 2010 – 2018 en las regiones de Lambayeque, Cajamarca, Piura, Tumbes y La Libertad. Las bases de datos fueron extraídas de la web de estadísticas sociales del Instituto Nacional de Estadística e Informática – INEI y analizadas en el programa Stata 14; con el fin de reducir los problemas de heterocedasticidad se transformaron las variables a logaritmo natural, logrando evaluar los modelos de efectos aleatorios, de efectos fijos y de datos agrupados, en donde para seleccionar el mejor modelo se hizo uso de la prueba del Multiplicador de Lagrange para Efectos Aleatorios, así como la Hausman; encontrando que el modelo de datos panel que pronostica mejor la desnutrición infantil es el de efectos aleatorios de sección cruzada tanto para región como para tiempo, con la corrección de Errores Estándar Corregidos para Panel – PCSE, por problemas de correlación contemporánea.

Palabras claves: Desnutrición infantil, Datos panel, Correlación contemporánea, Heterocedasticidad, Sección cruzada.

ABSTRACT

The objective of this research was to evaluate the panel data model that better predicts child malnutrition in five regions of Northern Peru in the 2010-2018 period, being longitudinal and spatial because data were obtained from the same population in different years and regions, in order to examine their variations over time; The sample was composed of the annual observations of chronic child malnutrition, incidence of total poverty and illiteracy rate for the female population aged 15 and over during the years 2010 - 2018 in the regions of Lambayeque, Cajamarca, Piura, Tumbes and Freedom. The databases were extracted from the social statistics website of the National Institute of Statistics and Informatics - INEI and analyzed in the Stata 14 program; In order to reduce the problems of heteroscedasticity, the variables were transformed to a natural logarithm, managing to evaluate the random effects, fixed effects and grouped data models, where to select the best model, the Lagrange multiplier test was used. for Random Effects, as well as the Hausman; finding that the panel data model that best predicts child malnutrition is the one with random effects of cross section for both region and time, with the Corrected Standard Errors correction for Panel - PCSE, due to contemporary correlation problems.

Keywords: Child malnutrition, Panel data, Random effects, Contemporary correlation, Heteroskedasticity, Cross section.

INTRODUCCIÓN

Se puede afirmar que la desnutrición infantil es considerada como una problemática que se incrementa de forma silenciosa, puesto que está comprobado que su padecimiento trae consigo efectos dañinos hacia las personas, y, que esto incluso se manifiesta durante todo el desarrollo de la persona, la bibliografía nos señala que una de las primeras señales es el peso por debajo de lo normal, al igual que se da con la estatura; sin embargo, son sólo algunas de las evidencias más resaltantes de este problema; ya que la mortalidad infantil se origina, directa o indirectamente, por un pobre estado nutricional (Banco Mundial, 2016).

Según el Instituto Nacional de Estadística e Informática – INEI (2017), en el Perú en los últimos cinco años se afirma que la desnutrición crónica en menores de 5 años ha decaído un 5,2%, asimismo, se manifestó que el mayor índice de desnutrición se alcanzó en la población infantil menor de tres años de edad (13,6%), así como en las niñas y niños con madres sin educación o con estudios de primaria (27,6%). Revelando que las cifras más altas para la desnutrición crónica en infantes de cinco años de edad se reportaron en Apurímac (20,9%), Huancavelica (31,2%), Loreto (23,8%), Pasco (22,8%), Cajamarca (26,6%) y Ayacucho (20%).

Hoy en día existen investigaciones que buscan dar solución a este tipo de problemáticas haciendo uso de las ciencias exactas, tal es el caso de la investigación que hace uso de los modelos lineales generalizados para pronosticar la anemia infantil a través de factores asociados, cuyo autor manifiesta que esta técnica puede ser empleada para presagiar convenientemente tanto la severidad de la anemia infantil como la prevalencia (Minchón, 2015).

En base a la problemática antes descrita, esta investigación abordó modelos econométricos que permitieron pronosticar el comportamiento de la desnutrición infantil en cinco regiones del Perú, desarrollando el modelo de datos panel ya que se tiene una serie temporal de los años 2010 – 2018 para las regiones de Lambayeque, Cajamarca, Tumbes, La Libertad y Piura.

La generación de un modelo adecuado para el pronóstico de la variable desnutrición infantil logrará que los dirigentes de estado puedan tomar decisiones a gran escala,

que permitirán construir alianzas con organizaciones de los sectores públicos y privados, quienes aportarán a la búsqueda de soluciones logrando la mejora en la sociedad.

Es por ello que se planteó el problema: ¿Qué modelo de datos panel pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018?, en base a esta problemática se formuló por hipótesis: El modelo de datos panel de efectos fijos pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018.

Presentando como objetivo general: Evaluar el modelo de datos panel que pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018, así mismo la población estuvo conformada por las observaciones anuales de la desnutrición infantil, incidencia de la pobreza total y tasa de analfabetismo para la población femenina con una edad de 15 a más en el Perú, mientras que la muestra estuvo compuesta por estas observaciones en las regiones de Lambayeque, Cajamarca, Piura, Tumbes y La Libertad.

Esta investigación hizo uso de datos secundarios, la técnica que se desarrolló fue el análisis de base de datos oficiales, donde la tasa de desnutrición infantil, incidencia de la pobreza total y tasa de analfabetismo de la población femenina de 15 años se obtuvieron de la web de estadísticas sociales del Instituto Nacional de Estadística e Informática – INEI.

En resumen, con ayuda de distintas técnicas econométricas cuyo fin es conectar y confrontar la teoría con la realidad, que actualmente no sólo se vienen aplicando al área de las finanzas, sino que ya existen estudios de este tipo de técnicas enfocadas a solucionar problemas en el ámbito de la salud, es que se pretendió predecir mediante un modelo de datos panel, la desnutrición crónica en infantes de una edad de 5 años para las regiones de Lambayeque, Piura, Tumbes, La Libertad y Cajamarca considerando como variables predictoras la Incidencia de la pobreza total y el analfabetismo en la población femenina con una edad de 15 a más años de edad, que según la literatura son factores que intervienen en la desnutrición crónica.

Este informe de investigación se encuentra estructurado en cinco capítulos:

En el Capítulo I se presenta el problema investigativo, iniciando por la formulación y sustento de la problemática, seguido de los antecedentes, planteamiento del problema, demarcación del estudio, importancia y justificación, así como los objetivos generales y específicos.

En el Capítulo II se muestran los elementos presentes en la teoría de la investigación, describiendo la teoría relacionada a los datos de panel, así mismo se contempla el Marco conceptual, en donde se muestran definiciones escogidas necesarias para exponer la hipótesis plasmada.

En el Capítulo III se describen las variables, así como sus indicadores, además del componente temporal y espacial, de igual forma se muestra como punto central a la hipótesis de este estudio, así mismo se presenta la muestra, población, el diseño y el método de la investigación, como también las acciones del proceso investigativo, las técnicas desarrolladas, el proceso que se realizó para obtener los datos y las diversas técnicas para el procesamiento y análisis de los mismos.

En el Capítulo IV se detallan los resultados y la respectiva discusión de los mismos comparados con los autores presentes en los antecedentes de estudio.

Por último, en el Capítulo V se presentan las conclusiones y recomendaciones de esta investigación.

Al finalizar se publican las referencias bibliográficas y los anexos.

CAPÍTULO I

PROBLEMA DE INVESTIGACIÓN

1.1. Planteamiento y fundamentación del problema de investigación

Según la Organización mundial de la salud - OMS (2018), hacía el 2017 en una de sus conferencias se manifestó que día a día el índice de personas que vienen padeciendo hambre en el mundo se presenta de forma creciente, logrando sobrepasar los 820 millones de personas en ese último año, por lo que se puede inferir que esta problemática está convirtiendo al mundo actual en una población con un gran índice de desnutrición.

Se puede afirmar que la desnutrición infantil es considerada como una problemática que se incrementa de forma silenciosa, puesto que está comprobado que su padecimiento trae consigo efectos dañinos hacia las personas, y, que esto incluso se manifiesta durante todo el desarrollo de la persona, la bibliografía nos señala que una de las primeras señales es el peso por debajo de lo normal, al igual que se da con la estatura; sin embargo, son sólo algunas de las evidencias más resaltantes de este problema; ya que la mortalidad infantil se origina, directa o indirectamente, por un pobre estado nutricional (Banco Mundial, 2016).

Es una cifra alarmante que en el mundo exista más de 8 millones de niños cuya edad es menor a cinco años que actualmente en países de América Latina se considera como desnutrición crónica. Si a tal problemática se le incrementa otros factores como la falta de atención médica, falta de educación tanto como para el progenitor y el hijo, así como también los problemas de violencia familiar, se estaría dañando aún más el desarrollo físico y mental de los niños en el mundo, en donde los daños pueden ser irreversibles. Es importante señalar que, si a un menor no se le explota las capacidades cognitivas, físicas, emocionales y sociales durante los cinco primeros años de vida, no podrá realizarlo nunca más en su etapa de adultez (Banco Mundial, 2016).

En Estados Unidos se manifiesta la malnutrición extrema llegando alrededor de los 16 millones de niños que son obligados a requerir cupones para alimentos gratuitos,

sumando a esto que más de la mitad de estudiantes de escuelas públicas y casi un 50% de menores afroamericanos cuyas edades oscilan entre los cinco y seis años que requieren de alimentación gratuita ya que carecen de medios económicos. Así mismo en el 2007, 12 de cada 100 niños hizo uso de esta clase de cupones. Hoy en día 20 de cada 100 niños de los Estados Unidos pertenece a la tasa de pobreza infantil significativamente más altas de entre los países el primer mundo, trayendo como consecuencia una elevada tasa de desnutrición (Rinaldi, 2015).

En Colombia, son 560.000 casos de hambre extrema en la primera infancia (entre los 0 y 5 años de edad) en todo el territorio nacional, en donde 1 de cada 9 niños tiene desnutrición crónica, cifra que es alarmante y según los estudios debido a factores políticos y educativos de los padres, en donde como parte de la solución para esta problemática está el apoyo de países desarrollados como Canadá (Resumen Latinoamericano, 2019).

En nuestro Perú hablar de la desnutrición es un problema muy grave, ya que se tiene una cifra muy en el nivel crónico. Según estándares Internacionales, cerca del 30% de menores peruanos cuyas edades fluctúan hasta los cinco años se encuentran padeciendo este problema. Desde hace ya muchos años se vienen llevando a cabo políticas públicas cuyo fin es disminuir esta cifra alarmante, ya que se reconoce que ello involucra altos costos económicos y sociales logrando cumplir el objetivo, pero de forma mínima; sin embargo, estos esfuerzos aún no son suficientes. Considerando que ya vienen pasando más de veinte años de programas y políticas que tienen por misión erradicar la desnutrición, el padecimiento de esta sigue siendo alta, adicionando a ello la gran diferencia que aún existe entre las personas de las distintas regiones y de distintos estatus sociales.

Según el INEI (2017), en el Perú en los últimos cinco años se afirma que la desnutrición crónica en menores de 5 años ha decaído un 5,2%, asimismo, se manifestó que el mayor índice de desnutrición se alcanzó en la población infantil menor de tres años de edad (13,6%), así como en las niñas y niños con madres sin educación o con estudios de primaria (27,6%). Revelando que las cifras más altas para la desnutrición crónica en infantes de cinco años de edad se reportaron en

Apurímac (20,9%), Huancavelica (31,2%), Loreto (23,8%), Pasco (22,8%), Cajamarca (26,6%) y Ayacucho (20%).

Hoy en día existen investigaciones que buscan dar solución a este tipo de problemáticas haciendo uso de las ciencias exactas, tal es el caso de la investigación que hace uso de los modelos lineales generalizados para pronosticar la anemia infantil a través de factores asociados, cuyo autor manifiesta que esta técnica puede ser empleada para presagiar convenientemente tanto la severidad de la anemia infantil como la prevalencia (Minchón, 2015).

En base a la problemática antes descrita, esta investigación abordó modelos econométricos que permitieron estimar el comportamiento de la desnutrición infantil en cinco regiones del Perú, desarrollando el modelo de datos panel ya que se tiene una serie temporal de los años 2010 – 2018 para las regiones de Lambayeque, Cajamarca, Tumbes, La Libertad y Piura.

Cabe resaltar que la generación de un modelo adecuado para el pronóstico de la variable desnutrición infantil permitirá tomar medidas que permitan mejorar la situación actual del país.

