

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

EL MODELO DE AUTORREGRESIÓN SIMULTÁNEA: UNA
APLICACIÓN PARA EL PROGRAMA SOCIAL AVANCEMOS

Tesis sometida a la consideración de la Comisión del Programa de
Estudios de Posgrado en Estadística para optar al grado y título de
Maestría Académica en Estadística

GREIVIN JOSÉ BARBOZA RETANA

Ciudad Universitaria Rodrigo Facio, Costa Rica

2015

DEDICATORIA Y AGRADECIMIENTOS

A Dios Padre Todopoderoso, a Dios Hijo Redentor del Mundo, a la Santísima Virgen María y al Espíritu Santo, por darme la oportunidad y salud para realizar este trabajo de investigación.

A mi mamá, Flor de María, a mi papá, José Ángel, a mi hermano, Irving Fernando y a mi novia, Ana Catalina, por ser mi gran soporte durante los momentos de desarrollo profesional de mi vida y por acompañarme en este proceso.

A mis familiares que han partido de este mundo y que se encuentran en el Cielo, por ser un apoyo incondicional desde mi niñez, ayudándome en lo que estaba a su alcance y por ser un ejemplo de lucha y tenacidad ante sus enfermedades.

Al M.Sc. Olman Ramírez Moreira, por impulsarme a continuar en el camino académico, brindándome consejos como Profesor Guía y como Director de esta Tesis en su etapa inicial, que Dios lo tenga en su Santa Gloria.

Al Ph.D. Gilbert Brenes Camacho por ser una excelente persona y gran ejemplo para mí como profesional y por ayudarme en todo momento, siempre estuvo a mi lado y pude contar con él cuando más lo necesitaba.

Al Ph.D. Édgar Gutiérrez Espeleta y al M.Sc. Juan Bautista Chavarría, por colaborar con sus apreciaciones durante el desarrollo de esta investigación.

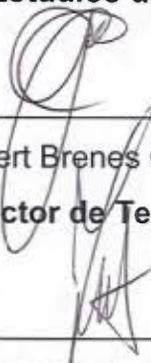
Al M.Sc. Fernando Ramírez y al M.Sc. Róger Bonilla por brindar sus valiosos aportes que robustecen el análisis efectuado.

Y a todos mis compañeros y amigos que estuvieron conmigo en esta etapa.

“Esta tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Estadística”.



M.Sc. Jorge Barquero Barquero
Representante del Decano
Sistema de Estudios de Posgrado

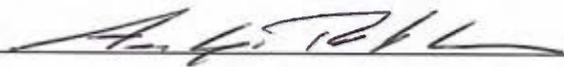


Ph.D. Gilbert Brenes Camacho
Director de Tesis

M.Sc. Fernando Ramírez Hernández
Asesor



M.Sc. Róger Bonilla Carrión
Asesor



Ph.D. Arodys Robles Soto
Representante del Director
Programa de Posgrado en Estadística



Greivin José Barboza Retana
Candidato

ÍNDICE GENERAL

DEDICATORIA Y AGRADECIMIENTOS	ii
RESUMEN	vii
LISTA DE CUADROS	viii
LISTA DE GRÁFICOS.....	x
LISTA DE MAPAS.....	xiii
LISTA DE ABREVIATURAS.....	xv
1. Presentación	1
2. Problema y objetivos	3
2. 1 Problema	3
2. 2 Justificación	5
2. 3 Objetivo general.....	9
2. 4 Objetivos específicos	9
3. Estado de la cuestión	10
3. 1 Introducción al análisis espacial.....	10
3. 2 Modelos de autorregresión espacial.....	11
3. 2. 1 Matriz de pesos (W).....	14
3. 2. 2 Fórmula general de la matriz de pesos W	16
3. 2. 3 Estimación de los parámetros del modelo SAR	17
3. 2. 4 Tipos de datos espaciales	19
3. 2. 5 Supuestos de los Modelos de Autorregresión Simultánea (SAR)	19
3. 2. 6 ¿Cómo se diagnostica la autocorrelación espacial?	21
3. 3 Estado de la cuestión sobre programas sociales y permanencia en el sistema educativo: aplicación del modelo SAR.....	23
3. 3. 1 Implementación de programas sociales.....	23

3. 3. 2 Estudios estadísticos para analizar la implementación de programas sociales.....	27
3. 3. 3 Programas sociales similares en América Latina.....	28
3. 3. 4 Factores que inciden en la no permanencia en el sistema educativo.....	29
3. 3. 5 Avancemos.....	35
3. 3. 6 El programa social Avancemos en cifras	37
4. Abordaje metodológico.....	43
4. 1 Análisis de la robustez del modelo SAR a la violación de supuestos	44
4. 1. 1 Revisión de literatura sobre simulaciones aplicadas al modelo SAR	44
4. 1. 2 Simulaciones planeadas para el proyecto	46
4. 1. 3 Características de la aplicación del modelo SAR a datos distritales costarricenses en términos geográficos	48
4. 2 Abordaje metodológico de la aplicación del SAR a un modelo que relaciona el programa social Avancemos con la permanencia en el sistema escolar.....	49
4. 2. 1 Población, unidad de estudio, alcance.....	49
4. 2. 2 Modelo de Avancemos	50
5. Evaluación de la robustez del modelo SAR ante cambios de escenarios de estimación.....	53
5. 1 Simulaciones para la autocorrelación espacial.....	54
5. 1. 1 Metodología.....	54
5. 1. 2 Análisis.....	56
5. 2 Simulaciones para diferentes matrices de pesos	64
5. 2. 1 Metodología.....	64
5. 2. 2 Análisis.....	65
5. 3 Simulaciones para evaluar la robustez del modelo SAR ante la falta de normalidad	78
5. 3. 1 Metodología.....	78
5. 3. 2 Análisis.....	79

5. 4 Síntesis preliminar	104
6. Aplicación del SAR al programa social Avancemos.....	106
6. 1 Análisis descriptivo	108
6. 2 Análisis del modelo SAR.....	124
6. 2. 1 Modelos con variable dependiente definida con datos administrativos ..	125
6. 2. 1. 1 Comportamiento de variables y supuestos del modelo SAR 2011 .	131
6. 2. 1. 2 Comportamiento de variables y supuestos del modelo SAR de las diferencias entre el 2007 y 2011.....	139
6. 2. 1. 3 Resultados de los modelos SAR 2011 con variable dependiente definida con datos administrativos.....	148
6. 2. 1. 4 El modelo SAR 2011 controlando por variables censales	153
6. 2. 2 Modelo SAR 2011 con variable dependiente definida con datos del Censo 2011.....	159
6. 2. 2. 1 Comportamiento de variables y supuestos del modelo SAR 2011 con variable dependiente del Censo 2011	163
6. 2. 2. 2 Resultados de los modelos SAR 2011 con variable dependiente definida con datos del Censo 2011.....	175
7. Conclusiones.....	180
7. 1. Conclusiones referidas a la utilización del modelo SAR con los datos distritales de Costa Rica.....	180
7. 2. Conclusiones referidas a la implementación del programa social Avancemos y la deserción colegial	182
8. Referencias bibliográficas	186

RESUMEN

En este trabajo de investigación se establecen las condiciones estadísticas necesarias para emplear el modelo de Autorregresión Simultánea (SAR) a los datos distritales de Costa Rica. Se evaluó su robustez ante cambios de escenarios de estimación, utilizando distintos grados de autocorrelación espacial con una distribución normal y beta. Este procedimiento se realizó también cambiando la matriz de pesos para analizar si se obtenían resultados distintos si se usaba otro tipo de matriz. Estos cálculos fueron realizados, utilizando el paquete estadístico R 2.14.

Con las simulaciones realizadas se comprobó la robustez de los supuestos del modelo SAR ante distintos niveles que determinan la autocorrelación espacial con diferentes tipos de matrices de pesos. Se concluyó que si se utiliza otro tipo de distribución que no sea la normal, los resultados siguen siendo coherentes ya que las estimaciones son cercanas a los parámetros teóricos propuestos, tomando en cuenta exclusivamente la estructura espacial de los distritos de Costa Rica.

Como aplicación se analizaron las variables relacionadas con el programa social *Avancemos* (tasa de la cantidad de becas y monto promedio de las becas) así como otras variables de control (población en edad colegial, consumo promedio de electricidad, cantidad de colegios y tasa del total de profesores) con la variable dependiente porcentaje de deserción colegial, todas ellas desagregadas a nivel distrital, para verificar la existencia de asociación entre ellas y de autocorrelación espacial que justifique el uso de este modelo SAR.

El modelo SAR logró demostrar la autocorrelación espacial presente en la estructura distrital del país hasta que el porcentaje de deserción fue estimado mediante variables censales. Con el porcentaje de deserción de registros administrativos, no se encontró autocorrelación espacial, ya que el sistema educativo costarricense sugiere que la autocorrelación existe, porque estudiantes que viven en un distrito pueden asistir al colegio de distritos vecinos.

LISTA DE CUADROS

Cuadro 1. Intervalos de la estimación del λ según valores teóricos utilizando distribución normal con matriz <i>queen</i> , 2011	59
Cuadro 2. Intervalos de confianza de los coeficientes, variando el λ y el σ utilizando distribución normal con matriz <i>queen</i> , 2011	61
Cuadro 3. Porcentaje de corridas para identificar la potencia de prueba con distribución normal utilizando matriz <i>queen</i> , 2011.....	64
Cuadro 4. Intervalos de la estimación del λ según valores teóricos utilizando distribución normal con matriz K vecinos, 2011	66
Cuadro 5. Intervalos de la estimación del λ según valores teóricos utilizando distribución normal con matriz de distancias, 2011	67
Cuadro 6. Intervalos de la estimación del coef. para X_1 según valores teóricos para coef. utilizando distrib. normal con matriz K vecinos, 2011	71
Cuadro 7. Intervalos de estimación del coef. para X_1 según valores teóricos para coef. utilizando distrib. normal con matriz de distancias, 2011	72
Cuadro 8. Porcentaje de corridas para identificar la potencia de prueba con distribución normal utilizando matriz de distancias, 2011	77
Cuadro 9. Porcentaje de corridas para identificar la potencia de prueba con distribución normal utilizando matriz K vecinos, 2011	78
Cuadro 10. Niveles de asimetría para λ y β teóricos utilizando distribución beta con matriz <i>queen</i> , 2011.....	80
Cuadro 11. Intervalos de la estimación del λ según valores teóricos utilizando distribución beta con matriz <i>queen</i> , 2011	81
Cuadro 12. Intervalos de la estimación del λ , según valores teóricos, utilizando distribución beta con matriz K vecinos, 2011.....	82

Cuadro 13. Intervalos de la estimación del λ , según valores teóricos, utilizando distribución beta con matriz de distancias, 2011	83
Cuadro 14. Intervalos de estimación del coeficiente para X_1 según valores teóricos para coeficientes, utilizando distribución beta con matriz <i>queen</i> , 2011.....	87
Cuadro 15. Intervalos de estimación del coeficiente para X_1 según valores teóricos para coeficientes, utilizando distribución beta con matriz <i>K</i> vecinos, 2011	88
Cuadro 16. Intervalos de estimación del coeficiente para X_1 según valores teóricos para coeficientes utilizando distribución beta con matriz de distancias, 2011.....	89
Cuadro 17. Porcentaje de corridas para identificar la potencia de prueba con distribución beta utilizando matriz <i>queen</i> , 2011.....	102
Cuadro 18. Porcentaje de corridas para identificar la potencia de prueba con distribución beta utilizando matriz <i>K</i> vecinos, 2011	103
Cuadro 19. Porcentaje de corridas para identificar la potencia de prueba con distribución beta utilizando matriz de distancias, 2011.....	103
Cuadro 20. Modelos SAR con variable dependiente obtenida de datos administrativos, 2011.....	149
Cuadro 21. Modelos SAR 2011 y 2007-2011.	150
Cuadro 22. Modelo SAR 2011 con el cambio en el porcentaje de deserción entre el 2012 y 2011.....	151
Cuadro 23. Modelo SAR 2011 y con variables independientes del Censo 2011.	155
Cuadro 24. Comparación de los modelos SAR 2011.	161
Cuadro 25. Modelos SAR 2011 con porcentaje de deserción según el censo.....	176
Cuadro 26. Modelos SAR 2011 con porcentaje de deserción según el censo por cantones.	177
Cuadro 27. Modelos lineales 2011 con porcentaje de deserción según el censo.....	178

LISTA DE GRÁFICOS

Gráfico 1. Intervalos de confianza simulados para lambda utilizando distribución normal con matriz <i>queen</i> , 2011.....	57
Gráfico 2. Gráfico de normalidad de los coeficientes con distribución normal y matriz <i>queen</i> , 2011.....	62
Gráfico 3. Intervalos de confianza simulados para lambda utilizando distribución normal con matriz K vecinos, 2011.....	68
Gráfico 4. Intervalos de confianza simulados para lambda utilizando distribución normal con matriz de distancias, 2011.....	69
Gráfico 5. Gráfico de normalidad de los coeficientes con distribución normal y matriz K vecinos, 2011.....	73
Gráfico 6. Gráfico de normalidad de los coeficientes con distribución normal y matriz de distancias, 2011.....	75
Gráfico 7. Intervalos de confianza simulados para lambda, utilizando distribución beta con matriz <i>queen</i> , 2011.....	84
Gráfico 8. Intervalos de confianza simulados para lambda, utilizando distribución beta con matriz K vecinos, 2011.....	85
Gráfico 9. Intervalos de confianza simulados para lambda, utilizando distribución beta con matriz de distancias, 2011.....	86
Gráfico 10. Gráfico de normalidad de los coeficientes con distribución beta y matriz <i>queen</i> , 2011.....	90
Gráfico 11. Gráfico de normalidad de los coeficientes con distribución beta y matriz K vecinos, 2011.....	94
Gráfico 12. Gráfico de normalidad de los coeficientes con distribución beta y matriz de distancias, 2011.....	98
Gráfico 13. Histogramas de las variables del modelo SAR 2011.....	126

Gráfico 14. Histogramas de las variables del modelo SAR de las diferencias 2007-2011.	128
Gráfico 15. Diagramas de dispersión del porcentaje de deserción en el 2011 por cada variable predictora, por distritos de Costa Rica.	133
Gráfico 16. Valores ajustados vs. Residuos del modelo SAR 2011.....	135
Gráfico 17. Variables predictoras vs. Residuos del modelo SAR 2011.....	136
Gráfico 18. Gráfico de Normalidad de los residuos del modelo SAR 2011.....	138
Gráfico 19. Diagramas de dispersión del porcentaje de deserción entre el 2007 y 2011 por cada variable predictora, por distritos de Costa Rica.....	140
Gráfico 20. Valores ajustados vs. Residuos del modelo SAR de las diferencias entre el 2007 y 2011.....	142
Gráfico 21. Variables predictoras vs. Residuos del modelo SAR 2007 y 2011.....	143
Gráfico 22. Gráfico de Normalidad de los residuos del modelo SAR 2007 y 2011.	145
Gráfico 23. Valores ajustados vs. Residuos del modelo SAR 2011 con variables del Censo 2011.	157
Gráfico 24. Gráfico de Normalidad de los residuos del modelo SAR 2011 con variables del Censo 2011.	158
Gráfico 25. Diagrama de dispersión del porcentaje de deserción de registros administrativos y datos censales.....	160
Gráfico 26. Histograma de la variable porcentaje de deserción según el censo, 2011 ..	163
Gráfico 27. Diagramas de dispersión del porcentaje de deserción según el censo en el 2011 por cada variable predictora, por distritos de Costa Rica.....	167
Gráfico 28. Valores ajustados vs. Residuos del modelo SAR 2011 con porcentaje de deserción según el censo.....	169
Gráfico 29. Variables predictoras vs. Residuos del modelo SAR 2011.....	170

Gráfico 30. Gráfico de Normalidad de los residuos del modelo SAR 2011 con porcentaje de deserción según el censo..... 172

LISTA DE MAPAS

Mapa 1. Costa Rica: porcentaje de deserción por rangos, 2011	109
Mapa 2. GAM: porcentaje de deserción por rangos, 2011	110
Mapa 3. Costa Rica: tasa de la cantidad de becas por cada mil habitantes con edades entre 12 y 21 años por rangos, 2011.....	111
Mapa 4. GAM: tasa de la cantidad de becas por cada mil habitantes con edades entre 12 y 21 años por rangos, 2011.....	112
Mapa 5. Costa Rica: monto promedio de becas en miles de colones por rangos, 2011	114
Mapa 6. GAM: monto promedio de becas en miles de colones por rangos, 2011	115
Mapa 7. Costa Rica: consumo promedio de electricidad por rangos, 2011	116
Mapa 8. GAM: consumo promedio de electricidad por rangos, 2011	117
Mapa 9. Costa Rica: cantidad de colegios por rangos, 2011.....	118
Mapa 10. GAM: cantidad de colegios por rangos, 2011	119
Mapa 11. Costa Rica: tasa del total de profesores por cada mil personas que asisten al colegio por rangos, 2011.....	120
Mapa 12. GAM: tasa del total de profesores por cada mil personas que asisten al colegio por rangos, 2011.....	121
Mapa 13. Costa Rica: población en edad colegial de 12 a 21 años por rangos, 2011 ...	122
Mapa 14. GAM: población en edad colegial de 12 a 21 años por rangos, 2011	123
Mapa 15. Índice de Moran a nivel local del porcentaje de deserción, 2011	131
Mapa 16. Coroplético de los Residuos del modelo SAR 2011.....	147
Mapa 17. Coroplético de los Residuos del modelo SAR 2007 y 2011.....	148
Mapa 18. Coroplético de los Residuos del modelo SAR 2011 con el cambio en el porcentaje de deserción entre el 2012 y 2011.....	152

Mapa 19. Índice de Moran a nivel local del porcentaje de deserción según el censo, 2011	165
Mapa 20. Coroplético de los Residuos del modelo SAR 2011 con porcentaje de deserción según el censo.....	174

LISTA DE ABREVIATURAS

CCP	Centro Centroamericano de Población
CCSS	Caja Costarricense del Seguro Social
CEN-Cinai	Centros de Educación y Nutrición y Centros Infantiles de Atención Integral
CGR	Contraloría General de la República
Desaf	Dirección de Asignaciones Familiares
EHPM	Encuesta de Hogares y Propósitos Múltiples
Enaho	Encuesta Nacional de Hogares
Fonabe	Fondo Nacional de Becas
GAM	Gran Área Metropolitana
IDS	Índice de Desarrollo Social
IMAS	Instituto Mixto de Ayuda Social
INA	Instituto Nacional de Aprendizaje
INEC	Instituto Nacional de Estadística y Censos
MEP	Ministerio de Educación Pública
Mideplan	Ministerio de Planificación Nacional y Política Económica
Mivah	Ministerio de Vivienda y Asentamientos Humanos
MTSS	Ministerio de Trabajo y Seguridad Social
NBI	Necesidades Básicas Insatisfechas
PISA	Programa Internacional para la Evaluación de Estudiantes
PNUD-Costa Rica	Programa de las Naciones Unidas para el Desarrollo en Costa Rica
SAR	Modelos de Autorregresión Simultánea
UCR	Universidad de Costa Rica
UGM	Unidad Geoestadística Mínima

1. Presentación

Los modelos de Autorregresión Simultánea (SAR) son una alternativa importante y de gran utilidad para analizar el comportamiento de distintos datos con características espaciales o geográficas. Asociado a ello, los modelos SAR presentan una valoración de ciertas variables independientes y su relación con la variable dependiente de interés, mediante un modelo de regresión, permitiendo estudiar su grado de asociación con cada una de las variables explicativas para identificar cuáles de ellas son las más importantes del modelo. A su vez, los modelos SAR ayudan a estimar de una mejor forma los parámetros obtenidos, ya que incorporan el efecto de la autocorrelación espacial que ocurre cuando los datos son espaciales, ubicados en áreas geográficas y para esta investigación los datos espaciales se encuentran en unidades geográficas pequeñas y específicamente en el caso de la aplicación por distrito. La autocorrelación espacial es uno de los problemas más comunes que presenta este tipo de análisis geográfico o espacial y dichos modelos tratan de corregir este inconveniente que ocurre por la naturaleza propiamente de la ubicación geográfica de las áreas espaciales.

En Costa Rica, en temáticas sociales y económicas, los datos están agrupados o reportados según distintos niveles de la división político-administrativa del país: provincias, cantones, distritos, segmentos censales, Unidades Geoestadísticas Mínimas (UGMs) según el censo de población del 2011, etc. La información distrital es profusamente usada porque permite un nivel de desagregación suficiente para describir desigualdades socioeconómicas a lo interno de los municipios. Por ejemplo, el Índice de Desarrollo Social (IDS) (Mideplan, 2013) se calcula para cada distrito y se utiliza para definir asignación de recursos públicos. El Atlas de Desarrollo Humano Cantonal de Costa Rica

(PNUD-Costa Rica y UCR, 2011) se estima en el nivel cantonal y es empleado por autoridades municipales para la definición de políticas locales. El presente trabajo estudia las características del modelo SAR cuando se aplica a datos del nivel distrital en Costa Rica. Se generan simulaciones para evaluar la robustez de este a la violación de algunos de los supuestos del modelo SAR. Las simulaciones calculadas en esta investigación pueden ser interpretadas únicamente con la estructura espacial de Costa Rica, ya que los datos espaciales corresponden a unidades político-administrativas que permanecen en la misma área geográfica durante el periodo de estudio.

Como ejercicio de aplicación, se utilizaron algunas variables que cumplen con dicha desagregación a nivel distrital afines al programa social *Avancemos*^{*}. Cabe señalar que la información recopilada para el programa social *Avancemos* corresponde al periodo del 2007 al 2011. Los datos están desagregados para cada uno de los 472 distritos de Costa Rica existentes al finalizar ese periodo. La información proviene de las bases de datos pertenecientes al Ministerio de Educación Pública (MEP), al Instituto Mixto de Ayuda Social (IMAS), al Instituto Nacional de Estadística y Censos (INEC), al Centro Centroamericano de Población (CCP) y al Ministerio de Planificación Nacional y Política Económica (Mideplan).

^{*} De ahora en adelante se utilizará un tratamiento de letra cursiva.

2. Problema y objetivos

2.1 Problema

El modelo lineal general o modelo lineal generalizado gaussiano –conocido popularmente como modelo de regresión lineal– requiere de ciertos supuestos para que sus estimaciones sean insesgadas y de variancia mínima: homoscedasticidad, normalidad de los errores, linealidad en los coeficientes, no autocorrelación de los errores, y ausencia de colinealidad perfecta (Kutner *et al.*, 2005). No obstante, si se dan ciertas condiciones estadísticas –diseños muestrales o experimentales específicos, grados diferenciales en la violación de supuestos, etc.–, las estimaciones del modelo lineal pueden seguir siendo confiables para los objetivos prácticos de una investigación. Por ejemplo, según las propiedades asintóticas del modelo lineal general, si se aumenta el tamaño de la muestra, la distribución muestral de los estadísticos de prueba se aproximan a una distribución normal, aun cuando la distribución de los errores no sea normal (Kutner *et al.*, 2005). Además, si se aplica a un diseño de experimentos, el modelo lineal general es robusto a la presencia de ciertos grados de heteroscedasticidad si el diseño experimental es balanceado (Montgomery, 2004).

No se recomienda utilizar el modelo lineal gaussiano a datos geográficos porque es altamente probable que se dé un proceso de autocorrelación: los errores de ciertas unidades geográficas estarán correlacionados con los errores de las unidades vecinas por la similitud entre estas o por los procesos de construcción social que delimitan las fronteras de estas. Además, no se puede variar el diseño muestral o experimental del que surgen los datos, porque estos generalmente están definidos por un sistema administrativo de recolección de

información. Por esta razón, se han desarrollado una serie de modelos estadísticos, agrupados según Anselin (1988, 2001) en el campo denominado “econometría espacial”. El modelo de autorregresión simultánea SAR es uno de ellos. En Costa Rica, los modelos de autorregresión simultánea se pueden utilizar cuando se tienen datos geográficos con distintos niveles de desagregación como, por ejemplo: cantonal, distrital, UGM (Unidad Geoestadística Mínima), etc.

Ahora bien, se han analizado las propiedades estadísticas del modelo SAR cuando se aplican principalmente a datos espaciales de EEUU y Europa (Cliff y Ord, 1972; Kelejian y Prucha, 1997; Lee, 1999). No obstante, no se sabe cuáles son las propiedades estadísticas de este modelo, cuando se aplica a estas desagregaciones que se tiene en Costa Rica. Entonces, este trabajo de investigación, lo que busca es tratar de analizar esas propiedades estadísticas del modelo a nivel de distrito. Si el modelo SAR más frecuentemente usado supone normalidad condicional de la variable dependiente, ¿es robusto dicho modelo a violaciones a este supuesto cuando la variable dependiente es una proporción? ¿Cuánto varían las estimaciones de los coeficientes del modelo SAR con distintos grados de autocorrelación espacial?

En el caso de la aplicación, se quiere analizar además ¿Qué debilidades o fortalezas tiene el modelo de Autorregresión Simultánea (SAR) cuando se aplica a datos distritales de Costa Rica sobre la asociación del programa social *Avancemos* y la deserción colegial? Dado que no se han establecido técnicas inferenciales para profundizar en la asociación estadística entre la implementación del programa social *Avancemos* y la deserción colegial, quiere observarse si con un modelo de Autorregresión Simultánea (SAR) puede medirse o analizarse esa relación controlando por variables relacionadas.

2. 2 Justificación

Los modelos de autorregresión simultánea son un caso específico de los modelos de regresión espacial y se aplican a todas las áreas. Son utilizados para estimar los parámetros de asociación espacial, en presencia de autocorrelación, representando así la relación entre áreas. Este tipo de análisis usa un parámetro que permite modelar la autocorrelación existente, lo cual ayuda a mejorar las estimaciones realizadas por el modelo de regresión. Por lo tanto, los modelos SAR son convenientes en este tipo de aplicaciones estadísticas debido a la naturaleza de los datos espaciales o geográficos y al trato del problema de la autocorrelación espacial. Precisamente, Fotheringham y Rogerson (1993) mencionan que la autocorrelación espacial tiene que ver con la relación de un valor asociado a un área respecto a otros valores situados en áreas distintas, ubicadas en los alrededores.

En esa misma línea, la autocorrelación espacial ocurre cuando los valores de las variables localizados en lugares cercanos no son independientes entre sí (Dormann *et al.*, 2007). Estos autores indican que existen factores en común que pueden ser identificados como causantes de la autocorrelación espacial desde el punto de vista biológico. Sin embargo, estos factores replanteados para la parte social, señalan que los fenómenos en general que se reflejan en los centros de estudio pueden estar relacionados con la distancia hacia ciertos lugares de interés como los centros educativos de secundaria.

En el caso de Costa Rica hay fenómenos sociales que se pueden describir usando datos geográficos y específicamente utilizando el modelo SAR cuando la información se presenta por distrito. En el momento de elaborar esta investigación, Costa Rica contaba con 472 distritos. Sin embargo, las características de estos distritos varían según el devenir histórico del país. La capital de Costa Rica es la ciudad de San José, ubicada en el cantón y la

provincia del mismo nombre. Sin embargo, el crecimiento urbano del siglo XX ha conformado de manera no planificada un “Gran San José” que, en términos administrativos, se denomina la Gran Área Metropolitana (GAM). La GAM se encuentra en el centro del país y abarca cantones de 4 provincias distintas: San José, Alajuela, Cartago y Heredia. Sus distritos son típicamente de pequeña extensión, altamente poblados y fácilmente interconectados por una red vial compuesta por caminos nacionales, municipales y vecinales. Además, es común encontrar más de una secundaria por cantón. Por el contrario, los distritos fuera de la GAM son menos poblados y de mayor extensión. En la mayoría de los casos, solo hay un colegio público por cantón. Todas estas características determinan las propiedades del modelo SAR cuando se aplican a Costa Rica.

El análisis de la implementación de programas sociales es un caso en el que se pueda aplicar algún modelo geoestadístico, porque las estadísticas oficiales se publican con desagregación distrital. Los programas sociales han sido aplicados desde hace varios años para mejorar a nivel general las condiciones de las personas en diversos campos. En este trabajo de investigación, la aplicación se enfoca específicamente en los procesos educativos mediante el programa social *Avancemos*.

¿Por qué el análisis del programa *Avancemos* (y otros programas sociales similares) se puede efectuar con un modelo SAR? Como ya se argumentó, este modelo geoestadístico es útil cuando se espera que haya un proceso de autocorrelación espacial. En otras palabras, que los errores en una unidad geográfica sean similares (“estén correlacionados”) con los errores de unidades vecinas. Estos procesos surgen por las similitudes entre las unidades geográficas contiguas. En el caso de *Avancemos*, se esperaría un proceso como el descrito por las siguientes razones:

- No existen instituciones de educación secundaria en todos los distritos. Los estudiantes pueden matricularse en los centros educativos más

cercanos, típicamente (pero no necesariamente) ubicados en el mismo cantón. En las estadísticas agregadas en el nivel distrital, dos distritos cercanos cualesquiera pueden tener grados similares de deserción o reprobación porque hay una alta probabilidad de que los estudiantes que residen en ellos asistan a la misma secundaria.

- Las becas de *Avancemos* son asignadas en el nivel distrital, pero la prontitud o efectividad con la que se asignen puede variar según la estructura geográfica de las instituciones político-administrativas: Direcciones Regionales del Ministerio de Educación Pública (MEP), municipalidades, u organizaciones comunales. Todas estas instancias agrupan a distritos cercanos.
- Los errores en los modelos lineales de regresión (sean geoestadísticos o no) pueden surgir por variables omitidas en el modelo pero que están asociadas con la variable dependiente. En el caso de los *indicadores meta* de *Avancemos*, como el porcentaje de deserción estudiantil o el porcentaje de reprobación, los fenómenos sociales que miden pueden estar determinados por factores que trascienden o cruzan los límites (“artificiales”) político-administrativos de los distritos en Costa Rica. Por ejemplo, barrios con procesos de exclusión social (pobreza, desempleo, crimen, etc.) pueden situarse en dos distritos contiguos. Si estas variables no son incluidas en la especificación del modelo, se esperaría que los errores de estos distritos contiguos estén correlacionados.

La inquietud de analizar el grado de asociación de diversas variables de interés para el programa social *Avancemos* surge de estudiar su comportamiento, durante el periodo 2007-2011 para ubicar esa relación a nivel distrital, utilizando los modelos SAR que son una opción importante para analizar datos espaciales o geográficos. Con el programa social *Avancemos*, el Gobierno busca brindar ayuda monetaria a aquellas familias que tengan un nivel socioeconómico bajo cercano a

la pobreza para que mantengan a sus hijos dentro del sistema educativo, de manera que, logre disminuir la deserción en el colegio (Mivah, 2007).

Este tipo de investigaciones son de interés para el país, ya que permiten el mejoramiento en la aplicación de programas sociales para beneficiar a aquellos sectores de la población que presentan condiciones de desventaja a nivel general en temas relacionados con la educación, el empleo, la economía, entre otros. Esta contribución académica sirve de insumo para fortalecer aspectos políticos relevantes como, por ejemplo, los diversos sectores que están incluidos en los planes de desarrollo, teniendo influencia, inclusive, en las metas estratégicas que puede plantear el Gobierno en distintas áreas de análisis.

Resulta necesario describir y analizar, las asociaciones existentes para el programa social *Avancemos* con sus variables dependientes (porcentaje de deserción en los colegios) e independientes (tasa de la cantidad de becas *Avancemos*, monto promedio de las becas *Avancemos*, consumo promedio de electricidad, población en edad colegial de 12 a 21 años, cantidad de colegios y tasa total de profesores), para identificar a nivel distrital aquellas zonas geográficas donde el programa social *Avancemos* ha contribuido de forma sustantiva en el territorio nacional durante el periodo 2007-2011.

2.3 Objetivo general

Establecer las propiedades estadísticas del modelo de Autorregresión Simultánea (SAR), cuando se aplica a los datos distritales de Costa Rica.

Se entenderá como propiedad estadística el grado de robustez de las estimaciones a la violación de supuestos seleccionados del modelo SAR.

2.4 Objetivos específicos

- Evaluar la sensibilidad de los supuestos de autorregresión espacial en análisis de variables referentes a distritos costarricenses, mediante simulaciones.
- Identificar las debilidades o fortalezas que tiene el modelo de autorregresión simultánea cuando se aplica a datos distritales de Costa Rica.
- Analizar la asociación entre medidas de implementación del programa social *Avancemos* y medidas de logro de ese programa, utilizando modelos de autorregresión espacial a nivel distrital.

3. Estado de la cuestión

3. 1 Introducción al análisis espacial

En el análisis espacial, los datos espaciales se refieren a las observaciones de una o más variables que están relacionadas con una zona o área geográfica (Acevedo y Velásquez, 2008). Es decir, para cada uno de los territorios geográficos existen diversos datos compuestos por las variables de interés en un momento determinado. Además, se conoce que los usuarios de información geográfica prefieren contar con estadísticas desagregadas a niveles menores, por ejemplo regiones, cantones, distritos y áreas geográficas más pequeñas, ya que, las utilizan comúnmente (INEC, 2006).

El análisis espacial de esta investigación se enfoca específicamente en áreas o zonas geográficas correspondientes a los distritos existentes del 2007 al 2011. Es así como los datos recolectados y analizados intentan visualizar la realidad espacial de manera conjunta, utilizando áreas o zonas geográficas.

En nuestro país existen diversas investigaciones que muestran datos a nivel geográfico que son importantes para la toma de decisiones y que sirven de insumo para visualizar la realidad nacional en un momento dado así como el nivel de desarrollo con desagregaciones regionales, cantonales y distritales. Es importante contar con este tipo de herramientas que ayudan a la generación de programas y proyectos para mejorar la situación de las zonas geográficas que se encuentran rezagadas en su nivel de desarrollo local.

Dentro de las investigaciones nacionales que utilizan datos geográficos se encuentran el Índice de Desarrollo Social que elabora el Mideplan, el Atlas de Desarrollo Humano Cantonal de Costa Rica calculado por la Universidad de Costa Rica (UCR) con el apoyo del Programa de las Naciones Unidas para el Desarrollo en Costa Rica (PNUD-Costa Rica), el Índice de Necesidades Básicas Insatisfechas (NBI) que se construye por parte del Instituto Nacional de Estadística y Censos, entre otros.

Con este tipo de enfoques se confirma la disparidad existente en las diversas regiones que componen el territorio nacional y permite ubicar aquellas zonas geográficas que están en una condición desventajosa, y así direccionar planes, programas y proyectos específicos que ayuden a disminuir la brecha existente en esos territorios.

3. 2 Modelos de autorregresión espacial

Un modelo lineal sin transformación de la variable dependiente se puede plantear como (Cliff y Ord, 1972; Anselin, 1988; Anselin 2003; Dormann *et al.*, 2007).

$$Y = X\beta + \varepsilon \quad (1)$$

Donde:

Y : es un vector de dimensión $n \times 1$ que representa la variable dependiente en el que hay un valor por cada una de las n unidades geográficas.

X : es una matriz de dimensión $n \times (k+1)$ que contiene la información de las k variables independientes junto con un vector columna de unos.

β : es un vector de dimensión $(k+1) \times 1$.

ε : es un vector de dimensión $n \times 1$.

En un contexto de análisis espacial, estos errores ε pueden no ser independientes entre sí violando uno de los supuestos de los modelos lineales. Esta falta de independencia se denomina autocorrelación espacial. En modelos espaciales, la autocorrelación es deseable porque se puede modelar.

La autocorrelación espacial es la correlación entre los valores de una sola variable aplicable a sus posiciones, localizadas relativamente cerca sobre una superficie de dos dimensiones, incumpliendo con el supuesto de independencia de las observaciones (Griffith, 2005).

La autocorrelación espacial según Ding y Fotheringham (1992) se refiere a la forma en que los objetos o actividades son similares a otros objetos o actividades localizados en sus cercanías. La autocorrelación espacial trata, simultáneamente, con la ubicación y la información de los atributos. Si los objetos que son similares en ubicación también tienden a ser similares en atributos, entonces, estamos en presencia de autocorrelación espacial positiva. En cambio, si los objetos que están muy juntos en el espacio tienden a ser más disímiles en los atributos que en los objetos que están más separados, entonces, estamos en presencia de autocorrelación negativa. Y, por último, no existe autocorrelación cuando la distribución de los atributos es independiente de la distribución de las ubicaciones.

Ante la presencia de la autocorrelación espacial, los errores del modelo no afectarán al insesgamiento de los estimadores mínimos cuadrados ordinarios, pero sí podrían afectar a su variabilidad, ocasionando como consecuencia

pruebas de significancia potencialmente engañosas (Fotheringham y Rogerson, 1993).

Para mostrar el efecto de la autocorrelación espacial, se puede plantear un modelo que tome en cuenta la asociación de la variable dependiente entre vecinos. Uno de esos modelos es el autorregresivo simultáneo.

Al respecto, Anselin (2003) indica que en un modelo autorregresivo, el vector que contiene los errores es:

$$\varepsilon = \lambda W\varepsilon + u \quad (2)$$

Donde:

ε : es un vector de dimensión $n \times 1$.

λ : es el parámetro espacial autorregresivo.

W : es una matriz de dimensión $n \times n$ que contiene pesos que se refieren a la relación entre la unidad geográfica y la otra.

u : es un vector de dimensión $n \times 1$ con errores independientes e idénticamente distribuidos con varianza σ^2 .

Si se resuelve la ecuación anterior (2), se obtiene (Anselin, 2001)

$$\begin{aligned} \varepsilon - \lambda W\varepsilon &= u \\ \varepsilon [I - \lambda W] &= u \\ \varepsilon &= [I - \lambda W]^{-1} u \end{aligned} \quad (3)$$

Y esto implica que la esperanza matemática de la matriz de las sumas de los errores al cuadrado, que sería equivalente a la matriz de varianza-covarianza de los estimadores SAR (Anselin, 2001), es:

$$E[\varepsilon\varepsilon'] = \sigma^2[(I - \lambda W)^{-1}(I - \lambda W)^{-1}] \quad (4)$$

Sustituyendo (3) en (1), el modelo lineal autorregresivo simultáneo se expresa como (Cliff y Ord, 1975; Anselin, 2003; Dormann, 2007):

$$\begin{aligned} Y &= X\beta + [I - \lambda W]^{-1} u \\ [I - \lambda W]^{-1} u &= \lambda WY - \lambda WX\beta + u \\ Y &= X\beta - \lambda WX\beta + \lambda WY + u \end{aligned} \quad (5)$$

Donde:

$\lambda WX\beta$: son las variables independientes rezagadas espacialmente.

λWY : es una variable dependiente espacialmente rezagada que depende de sus unidades geográficas vecinas.

3. 2. 1 Matriz de pesos (W)

El análisis espacial considera estudiar, en la medida de lo posible, la asociación que tienen los valores de las unidades de estudio entre sí. Intenta comprobar si la correlación entre las áreas de estudio que tienen frontera en común es más intensa que las áreas de estudio que están más distanciadas, estudiando su posición geográfica.

Una vez definidas todas las unidades espaciales de estudio, es necesario seleccionar de manera conveniente la matriz de pesos o de proximidad conocida

como W , tomando en cuenta la naturaleza de los datos y que esta responda a las necesidades específicas de la investigación.

Existen diferentes tipos de estas matrices de pesos. Su escogencia depende de la estructura de la correlación de los datos y de las posibilidades del algoritmo de estimación para converger. Algunas de las matrices de pesos más utilizadas son las siguientes:

- **Matriz binaria de contigüidad**

En este caso, la matriz de pesos de contigüidad geográfica toma únicamente valores de 0 y 1, donde el 1 indica que las unidades espaciales de estudio son vecinas (comparten frontera) y 0 en caso contrario. Su diagonal está compuesta de ceros porque un área no puede ser vecina de sí misma. Está definida como simétrica (Aroca, 2000; Patuelli *et al.*, 2006; Borges, 2009). Existen, a su vez, varios subtipos:

- **Matriz *rook* (torre)**

Conocida como matriz torre, su nombre proviene del movimiento de la pieza de ajedrez, con ello, se identifican las regiones vecinas siempre y cuando compartan parte de una frontera en común de una forma relativamente extensa (Viton, 2010). Asigna el peso de 1 a todas las unidades espaciales de estudio ubicadas arriba, abajo, a la izquierda y a la derecha de un área determinada y 0 a las restantes (Krämer, 2003).

- **Matriz *bishop* (alfil)**

Conocida como matriz alfil, siguiendo los movimientos de la pieza de ajedrez se identifican las regiones vecinas, cuya característica es que compartan un mismo vértice, indicando que la parte compartida entre las unidades espaciales de estudio puede ser muy pequeña (Viton, 2010).

- **Matriz *queen* (reina)**

La matriz reina asigna el valor de 1 a todas las unidades espaciales estudio vecinas de un área determinada, siempre y cuando compartan una frontera en común o al menos un vértice (Krämer, 2003; Furtado y van Oort, 2011).

- **Matriz de distancias**

Como su nombre lo indica, está compuesta por las distancias entre los centroides de las unidades espaciales de estudio, o por una función de las mismas. Algunos ejemplos frecuentemente utilizados son el exponencial negativo de las distancias y el inverso de las distancias, entre otros (Cliff y Ord, 1981).

- **Matriz K vecinos más cercanos**

Este tipo de matriz considera a los vecinos que estén más cercanos a las unidades espaciales de estudio. La cantidad de vecinos k es usualmente definida de forma arbitraria (Furtado y van Oort, 2011).

- **Matriz K vecinos a partir de una distancia máxima (punto de corte)**

Los puntos de corte a cierta distancia que contiene esta matriz son definidos por el investigador y de esta forma son identificados los vecinos de las unidades espaciales de estudio que se encuentren a una distancia menor o igual a este punto de corte. Los puntos de corte pueden incluir de forma inadecuada a una unidad espacial que esté separada por un límite natural (carreteras, ríos, montañas) como vecinos de otra unidad espacial (Furtado y van Oort, 2011).

3. 2. 2 Fórmula general de la matriz de pesos W

Una fórmula que generaliza la definición de todos los tipos de matrices antes señalados es (Patuelli *et al.*, 2006)

$$W_{[q]} = \frac{n}{\sum_{i=1}^n d_i^{q+1}} * D^q * B \quad (6)$$

Donde:

B : es una especificación binaria de la matriz de pesos geográficos.

D^q : es una especificación basada en distancias de la matriz de pesos geográficos; es una matriz diagonal que contiene los d_i^q componentes (d_1^q, \dots, d_n^q) pertenecientes al vector $d = B * 1$, representando el grado de vínculo del objeto espacial i .

Y, siguiendo con Patuelli *et al.* (2006), en un modelo de autorregresión simultánea, el esquema utilizado es $q = -1$ y esto es equivalente a usar la matriz binaria B como la matriz W .

Por otra parte, en el análisis empírico se recomienda normalizar las filas de la matriz de pesos, por ejemplo de la siguiente forma (Lee, 1999)

$$w_{i,n} = \frac{(d_{i1}, d_{i2}, \dots, d_{in})}{\sum_{j=1}^n d_{ij}} \text{ donde } d_{ij} \geq 0 \text{ indica distancias positivas} \quad (7)$$

La estandarización asegura que la suma de las filas y columnas de la matriz de pesos siempre esté limitada en un valor absoluto, lo cual se requiere para estimabilidad.

3. 2. 3 Estimación de los parámetros del modelo SAR

El modelo de Autorregresión Simultánea (SAR) con errores ε_i gaussianos se estima por máxima verosimilitud usando algoritmos de mínimos cuadrados

generalizados. La función de logverosimilitud para el modelo SAR es entonces (Anselin, 2001; Lauridsen y Kosfeld, 2004; Giraldo, 2011; Lee, 1999)

$$\ln L = -\left(\frac{n}{2}\right) \ln(2\pi) - \left(\frac{n}{2}\right) \ln \sigma^2 + \ln |I - \lambda W| - \left(\frac{1}{2} \sigma^2\right) (Y - X\beta)'(I - \lambda W)'(I - \lambda W)(Y - X\beta) \quad (8)$$

En este caso, la logverosimilitud del modelo se deriva de suponer que el vector con los errores ε_i se distribuyen como una normal multivariada con media 0 y varianza-covarianza $\sigma^2 = [(I - \lambda W)'(I - \lambda W)]^{-1}$.

Además, utilizando el método de mínimos cuadrados generalizados, la estimación del coeficiente $\hat{\beta}$ de la máxima verosimilitud es

$$\hat{\beta} = [(X - \lambda WX)'(X - \lambda WX)]^{-1}(X - \lambda WX)'(Y - \lambda WY) \quad (9)$$

Y, el estimador de la máxima verosimilitud de σ^2 es

$$\hat{\sigma}^2 = \frac{(\varepsilon - \lambda W\varepsilon)'(\varepsilon - \lambda W\varepsilon)}{n} \quad (10)$$

Cabe señalar que la matriz de varianza-covarianza de ε es (Kelejian y Prucha, 1997).

$$VarCov(\varepsilon) = (I - \lambda W)^{-1}(I - \lambda W')^{-1} \quad (11)$$

3. 2. 4 Tipos de datos espaciales

En el análisis espacial existen tres tipos de datos espaciales, los cuales son: datos en rejilla o área (*lattice data* o "*raster*"), datos de patrones de puntos (*point patterns data*) y datos geoestadísticos (*geostatistical data* o "*datos vectoriales*").

Primeramente, los datos en rejilla o área son observaciones ubicadas en regiones espaciales que presentan una estructura de vecindad. La rejilla o área está delimitada por un conjunto de lados y vértices que contienen al centroide de cada una de esas regiones espaciales. Segundo, los datos de patrones de puntos ocurren cuando la variable de interés a analizar es la localización de cada evento específico. Es importante estudiar la variación de los eventos en la región espacial para identificar los patrones de dicha variación. Y, por último, los datos geoestadísticos son valores obtenidos en un punto fijo que buscan ampliar la distribución espacial de esos valores en toda la región de estudio. Frecuentemente son utilizados en temas relacionados con la geología (Cressie, 1993).

Los datos espaciales a nivel de unidad político-administrativa son datos en rejilla o área. Por consiguiente, en esta investigación, como se tienen datos agregados de las distintas variables de interés por distrito, entonces son denominados como datos en rejilla o área.

3. 2. 5 Supuestos de los Modelos de Autorregresión Simultánea (SAR)

Los supuestos de los modelos SAR son los siguientes:

Supuesto 1: Los errores u_i son independientes e idénticamente distribuidos con media 0 y varianza σ^2 (Kelejian y Prucha, 1997; Lee, 1999).

Supuesto 2: Los elementos de la matriz de pesos W son constantes conocidas y el rango de $(I - \lambda W) = n$ para todo $|\lambda| < 1$ (Kelejian y Prucha, 1997; Lee, 1999).

Supuesto 3: Las sumas de la fila y la columna de W y de $(I - \lambda W)^{-1}(I - \lambda W')^{-1}$ están limitadas uniformemente en valor absoluto (Kelejian y Prucha, 1997; Lee, 1999). En otras palabras, la matriz de pesos no incluye valores infinitos. Una forma de ajustar los datos al supuesto es si se normaliza la matriz de pesos, logrando que el número máximo de no ceros siempre se mantenga entre ciertos rangos.

También es importante destacar que la suma de todas las correlaciones de los errores siempre va estar por debajo de una constante, es decir,

$$n^{-1} \sum_{i=1}^n \sum_{j=1}^n |corr(\varepsilon_i, \varepsilon_j)| \leq c_\omega < \infty \text{ para todo } n \geq 1 \quad (12)$$

Supuesto 4: Los elementos de la matriz X son no estocásticos y la X tiene el rango por columnas completo (Kelejian y Prucha, 1997). Esto quiere decir que las variables independientes no son obtenidas completamente de forma aleatoria y no están expuestas a errores muestrales, por lo tanto, estas variables independientes son constantes conocidas. Además, este supuesto implica que el modelo no debe tener multicolinealidad perfecta.

Supuesto 5: Si el modelo SAR se estima con máxima verosimilitud, se supone que el término de los errores u se distribuye normalmente con media y varianza definidas en el Supuesto 1 (Kelejian y Prucha, 1997).

Supuesto 6: Una vez estandarizada la matriz de pesos, la suma de los valores D_{ij} por cada columna o por cada fila es 1 (Kelley y Barry, 1997).

La robustez del análisis a la violación de estos supuestos se va a analizar con simulaciones y ellas se encuentran descritas en el apartado metodológico de esta investigación.

3. 2. 6 ¿Cómo se diagnostica la autocorrelación espacial?

Como introducción al tema, es importante tener en cuenta que a pesar de que los errores u son independientes, los errores ε son dependientes y es aquí donde ocurre la autocorrelación espacial.

El problema de la autocorrelación espacial debe ser analizado mediante valores de probabilidad asociada. Por lo tanto, si el parámetro espacial autorregresivo λ es significativo ($p < 0,05$) implica que sí hay autocorrelación espacial en los residuos ε y en caso contrario, no existe autocorrelación espacial.

Luego, para verificar si el parámetro espacial autorregresivo λ capturó toda la variación espacial se analizan los residuos del modelo con la prueba I de Moran (Moran, 1948), en donde, la hipótesis nula señala que no existe autocorrelación global en los residuos y la hipótesis alternativa indica que sí existe autocorrelación global en los residuos.

En notación matricial, el estadístico I de Moran es (Anselin, 2001)

$$I = \left(\frac{n}{Z_0} \right) \left(\frac{\hat{\varepsilon}' W \hat{\varepsilon}}{\hat{\varepsilon}' \hat{\varepsilon}} \right) \quad (13)$$

Donde:

Z_0 : es el factor de estandarización que corresponde a la suma de los pesos de los productos cruzados que no son cero.

ε : es el vector de dimensión $n \times 1$ con los residuos señalados en (1).

W : es la matriz de dimensión $n \times n$ señalada en (2).

n : es la cantidad de observaciones.

Cuando la I de Moran es calculada, se mapea¹ los residuos del modelo SAR.

¹ En el apartado metodológico se describe el *software* utilizado.

3. 3 Estado de la cuestión sobre programas sociales y permanencia en el sistema educativo: aplicación del modelo SAR

Primeramente, es necesario aclarar que en esta investigación, la aplicación que va a utilizarse para el modelo SAR es sobre la asociación entre indicadores relacionados con el programa social *Avancemos* y medidas de logro de dicho programa, en específico, con el objetivo puntual de la permanencia en el sistema educativo por parte de los estudiantes de educación secundaria.