1.2. Antecedentes de la investigación

En el estudio de Pizan (2018) en donde desarrolla un modelo de datos panel para la inversión pública regional y cómo esta influye en la pobreza del Perú, se utilizó como variable dependiente la tasa de pobreza, mientras que como variables independientes se consideraron la inversión pública regional en educación y en salud, así como la inversión pública regional en saneamiento, esta investigación estuvo formada por una dimensión temporal de cinco años, siendo extraídos los datos del Ministerio de Economía y Finanzas (MEF) así como del Instituto Nacional de Estadística e Informática (INEI), los mismos que fueron transformados en logaritmos naturales con el fin de no presentar heterocedasticidad. Se inició el modelamiento mediante el modelo de datos agrupados, estimando además el modelo de efectos fijos y aleatorios; con el fin de encontrar el mejor modelo se utilizaron la prueba del

multiplicador de Lagrange para Efectos Aleatorios y Hausman. En el análisis se encontró problemas de autocorrelación, el misma que fue solucionado con un modelo autorregresivo de orden 1. A pesar de la transformación de la serie en logaritmo natural, se evidenció problemas de heterocedasticidad, además de correlación contemporánea, para dar solución a estos problemas se desarrollaron Feasible Generalized Least Squares – FGLS y Panel Corrected Standard Errors – PCSE, en donde esta última generó parámetros significativos.

Según Romo (2016) en su investigación que propone un modelo de datos panel para las variables macroeconómicas en los procedimientos concursales de las empresas españolas, en donde propone como meta principal el análisis de la efectividad de variables macroeconómicas para con la diversificación del número de suspensiones de pago en una muestra de diez ciudades pertenecientes a España, en la que se plantea un modelo de riesgo basada en la ubicación geográfica. Para el estudio se hizo uso de la técnica de datos panel, que permitió el análisis de las 10 ciudades de España. Para desarrollar data panel en primer lugar se establecieron grupos para las diez ciudades antes mencionadas, considerando un contexto temporal comprendido entre los años 2004 al 2015, periodo extenso, en la que se halló que las variables estudiadas permiten lograr una explicación del 53,8% en relación a la suspensión de pago, dicha explicación demostrada por un R^2 producto de un modelo de datos panel de efectos fijos, considerando el uso de desviaciones típicas robustas. Cabe señalar además que se encontró un R^2 – intra al que se considera que está a nivel del coeficiente de determinación según individuos, a lo que en este estudio se denominó ciudades, evidenciando un 53,6%, lo que conlleva a confirmar que existe correspondencia entre cada una de las ciudades y la muestra en general.

En la investigación de Rodríguez *et al.* (2018) que desarrolla un modelo de datos panel con las variables gasto público sanitario y educativo para el bienestar de los países de la Organización para la Cooperación y el Desarrollo Económicos – OCDE, los autores de esta investigación señalaron que el objeto principal fue analizar la efectividad que se presenta en el sistema de gasto público en el sector educativo y el gasto nacional sanitario en el estado de bienestar de la población, que será medido a través del indicador IDH en países que según las estadísticas presentan un mayor ingreso, para lograr la evaluación de las políticas de disminución en su presupuesto

de educación y sanidad en el desarrollo de bienestar. Para este análisis se hizo uso de la información de 34 países que forman parte de la OCDE, el mismo que abarcó los años de 1980 al 2013, sin embargo, para el análisis se redujeron los años de 1980 al 2011, ello debido a que en los últimos años se presentó ausencia de datos. La investigación fue desarrollada bajo la técnica de datos de panel, la cual empleó estimadores necesarios para que puedan cumplir con los requisitos de un modelo econométrico de data panel, al realizar la estimación correspondiente los autores narraron que se presentó problemas de heteroscedasticidad y autocorrelación, así como la correlación contemporánea, en donde para dar solución a estas problemáticas y a favor de robustez del modelo, se acogieron a una orientación enfocada a distintos estimadores.

Martínez *et al.* (2019) en su investigación de datos panel para el capital social, aquí los autores consideraron analizar para los países de la Organización para la Cooperación y el Desarrollo Económicos - OCDE de qué manera el capital social sería considerado un factor del emprendimiento, para ellos se utilizó los datos de la GEM (Monitor Global Entrepreneurship) además de la WVS (World Values Survey) ambas bases de datos abiertos, para estudiar las variables relacionadas al emprendimiento los datos fueron extraídos de la GEM, mientras que para el estudio de la pertenencia a un grupo social, confianza y capital social, se hizo uso de la segunda base de datos – WVS. Los autores estimaron un modelo de datos considerando efectos aleatorios, la misma que fue determinada mediante la prueba de Hausman. Los resultados llevaron a que se concluya que las variables del capital social intervienen positivamente en el desarrollo del emprendimiento, logrando incrementar su posibilidad de éxito, no sólo en el inicio sino a nivel longitudinal (largo plazo).

En el estudio de Carbajal *et al.* (2018) que proponen un modelo de datos panel hacia la dinámica productiva del sector automotriz, se analizó la correspondencia entre el sector automotriz en los sujetos federales de la frontera norte de México (Baja California, Sonora, Chihuahua, Coahuila, Nuevo León y Tamaulipas) y la dinámica de crecimiento de la producción de la manufactura. Se contribuyó con certeza positiva en relación a cómo la dinámica de desarrollo de la industria automotriz ha promovido el progreso de la industria en la región. Se realizó una estimación con un modelo con datos de panel para la fabricación de la industria en las entidades de la frontera norte,

en donde para cada una de las propuestas de las ecuaciones se desarrolló mediante los modelos de efectos aleatorios y de efectos fijos, aplicando las pruebas de Breusch y Pagan para lograr seleccionar entre la consistencia de los modelos de datos agrupados y de efectos aleatorios, así mismo para determinar si el modelo de efectos fijos o aleatorios era el más adecuados se utilizó la prueba de Hausman , la misma que reveló una mejor estimación con el de efectos fijos. Los autores señalan que se presentaron problemas de heteroscedasticidad y de autocorrelación, así mismo indicaron que para dar solución a estos problemas utilizaron los PCSE o Panel Corrected Standard Errors - Errores Estándar Corregidos para Panel ya que son más exactos que los FGLS o Feasible Generalized Least Squares - Mínimos Cuadrados Generalizados Factibles. Como resultados resaltantes se encontró que para el sector automotriz interviene en el crecimiento manufacturero en el norte del país, revelando grandes beneficios, en variables como el empleo, producción, calificación de mano de obra, entre otros.

1.3. Formulación del problema de investigación

¿Qué modelo de datos panel pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018?

1.4. Delimitación del estudio

Delimitación espacial: Para el presente estudio se contó con los datos que corresponden a la medición de variables demográficas en salud como: la tasa de desnutrición crónica de niños (as) menores de 5 años, tasa de analfabetismo de la población femenina de 15 y más años de edad y la incidencia de la pobreza total para las regiones de Tumbes, Piura, Lambayeque, Cajamarca y La Libertad.

Delimitación temporal: Los datos se encuentran comprendidos entre los años 2010 al 2018.

1.5. Justificación e importancia de la investigación

Se conoce que uno de los grandes problemas mundiales es la desnutrición infantil, en donde las secuelas se revelan a lo largo de todo el periodo vital ya que tiene efectos negativos inmediatos como el desempeño escolar, la capacidad de trabajo e incluso muertes prematuras, y, que afecta principalmente a países del tercer mundo por su baja economía; es por ello que, siendo un problema de gran magnitud, debe ser afrontado en mayor participación por el estado y por ende toda la comunidad que alberga esta complicación.

La econometría utiliza métodos estadísticos para estudiar y cuantificar mediante datos reales los fenómenos económicos, hoy en día existen distintas técnicas econométricas que tienen por fin conectar y confrontar la teoría con la realidad, que actualmente no sólo se vienen aplicando al área de las finanzas, sino que ya existen estudios de este tipo de técnicas enfocadas a solucionar problemas en el ámbito de la salud, es por ello que frente a esta realidad esta investigación pretende predecir mediante un modelo de datos panel, la tasa de desnutrición crónica en menores de 5 años para las regiones de Lambayeque, Piura, Tumbes, La Libertad y Cajamarca considerando como variables predictoras la Incidencia de la pobreza total y la tasa de analfabetismo de la población femenina de 15 y más años de edad, que según la literatura son factores que intervienen sobre la tasa de desnutrición crónica.

Por lo antes mencionado cabe la necesidad de utilizar este tipo de técnicas estadísticas que ayuden a predecir mediante un modelo teórico el comportamiento de las variables antes mencionadas con el fin de tomar una decisión en base a un sustento teórico-científico, en donde además se obtendrán cifras exactas que permitan establecer políticas de salud en beneficio de la población en riesgo. Se resalta además que los efectos de esta investigación son de suma importancia porque permitirá a los dirigentes de estado tomar decisiones a gran escala, que permitirán construir alianzas con organizaciones de los sectores públicos y privados, quienes aportarán a la búsqueda de soluciones logrando la mejora en la sociedad.

1.6. Objetivos de la investigación:

Objetivo General

Evaluar el modelo de datos panel que pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018

Objetivos Específicos

- Analizar la evolución de la incidencia de la pobreza total y la tasa de analfabetismo de la población femenina de 15 y más años de edad en el norte peruano en el periodo 2010-2018
- Analizar la evolución de la desnutrición infantil en el norte peruano en el periodo 2010-2018
- Estimar el modelo de datos panel que pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018.
- Determinar el modelo de datos panel que pronostique mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018

CAPÍTULO II

MARCO TEÓRICO

2.1. Fundamentos teóricos de la investigación

Modelos Econométricos aplicados en el área de salud

En econometría las construcciones intelectuales que tienen como base la teoría con el fin de realizar estimaciones en función a efectos derivados de las permutas para los datos reales, son los llamados modelos. Se dice que un modelo econométrico conecta el mundo real con la teoría, en donde se puede contrastar si esta es adecuada para representar situaciones que tienen una explicación.

Los modelos econométricos se basan en procurados propios de una realidad y estructuralmente están desarrollados por ecuaciones matemáticas que dependen de diferentes variables. Por teoría estas variables se clasifican en variables exógenas y endógenas; una variable exógena es aquella en donde sus valores son extraídos de la realidad, mientras que la variable endógena, posee un valor que es derivado de la operación de la ecuación de un modelo. Estas clases de variables se relacionan a través de una serie de parámetros, los mismos que deben ser estimados (Valero, 1991).

Una de las bondades de los modelos es que acceden a cumplir pronósticos que hoy en día son aplicables a cualquier área del conocimiento, estando dispuestos a ser contrapuestas con la situación. Estos pronósticos se consideran probabilísticas mas no deterministas; esto quiere señalar, que mediante los modelos econométricos no se puede pronosticar con exactitud cuál sería, por ejemplo, el uso exacto que ejecutará cierto individuo, pero lo que sí es posible predecir es el procedimiento de grandes adheridos de consumidores considerando márgenes de error entre los que estará incluido, o lo que también podría decirse, estimar la probabilidad de cumplir con la predicción.

Se pueden clasificar los tipos de modelos econométricos según su objeto de análisis:

- a) Modelos Macroeconómicos, se encuentran en este rubro los modelos enfocados totalmente en la Economía, son aquellos modelos en que comúnmente se hace muy poco uso de detalles sectoriales que cuantifican los de políticas macroeconómicas, un ejemplo en este caso es el incremento del gasto público o el tributo directo. Estos modelos en su mayoría son utilizados por los Economistas para realizar sus pronósticos.
- b) Modelos Microeconómicos, a diferencia de la Macroeconomía en los modelos de este tipo se pueden estudiar otras áreas del conocimiento humano, tales como los casos de alguna institución, mercado, industrial, entre otros.

De acuerdo a su relación que existe entre las variables, se pueden diferenciar:

- i. Modelos Deterministas, aquí son exactas las crónicas entre las variables del modelo, partiendo de una o varias variables, a las cuales se les llamará inputs, que se encargan de estudiar el comportamiento de otras variables, llamadas output, que son producto de distintas evoluciones matemáticas. Un ejemplo muy claro de esta clase de modelos son las tablas Leontief de input-output.
- ii. Modelos Estocásticos, en este caso la relación entre variables no es exacta, debido a que existe una representación aleatoria como componente, al cual se le denomina término de error o perturbación aleatoria, que está incluida en la formulación del modelo. Este componente acopia los detalles que no están detallados en la relación de causa dada por el modelo, así como ciertos contextos acontecidos impredecibles de carácter (shocks) que entran los errores de la medición y la relación estudiada, así como la documentación y computación de las variables observables que se aproximan las variables teóricas (Valero, 1991).

Considerando el tipo de datos de las variables, se pueden encontrar modelos de:

- 1) Series Temporales, en donde podemos utilizar información (datos) que fueron acopiados longitudinalmente a través de un periodo de tiempo: pudiendo ser estos, años, días, trimestres, meses o, semanas. Podemos citar algunos casos en esta clasificación a las variables son el Índice de Precios al Consumo, las cotizaciones diarias de acciones, Producto Interior Bruto trimestral, y todos aquellos en que los datos son anuales y trimestrales, etc.
- 2) Series de Corte Transversal, se refiere hacer uso de disímiles sujetos en un equivalente momento del tiempo. Se puede considerar como ejemplos de este tipo a los productos que consumieron distintos linajes en cierto tiempo, las comercializaciones que ejecutan ciertas empresas en un trimestre, etc.
- 3) Datos de Panel, aquí armonizamos información de varios individuos que fueron acopiados durante un periodo largo de tiempo.

Evaluando si existen o no retardos de las variables que están presentes en el modelo, estas se pueden diversificar en:

- A. Modelos Estáticos, hace hincapié a la relación entre variables del modelo que tienen por característica un mismo instante de tiempo, siendo esta para todas las variables explicativas del modelo, así como para la variable endógena.
- B. Modelo Dinámico, se colocan en este tipo las variables que tienen un modelo en donde se refiere a diferentes momentos en el tiempo, de manera que al elaborar variables retardadas estaríamos hablando de un modelo dinámico.

Por último, teniendo en cuenta las variables endógenas y su número que deseemos manifestar, encontramos a los modelos:

- i. Uniecuacionales, una única variable endógena es característica de este tipo.

- ii. Multiecuacionales, al contrario del caso anterior aquí presentamos varias variables endógenas, pudiendo ser cualquiera de ellas a su vez variables explicativas en otros casos.

Modelo de datos panel

Cuando se trata de datos tipo panel, se presentan observaciones de series de tiempo en base a una muestra que contempla unidades individuales. En cierta variable Y_{it} , se presume que posee $i = 1, \dots, N$ análisis de tipo transversal y $t = 1, \dots, T$ observaciones en función al tiempo, donde i y t hacen referencia a los al periodo de tiempo y a los individuos, respectivamente.

En base a la cantidad de observaciones que se tiene por individuo, se considera una clasificatoria de dos tipos de panel, siendo estos el panel balanceado (donde existe un igual número de análisis para cada uno de los sujetos) y panel no balanceado (en este caso se compone de algún sujeto con un número distinto de observaciones) (De La Rosa, 2016).

Así mismo según las observaciones de acuerdo a cortes temporales y transversales, se puede hacer clasificar los siguientes tipos:

- a) Micropanel: se le llama así, cuando hay un mínimo número de observaciones temporales por cada sujeto, siendo además este número de sujetos muy elevados. Es por ello que para estudiar las características de los estimadores para los parámetros en esta clase de modelos se deben tener en cuenta con esta clase de datos, considerando que $N \rightarrow \infty$ y T es fijo.
- b) Macropanel: esta clasificación está dada cuando existen pocos individuos y un gran número de observaciones temporales. Es por ello que, para estudiar las propiedades con esta clase de datos, se tomará en cuenta que N es fijo y $T \rightarrow \infty$.
- c) Campo aleatorio: para este caso cuando se tiene un número elevado de individuos y observaciones temporales. Siendo necesario para esto que las

características que se esbocen con esta clase de datos, se tomará en cuenta que $N \rightarrow \infty$ y $T \rightarrow \infty$.

En función a su modelo clasificado como datos de panel se considera una ramificación en base al modelo clásico de regresión lineal, el cual se presenta tal como podemos observar:

$$Y_{it} = x'_{it}\beta + \varepsilon_{it} \quad ; i=1 \dots N; t=1 \dots t$$

En donde $x'_{it} = (x_{1it}, x_{2it}, \dots, x_{kit})$ es un vector $k \times 1$ teniendo en cuenta la información de los k regresores por cada sujeto i por cada periodo t , el vector de parámetros de interés de orden $k \times 1$ es β y ε_{it} la perturbación aleatoria (De La Rosa, 2016).

Es una característica muy preponderante de este modelo que las observaciones por cada variable considerando los subíndices: t , para la extensión temporal e i , para la extensión transversal. Para este caso, el estimador por MCO también llamado estimador POOLED (Mínimos de Cuadrados Ordinarios) será un estimador lineal, insesgado y óptimo (ELIO) considerando que para el caso de las perturbaciones cumplen las hipótesis clásicas.

En muchos momentos, está la clasificación de los sujetos que perturban a la variable endógena, que no toman en cuenta los regresores y que persisten firmes a lo largo del tiempo para cada sujeto, una muestra puede ser: la destreza, comprensión o sapiencia para el caso de que i esté alusiva a personas. Por tanto, resulta conveniente unir esta heterogeneidad individual que, habitualmente, es inobservable en el modelo que se presenta:

$$Y_{it} = x'_{it}\beta + n_i + \varepsilon_{it} \quad ; i=1 \dots N; t=1 \dots t$$

Donde n_i es la heterogeneidad inobservable constante e individual de forma longitudinal a través del tiempo por cada uno de los sujetos. Obstaculizando la

estimación del modelo puesto que si n_i está correlacionado con x'_{it} el estimador por MCO, en general, será sesgado.

Realizar un trabajo con estas características de panel es útil vigilar ciertos efectos que interceden en la variable endógena y que no son acopiados por las variables explicativas del modelo. De forma que, los modelos con datos de panel intervienen la heterogeneidad individual inobservable, es decir, la presencia de efectos latentes no observables específicos de cada agente, que habitualmente son firmes en el tiempo que inciden sobre el modo en que éste ocupa sus decisiones. Cuando los efectos ocultos no sean acopiados claramente en el modelo, se causará un problema exclusivamente grave al que se denominará variables omitidas, siendo esto cuando hay una correlación entre los regresores del modelo y los efectos latentes. Los datos de panel nos muestran otra ventaja que consta en proporcionar una mayor cantidad de información, mayor variabilidad, mayor precisión y una menor linealidad entre variables. Para culminar, los datos panel nos brindan una investigación valiosa de los sujetos siguiéndolos a través del tiempo, ofreciendo un punto de vista más completo del problema, demostrando una mejor dinámica de cambio para las unidades de corte transversal.

Se considera que uno de los problemas de trabajar con datos de panel es que los sujetos pueden caer en el abandono de la muestra, impidiendo el seguimiento de estos en un largo de tiempo. Teniendo en cuenta además que, pueden surgir inestabilidades para con el ejemplar, asumiendo mayor información de unos individuos que de otros, generando un panel incompleto o no equilibrado (De La Rosa, 2016).

Regresiones en datos de panel

Según Pérez (2008), los datos de panel se expresan mediante las siguientes regresiones:

Regresión agrupada (pooled ols)

Cuando se omiten dimensiones de tiempo para los datos agrupados y espacio, en la que sólo se calcula los Mínimos Cuadrados Ordinarios -MCO se considera la dirección más sencilla en data panel. Considerando que los coeficientes son iguales para cada uno de los elementos sociales relacionados a la muestra, tal como se muestra a continuación:

$$y_{it} = \beta_0 + \sum_{k=1}^k \beta_k X_{kit} + u_{it}$$

Teniendo en cuenta que $k= 1, \dots, K$ denominado índice de las variables exógenas, además de $i= 1, \dots, N$ elementos sociales y $t= 1, \dots, T$ utilizado para representar el tiempo, baso su representación matricial tendríamos:

$$Y_{it} = \beta' X_{it} + u_{it}$$

Bajo esta notación se estimarán los parámetros K , en las que estos parámetros son constantes para las unidades muestrales en su totalidad además de los periodos de tiempo. Para realizar su estimación esta se llevará a cabo mediante los MCO teniendo en cuenta el supuesto de que la varianza de los errores será la misma por cada observación, a lo que denominamos el supuesto de homocedasticidad, incorporando además la no correlación de los diferentes momentos de tiempo y los términos de error.

Hay varios contextos en la que los supuestos de estimación de los Mínimos Cuadrados Ordinarios pueden ser vulnerables, ya sea por la heterocedasticidad mediante el cruce de secciones, en donde la variabilidad de los elementos del error puede ser desigual por cada cruce de sección, mientras que la otra puede darse por la correlación de este cruce de secciones, a lo que llamamos correlación de los elementos del error mediante distintos elementos sociales en uno o más tiempos t , pero para un igual elemento social, los elementos del error quedan correlacionados.

Modelo de efectos fijos

La heterogeneidad transversal inobservable es una suposición de este modelo, correlacionada con los regresores y constante en el tiempo. Siendo este caso en donde n_i es considerado como un parámetro adicional que expresa independencia para cada uno de los sujetos del panel (De La Rosa, 2016).

Ante ello es que nombramos comúnmente el modelo de MCVF (Mínimos Cuadrados de Variables Ficticias), en donde podemos resaltar que la frase “Mínimos Cuadrados” nos hace hincapié a la técnica que se usa periódicamente para su estimación, no al modelo propiamente dicho. Presentamos a continuación un estimador consistente de β que así mismo podemos obtenerlo a partir de las evoluciones del modelo que elimina los efectos individuales partiendo de su forma matricial presentada anteriormente:

$$Y = Cn + X\beta + e$$

Para pronosticar β , se tiene en cuenta los parámetros de utilidad por MCO, usando los resultados de la regresión particionada. Así, el estimador de β por MCO se puede presentar de la siguiente forma:

$$\hat{\beta} = (X'\bar{M}X)^{-1}X'\bar{M}Y$$

$$\bar{M} = \begin{bmatrix} M^0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & M^0 \end{bmatrix}$$

Donde $\bar{M} = I_{NT} - C(C'C)^{-1}C' = I_N * M^0$ con $M^0 = I_T - \frac{T'T}{T}$. \bar{M} es una matriz

idempotente y simétrica $\bar{M}x = 0$

$\hat{\beta}$ es llamado estimador intragrupos, en donde la estabilidad de este pronosticador intra grupos, cuando T es fijo y $N \rightarrow \infty$, no es dependiente de la descripción de n_i debido a que los efectos siempre son suprimidos a causa de la transformación del modelo. La exigencia que avala la consistencia de este estimador es que las variables x'_{it} sean estrictamente exógenas, con respecto a ε_{it} , es decir,

$E(\varepsilon_{it}/x'_{it}) = 0, \forall t = 1, \dots, T$. Por otro lado, esta forma posee una limitante, considerando las variables X'_{it} no son variantes en el tiempo al hallar las desviaciones respecto a las medias temporales, estas desaparecerán del modelo, por lo que no podrá ser posible pronosticar los parámetros correspondientes (De La Rosa, 2016).

Modelo de efectos aleatorios

Para esta clase de modelos, se asume que n_i acumula los efectos inobservables de cada agente que se suponen constantes en el tiempo y que constan incorrelaciones con los regresores (De La Rosa, 2016).

$$Y_{it} = x'_{it}\beta + n_i + e_{it} \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T.$$

Es particular de esta clase de modelo que los efectos aleatorios de n_i llegando a contribuir parte del término error del modelo:

$$Y_{it} = x'_{it}\beta + \omega_{it} \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T.$$

Donde $\omega_{it} = n_i + \varepsilon_{it}$, trabajando bajo los siguientes supuestos:

$$E(\varepsilon_{it}) = E(n_i) = E(n_i \varepsilon_{it}) = 0 \quad \forall it$$

$$E(x_{it} \setminus n_i) = E(x_{it} \setminus \varepsilon_{it}) = 0 \quad \forall it$$

$$E(n_i n_j) = \begin{cases} \sigma_n^2 & \text{Si } t = s \quad i = j \\ 0 & \quad \quad \quad i \neq j \end{cases}$$

$$E(\varepsilon_{it} \varepsilon_{js}) = \begin{cases} \sigma_\varepsilon^2 & \text{Si } t = s \quad i = j \\ 0 & \quad \quad \quad i \neq j \end{cases}$$

La varianza del nuevo término error $\omega = n_i + \varepsilon_{it}$ es de la forma $E[\omega_{it}^2] = \sigma_\omega^2 = \sigma_n^2 + \sigma_\varepsilon^2; \forall i, t$. la covarianza para el término error para un mismo sujeto sería $Cov(\omega_{it} \omega_{is}) = \sigma_n^2$; por lo que la matriz de varianzas y covarianzas de los T valores proporcionados al i-ésimo individuo $\omega'_t = [\omega_{i1}, \omega_{i2}, \dots, \omega_{iT}]$ es una matriz (TxT).

$$\Omega = E[\omega_i \omega_i'] = \begin{bmatrix} \sigma_n^2 + \sigma_\varepsilon^2 & \sigma_n^2 & \cdots & \cdots & \sigma_n^2 \\ \sigma_n^2 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \ddots & \ddots & \sigma_n^2 \\ \sigma_n^2 & \cdots & \cdots & \sigma_n^2 & \sigma_n^2 + \sigma_\varepsilon^2 \end{bmatrix} = \sigma_\varepsilon^2 I + \sigma_n^2 i_T i_T'$$

La matriz que se muestra a continuación presenta la correlación entre dos observaciones originarias que parten del mismo sujeto, siendo este constante y que no desaparece con el tiempo, ya que esto se debe a la presencia del efecto individual n_i , σ_n^2 .