Por esta razón, se hace una presentación resumida sobre programas sociales de manera general, luego se hace una descripción del programa social *Avancemos* de forma específica ya que constituye la aplicación de los modelos SAR. Se mencionan además aspectos importantes relacionados con la permanencia de los estudiantes dentro del sistema educativo costarricense, específicamente, en la educación secundaria.

3. 3. 1 Implementación de programas sociales

En el transcurso de los años, Costa Rica ha implementado diversos programas sociales con la intención de ayudar a la población, específicamente aquella que tiene problemas económicos, educativos, sociales, laborales. Por ello, con la colaboración del Gobierno de Costa Rica, muchas familias logran tener una condición de vida mejorada gracias a los programas tales como las becas de *Avancemos* para que los niños y niñas puedan seguir asistiendo al colegio, donde existen diferenciales por distrito.

Entre los programas sociales que han sido ejecutados en nuestro país, Sánchez y Sauma (2011) mencionan las pensiones del régimen no contributivo de la Caja Costarricense del Seguro Social (CCSS) concretamente dirigidas a personas con 65 años y más que estén en condiciones de pobreza, y que no hayan cotizado para ninguno otro, o exista un faltante de cuotas por cumplir. Este programa fue creado en 1974, teniendo un impacto fuerte en las condiciones de vida de dichas personas con el paso de los años. Otro programa que ha dejado huella en nuestro país es el Bono Familiar de Vivienda, el cual otorga recursos económicos a las familias más necesitadas, para construir su propia casa. Además, los Centros de Educación y Nutrición y Centros Infantiles de Atención Integral (CEN-Cinai) que sirven como hogares de cuidado para los niños y niñas de aquellas familias que no pueden hacerlo durante todo el día por motivos laborales, además de estar en un lugar seguro, reciben educación y alimentación. También, las becas brindadas por el Fondo Nacional de Becas (Fonabe) que ayudan a las familias de escasos recursos para mantener a los miembros que están estudiando dentro del sistema educativo.

Los programas sociales no son ideas recientes, algunos fueron desarrollados desde la mitad del siglo pasado. El Estado costarricense adquirió nuevas funciones en los ámbitos económicos, sociales y ambientales, haciendo ideas sobre la ejecución de políticas públicas, principalmente, en el ámbito social que beneficiaron a los ciudadanos de esa época.

Al inicio de la década de los cuarenta las políticas públicas direccionadas al tema social eran más limitadas en la Costa Rica de esa época. Pero, por los problemas reflejados en las condiciones económicas y sociales, el país tuvo que analizar la situación particular de dicha coyuntura para mejorar el panorama general y considerar la importancia de la creación de políticas públicas enfocadas en la parte social y económica. Con el paso de los años, el modelo de desarrollo

costarricense adquirió fuerza con la industrialización sustitutiva de importaciones, repercutiendo de manera favorable en la economía nacional. Pero en la década de los setenta, empezó a notarse un agotamiento del modelo de sustitución de importaciones y para contrarrestar lo anterior, los gobiernos de la época recurrieron a la creación de instituciones y ministerios para fortalecer de cierta forma la generalización del desarrollo social del país brindando beneficios generales a toda la población (Barahona, 1999).

No obstante, este esfuerzo político no resistió a la crisis de la década de los ochenta, que golpeó fuertemente a nuestro país, ocasionando graves problemas en el modelo de sustitución de importaciones, impactando en la condición social de los costarricenses, deteriorando sus servicios, la calidad de vida y aumentando la desigualdad de ingresos. Posteriormente, después de la crisis de los ochentas, el país tuvo un cambio en las políticas sociales, implementando programas sociales selectivos con énfasis en la focalización, dirigidos a grupos con altas necesidades en temas económicos, de vivienda, educación y salud, principalmente, para luchar contra la pobreza para que exista una adecuada equidad social. Luego, a inicios de la década de los noventa, el país tuvo una reestructuración del mercado laboral para seguir estabilizando la economía nacional (Barahona, 1999).

Por otra parte, el diseño de las políticas públicas puede ser elaborado por instituciones que pertenezcan o no al Estado, pero su aplicación y ejecución debe ser comandada por parte del Gobierno. La idea o propósito general de la creación de una política pública se basa en buscar o mejorar las condiciones de los potenciales beneficiarios y de la ciudadanía en general. Ante esta situación, los programas públicos deben ser evaluados constantemente para garantizar resultados óptimos en términos de desarrollo y en los casos que sea necesario, se deben rediseñar y mejorar los programas que estén siendo ejecutados para la toma de decisiones (Galiani, 2006).

Existen diversas ideas del concepto de política pública que engloban aspectos relacionados con el motivo de su creación y ejecución que se basa, principalmente, en identificar objetivos considerados de importancia para la sociedad así como la resolución de problemas que son de beneficio público, los cuales están políticamente definidos como colectivos. En esa misma línea, las políticas públicas buscan ser una opción para solucionar los problemas sociales y colectivos que vive la ciudadanía, que son caracterizados desde el punto de vista político como públicos. Además, en una política pública pueden tener participación actores privados siempre y cuando exista una normativa que sustente dicha participación o, en su defecto, que un actor público le atribuya el derecho de participar en la política pública (Aguilar, 2009; Subirats, Knoepfel, Larruel y Varone, 2008).

Las políticas públicas son creadas para ayudar a las poblaciones que viven en condiciones de pobreza para mejorar su calidad de vida en términos de que sean menos vulnerables a diversas situaciones y a reducir la brecha de desigualdad social y económica existente, según sea la realidad de cada uno de los países. Estas brechas pueden ser vistas en función del ingreso laboral que perciben los ciudadanos, dependiendo de la ocupación en que se desempeñen. Como es lógico pensar, cuanto menos ingreso tenga una familia, menor es la probabilidad de salir adelante con los diversos gastos mensuales. En el contexto latinoamericano, los programas sociales presentaron su mayor auge en la década de los noventa. Por su parte, la implementación de los programas sociales no solamente sirve para mejorar el contexto nacional del país, sino que pueden ser utilizados como opción para soportar las crisis presentes en el entorno internacional, y así que estas no castiguen de una forma muy fuerte a los países más vulnerables. Un ejemplo de crisis internacional fue la acontecida a mediados de 2008 (Sánchez y Sauma, 2011).

Para profundizar con más detalle, el programa social *Avancemos* es caracterizado posteriormente en esta investigación.

3. 3. 2 Estudios estadísticos para analizar la implementación de programas sociales

Cuando un programa social es ejecutado, se espera que los resultados sean favorables en un futuro predefinido por el programa, logrando alcanzar los objetivos para los que fue creado. La idea anteriormente descrita está presente en la mente de las personas que hacen posible la generación de diversos programas sociales, ya que busca mejorarse las condiciones sociales de una manera positiva. Una forma de averiguar si el programa social está dando resultados positivos es mediante el análisis de datos. Incluir diversos indicadores que puedan contabilizarse en un momento y tiempo dado es de gran ayuda para analizar, entender y explicar el comportamiento en un periodo determinado.

El programa social *Avancemos* se considera como una política social selectiva que transfiere dinero a las familias de escasos recursos. Ha tenido un crecimiento importante, pasando de seis mil a 187 mil beneficiarios del 2006 al 2010, respectivamente. Además, ha contribuido a aumentar la presencia de adolescentes en las aulas cambiando el trabajo por estudio. Complementariamente a las ayudas monetarias, debe mejorarse la educación secundaria en las zonas rurales, aplicando la enseñanza del inglés, informática y cursos técnicos especializados para optar por tener empleos con mayores ingresos monetarios (Trejos, 2012a).

3. 3. 3 Programas sociales similares en América Latina

Por otra parte, no solamente existe el caso del programa *Avancemos* en Costa Rica, sino que a nivel latinoamericano, en varios países se han implementado programas sociales similares, los cuales buscan ayudar a aquellas familias de escasos recursos para que los hijos no abandonen los estudios y sigan en el sistema educativo para un mejor futuro a nivel familiar, económico, social e intelectual. Por lo tanto, es interesante conocer algunos ejemplos de programas parecidos al nuestro, para contextualizar las realidades presentes en los diversos países latinoamericanos que han utilizado este tipo de ayudas sociales, según sean sus casos particulares.

El auge de los programas sociales específicos de transferencias monetarias ocurrió desde mediados de la década de los noventa y se han ido extendiendo paulatinamente, ofreciendo ayuda a los hogares que tienen pocos ingresos, y que se encuentran en condiciones de pobreza, para que estas familias puedan cumplir con las metas direccionadas a la educación, salud y nutrición. Por ejemplo, en Brasil, se inició con el programa "*Bolsa scola*" y, posteriormente, fue modificado a "*Bolsa familia*" que estuvo direccionado a familias, en Ecuador el programa "*El Bono solidario*", en México el programa de desarrollo humano "*Oportunidades*" y en Nicaragua el programa "*Red de protección social*". Además, existen programas que en algunas ocasiones estaban dirigidos a niños y adolescentes, como los casos en Bolivia con el programa "*Bono Juancito Pinto*", en Colombia con el programa "*Familias en acción*" y en Guatemala con el programa "*Mi familia progresa*" (Sánchez y Sauma, 2011).

Estos países intentan disminuir la problemática con respecto a la pobreza en la que viven grandes cantidades de familias que se ven obligadas a quitar a sus hijos del sistema educativo para insertarlos en un mercado laboral que no es bien

remunerado por la naturaleza de los empleos que puede tener un adolescente sin formación académica completa.

Los beneficios que han recibido estos países latinoamericanos son muy importantes en cada uno de sus ámbitos, logrando aumentar los ingresos de las familias pobres beneficiadas, teniendo así una opción más para que sus hijos sigan preparándose académicamente, en busca de un mejor futuro profesional, obteniendo así una herramienta importante para salir de los niveles de pobreza de sus respectivas familias en un mediano y largo plazo. Estos esfuerzos gubernamentales son trascendentales para ayudar a la disminución de la deserción y repitencia, que son temas problemáticos presentes en los diversos contextos latinoamericanos.

3. 3. 4 Factores que inciden en la no permanencia en el sistema educativo

A través de los años, diversos análisis se han concentrado en estudiar aquellos fenómenos que tienen repercusión en la no asistencia al sistema educativo costarricense. Las siguientes investigaciones intentan demostrar los factores anidados a esta problemática nacional.

A inicios de la década de los ochenta, en el país existía una crisis socioeconómica que afectó el sistema educativo. En su momento, Sánchez *et al.* (1984) encontraron que la crisis económica de esa época fue la causa principal de la deserción en el tercer ciclo y educación diversificada. También, se observó que la deserción en colegios académicos fue de 33% en el área urbana y de 41% en el área rural. Además, comentan que las causas de la deserción pueden estar asociadas con variables: económicas de la familia que los lleva en ciertos casos a

dejar los estudios para ir a trabajar, sociales, demográficas ya que las edades donde había más desertores son los 13, 15 y 16 años; educativas como por ejemplo la carencia de libros para el estudio, problemas en la metodología de enseñanza empleada por los profesores, entre otras. Por su parte, los porcentajes promedio de deserción presentan valores más altos en el inicio de los ciclos, es decir en sétimo y décimo año, en colegios académicos, diurnos y nocturnos, siendo mayor para estos últimos. Se registra durante el periodo comprendido entre los años 1970 y 1977, para el tercer ciclo y educación diversificada, un promedio de deserción de 10,61% en el sector oficial y de 6,69% en el privado.

De una forma más específica, los desertores rurales estuvieron afectados por problemas económicos familiares que los obligó a trabajar obteniendo el permiso por parte de los padres (esto ocurrió en la mitad de los casos encontrados), así como, el bajo rendimiento académico, falta de interés en el estudio, carencia de útiles y materiales, metodología utilizada por los profesores. Por otra parte, los desertores urbanos tuvieron los siguientes inconvenientes: abandonaron sus estudios mayoritariamente en los primeros años del tercer ciclo (sétimo, octavo y noveno año), los problemas económicos propios, la falta de interés en el estudio, la carencia de amigos en las aulas, carencia de útiles y materiales, bajo rendimiento académico, casi la mitad de los desertores tenían interés en seguir estudiando y la familia no quería que abandonaran los estudios en casi la mitad de los casos. Y, de manera general, el bajo ingreso de la familia ocasiona problemas importantes en la situación económica del alumno desertor (Sánchez *et al.*, 1984).

El efecto de este fenómeno está relacionado, entre otras cosas, con la ubicación geográfica de las áreas o regiones (Collado, 2004). Al respecto, la autora señala que existe un patrón geográfico en la deserción escolar encontrando conglomerados de alta y baja no asistencia escolar en Río Azul-Patarrá, Barrios del Sur y Piedades con parte de la GAM y La Aurora,

respectivamente. De esta manera, se confirma el supuesto de existencia de autocorrelación espacial, relacionado con el programa social *Avancemos* y la no permanencia en el sistema educativo por parte de los alumnos.

En esa misma línea, según investigaciones realizadas por las direcciones regionales del MEP, tanto Desamparados como Limón y Puntarenas muestran mayores cifras en deserción y repitencia en la educación secundaria (Programa Estado de la Nación, 2003). También, se señala que la deserción en el tercer ciclo y educación diversificada afecta de una forma más directa a San José, Alajuela, Heredia, Cartago, Limón y Desamparados, teniendo estos dos últimos cifras de deserción mayores al promedio nacional (Gonzaga *et al.*, 2008).

También, se ha encontrado que los adolescentes del 25% de hogares ubicados en la zona urbana con menores ingresos contabilizan tasas de abandono escolar que son tres veces mayores a la de los adolescentes del 25% de hogares de ingresos más altos (Preal, 2003).

La deserción en el III y IV ciclo se produce por factores del estudiante en temas económicos, familiares, sociales, entre otros, y del sistema educativo. Según estudios realizados por el MEP, casi la mitad de los desertores están trabajando en labores que generan muy pocos ingresos, con una estabilidad muy reducida, cuyos padres poseen un nivel de escolaridad bajo. Desde el punto de vista del sistema educativo se indica que existe poca capacitación de los docentes, falta de recursos didácticos en el aula, poca asesoría a los estudiantes, no existe un claro conocimiento de las técnicas de estudio que pueden ser empleadas para ayudar a los estudiantes, entre otros (Programa Estado de la Nación, 2003).

Los estudios de Vargas (1987) y Brenes (1995) indican que los factores asociados a la deserción están relacionados con las condiciones de la vivienda

donde habita el alumno, condiciones de transporte, el ingreso familiar, acceso a servicios básicos como la electricidad, salud, agua; el rendimiento académico y aquellos factores que repercuten en el ambiente pedagógico del aula en el centro educativo; además, indica que los desertores provienen de hogares con escasos recursos económicos y en condiciones de pobreza, existe una correlación entre el acceso a servicios básicos y la deserción, los padres tienen un nivel educativo bajo ocasionando una menor calidad de vida familiar (Vargas, 1987). Además, resultados de diversas investigaciones manifiestan que la deserción escolar ocurre por causas económicas, en donde esos estudiantes pertenecen a familias de bajo estrato educativo y económico, ocasionando que los padres de familia no puedan mantener a sus hijos en el sistema educativo y algunos de ellos decidan que sus hijos trabajen para tener otra fuente de ingreso (Brenes, 1995).

Otra de las investigaciones (Hernández y Barrantes, 2002) plantea que los determinantes que originan el abandono del sistema educativo tienen que ver con factores del estudiante: relacionados con la falta de interés por el estudio, tensión a los exámenes, bajo rendimiento académico. Factores familiares: los estudiantes no tienen el apoyo familiar, no reciben ayuda, el ambiente del hogar no permite concentración. Factores del sistema educativo: las asignaturas son difíciles, gran cantidad de exámenes, la metodología utilizada por los docentes no es la adecuada.

Además, se ha reconocido que algunas variables que tienen su influencia en la deserción en secundaria están relacionadas con la incorporación de los adolescentes al trabajo asalariado que generalmente es poco remunerado debido a la falta de estudios, la escolaridad de los familiares, apoyo por parte de los padres y problemas personales como drogadicción, embarazos no deseados, estrés de la familia, condiciones de vida desfavorables, entre otros (Ruiz, 2006).

En combinación a los factores anteriormente citados, existe una época de transición que experimenta el niño al pasar de la escuela al colegio. Para algunos estudiantes, este cambio es difícil de asimilar debido a las diferencias entre ambos sistemas educativos, así como el desarrollo propiamente del niño que lo llena de ciertas incertidumbres sobre su personalidad, expectativas de vida, aceptación por parte de la sociedad, entre otros, los cuales debe manejar en ese momento. El hecho de aprobar el séptimo año ayuda a que no llegue al fracaso para el estudiante de forma inmediata, labor que debe ir de la mano con ayuda de los docentes y así evitar la exclusión del sistema educativo (Alfaro *et al.*, 2008).

Cuando se abandona el sistema educativo, existen diversas repercusiones que afectan al desarrollo del adolescente en el corto, mediano y largo plazo. Se reducen las oportunidades laborales, existe un mayor riesgo de vivir en la pobreza, la calidad de vida puede ser limitada, entre otras. También, existen factores anidados al abandono escolar que están relacionados desde el punto de vista sociocultural: familia, nivel educativo de los padres y aspiraciones ocupacionales; pedagógico: bajo rendimiento académico, metodología de enseñanza, falta de interés por el estudiante, etc.; socioeconómico y geográfico: familias con ingresos insuficientes que dificultan completar los estudios de sus hijos (Arguedas y Jiménez, 2008).

Es importante mencionar que en una de las investigaciones realizadas a partir del diagnóstico del Programa Internacional para la Evaluación de Estudiantes (PISA), en las áreas de Matemática, Ciencias y Lectura, presente en el cuarto informe del Estado de la Educación, revela que los elementos que inciden en el aumento o disminución del rendimiento académico están asociados a factores socioeconómicos de la familia, las actitudes hacia la lectura, las técnicas de estudio aplicadas por el estudiante para la comprensión y habilidades para resumir textos. También, contribuyen elementos relacionados con las características del centro educativo, así como el nivel de desarrollo

socioeconómico de los distritos². Según el análisis efectuado por el Programa Estado de la Nación (2013), si se logra mejorar las estrategias de lectura en los colegios públicos, la brecha que existe entre estudiantes de colegios públicos y privados sería más pequeña en el desempeño en las pruebas PISA. Dicha disminución sería superior del 67% en competencia lectora y mayor al 70% en competencia matemática (Programa Estado de la Nación, 2013).

Del mismo modo, otra investigación elaborada a partir del diagnóstico de las pruebas PISA, indica que la brecha entre estudiantes de colegios públicos y privados es amplia. Por ejemplo, un 26% de los estudiantes evaluados pertenecientes a colegios públicos han repetido al menos un año en secundaria, en contraste, apenas un 10% de los estudiantes de colegios privados han repetido un año educativo. Esta disparidad en el rendimiento de los estudiantes se debe a diferencias en los factores familiares y personales que aquejan al alumno. Sin duda, reducir la repitencia sería muy importante para mejorar la calidad del sistema educativo de Costa Rica. Además, la educación impartida en colegios privados en comparación con los colegios públicos no incide fuertemente en la brecha del rendimiento existente entre los estudiantes de colegios públicos y privados. Por lo tanto, las diferencias en las condiciones socioeconómicas pesan mucho más que las diferencias en los recursos escolares, esto es aplicable cuando se intenta explicar la desigualdad existente en el rendimiento educativo (Fernández y Del Valle, 2013).

² El Índice de Desarrollo Social (IDS) mide el nivel de desarrollo social a nivel distrital.

3. 3. 5 Avancemos

En el caso de Costa Rica, el programa social *Avancemos* es un impulso brindado por el Gobierno para ayudar a sobrellevar la carga monetaria que involucra la permanencia de hijos e hijas en la educación secundaria. Está dirigido especialmente a aquellas familias en condición de pobreza y vulnerabilidad, que tengan hijos con edades de los 12 a los 21 años que estén matriculados y cursando la educación secundaria, pero si existe un caso especial de un adolescente que sobrepase este intervalo de edad, puede ser considerado como beneficiario si la condición lo amerita. Este esfuerzo pretende mantener a los estudiantes dentro de la educación secundaria disminuyendo así la salida de estudiantes del sistema educativo, problema que es recurrente, en nuestro país, en la última década (Mivah, 2007).

Este programa social fue creado el 19 de mayo de 2006 y ejecutado en el Gobierno de Oscar Arias Sánchez 2006-2010, según lo cita el Decreto Ejecutivo N°. 33154-MP-Mideplan-MEP-Mtssmivah (Mivah, 2007). Durante dicha Administración, el Ministro de Educación Pública fue Leonardo Garnier Rímolo y continuó en el mismo cargo durante el Gobierno de Laura Chinchilla Miranda 2010-2014, teniendo una participación activa como jerarca del MEP desde el inicio de *Avancemos* hasta la conclusión de la Administración Chinchilla Miranda.

El objetivo general de *Avancemos* consta de motivar a los adolescentes que están en condiciones de pobreza por continuar en el sistema educativo. Los objetivos específicos están relacionados con dotar a las familias de escasos recursos de un ingreso adicional para la permanencia de los adolescentes dentro del sistema educativo, mejorar el futuro de dichas familias mediante la preparación académica de sus hijos a mediano plazo, contribuir en la reducción

de las tasas de deserción y repitencia en la educación secundaria y así evitar, en alguna medida, el trabajo infantil y adolescente (Mivah, 2007).

Desde finales de la década de los noventa hasta la aparición del programa social *Avancemos*, existieron dos programas sociales que brindaban transferencias condicionadas, los cuales son: *Superémonos*, que estuvo vigente de 1999 hasta el 2002 y el programa *Acceso a la Educación*, presente del 2002 al 2006. Pero, el programa social *Avancemos* es diferente a estos dos anteriores en términos de una gran cobertura y de mayores transferencias por estudiante (Céspedes y Jiménez, 2010).

Políticas públicas, como el programa *Avancemos*, ayudan a disminuir la población adolescente que puede ser excluida dentro del tercer ciclo y educación diversificada. Por ejemplo, en el tercer informe del Estado de la Educación se destaca que en el 2009, el porcentaje de asistencia a la educación tradicional en la población de 13 a 17 años de edad, tomando en cuenta los hogares representaba, en su momento, un 81,9% (Programa Estado de la Nación, 2011).

En sus inicios, el programa social *Avancemos* contaba con la participación de Fonabe y del IMAS, contabilizando un total de 41 691 beneficiarios durante el 2007, dato estimado de la Encuesta de Hogares y Propósitos Múltiples del mismo año (Trejos, 2008). No obstante, a partir del 2009, solamente el IMAS se encarga de otorgar las becas a los beneficiarios (Trejos, 2012a).

Con respecto al trámite necesario para optar por la beca, una vez que la solicitud es aprobada, cada familia beneficiaria deberá firmar un documento, haciéndose responsable de que el adolescente seguirá asistiendo a la secundaria y deberá colaborar activamente para que apruebe el curso lectivo. Luego, la transferencia es recibida por las familias de forma mensual, hasta que finalice el estudiante o se mantenga en el sistema educativo.

Diversas instituciones nacionales han trabajado para que el programa social *Avancemos* continúe brindando ayudas a las familias que así lo requieran; entre ellas están el Ministerio de Educación Pública (MEP), el Ministerio de Vivienda y Asentamientos Humanos (Mivah), el Ministerio de Trabajo y Seguridad Social (MTSS) representado por la Dirección de Asignaciones Familiares (Desaf), la Caja Costarricense del Seguro Social (CCSS), el Instituto Mixto de Ayuda Social (IMAS), el Instituto Nacional de Aprendizaje (INA) y el Fondo Nacional de Becas (Fonabe) (Mivah, 2007).

3. 3. 6 El programa social *Avancemos* en cifras

Existen diversas investigaciones y estadísticas sobre la importancia del programa social *Avancemos* desde su implementación. Este programa es considerado como una de las fortalezas que ha ayudado, de cierta forma, a mantener a los estudiantes dentro del sistema educativo de la secundaria costarricense.

Un estudio realizado por Arce señala que el programa social *Avancemos* "...aumenta la probabilidad de asistencia de los jóvenes a los colegios, lo cual es un indicador de su efectividad para aumentar la cobertura³. Sin embargo, el citado autor concluye que el programa no aumenta la probabilidad de aprobación de los cursos" (Céspedes y Jiménez, 2010, pp. 17-18).

³ Gilberto Arce. Políticas educativas y pobreza en Costa Rica. En las políticas de combate a la pobreza en Costa Rica. Editores Víctor Hugo Céspedes y Ronulfo Jiménez. Academia de Centroamérica, por publicar.

Según los datos recopilados, en el 2007, casi ocho de cada diez personas con edades entre 13 y 18 años tuvieron la oportunidad de ir a un centro educativo. Y si el dato anterior se divide en población con pobreza extrema y pobre, el resultado de asistencia a un centro educativo es de 71 y 77%, respectivamente. La asistencia ha aumentado, ya que el 37% de la población entre 17 y 18 años ubicada en el quintil de hogares con menor ingreso tenía 9 años de educación aprobados y solamente 1 de cada 5 jóvenes con edades entre los 20 y 21 años se habían graduado como Bachilleres en Educación Media (Céspedes y Jiménez, 2010).

Otra de las investigaciones analiza los resultados del programa *Avancemos*, utilizando la Encuesta de Hogares y Propósitos Múltiples (EHPM) y la Encuesta Nacional de Hogares (Enaho), entre el 2007 y 2011. El inconveniente de esta encuesta es que imposibilita separar las ayudas económicas de *Avancemos* y las becas tradicionales, pero para efectos generales es de cierta forma un acercamiento. Además, el hecho de descentralizar las becas *Avancemos* completamente en el IMAS a partir del 2009, produce que la ayuda de Fonabe sea mucho mayor en el nivel educativo de primaria (Trejos, 2009). Durante el periodo de estudio comprendido entre el 2007 y 2011, el porcentaje de beneficiarios del programa *Avancemos* con respecto a otras instituciones ejecutoras, es del 45,6 y 46,4%, respectivamente, manteniendo un comportamiento similar en el periodo, con excepción del 34,1% obtenido en el 2008. Los beneficiarios de ayudas económicas de *Avancemos* y las becas tradicionales están ubicados, en su mayoría, en secundaria, clasificados como grupos pobres y vulnerables, repercutiendo en el aumento entre los asistentes a la secundaria, pasando de 17,2% en el 2007 a 39,4% en el 2011 (Trejos, 2012b).

En el Plan Nacional de Desarrollo 2006-2010, el programa social *Avancemos* tuvo su espacio dedicado a la atención aproximada de 130 000 estudiantes beneficiarios. Aunado a lo anterior, se estableció hacer un esfuerzo por reducir la

desigualdad existente en la educación, otorgando, de manera general, cerca de 24 000 becas, 17 000 beneficiarios de servicio de transporte gratis y 140 000 beneficiarios de los servicios que brinda el programa de comedores escolares (Mideplan, 2007).

Posteriormente, con el Plan Nacional de Desarrollo 2011-2014, en el tema de bienestar social, se menciona que el programa social *Avancemos* recibirá cooperación del MEP, IMAS y Fonabe; logrando un total aproximado de 160 000 beneficiarios, prometiendo una mejora en la gestión administrativa de dicho programa. Para lograr esa cifra, la estimación presupuestaria es de 240 000 millones de colones (Mideplan, 2010).

Dentro de los aspectos positivos del programa *Avancemos*, se comenta que los orientadores de los centros educativos están de acuerdo con este tipo de ayudas ya que ellos están en un contacto más cercano con los estudiantes más necesitados (Jiménez y Gaete, 2010).

No obstante, existen algunas críticas sobre el programa *Avancemos* que apuntan hacia un trámite largo para el otorgamiento de las becas así como una mala asignación de las mismas que ocurre en algunos casos debido a que las familias no las necesitaba. Además, se cuestiona el uso que dan tanto las familias como los estudiantes beneficiados ya que el único requisito para optar por una beca es estar matriculados en el colegio, dejando de lado el rendimiento académico y el compromiso tanto del estudiante como de la familia involucrada, haciendo que exista un interés para que el estudiante esté matriculado y así seguir recibiendo la ayuda y no se preocupan para que cumpla con sus deberes estudiantiles. También, existen ejemplos que manifiestan su opinión en el sentido que tanto la ayuda de *Avancemos* como el comedor escolar y el transporte para algunas instituciones son insuficientes para que los estudiantes no abandonen sus estudios (Jiménez y Gaete, 2010).

Según un estudio realizado por la Contraloría General de la República (CGR), se aplicaron cuestionarios a 2 687 estudiantes y 762 docentes de 84 colegios académicos diurnos y nocturnos para conocer la opinión de los informantes sobre temas relacionados con las estrategias implementadas por el MEP para mejorar la cobertura en educación secundaria y entre ellas, el programa social *Avancemos* tuvo su espacio de análisis.

Los resultados obtenidos de dicha investigación señalan que el programa social *Avancemos*, según la opinión de los entrevistados, es visto como más necesario de implementar en los colegios nocturnos y en los colegios de zona rural. Además, el 84% de los estudiantes entrevistados manifestó que la beca ayuda mucho o demasiado en la permanencia dentro del sistema educativo. También, el 53,9% de los estudiantes entrevistados opinaron que la influencia de la beca es demasiada para los colegios nocturnos y un 37,6% expresó lo mismo para los colegios diurnos. Por último, 4 de cada 10 estudiantes entrevistados manifestó que pocas veces la beca se otorga a los jóvenes que verdaderamente la necesitan. Con respecto a la asignación de las becas, un 93,3% de los docentes entrevistados manifiestan que existen estudiantes que necesitan la beca y no la tienen. De la misma forma, en este estudio se rescata la percepción por parte de los entrevistados sobre la cantidad de recursos económicos de las becas que se asignan a jóvenes que no las necesitan y el dinero lo gastan en otras cosas y no en el estudio propiamente, como por ejemplo, de ayuda con los gastos del hogar, por lo que se exigen mecanismos de control y de seguimiento del programa que asegure el buen destino de las becas, como por ejemplo, que exista una participación más directa tanto de los directores como de los docentes de los colegios ya que conocen de una forma más cercana la realidad de sus estudiantes (CGR, 2011).

El programa social *Avancemos* es uno de los temas que se encuentra en las investigaciones que realiza el Estado de la Educación. Si bien es cierto, en el primer informe no estaba en ejecución dicho programa social, se mencionaba que existía una expulsión de adolescentes del sistema educativo debido a la falta de recursos económicos. De ahí la necesidad de contar con una ayuda económica para aquellas familias necesitadas y así, los estudiantes pudieran mantenerse en el colegio y terminar la educación secundaria. En el segundo informe del Estado de la Educación, se plantea que los montos de las transferencias son los siguientes: 15 000 colones para aquellos estudiantes que estén cursando el séptimo año, 20 000 colones en octavo, 25 000 colones en noveno, 35 000 colones en décimo, 45 000 en undécimo y 50 000 en duodécimo. El límite por familia es de 80 000 colones en el caso que más de un miembro del mismo hogar sea beneficiario del programa (Programa Estado de la Nación, 2008).

Se menciona que *Avancemos* es prioridad en aquellas zonas geográficas que posean mayores niveles de pobreza, siendo para el 2007, la provincia de Puntarenas la que tuvo un mayor porcentaje de beneficiarios con el 30%, seguida de Guanacaste con el 22,7%. El programa social *Avancemos* es financiado con recursos del Estado. Inicialmente para el 2006 se destinó un total de 3 459,7 millones de colones, para el 2007 un total de 21 241,2 millones de colones y para el siguiente año estaban presupuestados 39 157,6 millones de colones. Para el 2007, cinco de cada diez beneficiarios tenían una edad de 15 a 17 años, tres de cada diez eran menores de 15 años y un 17,1% se situaban entre 18 y 20 años (Programa Estado de la Nación, 2008).

En el tercer informe del Estado de la Educación, se destaca que aumentó la importancia relativa del programa social *Avancemos* dentro de la inversión social pública. Para el 2010, existen 185 229 beneficiarios de aproximadamente 138 415 familias que significa un total de 49 516 millones de colones. Además, para el 2009 se registra que casi la mitad de los beneficiarios provienen de hogares

pobres y casi un 15% son de hogares vulnerables, pero uno de cada tres beneficiarios pertenece a hogares situados en los deciles del cuarto al décimo; esto tomando en cuenta la distribución de los beneficiarios por deciles de ingreso per cápita. Además, se señala que no existe un seguimiento constante a los estudiantes que reciben el beneficio, detalle que debe ser manejado en el IMAS y MEP. También, se destaca que las becas de Fonabe y *Avancemos* están ubicadas en zonas rurales y cerca del 58% de los beneficiarios no es pobre (Programa Estado de la Nación, 2011).

Y, por último, en el cuarto informe del Estado de la Educación, se contabiliza para el 2011 un total de 185 315 beneficiarios que están ubicados aproximadamente en 139 665 familias, representando al 43,4% de la matrícula del MEP. Sigue manteniéndose un predominio de beneficiarios ubicados en las zonas más alejadas del centro del país. En el 2011, alrededor de la mitad de las becas iba dirigida al 20% de las familias más pobres y cerca del 80% estuvo ubicada en los dos primeros quintiles del ingreso familiar per cápita. En el mismo año, se entregaron un total de 11 708 becas a adolescentes que no estaban asociados a ningún centro educativo, representando el 6,3% del total de becas y a un 77% de ellos no se les siguió otorgando la transferencia monetaria debido a que no asistían al colegio. Deben realizarse estudios sobre las implicaciones de las becas en tasas de deserción para tener un control más preciso de los beneficiarios así como conocer las razones por las que 9 523 estudiantes abandonaron el sistema educativo y recibieron la beca en el 2011. Además, debe valorarse los montos de las transferencias ya que desde el 2007 permanecen iguales para cada uno de los niveles académicos (Programa Estado de la Nación, 2013).

4. Abordaje metodológico

La investigación propone la creación de un modelo de Autorregresión Simultánea (SAR), involucrando variables de interés desagregadas a nivel distrital, para el programa social *Avancemos*. Además, para medir la influencia de dicho programa social en los cambios de las variables, se utilizó el modelo SAR, en el cual, los coeficientes miden el cambio en un periodo determinado. Y, precisamente, el periodo de estudio para esta investigación inicia en el 2007 y finaliza en el 2011, debido a que durante estos años se logró obtener información al respecto a nivel distrital.

Tal como se ha referido en el Estado de la Cuestión, el uso de los modelos de autorregresión simultánea radica en la presencia de un componente espacial en la variable dependiente que puede estar siendo explicado por alguna variable independiente que implícitamente esté reflejando el patrón espacial contenido en la variable dependiente. En el caso de que aun existan residuos autocorrelacionados después de intentar resolver el problema con las variables que fueron utilizadas en la regresión espacial, es necesario incluir al modelo otras variables independientes que ayuden a explicar el componente espacial de la variable dependiente. Si el factor geográfico no puede ser medido, entonces, la ventaja de los modelos de autorregresión simultánea es que ayuda a controlar dicha variación espacial, tomando en cuenta un parámetro para tratar de absorber la autocorrelación.

4. 1 Análisis de la robustez del modelo SAR a la violación de supuestos

4. 1. 1 Revisión de literatura sobre simulaciones aplicadas al modelo SAR

El análisis de las simulaciones ha estado involucrado en investigaciones científicas que buscan valorar la robustez de sus datos. Según Maisel y Gnugnoli, la definición del término simulación es la siguiente: *"...es una técnica numérica para realizar experimentos en una computadora digital. Estos experimentos involucran ciertos tipos de modelos matemáticos y lógicos que describen el comportamiento de sistemas de negocios, económicos, sociales, biológicos, físicos o químicos a través de largos periodos de tiempo..."* (Coss, 2003, p.12).

Uno de los modelos de simulación conocidos es el Monte Carlo, que consiste en definir un modelo que sea el objeto de la investigación y los eventos aleatorios son utilizados para estudiar dicho modelo. A menudo, en la aplicación de los métodos Monte Carlo, el problema que se estudia no tiene un componente aleatorio explícito, sin embargo, en esos casos, un parámetro determinístico del problema es expresado como un parámetro de la distribución aleatoria y luego esa distribución es simulada (Gentle, 1982).

También, los métodos Monte Carlo son ampliamente usados en estudios de robustez de procedimientos estadísticos y en este caso, el método se refiere a simular observaciones de una distribución alternativa y calcular a partir de estas observaciones las estadísticas para el procedimiento de la forma habitual. La robustez del procedimiento estadístico ordinario se puede evaluar desde las distribuciones empíricas para las estadísticas obtenidas de dicha manera (Gentle, 1982).

A continuación, se describen algunas investigaciones que han utilizado diferentes ejemplos de simulaciones aplicadas a los modelos SAR.

En el caso específico de los modelos autorregresivos Kelley y Barry (1997) realizaron simulaciones Monte Carlo donde obtuvieron datos para distintos valores de λ y σ para generar la variable dependiente autocorrelacionada. Dentro de las conclusiones se destaca que el método de máxima verosimilitud es más robusto en comparación con los resultados obtenidos utilizando los modelos lineales generalizados en dos etapas y los modelos lineales generalizados en una etapa, ya que, en este caso, el método de máxima verosimilitud estima casi de una forma idéntica los valores de λ .

A su vez, Epperson y Li (1996) utilizaron simulaciones Monte Carlo para modelos estocásticos en patrones de variación genética entre poblaciones. Analizaron la autocorrelación espacial utilizando el estadístico I de Moran. En las conclusiones obtenidas se menciona que el estadístico I de Moran proporciona expectativas de robustez y los resultados indican que el tema estocástico unido al muestreo de variación genera diferencias muy pequeñas en las estadísticas relacionadas con la autocorrelación.

También, Dormann et al. (2007) estudiaron diferentes simulaciones para cada uno de los enfoques estadísticos planteados utilizando los mismos datos en distribuciones: Normal, Binomial y Poisson. Llegaron a conclusiones relacionadas con diferencias en los residuos entre modelos espaciales y no espaciales, además, se señalan algunas de las limitaciones de las simulaciones ya que existen factores que pueden influir en la autocorrelación espacial, específicamente con los datos utilizados en la investigación relacionados con distribución de especies basados en mapas.

En otro estudio, Anselin y Kelejian (1997) realizaron simulaciones Monte Carlo para comparar el efecto que puede tener distintos tamaños de muestra para las distribuciones: Normal, Lognormal, Uniforme y Chi-cuadrado. La prueba con mejores resultados presenta el modelo con una variable dependiente espacial rezagada y por último, se concluye que tanto la prueba I de Moran como la prueba del multiplicador de Lagrange para estudiar la autocorrelación espacial son asintóticamente equivalentes cuando solamente los regresores exógenos están presentes sin requerir que los errores estén distribuidos normalmente.

Además, Lauridsen y Kosfeld (2004) calcularon simulaciones Monte Carlo para un ejemplo empírico de una distribución de muestra finita que involucra la prueba de Wald con no estacionariedad espacial. Dentro de los resultados que arroja la simulación, se analizó para un conjunto general de datos con algunos tipos de matrices de contigüidad con diferentes tamaños muestrales finitos. Se encontró que los valores críticos para la prueba de Wald con no estacionariedad son generalmente más altos que los valores críticos de la distribución Chi-cuadrado.

4. 1. 2 Simulaciones planeadas para el proyecto

En el presente proyecto, se analizará mediante simulaciones la robustez del modelo SAR ante violaciones a los siguientes supuestos del SAR:

- Normalidad de los errores u .
- El grado de independencia de los errores ε con distintos niveles de autocorrelación espacial λ .

En ese sentido, se pretende calcular cuánto cambian las estimaciones de los coeficientes del modelo y sus errores estándar, con errores no normales si la variable dependiente no presenta un comportamiento normal.

En esta investigación, los pasos de la simulación utilizada son los siguientes:

1. Generar números aleatorios para cada una de las columnas de la matriz que representan valores de distintos coeficientes para los modelos SAR.
2. Construir una matriz β de unos.
3. Obtener cálculos para distintos valores de λ .
4. Obtener cálculos para distintos valores de σ .
5. Los errores aleatorios se generaron con distintas distribuciones asimétricas similares a la normal como por ejemplo la distribución beta; exceptuando la distribución normal. Para identificar los cambios obtenidos para cada una de las distribuciones calculadas, se utiliza el coeficiente de asimetría y un estadístico de una prueba de hipótesis.

Y, para analizar el supuesto de independencia, se realizan los mismos pasos utilizados para la simulación propuesta, aumentando los valores de λ .

Además, se simulan los distintos valores de λ para analizar el comportamiento de los datos tomando en cuenta varios valores definidos previamente por el investigador, para los 471 distritos de Costa Rica existentes en el 2011 excluyendo la Isla de Chira ya que no posee vecinos contiguos y de esta forma estimar los valores de la variable dependiente. Es decir, las simulaciones sirven para evaluar la robustez del modelo SAR a los datos distritales costarricenses.

Cada una de las simulaciones realizadas son concluyentes únicamente con la estructura espacial de Costa Rica ya que los datos espaciales a nivel de unidad político-administrativa estarán en el mismo lugar sin modificaciones.

4. 1. 3 Características de la aplicación del modelo SAR a datos distritales costarricenses en términos geográficos

La investigación toma en cuenta un total de 472 distritos de Costa Rica. Para el 2007, existían 469 distritos y durante el 2009 se crearon 3 distritos más (Palmitos, Katira y Rincón de Sabanilla), de manera que, al final del periodo de estudio, se contabiliza un total de 472 distritos. Para efectos de los mapas distritales, las coordenadas utilizadas están basadas en la proyección CRTM05 proporcionadas por el Instituto Nacional de Estadística y Censos (INEC).

A nivel geográfico, nuestro país tiene una particularidad ya que agrupa una gran cantidad de distritos en la región Central (alrededor del 58% del total de distritos del país) y en su mayoría, se caracterizan por poseer áreas relativamente pequeñas, por ejemplo de los 274 distritos que conforman la región Central, un total de 214 no sobrepasan los 30 km² de extensión. Por el contrario, los distritos localizados en la periferia del país, ubicados en las demás regiones de planificación, presentan áreas más grandes. Por ejemplo, de los restantes 198 distritos del territorio nacional, un total de 54 sobrepasan los 250 km² de extensión.

4. 2 Abordaje metodológico de la aplicación del SAR a un modelo que relaciona el programa social *Avancemos* con la permanencia en el sistema escolar

4. 2. 1 Población, unidad de estudio, alcance

La población objetivo corresponde a todas las personas, sean nacionales o extranjeras que hayan tenido una participación directa o indirecta con el programa social *Avancemos*. Para este caso particular, son aquellas familias que solicitaron ayuda económica para la permanencia de sus hijos en el sistema educativo costarricense. No obstante, es importante señalar que la unidad de análisis es el distrito, ya que la información está desagregada a dicho nivel geográfico. Para el programa social *Avancemos*, se analiza la presencia del mismo a nivel distrital y su asociación con diversas variables de interés.

Las variables involucradas en la investigación cumplen con el requisito de desagregación a nivel distrital. La disponibilidad de información es limitada, pero sin embargo el país cuenta con diversas bases de datos, cuyas variables están desagregadas a nivel distrital, facilitando el acceso a dicha información y solicitando ayuda a las instituciones que son fuente primaria de datos cuando el caso lo permita.

Con respecto a los resultados, se espera encontrar que la aplicación y el tiempo de ejecución del programa social *Avancemos* esté significativamente asociado con la deserción en secundaria en Costa Rica. Además, el *software* utilizado para la elaboración de los mapas fue el programa R 2.14 y para el

análisis estadístico además del mencionado anteriormente se usó el programa SPSS 19.

En el caso propio de esta investigación, el modelo de regresión utilizado y sus respectivas variables de interés para el programa social *Avancemos*, se describen a continuación.

4. 2. 2 Modelo de Avancemos

Con respecto a las variables utilizadas en el modelo, se le dio prioridad a las variables relacionadas con el programa social *Avancemos* y las otras variables están justificadas en la metodología, pero se seleccionaron también por la disponibilidad de información a nivel distrital.

Variable dependiente

La variable dependiente es el porcentaje de deserción en los colegios desde el 2007 hasta el 2011. En ese sentido, es la proporción que representa los estudiantes que desertaron en un año. Se considera desertor al estudiante que abandona el curso lectivo, es decir, la diferencia entre la matrícula inicial y la matrícula final de un año. La variable pretende responder al principal objetivo del programa social *Avancemos* que consiste en mantener a los estudiantes dentro del sistema educativo para que no se ausenten de las aulas de educación secundaria (Mivah, 2007), (Céspedes y Jiménez, 2010), (Trejos, 2012a), (Trejos, 2012b), (Programa Estado de la Nación, 2013). Los datos administrativos fueron suministrados por el Ministerio de Educación Pública (MEP) y los datos censales fueron obtenidos del Instituto Nacional de Estadística y Censos (INEC).

Variables independientes

- i. Tasa de la cantidad de las becas *Avancemos* que han sido suministradas por distrito que registra un conteo de las becas entregadas por distrito por cada mil personas con edades de 12 a 21 años del 2007 al 2011, proporcionadas por el Instituto Mixto de Ayuda Social (IMAS). Para contabilizar esta variable se toma en cuenta el lugar de residencia del beneficiario para cada uno de los años en particular. Existen registros que indican el otorgamiento de miles de becas, lo cual hace interesante su inclusión para el análisis (Trejos, 2012a), (Programa Estado de la Nación, 2008), (Programa Estado de la Nación, 2011), (Programa Estado de la Nación, 2013).
- ii. Monto promedio de las becas *Avancemos* en miles de colones que han sido suministradas por distrito, con colaboración del Instituto Mixto de Ayuda Social (IMAS), calculado con un promedio simple de los montos de las becas por distrito del 2007 al 2011. Desde su implementación, el Estado ha otorgado miles de millones de colones para distribuir estas transferencias condicionadas (Programa Estado de la Nación, 2008), (Programa Estado de la Nación, 2011), (Programa Estado de la Nación, 2013).
- iii. Consumo promedio de electricidad por distrito. Esta variable es importante, ya que es utilizada como *proxy* del ingreso por distrito del 2007 al 2011. Es importante tener una variable de esta índole debido a que los problemas socioeconómicos en el seno familiar inciden de cierta forma en la deserción (Vargas, 1987), (Brenes, 1995), (Preal, 2003), (Arguedas y Jiménez, 2008), (Fernández y Del Valle, 2013). Estos datos fueron obtenidos de investigaciones realizadas por el Ministerio de Planificación Nacional y Política Económica (Mideplan). Cabe señalar que los datos de los distritos costeros con alto consumo de electricidad fueron depurados por medio de estadísticos de tendencia central.

- iv. Población en edad colegial de 12 a 21 años, obtenida de la base de datos del Centro Centroamericano de Población (CCP) del 2007 al 2011. Esta variable está siendo utilizada como un control para la variable dependiente. Se decidió utilizar la población de 12 a 21 años porque precisamente entre estas edades se encuentra la población meta del programa social *Avancemos*.

- v. Cantidad de colegios y tasa del total de profesores por distrito por cada mil personas que asisten al colegio del 2007 al 2011. El registro de estos datos por parte del Ministerio de Educación Pública (MEP), puede ser utilizado para observar si en los distritos con mayor cantidad de colegios y profesores existen mayor cantidad de becas *Avancemos*. Existen datos disponibles del 2007 al 2011. Estas variables se utilizan como un control para la variable dependiente.

5. Evaluación de la robustez del modelo SAR ante cambios de escenarios de estimación

Para evaluar la robustez de las estimaciones del modelo SAR con distintos grados de autocorrelación espacial, se realizaron simulaciones tomando en cuenta distintos niveles que determinan la autocorrelación espacial λ . Para ello fue necesario aplicar una distribución normal con media cero y varianza σ dependiendo de los valores teóricos representados como sigma. Posteriormente, se estudia la robustez del modelo ante la violación del supuesto de normalidad, simulando residuos de distribuciones beta con distintos niveles de asimetría.

Este procedimiento se calculó en tres ocasiones para la distribución normal y para la distribución beta. También se cambió el tipo de matriz de pesos, para analizar si se obtenían resultados distintos si se usaba otro tipo de matriz. El modelo SAR resultante de esta investigación posee la matriz de pesos reina o *queen* cuyas unidades espaciales comparten una frontera en común o al menos un vértice. El otro par de alternativas aplicadas fueron la matriz de distancias que está compuesta precisamente por las distancias entre los centroides de las unidades espaciales de estudio y la matriz K vecinos más cercanos que considera a un cierto número predefinido K de vecinos que estén más cercanos a las unidades espaciales de estudio. De esta forma, también se evaluó la variabilidad de las estimaciones al usar distintas matrices de distancias.

Una vez realizado el procedimiento se procede a comparar los resultados obtenidos tomando en cuenta cada una de las opciones desarrolladas. Esta

comparación permite identificar si las estimaciones están cercanas a los parámetros teóricos, tomando en cuenta específicamente la estructura espacial de los distritos de Costa Rica para el periodo de estudio delimitado en esta investigación. Es importante indicar que el *software* utilizado para realizar las simulaciones fue el paquete estadístico R 2.14.

5. 1 Simulaciones para la autocorrelación espacial

5. 1. 1 Metodología

Las simulaciones utilizadas están compuestas por la información de 471 distritos porque el distrito de Chira, al ser una isla, tuvo que ser excluido del análisis debido a que no posee vecinos totalmente contiguos. La matriz de pesos reina o *queen* representada como W está definida según el método de fronteras en común. Como el modelo SAR analizado contiene 6 variables independientes, fue necesario construir un vector β de 7 unos incluyendo el espacio del β_0 correspondiente.

Cabe señalar que la matriz de pesos reina o *queen* fue la utilizada para este trabajo de investigación porque en la aplicación detallada en el Capítulo correspondiente, se considera que los jóvenes tienen que ir a colegios que sean vecinos de su distrito aun cuando solo compartan una pequeña parte de frontera con el distrito, por lo tanto se pensó en usar este tipo de matriz de pesos específicamente por la aplicación, ya que, los estudiantes asisten a los colegios situados en los distritos más cercanos, generalmente, en distritos que comparten frontera en común.

Además, los resultados de las simulaciones fueron obtenidos después de utilizar 7 y 8 valores teóricos para los parámetros sigma y lambda, respectivamente. También, fue necesario crear la matriz identidad con 471 filas y columnas correspondientes al total de distritos analizados.

El total de corridas de las simulaciones fue de 500 repeticiones, la matriz de información creada está compuesta por 471 filas que conforman 3 297 valores de las 7 variables simuladas (incluyendo el β_0) utilizando una distribución uniforme. Para replicar el modelo SAR en esta simulación, es necesario que los errores sean independientes e idénticamente distribuidos con media cero y varianza constante, en este caso los errores están en una matriz compuesta por una distribución normal con media cero y varianza sigma para cada uno de los 7 valores teóricos escogidos. El modelo teórico que se está simulando es el siguiente:

$$Y_i = 1 + X_{1i} + X_{2i} + X_{3i} + X_{4i} + X_{5i} + X_{6i} + \varepsilon_i \quad (14)$$

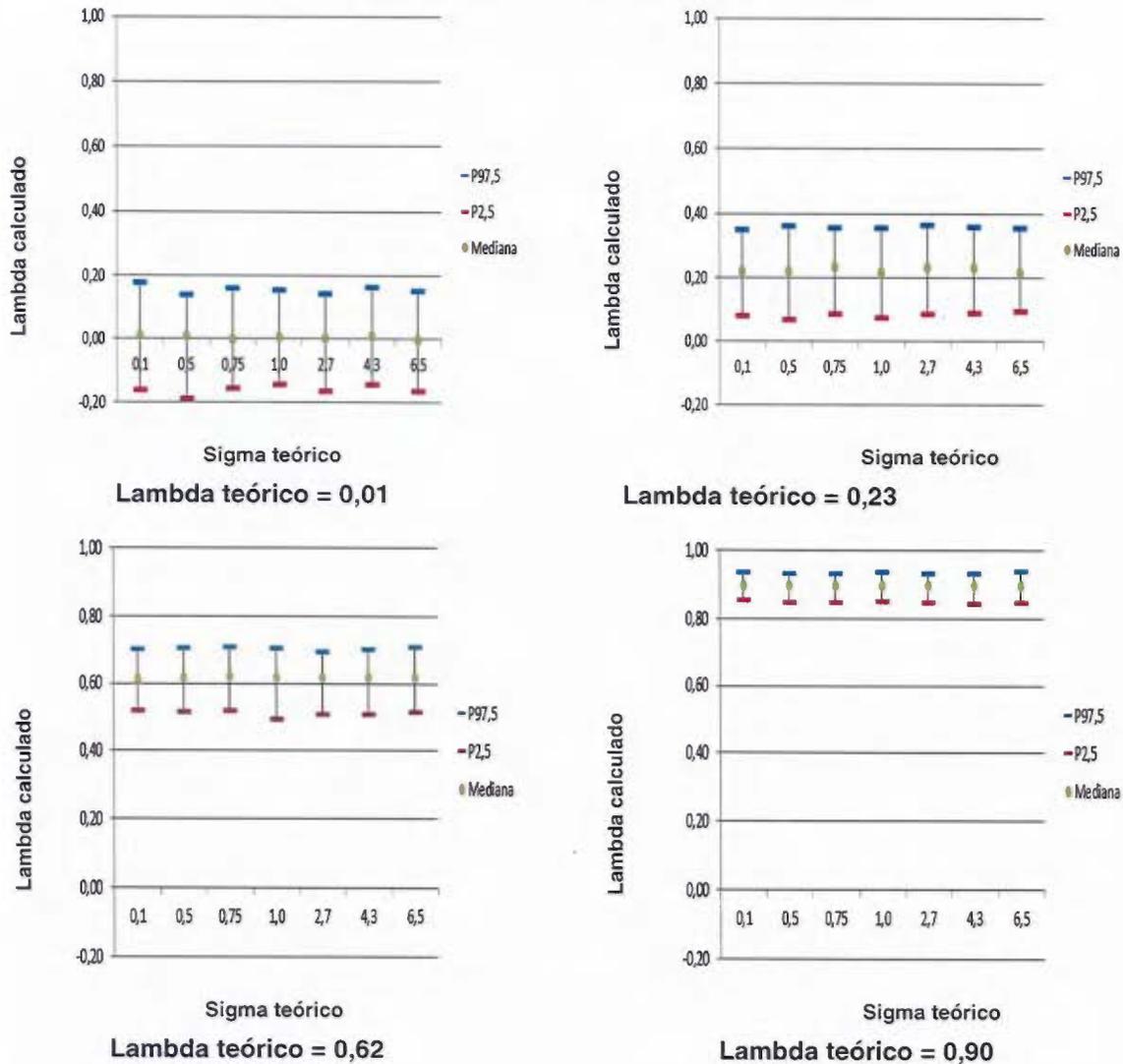
Es así como se forma el modelo SAR de la simulación, multiplicando la matriz de información por el vector β sumando el resultado de la inversa de la matriz identidad menos la multiplicación de los valores escogidos de λ y la matriz de pesos W . Además, los resultados de las simulaciones del modelo SAR permiten evaluar la normalidad de los coeficientes para cada caso según sean los valores teóricos de los parámetros de λ y σ , así como calcular el nivel de potencia para identificar el valor teórico de lambda que asegure con certeza la existencia de autocorrelación espacial utilizando la estructura geográfica de los distritos de Costa Rica. Este proceso también fue desarrollado paralelamente cambiando la composición de los errores con el uso de una distribución beta para probar el supuesto de normalidad. La evaluación de la normalidad de los coeficientes se realiza suponiendo que, al tener un tamaño de muestra de 471 distritos, la

distribución de los coeficientes se aproxima asintóticamente a la distribución normal, por el teorema del límite central con media igual a 1, el valor del parámetro que se usó en los modelos simulados.

5. 1. 2 Análisis

En este primer caso utilizando el modelo para lambda con distribución normal, se decide obtener valores para la mediana y para los intervalos de confianza simulados representados por los percentiles 2,5 y 97,5 tomando en cuenta distintos valores teóricos para los parámetros λ y σ . Los resultados obtenidos indican que conforme aumentan los valores de lambda, los intervalos de confianza simulados son más angostos comparando los resultados de los percentiles 2,5 y 97,5 y el valor de la mediana aumenta.

Gráfico 1. Intervalos de confianza simulados para lambda utilizando distribución normal con matriz *queen*, 2011



Fuente: Elaboración propia, 2011.

Con respecto a los intervalos de confianza simulados, el porcentaje de corridas que son significativas con un nivel de significancia del 5% es mayor conforme aumentan los valores teóricos del parámetro λ . Para que el intervalo sea significativo, el percentil 2,5 y 97,5 no debe de incluir al cero. A partir del valor de lambda de 0,23 y para cualquier valor de sigma, los intervalos de confianza

simulados no incluyen al cero, por lo tanto son significativos verificando de esta forma la autocorrelación espacial (Gráfico 1 y Cuadro 1). Es importante mencionar que para valores de lambda menores a 0,23 no es posible confirmar la presencia o no de autocorrelación espacial con la prueba de hipótesis respectiva para la estructura espacial de distritos de Costa Rica.

**Cuadro 1. Intervalos de la estimación del lambda según valores teóricos
utilizando distribución normal con matriz *queen*, 2011**

Sigma teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,10	P2,5	-0,1640	-0,0643	0,0777	0,2272	0,3619	0,5160	0,6660	0,8501
	Me	0,0101	0,0962	0,2180	0,3542	0,4837	0,6154	0,7430	0,8970
	P97,5	0,1743	0,2295	0,3482	0,4794	0,5789	0,7000	0,8072	0,9337
0,50	P2,5	-0,1906	-0,0488	0,0668	0,2228	0,3586	0,5123	0,6626	0,8448
	Me	0,0069	0,0963	0,2166	0,3517	0,4854	0,6156	0,7460	0,8953
	P97,5	0,1366	0,2286	0,3601	0,4717	0,5781	0,7030	0,8166	0,9298
0,75	P2,5	-0,1584	-0,0778	0,0842	0,2217	0,3807	0,5157	0,6618	0,8428
	Me	-0,0034	0,1001	0,2315	0,3544	0,4871	0,6193	0,7442	0,8949
	P97,5	0,1564	0,2414	0,3552	0,4653	0,5878	0,7051	0,8147	0,9284
1,00	P2,5	-0,1446	-0,0739	0,0720	0,2161	0,3685	0,4922	0,6643	0,8470
	Me	0,0043	0,0911	0,2155	0,3605	0,4803	0,6158	0,7435	0,8959
	P97,5	0,1517	0,2337	0,3532	0,4763	0,5858	0,7044	0,8165	0,9327
2,70	P2,5	-0,1653	-0,0631	0,0848	0,2140	0,3688	0,5066	0,6601	0,8428
	Me	-0,0009	0,0924	0,2280	0,3520	0,4862	0,6158	0,7453	0,8943
	P97,5	0,1403	0,2384	0,3636	0,4662	0,5786	0,6916	0,8089	0,9282
4,30	P2,5	-0,1443	-0,0748	0,0872	0,2280	0,3699	0,5071	0,6554	0,8418
	Me	0,0047	0,0951	0,2295	0,3548	0,4886	0,6192	0,7454	0,8955
	P97,5	0,1615	0,2261	0,3570	0,4766	0,5904	0,7008	0,8107	0,9295
6,50	P2,5	-0,1656	-0,0585	0,0912	0,2301	0,3696	0,5126	0,6564	0,8437
	Me	-0,0047	0,0972	0,2156	0,3533	0,4868	0,6182	0,7471	0,8950
	P97,5	0,1493	0,2472	0,3527	0,4715	0,5930	0,7057	0,8142	0,9360

Fuente: Elaboración propia, 2011.

A medida que aumenta el valor de sigma existe menos variabilidad explicada disminuyendo el R^2 y viceversa. Además, con un lambda teórico alto de 0,90 por ejemplo, el valor del R^2 disminuye considerablemente por la correlación existente.

Con respecto a los coeficientes del modelo SAR simulado, todos los intervalos de confianza simulados contienen al valor 1, lo cual confirma que se obtienen los resultados esperados cuando se cumplen los supuestos. Con valores de sigma mayores o iguales a 4,3 se tiene coeficientes estimados no significativos porque los intervalos contienen el valor 0 (Cuadro 2). Además, cuanto mayor es el valor de sigma, más amplios son los intervalos de confianza simulados quitando precisión en los resultados. Según los gráficos de normalidad *qqplot* realizados se puede afirmar que los coeficientes se comportan normalmente ya que los puntos de los gráficos forman una línea recta para todos los valores de lambda confirmándose el supuesto de normalidad de los coeficientes (Gráfico 2).

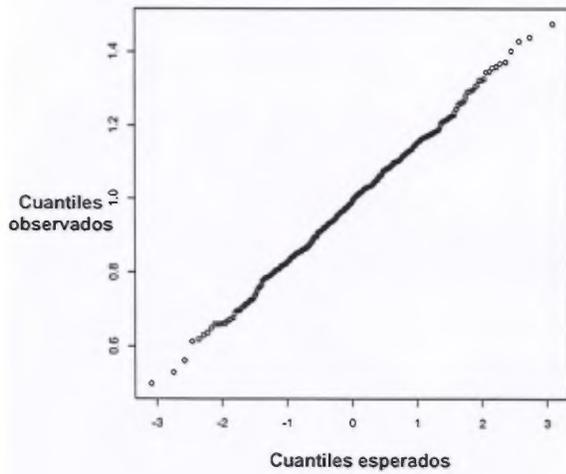
Cuadro 2. Intervalos de confianza de los coeficientes, variando el lambda y el sigma utilizando distribución normal con matriz *queen*, 2011

Sigma teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,10	P2,5	0,9700	0,9666	0,9681	0,9677	0,9707	0,9713	0,9681	0,9718
	Me	1,0004	1,0010	0,9983	1,0010	0,9992	0,9995	0,9995	0,9998
	P97,5	1,0348	1,0320	1,0313	1,0337	1,0316	1,0290	1,0264	1,0268
0,50	P2,5	0,8427	0,8341	0,8533	0,8416	0,8550	0,8450	0,8626	0,8471
	Me	1,0036	0,9980	0,9984	1,0004	1,0011	0,9995	0,9973	0,9983
	P97,5	1,1720	1,1431	1,1456	1,1567	1,1620	1,1511	1,1382	1,1377
0,75	P2,5	0,7672	0,7540	0,7632	0,7557	0,7400	0,7761	0,8003	0,7696
	Me	0,9977	1,0001	1,0057	1,0021	1,0026	1,0008	0,9945	0,9930
	P97,5	1,2519	1,2159	1,2131	1,2361	1,2361	1,2071	1,2060	1,2234
1,00	P2,5	0,6419	0,6679	0,6577	0,6949	0,7125	0,7223	0,7079	0,7163
	Me	0,9935	0,9962	0,9998	0,9975	1,0093	1,0098	1,0044	1,0040
	P97,5	1,3048	1,3175	1,3423	1,3150	1,2768	1,2963	1,3217	1,3307
2,70	P2,5	0,1545	0,2018	0,2465	0,1046	0,2075	0,1634	0,2323	0,2247
	Me	0,9614	1,0049	1,0177	1,0094	0,9962	1,0140	1,0086	0,9872
	P97,5	1,9117	1,9530	1,8713	1,8628	1,6903	1,8850	1,8474	1,9086
4,30	P2,5	-0,2906	-0,2287	-0,4098	-0,1357	-0,2602	-0,2591	-0,2596	-0,2531
	Me	0,9816	0,9871	0,9306	1,0316	0,9861	0,9932	1,0403	1,0429
	P97,5	2,3921	2,4169	2,2590	2,3670	2,3167	2,3275	2,3931	2,2935
6,50	P2,5	-1,0774	-1,1594	-1,0195	-0,8760	-1,0535	-0,8199	-0,7855	-0,7568
	Me	1,0339	0,9592	0,9599	0,9545	1,0409	0,9735	1,1389	1,1073
	P97,5	2,9839	2,8083	2,7562	3,0151	2,8498	3,1118	3,0782	2,7405

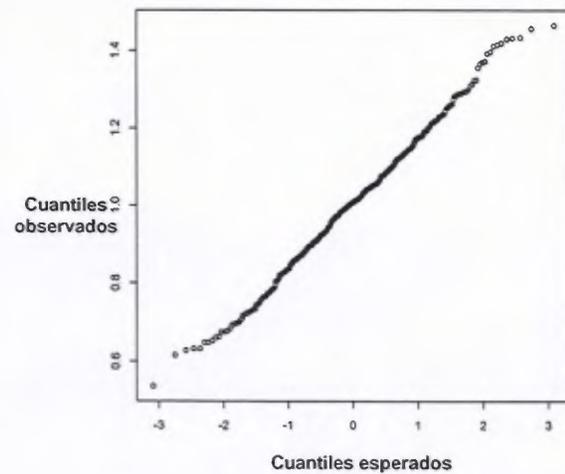
Nota: La media teórica es 4 y las equivalencias en coeficiente de variación se obtienen dividiendo el valor de sigma entre la media teórica.

Fuente: Elaboración propia, 2011.

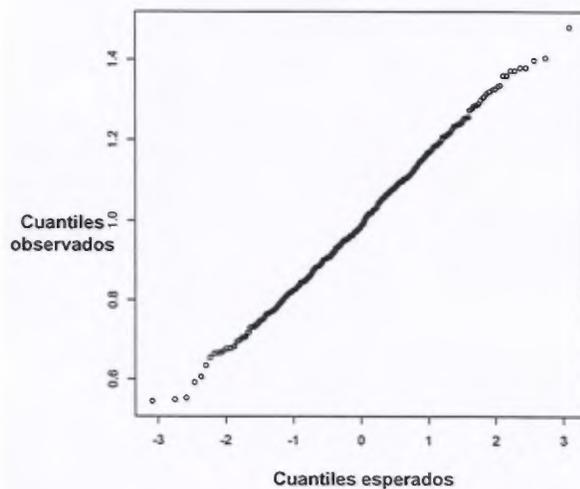
Gráfico 2. Gráfico de normalidad de los coeficientes con distribución normal y matriz queen, 2011



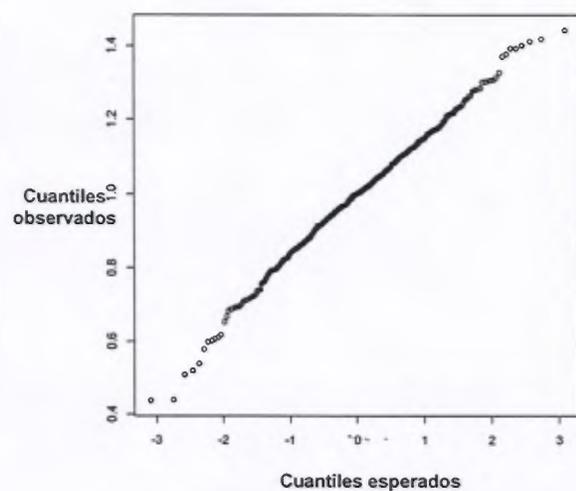
Lambda teórico = 0,01 y sigma teórico = 1



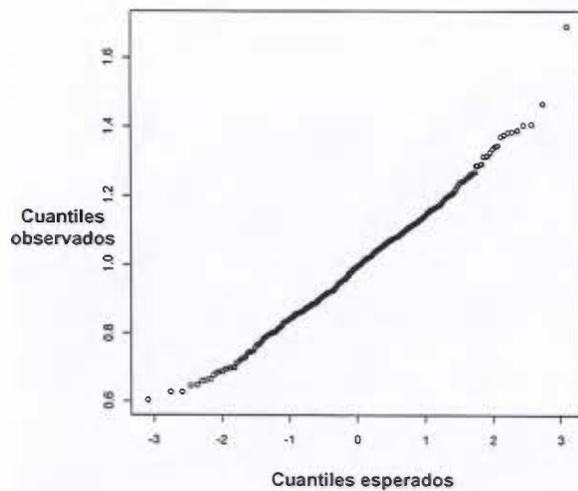
Lambda teórico = 0,10 y sigma teórico = 1



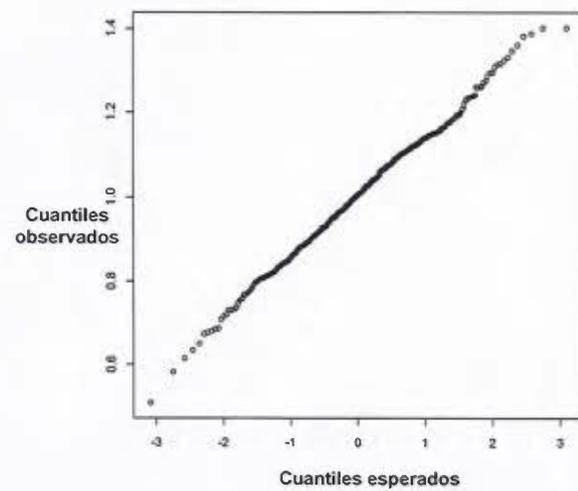
Lambda teórico = 0,23 y sigma teórico = 1



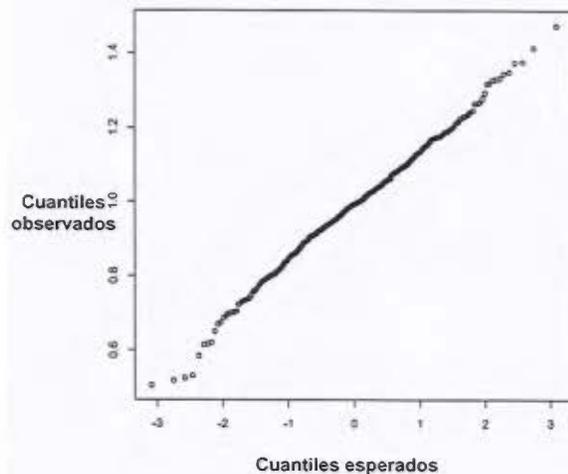
Lambda teórico = 0,36 y sigma teórico = 1



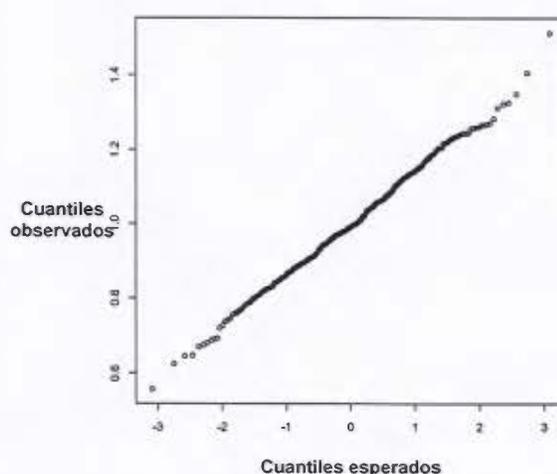
Lambda teórico = 0,49 y sigma teórico = 1



Lambda teórico = 0,62 y sigma teórico = 1



Lambda teórico = 0,75 y sigma teórico = 1



Lambda teórico = 0,90 y sigma teórico = 1

Fuente: Elaboración propia, 2011.

Las propiedades asintóticas de los estimadores de máxima verosimilitud permiten identificar la tasa de rechazo con la que se puede confiar en los resultados obtenidos para concluir sobre la existencia de autocorrelación espacial. En ese sentido, en el modelo para lambda con distribución normal se puede afirmar que existe una potencia de prueba mayor al 80% de detectar a lambda como significativamente distinto de cero cuando el lambda teórico es mayor o igual a 0,23 (Cuadro 3). Esta potencia de prueba es alta debido a que las unidades estadísticas son 471 distritos, siendo una muestra suficientemente

grande y también, por el teorema del límite central el estadístico de prueba tendría una distribución normal, en este caso, la potencia varía poco cuando se cambia el valor del sigma como error muestral.

Cuadro 3. Porcentaje de corridas para identificar la potencia de prueba con distribución normal utilizando matriz *queen*, 2011

Sigma teórico	Lambda teórico							
	0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,10	0,080	0,248	0,842	1,000	1,000	1,000	1,000	1,000
0,50	0,062	0,248	0,836	1,000	1,000	1,000	1,000	1,000
0,75	0,062	0,250	0,884	0,998	1,000	1,000	1,000	1,000
1,00	0,054	0,232	0,840	1,000	1,000	1,000	1,000	1,000
2,70	0,056	0,248	0,876	0,998	1,000	1,000	1,000	1,000
4,30	0,058	0,248	0,866	1,000	1,000	1,000	1,000	1,000
6,50	0,060	0,252	0,868	1,000	1,000	1,000	1,000	1,000

Fuente: Elaboración propia, 2011.

5. 2 Simulaciones para diferentes matrices de pesos

5. 2. 1 Metodología

Las simulaciones elaboradas cambiando el tipo de matriz de pesos contienen el mismo procedimiento utilizado para evaluar la autocorrelación espacial. Sin embargo se quiso probar si las condiciones encontradas anteriormente siguen el mismo comportamiento cuando el modelo SAR deja de utilizar la matriz reina o *queen* y es cambiada por la matriz de pesos K vecinos más cercanos cuyo valor para la simulación es de 6 ya que ese es el promedio de vecinos que presenta la estructura espacial de los distritos de Costa Rica. La matriz de distancias también

fue utilizada para este procedimiento, escogiendo de manera arbitraria aquellos distritos que estuviesen entre 1km y 30km de su respectivo centroide.

5. 2. 2 Análisis

Para la matriz de pesos K vecinos más cercanos y la matriz de pesos de distancias, existen algunas semejanzas y diferencias en los resultados obtenidos. Una de las diferencias radica en que la matriz K vecinos más cercanos presenta un valor más pequeño del lambda teórico (0,23) a partir del que los intervalos de confianza simulados no incluyen el valor de cero con un 95% de confianza afirmando que existe suficiente potencia en las estimaciones para detectar autocorrelación espacial con los datos distritales de Costa Rica (Cuadro 4). Por su parte, la matriz de pesos de distancias sugiere que a partir de un lambda teórico de 0,36 existe un 95% de confianza para decir que hay evidencia estadística de autocorrelación espacial (Cuadro 5).

Observando los gráficos de los intervalos de confianza simulados se obtienen resultados similares con respecto a la amplitud de dichos intervalos. No obstante, existe una leve amplitud mayor en los intervalos de la matriz de pesos de distancias comparando con los intervalos de confianza simulados de la matriz de pesos K vecinos más cercanos (Gráficos 3 y 4).

**Cuadro 4. Intervalos de la estimación del lambda según valores teóricos
utilizando distribución normal con matriz *K* vecinos, 2011**

Sigma teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,10	P2,5	-0,1841	-0,0963	0,0613	0,2107	0,3625	0,5219	0,6551	0,8456
	Me	0,0092	0,0921	0,2137	0,3551	0,4813	0,6188	0,7432	0,8962
	P97,5	0,1585	0,2455	0,3739	0,4768	0,5899	0,6977	0,8123	0,9270
0,50	P2,5	-0,1691	-0,0823	0,0602	0,2237	0,3497	0,5007	0,6481	0,8372
	Me	0,0030	0,0906	0,2319	0,3566	0,4852	0,6169	0,7416	0,8949
	P97,5	0,1645	0,2416	0,3625	0,4825	0,5958	0,7028	0,8092	0,9313
0,75	P2,5	-0,1802	-0,0747	0,0657	0,1973	0,3571	0,5080	0,6583	0,8509
	Me	0,0019	0,0846	0,2196	0,3531	0,4882	0,6131	0,7440	0,8959
	P97,5	0,1662	0,2345	0,3561	0,4885	0,5926	0,7055	0,8136	0,9280
1,00	P2,5	-0,1822	-0,0898	0,0780	0,2052	0,3447	0,5016	0,6510	0,8405
	Me	-0,0020	0,0891	0,2238	0,3560	0,4846	0,6178	0,7465	0,8936
	P97,5	0,1747	0,2301	0,3569	0,4704	0,5975	0,7002	0,8111	0,9309
2,70	P2,5	-0,1737	-0,0879	0,0651	0,2028	0,3407	0,5160	0,6473	0,8470
	Me	-0,0139	0,0845	0,2251	0,3573	0,4868	0,6117	0,7415	0,8952
	P97,5	0,1561	0,2377	0,3637	0,4660	0,5961	0,7091	0,8030	0,9265
4,30	P2,5	-0,1736	-0,0862	0,0471	0,2093	0,3430	0,4969	0,6562	0,8496
	Me	-0,0067	0,0877	0,2219	0,3532	0,4783	0,6113	0,7468	0,8976
	P97,5	0,1557	0,2301	0,3608	0,4949	0,5830	0,6973	0,8141	0,9294
6,50	P2,5	-0,1806	-0,0680	0,0432	0,1900	0,3728	0,5178	0,6634	0,8432
	Me	-0,0072	0,0955	0,2194	0,3566	0,4830	0,6163	0,7484	0,8953
	P97,5	0,1366	0,2371	0,3611	0,4787	0,5934	0,6976	0,8074	0,9291

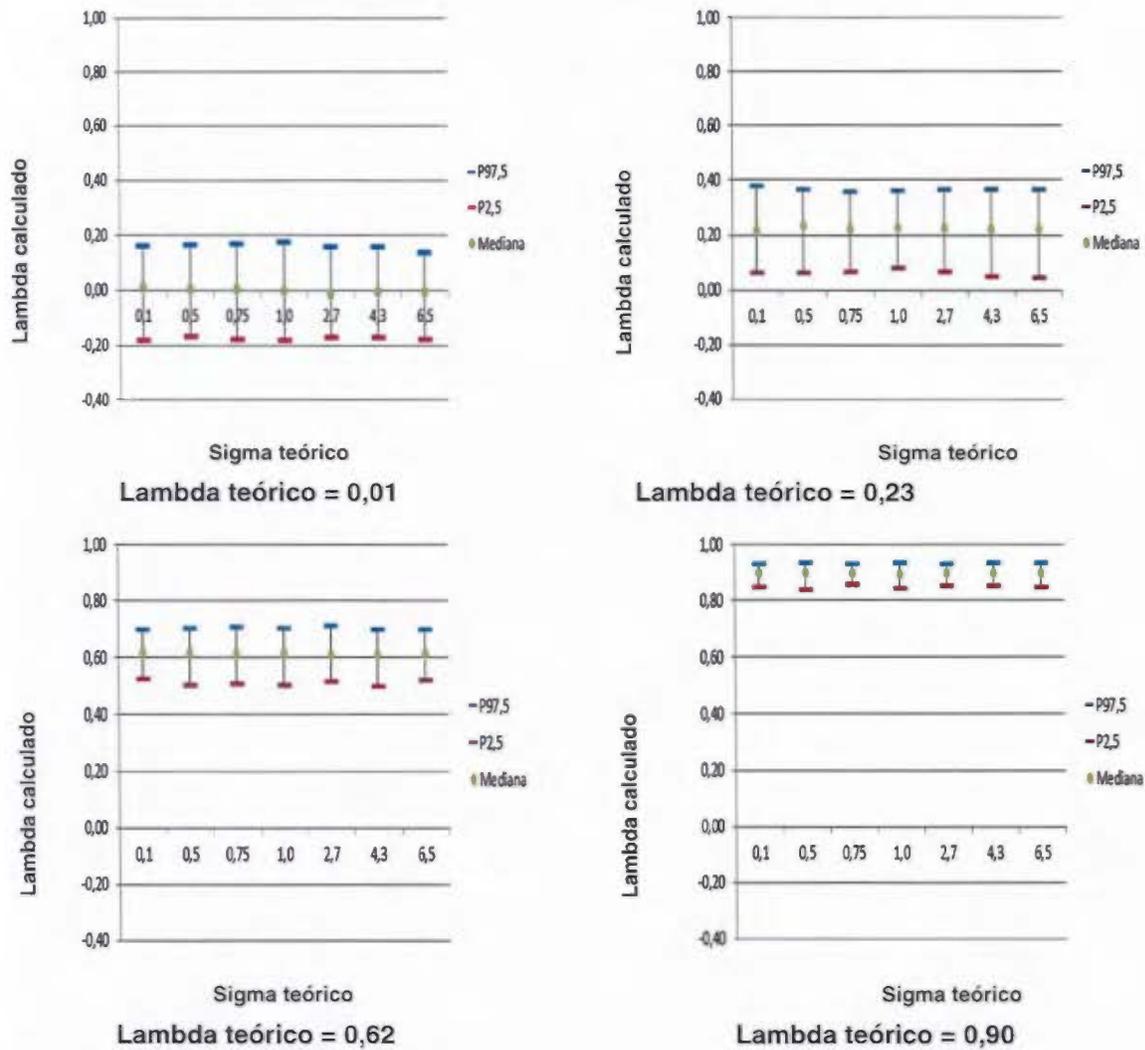
Fuente: Elaboración propia, 2011.

Cuadro 5. Intervalos de la estimación del lambda según valores teóricos utilizando distribución normal con matriz de distancias, 2011

Sigma teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,10	P2,5	-0,3321	-0,2614	-0,1398	0,0266	0,2232	0,3780	0,5427	0,7585
	Me	-0,0168	0,0793	0,2174	0,3440	0,4518	0,5933	0,7255	0,8816
	P97,5	0,2456	0,3283	0,4368	0,5291	0,6367	0,7333	0,8355	0,9384
0,50	P2,5	-0,3588	-0,2460	-0,0829	0,0737	0,2309	0,3809	0,5637	0,7581
	Me	-0,0206	0,0774	0,2029	0,3344	0,4747	0,5981	0,7289	0,8781
	P97,5	0,2443	0,3148	0,4472	0,5219	0,6380	0,7376	0,8322	0,9355
0,75	P2,5	-0,3317	-0,2606	-0,1128	-0,0151	0,2149	0,3624	0,5464	0,7655
	Me	-0,0164	0,0578	0,1898	0,3373	0,4703	0,5908	0,7267	0,8844
	P97,5	0,2568	0,2977	0,4291	0,5609	0,6496	0,7360	0,8295	0,9404
1,00	P2,5	-0,3668	-0,2716	-0,1096	0,0383	0,1985	0,3546	0,5340	0,7596
	Me	-0,0122	0,0712	0,1976	0,3332	0,4640	0,5988	0,7272	0,8814
	P97,5	0,2532	0,3045	0,4147	0,5318	0,6381	0,7236	0,8313	0,9400
2,70	P2,5	-0,3476	-0,2431	-0,1263	0,0369	0,1610	0,3625	0,5332	0,7403
	Me	-0,0195	0,0719	0,2049	0,3344	0,4658	0,5978	0,7284	0,8823
	P97,5	0,2528	0,3224	0,4254	0,5401	0,6366	0,7386	0,8383	0,9410
4,30	P2,5	-0,3436	-0,2702	-0,0830	0,0799	0,2066	0,4089	0,5427	0,7695
	Me	-0,0170	0,0656	0,2041	0,3258	0,4573	0,5936	0,7270	0,8811
	P97,5	0,2655	0,2993	0,4394	0,5286	0,6332	0,7409	0,8364	0,9330
6,50	P2,5	-0,3274	-0,2248	-0,0738	0,0202	0,2043	0,3717	0,5581	0,7694
	Me	-0,0148	0,0776	0,2053	0,3314	0,4599	0,5883	0,7275	0,8767
	P97,5	0,2574	0,3288	0,4281	0,5382	0,6317	0,7391	0,8289	0,9351

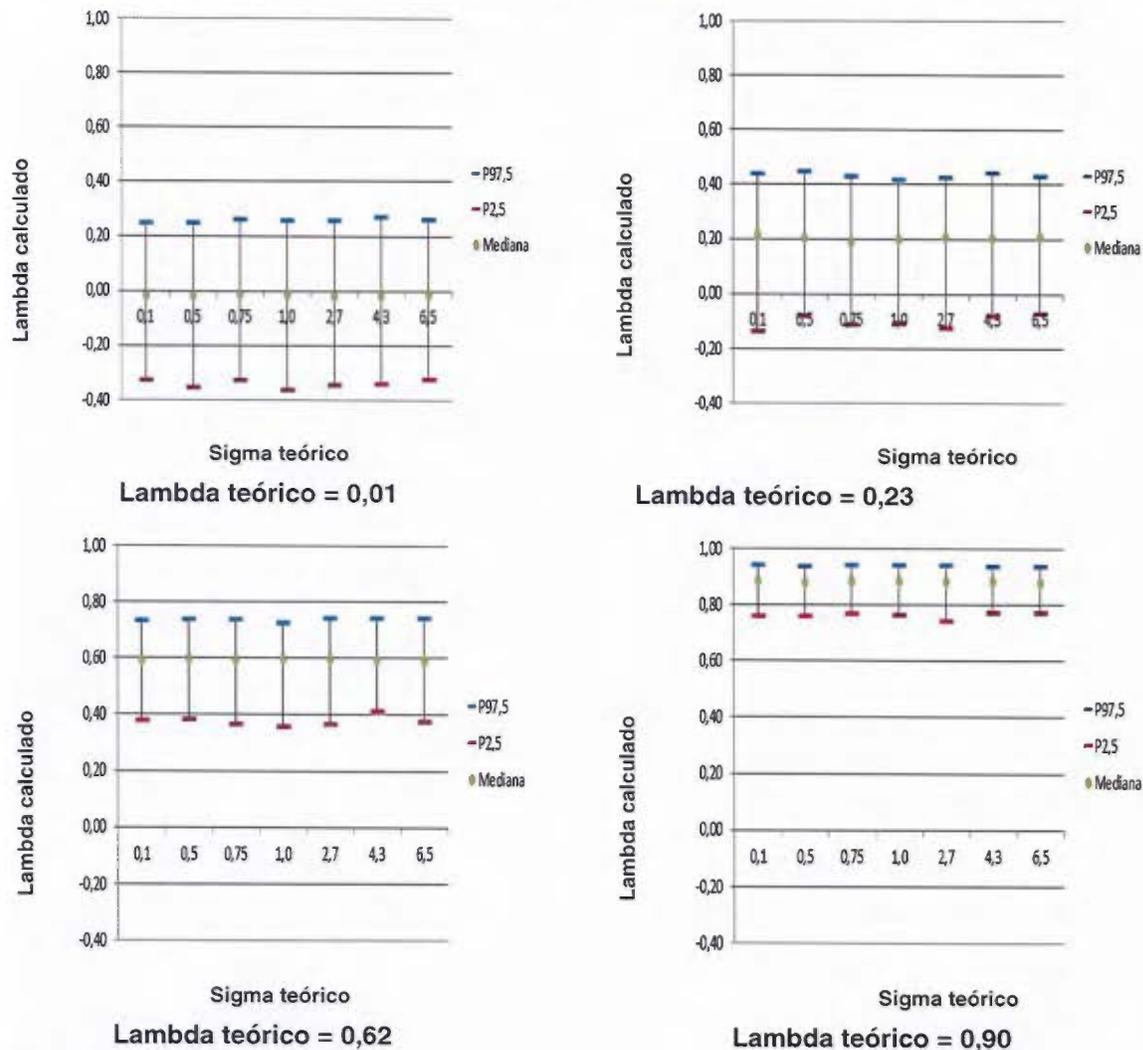
Fuente: Elaboración propia, 2011.

Gráfico 3. Intervalos de confianza simulados para lambda utilizando distribución normal con matriz K vecinos, 2011



Fuente: Elaboración propia, 2011.

Gráfico 4. Intervalos de confianza simulados para lambda utilizando distribución normal con matriz de distancias, 2011



Fuente: Elaboración propia, 2011.

Con respecto a los coeficientes del modelo SAR simulado, todos los intervalos de confianza simulados contienen al valor 1, lo cual confirma que se obtienen los resultados esperados cuando se cumplen los supuestos sin importar si se utiliza la matriz de distancias o la matriz K vecinos más cercanos. De la misma manera, con valores de sigma mayores o iguales a 4,3 tenemos coeficientes estimados no

significativos, porque los intervalos contienen el valor 0 para ambas matrices. Además, cuanto mayor es el valor de sigma, más amplios son los intervalos de confianza simulados quitando precisión en los resultados tanto para la matriz de distancias como para la matriz K vecinos más cercanos (Cuadros 6 y 7). Según los gráficos de normalidad *qqplot* realizados puede afirmarse que los coeficientes se comportan normalmente ya que los puntos de los gráficos forman una línea recta para todos los valores de lambda confirmándose el supuesto de normalidad de los coeficientes. Estos resultados igualmente fueron obtenidos utilizando la matriz *queen* con lo cual se concluye que para los coeficientes del modelo SAR simulado, los distintos tipos de matrices tienen igual comportamiento (Gráficos 5 y 6).

Cuadro 6. Intervalos de la estimación del coef. para X_1 según valores teóricos para coef. utilizando distrib. normal con matriz K vecinos, 2011

Sigma teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,10	P2,5	0,9717	0,9692	0,9688	0,9664	0,9707	0,9711	0,9711	0,9722
	Me	1,0003	1,0003	1,0002	1,0007	1,0003	1,0005	0,9997	1,0007
	P97,5	1,0328	1,0328	1,0314	1,0332	1,0310	1,0310	1,0295	1,0316
0,50	P2,5	0,8454	0,8433	0,8659	0,8490	0,8493	0,8465	0,8437	0,8597
	Me	1,0016	0,9983	1,0040	1,0015	0,9976	1,0023	0,9982	0,9996
	P97,5	1,1538	1,1403	1,1530	1,1588	1,1582	1,1436	1,1456	1,1381
0,75	P2,5	0,7407	0,7910	0,7586	0,7869	0,7723	0,7601	0,7710	0,7960
	Me	0,9835	0,9981	1,0111	0,9997	0,9918	1,0026	1,0071	0,9999
	P97,5	1,2552	1,2312	1,2532	1,2290	1,2339	1,2510	1,2217	1,2269
1,00	P2,5	0,6854	0,6869	0,6763	0,7296	0,6809	0,6605	0,7010	0,7079
	Me	1,0032	1,0066	1,0188	1,0202	0,9993	0,9826	0,9936	0,9785
	P97,5	1,2912	1,3202	1,3210	1,3096	1,3171	1,2963	1,2787	1,3305
2,70	P2,5	0,0839	0,1602	0,2425	0,1765	0,2473	0,1445	0,1646	0,2029
	Me	1,0044	0,9905	1,0124	1,0137	1,0283	0,9816	1,0413	0,9876
	P97,5	1,9106	1,8071	1,8357	1,8965	1,8155	1,7304	1,7755	1,7923
4,30	P2,5	-0,3969	-0,3338	-0,2855	-0,4156	-0,3593	-0,3099	-0,2927	-0,3387
	Me	1,0700	0,9655	1,0460	1,0417	0,9402	1,0546	1,0146	0,9743
	P97,5	2,4022	2,2722	2,4596	2,3756	2,2163	2,4924	2,4226	2,2040
6,50	P2,5	-1,0292	-1,0017	-1,0416	-1,2470	-1,2010	-1,0977	-0,9006	-1,0317
	Me	0,9435	0,9939	0,9625	1,0007	0,9112	0,9010	0,9918	0,9919
	P97,5	2,9010	3,1249	3,0234	3,1559	3,0064	2,8588	2,8442	2,8386

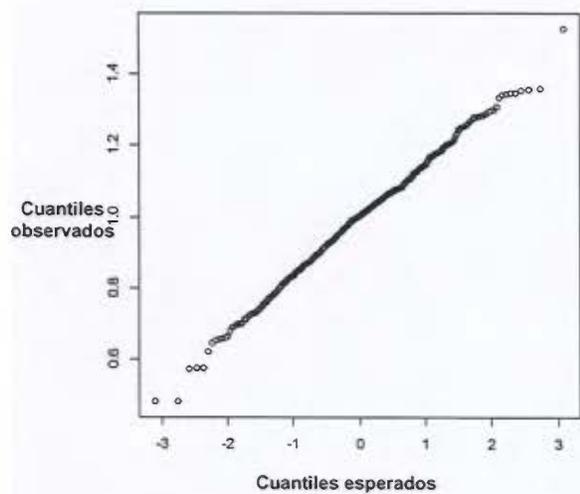
Fuente: Elaboración propia, 2011.

Cuadro 7. Intervalos de estimación del coef. para X_1 según valores teóricos para coef. utilizando distrib. normal con matriz de distancias, 2011

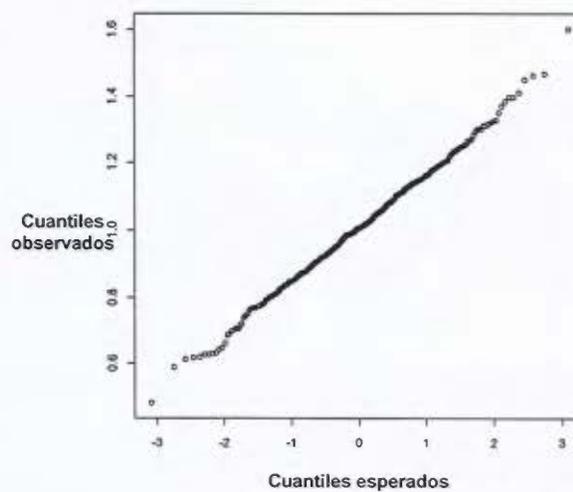
Sigma teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,10	P2,5	0,9724	0,9680	0,9680	0,9707	0,9700	0,9704	0,9695	0,9655
	Me	1,0007	0,9995	0,9999	1,0004	1,0017	1,0012	1,0003	0,9999
	P97,5	1,0314	1,0300	1,0304	1,0272	1,0293	1,0305	1,0333	1,0327
0,50	P2,5	0,8469	0,8420	0,8486	0,8438	0,8515	0,8271	0,8471	0,8437
	Me	0,9966	0,9975	1,0015	1,0048	0,9968	1,0003	1,0017	0,9976
	P97,5	1,1599	1,1726	1,1584	1,1460	1,1457	1,1689	1,1622	1,1497
0,75	P2,5	0,7569	0,7644	0,7682	0,7466	0,7693	0,7672	0,7662	0,7511
	Me	0,9898	0,9908	0,9968	0,9908	0,9945	0,9948	0,9835	1,0019
	P97,5	1,2134	1,2223	1,2463	1,2137	1,2430	1,1977	1,2092	1,2122
1,00	P2,5	0,7262	0,7031	0,7022	0,6252	0,6857	0,6797	0,6929	0,6945
	Me	1,0164	0,9996	1,0057	0,9923	1,0144	0,9934	1,0018	0,9947
	P97,5	1,3182	1,3263	1,3121	1,3008	1,3055	1,3144	1,3398	1,2999
2,70	P2,5	0,2340	0,1426	0,1834	0,1707	0,1365	0,1085	0,0288	0,2284
	Me	0,9747	0,9879	0,9726	0,9961	0,9928	0,9722	1,0254	0,9781
	P97,5	1,8556	1,7604	1,9006	1,8428	1,7226	1,8043	1,8381	1,8723
4,30	P2,5	-0,2316	-0,3411	-0,2970	-0,2672	-0,3825	-0,4582	-0,3601	-0,3534
	Me	1,0318	0,9744	1,0670	1,0121	1,0221	1,0376	0,9441	0,9894
	P97,5	2,3085	2,3176	2,4154	2,3497	2,2186	2,3795	2,3632	2,2756
6,50	P2,5	-0,8986	-1,1205	-0,8430	-0,8626	-0,8625	-1,0014	-1,0182	-0,8151
	Me	1,0333	1,0766	0,9427	0,9985	0,9358	0,9562	1,1687	1,0108
	P97,5	3,0254	3,1647	2,9888	2,8854	2,9196	2,9833	3,3261	2,6986

Fuente: Elaboración propia, 2011.

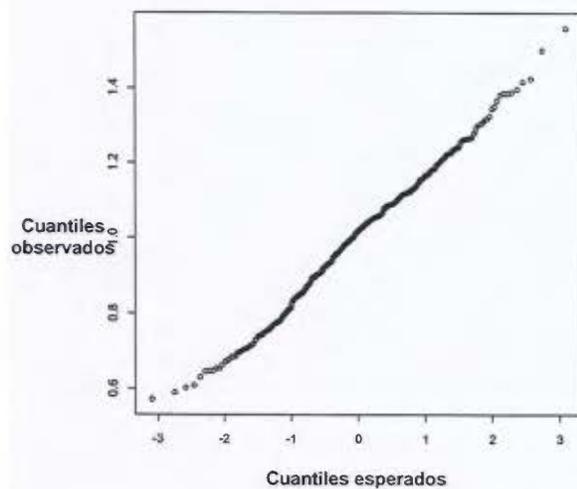
Gráfico 5. Gráfico de normalidad de los coeficientes con distribución normal y matriz K vecinos, 2011



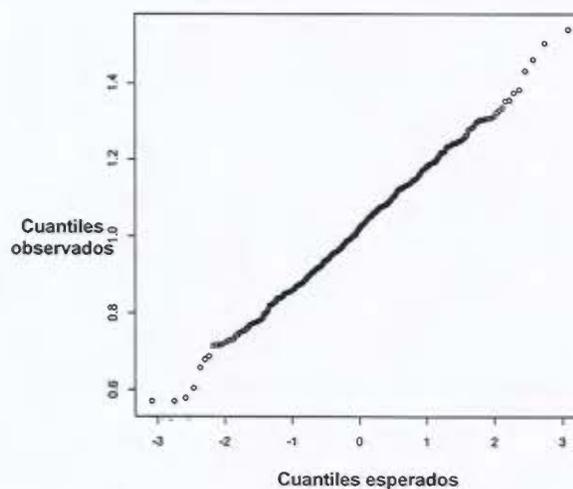
Lambda teórico = 0,01 y sigma teórico = 1



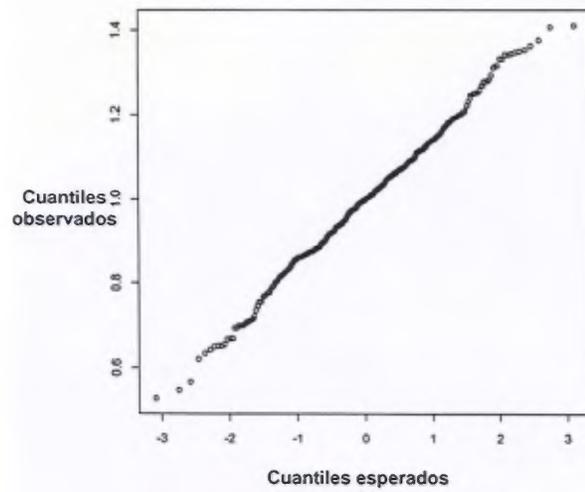
Lambda teórico = 0,10 y sigma teórico = 1



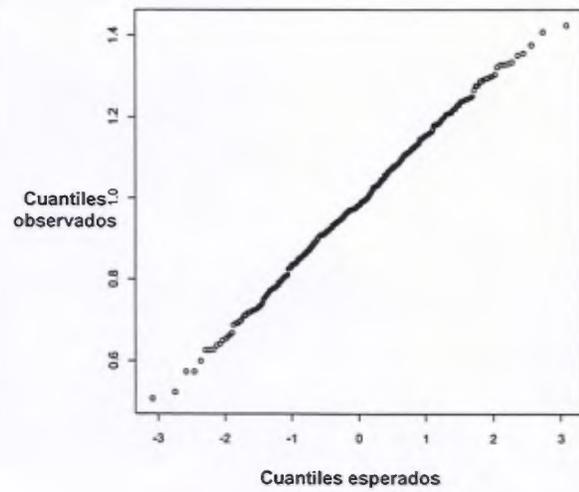
Lambda teórico = 0,23 y sigma teórico = 1



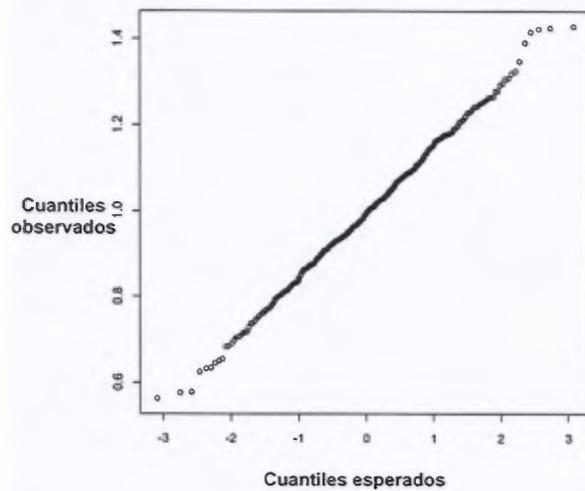
Lambda teórico = 0,36 y sigma teórico = 1



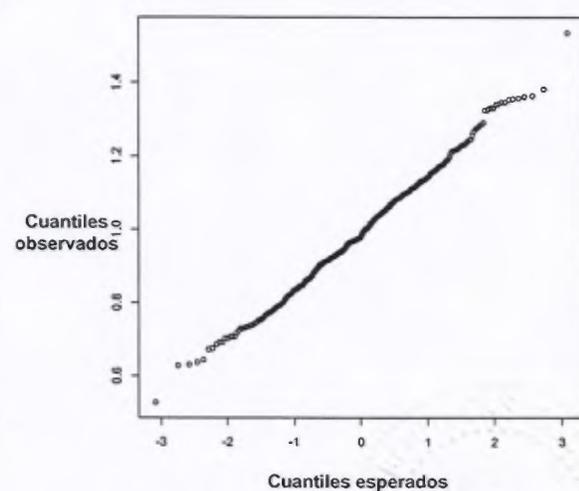
Lambda teórico = 0,49 y sigma teórico = 1



Lambda teórico = 0,62 y sigma teórico = 1



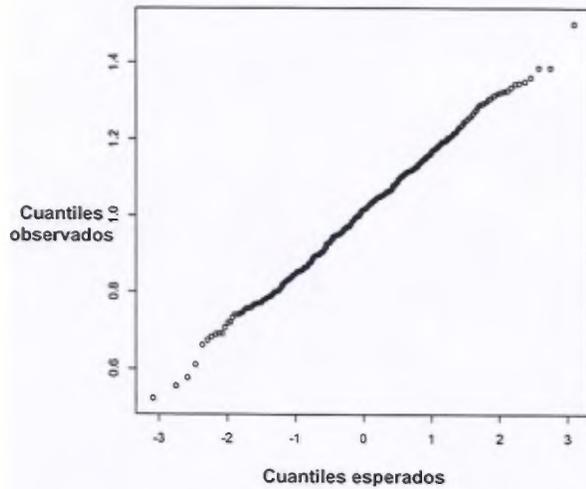
Lambda teórico = 0,75 y sigma teórico = 1



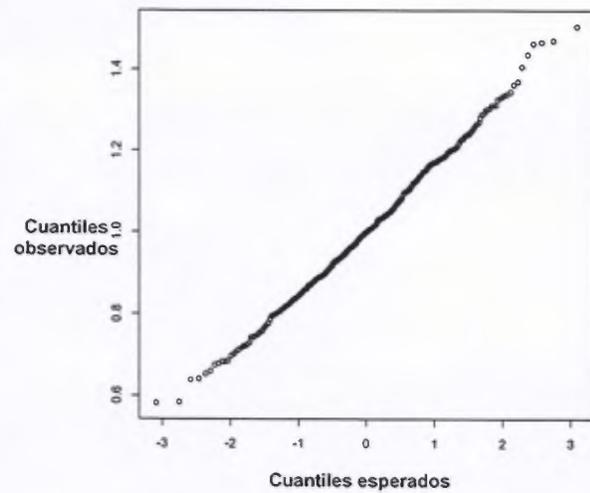
Lambda teórico = 0,90 y sigma teórico = 1

Fuente: Elaboración propia, 2011.