Es por ello que, la matriz covarianzas y varianzas de cada una de las perturbaciones del modelo completo $\omega = [\omega_1, \omega_2, \dots, \omega_N]$ es diagonal a bloques y se expresa de la siguiente forma:

$$E[\omega \omega'] = V = I_N \otimes \Omega = \begin{bmatrix} \Omega & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Omega \end{bmatrix}$$

Balestra y Nivole (1966) proponen una manera de pronosticar $\hat{\sigma}_\varepsilon^2$ y $\hat{\sigma}_n^2$ siendo la siguiente:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=1}^T (\tilde{y}_{it} + \hat{\beta}' \tilde{x}_{it})^2$$

$$\hat{\sigma}_n^2 = \frac{1}{N} \sum_{i=1}^N (\bar{y}_i + \hat{\beta}' \bar{x}_i)^2 + \frac{1}{T} \hat{\sigma}_\varepsilon^2$$

Donde $\tilde{y}_{it} = y_{it} - \bar{y}_i$; $\tilde{x}_{it} = x_{it} - \bar{x}_i$ y $\hat{\beta}$ estimador intragrupos.

En este caso el estimador por MCGF es consistente y asintóticamente ineficiente.

Prueba de Hausman

Esta prueba permite determinar qué modelo es el más apropiado para el panel de datos que se está analizando, si el de efectos fijos o de efectos aleatorios.

Utilizando para ello una prueba Chi-cuadrado con la hipótesis nula de que el modelo de efectos aleatorios es el que mejor explica la relación de la variable

dependiente con las explicativas, y por tanto se tiene la hipótesis alternativa de que el mejor método que se ajusta es el de efectos fijos (caso contrario).

Supongamos que tenemos los estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ y sabiendo además que para $\hat{\theta}_2$, siendo el más eficiente (ello por tener menor varianza). Al realizar el cálculo en el test con una formulación especial, que persigue una distribución chi cuadrado, las diferencias en las estimaciones comunes a ambos modelos. Si las diferencias, por más mínimas que sean, no son consideradas sistemáticas, ello por no tener un sesgo definido, en aquel momento ambos estimadores son consistentes ya que su estimación muestral tiende al parámetro poblacional y optaríamos por el más eficiente $\hat{\theta}_2$. En el caso de que si las diferencias son sistemáticas entonces nuestra hipótesis no se estaría cumpliendo, encontrando que los dos no son consistentes y generando un conflicto: considerar que el modelo está mal desarrollado para ambos casos o quedarnos con el estimador consistente, que es $\hat{\theta}_1$ (Montero, 2005).

Autocorrelación

Al no existir independencia entre los errores U_{it} en relación al tiempo, se llega a afirmar que se presenta un inconveniente de autocorrelación o correlación serial. Actualmente se dice que hay varias formas de detectar los inconvenientes de correlación serial, sin embargo, estas formas están sujetas a ciertos supuestos que dependen de la naturaleza de la individualidad de los efectos. La prueba más utilizada enfocada en apócrifos mínimos y que además es muy flexible fue desarrollada por Wooldridge, en la que dentro de su prueba propone por hipótesis nula que no existe autocorrelación, obviamente se esta es rechazada, concluiríamos que si hay evidencia de autocorrelación.

Wooldridge desarrolló esta prueba haciendo uso de una regresión de las primeras diferencias a través de los elementos residuales, analizando que si los U_{it} no están consecutivamente correlacionado; entonces la correspondencia entre los errores U_{it} diferenciados en el periodo $t-1$ y t es equivalente a -0.5 . En consecuencia, la

prueba que propone Wooldridge está enfocada a experimentar esta identidad (Pérez, 2008).

Correlación contemporánea

Podemos presentar inconvenientes de correlación contemporánea en las apreciaciones de data panel, si llegamos a encontrar que la información de dichas unidades está correlacionada con los elementos de esas informaciones en un mismo periodo temporal. Esta problemática de correlación contemporánea se enfoca a la correspondencia de los errores de al menos dos o más unidades en un mismo tiempo t . En resumen, podemos tener errores contemporáneamente correlacionados cuando existen particularidades inobservables de ciertas unidades que se relacionan con las características que no pueden ser observadas de otros elementos. Un ejemplo claro es para este estudio, cuando encontramos correspondencia entre los errores de dos regiones, pero conservándose la independencia en los errores de las demás regiones.

Para analizar inconvenientes de este tipo la prueba de Breusch y Pagan es muy utilizada para identificar problemas de correlación contemporánea en los residuales de un modelo de efectos fijos. La hipótesis nula es que existe “independencia transversal” (cross-sectional independence); es decir, que los errores entre las unidades son independientes entre sí. Si la H_0 se rechaza, entonces existe un problema de correlación contemporánea (Pérez, 2008).

Heterocedasticidad

Gauss-Markov señalan que la variabilidad de los errores para cada elemento transversal es constante, en datos de panel es común que este supuesto se vea afectado, por lo que una de las formas de analizar si se tiene este problema es mediante la prueba del Multiplicador de Lagrange de Breusch y Pagan, no obstante la que se acaba de referenciar junto a algunas más son sensibles en relación a la normalidad de los errores; favorablemente, se tiene una alternativa a este problema, es ahí donde la prueba Modificada de Wald para Heterocedasticidad trabaja incluso cuando no se cumple este supuesto. Para desarrollar esta prueba

se tiene por hipótesis nula que no hay evidencia de heterocedasticidad, considerando que $\sigma_i^2 = \sigma^2$ para toda $i=1\dots N$, en donde N es considerado el número de elementos transversales, al rechazar esta hipótesis nula nos encontramos ante un dilema de heterocedasticidad (Pérez, 2008).

Errores estándar corregidos para panel (PCSE)

Los problemas de heteroscedasticidad, correlación contemporánea y autocorrelación que hemos inspeccionado se pueden solucionar en conjunto con Feasible Generalized Least Squares ó FGLS (estimadores de Mínimos Cuadrados Generalizados Factibles), o bien con Panel Corrected Standard Errors ó PCSE (Errores Estándar Corregidos para Panel). Beck y Katz (What to do and not to do with time-series cross-section data, 1995) demostraron que los errores estándar de PCSE son más precisos que los de FGLS. Desde entonces, muchos trabajos en la disciplina han utilizado PCSE en sus estimaciones para panel.

Beck y Katz (1995) realizaron pruebas en base a los errores, en las que concluyen que los errores estándar corregidos para panel evidencian con exactitud la variabilidad de muestreo de los $\hat{\beta}$ de MCO, considerando además la no presencia de heterocedasticidad para data panel, los Panel Corrected Standard Errors están incluidos dentro de un porcentaje pequeño de los errores estándar de MCO, considerando que cuando la organización de datos panel lleva a errores estándar erróneos, los Panel Corrected Standard Errors son aún más puntuales.

Llegando a considerar que los Panel Corrected Standard Errors son considerados mejores a los errores estándar de Mínimos Cuadrados Ordinarios, logrando de esta manera realizar una adecuada deducción. Conociendo que no aborda la deficiencia viable de Mínimos Cuadrados Ordinarios, hay que resaltar que no producen inconvenientes en los estimadores Mínimos cuadrados generalizados factibles - MCGF. Se considera que por solución práctica el modelamiento de las variables que dirigen a dificultades de error, sin embargo, en el desarrollo de las estimaciones puede resultar complicado de realizar, es por ello que los PCSE (Panel Corrected Standard Errors) brindan una gran alternativa práctica para dar

soluciones a inconvenientes habituales para los data panel (Pérez, 2008).

Modelos presentes en la investigación presentada

El presente estudio plantea los modelos de: Modelo de MCO agrupados, Modelo de efectos fijos y el Modelo de efectos aleatorios. Prediciendo que el modelo de efectos fijos dentro del grupo es el adecuado para esta investigación

Modelo de MCO agrupados

$$Y_{it} = \beta_1 + \beta_2 X_{1,it} + \beta_3 X_{2,it} + \dots + \beta_k X_{k+1,it} + \epsilon_{it} \quad i=1,2,\dots,N. \quad t=1,2,\dots,T.$$

Donde i es el i-ésima región

Donde t es el periodo de tiempo

Modelo de efectos fijos

$$Y_{it} = \beta_2 X_{1,it} + \beta_3 X_{2,it} + \dots + \beta_k X_{k+1,it} + \epsilon_{it} \quad i=1,2,\dots,N. \quad t=1,2,\dots,T.$$

Considerando los valores corregidos por la media de las variables x e y

Donde i es el i-ésima región

Donde t es el periodo de tiempo

Modelo de efectos aleatorios

$$Y_{it} = \beta_{1i} + \beta_2 X_{1,it} + \beta_3 X_{2,it} + \dots + \beta_k X_{k+1,it} + \epsilon_{it} \quad i=1,2,\dots,N. \quad t=1,2,\dots,T.$$

Donde i es el i-ésima región

Donde t es el periodo de tiempo

En donde para todos los casos:

Y = Tasa de desnutrición crónica de niños (as) menores de 5 años (Porcentaje)

X₁= Incidencia de la pobreza total (Porcentaje)

X₂= Tasa de analfabetismo de la población femenina de 15 y más años de edad (Porcentaje)

2.2. Marco conceptual

Desnutrición: La desnutrición es el resultado del consumo insuficiente de alimentos y de la aparición repetida de enfermedades infecciosas. La desnutrición puede ser crónica, aguda y desnutrición según peso para la edad.

Analfabetismo: hace referencia a aquella persona que no sabe leer ni escribir. El término suele tener un uso extendido y se utiliza para nombrar a los individuos que son ignorantes o que carecen de los conocimientos más básicos en alguna disciplina.

Pobreza: condición caracterizada por una privación severa de necesidades humanas básicas, incluyendo alimentos, agua potable, instalaciones sanitarias, salud, vivienda, educación e información.

Modelo de efecto fijo: este modelo toma en cuenta que hay un término invariable desigual por cada individuo y conjetura que los efectos individuales guardan una independencia entre sí. El desarrollo de este modelo toma en que las variables independientes perturban de igual manera a las unidades de corte transversal y que estas se diversifican por tipologías adecuadas de cada una de ellas, moderadas a través del intercepto.

Modelo de efecto aleatorio: considera que los efectos individuales no son independientes entre sí, es decir que considera que los efectos individuales están distribuidos alrededor de un valor dado, sino que estén distribuidos aleatoriamente alrededor de un valor dado.

Sesgo: se llama sesgo de un estimador a la diferencia entre su esperanza matemática y el valor numérico del parámetro que estima. Un estimador cuyo sesgo es nulo se llama insesgado o centrado.

Heteroscedasticidad: surge de la unión de dos palabras: “hetero” (diferente) y “cedasticidad” (dispersión), es decir, varianza diferente.

CAPÍTULO III

MARCO METODOLÓGICO

3.1. Hipótesis central de la investigación

El modelo de datos panel a efectos fijos pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018.

3.2. Variables e indicadores de la investigación

Variable	Definición	Indicador	Escala
Independiente			
Pobreza total	Estado representado por una falta rígida de escaseces humanas básicas, circunscribiendo agua potable, alimentos, salud, instalaciones sanitarias, vivienda, educación e información.	Porcentaje (%)	Razón
Analfabetismo de la población femenina de 15 y más años	Concibe reseña de aquella persona que no sabe escribir ni leer. La expresión enlase asumir un uso amplio y es utilizado para mencionar a los sujetos que son iletrados o que escasean de los conocimientos más elementales en alguna disciplina.	Porcentaje (%)	Razón
Componentes			
Espacial	Considera a una fracción de espacio determinado por algunas particularidades, contextos específicos o habituales como	Región	Nominal

	menciónese el clima, la forma de régimen o la topografía.		
Temporal	Periodo de tiempo que la tierra tarda en dar una vuelta completa alrededor del sol, que a su vez está compuesto por 12 meses.	Año	Nominal
Dependiente			
Desnutrición infantil	La desnutrición puede definirse como la consecuencia del consumo escaso e ineficiente de alimentos y de la muestra consecutiva de malestares patógenos. La desnutrición puede lograr ser crónica y aguda, en base a la edad y el peso.	Porcentaje (%)	Razón

3.3. Métodos de la investigación

La presente investigación se enfoca en el método deductivo, ya que se desarrolla en el procedimiento fundado que va de lo general a lo particular. Tiene la peculiaridad de que las conclusiones de la deducción son verdaderas, si las premisas de las que se originan también lo son. Por lo tanto, toda inclinación deductiva nos lleva de lo general a lo individual. De esta manera, si un fenómeno se ha probado para cierto conjunto de personas, se puede derivar que tal fenómeno se emplea a cada uno de estos sujetos (Gómez, 2012).

3.4. Diseño o esquema de la investigación

Considerando que se analizará la desnutrición infantil en cinco regiones, para 8 años, bajo el modelo de data panel el diseño es de tipo longitudinal y espacial, ya se

obtuvieron datos de la misma población en distintos años y regiones, con la finalidad de examinar sus variaciones en el tiempo (Bernal, 2010).

Considerando como dimensión temporal a los años en los que se realizaron las observaciones, siendo estas desde el año 2010 al 2018; la dimensión espacial se reflejó mediante las mediciones de la desnutrición infantil por región para cada año de estudio, siendo las regiones de Tumbes, Piura, Lambayeque, La Libertad y Cajamarca.

3.5. Población y muestra

La población estuvo conformada por los indicadores anuales de la desnutrición infantil, incidencia de la pobreza total y analfabetismo de la población femenina de 15 y más años de edad en cinco regiones del Perú, correspondientes a periodos anuales.

La muestra estuvo compuesta por los indicadores anuales de la desnutrición infantil, incidencia de la pobreza total y analfabetismo de la población femenina de 15 y más años de edad en las regiones de Lambayeque, Cajamarca, Piura, Tumbes y La Libertad, correspondientes al periodo 2010 – 2018, haciendo una muestra total de 45 observaciones anuales de cada variable.

3.6. Actividades del proceso investigativo

Para el desarrollo de un modelo de datos panel, este se llevó a cabo de cinco etapas, siendo estas las siguientes:

Primera: posterior a la recolección de las observaciones de las variables, se organizó los datos por tiempo, región, variable dependiente e independientes, ello con el fin de analizar si se contaba con datos balanceados.

Segunda: Mediante una tabla de distribución de frecuencias se analizó el comportamiento de la serie temporal: desnutrición infantil, incidencia de la pobreza total y tasa de analfabetismo de la población femenina de 15 años, para un periodo

de 2010 – 2018 en las regiones de Lambayeque, Cajamarca, Piura, Tumbes y La Libertad.

Tercera: Se estimó el modelo de datos panel de datos agrupados y efectos fijos para las variables y tiempos descritos en la etapa anterior.

Cuarta: Se estimó el modelo de datos panel de efectos aleatorios para las variables de esta investigación, en el mismo intervalo de tiempo.

Quinta: Se hizo uso de la prueba de Hausman con el fin de determinar cuál de los modelos de efectos aleatorios o fijos era más apropiado, además se dio solución al problema de autocorrelación contemporánea presente en el modelo.

3.7. Técnicas e instrumentos de la investigación

Considerando que la investigación hace uso de datos secundarios, la técnica que se desarrolló fue el análisis de base de datos oficiales, donde la tasa de desnutrición infantil, incidencia de la pobreza total y tasa de analfabetismo de la población femenina de 15 años se obtuvieron de la web de estadísticas sociales del Instituto Nacional de Estadística e Informática – INEI.

Las bases de datos fueron recolectadas en una hoja de cálculo de Excel, lo que configuraría como instrumento de recolección (extracción) de datos a la hoja de cálculo.

3.8. Procedimiento para la recolección de datos

Los datos de la tasa de desnutrición infantil, incidencia de la pobreza total y tasa de analfabetismo de la población femenina de 15 años a más, fueron extraídas de la web de estadísticas sociales del Instituto Nacional de Estadística e Informática – INEI, descargadas en formato de Microsoft Excel. La extracción de los datos se realizó en los últimos doce meses calendario.

Una vez organizados los datos de las variables en estudio, desde el programa Stata 14 se importó los datos, manteniendo las propiedades de la base de datos original.

3.9. Técnicas de procesamiento y análisis de los datos.

Haciendo uso del programa econométrico Stata 14, se transformaron las variables procedió a transformar las variables a logaritmo natural para reducir los problemas de heterocedasticidad y llevar los datos a una escala lineal, posterior a ello se evaluaron los siguientes modelos:

1. Modelo de datos agrupados
2. Modelo de efectos fijos
3. Modelo de efectos aleatorios

Las técnicas empleadas para seleccionar el mejor modelo consideraron el no cumplimiento de:

- a) Autocorrelación: que contrasta como hipótesis nula que no hay evidencias de autocorrelación
- b) Heterocedasticidad: que señala por hipótesis nula que no hay evidencias de problema de heteroscedasticidad
- c) Correlación contemporánea: que indica por hipótesis nula que no hay evidencias de problema de correlación contemporánea

4. Prueba de Hausman

Esta prueba permitió seleccionar el mejor modelo, ya que evaluó la correlación entre los efectos individuales y las regresoras, con el fin de determinar si el modelo de efectos fijos o aleatorios sería el indicado, considerando la siguiente contrastación de hipótesis:

H_0 : Los estimadores MEF y MCE no difieren considerablemente
(considerar efectos aleatorios)

H_1 : Los estimadores MEF y MCE difieren considerablemente
(considerar efectos fijos)

5. Se solucionó el problema de correlación contemporánea haciendo uso de los Errores Estándar Corregidos para Panel (Panel Corrected Standard Errors ó PCSE).

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

Resultados

La pobreza total y la tasa de analfabetismo de la población femenina de 15 a más años en el periodo 2010-2018 ha venido mostrando una tendencia decreciente en las cinco regiones del norte del Perú; en forma similar, la desnutrición infantil crónica también ha venido disminuyendo (tabla1). Las regiones de Cajamarca (49.1%) y Piura (42.5%) son las que presentaron el 2010 la mayor pobreza; y, al 2018, Cajamarca continuó presentando la mayor pobreza (41.9%), seguidos de Piura (23.1%) y La Libertad (23.1%). En cuanto a la pobreza total, Cajamarca (54.9%) se encuentra en el primer lugar para el año 2011, sin embargo, ha venido presentando un gran decrecimiento al 2018 (41.9%), el mismo escenario se muestra en la región de Piura (42.5%) que para el año 2010 mostró su mayor pobreza, llegando a decrecer al año 2018 en 23.1%, la región con menor tasa de pobreza fue Tumbes, de 20.1% en al 2010 a 12.6% al 2018. En relación a la Tasa de analfabetismo de la población femenina de 15 a más años las regiones con mayor analfabetismo fueron Cajamarca (23.2%) y Piura (13.3%) para el año 2010, regiones que mostraron decrecimiento al año 2018, sin embargo en el año 2013, Cajamarca evidenció un crecimiento al 21.0%, para luego llegar al 17.2% en el año 2018, panorama similar se aconteció en las regiones de Lambayeque (11.9%) y La Libertad (11.7%), en donde su mayor analfabetismo se presentó en el 2010 con una tasa decreciente para los años siguientes, no obstante se manifestó un incremento al 9.4% para Lambayeque en el año 2016 y de 9.3% para La Libertad en el año 2017, la región que muestra una menor tasa de analfabetismo de la población femenina de 15 a más es Tumbes (5.3%) al 2010 ,llegando al 2013 en un 3.0%, para luego incrementar al año 2014 en 4.2%, llegando al 2018 en un 3.3%.

Tabla 1. Incidencia de la pobreza total, tasa de analfabetismo de la población femenina de 15 a más años de edad y desnutrición infantil en el norte peruano en el periodo 2010-2018

Región	2010	2011	2012	2013	2014	2015	2016	2017	2018
Pobreza (%)									
Tumbes	20.1	11.8	14.0	13.1	12.1	11.2	10.8	13.3	12.6
Piura	42.5	40.1	38.9	37.6	29.5	29.9	34.2	24.6	23.1
Lambayeque	35.3	28.5	26.4	21.5	21.2	18.8	16.0	13.3	12.6
La Libertad	32.6	28.5	26.4	30.6	29.5	29.9	22.7	24.6	23.1
Cajamarca	49.1	54.9	53.3	50.5	50.3	48.2	47.4	47.5	41.9
Analfabetismo (%)									
Tumbes	5.3	4.6	4.5	3.0	4.2	3.4	3.5	3.2	3.3
Piura	13.3	10.8	9.7	10.4	10.6	10.7	10.4	10.5	10.4
Lambayeque	11.9	10.2	9.2	9.1	8.6	8.2	9.4	7.9	7.6
La Libertad	11.7	10.8	10.3	8.8	9.1	9.0	7.3	9.3	8.1
Cajamarca	23.2	20.8	18.2	21.0	19.7	19.1	21.0	17.9	17.2
Desnutrición infantil crónica (%)									
Tumbes	12.8	11	8.5	9.1	8.3	9.1	7.4	8.2	8.1
Piura	22.9	20.9	20.8	24.9	21.7	20.3	15.3	15.9	13.1
Lambayeque	17.8	16.2	13.3	14.2	14.3	14.1	11.8	10.5	9.3
La Libertad	26.1	23.2	23.5	22.1	19.9	15.9	12.2	15.6	14.8
Cajamarca	40.5	37.6	36.1	35.6	32.2	23.9	26	26.6	27.4

Fuente: Instituto Nacional de Estadística e Informática – INEI

Con el fin de determinar el modelo de datos panel que pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018 se estimaron tres modelos (Tabla 2), el modelo de datos agrupados que presentó coeficientes de 0.5398203, 0.5398203 y 0.2767282 para la Incidencia de la pobreza, Tasa de analfabetismo y la constante respectivamente, se muestra además el modelo de efectos fijos con coeficiente de 0.461336 para la incidencia de la pobreza, 0.6306148 para la tasa de analfabetismo en mujeres de 15 a más años de edad y de -0.0860082 para la constante, para el caso del modelo de efectos aleatorios, los coeficientes resultaron igual al del modelo de datos agrupados, se sustenta dicho acontecimiento debido a que la varianza de u_i es igual a cero, es decir $\sigma_u^2 = 0$, por lo que no existe diferencia relevante entre los modelos. El modelo de datos agrupados reveló un $R^2 = 0.9020$ al igual que el de efectos aleatorios, mientras que el modelo de efectos fijos evidencia un $R^2 = 0.8915$

Tabla 2. Modelos de datos panel que pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018.

	Modelo de datos panel: Coeficientes de regresión		
	MCO agrupados	Efectos fijos	Efectos aleatorios
Incidencia de la pobreza	0.5398203	0.461336	0.5398203
Tasa de analfabetismo	0.3535067	0.6306148	0.3535067
Constante	0.2767282	-0.0860082	0.2767282
R^2	0.9020	0.8915	0.9020
F		28.46	
Sigma_u	0.06149364	0.16019034	0.06149364
Sigma_e	0.1360752	0.1360752	0.1360752
rho	0.16958831	0.5808613	0.16958831

Considerando que los coeficientes del modelo de datos agrupados y el de efectos aleatorios resultaron iguales (Tabla 2), se contrastó la prueba del Multiplicador de Lagrange para Efectos Aleatorios, en la que se confirma que no existe diferencia significativa entre los modelos ($p > 5\%$) (Tabla 3), ratificando que la varianza de $u_i = 0$

Tabla 3. Prueba del Multiplicador de Lagrange para Efectos Aleatorios

Breusch y Pagan Multiplicador de Lagrange para Efectos Aleatorios		
lnDesnutrición [Región,t] = Xb + u[Región] + e[Región,t] Resultados estimados:		
	Varianza	Desviación estándar
lnDesnutrición	0.214698	0.463355
e	0.018517	0.136075
u	0.003782	0.061494
Test: Var(u) = 0 Prob>chi ² = 0.2108		

Se estimó el modelo de efectos fijos, y, con el fin de corroborar que el modelo de efectos aleatorios o de datos agrupados era el adecuado, se desarrolló la prueba de Hausman (Tabla 4), concluyendo que los estimadores de efectos aleatorios y de efectos fijos no difieren sustancialmente, confirmando que el modelo de efectos aleatorios es el conveniente.