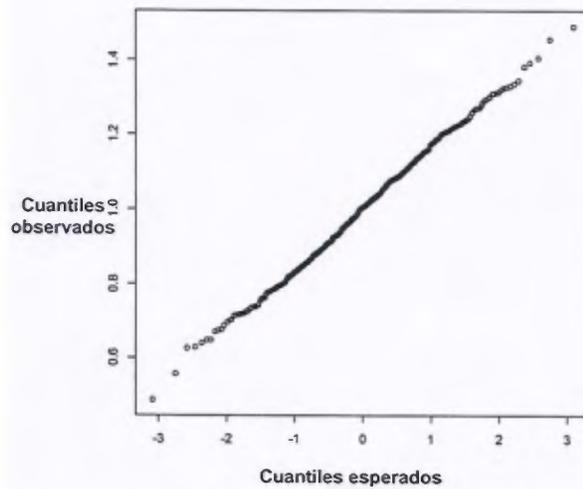
Gráfico 6. Gráfico de normalidad de los coeficientes con distribución normal y matriz de distancias, 2011



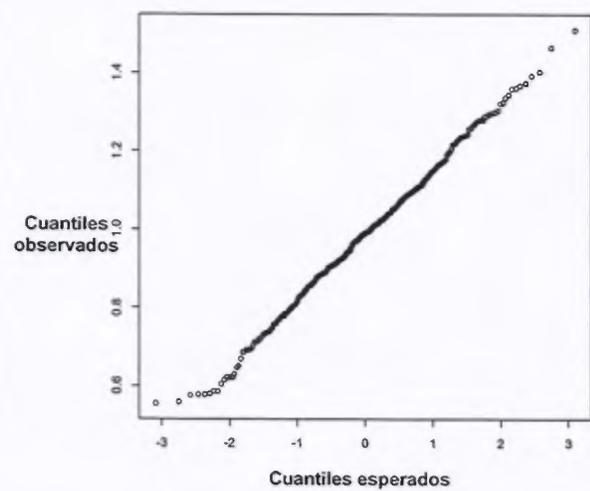
Lambda teórico = 0,01 y sigma teórico = 1



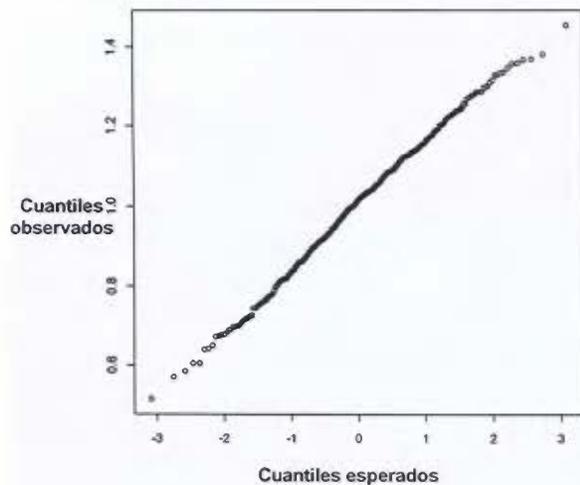
Lambda teórico = 0,10 y sigma teórico = 1



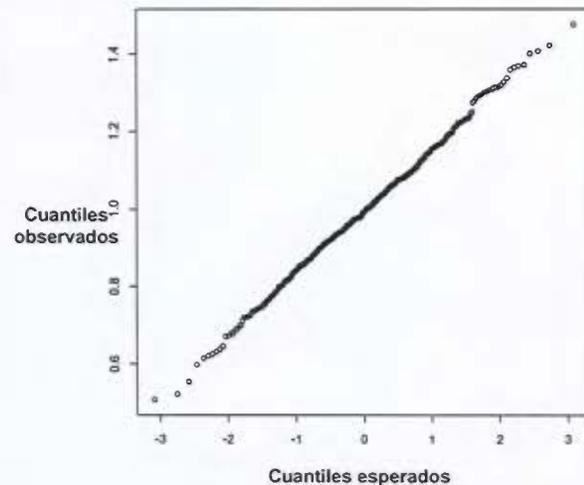
Lambda teórico = 0,23 y sigma teórico = 1



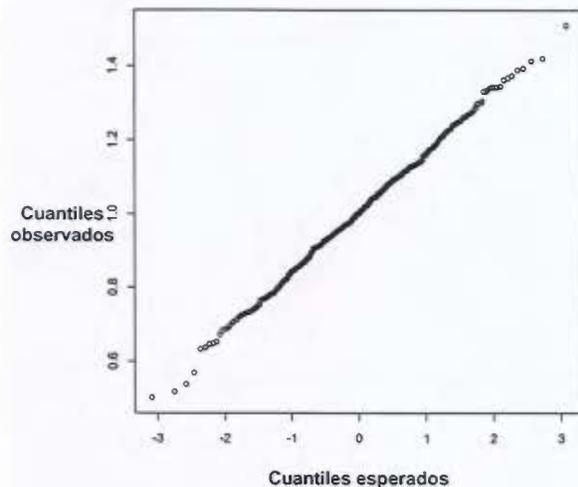
Lambda teórico = 0,36 y sigma teórico = 1



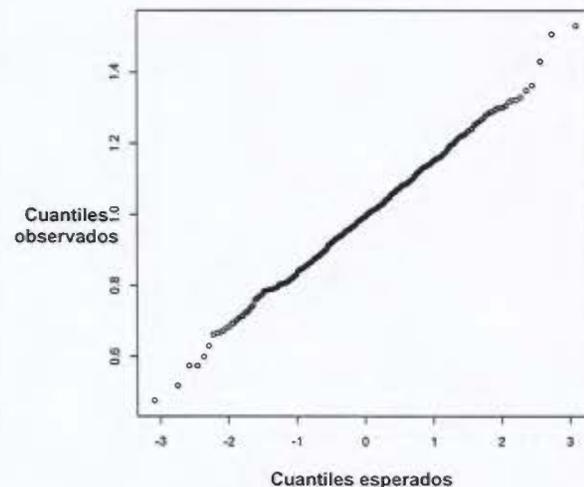
Lambda teórico = 0,49 y sigma teórico = 1



Lambda teórico = 0,62 y sigma teórico = 1



Lambda teórico = 0,75 y sigma teórico = 1



Lambda teórico = 0,90 y sigma teórico = 1

Fuente: Elaboración propia, 2011.

Por otra parte, el modelo simulado con una matriz de pesos de distancias necesita un valor de lambda más alto que el modelo simulado con una matriz de pesos K vecinos más cercanos para poder inferir la existencia de autocorrelación espacial observando la tasa de rechazo obtenida para cada uno de los valores del lambda calculado. Es decir, en el modelo para lambda con distribución normal utilizando una matriz de distancias se puede afirmar que existe una potencia mayor al 93% de detectar a lambda como significativamente distinto de cero cuando el lambda teórico es mayor a 0,49. En cambio, si se utiliza la matriz de

pesos K vecinos más cercanos, el valor de la potencia baja al 77% y los valores del lambda teórico que permiten concluir sobre la existencia de autocorrelación espacial deben ser mayores a 0,23 resultando el lambda teórico menor al obtenido usando una matriz de distancias (Cuadros 8 y 9). Para finalizar, en ambas matrices de pesos queda demostrado que hay un valor umbral relativamente bajo de los lambdas (entre 0,23 en la matriz K vecinos más cercanos y 0,36 en la matriz de distancias) para poder detectarlo como significativo con una potencia alta.

Cuadro 8. Porcentaje de corridas para identificar la potencia de prueba con distribución normal utilizando matriz de distancias, 2011

Sigma teórico	Lambda teórico							
	0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,10	0,034	0,082	0,372	0,730	0,934	1,000	1,000	1,000
0,50	0,054	0,086	0,318	0,686	0,962	0,998	1,000	1,000
0,75	0,046	0,068	0,308	0,702	0,946	0,992	1,000	1,000
1,00	0,068	0,076	0,290	0,694	0,950	0,994	1,000	1,000
2,70	0,056	0,096	0,340	0,714	0,934	0,996	1,000	1,000
4,30	0,062	0,084	0,320	0,698	0,940	0,992	1,000	1,000
6,50	0,050	0,092	0,344	0,716	0,948	0,996	1,000	1,000

Fuente: Elaboración propia, 2011.

Cuadro 9. Porcentaje de corridas para identificar la potencia de prueba con distribución normal utilizando matriz K vecinos, 2011

Sigma teórico	Lambda teórico							
	0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,10	0,050	0,200	0,772	0,994	1,000	1,000	1,000	1,000
0,50	0,046	0,212	0,796	0,996	1,000	1,000	1,000	1,000
0,75	0,054	0,210	0,794	0,998	1,000	1,000	1,000	1,000
1,00	0,058	0,182	0,816	0,996	1,000	1,000	1,000	1,000
2,70	0,044	0,210	0,800	0,990	1,000	1,000	1,000	1,000
4,30	0,046	0,196	0,784	0,996	1,000	1,000	1,000	1,000
6,50	0,038	0,236	0,770	0,996	1,000	1,000	1,000	1,000

Fuente: Elaboración propia, 2011.

5. 3 Simulaciones para evaluar la robustez del modelo SAR ante la falta de normalidad

5. 3. 1 Metodología

En el análisis de la robustez del modelo ante la violación del supuesto de normalidad se simularon residuos de distribuciones beta con distintos niveles de asimetría para observar si el comportamiento de los datos obtenidos difiere de los resultados del modelo simulado usando una distribución uniforme. En ese sentido, la distribución del modelo utilizada es la distribución beta para los tres diferentes tipos de matrices de pesos: reina o *queen*, K vecinos más cercanos y de distancias. Los parámetros del beta teórico utilizados para la distribución beta fueron los valores de 0,05; 0,20; 0,30; 0,40 y 5,00; manteniendo el primer parámetro fijo con un valor de 0,30 de forma arbitraria y el segundo parámetro fue modificándose simultáneamente para cada uno de los valores utilizados para la

distribución beta, propuestos anteriormente. Estos valores significan diferentes asimetrías tomando en cuenta cada uno de los valores del lambda teórico.

Una vez realizadas las simulaciones se comparan los resultados obtenidos tomando en cuenta las distintas matrices de pesos. Esta comparación permite identificar si las estimaciones están cercanas a los parámetros teóricos establecidos, tomando en cuenta específicamente la estructura espacial de los distritos de Costa Rica para el periodo de estudio delimitado en esta investigación.

5. 3. 2 Análisis

Primeramente es necesario referirse a los niveles de asimetría que se generan a partir de los valores de lambda y beta teóricos utilizando una distribución beta con matriz de pesos reina o *queen*. Los resultados señalan que las simulaciones de los residuos de distribuciones beta con distintos niveles de asimetría para lambda y beta teóricos en los tres tipos de matrices de pesos son similares entre sí tomando en cuenta los beta teóricos de 0,20; 0,30 y 0,40. Solamente existen diferencias que son leves en los valores de asimetría para los beta teóricos de 0,05 y 5,00 que son los extremos de los valores seleccionados para las simulaciones (Cuadro 10).

Cuadro 10. Niveles de asimetría para lambda y beta teóricos utilizando distribución beta con matriz *queen*, 2011

Beta teórico	Lambda teórico							
	0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,05	-1,09	-1,08	-1,09	-1,08	-1,06	-1,05	-0,99	-0,76
0,20	-0,22	-0,22	-0,21	-0,22	-0,22	-0,20	-0,20	-0,15
0,30	-0,01	0,00	0,01	0,00	-0,01	0,00	0,00	0,01
0,40	0,15	0,16	0,15	0,15	0,15	0,15	0,13	0,13
5,00	1,40	1,40	1,39	1,39	1,36	1,34	1,25	0,96

Fuente: Elaboración propia, 2011.

Por su parte, con cualquiera de las tres matrices analizadas todos los intervalos de confianza simulados incluyen el valor del lambda teórico seleccionado y se observa que cuanto mayor es el valor del lambda teórico mejor es la estimación del lambda calculado siendo más precisa, porque los intervalos de confianza simulados son más pequeños; no obstante, para el modelo con matriz de distancias, la amplitud de sus intervalos de confianza simulados es más grande que lo obtenido utilizando las matrices reina o *queen* y K vecinos más cercanos, tal y como se observó en los análisis con errores normales (Gráficos 7, 8 y 9).

Además, tanto para la matriz reina o *queen* como para la matriz K vecinos más cercanos es a partir del lambda teórico de 0,23 que los intervalos a un 95% de confianza no incluyen el valor de cero lo cual indica que existe suficiente potencia en las estimaciones para detectar autocorrelación espacial con los datos distritales de Costa Rica (Cuadros 11 y 12). En el caso de la matriz de distancias, esta situación ocurre a partir del lambda teórico de 0,36. Cabe señalar que la amplitud de los intervalos es similar a la de los modelos con errores normales (Cuadro 13).

Cuadro 11. Intervalos de la estimación del lambda según valores teóricos utilizando distribución beta con matriz *queen*, 2011

Beta teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,05	P2,5	-0,1568	-0,0596	0,0868	0,2153	0,3671	0,5102	0,6604	0,8482
	Me	0,0044	0,0949	0,2252	0,3559	0,4794	0,6109	0,7484	0,8965
	P97,5	0,1750	0,2355	0,3557	0,4808	0,5846	0,6967	0,8069	0,9311
0,20	P2,5	-0,1512	-0,0529	0,0803	0,2282	0,3679	0,5011	0,6688	0,8362
	Me	0,0070	0,0997	0,2253	0,3533	0,4863	0,6171	0,7472	0,8960
	P97,5	0,1560	0,2306	0,3557	0,4816	0,5822	0,7079	0,8151	0,9306
0,30	P2,5	-0,1448	-0,0568	0,0667	0,2259	0,3683	0,5060	0,6519	0,8445
	Me	0,0022	0,0988	0,2185	0,3579	0,4800	0,6173	0,7463	0,8963
	P97,5	0,1400	0,2326	0,3492	0,4718	0,5840	0,6990	0,8126	0,9325
0,40	P2,5	-0,1419	-0,0655	0,0845	0,2244	0,3600	0,5100	0,6481	0,8378
	Me	0,0053	0,0940	0,2250	0,3474	0,4790	0,6149	0,7447	0,8940
	P97,5	0,1380	0,2223	0,3533	0,4743	0,5807	0,7028	0,8160	0,9314
5,00	P2,5	-0,1387	-0,0476	0,0839	0,2200	0,3587	0,5026	0,6690	0,8453
	Me	0,0012	0,1027	0,2259	0,3575	0,4803	0,6071	0,7426	0,8942
	P97,5	0,1541	0,2413	0,3512	0,4751	0,5843	0,7001	0,8122	0,9309

Fuente: Elaboración propia, 2011.

Cuadro 12. Intervalos de la estimación del lambda, según valores teóricos, utilizando distribución beta con matriz K vecinos, 2011

Beta teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,05	P2,5	-0,1877	-0,0733	0,0616	0,2129	0,3493	0,5044	0,6620	0,8491
	Me	0,0090	0,0854	0,2292	0,3487	0,4806	0,6129	0,7464	0,8956
	P97,5	0,1607	0,2326	0,3550	0,4754	0,5930	0,7009	0,8106	0,9273
0,20	P2,5	-0,1845	-0,0887	0,0508	0,2033	0,3523	0,5089	0,6667	0,8442
	Me	0,0088	0,0829	0,2173	0,3500	0,4827	0,6199	0,7457	0,8968
	P97,5	0,1573	0,2355	0,3494	0,4770	0,5821	0,7034	0,8078	0,9255
0,30	P2,5	-0,1669	-0,0847	0,0385	0,2122	0,3546	0,5007	0,6696	0,8490
	Me	-0,0032	0,0957	0,2199	0,3482	0,4877	0,6139	0,7448	0,8958
	P97,5	0,1566	0,2441	0,3646	0,4627	0,5965	0,6936	0,8093	0,9269
0,40	P2,5	-0,1662	-0,0767	0,0677	0,2203	0,3526	0,5094	0,6536	0,8444
	Me	-0,0035	0,0933	0,2231	0,3542	0,4900	0,6183	0,7439	0,8944
	P97,5	0,1552	0,2393	0,3720	0,4660	0,5944	0,7042	0,8075	0,9273
5,00	P2,5	-0,1558	-0,0821	0,0742	0,1952	0,3528	0,5120	0,6505	0,8439
	Me	0,0004	0,0862	0,2117	0,3548	0,4821	0,6157	0,7421	0,8960
	P97,5	0,1624	0,2472	0,3721	0,4849	0,5865	0,7106	0,8084	0,9300

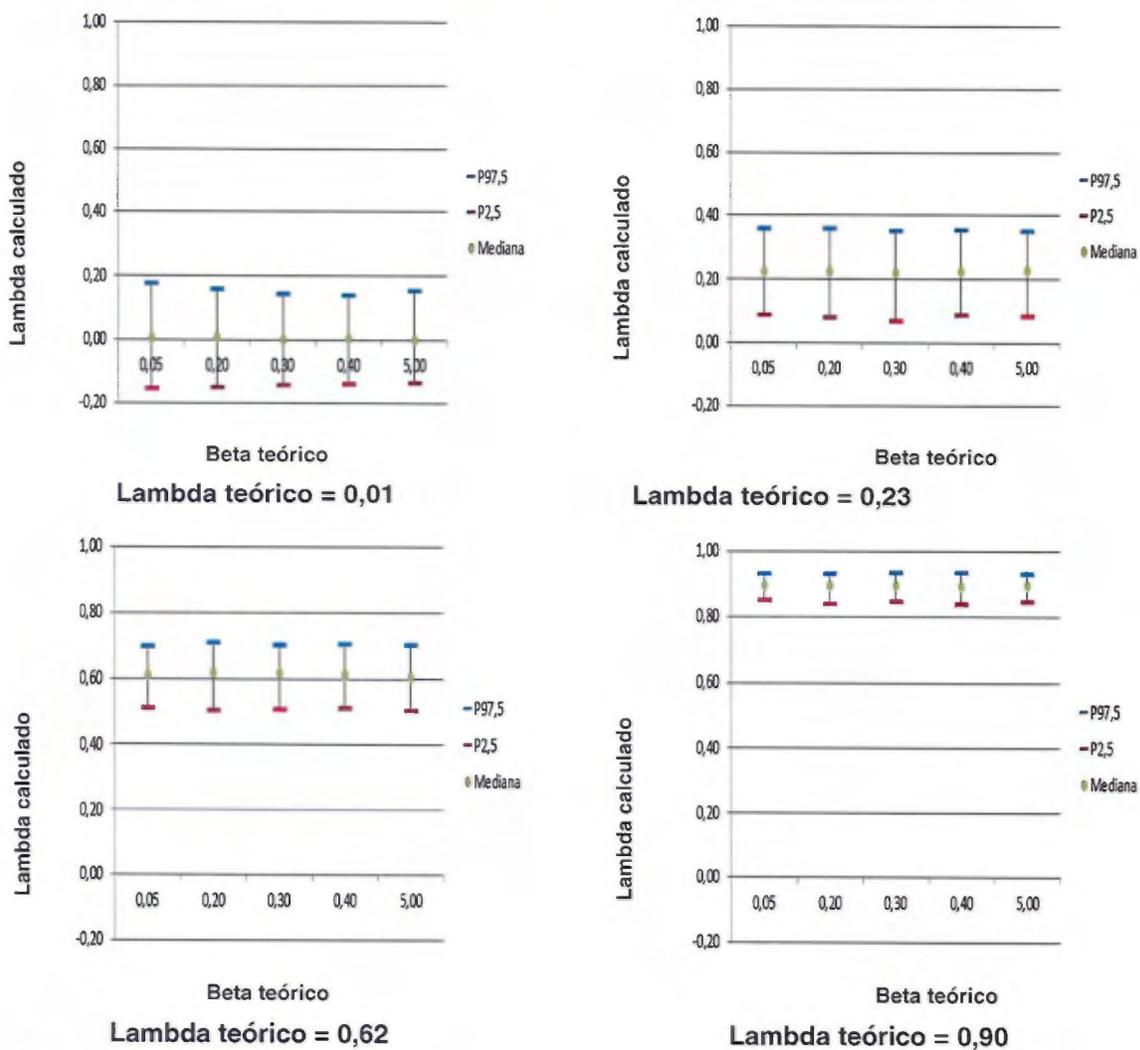
Fuente: Elaboración propia, 2011.

Cuadro 13. Intervalos de la estimación del lambda, según valores teóricos, utilizando distribución beta con matriz de distancias, 2011

Beta teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,05	P2,5	-0,3525	-0,2360	-0,1090	0,0315	0,2390	0,3871	0,5708	0,7602
	Me	-0,0340	0,0585	0,1939	0,3215	0,4581	0,5924	0,7284	0,8821
	P97,5	0,2415	0,3356	0,4550	0,5397	0,6377	0,7420	0,8341	0,9449
0,20	P2,5	-0,3198	-0,2394	-0,1246	0,0532	0,1820	0,3820	0,5304	0,7686
	Me	-0,0086	0,0721	0,2092	0,3244	0,4574	0,5950	0,7316	0,8809
	P97,5	0,2692	0,3115	0,4312	0,5443	0,6244	0,7355	0,8352	0,9394
0,30	P2,5	-0,3240	-0,2469	-0,1512	0,0311	0,2122	0,3975	0,5356	0,7738
	Me	-0,0162	0,0609	0,1861	0,3279	0,4631	0,5997	0,7261	0,8855
	P97,5	0,2572	0,3373	0,4441	0,5221	0,6444	0,7324	0,8370	0,9399
0,40	P2,5	-0,3234	-0,3094	-0,0878	0,0509	0,2091	0,3617	0,5363	0,7706
	Me	-0,0231	0,0603	0,1902	0,3289	0,4621	0,5950	0,7296	0,8849
	P97,5	0,2471	0,3385	0,4360	0,5299	0,6238	0,7408	0,8303	0,9388
5,00	P2,5	-0,3520	-0,2197	-0,1064	0,0587	0,2173	0,3684	0,5220	0,7696
	Me	-0,0344	0,0605	0,1924	0,3127	0,4577	0,5868	0,7186	0,8788
	P97,5	0,2445	0,3287	0,4359	0,5641	0,6437	0,7506	0,8515	0,9472

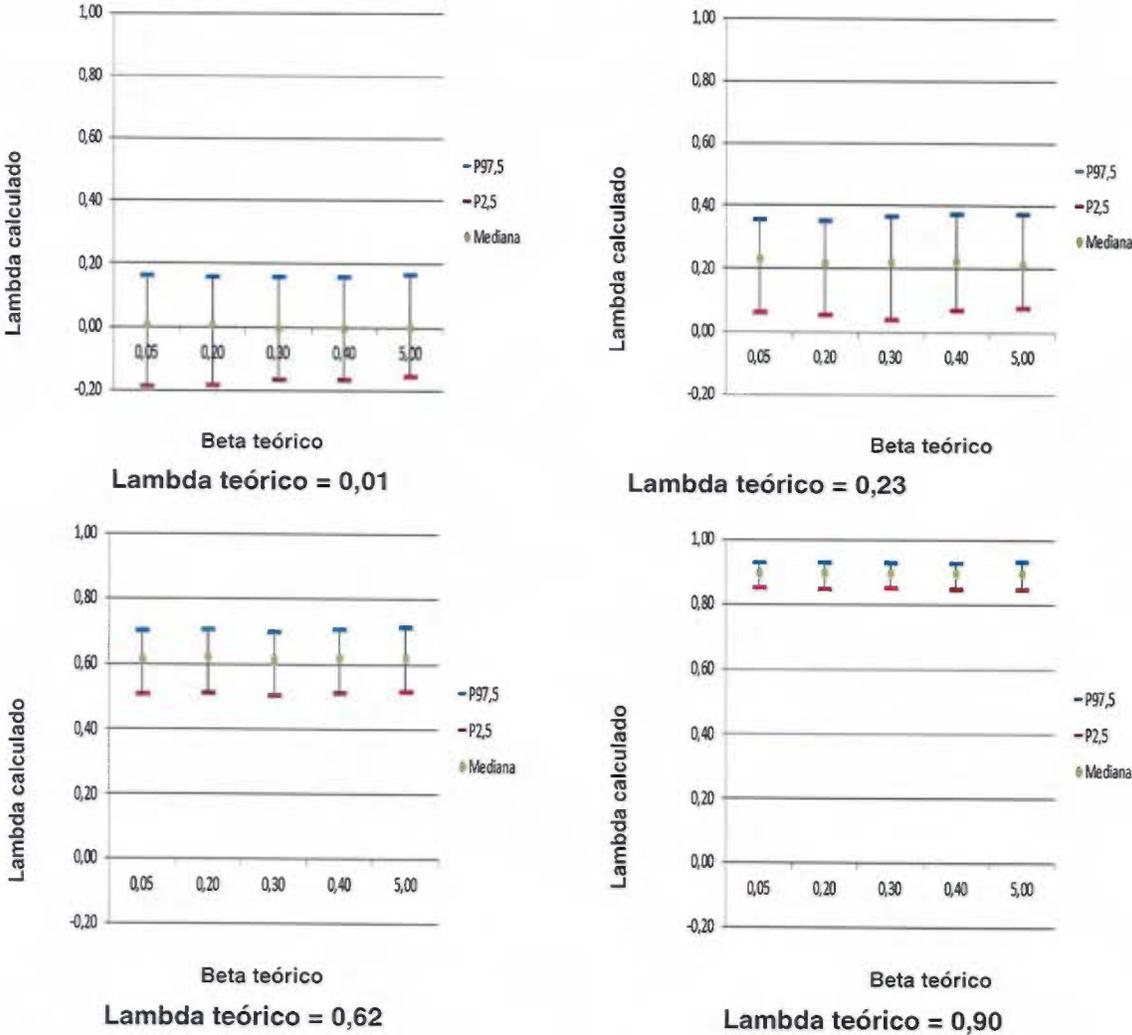
Fuente: Elaboración propia, 2011.

Gráfico 7. Intervalos de confianza simulados para lambda, utilizando distribución beta con matriz *queen*, 2011



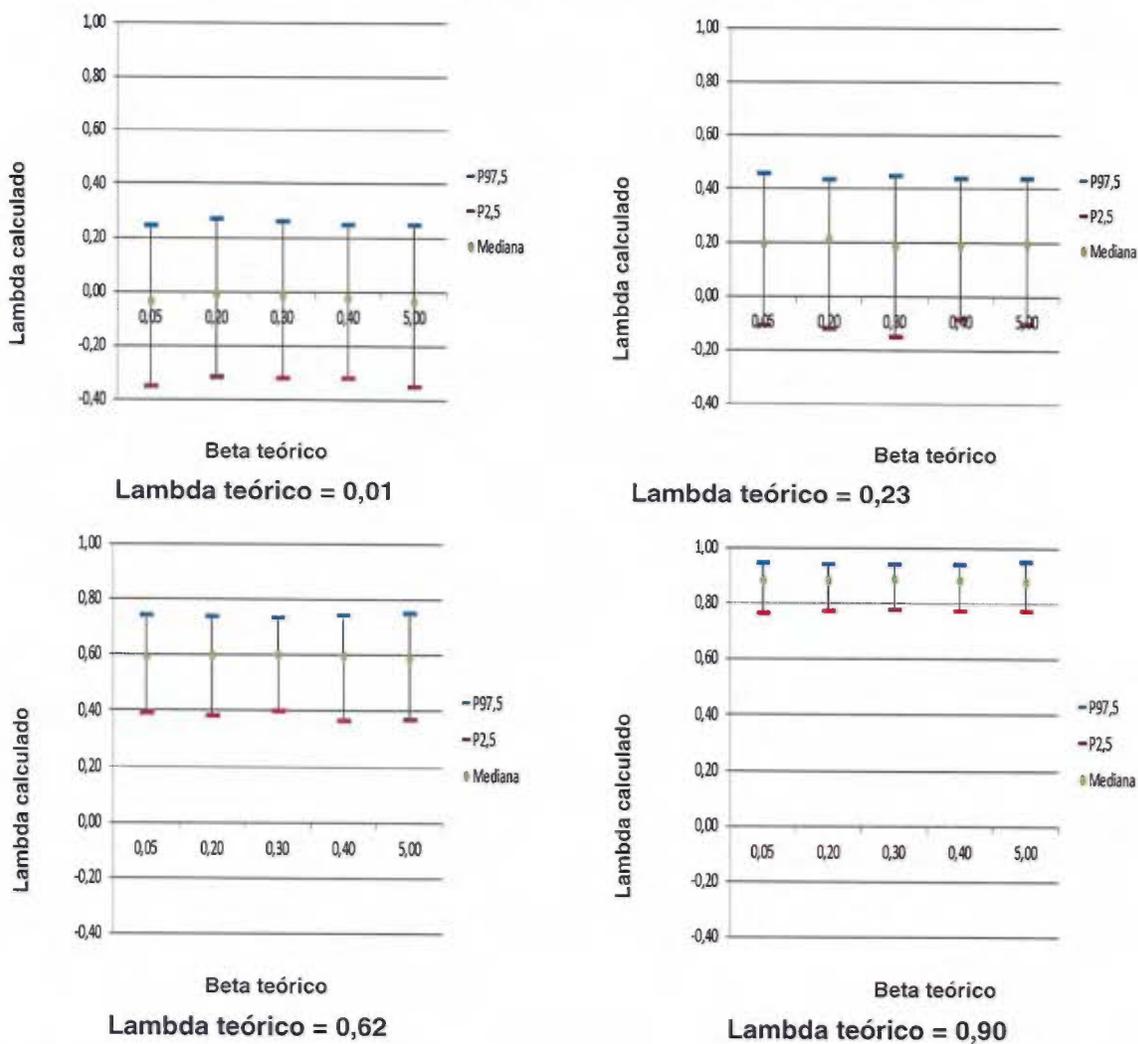
Fuente: Elaboración propia, 2011.

Gráfico 8. Intervalos de confianza simulados para lambda, utilizando distribución beta con matriz K vecinos, 2011



Fuente: Elaboración propia, 2011.

Gráfico 9. Intervalos de confianza simulados para lambda, utilizando distribución beta con matriz de distancias, 2011



Fuente: Elaboración propia, 2011.

Por otra parte, en los coeficientes de los modelos SAR simulados con los tres tipos de matrices se obtiene que todos los intervalos de confianza simulados contienen al valor 1, por lo tanto se consiguen los resultados esperados usando cualquier matriz de pesos. De igual forma, con cualquier valor de beta teórico existen coeficientes significativos ya que los intervalos de confianza simulados no incluyen el valor 0. Además, los intervalos de confianza simulados mantienen una amplitud similar con distintos valores de beta teórico lo cual indica que la precisión

es prácticamente la misma para cada uno de ellos en las tres matrices de pesos utilizadas para las simulaciones (Cuadros 14, 15 y 16).

Cuadro 14. Intervalos de estimación del coeficiente para X_1 según valores teóricos para coeficientes, utilizando distribución beta con matriz *queen*, 2011

Beta teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,05	P2,5	0,6786	0,6817	0,6749	0,6564	0,6493	0,7130	0,6976	0,6963
	Me	0,9957	0,9991	1,0047	1,0059	0,9863	0,9980	0,9978	0,9917
	P97,5	1,3238	1,3476	1,3109	1,2955	1,2924	1,3192	1,3066	1,3166
0,20	P2,5	0,7045	0,6935	0,6687	0,6967	0,7346	0,7215	0,6952	0,7313
	Me	0,9944	1,0066	0,9983	0,9856	0,9896	0,9946	0,9801	0,9896
	P97,5	1,2954	1,3097	1,2871	1,3199	1,3152	1,3173	1,2998	1,2782
0,30	P2,5	0,6919	0,6567	0,6706	0,7233	0,6972	0,7056	0,6852	0,6912
	Me	1,0011	0,9917	0,9910	0,9927	0,9978	0,9915	0,9999	1,0022
	P97,5	1,3048	1,3151	1,3102	1,3054	1,3162	1,3012	1,2587	1,2687
0,40	P2,5	0,6532	0,7076	0,7087	0,7085	0,6970	0,7091	0,6918	0,6982
	Me	1,0025	1,0011	0,9966	1,0031	1,0172	1,0092	1,0083	0,9958
	P97,5	1,2910	1,3287	1,3041	1,3031	1,3219	1,3007	1,2981	1,2495
5,00	P2,5	0,6922	0,7001	0,6690	0,7002	0,7167	0,6845	0,7227	0,7285
	Me	1,0024	0,9986	1,0148	0,9998	1,0041	0,9857	0,9984	1,0013
	P97,5	1,2917	1,3502	1,3102	1,2990	1,3227	1,3046	1,2850	1,2821

Fuente: Elaboración propia, 2011.

Cuadro 15. Intervalos de estimación del coeficiente para X_1 según valores teóricos para coeficientes, utilizando distribución beta con matriz K vecinos, 2011

Beta teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,05	P2,5	0,6973	0,6512	0,6985	0,6773	0,6853	0,6651	0,7300	0,6807
	Me	0,9942	1,0028	1,0215	0,9851	1,0129	1,0014	1,0242	0,9986
	P97,5	1,2999	1,3159	1,3006	1,3031	1,3016	1,3102	1,3524	1,2873
0,20	P2,5	0,6379	0,7186	0,7296	0,7010	0,6785	0,7349	0,6668	0,7210
	Me	1,0057	0,9829	1,0033	0,9969	0,9904	0,9911	1,0033	0,9883
	P97,5	1,3082	1,2851	1,3252	1,2763	1,3228	1,3168	1,2871	1,2720
0,30	P2,5	0,6757	0,6826	0,6846	0,6970	0,7224	0,6645	0,6733	0,7102
	Me	0,9922	1,0069	1,0102	1,0037	1,0040	0,9998	1,0149	0,9827
	P97,5	1,3046	1,3333	1,3238	1,3002	1,3043	1,3301	1,3106	1,2630
0,40	P2,5	0,6749	0,6937	0,6674	0,6683	0,6708	0,7326	0,7036	0,6994
	Me	1,0018	1,0079	1,0188	1,0163	1,0053	1,0157	1,0014	1,0068
	P97,5	1,2923	1,3145	1,2980	1,2932	1,3072	1,3043	1,2774	1,3119
5,00	P2,5	0,7034	0,6821	0,6953	0,7002	0,6848	0,6910	0,7377	0,7117
	Me	0,9986	1,0063	1,0039	0,9840	0,9978	1,0023	1,0077	0,9925
	P97,5	1,2869	1,3131	1,3242	1,2868	1,2766	1,3049	1,3177	1,3091

Fuente: Elaboración propia, 2011.

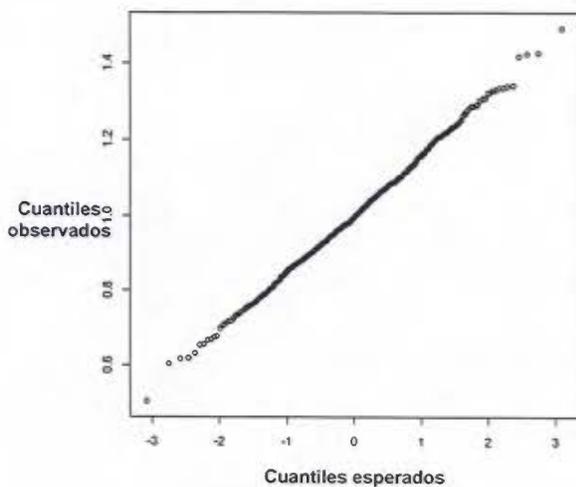
Cuadro 16. Intervalos de estimación del coeficiente para X_1 según valores teóricos para coeficientes utilizando distribución beta con matriz de distancias, 2011

Beta teórico	Percentiles	Lambda teórico							
		0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,05	P2,5	0,6442	0,6855	0,7060	0,7006	0,7054	0,6762	0,7037	0,6766
	Me	0,9984	1,0084	1,0080	1,0046	1,0047	0,9863	1,0016	0,9982
	P97,5	1,3231	1,2674	1,3044	1,3099	1,3354	1,3122	1,2927	1,2729
0,20	P2,5	0,7047	0,7084	0,6751	0,6634	0,6819	0,6584	0,6965	0,7015
	Me	0,9970	1,0093	1,0025	0,9983	0,9905	0,9877	0,9982	1,0047
	P97,5	1,3410	1,3184	1,3275	1,2837	1,2955	1,2715	1,3127	1,3230
0,30	P2,5	0,7045	0,6808	0,6686	0,7221	0,6979	0,6768	0,6681	0,6747
	Me	0,9926	0,9876	0,9860	0,9902	0,9943	0,9955	0,9957	0,9970
	P97,5	1,2899	1,3032	1,2907	1,3344	1,3251	1,2913	1,3191	1,3365
0,40	P2,5	0,6586	0,6752	0,6966	0,7134	0,6501	0,6955	0,6871	0,6806
	Me	0,9995	1,0059	1,0155	1,0043	0,9993	0,9937	1,0069	0,9861
	P97,5	1,3013	1,2975	1,3164	1,3487	1,2977	1,3188	1,3249	1,2939
5,00	P2,5	0,6773	0,6704	0,6873	0,6863	0,6824	0,6751	0,7093	0,6881
	Me	1,0127	0,9998	0,9815	0,9991	0,9995	1,0065	1,0067	1,0054
	P97,5	1,3035	1,3306	1,2972	1,3382	1,2891	1,3154	1,2665	1,3282

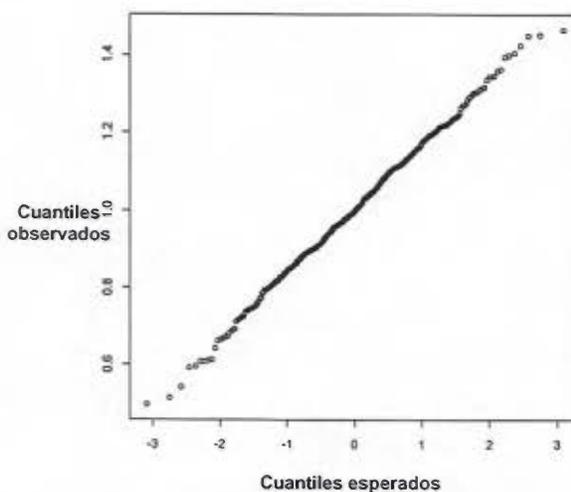
Fuente: Elaboración propia, 2011.

Los gráficos de normalidad *qqplot* indican que los coeficientes tienen un comportamiento normal porque los puntos de los gráficos forman una línea recta para todos los valores de lambda, confirmándose el supuesto de normalidad de los coeficientes independientemente del tipo de matriz de pesos utilizada para las simulaciones (Gráficos 10, 11 y 12).

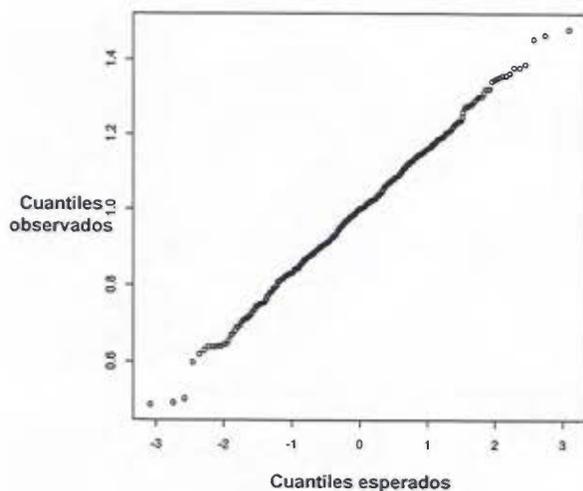
Gráfico 10. Gráfico de normalidad de los coeficientes con distribución beta y matriz *queen*, 2011



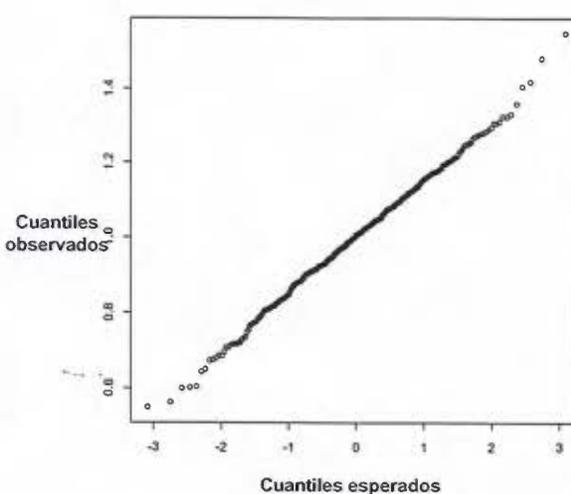
Lambda teórico = 0,01 y beta teórico = 0,05



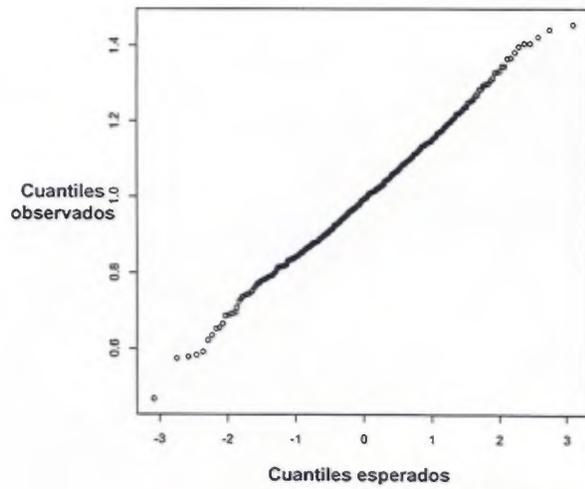
Lambda teórico = 0,01 y beta teórico = 5,00



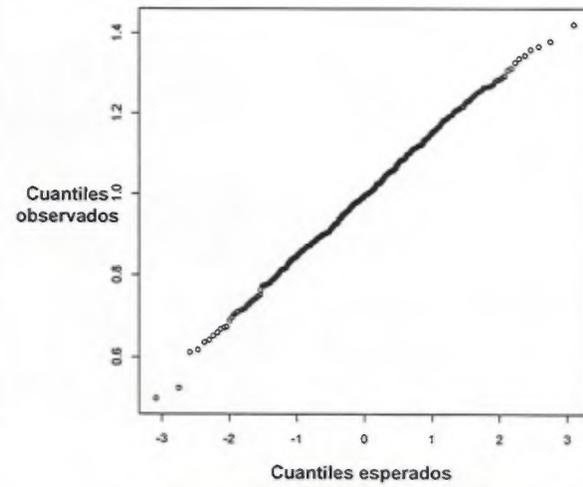
Lambda teórico = 0,10 y beta teórico = 0,05



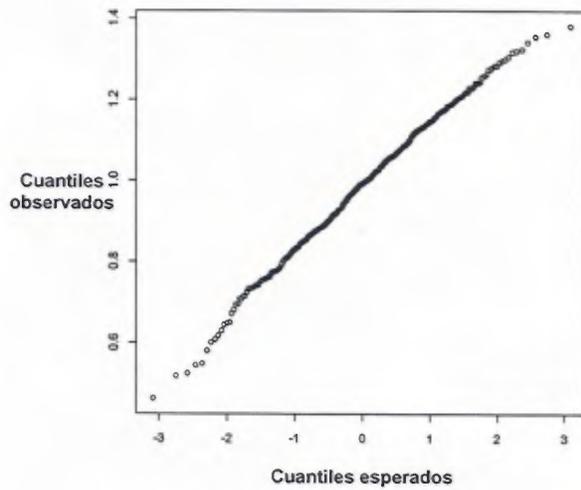
Lambda teórico = 0,10 y beta teórico = 5,00



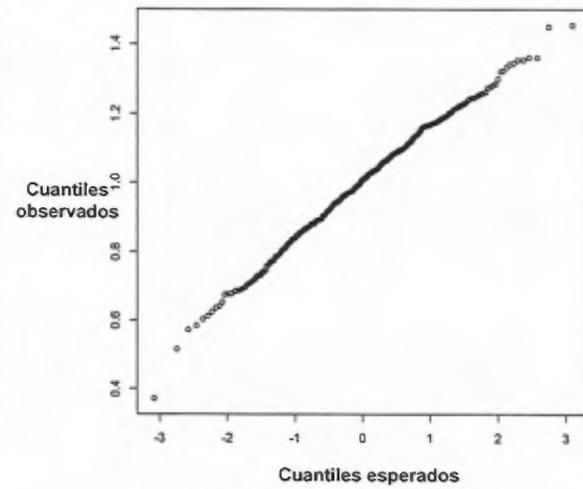
Lambda teórico = 0,23 y beta teórico = 0,05



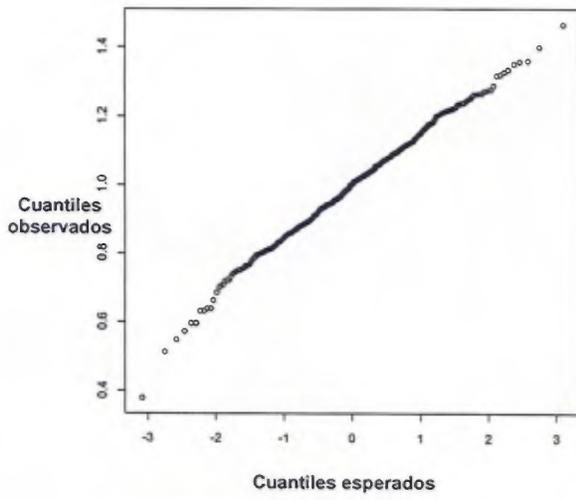
Lambda teórico = 0,23 y beta teórico = 5,00



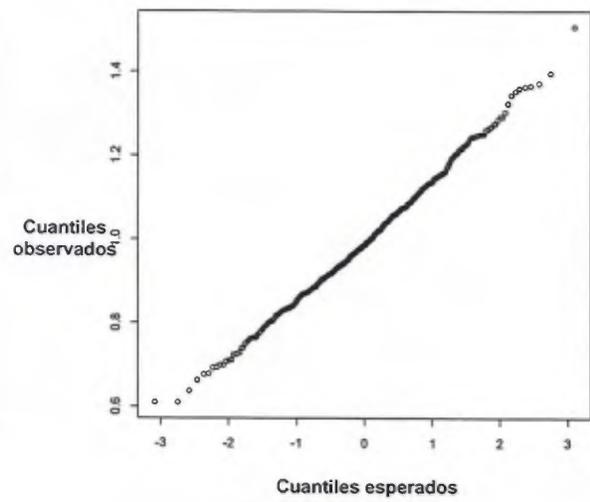
Lambda teórico = 0,36 y beta teórico = 0,05



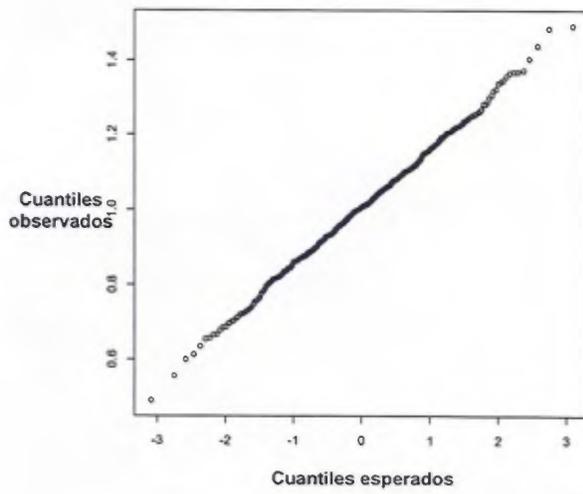
Lambda teórico = 0,36 y beta teórico = 5,00



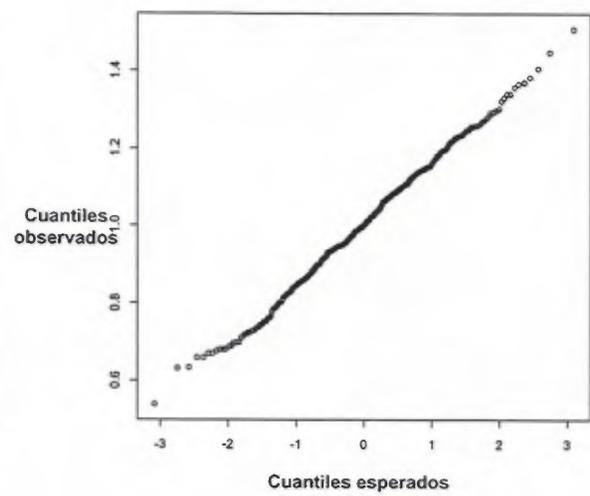
Lambda teórico = 0,49 y beta teórico = 0,05



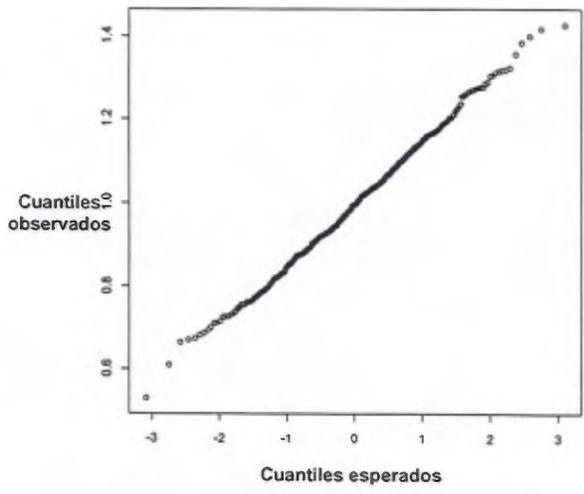
Lambda teórico = 0,49 y beta teórico = 5,00



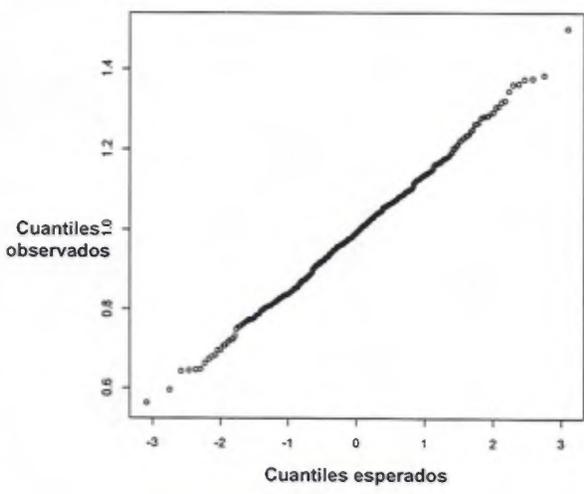
Lambda teórico = 0,62 y beta teórico = 0,05



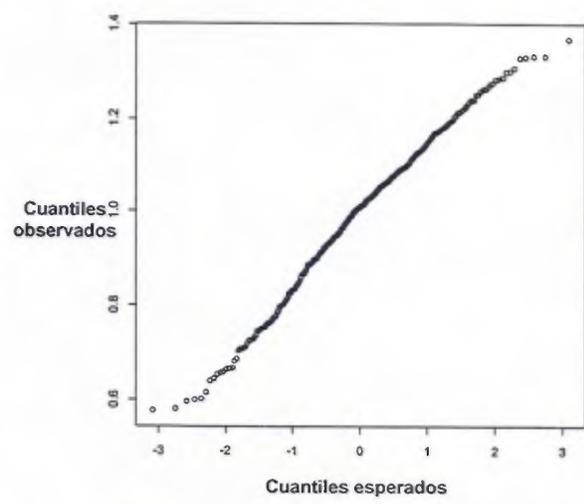
Lambda teórico = 0,62 y beta teórico = 5,00



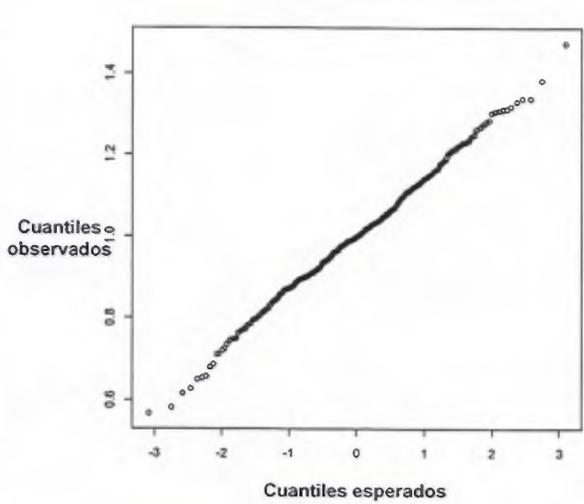
Lambda teórico = 0,75 y beta teórico = 0,05



Lambda teórico = 0,75 y beta teórico = 5,00



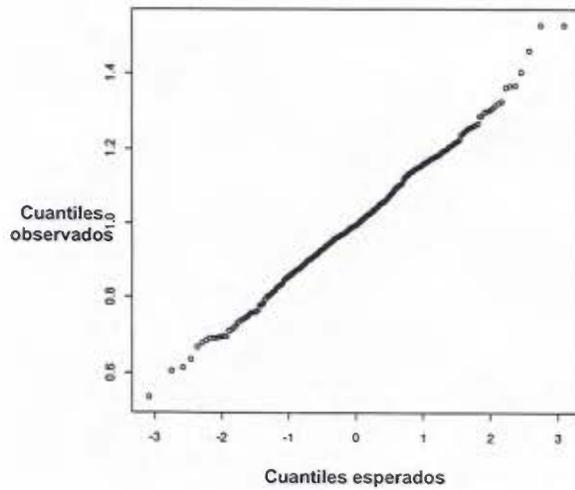
Lambda teórico = 0,90 y beta teórico = 0,05



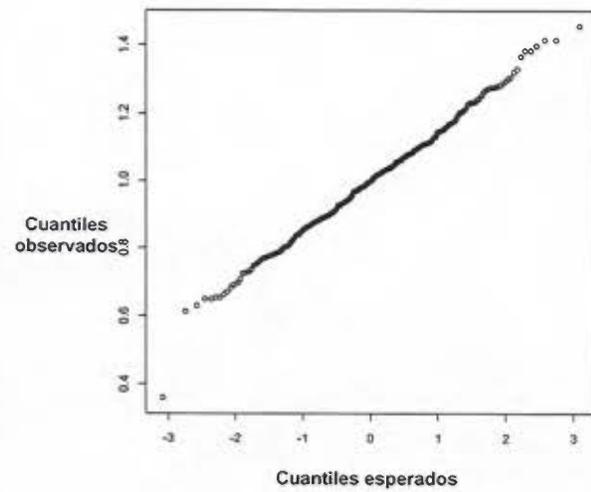
Lambda teórico = 0,90 y beta teórico = 5,00

Fuente: Elaboración propia, 2011.