Tabla 4. Prueba de Hausman para determinar qué modelo pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018

Coefficientes				
	(b) fijo	(B) aleatorio	(b-B) Diferencia	(diag(V _b -V _B)) ² Suma de errores
lnPobreza	0.461336	0.53982	-0.07848	0.045508
lnAnalfabetismo	0.630615	0.353507	0.277108	0.166073
Prueba: Ho: la diferencia en los coeficientes no es sistemática Chi ² = 3.89 Prob>chi ² = 0.1429				

Cuando los errores dentro de cada unidad se correlacionan temporalmente, se denomina autocorrelación o correlación serial; para analizar este supuesto en esta investigación, se trabajó con el test de Wooldridge, el mismo que reveló que no existen problemas de autocorrelación ($p > 5\%$) (Tabla 5).

Tabla 5. Test de Wooldridge para analizar autocorrelación

Prueba de Wooldridge para autocorrelación en datos de panel	
Ho: sin autocorrelación de primer orden	
F (1, 4) =	2.266
Prob > F =	0.2067

Para analizar si nuestra estimación presentaba problemas de heterocedasticidad, se hizo uso de la prueba Modificada de Wald para Heterocedasticidad, y, tal como se muestra, la estimación en este estudio no presenta problemas de heterocedasticidad ($p > 5\%$) (Tabla 6).

Tabla 6. Prueba Modificada de Wald para analizar Heterocedasticidad

Prueba de Wald modificada para heterocedasticidad grupal	
En el modelo de regresión de efectos fijos	
Ho: $\sigma^2(i) = \sigma^2$ para todo i	
chi ² (5) =	8.71
Prob > chi ² =	0.1214

Cuando existe correlación de los errores, de al menos dos o más unidades en el mismo tiempo t , se presenta un inconveniente de correlación contemporánea, en este estudio se probó este supuesto mediante la prueba de Breusch y Pagan (Tabla 7), en la que se confirma que la estimación está presentando el problema antes descrito. ($p < 5\%$)

Tabla 7. Prueba de Breusch y Pagan para analizar correlación contemporánea

Prueba de independencia de Breusch-Pagan LM:

$\chi^2(10) = 24.321$, $Pr = 0.0068$

Basado en 9 observaciones completas sobre unidades de panel

Habiendo detectado sólo problemas de correlación contemporánea, este se corrigió a través de Errores Estándar Corregidos para Panel – PCSE (Tabla 8), para así determinar que el modelo de efectos aleatorios en región y tiempo pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018.

Tabla 8. Modelo de efectos aleatorios corregido a través de Errores Estándar Corregidos para Panel – PCSE, por problemas de correlación contemporánea

Modelo de efectos aleatorios		
	Coefficientes	Significancia
Incidencia de la pobreza	0.6163622	0.000
Tasa de analfabetismo	0.2683425	0.002
Constante	0.2176073	0.219
R^2		0.9034
Wald $\chi^2(2)$		459.1

Discusión

Para determinar qué modelo pronostica mejor la desnutrición infantil en cinco regiones del norte peruano, se estimaron tres modelos: de datos agrupados, de efectos fijos y efectos aleatorios; tal como lo desarrolla Pizan (2018) cuando analiza la influencia de la inversión pública regional en la pobreza del Perú mediante un modelo econométrico panel data, donde la variable dependiente fue la tasa de pobreza, y las variables independientes fueron la inversión pública regional en educación y en salud, y la inversión pública regional en saneamiento, la dimensión temporal comprendió el periodo 2012-2016; de igual manera Carbajal *et al.* (2018), analizó la correspondencia entre el sector automotriz en los sujetos federales de la frontera norte de México (Baja California, Sonora, Chihuahua, Coahuila, Nuevo León y Tamaulipas) y la dinámica de crecimiento de la producción de la manufactura.

Con el fin de analizar si el modelo de datos agrupados o de efectos aleatorios sería el más conveniente, se aplicó la prueba del Multiplicador de Lagrange para Efectos Aleatorios, en la que se determinó que no existe diferencia significativa entre el modelo de datos agrupados y el de efectos aleatorios ($p = 0.2108 < 0.05$), prueba que también se desarrolló en la investigación de Pizan (2018) en donde concluye que se debe optar por el modelo de efectos aleatorios, sin embargo en la investigación de Carbajal *et al.* (2018) al igual que en este estudio no hubo diferencia significativa entre los modelos antes descritos.

La prueba de Hausman fue empleada para decidir entre el modelo de efectos aleatorios contra el de efectos fijos, encontrándose que el primero fue el más conveniente ($p = 0.1429 < 0.05$); coincidiendo con Martínez *et al.* (2019) cuando analizaron si el capital social es un factor del emprendimiento para los países de la OCDE en cuanto a emplear un modelo de efectos aleatorios; pero Pizan y Carbajal, *et al.* (2018) seleccionaron como adecuado un modelo de efectos fijos.

Como parte de los supuestos en el desarrollo de esta investigación se llevó a cabo el test de Wooldridge con el fin de analizar problemas de autocorrelación, problema que esta investigación no presentó ($p > 5\%$, donde $p = 0.2067$), sin embargo este fue un problema en el estudio de Romo (2016) en donde propuso como meta principal el análisis de la efectividad de variables macroeconómicas para con la diversificación del número de suspensiones de pago en una muestra de diez ciudades pertenecientes a España, mencionado autor estimó un modelo de datos panel de efectos fijos, sin embargo presentó problemas de autocorrelación, que solucionó haciendo uso de desviaciones típicas robustas; escenario similar presentó la investigación de Rodríguez *et al.* (2018) que analizó la efectividad que se presenta en el sistema de gasto público en el sector educativo y el gasto nacional sanitario en el estado de bienestar de la población, en este estudio señalaron los autores que presentaron problemas de autocorrelación y optaron por la robustez del modelo para dar solución a este inconveniente. Señalan Carbajal *et al.* (2018) que en su investigación también tuvieron problemas de autocorrelación y que lo corrigieron a través de Errores Estándar Corregidos para Panel (Panel Corrected Standard Errors o PCSE), sin embargo, Pizan (2018) hizo uso de un modelo autorregresivo de orden 1 para dar solución a la misma.

La presente investigación también hizo uso de la prueba modificada de Wald para estudiar la heterocedasticidad, la misma que reveló no tener este inconveniente ($p > 5\%$, donde $p = 0.1214$), resultado que no se encontró en los trabajos de Romo (2016); Rodríguez *et al.* (2017); Carbajal *et al.* (2018) y en el estudio de Pizan (2018) ya que estos si presentaron problemas de heterocedasticidad, que corrigieron con estimadores Errores Estándar Corregidos para Panel – PCSE, este último autor desarrolló además Mínimos Cuadrados Generalizados Factibles.

Se analizó también la existencia de correlación contemporánea mediante la prueba de Breusch y Pagan, encontrando que este estudio presenta el problema antes mencionado ($p < 5\%$, donde $p = 0.0068$), para la solución de la misma se aplicó los Errores Estándar Corregidos para Panel – PCSE; tal como lo hicieron los autores Romo (2016); Rodríguez *et al.* (2017); Carbajal *et al.* (2018) y en el estudio de Pizan (2018), señalando además estos dos últimos autores que los Errores Estándar

Corregidos para Panel – PCSE genera parámetros más significativos que los Mínimos Cuadrados Generalizados Factibles.

Esta investigación encontró como mejor modelo el de efectos aleatorios con la corrección de Errores Estándar Corregidos para Panel – PCSE considerando un $R^2 = 90.4\%$, mientras que Romo (2016) encontró que las variables propuestas logran explicar el 53,8% de la variación de los procedimientos concursales demostrado por el R^2 obtenido en el modelo por efectos fijos usando desviaciones típicas robustas, sin embargo, el mejor modelo para Carbajal *et al.* (2018) fue el de efectos aleatorios Corregidos con un $R^2 = 85.7\%$.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

1. En general, la incidencia de la pobreza total en las regiones en estudio se ha reducido entre el periodo 2010-2018, Tumbes disminuyó de 20.1% a 12.6%, Piura de 42.5% a 23.1%, Lambayeque de 35.3% a 12.6%, La Libertad de 32.6% a 23.1%, y Cajamarca de 49.1% a 41.9%; en relación a la tasa de analfabetismo de la población femenina de 15 a más años de edad fue decreciendo para todos los periodos en estudio, en el caso de Tumbes de 5.3% a 3.3%, Piura de 13.3% a 10.4%, Lambayeque de 11.9% a 7.6%, La Libertad de 11.7% a 8.1% y Cajamarca de 23.2% a 17.2%. Se resalta que la disminución fue permanente en todas las regiones.
2. Al analizar la evolución de la desnutrición infantil, se evidenció un notable decrecimiento en las regiones para el periodo el 2010 – 2018, en Tumbes de 12.8% a 8.1%, Piura de 22.9% a 13.1%, Lambayeque de 17.8% a 9.3%, La Libertad de 26.1% a 14.8% y Cajamarca de 40.5% a 23.9% al 2015. Se presentaron algunas variantes de incremento para todas las regiones en diferentes periodos.
3. El modelo de efectos aleatorios de sección cruzada tanto para región como para tiempo fue el mejor de acuerdo a la prueba de Hausman, siendo estimado como:
$$\text{LnDesnutrición}_{it} = 0.2176073 + 0.6163622 * \text{LnPobreza}_{it} + 0.2683425 * \text{LnAnalfabetismo}_{it}$$

Donde *i* representa la región y *t* el año del periodo 2010-2108.
4. El modelo estimado no presentó problemas de estimación e inferencia como la autocorrelación y heterocedasticidad, no obstante, se encontró inconvenientes de correlación contemporánea según la prueba de Breusch y Pagan ($p < 5\%$), la misma que se corrigió con Errores Estándar Corregidos para Panel – PCSE, generando coeficientes significativos a excepción de la constante.

5.2. Recomendaciones

En base a los resultados obtenidos se encontró que Cajamarca registra los porcentajes más altos en la incidencia de la pobreza total, tasa de analfabetismo de la población femenina de 15 a más años y la desnutrición infantil para todos los años, es por ello que se recomienda que las instituciones públicas y privadas propongan alternativas de solución para mejorar estos indicadores, ya que interviniendo en la educación y la pobreza se podrá corregir la desnutrición, considerando además que con este aporte investigativo a través del modelo de datos panel se podrá anticipar la desnutrición infantil.

Se recomienda a los investigadores que propongan más modelos econométricos enfocados no sólo a la gestión empresarial, ya que como se ha demostrado en este estudio, se pueden estimar modelos econométricos en todas las áreas del conocimiento con el fin de aportar a la sociedad en busca de mejoras a las distintas problemáticas existentes.

Por último, recomiendo a las entidades públicas y privadas considerar los resultados de este estudio para mejorar los indicadores de la desnutrición crónica infantil, ya que como se conoce la desnutrición llega incluso a afectar el buen desarrollo cerebral del niño, disminuyendo su capacidad intelectual, es por ello que si se conocen con anticipación pronósticos de como esta se podría presentar, lograrían realizar grandes cambios en nuestra sociedad.

REFERENCIAS BIBLIOGRÁFICAS

- Alonso, A., Fernández, F. y Gallastegui, J. (2005), *Econometría*. España: Prentice Hall Pearson
- Banco Mundial. (2016). *Nutrición infantil: la inversión más rentable del mundo*. Recuperado de: <http://www.bancomundial.org/es/news/feature/2016/05/03/nutricion-infantil-la-inversion-mas-rentable-del-mundo-america-latina>
- Beck, Nathaniel & Katz, Jonathan. (1995). What to Do (and Not to Do) with Times-Series Cross-Section Data. *American Political Science Review*. 89. 10.2307/2082979.
- Bernal, C. (2010). *Metodología de la investigación*. (3ra edición). Colombia: Pearson educación
- Carbajal, Y., Carrillo, B. y De Jesús, L. (2018). Dinámica productiva del sector automotriz y la manufactura en la frontera norte de México: Un análisis con datos de panel, 1980-2014. *Frontera Norte*, 30 (59), 29 - 56. <https://doi.org/10.17428/rfn.v30i59.511>
- Castillo, S. y Calsin, G. (2017). Factores socioeconómicos que determinan la desnutrición crónica infantil en niños menores de cinco años en el Perú – 2014. *Semestre económico*, 6 (1).
- Carracedo, P. (2017). *Metodología espacio-temporal con datos de panel. Estudio de la mortalidad Europea*, (Tesis de postgrado). Universidad Politécnica de Valencia. España
- Courtois, V. (2014). *Situación nutricional en escolares de Santiago el Pinar, Chiapas (México)*, (Tesis de postgrado). Universidad Complutense de Madrid. España

- De La Rosa, M. (2016). *Introducción a modelos de datos de panel*. Universidad de Valladolid. Recuperado de: <https://uvadoc.uva.es/bitstream/10324/21944/1/TFG-E-321.pdf>
- Ediassa,(s.f.). *La econometría: historia y metodología*. Recuperado de: <https://www.ediasa.es/files/capitulos/9788496062948.pdf>
- Estevan, V.,Moral, P.,et al. (2009). *Econometría básica Aplicada con Gretl*. España: Universidad del País Vasco
- García, L. (2018). *Factores asociados a desnutrición crónica infantil en Mozambique*, (Tesis de postgrado). Universidad de Las Palmas de Gran Canaria. España
- Gómez, S. (2012). *Metodología de la investigación*. Recuperado de : http://www.aliat.org.mx/BibliotecasDigitales/Axiologicas/Metodologia_de_la_investigacion.pdf
- Gujarati. D.y Porter.D. (2010). *Econometría*. México: Editorial Mc Graw Hill
- Instituto Nacional de Estadística e Informática-INEI. (2017). *INEI: desnutrición infantil disminuyó 5,2% en los últimos 5 años en el Perú*. Recuperado de: <https://andina.pe/agencia/noticia-inei-desnutricion-infantil-disminuyo-52-los-ultimos-5-anos-el-peru-711991.aspx>
- Marrero, G. (2002). *Econometría: fundamentos, métodos y aplicaciones*. España: Universidad Complutense de Madrid
- Mayorga, M. y Muñoz, E. (2000). *La técnica de datos de panel: Una guía para su uso e interpretación*. Costa Rica: Banco Central de Costa Rica
- Martínez,G., Saavedra, R. y Morones, A. (2019). Capital Social como factor de emprendimiento en los países de la OCDE: implementación de un modelo con datos de panel. *Revista de métodos cuantitativos para la economía y la empresa*, (30), 312 – 326. <https://doi.org/10.46661/revmetodoscuanteconemp resa.3512>

Minchón, C., Minchón, D., Minchón, M. y Vizconde, T. (2015). Modelos lineales generalizados para pronóstico de la anemia infantil mediante factores asociados. *UCV-Scientia*, 7 (2).

Montero, R (2005): Test de Hausman. Documentos de Trabajo en Economía Aplicada. Universidad de Granada.:España

Moreno, F., He, Y., Merino, C. (2019). *Manual Práctico para Datos de Panel*. España: Universidad Complutense de Madrid

Organización mundial de la salud. (2018). *El hambre en el mundo sigue aumentando, advierte un nuevo informe de la ONU*. Recuperado de: <https://www.who.int/es/news-room/detail/11-09-2018-global-hunger-continue-to-rise---new-un-report-says>

Pérez, C. (2008). *Econometría Avanzada Técnicas y Herramientas*. Pearson

Pizan, J. (2018). *Inversión pública regional y su influencia en la pobreza del Perú mediante un modelo econométrico panel data, periodo 2012 – 2016*. Universidad Inca Garcilaso de la Vega. Perú

Resumen Latinoamericano (2019). Colombia: Desnutrición infantil supera los 500 mil casos. Recuperado de: <http://www.resumenlatinoamericano.org/2019/03/07/colombia-desnutricion-infantil-supera-los-500-mil-casos/>

Rinaldi, J. (2015). *Los inesperados datos de la pobreza infantil en EE.UU*. Recuperado de: <https://actualidad.rt.com/sociedad/172091-cifras-pobreza-infantil-eeuu>

Rodríguez, M., Freire, M. y Pais, C. (2018). El efecto del gasto público sanitario y educativo en la determinación del bienestar de los países de la OCDE: un modelo con datos de panel. *Cuadernos de Economía*, 41, 104-118

- Romo, B. (2016). *Modelo de datos de panel para el análisis del efecto de variables macroeconómicas en los procedimientos concursales de empresas españolas*. Universidad Pontificia Comillas ICAI-ICADE. España
- Santamargarita, S., Sancho, C., Vivas, D. y Uso, R. (2013). Evaluación de un modelo de predicción del gasto farmacéutico en atención primaria de salud basado en variables demográficas. *Gerenc. Polit. Salud -Colombia*, 12 (25),120 - 132
- Stock, J. y Watson, M. (2012). *Introducción a la econometría*. España: Prentice-Hall Pearson
- Valero, Jorge. (1991). MARY S. MORGAN: The History of Econometric Ideas, Cambridge University Press, 1990, 296 pp.. *El Trimestre Económico*. LVIII (4). 863-867.
- Wooldridge, J. (2010). *Introducción a la econometría: un enfoque moderno*.4ta edición. Usa: South-Western

ANEXOS

Anexo 1. Cuadro de operacionalización de variables

Variable	Función	Definición	Indicador	Tipo	Escala	Instrumento
Incidencia de pobreza total	Independiente	Estado representado por una falta rígida de escaseces humanas básicas, circunscribiendo agua potable, alimentos, salud, instalaciones sanitarias, vivienda, educación e información.	Porcentaje (%)	Cuantitativa	Razón	Estadística INEI
Tasa de analfabetismo de la población femenina de 15 y más años	Independiente	Concibe reseña de aquella persona que no sabe escribir ni leer. La expresión enlase asumir un uso amplio y es utilizado para mencionar a los sujetos que son iletrados o que escasean de los conocimientos más elementales en alguna disciplina.	Porcentaje (%)	Cuantitativa	Razón	Estadística INEI
Tasa de desnutrición infantil	Dependiente	La desnutrición puede definirse como la consecuencia del consumo escaso e ineficiente de alimentos y de la muestra consecutiva de malestares patógenos. La desnutrición puede lograr ser crónica y aguda, en base a la edad y el peso.	Porcentaje (%)	Cuantitativa	Razón	Estadística INEI
Espacial	Componente	Considera a una fracción de espacio determinado por algunas particularidades, contextos específicos o habituales como menciónese el clima, la forma de régimen o la topografía.	Región	Cualitativa	Nominal	Estadística INEI
Temporal	Componente	Periodo de tiempo que la tierra tarda en dar una vuelta completa alrededor del sol, que a su vez está compuesto por 12 meses.	Año	Cualitativa	Nominal	Estadística INEI

Anexo 2. Matriz de consistencia

Problema	Objetivos	Hipótesis	Variables e indicadores	Metodología
¿Qué modelo de datos panel pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018?	<p>Objetivo General Evaluar el modelo de datos panel que pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018</p> <p>Objetivos Específicos</p> <ul style="list-style-type: none"> • Analizar la evolución de la incidencia de la pobreza total y la tasa de analfabetismo de la población femenina de 15 y más años de edad en el norte peruano en el periodo 2010-2018 • Analizar la evolución de la desnutrición infantil en el norte peruano en el periodo 2010-2018 • Estimar el modelo de datos panel que pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018. • Determinar el modelo de datos panel que pronostique mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018 	El modelo de datos panel a efectos fijos pronostica mejor la desnutrición infantil en cinco regiones del Norte Peruano en el periodo 2010-2018	<p>Variabes Independientes</p> <ul style="list-style-type: none"> • Pobreza total • Analfabetismo de la población femenina de 15 y más años <p>Indicadores de las V.I.</p> <ul style="list-style-type: none"> • Porcentaje (%) • Porcentaje (%) <p>Variable Dependiente Desnutrición infantil</p> <p>Indicador de las V.D. Porcentaje (%)</p> <p>Componentes:</p> <ul style="list-style-type: none"> • Temporal • Espacial 	<p>Método de Investigación: Deductivo</p> <p>Diseño de Investigación: Longitudinal y espacial</p> <p>Población y muestra:</p> <ul style="list-style-type: none"> • Población: indicadores anuales de la desnutrición infantil, pobreza total y analfabetismo de la población femenina de 15 y más años de edad en cinco regiones del Perú. • Muestra: 48 observaciones anuales de la desnutrición infantil, pobreza total y analfabetismo de la población femenina de 15 años a más, para el periodo 2010 – 2018 en la región de Lambayeque, Cajamarca, Piura, Tumbes y La Libertad. <p>Técnica e Instrumento de recolección de datos</p> <ul style="list-style-type: none"> • Técnica: análisis de base de datos oficiales • Instrumento: Estadísticas-INEI

Anexo 3. Datos analizados en logaritmo natural

Año	Región	% Desnutrición	% Pobreza	% Analfabetismo Mujeres
2010	Tumbes	2.54944517	3.00071982	1.66770682
2011	Tumbes	2.39789527	2.47148363	1.5260563
2012	Tumbes	2.14006616	2.64048488	1.5040774
2013	Tumbes	2.20827441	2.5733753	1.36097655
2014	Tumbes	2.11625551	2.49568172	1.43508453
2015	Tumbes	2.20827441	2.41591378	1.22377543
2016	Tumbes	2.00148	2.37954613	1.25276297
2017	Tumbes	2.10413415	2.59076704	1.16315081
2018	Tumbes	2.09186406	2.52972067	1.19392247
2010	Piura	3.13113691	3.74950408	2.58776404
2011	Piura	3.03974916	3.69162568	2.37954613
2012	Piura	3.03495299	3.66047998	2.27212589
2013	Piura	3.2148678	3.62647199	2.34180581
2014	Piura	3.07731226	3.38574528	2.360854
2015	Piura	3.01062089	3.39785848	2.37024374
2016	Piura	2.72785283	3.53281027	2.34180581
2017	Piura	2.76631911	3.20437114	2.35137526
2018	Piura	2.57261223	3.13983262	2.34180581
2010	Lambayeque	2.87919846	3.56388296	2.4765384
2011	Lambayeque	2.78501124	3.3502549	2.32238772
2012	Lambayeque	2.58776404	3.27487801	2.21920348
2013	Lambayeque	2.65324196	3.06758771	2.20827441
2014	Lambayeque	2.66025954	3.05400118	2.1517622
2015	Lambayeque	2.6461748	2.93438864	2.10413415
2016	Lambayeque	2.46809953	2.77321353	2.24070969
2017	Lambayeque	2.35137526	2.59076704	2.06686276
2018	Lambayeque	2.2300144	2.52972067	2.02814825
2010	La Libertad	3.26193531	3.48431229	2.45958884
2011	La Libertad	3.14415228	3.3502549	2.37954613
2012	La Libertad	3.15700042	3.27487801	2.3321439
2013	La Libertad	3.09557761	3.42132675	2.17475172
2014	La Libertad	2.99071973	3.38574528	2.20827441
2015	La Libertad	2.76631911	3.39785848	2.19722458
2016	La Libertad	2.50143595	3.12104246	1.98787435
2017	La Libertad	2.74727091	3.20437114	2.2300144
2018	La Libertad	2.69462718	3.13983262	2.09186406
2010	Cajamarca	3.70130197	3.89385903	3.14415228
2011	Cajamarca	3.62700405	4.00587758	3.03495299
2012	Cajamarca	3.58629287	3.9763115	2.90142159
2013	Cajamarca	3.57234564	3.9227651	3.04452244
2014	Cajamarca	3.47196645	3.91760738	2.98061864
2015	Cajamarca	3.17387846	3.87515153	2.94968834
2016	Cajamarca	3.25809654	3.85777799	3.04452244
2017	Cajamarca	3.28091122	3.86157146	2.88480071
2018	Cajamarca	3.31054301	3.7340918	2.84490938

Anexo 4. Modelo de datos agrupados

Random-effects GLS regression		Number of obs	=	45		
Group variable: Región		Number of groups	=	5		
R-sq:		Obs per group:				
within	= 0.5834		min	=	9	
between	= 0.9762		avg	=	9.0	
overall	= 0.9020		max	=	9	
corr(u_i, X) = 0 (assumed)		Wald chi2(2)	=	190.99		
		Prob > chi2	=	0.0000		
lnDesnutrición	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lnPobreza	.5398203	.1190608	4.53	0.000	.3064655	.7731752
lnAnalfabetismo	.3535067	.1197923	2.95	0.003	.1187182	.5882952
_cons	.2767282	.2148254	1.29	0.198	-.1443218	.6977783
sigma_u	.06149364					
sigma_e	.1360752					
rho	.16958831	(fraction of variance due to u_i)				