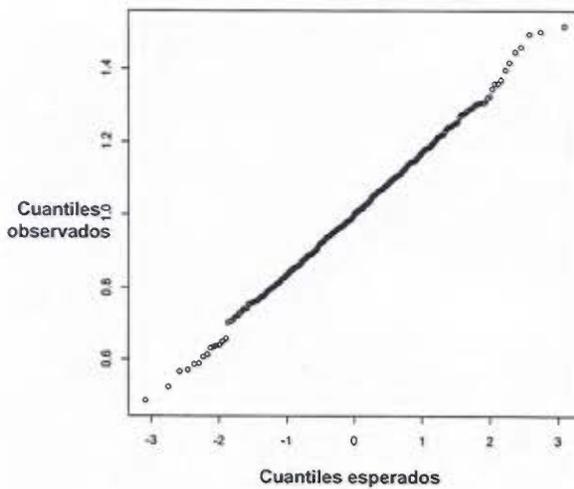
Gráfico 11. Gráfico de normalidad de los coeficientes con distribución beta y matriz K vecinos, 2011



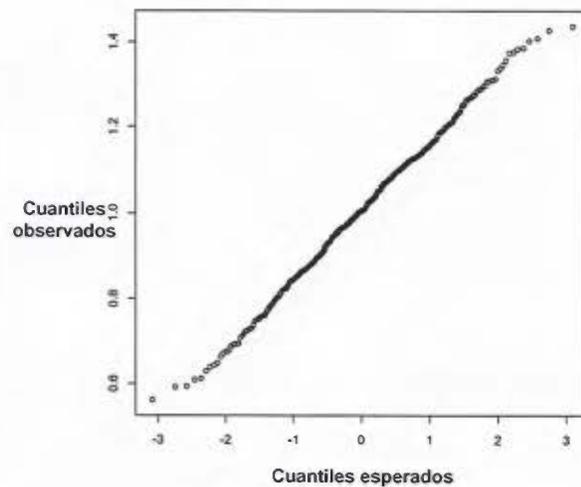
Lambda teórico = 0,01 y beta teórico = 0,05



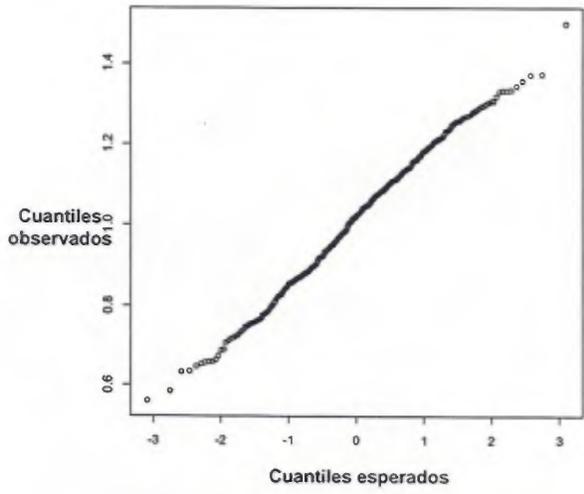
Lambda teórico = 0,01 y beta teórico = 5,00



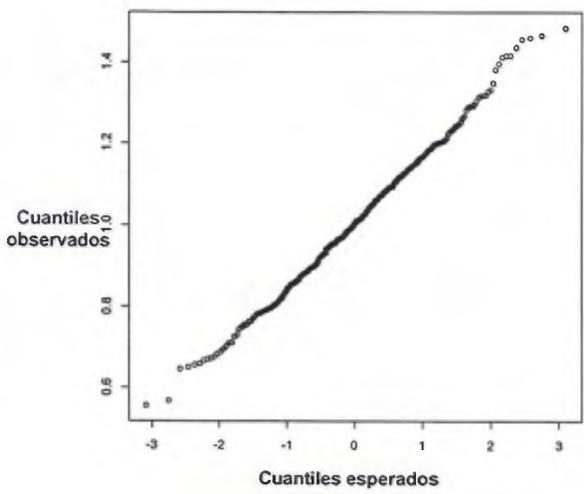
Lambda teórico = 0,10 y beta teórico = 0,05



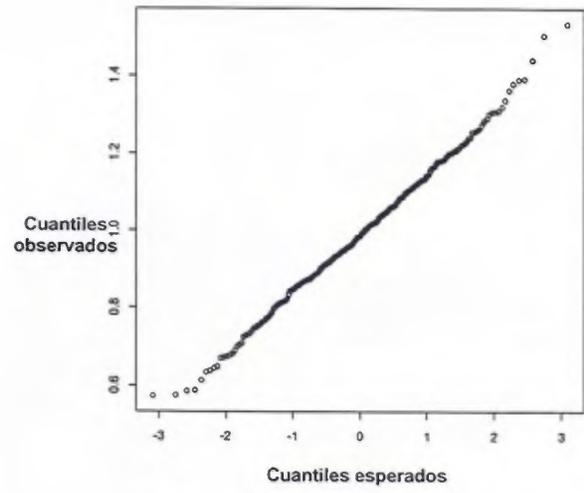
Lambda teórico = 0,10 y beta teórico = 5,00



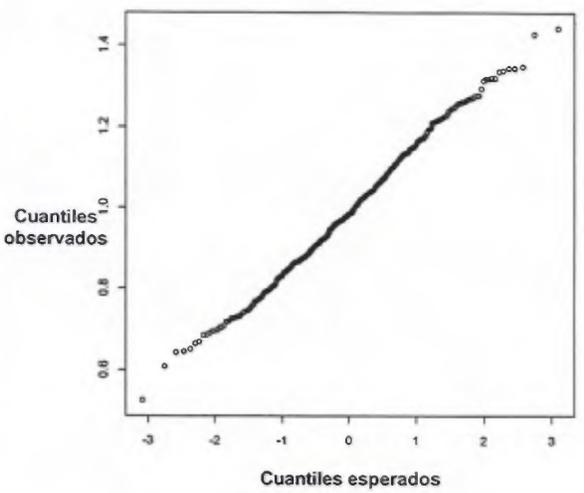
Lambda teórico = 0,23 y beta teórico = 0,05



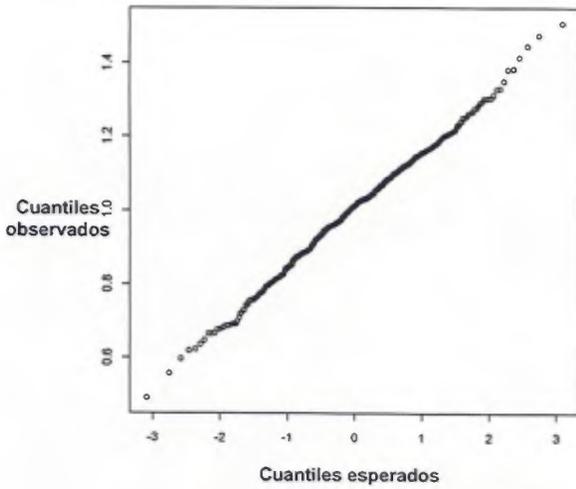
Lambda teórico = 0,23 y beta teórico = 5,00



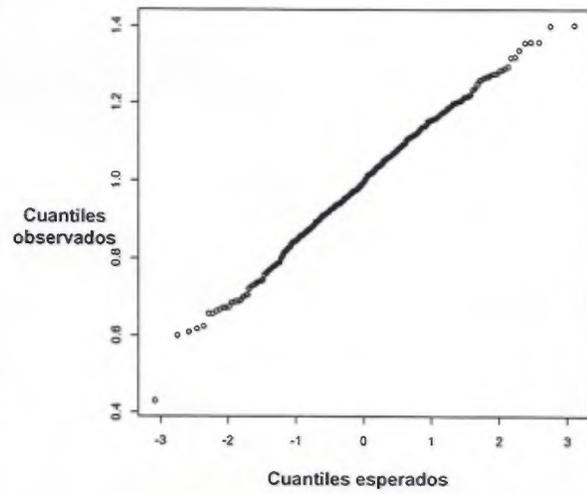
Lambda teórico = 0,36 y beta teórico = 0,05



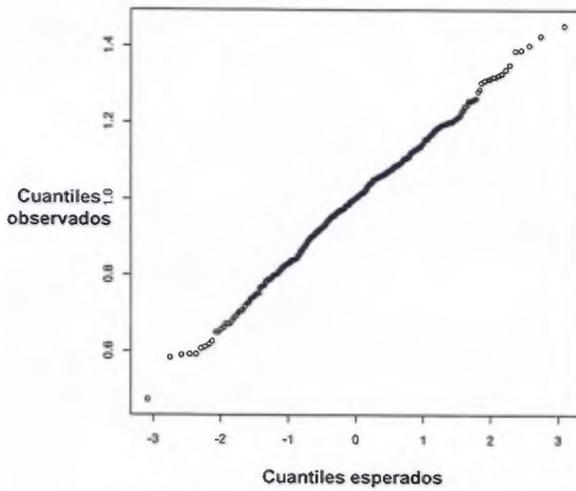
Lambda teórico = 0,36 y beta teórico = 5,00



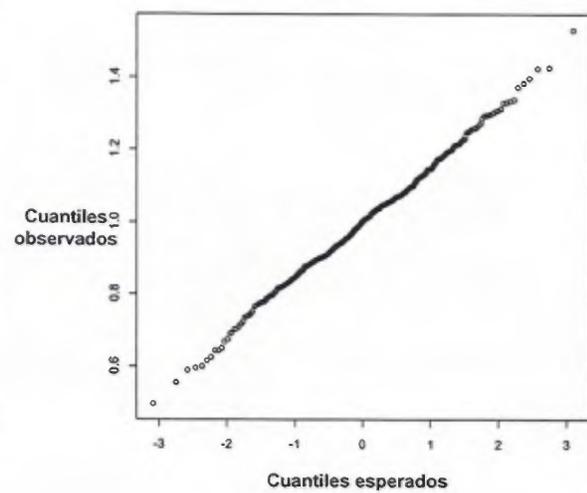
Lambda teórico = 0,49 y beta teórico = 0,05



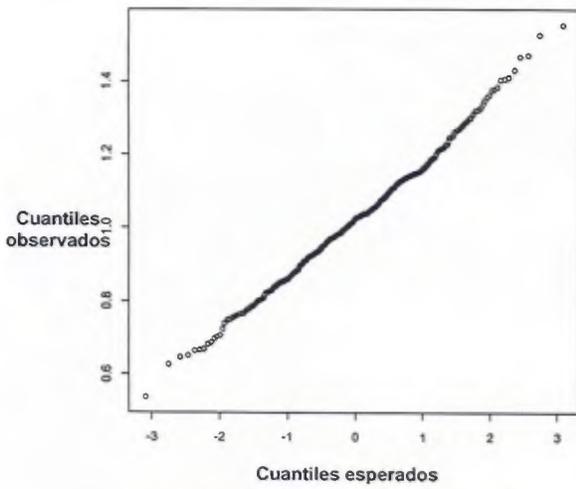
Lambda teórico = 0,49 y beta teórico = 5,00



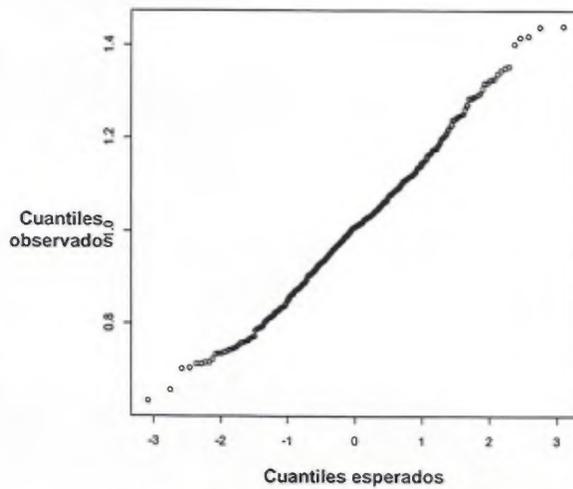
Lambda teórico = 0,62 y beta teórico = 0,05



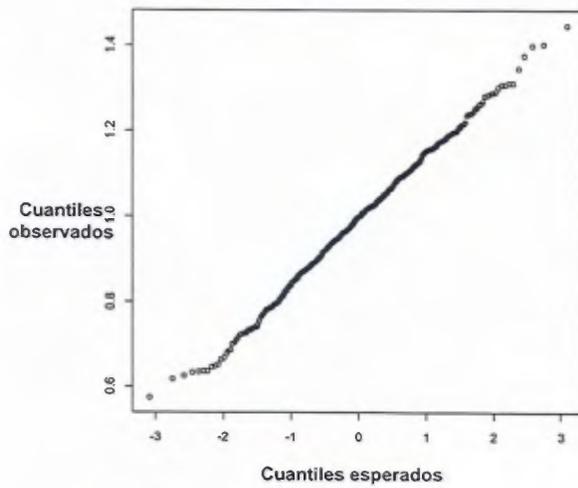
Lambda teórico = 0,62 y beta teórico = 5,00



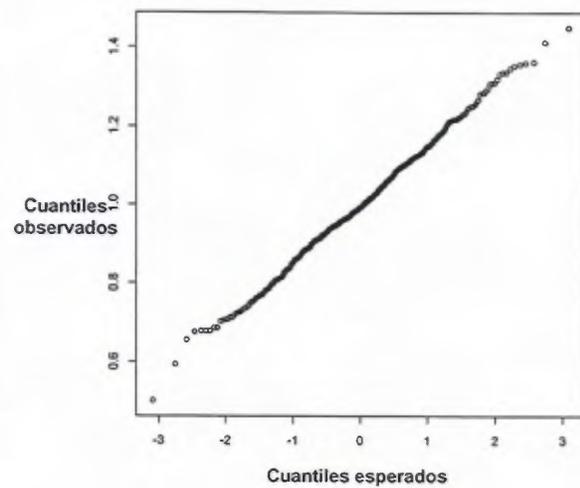
Lambda teórico = 0,75 y beta teórico = 0,05



Lambda teórico = 0,75 y beta teórico = 5,00



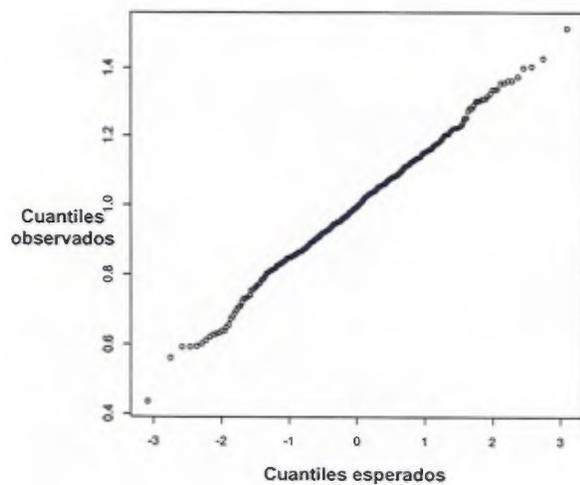
Lambda teórico = 0,90 y beta teórico = 0,05



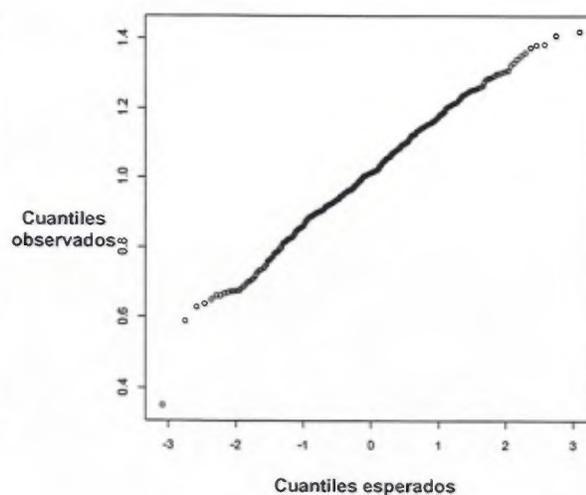
Lambda teórico = 0,90 y beta teórico = 5,00

Fuente: Elaboración propia, 2011.

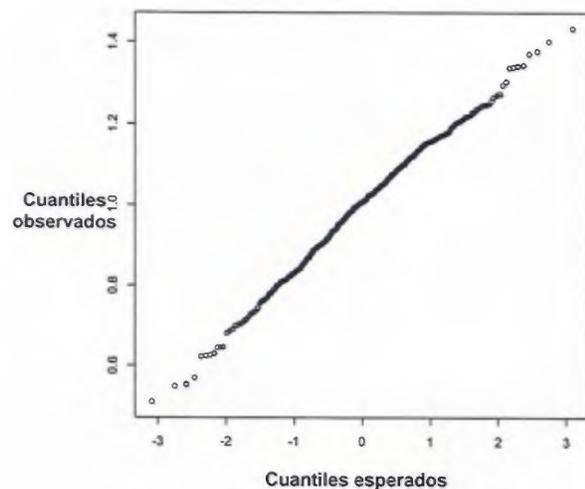
Gráfico 12. Gráfico de normalidad de los coeficientes con distribución beta y matriz de distancias, 2011



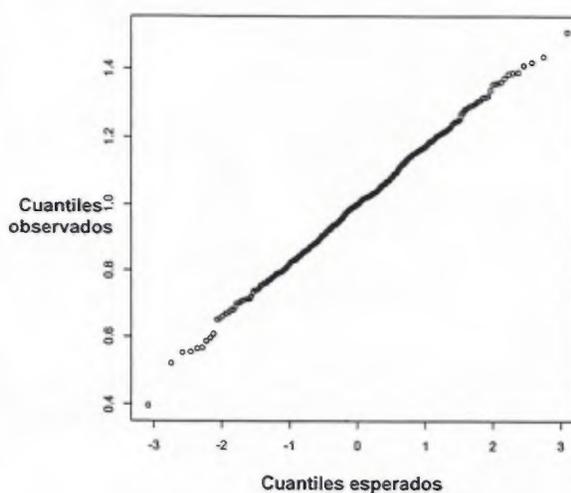
Lambda teórico = 0,01 y beta teórico = 0,05



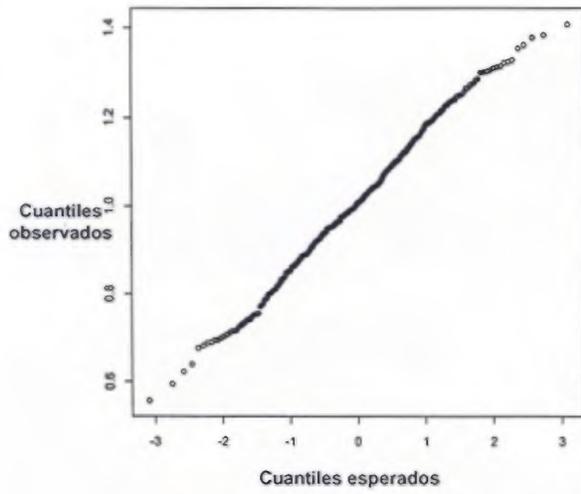
Lambda teórico = 0,01 y beta teórico = 5,00



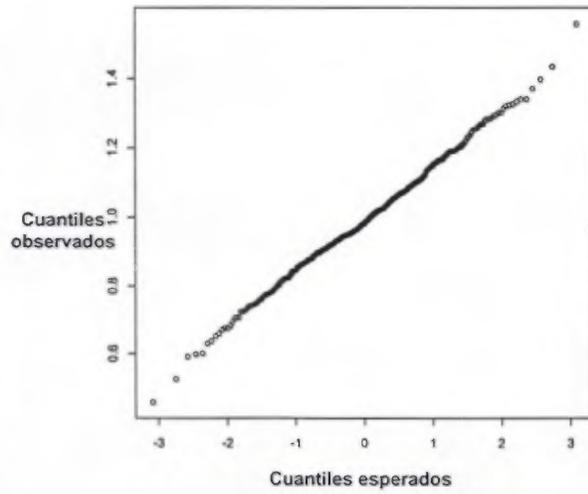
Lambda teórico = 0,10 y beta teórico = 0,05



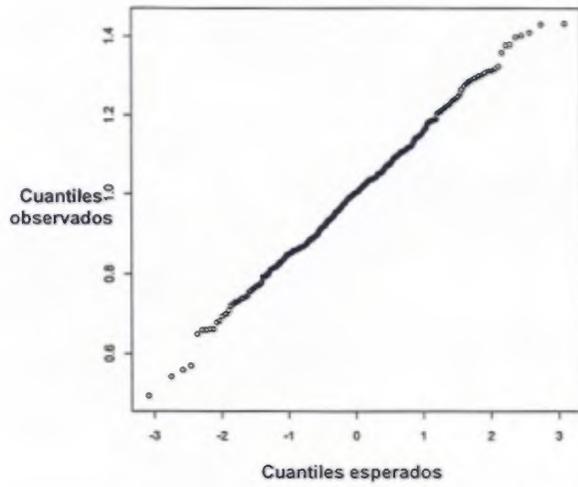
Lambda teórico = 0,10 y beta teórico = 5,00



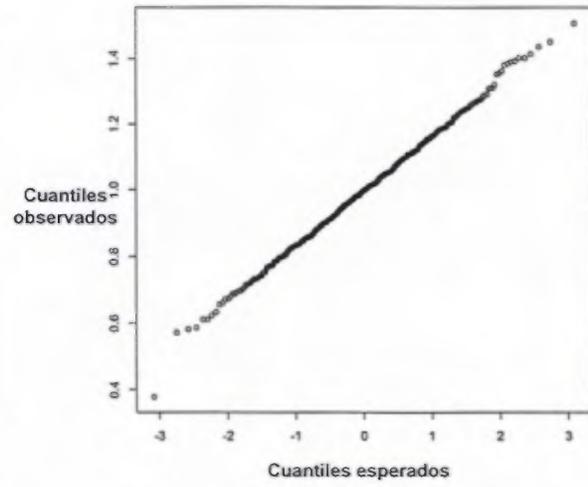
Lambda teórico = 0,23 y beta teórico = 0,05



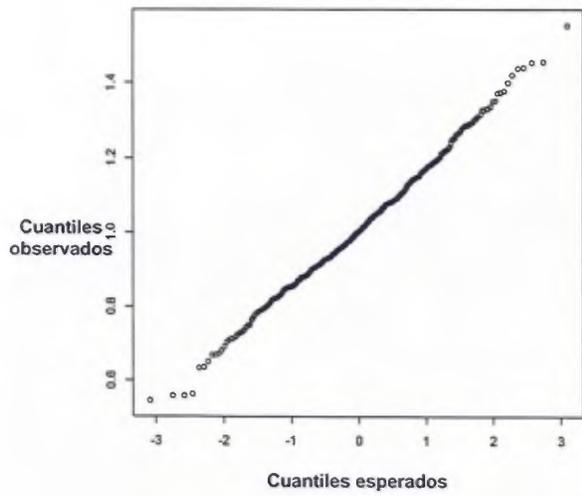
Lambda teórico = 0,23 y beta teórico = 5,00



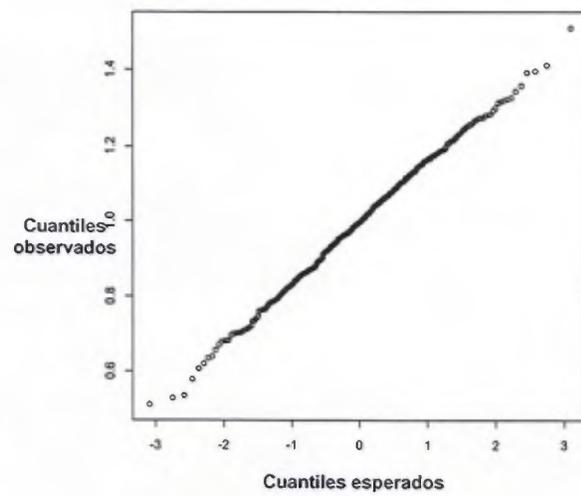
Lambda teórico = 0,36 y beta teórico = 0,05



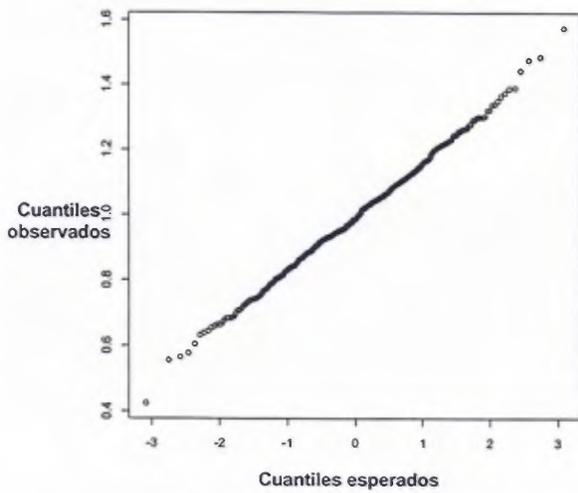
Lambda teórico = 0,36 y beta teórico = 5,00



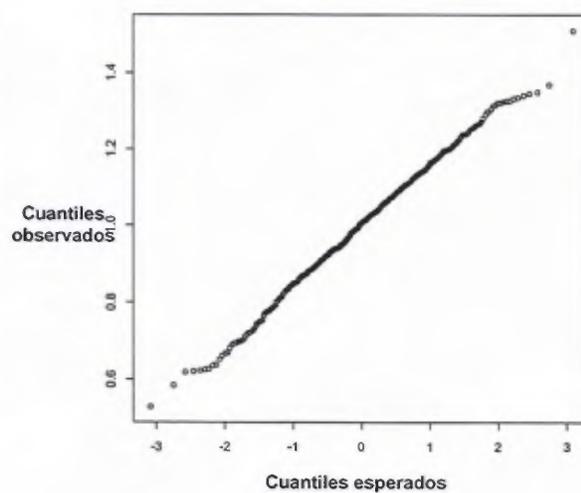
Lambda teórico = 0,49 y beta teórico = 0,05



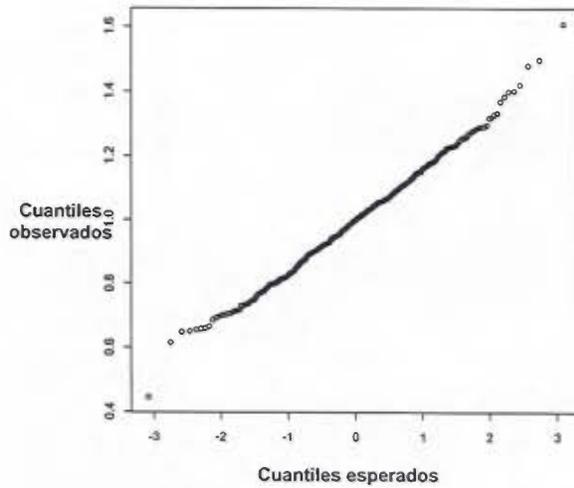
Lambda teórico = 0,49 y beta teórico = 5,00



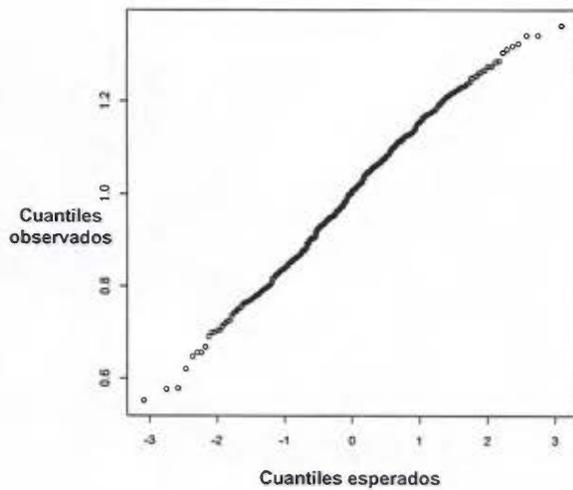
Lambda teórico = 0,62 y beta teórico = 0,05



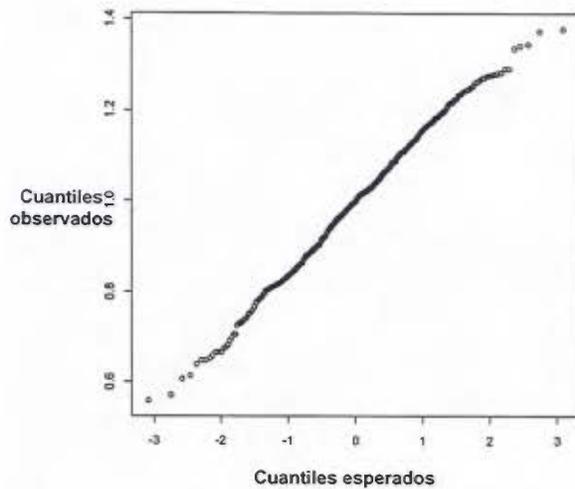
Lambda teórico = 0,62 y beta teórico = 5,00



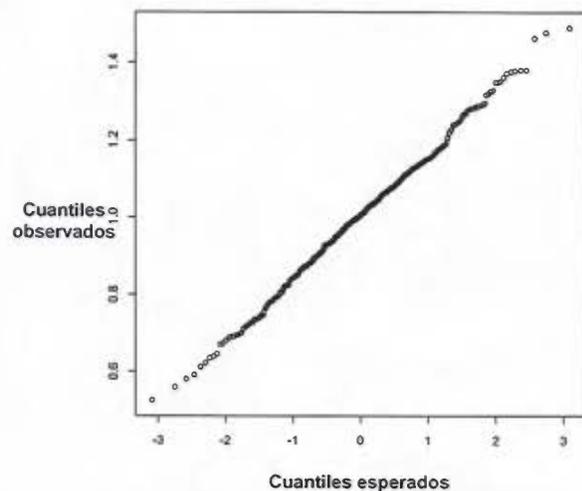
Lambda teórico = 0,75 y beta teórico = 0,05



Lambda teórico = 0,75 y beta teórico = 5,00



Lambda teórico = 0,90 y beta teórico = 0,05



Lambda teórico = 0,90 y beta teórico = 5,00

Fuente: Elaboración propia, 2011.

Con respecto a la tasa de rechazo obtenida para cada uno de los modelos simulados utilizando los tres tipos de matrices se observa que con la matriz de distancias existe una potencia mayor al 65% de detectar a lambda como significativamente distinto de cero cuando el lambda teórico es mayor a 0,36 (Cuadro 19). Por el contrario, tanto la matriz reina o *queen* como la matriz *K* vecinos más cercanos contabilizan una potencia mayor al 80% y 77% de detectar a lambda como significativamente distinto de cero cuando el lambda teórico es mayor a 0,23 respectivamente en ambos casos (Cuadros 17 y 18). Esto significa

que con la matriz de distancias el valor de la potencia de la prueba es menor aun cuando el lambda teórico es mayor, esto comparando los resultados obtenidos utilizando la matriz reina o *queen* y la matriz K vecinos más cercanos. Es importante señalar que en todos los cuadros y gráficos con los parámetros de beta, la distribución beta con beta igual a 0 sería muy parecida a una distribución normal de los errores. Por último, sin importar cual tipo de matriz de pesos se utilice, queda demostrado que hay un valor umbral relativamente bajo de los lambdas (entre 0,23 en las matrices reina y K vecinos, y 0,36 en la matriz de distancias) para poder detectarlo como significativo con una potencia alta.

Cuadro 17. Porcentaje de corridas para identificar la potencia de prueba con distribución beta utilizando matriz *queen*, 2011

Beta teórico	Lambda teórico							
	0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,05	0,068	0,234	0,856	1,000	1,000	1,000	1,000	1,000
0,20	0,058	0,290	0,860	0,998	1,000	1,000	1,000	1,000
0,30	0,042	0,242	0,816	0,998	1,000	1,000	1,000	1,000
0,40	0,042	0,246	0,862	0,996	1,000	1,000	1,000	1,000
5,00	0,046	0,280	0,860	0,998	1,000	1,000	1,000	1,000

Fuente: Elaboración propia, 2011.

Cuadro 18. Porcentaje de corridas para identificar la potencia de prueba con distribución beta utilizando matriz K vecinos, 2011

Beta teórico	Lambda teórico							
	0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,05	0,054	0,192	0,794	0,996	1,000	1,000	1,000	1,000
0,20	0,058	0,200	0,786	0,990	1,000	1,000	1,000	1,000
0,30	0,048	0,216	0,780	1,000	1,000	1,000	1,000	1,000
0,40	0,044	0,222	0,800	0,996	1,000	1,000	1,000	1,000
5,00	0,040	0,190	0,768	0,996	1,000	1,000	1,000	1,000

Fuente: Elaboración propia, 2011.

Cuadro 19. Porcentaje de corridas para identificar la potencia de prueba con distribución beta utilizando matriz de distancias, 2011

Beta teórico	Lambda teórico							
	0,01	0,10	0,23	0,36	0,49	0,62	0,75	0,90
0,05	0,046	0,080	0,302	0,664	0,960	0,998	1,000	1,000
0,20	0,052	0,076	0,360	0,710	0,934	0,998	1,000	1,000
0,30	0,042	0,098	0,310	0,694	0,934	0,998	1,000	1,000
0,40	0,046	0,102	0,290	0,666	0,942	0,996	1,000	1,000
5,00	0,056	0,078	0,292	0,648	0,946	0,998	1,000	1,000

Fuente: Elaboración propia, 2011.

5. 4 Síntesis preliminar

¿Cuáles son las propiedades estadísticas del modelo SAR cuando se aplica a los datos distritales de Costa Rica? Se halló que el modelo SAR es robusto a la asimetría de los errores (o sea, a la falta de normalidad) así como a distintos tamaños del error residual. Esto quiere decir que, no importa que los errores no sean normales o que el cuadrado medio de error sea alto o pequeño, las estimaciones de los coeficientes de regresión van a ser precisas e insesgadas. Este hallazgo puede deberse a las características implícitas de la estructura geográfica de Costa Rica. El contar con 471 distritos (si se excluye la Isla de Chira) es un tamaño de muestra suficientemente grande para que las estimaciones de los coeficientes se distribuyan normalmente y para que se tenga suficiente precisión para llegar a conclusiones prácticas a partir de las pruebas de hipótesis o intervalos de confianza.

Además, también se encontró que los estimadores SAR son robustos a la selección de la matriz de pesos que describe la autocorrelación. No importa cuál matriz de pesos se elija, las estimaciones de los coeficientes van a ser insesgadas, ya que las estimaciones son cercanas a los parámetros teóricos propuestos.

La robustez del SAR a los datos distritales sí varía según el grado de autocorrelación, si se desea analizar la potencia de la estimación del λ , con una significancia del 5%. La potencia de la prueba de hipótesis que se refiere a la existencia de autocorrelación es mayor a 75% a partir de valores de λ de 0,36, salvo con una matriz de pesos definida según las distancias, pues la potencia es menor. Este hallazgo es muy importante porque los analistas deciden usar un SAR en lugar de un modelo lineal general si se rechaza la hipótesis nula de que el λ poblacional es igual a cero. Las simulaciones sugieren que, si el λ poblacional es diferente de cero pero menor a 0,36, únicamente en una

fracción muy pequeña de las muestras se va a concluir que existe autocorrelación; el analista usa erróneamente un modelo lineal generalizado, con la subsecuente subestimación de los errores estándar. Este problema se agravaría si se escoge una matriz de distancias para la matriz de pesos.

Posteriormente, en la siguiente sección se expondrán los diferentes modelos SAR que fueron calculados para identificar si existe o no asociación estadística entre el programa social *Avancemos* y la deserción colegial tomando en cuenta la autocorrelación espacial.

6. Aplicación del SAR al programa social Avancemos

En este capítulo, se analiza el modelo SAR propuesto para investigar la asociación entre la implementación del programa social *Avancemos* y el nivel de deserción. Se establecería que la implementación del programa social *Avancemos* sí está asociada con la reducción de la deserción si las variables referidas a *Avancemos* (tasa de la cantidad de becas y monto promedio de las becas en miles de colones) tienen coeficientes negativos y significativamente distintos de cero.

Se va a iniciar con un análisis descriptivo de las variables que conforman el modelo SAR. Seguidamente, se desarrolla un análisis compuesto por 5 modelos SAR, tomando en cuenta la información recopilada en el periodo de estudio:

1. Modelo SAR 2011 con las variables planteadas inicialmente en la investigación.
2. Modelo SAR de las diferencias entre los valores de esas variables correspondientes al 2007 y 2011 cuyos años son el inicio y final del periodo de estudio.
3. Modelo SAR parecido al planteado en el 2011 pero modificando solamente la variable dependiente recalculada como el cambio en el porcentaje de deserción entre el 2012 y 2011 para observar el comportamiento de los resultados tomando en cuenta el año posterior al finalizar el periodo de estudio.

Posteriormente, se analizaron otros dos modelos SAR involucrando información del Censo 2011:

4. Modelo SAR 2011 controlando por variables censales el cual analiza otras variables independientes que pueden tener influencia en los resultados y así crear un modelo más robusto.
5. Modelo SAR 2011 cuya variable dependiente se estima de forma aproximada utilizando variables censales creando así el porcentaje de deserción según el censo. Este modelo fue calculado con esta variante y también se calculó otro modelo SAR 2011 excluyendo la variable monto promedio de becas en miles de colones, ya que en el primer caso, provoca que la tasa de becas sea no significativa, por la correlación existente entre ellas.

Originalmente, se había planteado que el análisis debería usar los datos administrativos que se recolectan y publican anualmente. Sin embargo, como se va a apreciar en el texto, el análisis dejó claro que los datos administrativos contienen un sesgo de medición que genera una gran cantidad de información en cero para distritos que en la realidad sí cuentan con estudiantes beneficiarios. El cálculo del porcentaje de deserción según el censo se fundamenta en que las variables independientes relacionadas con *Avanceamos* están referidas al distrito de residencia del beneficiario; en cambio el porcentaje de deserción según registros administrativos está referido al distrito en donde está localizado el colegio generando los "falsos ceros" mencionados. Por el contrario, el porcentaje de deserción según el censo sí logra captar el porcentaje de deserción de los estudiantes que residen en el distrito.

Cabe señalar que todos los modelos fueron estimados debido a la inquietud de profundizar en la relación estadística de las variables correspondientes de *Avanceamos* y el porcentaje de deserción con distintas especificaciones para obtener los resultados más óptimos que expliquen el grado de asociación existente entre ellas. La falta de variables censales que puedan ser utilizadas como control de las relacionadas directamente con *Avanceamos* es una desventaja presente en los tres modelos iniciales que logra ser corregida más adelante con el

análisis de los modelos controlando por variables censales y operacionalizando el porcentaje de deserción a partir de la información censal. Dicho cálculo tiene cierto grado de sesgo de selección porque no se sabe el año en que desertó el estudiante y aun presentando este inconveniente, el porcentaje de deserción obtenido a partir de la información censal es mejor que el porcentaje de deserción calculado por el MEP.

6. 1 Análisis descriptivo

De manera general, en el 2011 existen 115 distritos que no poseen del todo un colegio, lo cual representa el 24,4% del universo de estudio compuesto por 472 distritos. Es importante señalar que la región Central tiene 80 distritos sin colegio, seguido muy distante por la región Chorotega con 17 distritos sin colegio, como contraste, en la región Brunca todos sus distritos poseen al menos un colegio.

Cada una de las variables que forman parte del modelo SAR presenta las siguientes características:

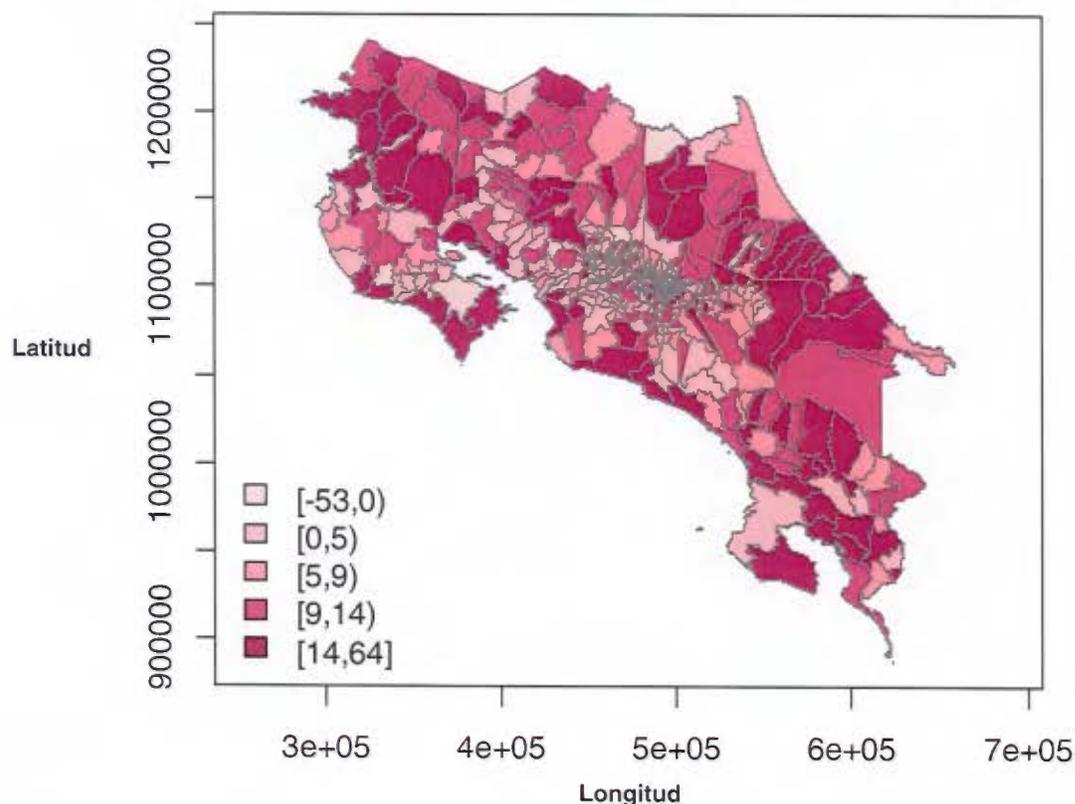
Porcentaje de deserción según datos administrativos

En el 2011, el porcentaje de deserción resultó negativo en 18 distritos, esto ocurre debido a una mayor cantidad de estudiantes contabilizados en la matrícula final con respecto a los reportados en la matrícula inicial. Las posibles causas de estos inconvenientes están relacionadas con la migración de estudiantes hacia otros distritos convirtiéndose en nuevos alumnos de otras instituciones públicas. Esta es una clara limitación de utilizar los registros administrativos del MEP para calcular el porcentaje de deserción.

Solamente en 8 distritos, la matrícula inicial y final estuvo compuesta por la misma cantidad de estudiantes, lo cual señala que para estos distritos no hubo deserción colegial. Este valor representa el 2,2% de los distritos que tienen al menos un colegio.

Además, esta variable presenta sus mayores valores en la periferia del país y en algunos distritos de la Gran Área Metropolitana (GAM). A nivel regional, se observa gran cantidad de distritos con deserción colegial mayor en las regiones Huetar Caribe, Brunca y Chorotega. La región que presenta mejores condiciones con respecto a esta variable es la Central.

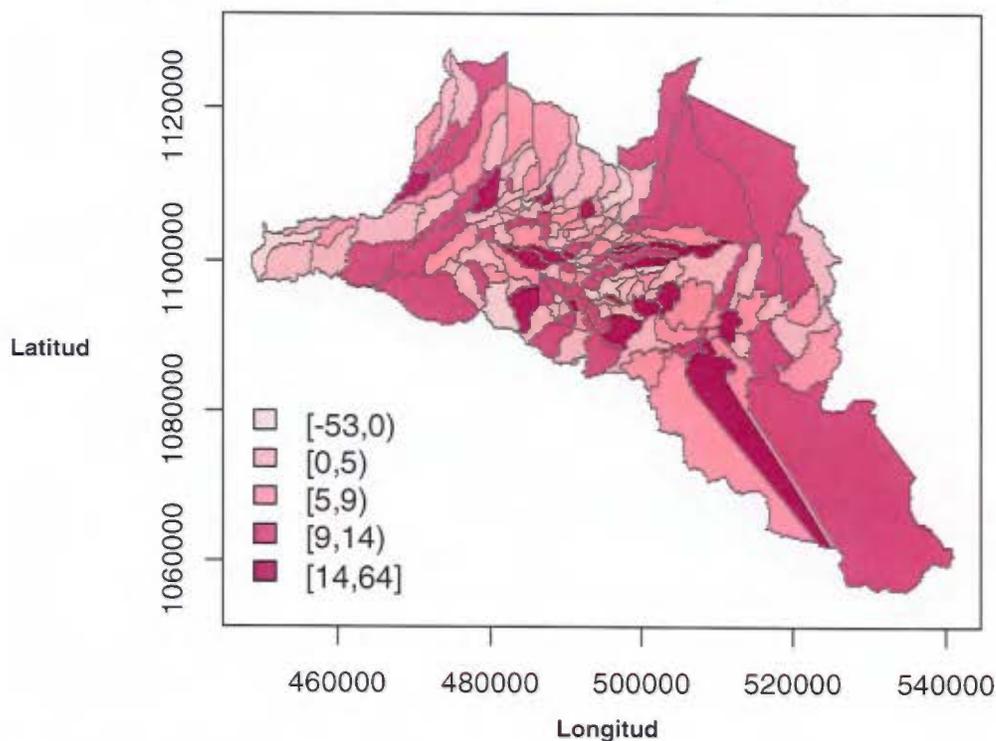
Mapa 1. Costa Rica: porcentaje de deserción por rangos, 2011



Fuente: Estadísticas del Ministerio de Educación Pública (MEP), 2011.

Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

Mapa 2. GAM: porcentaje de deserción por rangos, 2011



Fuente: Estadísticas del Ministerio de Educación Pública (MEP), 2011.

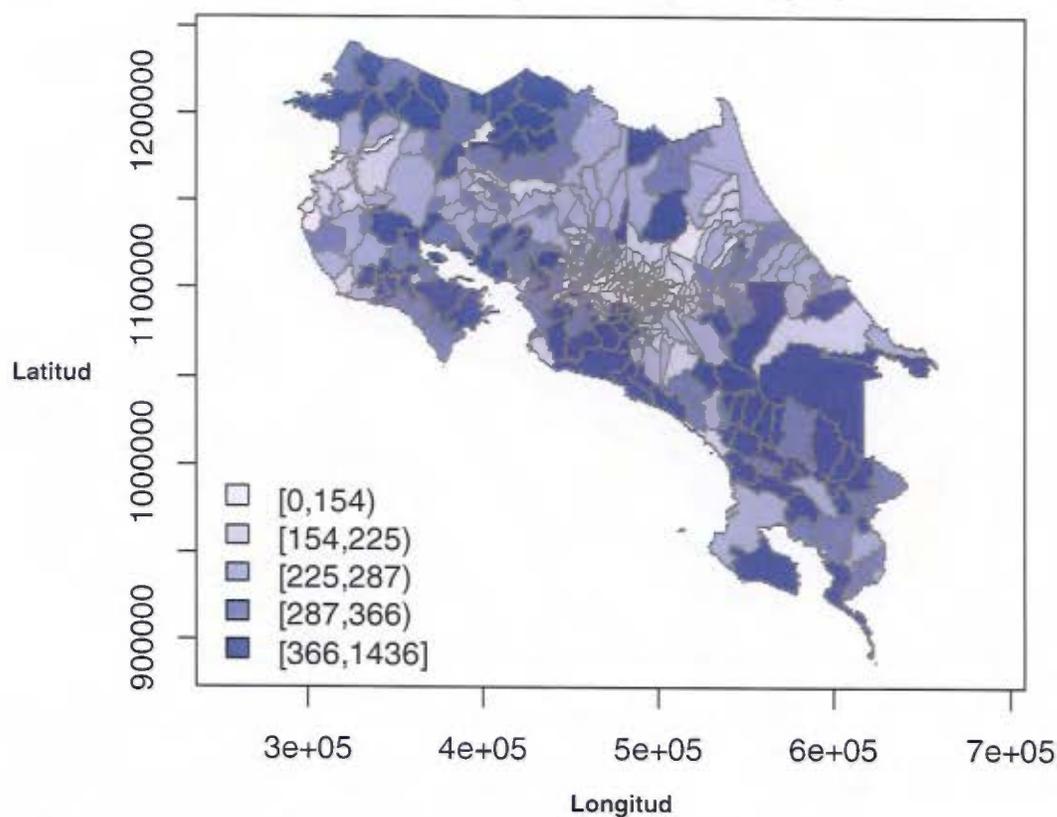
Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

Tasa de la cantidad de becas

En el 2011, en promedio se otorgaron 267 becas por cada mil habitantes con edades entre 12 y 21 años. Esta variable presenta un comportamiento normal porque existen pocos valores extremos y esto genera un agrupamiento en los demás valores. Con respecto a lo anterior, el valor máximo y mínimo de esta variable corresponden al distrito de Buenavista perteneciente al cantón de Guatuso con 1 435 becas por cada mil habitantes con edades entre 12 y 21 años y al distrito de Sánchez que se ubica en el cantón de Curridabat con un total de 7 becas por cada mil habitantes con edades entre 12 y 21 años respectivamente.

También, se observa que en la periferia del país están ubicadas las mayores tasas de becas principalmente en la parte norte y sur, así como en algunos distritos de la región Pacífico Central. Cabe destacar que la GAM en su mayoría presenta menores valores correspondientes a las becas otorgadas a estudiantes entre 12 y 21 años.

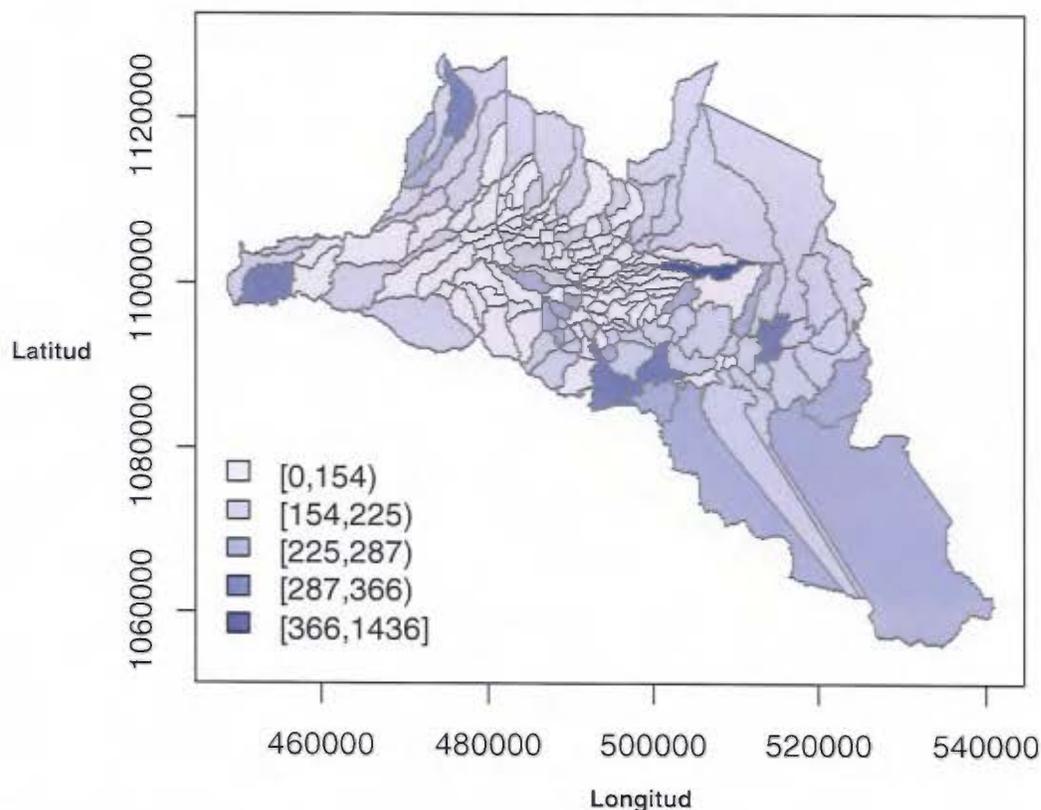
Mapa 3. Costa Rica: tasa de la cantidad de becas por cada mil habitantes con edades entre 12 y 21 años por rangos, 2011



Fuente: Estadísticas del Instituto Mixto de Ayuda Social (IMAS), 2011.

Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

Mapa 4. GAM: tasa de la cantidad de becas por cada mil habitantes con edades entre 12 y 21 años por rangos, 2011



Fuente: Estadísticas del Instituto Mixto de Ayuda Social (IMAS), 2011.

Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

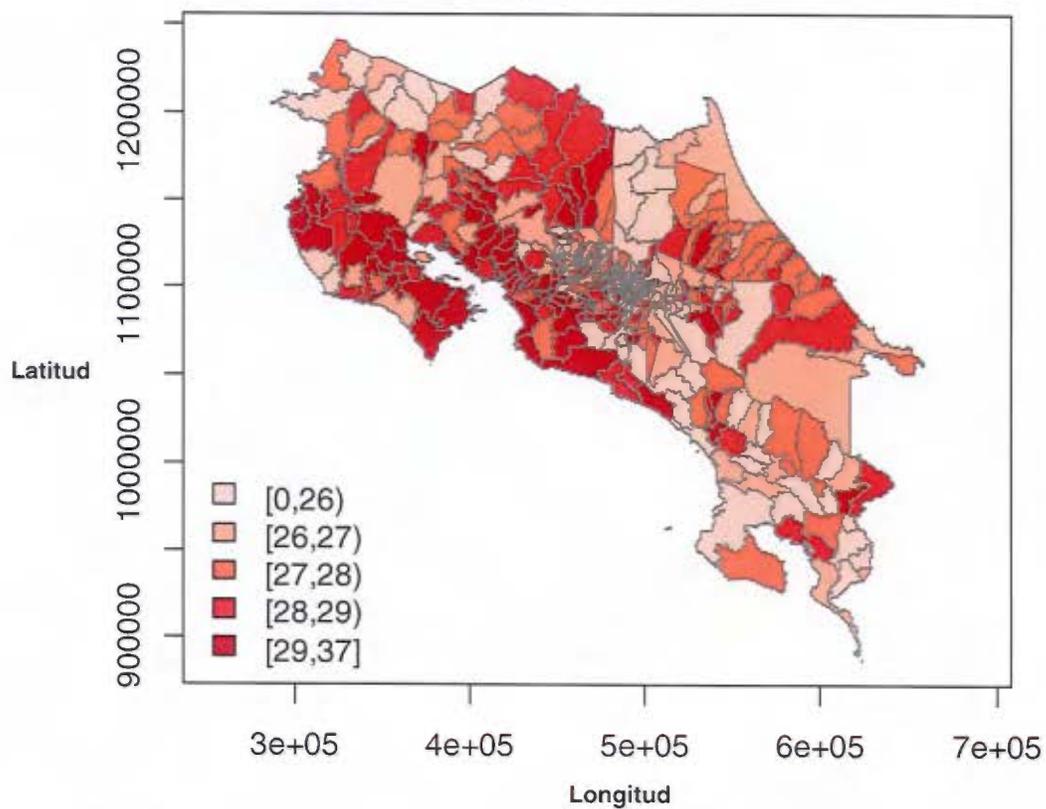
Monto promedio de becas en miles de colones

En el 2011, el monto promedio de las transferencias otorgadas fue de ₡27 230 colones. Esta variable presenta poca variabilidad ya que el percentil 25 y 75 poseen valores de ₡25 939 y ₡28 338 colones respectivamente, evidenciando una gran similitud en el monto promedio de las becas. No obstante, se visualiza un comportamiento normal en los valores centralizados. El distrito que en promedio recibió mayor monto de becas es Cabo Velas perteneciente al cantón

de Santa Cruz y como contraparte, el distrito con menor monto promedio de becas otorgadas es Curubandé, ubicado en el cantón de Liberia.

A nivel nacional, el monto promedio de las becas está distribuido de una forma muy similar y por esta razón presenta poca variabilidad en los quintiles. Como se observa, los valores más altos están distribuidos en las regiones Chorotega y Pacífico Central. Por otra parte, valores cercanos al promedio están situados en algunos distritos de la parte Norte y del Caribe de nuestro país. Con respecto a la GAM, en su mayoría, los distritos presentan valores intermedios relacionados con las transferencias monetarias recibidas por los beneficiarios de las becas.

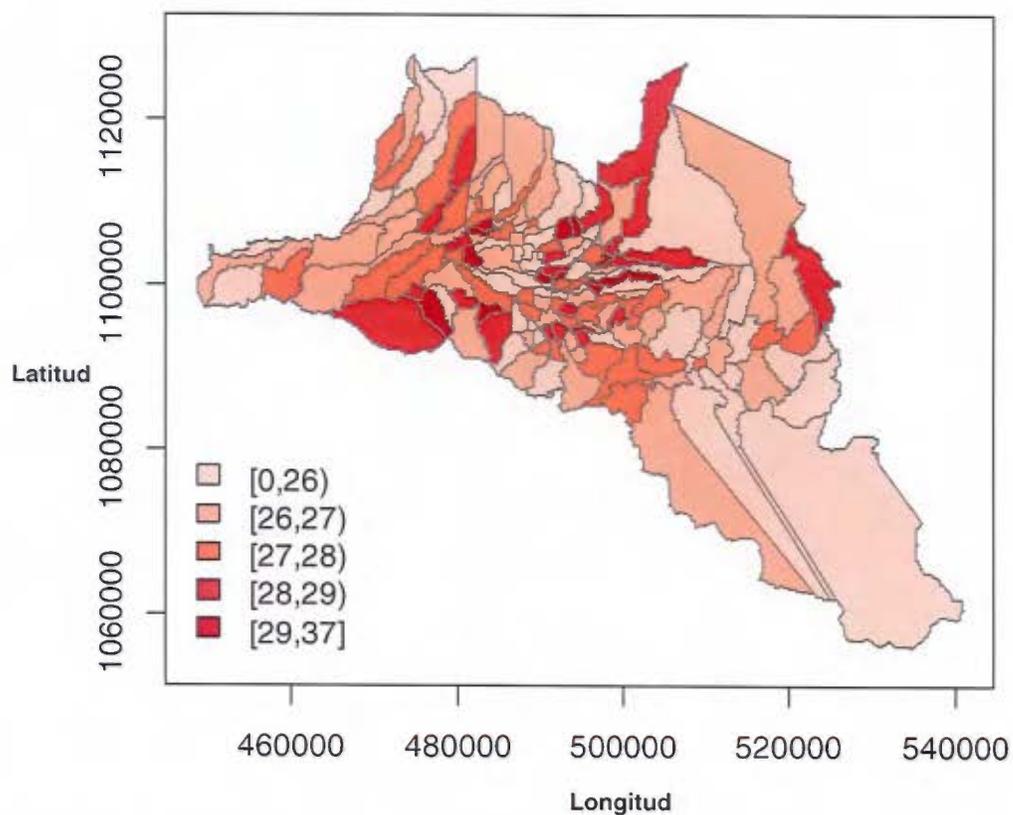
Mapa 5. Costa Rica: monto promedio de becas en miles de colones por rangos, 2011



Fuente: Estadísticas del Instituto Mixto de Ayuda Social (IMAS), 2011.

Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

Mapa 6. GAM: monto promedio de becas en miles de colones por rangos, 2011



Fuente: Estadísticas del Instituto Mixto de Ayuda Social (IMAS), 2011.

Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

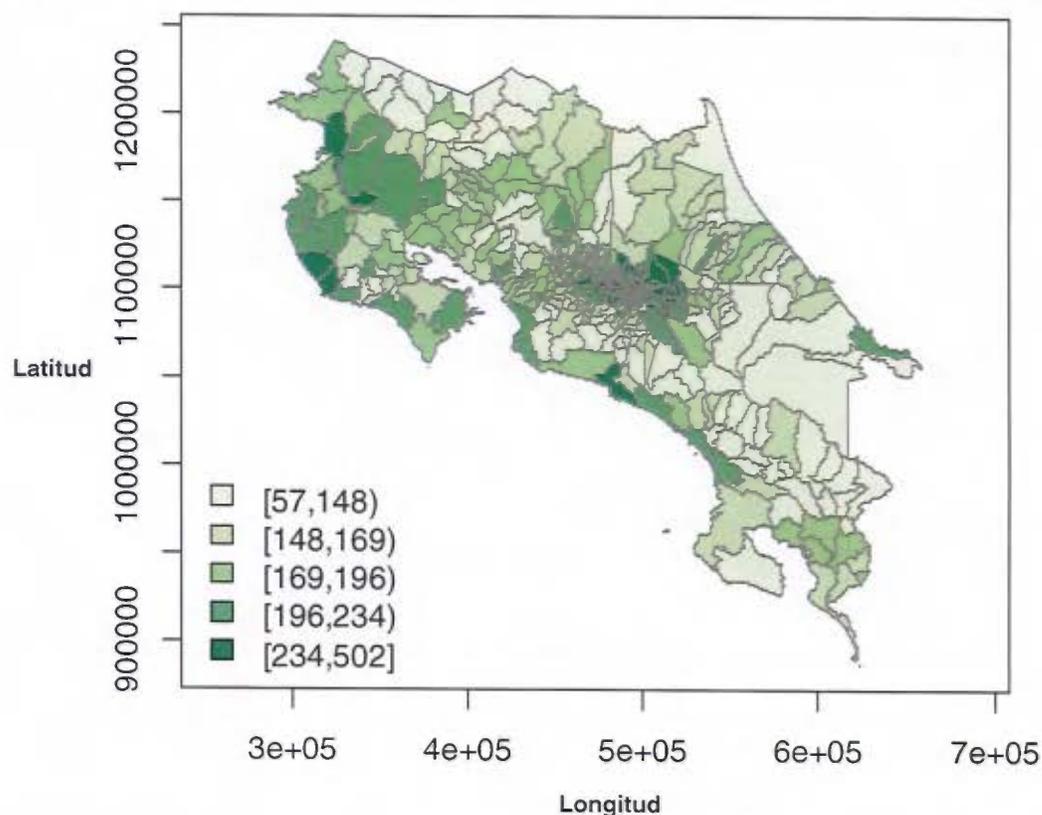
Consumo promedio de electricidad

En el 2011, el consumo promedio de electricidad fue de 192 kilowatts/hora. El valor modal estuvo cercano a los 200 kilowatts/hora superando por muy poco el promedio obtenido. El distrito que presenta el mayor valor es San Rafael perteneciente al cantón de Escazú, en contraste con el distrito de Chirripó ubicado en el cantón de Turrialba que obtuvo el menor consumo promedio de electricidad.

Es importante indicar que esta variable tiene un comportamiento normal con muy pocos valores extremos.

De manera general, el consumo promedio de electricidad tiene valores bajos en gran cantidad del territorio nacional exceptuando algunas zonas de la provincia de Guanacaste y del Pacífico en menor medida. Como contraparte, la GAM es la zona que registra los consumos promedio de electricidad más altos que están ubicados mayoritariamente con valores mayores o iguales a 196 kilowatts/hora (III quintil).

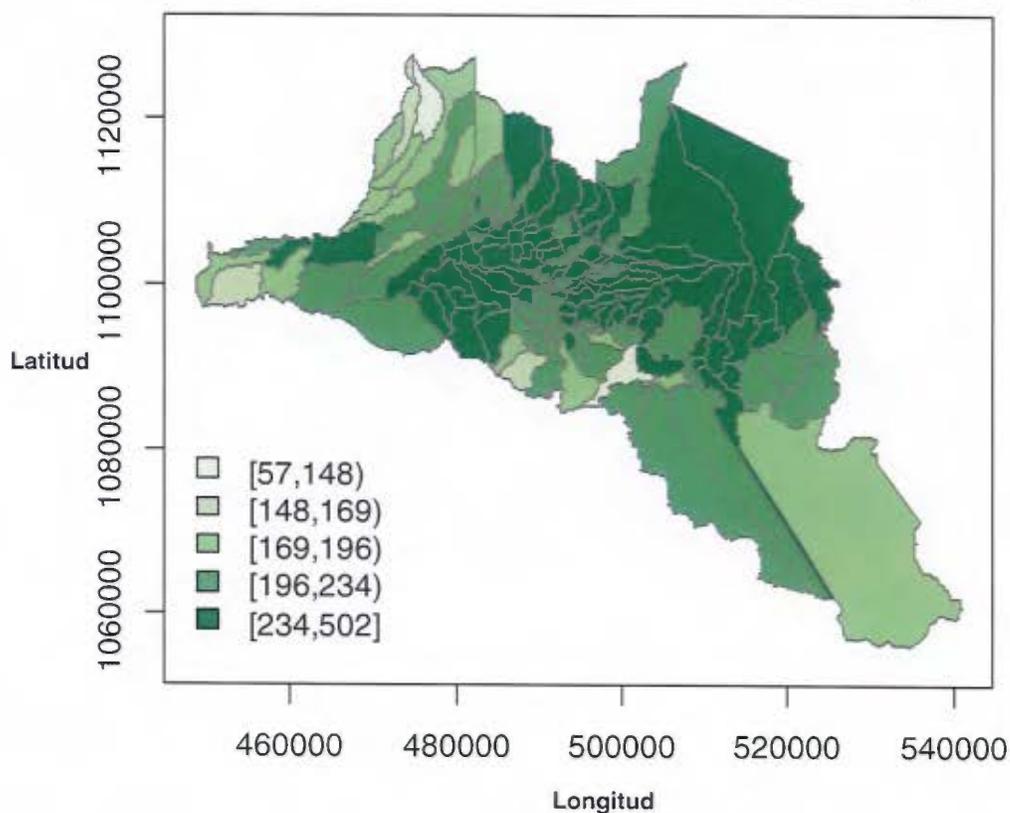
Mapa 7. Costa Rica: consumo promedio de electricidad por rangos, 2011



Fuente: Datos de investigaciones realizadas por el Ministerio de Planificación Nacional y Política Económica (Mideplan), 2011.

Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

Mapa 8. GAM: consumo promedio de electricidad por rangos, 2011



Fuente: Datos de investigaciones realizadas por el Ministerio de Planificación Nacional y Política Económica (Mideplan), 2011.

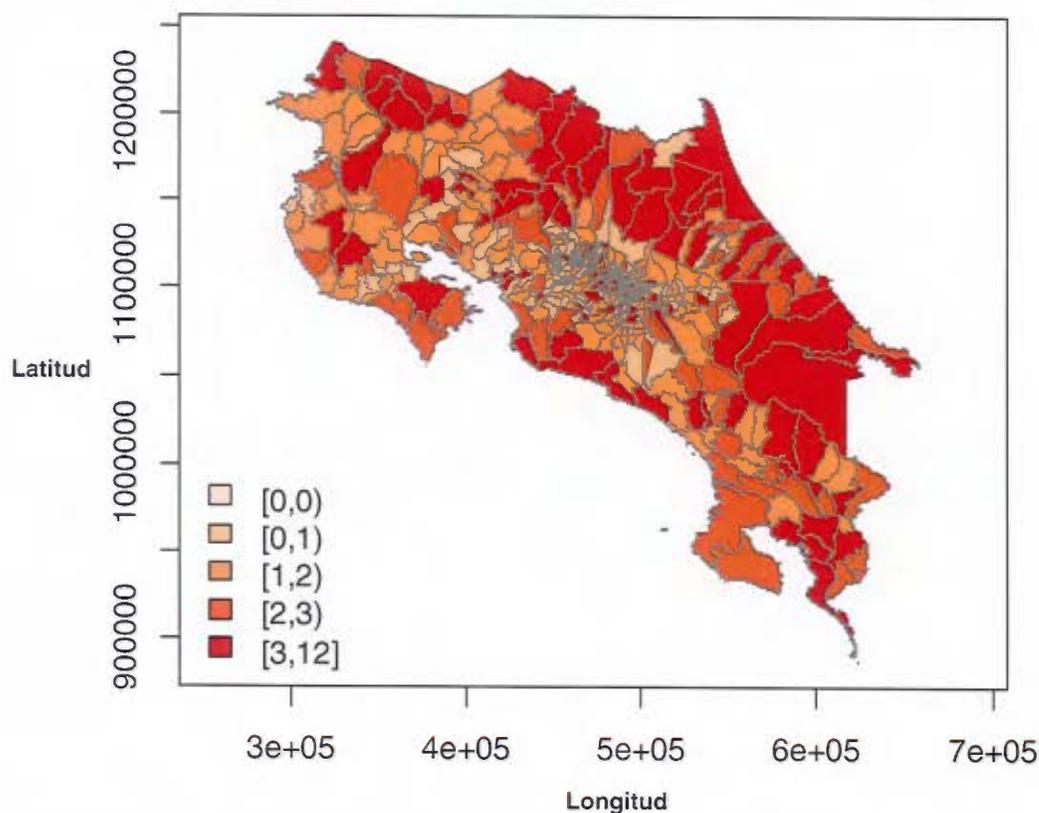
Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

Cantidad de colegios

En el 2011, un total de 357 distritos poseen al menos un colegio, representando el 75,6% de los 472 distritos existentes en nuestro país. Tomando en cuenta la cantidad de colegios, 165 distritos poseen solamente un colegio, 72 distritos tienen dos colegios y 49 distritos presentan tres colegios, resaltando los primeros lugares. Dadas las características de esta variable, presenta un comportamiento asimétrico ya que los valores están mayormente concentrados al inicio de la distribución.

A nivel nacional, existe una mayor cantidad de colegios en las zonas Norte, Este y Sur de la periferia del país. Esta frecuencia es menor en ciertos lugares de las regiones Chorotega, Brunca y Pacífico Central. Por otra parte, en la zona central de la GAM destaca la existencia de más colegios, en contraste con la periferia de la GAM donde la cantidad de colegios es menor.

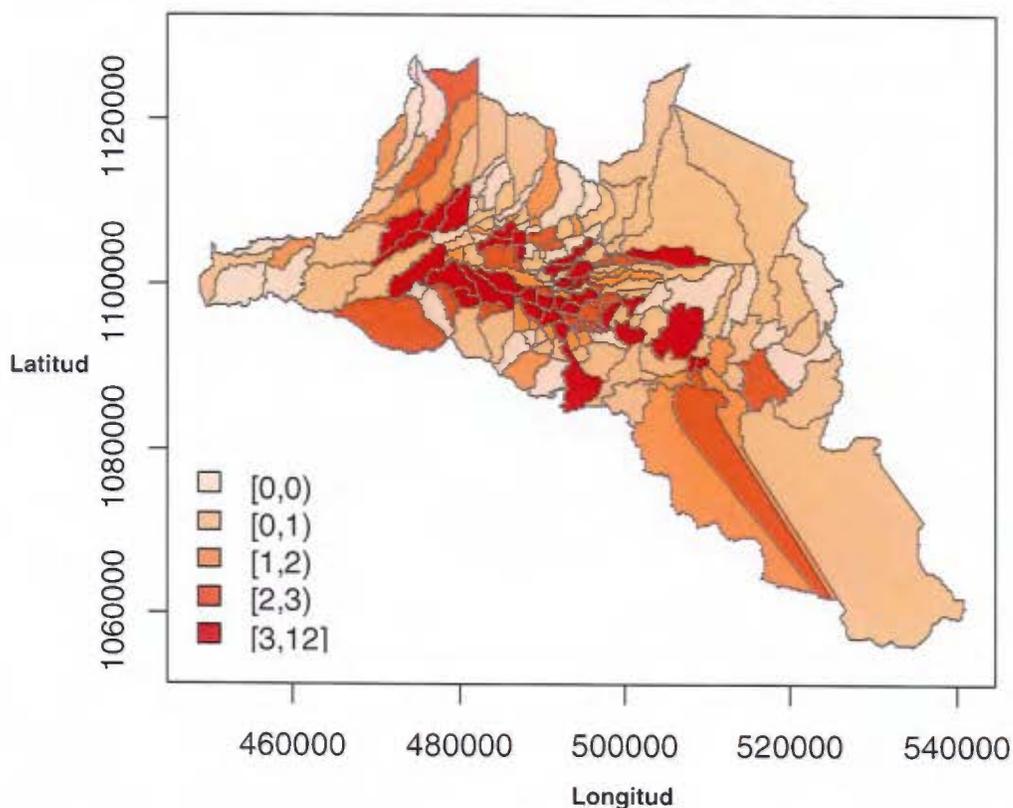
Mapa 9. Costa Rica: cantidad de colegios por rangos, 2011



Fuente: Estadísticas del Ministerio de Educación Pública (MEP), 2011.

Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

Mapa 10. GAM: cantidad de colegios por rangos, 2011



Fuente: Estadísticas del Ministerio de Educación Pública (MEP), 2011.

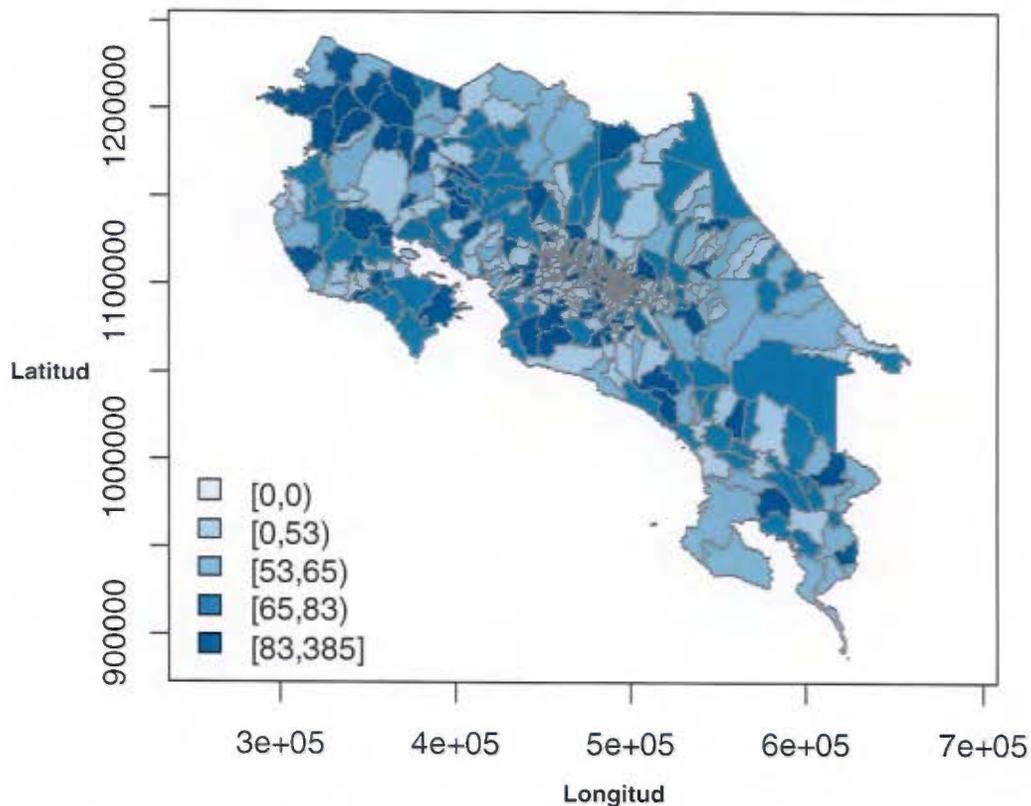
Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

Tasa del total de profesores

La tasa del total de profesores fue calculada con respecto a la población que asiste al colegio o también conocida como matrícula inicial al comienzo del curso lectivo. De manera general, en el 2011 algunos de los distritos de la región Chorotega están ubicados con valores mayores o iguales a 65 profesores por cada mil personas que asisten al colegio (III quintil). En el resto del territorio nacional predominan los distritos pertenecientes a los quintiles más bajos. Esta

variable presenta un comportamiento asimétrico agrupando la mayor cantidad de datos en los valores iniciales.

Mapa 11. Costa Rica: tasa del total de profesores por cada mil personas que asisten al colegio por rangos, 2011

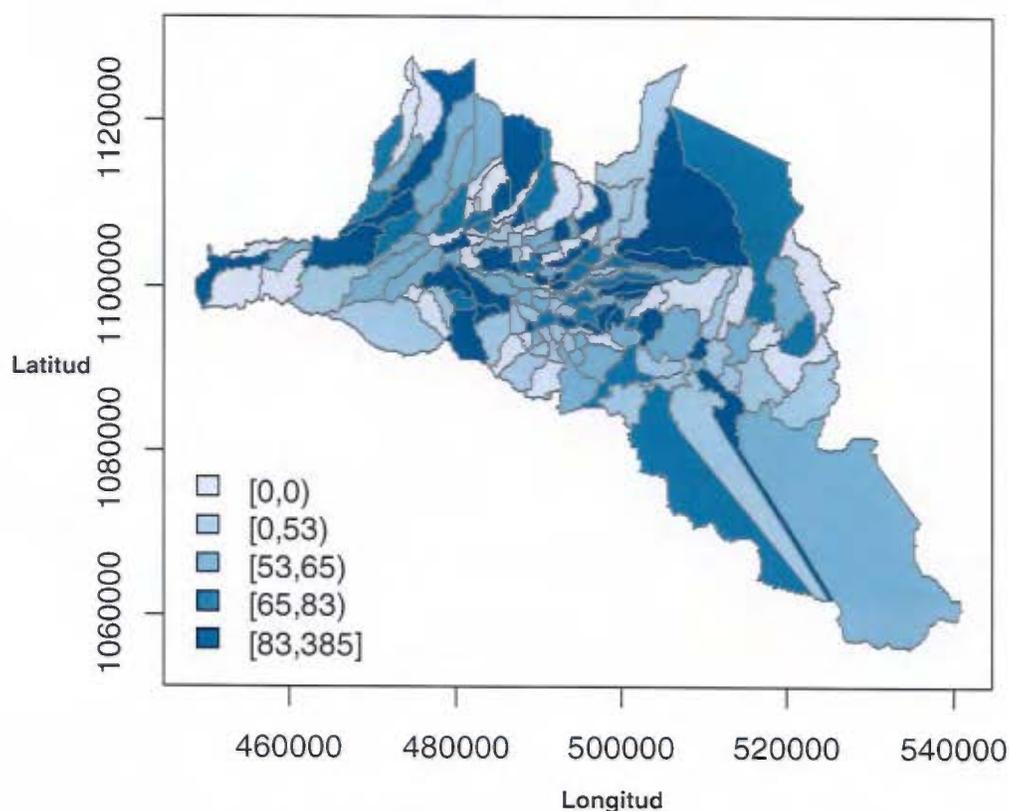


Fuente: Estadísticas del Ministerio de Educación Pública (MEP), 2011.

Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

En el caso de la GAM, en la parte central existen distritos que tienen valores más altos y los más alejados poseen quintiles más bajos principalmente en la parte sur y este.

Mapa 12. GAM: tasa del total de profesores por cada mil personas que asisten al colegio por rangos, 2011



Fuente: Estadísticas del Ministerio de Educación Pública (MEP), 2011.

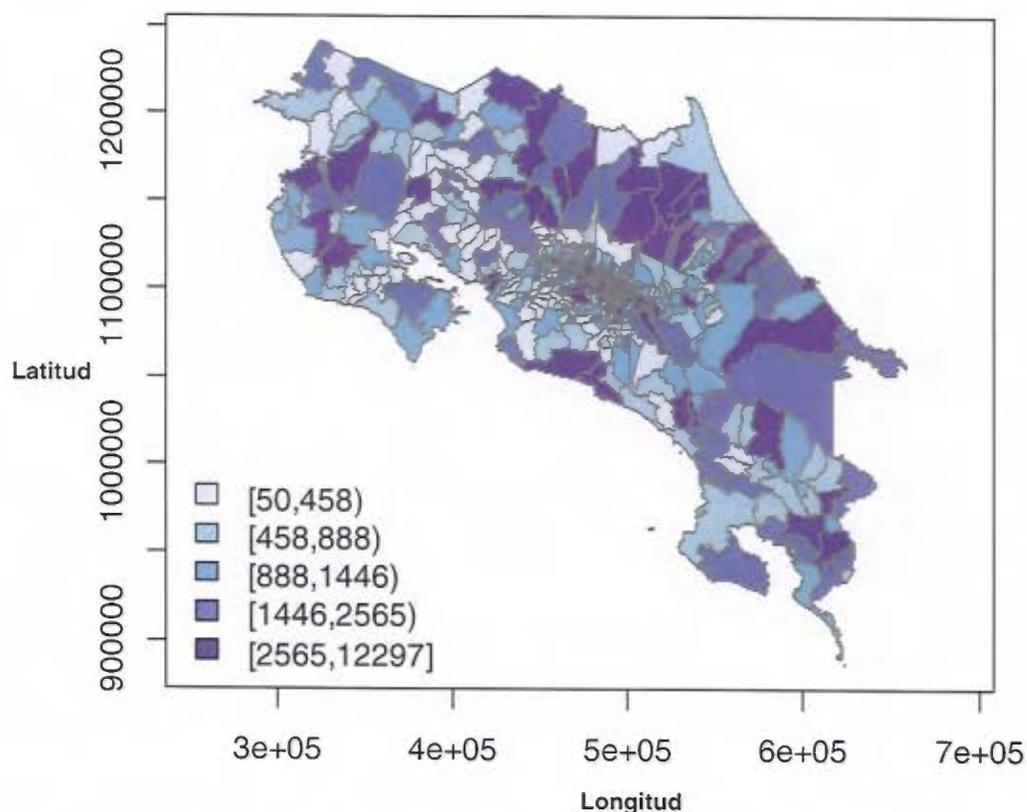
Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

Población en edad colegial de 12 a 21 años

Es importante recordar que se decidió utilizar la población de 12 a 21 años porque el programa social *Avancemos* está orientado a personas ubicadas entre esas edades. En el 2011, la población en edad colegial de 12 a 21 años estuvo ubicada mayoritariamente en la región Central y en varios lugares de la zona Norte y Atlántica del país. Cabe señalar que son pocos los distritos de las regiones Pacífico Central, Brunca y Chorotega que están ubicados con valores

mayores o iguales a 1 446 personas (III quintil). Con el ajuste correspondiente, esta variable presenta un comportamiento normal, agrupando la mayor cantidad de datos en el centro de la distribución. El distrito que tiene más personas con edad colegial de 12 a 21 años es Pavas con 12 297 habitantes y por el contrario, el distrito de Toro Amarillo presenta la menor cantidad de personas con edades entre 12 a 21 años con un total de 50 habitantes.

Mapa 13. Costa Rica: población en edad colegial de 12 a 21 años por rangos, 2011

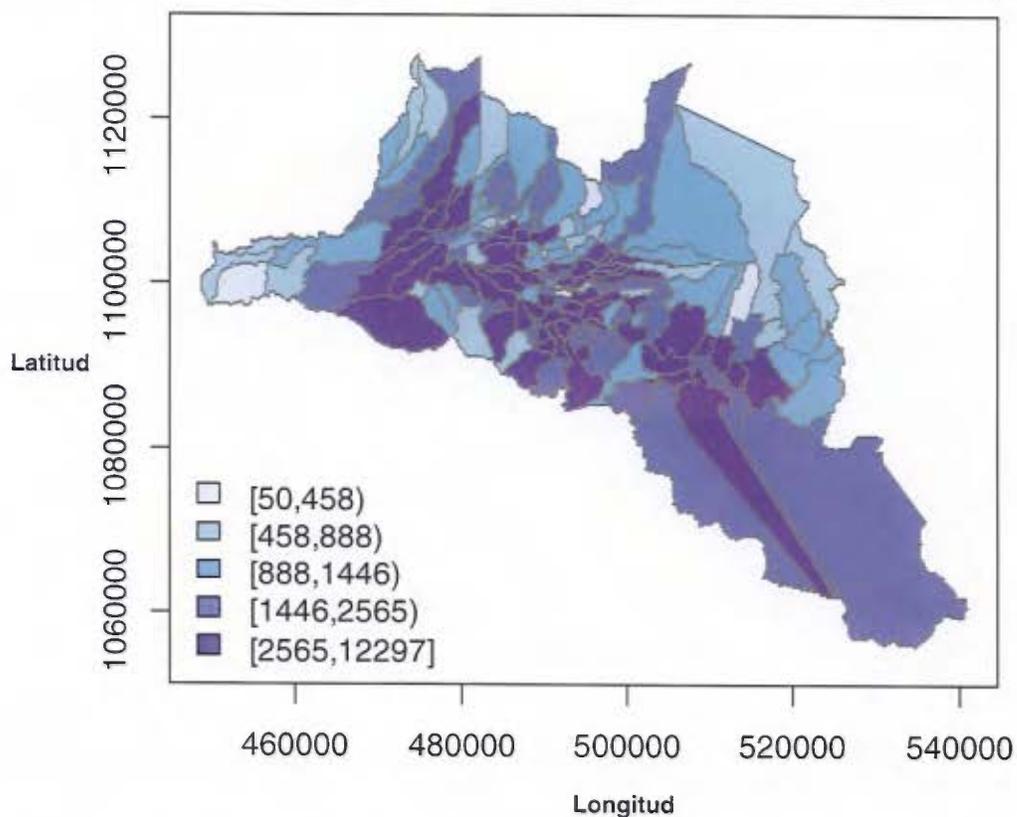


Fuente: Datos del Centro Centroamericano de Población (CCP), 2011.

Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

Es así como se identifica que la GAM contiene la mayor cantidad de distritos con población en edad colegial ubicados con valores mayores o iguales a 1 446 personas (III quintil). Este comportamiento disminuye de forma leve en la periferia de la GAM, principalmente en los sectores Norte y Oeste.

Mapa 14. GAM: población en edad colegial de 12 a 21 años por rangos, 2011



Fuente: Datos del Centro Centroamericano de Población (CCP), 2011.

Nota: Los límites de los rangos están definidos por los quintiles de la variable. Tanto la latitud como la longitud están medidas en términos de la proyección CRTM05.

6. 2 Análisis del modelo SAR

Recuérdese que en los modelos presentados en esta sección se busca identificar la asociación entre el porcentaje de deserción colegial con variables propias del programa social *Avancemos* y con otras variables de interés utilizadas como *proxy* y de control. Además, se tuvo que aproximar el valor del porcentaje de deserción colegial con base en los datos del Censo 2011 ya que los registros administrativos identificaban la deserción de acuerdo con la ubicación de colegios por distrito.

Antes de detallar el análisis del modelo SAR es conveniente aclarar que debido al estudio de la variable dependiente y las variables independientes, se decidió elaborar 6 modelos SAR descritos al inicio de este capítulo. Primeramente, un modelo SAR para el 2011, seguido de un modelo SAR de las diferencias entre los valores correspondientes al 2007 y 2011, además de un modelo SAR similar al planteado para el 2011 pero modificando la variable dependiente tomando en cuenta el cambio en el porcentaje de deserción entre el 2012 y 2011. Posteriormente, debido al estudio de las variables relacionadas con *Avancemos* se decidió calcular otros tres modelos SAR 2011 controlando por variables censales manteniendo la variable dependiente con datos administrativos y luego, cambiando la variable dependiente por el porcentaje de deserción según el censo ya que se buscaba refinar la especificación del modelo. Además, se ensayó con la inclusión y exclusión de la variable independiente monto promedio de becas en miles de colones para observar si la variable tasa de la cantidad de becas resultaba ser significativa en ausencia de la primera.

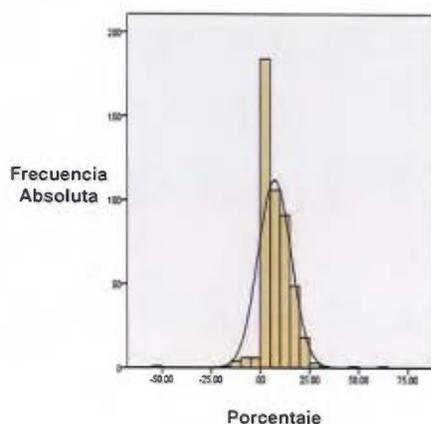
6. 2. 1 Modelos con variable dependiente definida con datos administrativos

En esta sección, se realizaron 4 modelos SAR a saber: el modelo SAR 2011, el modelo SAR de las diferencias al 2007 y 2011, el modelo SAR 2011 con variable dependiente correspondiente al cambio en el porcentaje de deserción entre el 2012 y 2011, y por último, el modelo SAR 2011 incluyendo variables independientes obtenidas del Censo 2011 utilizadas como variables de control. Para estos 4 modelos, la variable dependiente porcentaje de deserción fue obtenida utilizando los datos administrativos del MEP. Si bien es cierto los coeficientes de los modelos y sus valores de probabilidad son distintos, el común denominador es que en ninguno de estos modelos propuestos se encuentra asociación entre el programa social *Avance*mos y el porcentaje de deserción debido a que el valor de lambda resultó ser muy bajo (menor a 0,130) y a su vez no significativo en todos los modelos con un nivel de significancia del 5%, lo cual indica que no existe autocorrelación espacial en los residuos de estos modelos SAR propuestos en un inicio.

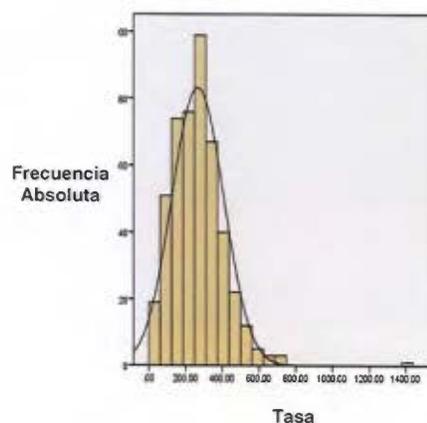
En general, las variables incluidas para el modelo SAR 2011 presentan las siguientes características: los valores que tiene el porcentaje de deserción están centralizados y poseen poca variabilidad, con la particularidad de la existencia de muchos ceros debido a la ausencia de al menos un colegio en 115 distritos, como se describió al inicio de este Capítulo. La tasa de la cantidad de becas en su mayoría agrupa valores hacia al inicio de la distribución, caso contrario sucede con el monto promedio de becas en miles de colones ya que los valores se sitúan a la derecha de la distribución. El consumo promedio de electricidad aglutina valores centralizados y posee pocos valores extremos. Por otra parte, tanto la cantidad de colegios como la tasa total de profesores tienen la mayor cantidad de valores al inicio de la distribución presentando un comportamiento asimétrico en ambos casos. Por último, para controlar valores extremos en la población en edad colegial de 12 a 21 años se calculó el logaritmo natural para usarla como variable

independiente en todos los modelos. Con respecto a la prueba de normalidad Kolmogorov-Smirnoff, solamente para la variable población en edad colegial (logaritmo natural) no se rechaza la hipótesis nula, comprobándose en este caso que el conjunto de datos sigue una distribución normal (Gráfico 13).

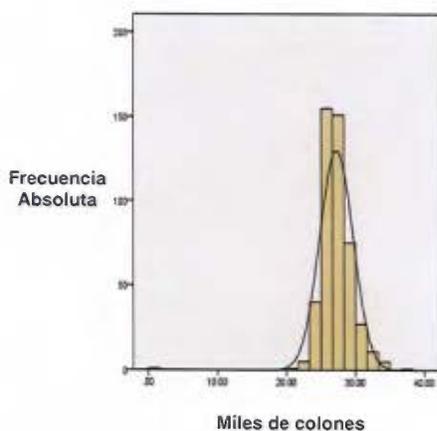
Gráfico 13. Histogramas de las variables del modelo SAR 2011.



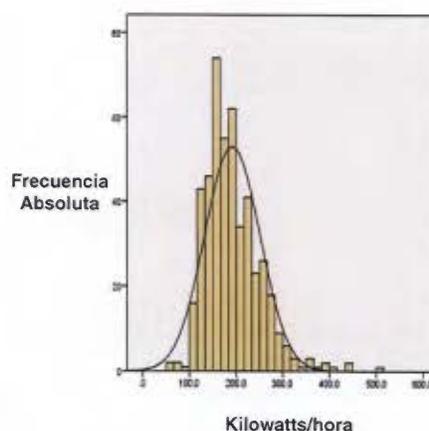
Porcentaje de deserción
(p-value de K-S = 0,000)



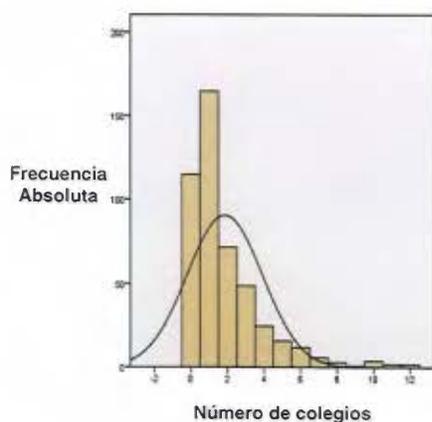
Tasa cantidad de becas
(p-value de K-S = 0,000)



Monto promedio de becas
(p-value de K-S = 0,000)

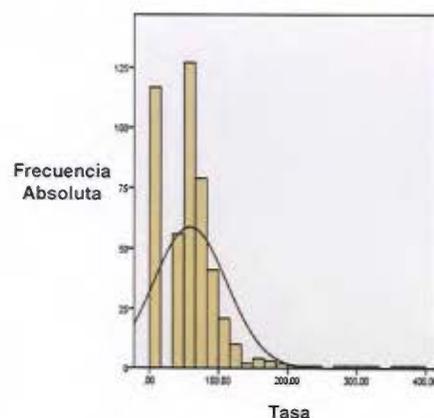


Consumo promedio de electricidad
(p-value de K-S = 0,000)



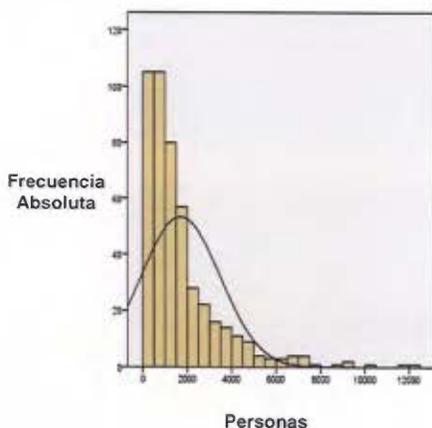
Cantidad de colegios

(p-value de K-S = 0,000)



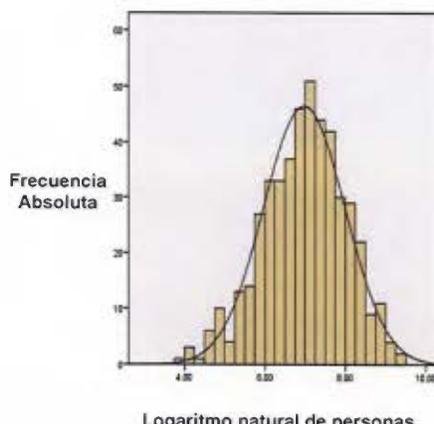
Tasa total de profesores

(p-value de K-S = 0,000)



Población en edad colegial

(p-value de K-S = 0,000)



Población en edad colegial (ln)

(p-value de K-S = 0,200)

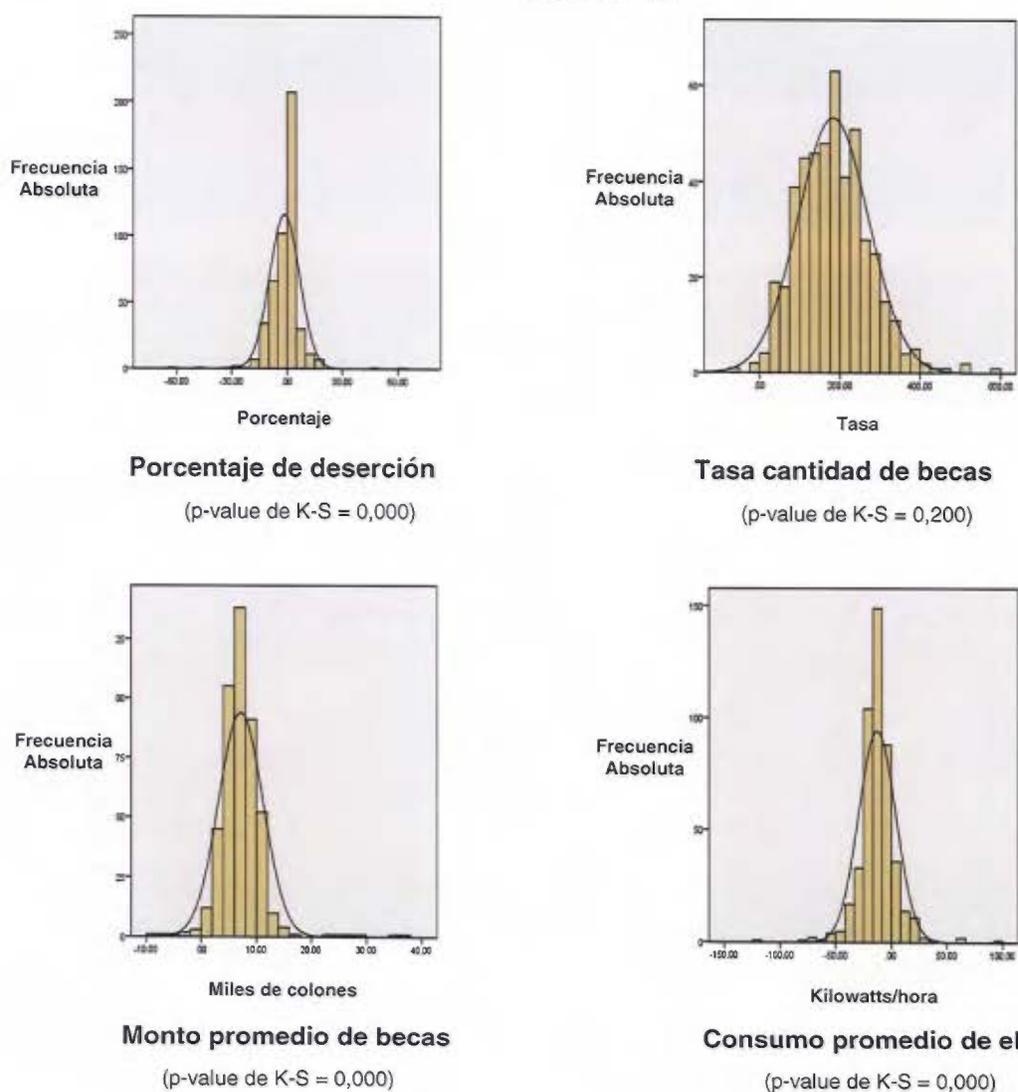
Fuente: Elaboración propia con base en estadísticas del Ministerio de Educación Pública (MEP), Instituto Mixto de Ayuda Social (IMAS), Ministerio de Planificación Nacional y Política Económica (Mideplan), Centro Centroamericano de Población (CCP), 2011.

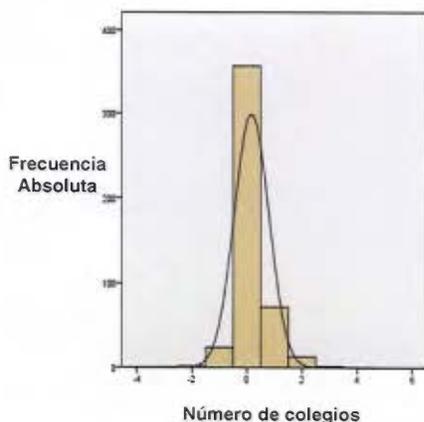
Nota: Se presentan las probabilidades asociadas a la prueba de Kolmogorov-Smirnov de bondad de ajuste de la normalidad debajo de cada variable.

En cambio, las variables incluidas para el modelo SAR de las diferencias entre los valores correspondientes al 2007 y 2011 contienen las siguientes características: el porcentaje de deserción mantiene los datos centralizados debido a la naturaleza y comportamiento implícito por la ausencia de al menos un colegio en los distritos. La tasa de la cantidad de becas logra agrupar sus valores

en el centro y alrededores de la distribución. Por otra parte, las variables cantidad de colegios y la tasa total de profesores gráficamente no presentan un comportamiento normal. Pero si se toma en cuenta la prueba de normalidad Kolmogorov-Smirnoff, para las variables tasa de la cantidad de becas y la población en edad colegial (logaritmo natural) no se rechaza la hipótesis nula, validándose que los datos siguen una distribución normal (Gráfico 14).

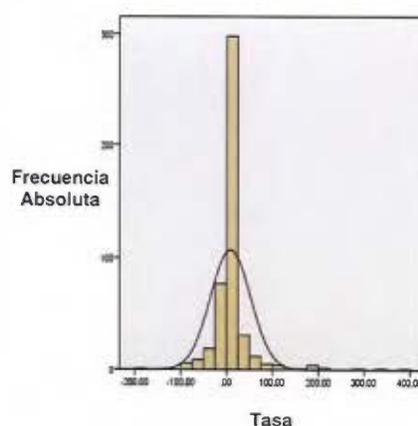
Gráfico 14. Histogramas de las variables del modelo SAR de las diferencias 2007-2011.





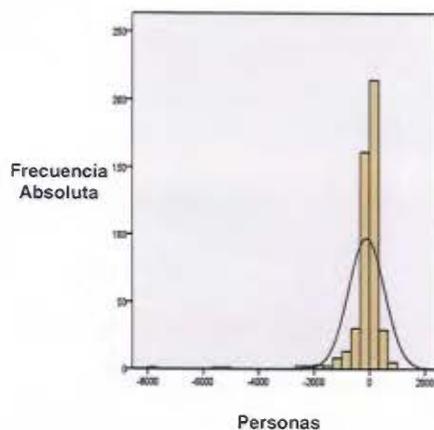
Cantidad de colegios

(p-value de K-S = 0,000)



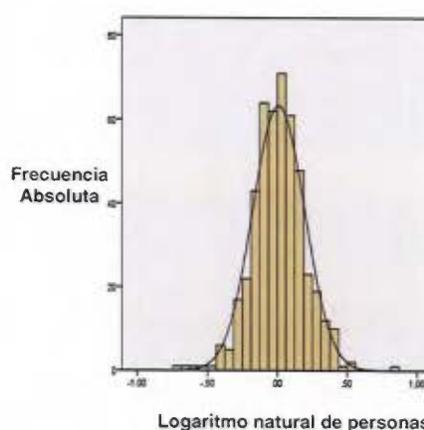
Tasa total de profesores

(p-value de K-S = 0,000)



Población en edad colegial

(p-value de K-S = 0,000)



Población en edad colegial (ln)

(p-value de K-S = 0,052)

Fuente: Elaboración propia con base en estadísticas del Ministerio de Educación Pública (MEP), Instituto Mixto de Ayuda Social (IMAS), Ministerio de Planificación Nacional y Política Económica (Mideplan), Centro Centroamericano de Población (CCP), 2011.

Nota: Se presentan las probabilidades asociadas a la prueba de Kolmogorov-Smirnov de bondad de ajuste de la normalidad debajo de cada variable.

Siguiendo con el análisis espacial, cabe señalar que debido a la naturaleza de la información geográfica presente en esta investigación se decidió utilizar la matriz binaria de contigüidad del tipo reina o *queen*, cuya característica principal es que las unidades espaciales comparten frontera en común que en este caso son los distritos de Costa Rica y los pesos utilizados están estandarizados por fila

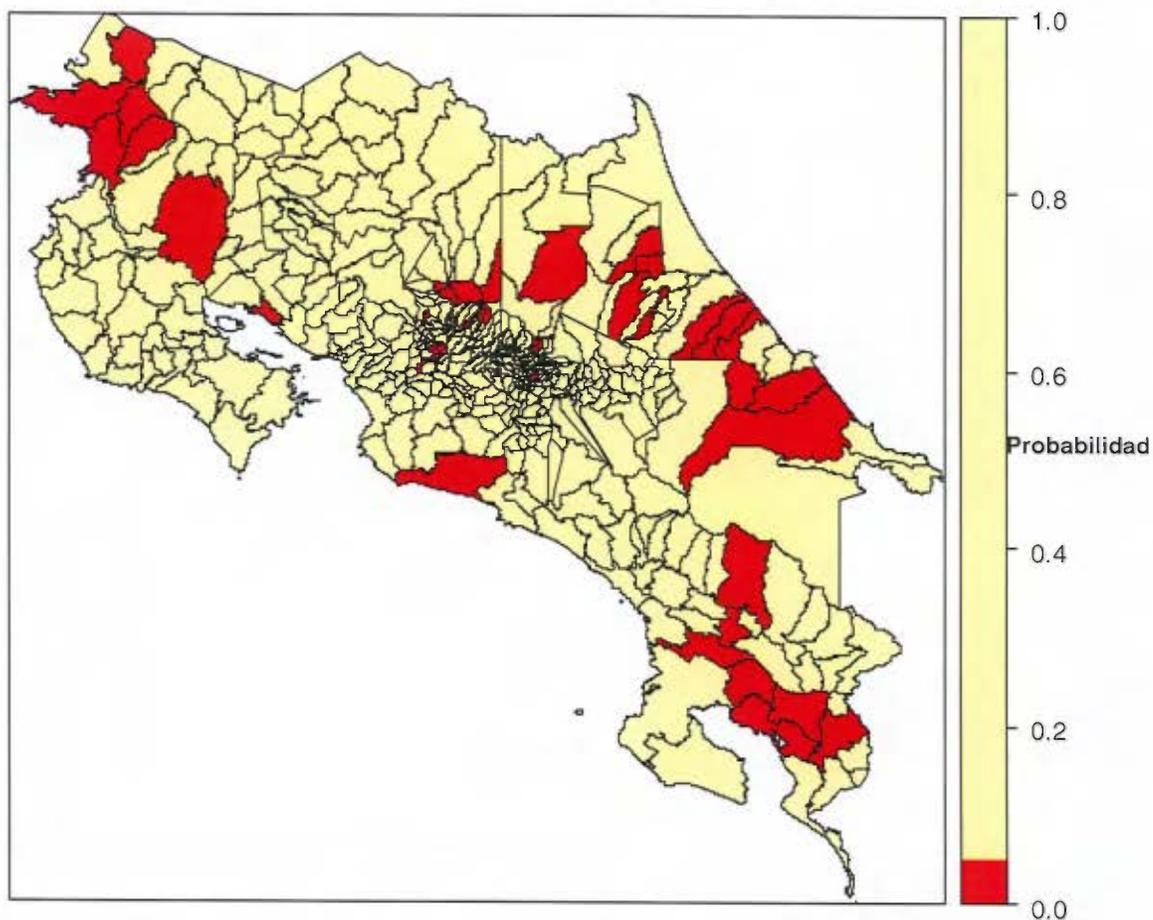
que contiene la suma de todos los pesos, la cual es igual al número de áreas. A su vez, se recuerda que para el caso de aplicación, se considera que los estudiantes acuden a los colegios ubicados en los distritos más cercanos, los cuales comparten frontera en común.

Es necesario calcular los índices de autocorrelación I de Moran para verificar si existe o no autocorrelación en el patrón espacial del porcentaje de deserción bajo el supuesto de vecinos con frontera en común. La I de Moran bajo el supuesto de normalidad y de asignación aleatoria obtuvo un valor de 3,77 y 3,81 respectivamente, resultando ambos significativos, lo cual indica que existe autocorrelación en el patrón espacial del porcentaje de deserción, es decir, se confirma el rechazo de la hipótesis de no autocorrelación, por lo que se concluye que hay evidencia estadística para afirmar que existe autocorrelación espacial entre los distritos de Costa Rica en cuanto al porcentaje de deserción en el 2011.

Como la autocorrelación espacial puede surgir por un proceso de aglomeración, con la autocorrelación local se puede identificar algunas zonas geográficas o distritos que presenten este tipo de problema. En el mapa de Costa Rica del Índice de Moran a nivel local, las zonas de color rojo indican las aglomeraciones de una forma más clara y éstas se ubican principalmente en la zona Atlántica, la parte Sur y algunas zonas del Norte, Centro y Noroeste de nuestro país. Las aglomeraciones son aquellas zonas geográficas donde hay presencia de autocorrelación espacial, es decir, donde existen más problemas con los distritos localizados en las zonas geográficas de color rojo, o sea, los niveles de la deserción colegial son similares en distritos vecinos, por lo que hay una agrupación geográfica que puede ser sujeta a política pública en forma más eficiente. El aporte que brinda este análisis puede ser visto de dos formas: en primer lugar, los datos a nivel agregado son más fáciles de conseguir por lo que un análisis así se puede repetir más fácilmente sin necesidad de encuestas y, en segundo lugar, detectar que hay un proceso de aglomeración geográfica debería

facilitar la implementación de políticas pues se puede focalizar a distritos similares.

Mapa 15. Índice de Moran a nivel local del porcentaje de deserción, 2011



Fuente: Elaboración propia, 2011.

6. 2. 1. 1 Comportamiento de variables y supuestos del modelo SAR 2011

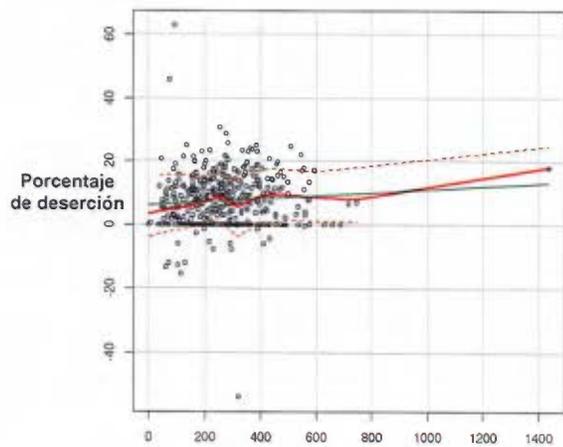
Primeramente, se analiza de manera visual la dispersión existente entre cada una de las variables independientes y la variable dependiente. Esto sirve para identificar la relación que tienen las variables entre sí (Gráfico 15).

Cabe resaltar que se utilizó una matriz reina o *matrix queen* para los modelos. El análisis de la robustez del modelo SAR mostró que la selección de la matriz de pesos no afecta los resultados del modelo SAR, así que hubiera sido indiferente el tipo seleccionado de matriz. Adicionalmente, se analizará la normalidad de las variables, así como la normalidad de los residuos. No obstante, las simulaciones también mostraron que el modelo es robusto al no cumplimiento del supuesto de normalidad. Esto quiere decir que, aun cuando no se cumpliera el supuesto distribucional, las conclusiones inferenciales de los modelos SAR estimados siguen siendo interpretables.

Pasando ya a los datos empíricos, la tasa de la cantidad de becas y el porcentaje de deserción mantienen un comportamiento estable conforme se aumenta la primera de ellas. El monto promedio de becas en miles de colones presenta poca variabilidad y esto refleja el agrupamiento existente con respecto al porcentaje de deserción. También existe estabilidad entre el consumo promedio de electricidad y el porcentaje de deserción, solamente con valores altos en el consumo promedio de electricidad se observa una disminución en el porcentaje de deserción.

Luego, se visualiza que entre menor sea la cantidad de colegios en el distrito existe más variabilidad en el porcentaje de deserción. Con respecto a la tasa total de profesores, el porcentaje de deserción se ubica mayoritariamente en los valores más pequeños de dicha tasa. Y, por último, se observa un comportamiento creciente entre la población en edad colegial de 12 a 21 años (logaritmo natural) y el porcentaje de deserción, es decir, conforme aumenta la población en edad colegial, el porcentaje de deserción tiende a aumentar. Cabe señalar que el programa social *Avancemos* está enfocado en atender a estudiantes entre 12 a 21 años que estén dentro del sistema educativo y que posiblemente los mayores de edad sean destinatarios importantes del programa para que logren concluir sus estudios de forma completa.

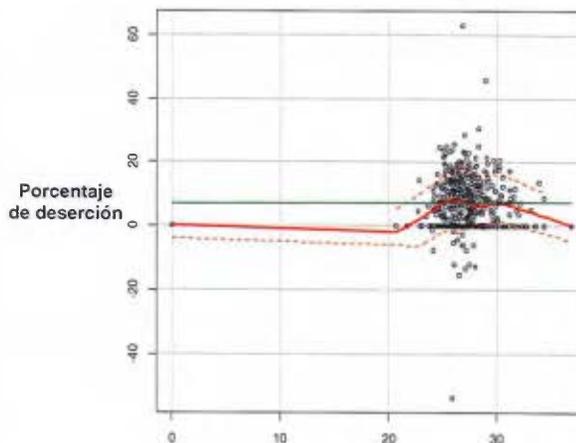
Gráfico 15. Diagramas de dispersión del porcentaje de deserción en el 2011 por cada variable predictora, por distritos de Costa Rica.



Tasa

Tasa cantidad de becas

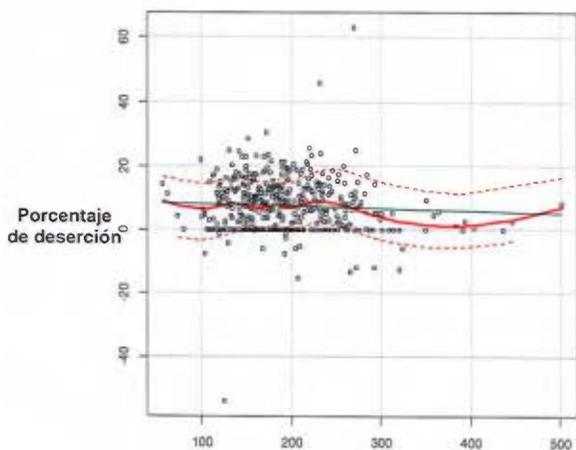
(r de Pearson = 0,0839)



Miles de colones

Monto promedio de becas

(r de Pearson = 0,0049)

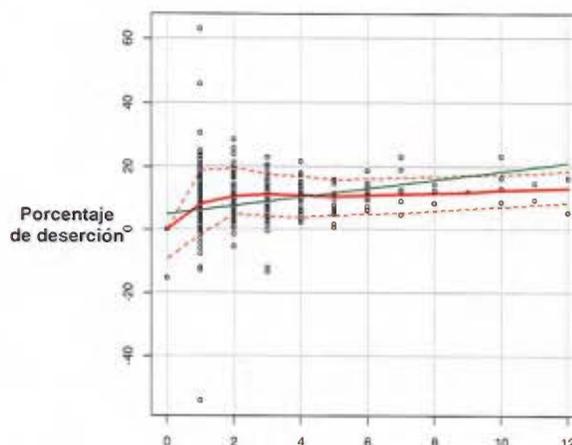


Porcentaje de deserción

Kilowatts/hora

Consumo promedio de electricidad

(r de Pearson = -0,0504)

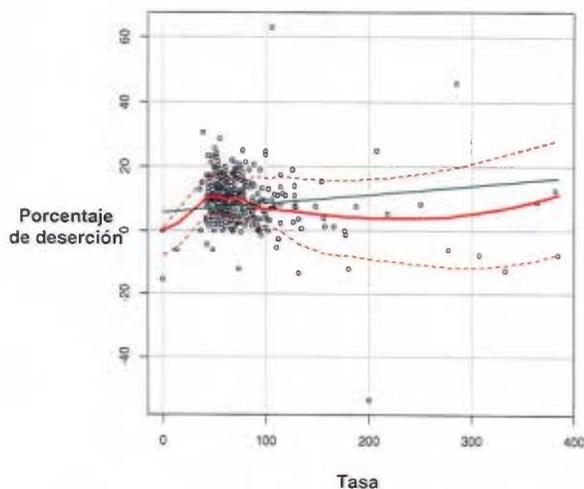


Porcentaje de deserción

Número de colegios

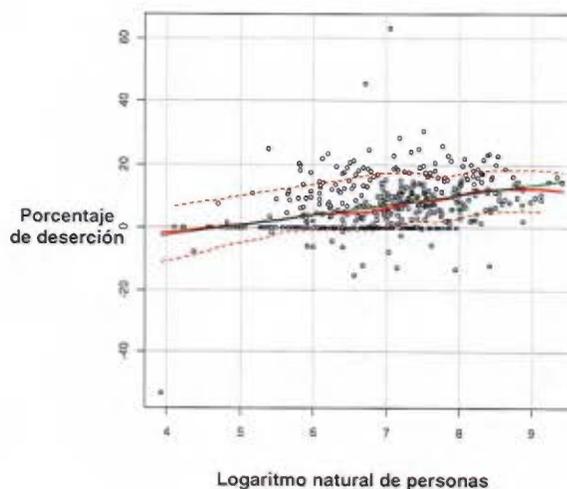
Cantidad de colegios

(r de Pearson = 0,3349)



Tasa total de profesores

(r de Pearson = 0,1728)



Población en edad colegial (ln)

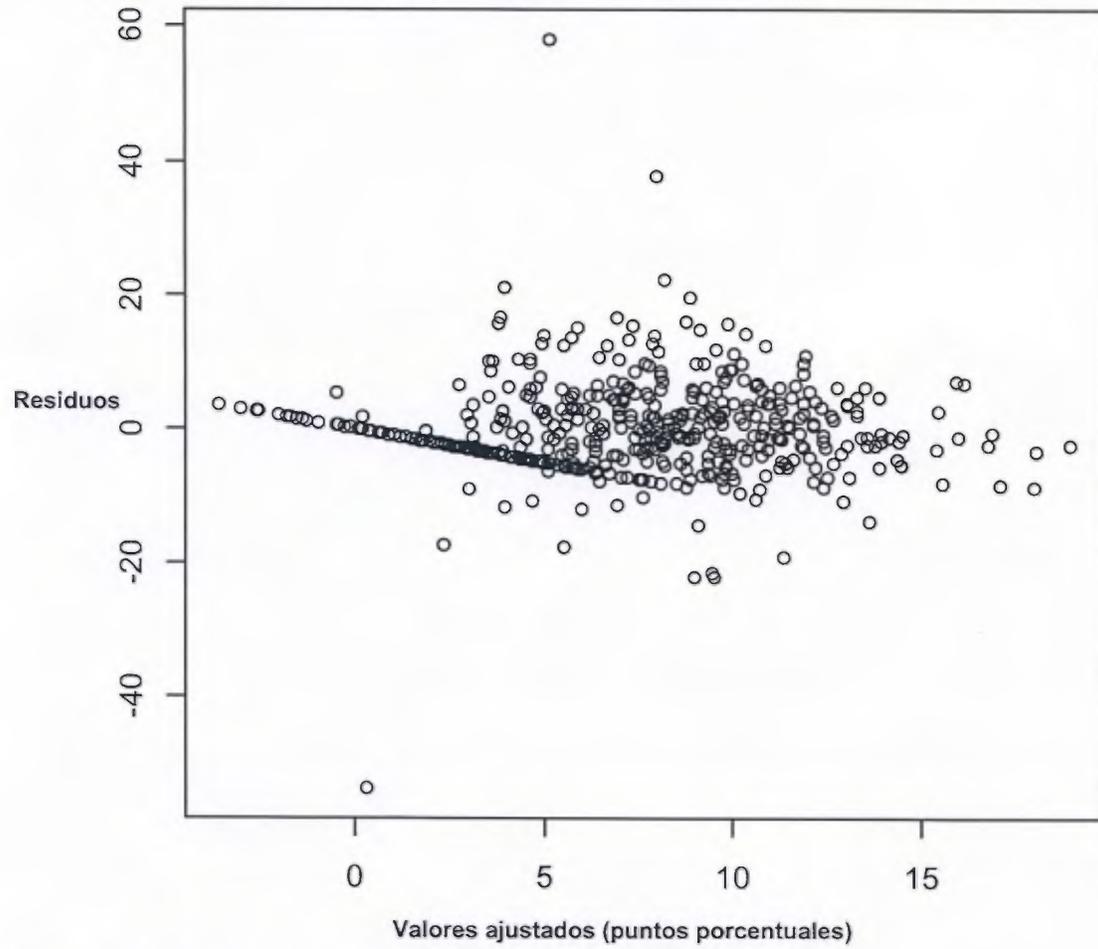
(r de Pearson = 0,3840)

Fuente: Elaboración propia, 2011.

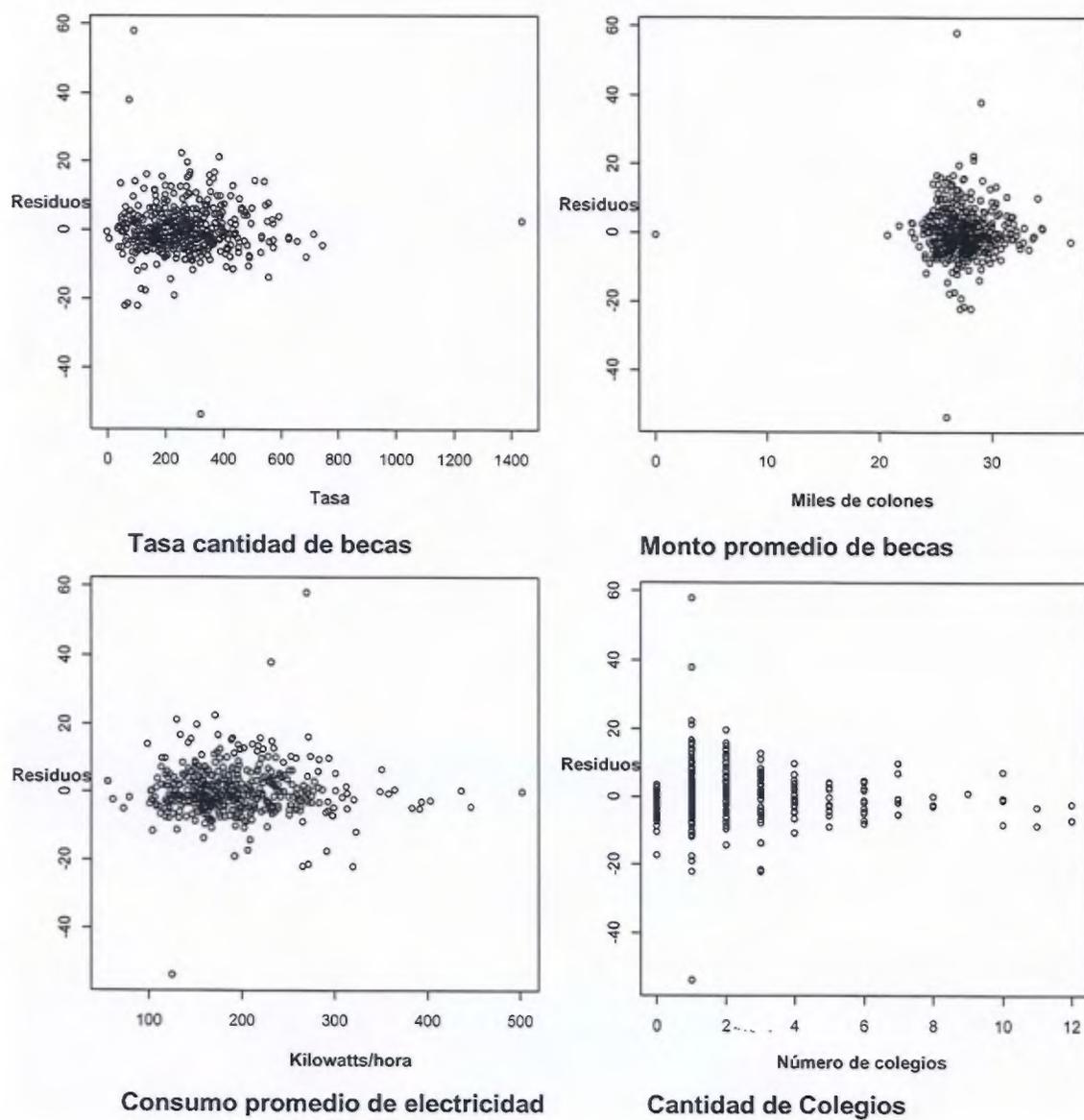
Nota: Se presentan los valores del coeficiente de correlación r de Pearson debajo de cada variable.

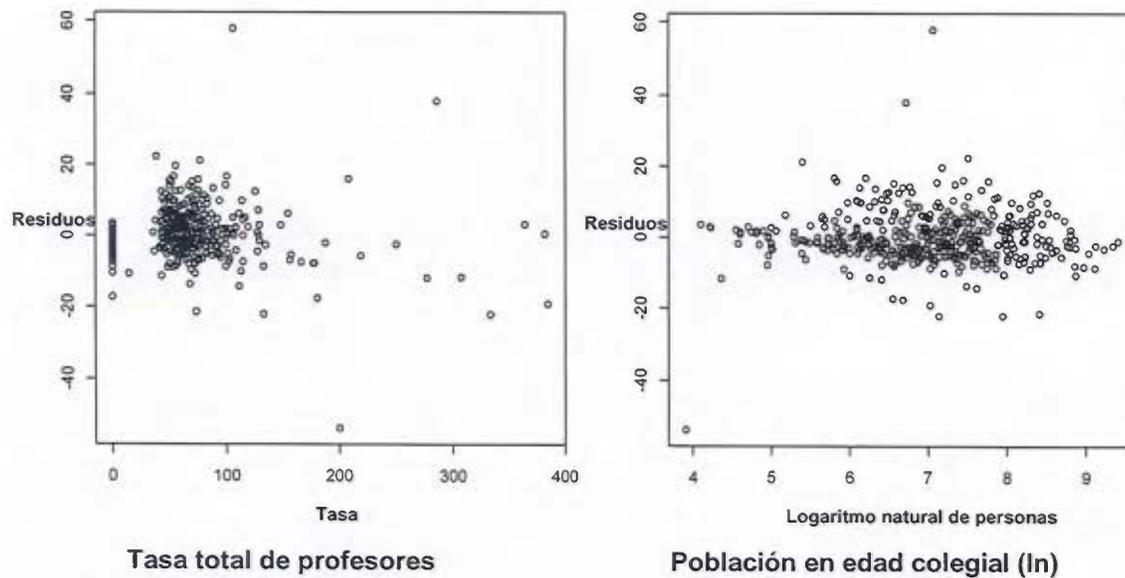
Siguiendo con el análisis correspondiente, los supuestos de los residuos del modelo SAR 2011 con respecto a sus valores ajustados se comprueban en el gráfico que compara ambos datos. Es importante señalar que se grafican los residuos y valores ajustados, aun sin presentar los resultados finales de los modelos, pues se está analizando los supuestos de los modelos. En el escenario idóneo, el gráfico debería mostrar una nube de puntos sin patrón alguno lo cual significa que los residuos son aleatorios con varianza constante, no obstante, el resultado obtenido no parece ser tan evidente, ya que la secuencia casi rectilínea presente hacia la derecha del gráfico es producto de aquellos distritos que no tienen presencia de al menos un colegio, lo cual repercute en la variable dependiente porcentaje de deserción (Gráfico 16).

Gráfico 16. Valores ajustados vs. Residuos del modelo SAR 2011.



Fuente: Elaboración propia, 2011.

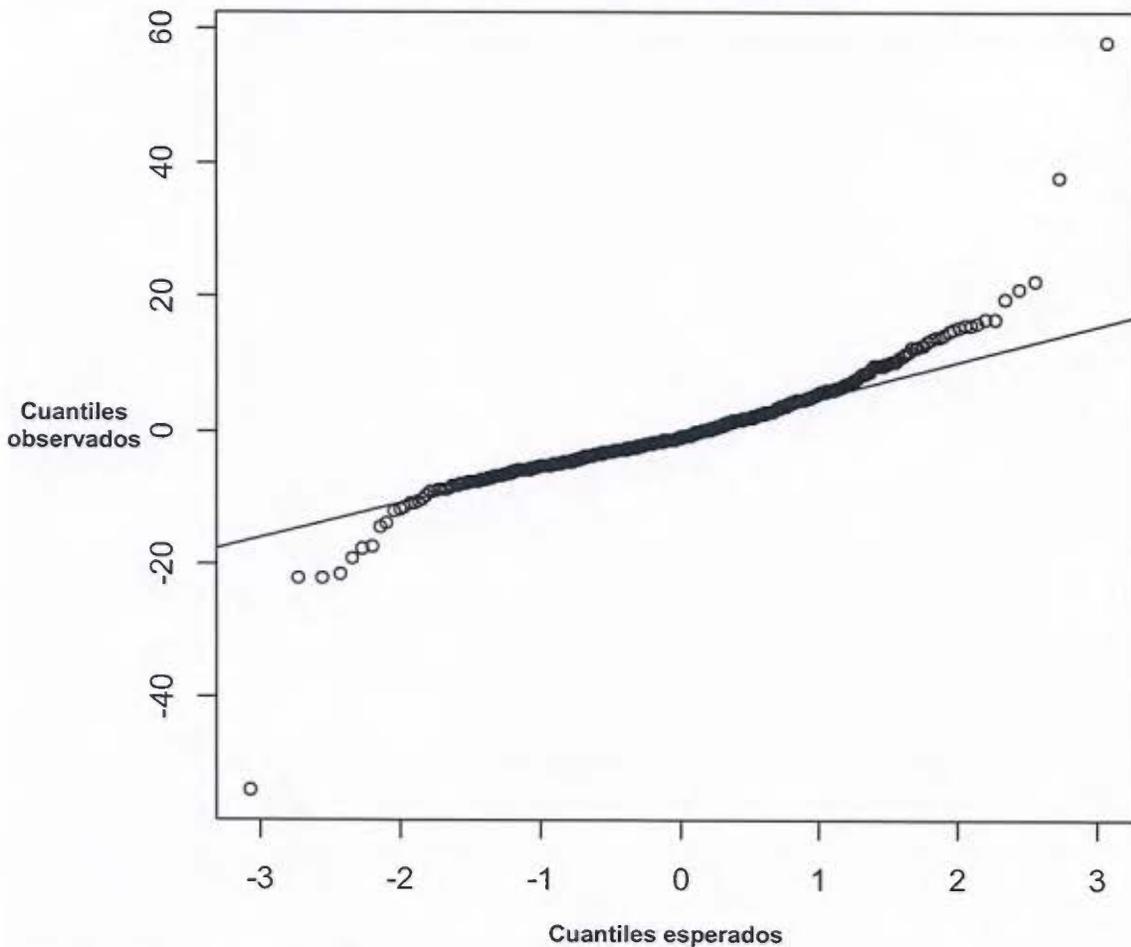
Gráfico 17. Variables predictoras vs. Residuos del modelo SAR 2011.



Fuente: Elaboración propia, 2011.

Con respecto al supuesto de normalidad de los residuos, el gráfico de cuantiles normales debería ser una diagonal bien marcada por los puntos. En este caso, los puntos no están completamente alineados en forma diagonal, por lo tanto de forma visual se descarta el supuesto de normalidad. Además, los residuos parecen tener una distribución simétrica, pero con muchos valores extremos hacia ambas colas (Gráfico 18). A su vez, el supuesto de normalidad de los residuos fue examinado con la prueba Kolmogorov-Smirnoff, los resultados arrojaron que el valor de probabilidad asociado fue de 0,000 por lo que con un nivel de significancia del 5% se rechaza la hipótesis nula de normalidad.

Gráfico 18. Gráfico de Normalidad de los residuos del modelo SAR 2011.



Fuente: Elaboración propia, 2011.

Recapitulando, se analizó el supuesto de normalidad para cada una de las variables que conforman el modelo SAR 2011, mediante la prueba Kolmogorov-Smirnoff, cuya hipótesis nula es que los datos tienen una distribución normal. Los resultados para todas las variables en cuestión arrojaron que existe evidencia estadística para rechazar la hipótesis nula, por lo tanto, se concluye que los datos de cada una de las variables no tienen una distribución normal, excepto para la variable población en edad colegial (logaritmo natural), ya que con un nivel de significancia del 5% no se rechazó la hipótesis nula.

Y para finalizar este análisis previo del modelo SAR 2011, se calculó la I de Moran para cada una de las variables correspondientes. Los resultados señalan que existe evidencia estadística para afirmar que hay autocorrelación espacial entre los distritos de Costa Rica para cada una de las variables analizadas.

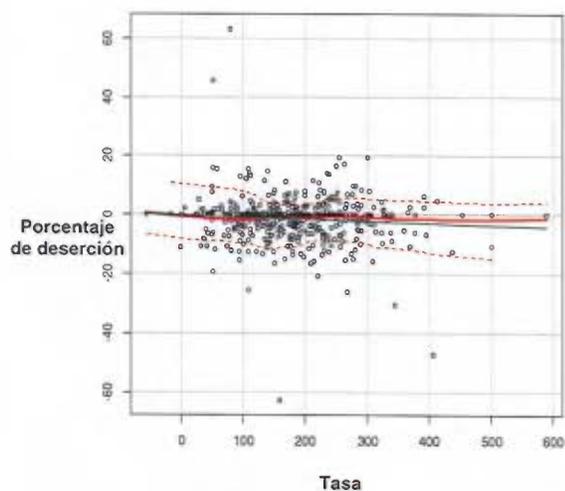
6. 2. 1. 2 Comportamiento de variables y supuestos del modelo SAR de las diferencias entre el 2007 y 2011

Para iniciar este apartado, es necesario tomar en cuenta la dispersión que tienen las variables predictoras involucradas en este modelo SAR de las diferencias entre el 2007 y 2011 para identificar la relación que pueden tener con la variable dependiente que en este caso es la diferencia en el porcentaje de deserción entre el 2007 y 2011 (Gráfico 19).

Se observa que la diferencia en la tasa de la cantidad de becas mantiene un comportamiento estable de forma horizontal con respecto a la diferencia en el porcentaje de deserción. También, la diferencia en el monto promedio de becas en miles de colones agrupa valores en un intervalo pequeño debido a la poca variabilidad existente en relación con la diferencia en el porcentaje de deserción, esto también ocurre observando lo que sucede con la diferencia en el consumo promedio de electricidad.

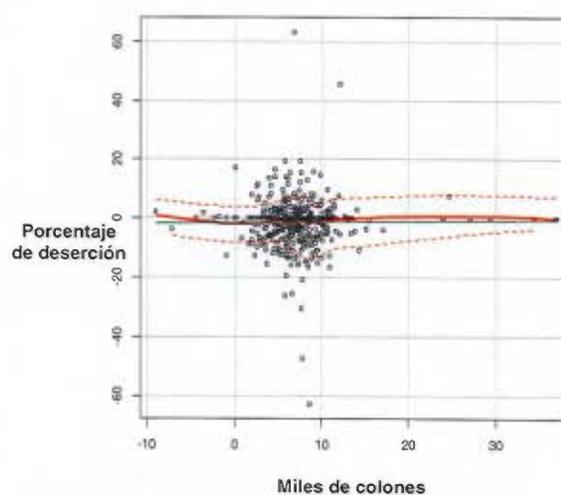
Además, con respecto a la diferencia en la cantidad de colegios, son pocos los casos ubicados en dos o más colegios debido a que los distritos mantuvieron en su mayoría la misma cantidad de colegios durante el periodo. La relación existente comparando los datos de la población en edad colegial (logaritmo natural) y la diferencia en el porcentaje de deserción es estable sin grandes variaciones. Por último, la diferencia en la tasa total de profesores ubica la mayoría de sus datos en un intervalo pequeño.

Gráfico 19. Diagramas de dispersión del porcentaje de deserción entre el 2007 y 2011 por cada variable predictora, por distritos de Costa Rica.



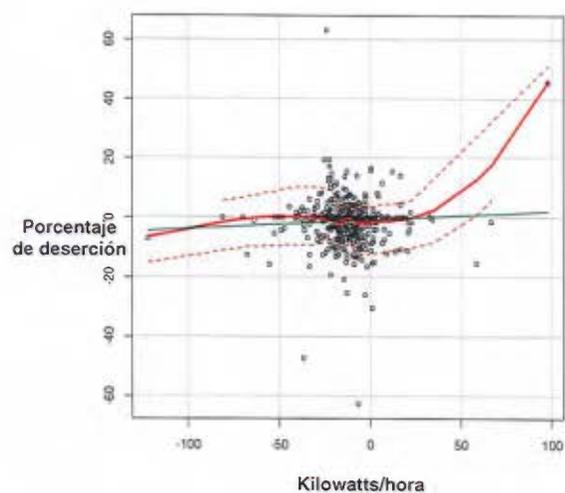
Tasa cantidad de becas

(r de Pearson = -0,0749)



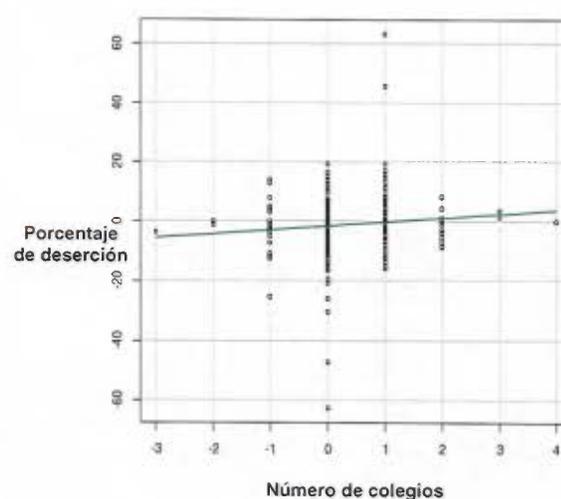
Monto promedio de becas

(r de Pearson = 0,0131)



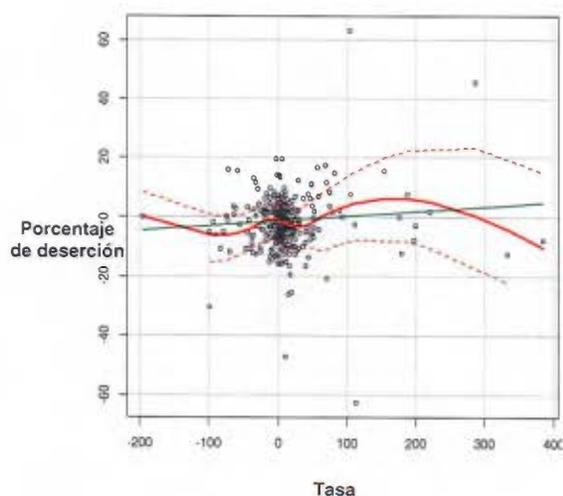
Consumo promedio de electricidad

(r de Pearson = 0,0621)



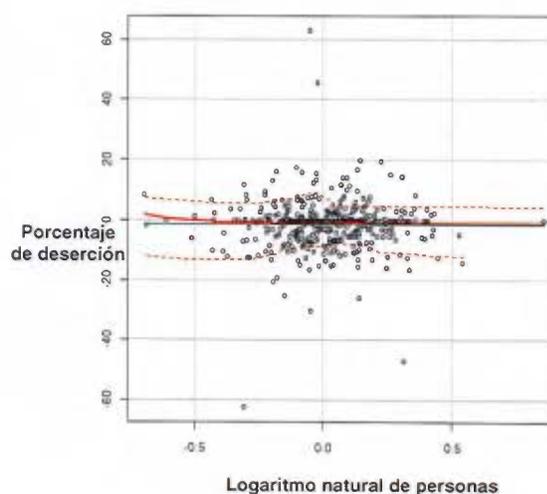
Cantidad de colegios

(r de Pearson = 0,1006)



Tasa total de profesores

(r de Pearson = 0,0862)



Población en edad colegial (ln)

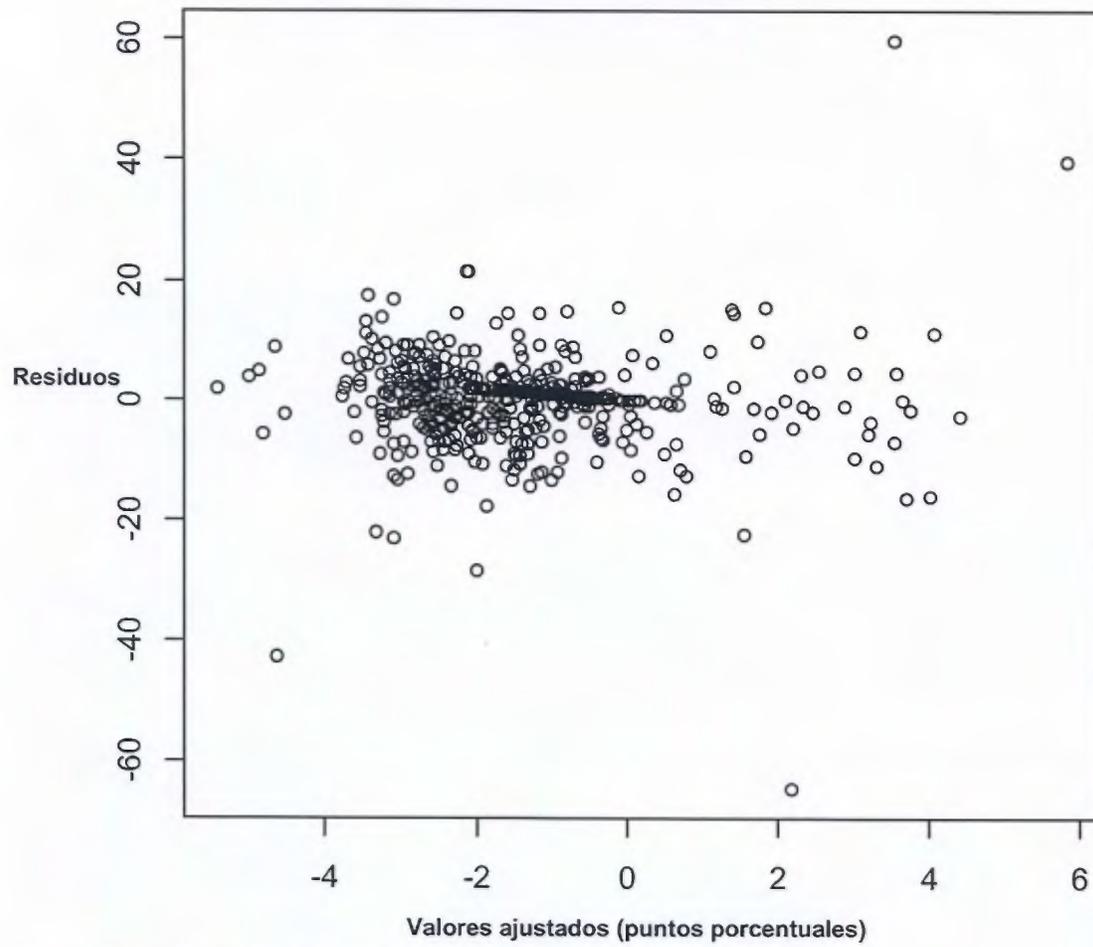
(r de Pearson = -0,003)

Fuente: Elaboración propia, 2011.

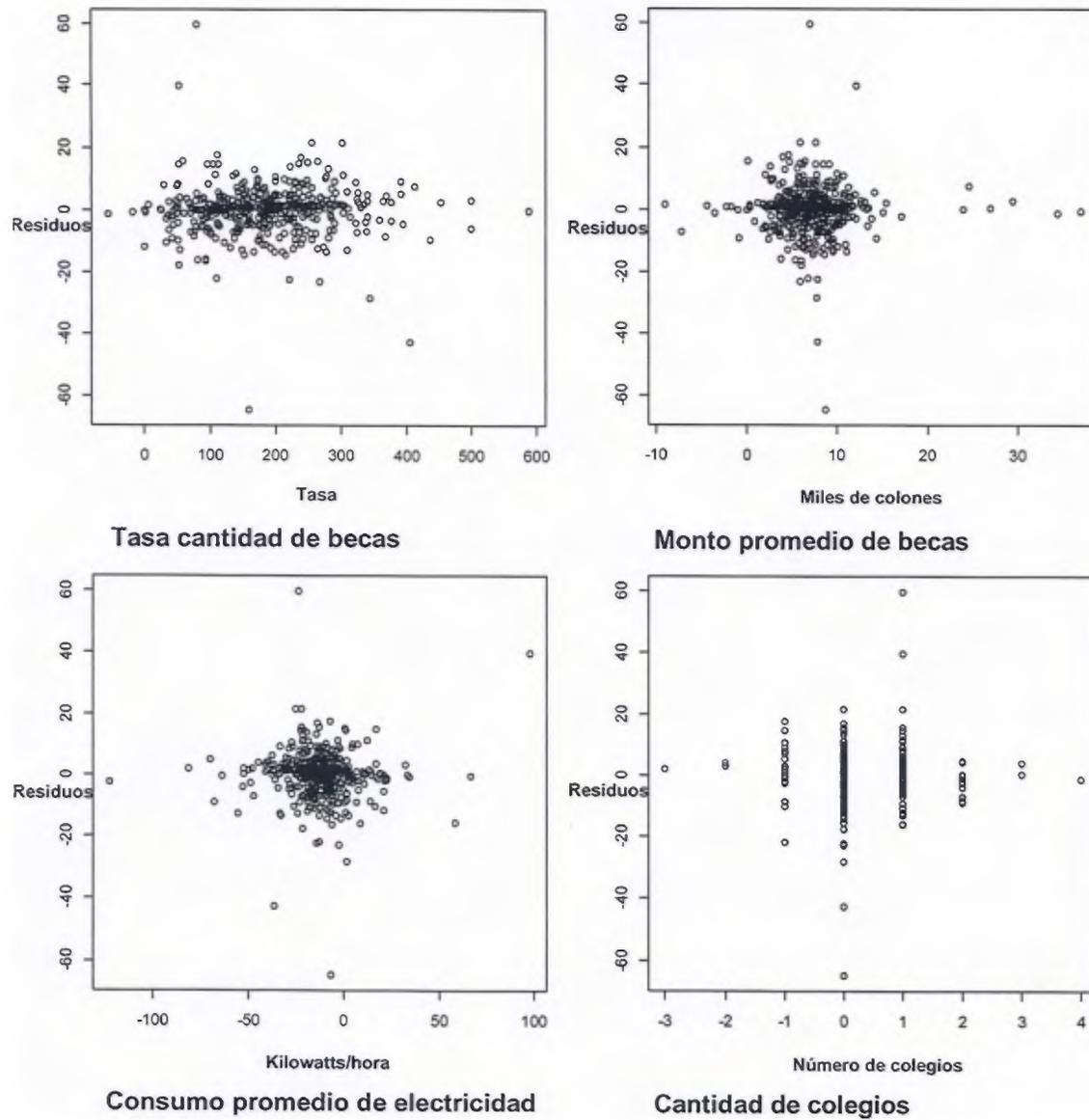
Nota: Se presentan los valores del coeficiente de correlación r de Pearson debajo de cada variable.