Anexo 5. Modelo de efectos aleatorios

Random-effects GLS regression		Number of obs	=	45		
Group variable: Región		Number of groups	=	5		
R-sq:		Obs per group:				
within	= 0.5834		min	=	9	
between	= 0.9762		avg	=	9.0	
overall	= 0.9020		max	=	9	
corr(u_i, X) = 0 (assumed)		Wald chi2(2)	=	190.99		
		Prob > chi2	=	0.0000		
lnDesnutrición	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lnPobreza	.5398203	.1190608	4.53	0.000	.3064655	.7731752
lnAnalfabetismo	.3535067	.1197923	2.95	0.003	.1187182	.5882952
_cons	.2767282	.2148254	1.29	0.198	-.1443218	.6977783
sigma_u	.06149364					
sigma_e	.1360752					
rho	.16958831	(fraction of variance due to u_i)				

Anexo 6. Prueba del Multiplicador de Lagrange para Efectos Aleatorios

```

Breusch and Pagan Lagrangian multiplier test for random effects

lnDesnutrición[Región,t] = Xb + u[Región] + e[Región,t]

Estimated results:

```

	Var	sd = sqrt(Var)
lnDesnu~n	.2146982	.4633554
e	.0185165	.1360752
u	.0037815	.0614936

```

Test:  Var(u) = 0
          chibar2(01) =    0.65
          Prob > chibar2 =    0.2108

```

La hipótesis nula de la prueba es $\sigma_u^2 = 0$, la misma que no se rechaza. No existe diferencia entre el modelo de datos agrupados y el de efectos aleatorios.

Anexo 7. Modelo de efectos fijos

Fixed-effects (within) regression		Number of obs	=	45	
Group variable: Región		Number of groups	=	5	
R-sq:		Obs per group:			
within	= 0.5997	min	=	9	
between	= 0.9648	avg	=	9.0	
overall	= 0.8915	max	=	9	
corr(u_i, Xb) = -0.8075		F(2,38)	=	28.46	
		Prob > F	=	0.0000	
lnDesnutrición	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lnPobreza	.461336	.1274615	3.62	0.001	.2033037 .7193683
lnAnalfabetismo	.6306148	.2047691	3.08	0.004	.2160814 1.045148
_cons	-.0860082	.3895931	-0.22	0.826	-.8746982 .7026818
sigma_u	.16019034				
sigma_e	.1360752				
rho	.5808613	(fraction of variance due to u_i)			
F test that all u_i=0: F(4, 38) = 2.81				Prob > F = 0.0386	

Anexo 8. Test de Hausman

	—— Coefficients ——		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) fe	(B) re		
lnPobreza	.461336	.5398203	-.0784843	.0455079
lnAnalfabe~o	.6306148	.3535067	.2771081	.1660729

b = consistent under Ho and Ha; obtained from xtreg
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

$\chi^2(2) = (b-B)' [(V_b-V_B)^{-1}] (b-B)$
 = 3.89
 Prob>chi2 = 0.1429

La hipótesis nula de la prueba es que los estimadores de efectos aleatorios y de efectos fijos no difieren sustancialmente. H_0 no se rechaza, por lo que se corrobora que es conveniente optar por el modelo de efectos aleatorios.

Anexo 9. Test de Wooldridge

Wooldridge test for autocorrelation in panel data			
H0: no first order autocorrelation			
F(1,	4)	= 2.266
Prob > F	=		0.2067

La hipótesis nula de la prueba es que no existe autocorrelación, tal como se muestra no se rechaza H_0 , por lo que se afirma que no existe problema de autocorrelación.

Anexo 10. Prueba de Breusch y Pagan

Breusch-Pagan LM test of independence: $\chi^2(10) =$	24.321, Pr = 0.0068
Based on 9 complete observations over panel units	

La hipótesis nula de la prueba es que existe “independencia transversal” (cross-sectional independence). Se rechaza H_0 , por lo que se afirma que existe un problema de correlación contemporánea.

Anexo 11. Prueba modificada de Wald para Heterocedasticidad

```

Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (5) =      8.71
Prob>chi2 =     0.1214
    
```

La hipótesis nula de la prueba es que no existe heterocedasticidad, siendo $\sigma_i^2 = \sigma^2$, como se muestra no se rechaza H_0 , no evidenciando problemas de heterocedasticidad.

Anexo 12. Modelo de efectos aleatorios con errores estándar corregidos

Linear regression, correlated panels corrected standard errors (PCSEs)						
Group variable:	Región	Number of obs	=	45		
Time variable:	Año	Number of groups	=	5		
Panels:	correlated (balanced)	Obs per group:				
Autocorrelation:	no autocorrelation	min	=	9		
		avg	=	9		
		max	=	9		
Estimated covariances	=	15	R-squared	=	0.9034	
Estimated autocorrelations	=	0	Wald chi2(2)	=	459.10	
Estimated coefficients	=	3	Prob > chi2	=	0.0000	
lnDesnutrición	Panel-corrected					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lnPobreza	.6163622	.1067656	5.77	0.000	.4071055	.825619
lnAnalfabetismo	.2683425	.0875076	3.07	0.002	.0968307	.4398543
_cons	.2176073	.1769925	1.23	0.219	-.1292917	.5645063

MODELOS DE DATOS PANEL PARA PRONÓSTICO DE LA DESNUTRICIÓN INFANTIL EN CINCO REGIONES DEL NORTE PERUANO, PERIODO 2010 – 2018

by Betsy Arrunategui Huaman

Submission date: 18-Aug-2021 03:48PM (UTC-0500)

Submission ID: 1632955312

File name: Informe_Final-Betsy_Pamela_Arrunategui_Huam_n.docx (315.59K)

Word count: 12091

Character count: 64312

MODELOS DE DATOS PANEL PARA PRONÓSTICO DE LA DESNUTRICIÓN INFANTIL EN CINCO REGIONES DEL NORTE PERUANO, PERIODO 2010 – 2018

ORIGINALITY REPORT

22%

SIMILARITY INDEX

21%

INTERNET SOURCES

7%

PUBLICATIONS

9%

STUDENT PAPERS

PRIMARY SOURCES

1	repositorio.uns.edu.pe Internet Source	2%
2	www.inei.gob.pe Internet Source	2%
3	www.coursehero.com Internet Source	2%
4	hdl.handle.net Internet Source	1%
5	repositorio.usmp.edu.pe Internet Source	1%
6	bibdigital.epn.edu.ec Internet Source	1%
7	1library.co Internet Source	1%
8	investigadores.cide.edu Internet Source	1%

issuu.com

9	Internet Source	1 %
10	andina.pe Internet Source	1 %
11	doaj.org Internet Source	1 %
12	Submitted to UNIV DE LAS AMERICAS Student Paper	<1 %
13	edoc.pub Internet Source	<1 %
14	www.clubensayos.com Internet Source	<1 %
15	www.scielo.org.mx Internet Source	<1 %
16	peru.com Internet Source	<1 %
17	Submitted to Universidad de Ciencias y Humanidades Student Paper	<1 %
18	repositorio.urp.edu.pe Internet Source	<1 %
19	Submitted to Universidad San Ignacio de Loyola Student Paper	<1 %

20	es.scribd.com Internet Source	<1 %
21	www.grade.org.pe Internet Source	<1 %
22	Submitted to Infile Student Paper	<1 %
23	repositorio.uoosevelt.edu.pe Internet Source	<1 %
24	bdigital.unal.edu.co Internet Source	<1 %
25	expansion.mx Internet Source	<1 %
26	www.buenastareas.com Internet Source	<1 %
27	www.camcopiura.org.pe Internet Source	<1 %
28	Eduardo Rodríguez Juárez. "Efecto de la educación en la precarización del trabajo en tiempos de la COVID-19 en México", Revista Mexicana de Economía y Finanzas, 2021 Publication	<1 %
29	idoc.pub Internet Source	<1 %
30	ri.ues.edu.sv Internet Source	<1 %

31 Jorge Eduardo Mendoza Cota. "Impacto de la inversión extranjera directa en el crecimiento manufacturero en México", Problemas del Desarrollo. Revista Latinoamericana de Economía, 2011
Publication <1 %

32 gotitasdeamorbendecidas.com
Internet Source <1 %

33 repositorio.upao.edu.pe
Internet Source <1 %

34 docplayer.es
Internet Source <1 %

35 gestion.pe
Internet Source <1 %

36 repositorio.uigv.edu.pe
Internet Source <1 %

37 ri.uaemex.mx
Internet Source <1 %

38 pirhua.udep.edu.pe
Internet Source <1 %

39 www.itson.mx
Internet Source <1 %

40 documents.mx
Internet Source <1 %

41	Submitted to Universidad Andina del Cusco Student Paper	<1 %
42	pesquisa.bvsalud.org Internet Source	<1 %
43	Submitted to unsaac Student Paper	<1 %
44	www.elperulegal.com Internet Source	<1 %
45	www.tdx.cat Internet Source	<1 %
46	webinei.inei.gob.pe Internet Source	<1 %
47	archive.org Internet Source	<1 %
48	Eduardo Rosas Rojas. "Intervención esterilizada en los mercados de cambios de América Latina: Brasil, Chile y México", Problemas del Desarrollo. Revista Latinoamericana de Economía, 2011 Publication	<1 %
49	biblioteca.ucm.es Internet Source	<1 %
50	bucserver01.unican.es Internet Source	<1 %

51	core.ac.uk Internet Source	<1 %
52	latrompetadejerico.com Internet Source	<1 %
53	members.tripod.com.ve Internet Source	<1 %
54	Submitted to Pontificia Universidad Catolica del Peru Student Paper	<1 %
55	Submitted to Universidad de Lima Student Paper	<1 %
56	repositorio.ucp.edu.co Internet Source	<1 %
57	www.perupolitico.com Internet Source	<1 %
58	prezi.com Internet Source	<1 %
59	pt.scribd.com Internet Source	<1 %
60	rabida.uhu.es Internet Source	<1 %
61	repositorio.unican.es Internet Source	<1 %
62	repositorioacademico.upc.edu.pe	

Internet Source

<1 %

63

studylib.es

Internet Source

<1 %

64

worldwidescience.org

Internet Source

<1 %

65

www.mysciencework.com

Internet Source

<1 %

66

Submitted to Facultad Latinoamericana de Ciencias Sociales (FLACSO) - Sede Ecuador

Student Paper

<1 %

67

ingbioquimica.blogspot.com

Internet Source

<1 %

68

repositorio.uchile.cl

Internet Source

<1 %

69

www.cochranlibrary.com

Internet Source

<1 %

70

www.mdpi.com

Internet Source

<1 %

71

ideas.repec.org

Internet Source

<1 %

72

repositorio.flacsoandes.edu.ec

Internet Source

<1 %

73

repositorio.usil.edu.pe

Internet Source

<1 %

74

repositorio.utelesup.edu.pe

Internet Source

<1 %

75

www.haciendachiapas.gob.mx

Internet Source

<1 %

76

www.pnudguatemala.org

Internet Source

<1 %

77

BLADIMIR DELAHOZ ROSALES, JOSE ANTONIO CAMACHO BALLESTA, IGNACIO TAMAYO TORRES. "HUMAN DEVELOPMENT EFFECTS IN ENTREPRENEURSHIP: AN INTERNATIONAL ANALYSIS", DYNA MANAGEMENT, 2021

Publication

<1 %

78

Luisa Fernanda Gutiérrez. "Determinantes de la concentración de deuda en el corto plazo para empresas del sector real de Colombia : análisis de datos panel basado en información pública", Ensayos sobre Política Económica, 2010

Publication

<1 %

79

desnutricionmundialinfantil.wordpress.com

Internet Source

<1 %

80

doku.pub

Internet Source

<1 %

moam.info

81

Internet Source

<1 %

82

repositorio.untumbes.edu.pe

Internet Source

<1 %

83

www.canacero.org.mx

Internet Source

<1 %

84

www.researchgate.net

Internet Source

<1 %

85

www.scribd.com

Internet Source

<1 %

86

Horacio Catalán. "Curva ambiental de Kuznets: implicaciones para un crecimiento sustentable", Economía Informa, 2014

Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off