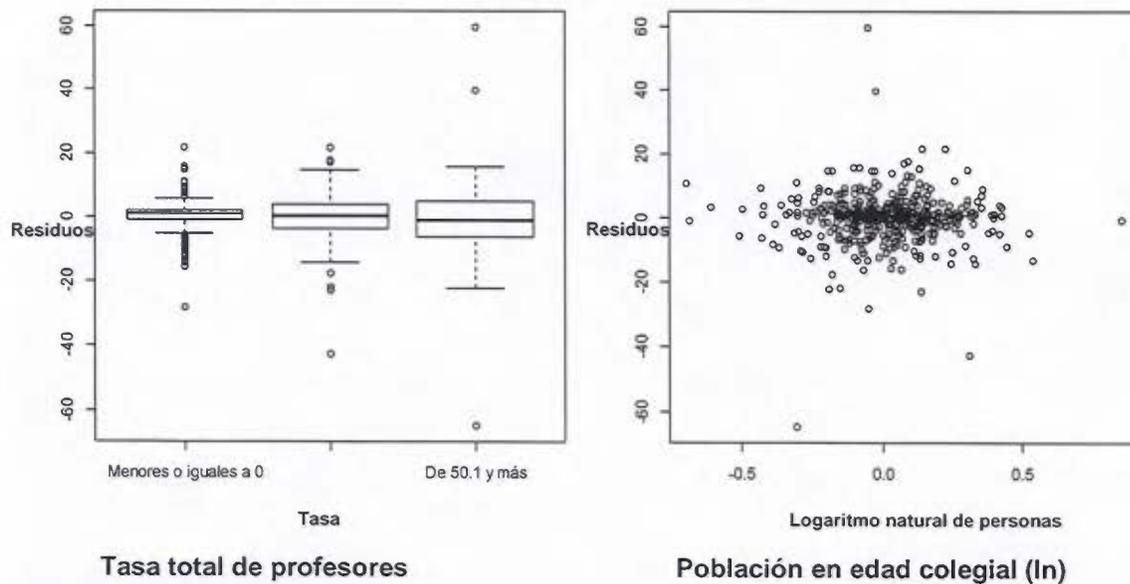
Posteriormente, es necesario analizar los supuestos de los residuos del modelo SAR de las diferencias entre el 2007 y 2011, y sus correspondientes valores ajustados de forma gráfica. En el caso óptimo, es necesario de que los puntos del gráfico no presenten ningún patrón para afirmar que los residuos son aleatorios con varianza constante. Al igual que sucedió con el modelo SAR 2011, lo obtenido difiere del caso óptimo, aunque en este caso, se observa más aleatorio (Gráficos 20 y 21).

Gráfico 20. Valores ajustados vs. Residuos del modelo SAR de las diferencias entre el 2007 y 2011.



Fuente: Elaboración propia, 2011.

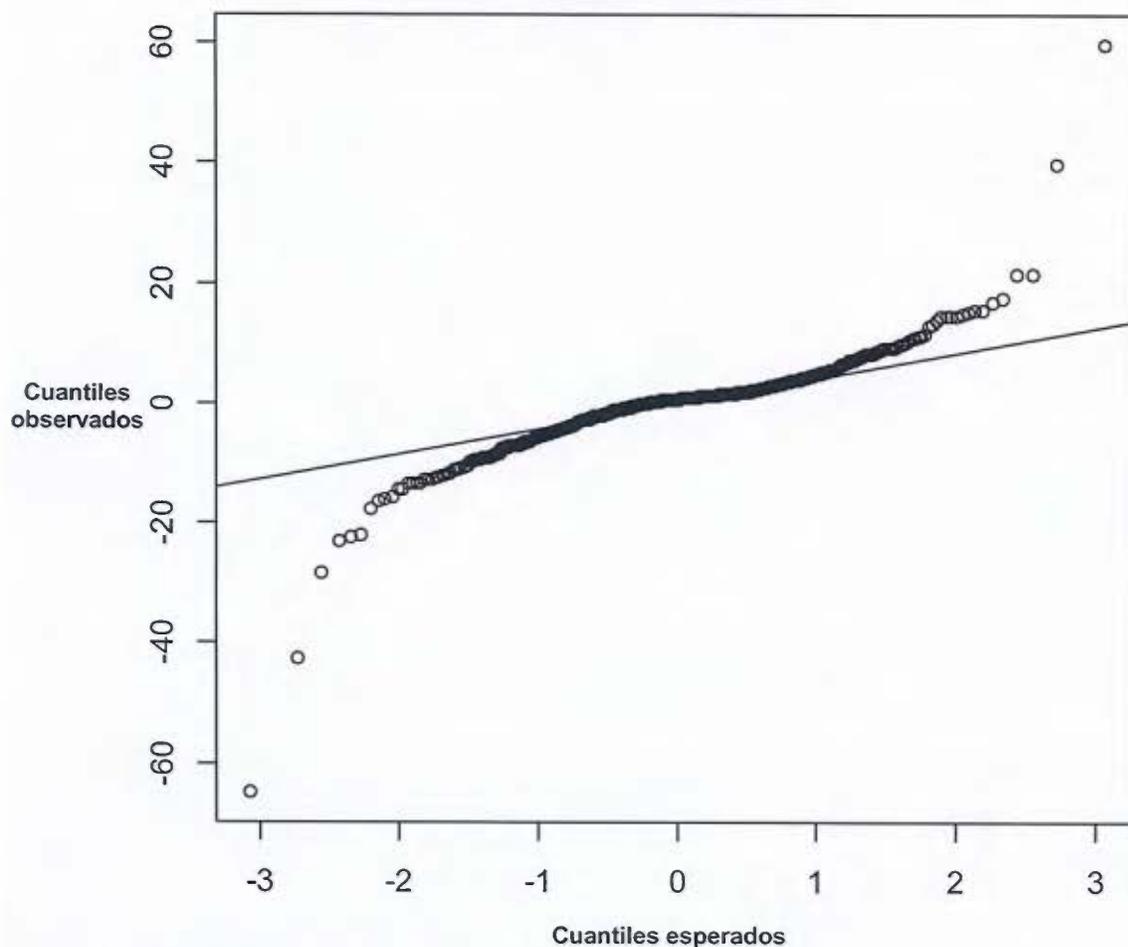
Gráfico 21. Variables predictoras vs. Residuos del modelo SAR 2007 y 2011.



Fuente: Elaboración propia, 2011.

Seguidamente, el supuesto de normalidad puede ser valorado de forma visual utilizando el gráfico de cuantiles normales y es necesario que los puntos estén alineados en diagonal. Por lo tanto, observando el gráfico se descarta el supuesto de normalidad debido a que los puntos no siguen la línea diagonal (Gráfico 22). Además, para comprobar este supuesto fue necesario utilizar la prueba de Kolmogorov-Smirnoff, la cual indica que la hipótesis nula es que los datos tienen una distribución normal, el valor de la probabilidad asociada fue de 0,000 y por lo tanto, se rechaza la hipótesis nula con un nivel de significancia del 5%, comprobándose que los residuos del modelo no presentan una distribución normal.

Gráfico 22. Gráfico de Normalidad de los residuos del modelo SAR 2007 y 2011.



Fuente: Elaboración propia, 2011.

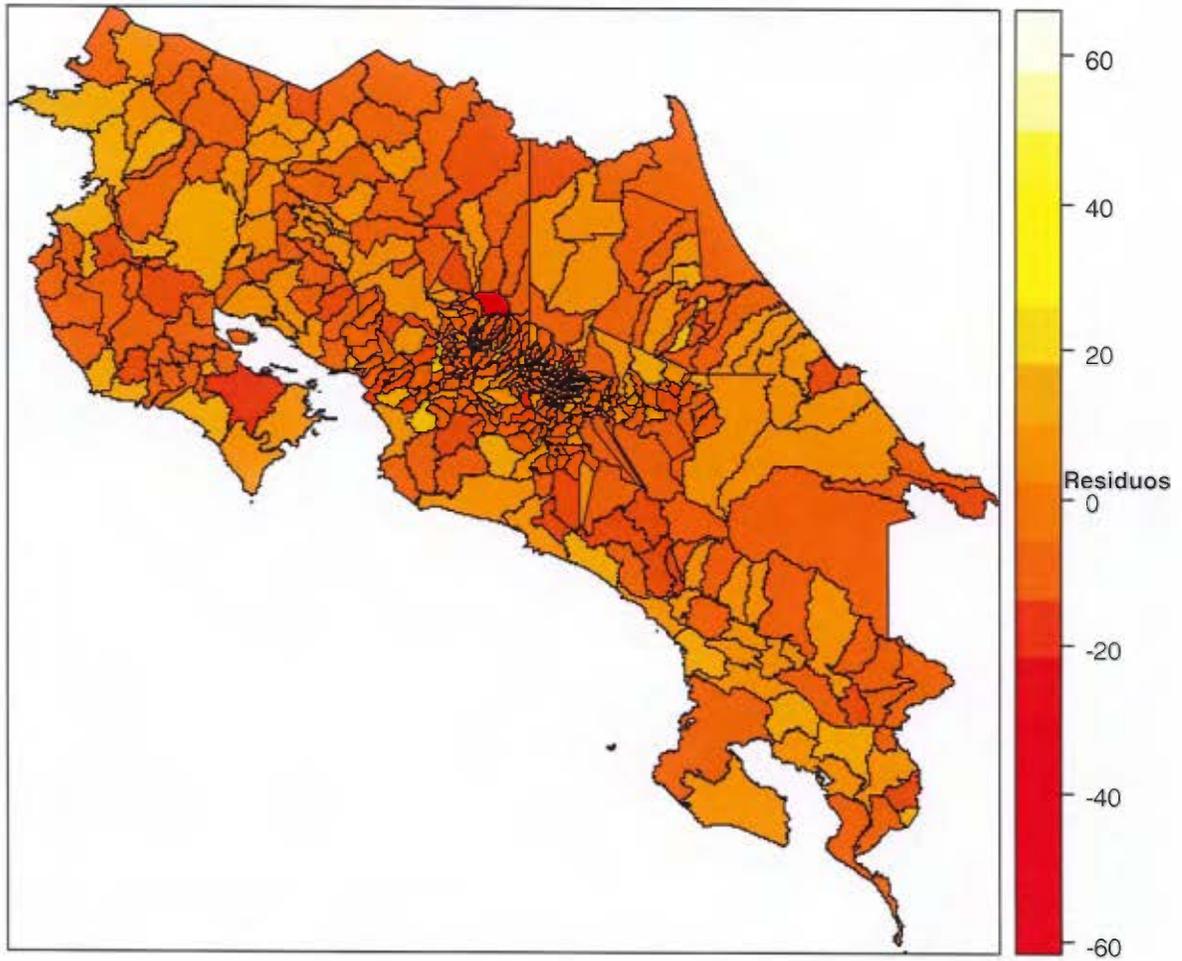
Resumiendo lo analizado, el supuesto de normalidad fue verificado para cada una de las variables del modelo SAR de las diferencias entre el 2007 y 2011, utilizando la prueba Kolmogorov-Smirnoff ya que su hipótesis nula hace referencia a que los datos poseen una distribución normal. Los resultados obtenidos de dicha prueba afirman que existe evidencia estadística para rechazar la hipótesis nula, es decir, los datos de cada una de las variables no presentan una distribución normal. Por el contrario, solamente para las variables tasa de la cantidad de becas y población en edad colegial (logaritmo natural) no se rechazó

el supuesto de normalidad, por lo tanto se concluye que se puede suponer que los datos de esas dos variables provienen de una distribución normal.

Continuando con el análisis previo del modelo SAR de las diferencias entre el 2007 y 2011, la I de Moran indica que existe evidencia estadística para afirmar que se cuenta con autocorrelación espacial entre los distritos de Costa Rica para cada una de las variables del modelo. Además, para comprobar si el parámetro lambda capturó toda la variación espacial, se estudiaron los residuos de los modelos SAR, utilizando la prueba I de Moran con Simulaciones Monte Carlo. El valor de la probabilidad obtenido para el modelo SAR 2011 y para el modelo SAR de las diferencias entre el 2007 y 2011 es de 0,485 y 0,447, respectivamente, resultando ser no significativa en ambos modelos, por lo tanto, se concluye que no hay ninguna estructura espacial en los residuos, en otras palabras, los residuos no dan indicios de autocorrelación global, o sea, la correlación entre los errores de dos distritos vecinos. Esto puede ser corroborado observando el mapa coroplético de los residuos de ambos modelos, en los cuales, no se visualizan aglomeraciones globales porque los colores están distribuidos aleatoriamente.

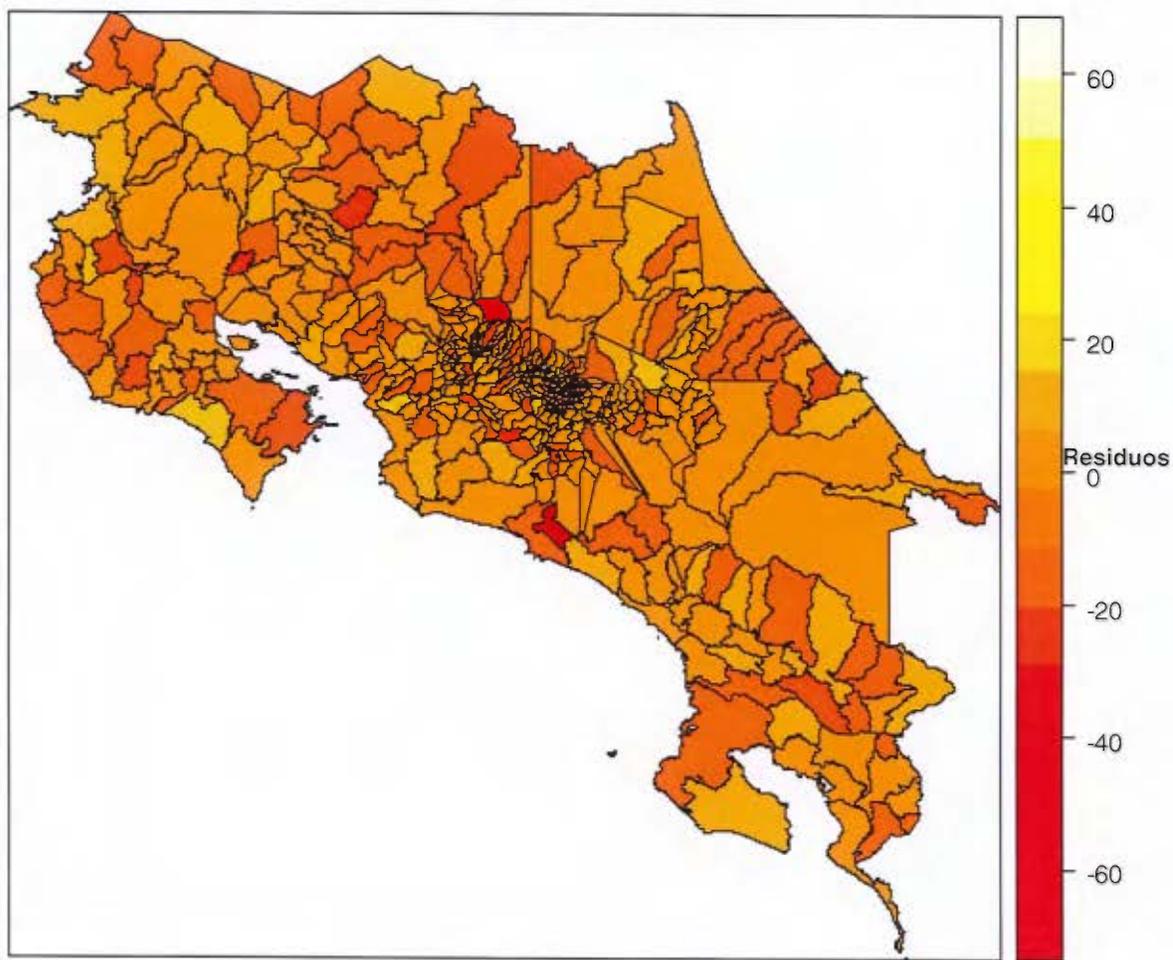
Los resultados encontrados en esta sección correspondiente a las variables y supuestos tomados en cuenta para el modelo SAR de las diferencias entre el 2007 y 2011 no fueron los esperados ya que los residuos del modelo no pueden ser considerados como aleatorios con varianza constante y, de igual forma, también quedó descartado el supuesto de normalidad de los residuos del modelo. Por consiguiente, fue necesario analizar los siguientes modelos SAR para mejorar estos problemas presentados hasta el momento.

Mapa 16. Coroplético de los Residuos del modelo SAR 2011.



Fuente: Elaboración propia, 2011.

Mapa 17. Coroplético de los Residuos del modelo SAR 2007 y 2011.



Fuente: Elaboración propia, 2011.

6. 2. 1. 3 Resultados de los modelos SAR 2011 con variable dependiente definida con datos administrativos

En este apartado, se comparan los resultados obtenidos en los modelos SAR del 2011 y de las diferencias obtenidas entre el 2007 y 2011. También, se muestran los resultados del modelo SAR cuya variable dependiente se ajustó al cambio en el porcentaje de deserción colegial entre el 2011 y 2012 (Cuadro 20).

Cuadro 20. Modelos SAR con variable dependiente obtenida de datos administrativos, 2011

Variables	Modelos SAR						
	Modelo 2011			Modelo 2007-2011			
	Coef.	Error est.	Prob.	Coef.	Error est.	Prob.	
Constante	-19,908	5,406	*	0,230	1,153		
Tasa de becas	0,010	0,003	*	-0,007	0,004		
Monto promedio de becas	0,061	0,146		0,040	0,091		
Consumo promedio de electric.	-0,013	0,008		0,017	0,022		
Población en edad colegial (ln)	3,375	0,465	*	-0,010	1,965		
Cantidad de colegios	0,375	0,215		1,155	0,606		
Tasa total de profesores	0,019	0,006	*				
Tasa total de profesores (0,1 a 50)				-1,590	0,771		*
Tasa total de profesores (50,1 y más)				2,865	1,584		
<i>Lambda</i>	0,128	0,070		-0,034	0,077		
	Modelo 2011 con cambio en porcentaje de deserción			Modelo con variables del Censo 2011			
	Coef.	Error est.	Prob.	Coef.	Error est.	Prob.	
Constante	0,546	1,495		-1,235	8,296		
Tasa de becas	-0,001	0,001		0,003	0,003		
Monto promedio de becas	0,009	0,040		-0,062	0,134		
Consumo promedio de electric.	-0,002	0,002		0,019	0,010		
Población en edad colegial (ln)	0,004	0,129		5,871	3,600		
Cantidad de colegios	-0,009	0,060		-0,225	0,223		
Tasa total de profesores	-0,002	0,002		-0,007	0,007		
% de viviendas con cable				-0,033	0,027		
% de viviendas con sistema agua				-0,173	0,066		*
% de personas con discapacidad				0,065	2,250		
Población total (ln)				-4,547	3,728		
Matrícula inicial (ln)				1,545	0,179		*
<i>Lambda</i>	0,026	0,065		0,036	0,054		

*Significativo al 5%.

Fuente: Elaboración propia, 2011.

En el modelo SAR 2011, los coeficientes de las variables que resultaron ser significativos con un valor de significancia del 5% son la tasa de la cantidad de becas, la población en edad colegial (logaritmo natural) y la tasa total de profesores. Se concluye que conforme aumenta la tasa de la cantidad de becas aumenta el porcentaje de deserción, lo mismo sucede con las variables población en edad colegial y la tasa total de profesores. No obstante, la asociación positiva

entre la tasa de la cantidad de becas y el porcentaje de deserción contradice lo que se espera según los objetivos de *Avancemos*. Además, cabe señalar que los coeficientes de las variables monto promedio de las becas en miles de colones, el consumo promedio de electricidad y la cantidad de colegios no son significativos.

En cambio, los resultados del modelo SAR de las diferencias entre el 2007 y 2011 muestran que solamente la variable tasa total de profesores categorizada según los valores de 0,1 a 50 tiene un coeficiente significativo con un nivel de significancia del 5% y se concluye que conforme aumenta la tasa total de profesores en esa categoría, el porcentaje de deserción colegial disminuye, es decir, existe una relación inversa entre dichas variables.

El valor de lambda del modelo SAR 2011 es de 0,128 y no es significativo, lo que sugiere que no existe correlación espacial en los residuos. Para el modelo SAR de las diferencias entre el 2007 y 2011, el valor de lambda es de -0,034 y tampoco es significativo, por lo tanto, se puede suponer que no existe correlación espacial en los residuos del modelo (Cuadro 21).

Cuadro 21. Modelos SAR 2011 y 2007-2011.

Variables	Modelo 2011			Modelo 2007-2011		
	Coef.	Error est.	Prob.	Coef.	Error est.	Prob.
Constante	-19,908	5,406	*	0,230	1,153	
Tasa de becas	0,010	0,003	*	-0,007	0,004	
Monto promedio de becas	0,061	0,146		0,040	0,091	
Consumo promedio de electric.	-0,013	0,008		0,017	0,022	
Población en edad colegial (ln)	3,375	0,465	*	-0,010	1,965	
Cantidad de colegios	0,375	0,215		1,155	0,606	
Tasa total de profesores	0,019	0,006	*			
Tasa total de profesores (0,1 a 50)				-1,590	0,771	*
Tasa total de profesores (50,1 y más)				2,865	1,584	
<i>Lambda</i>	0,128	0,070		-0,034	0,077	

*Significativo al 5%.

Fuente: Elaboración propia, 2011.

Con respecto al modelo SAR 2011 cuya variable dependiente es el cambio en el porcentaje de deserción entre el 2012 y 2011, ninguno de los coeficientes de las variables que componen dicho modelo resultaron ser significativos con un nivel de significancia del 5%. Además, el valor de lambda del modelo es de 0,026 y tampoco es significativo, por lo tanto no existe correlación espacial en los residuos (Cuadro 22).

Cuadro 22. Modelo SAR 2011 con el cambio en el porcentaje de deserción entre el 2012 y 2011.

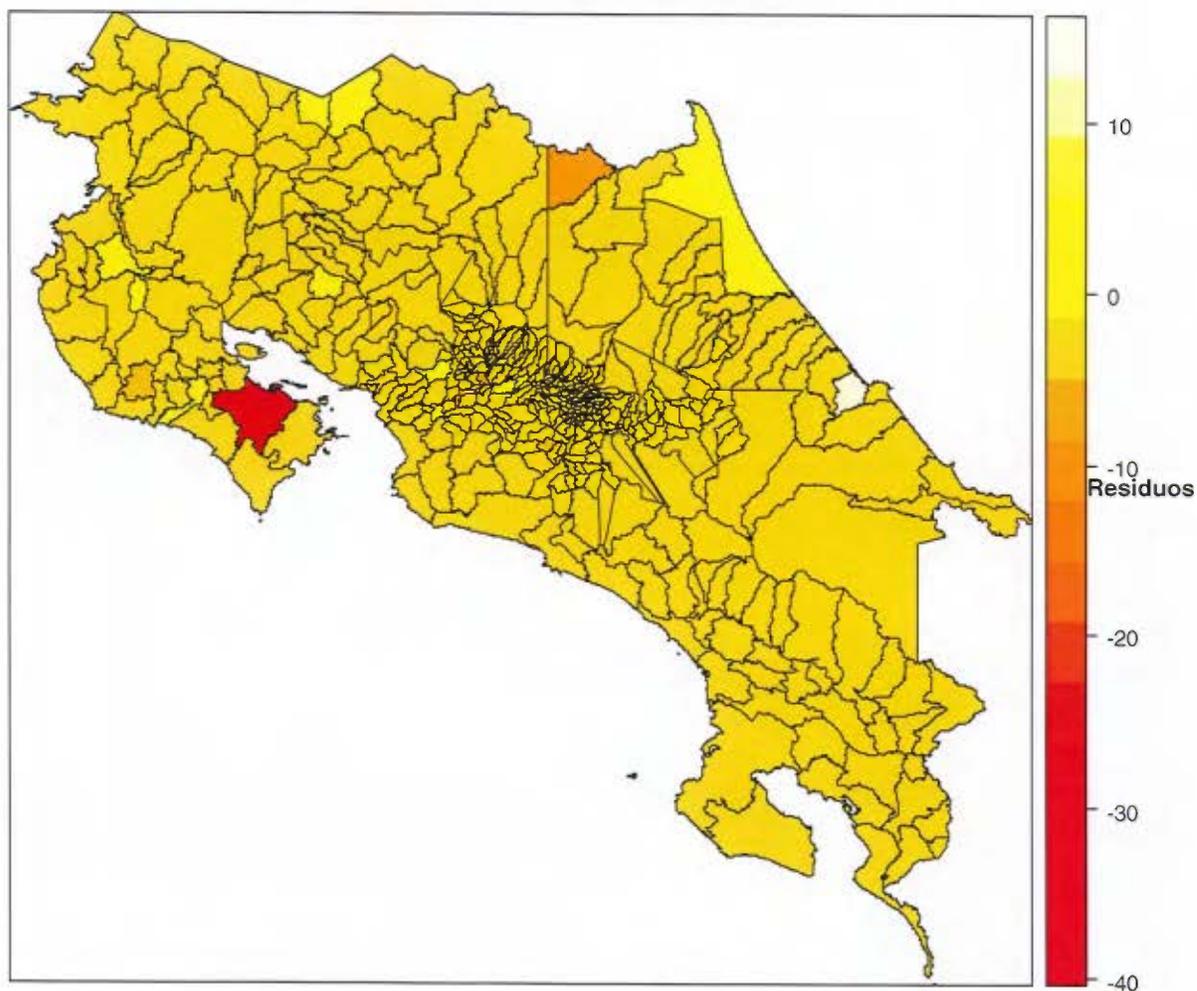
Variables	Coef.	Error est.	Prob.
Constante	0,546	1,495	
Tasa de becas	-0,001	0,001	
Monto promedio de becas	0,009	0,040	
Consumo promedio de electricidad	-0,002	0,002	
Población en edad colegial (ln)	0,004	0,129	
Cantidad de colegios	-0,009	0,060	
Tasa total de profesores	-0,002	0,002	
<i>Lambda</i>	<i>0,026</i>	<i>0,065</i>	

*Significativo al 5%.

Fuente: Elaboración propia, 2011.

Asimismo, se calculó la prueba I de Moran con Simulaciones Monte Carlo para verificar si el parámetro lambda capturó toda la variación espacial. Se realizaron 1 000 simulaciones obteniendo una probabilidad de 0,495 la cual es no significativa, entonces se concluye que los residuos del modelo con el cambio en la deserción entre el 2012 y 2011 no presentan ninguna estructura espacial o autocorrelación global. Sin embargo, el mapa coroplético de los residuos muestra casi una misma tonalidad de color lo cual indica que existen algunos valores extremos.

Mapa 18. Coroplético de los Residuos del modelo SAR 2011 con el cambio en el porcentaje de deserción entre el 2012 y 2011.



Fuente: Elaboración propia, 2011.

Es importante indicar que se hicieron diversos ejercicios para contrarrestar la cantidad de ceros existentes en la variable porcentaje de deserción debido a la inexistencia de colegios en más de 100 distritos y su influencia en los modelos SAR. Uno de esos ejercicios fue la creación de una variable derivada de un modelo probit, en donde la variable dependiente porcentaje de deserción es dicotómica (0 indica que no existe porcentaje de deserción y 1 para todos los demás casos) y todas las variables independientes del modelo SAR permanecían

iguales, esa variable derivada corresponde a la razón inversa de Mills y luego ésta pasaba a ser variable independiente del modelo SAR, pero los resultados obtenidos no fueron los esperados, así como las estimaciones y valores de probabilidad obtenidos, por lo que fue necesario analizar los siguientes modelos SAR.

6. 2. 1. 4 El modelo SAR 2011 controlando por variables censales

Posteriormente, se realizó un modelo SAR incorporando algunas variables independientes censales del 2011, obtenidas de la investigación realizada por Brenes (2014) para utilizarlas como variables control en el modelo SAR 2011 y así, determinar la asociación con las variables propiamente de *Avancemos*, confeccionando un modelo más robusto para no tener sesgo de variable omitida. Es importante señalar que para este modelo se mantuvo la misma variable dependiente con datos administrativos. Los resultados siguen siendo similares a los obtenidos en los modelos anteriores para las variables de *Avancemos* cuyos coeficientes no son significativos con un nivel de significancia del 5%, indicando que no hay asociación estadística entre la tasa de becas y el monto promedio de becas en miles de colones con el porcentaje de deserción para este modelo SAR. No obstante, los coeficientes de las variables porcentaje de viviendas con sistema de agua caliente en toda la casa y el logaritmo natural de la matrícula inicial resultaron ser significativos al 5%, indicando para la primera, cuanto mayor es el porcentaje de viviendas con sistema de agua caliente en toda la casa menor es el porcentaje de deserción, siendo *proxy* del nivel socioeconómico y para la segunda, cuanto mayor es la cantidad de población en la matrícula inicial en los colegios, mayor es el porcentaje de deserción, representando un problema de sobrepoblación estudiantil colegial. Además, el valor de lambda del modelo es de 0,036 y tampoco es significativo, por lo tanto se puede suponer que no existe correlación espacial en los residuos (Cuadro 23).

Comparando los resultados obtenidos entre el modelo SAR 2011 sin variables del Censo 2011 y el modelo SAR con variables del Censo 2011, se concluye que para el primero, el coeficiente de la tasa de becas es significativo al 5% así como los coeficientes de la población en edad colegial con logaritmo natural y la tasa total de profesores. No obstante, para el segundo, los coeficientes de las tres variables anteriormente citadas son no significativos al 5% y en este caso, los coeficientes significativos son los correspondientes al porcentaje de viviendas con sistema de agua caliente y el logaritmo natural de la matrícula inicial. Y como anteriormente fue señalado, el valor de lambda es no significativo en ambos modelos, lo cual es indicio de que no existe correlación espacial en los residuos.

Cuadro 23. Modelo SAR 2011 y con variables independientes del Censo 2011.

Variabes	Coef.	Error est.	Prob.
<i>Sin variables independientes del Censo 2011:</i>			
Constante	-19,908	5,406	*
Tasa de becas	0,010	0,003	*
Monto promedio de becas	0,061	0,146	
Consumo promedio de electricidad	-0,013	0,008	
Población en edad colegial (ln)	3,375	0,465	*
Cantidad de colegios	0,375	0,215	
Tasa total de profesores	0,019	0,006	*
<i>Lambda</i>	<i>0,128</i>	<i>0,070</i>	
<i>Con variables independientes del Censo 2011:</i>			
Constante	-1,235	8,296	
Tasa de becas	0,003	0,003	
Monto promedio de becas	-0,062	0,134	
Consumo promedio de electricidad	0,019	0,010	
Población en edad colegial (ln)	5,871	3,600	
Cantidad de colegios	-0,225	0,223	
Tasa total de profesores	-0,007	0,007	
Porcentaje de viviendas con cable	-0,033	0,027	
Porcentaje de viviendas con sistema agua	-0,173	0,066	*
Porcentaje de personas con discapacidad	0,065	2,250	
Población total (ln)	-4,547	3,728	
Matrícula inicial (ln)	1,545	0,179	*
<i>Lambda</i>	<i>0,036</i>	<i>0,054</i>	

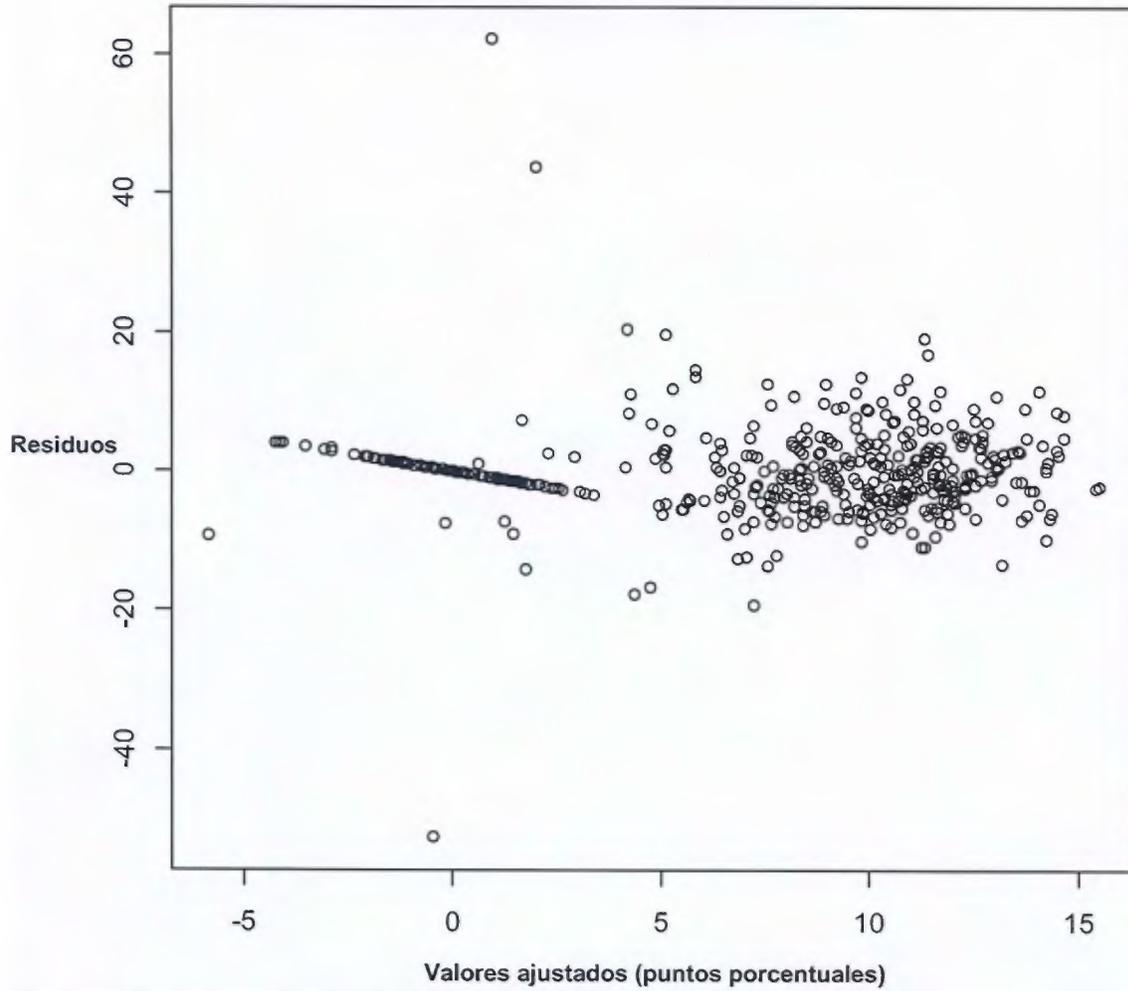
*Significativo al 5%.

Fuente: Elaboración propia, 2011.

Asimismo, se analizaron los supuestos de los residuos de este modelo SAR 2011 con variables del Censo 2011 y sus valores ajustados. Como se observa, los puntos del gráfico no siguen presentando ningún patrón para afirmar que los residuos son aleatorios con varianza constante porque se agrupan mayormente en el centro del gráfico (Gráfico 23). Este comportamiento es similar al presentado en los análisis homólogos de los anteriores modelos SAR, inconvenientes que también siguen siendo similares con cada una de las variables predictoras y los residuos de este modelo SAR 2011 con variables del Censo 2011.

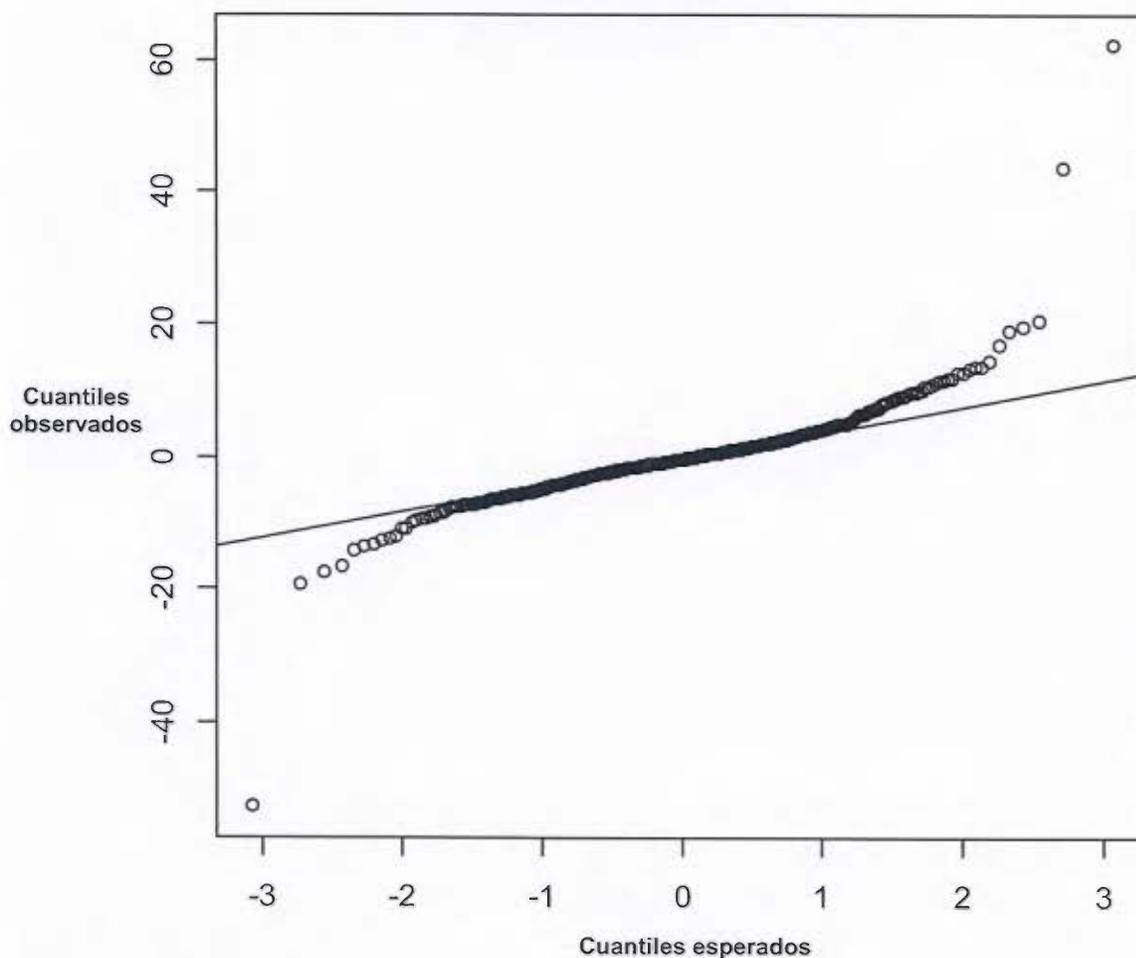
El supuesto de normalidad también fue analizado tomando en cuenta el gráfico de cuantiles normales, siendo necesario que los puntos estén alineados en diagonal. Por lo tanto, se obtiene la misma conclusión de los modelos SAR anteriores descartándose el supuesto de normalidad ya que los puntos no están alineados correctamente en línea recta. Este supuesto de normalidad de los residuos fue verificado usando la prueba Kolmogorov-Smirnoff, la cual tiene como hipótesis nula que los datos tienen una distribución normal. Con los resultados obtenidos se rechaza la hipótesis nula con un nivel de significancia del 5%, ya que el valor de probabilidad asociada fue de 0,000 y, por lo tanto, se concluye que los datos no tienen una distribución normal (Gráfico 24).

Gráfico 23. Valores ajustados vs. Residuos del modelo SAR 2011 con variables del Censo 2011.



Fuente: Elaboración propia, 2011.

Gráfico 24. Gráfico de Normalidad de los residuos del modelo SAR 2011 con variables del Censo 2011.



Fuente: Elaboración propia, 2011.

Nótese que en esta sección no se encontraron los resultados esperados ya que los modelos SAR planteados no lograron mostrar asociación entre el porcentaje de deserción colegial y las variables referidas al programa social *Avancemos*, y tampoco pudo validarse la correlación espacial en los residuos de los modelos SAR, con un nivel de significancia del 5%. Seguidamente, con la aproximación del porcentaje de deserción colegial mediante el Censo 2011, estos inconvenientes fueron solucionados satisfactoriamente.

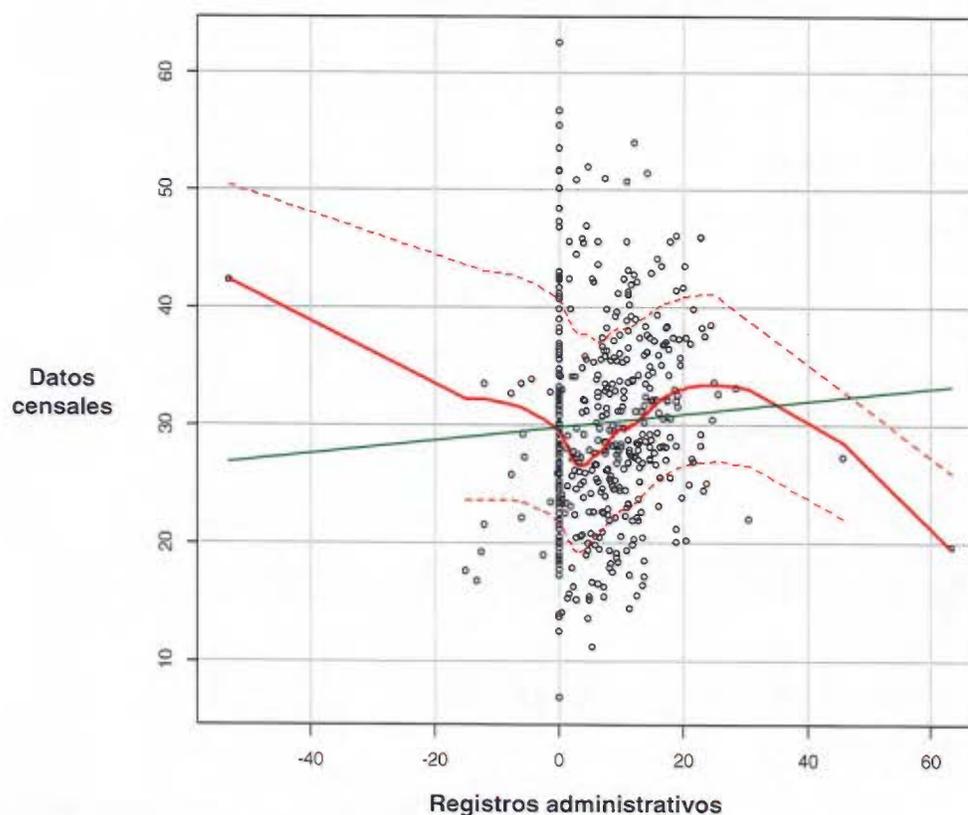
6. 2. 2 Modelo SAR 2011 con variable dependiente definida con datos del Censo 2011

Para inspeccionar si los resultados de la sección anterior se mantienen con una operacionalización distinta de la variable dependiente, se decidió aproximar el valor del porcentaje de deserción colegial tomando en cuenta los datos del Censo 2011 para evaluar los resultados obtenidos utilizando la variable dependiente porcentaje de deserción colegial extraída de los registros administrativos del MEP. La construcción del porcentaje de deserción según el censo fue obtenida del cociente resultante entre la población en edad colegial que no asiste a ningún centro educativo que tiene secundaria incompleta y la población en edad colegial que tiene secundaria incompleta.

Esto fue realizado debido a que se cuestionó que los datos administrativos del MEP están centrados en el distrito donde está ubicado el colegio y no en el distrito de residencia del estudiante y esto oculta el proceso de autocorrelación espacial. Por lo tanto los modelos SAR siguientes toman en cuenta como variable dependiente el porcentaje de deserción según el censo, manteniendo las mismas variables originales de los modelos anteriores. Evidentemente, el porcentaje de deserción con datos administrativos estaba afectando los resultados de los modelos debido a que en 115 distritos no existe ningún colegio por lo que el porcentaje de deserción con datos administrativos para esos distritos corresponde a cero para los modelos anteriores.

Analizando el coeficiente de correlación entre ambos porcentajes de deserción se obtuvo un valor de 0,055 lo cual indica que existen problemas con el porcentaje de deserción proveniente de registros administrativos del MEP ya que la correlación no es alta (Gráfico 25).

Gráfico 25. Diagrama de dispersión del porcentaje de deserción de registros administrativos y datos censales.



Fuente: Elaboración propia, 2011.

Con estas modificaciones en la variable dependiente tomando en cuenta datos censales se lograron corregir la dirección y los valores de la probabilidad asociada para cada uno de los coeficientes del modelo SAR 2011 de las variables relacionadas con el programa social *Avancemos*, y el valor de lambda del modelo SAR 2011 pasa a ser significativo con lo cual se concluye que existe autocorrelación espacial en los residuos de los modelos con esta nueva variable dependiente construida a partir de datos censales. Esto quiere decir que en el programa social *Avancemos* tanto la tasa de la cantidad de becas como el monto promedio de las becas en miles de colones están asociados con la deserción colegial de forma significativa al 5%.

Cuadro 24. Comparación de los modelos SAR 2011.

Variables	Modelos SAR			
	Modelo 2011 con becas y monto promedio		Modelo 2011 sin monto promedio	
<u>Con porcentaje de deserción según censo</u>	Coef.	Prob.	Coef.	Prob.
Constante	65,201	*	49,044	*
Tasa de becas	-0,005		-0,006	*
Monto promedio de becas	-0,653	*		
Consumo promedio de electric.	-0,049	*	-0,054	*
Población en edad colegial (ln)	-0,797		-0,845	
Cantidad de colegios	-0,387	*	-0,454	*
Tasa total de profesores	-0,012	*	-0,013	*
<i>Lambda</i>	0,609	*	0,617	*
<u>Con porcentaje de deserción según registros administrativos</u>				
	Modelo 2011		Modelo 2007-2011	
	Coef.	Prob.	Coef.	Prob.
Constante	-19,908	*	0,230	
Tasa de becas	0,010	*	-0,007	
Monto promedio de becas	0,061		0,040	
Consumo promedio de electric.	-0,013		0,017	
Población en edad colegial (ln)	3,375	*	-0,010	
Cantidad de colegios	0,375		1,155	
Tasa total de profesores	0,019	*		
Tasa total de profesores (0,1 a 50)			-1,590	*
Tasa total de profesores (50,1 y más)			2,865	
<i>Lambda</i>	0,128		-0,034	
	Modelo 2011 con cambio en % de deserción		Modelo con variables del Censo 2011	
	Coef.	Prob.	Coef.	Prob.
Constante	0,546		-1,235	
Tasa de becas	-0,001		0,003	
Monto promedio de becas	0,009		-0,062	
Consumo promedio de electric.	-0,002		0,019	
Población en edad colegial (ln)	0,004		5,871	
Cantidad de colegios	-0,009		-0,225	
Tasa total de profesores	-0,002		-0,007	
% de viviendas con cable			-0,033	
% de viviendas con sistema agua			-0,173	*
% de personas con discapacidad			0,065	
Población total (ln)			-4,547	
Matrícula inicial (ln)			1,545	*
<i>Lambda</i>	0,026		0,036	

*Significativo al 5%.

Fuente: Elaboración propia, 2011.

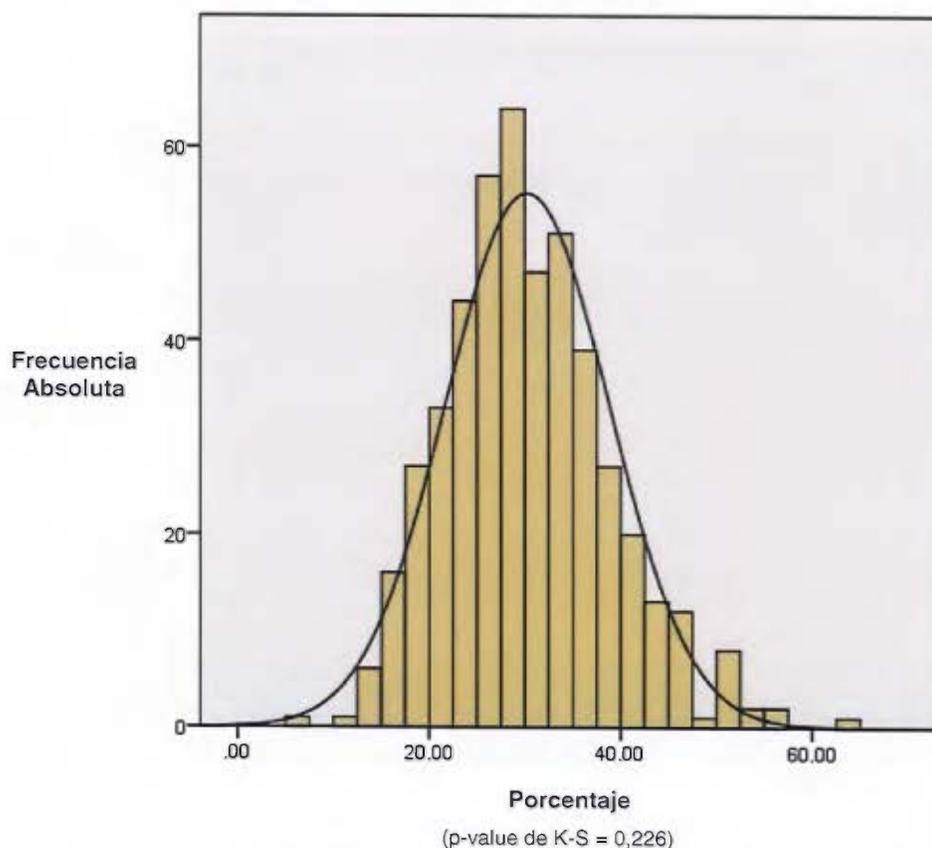
Además, cabe señalar que para los dos primeros modelos SAR con el porcentaje de deserción calculado con datos censales, todos los diagnósticos relacionados con la normalidad y comportamiento de los residuos, así como el tema de la autocorrelación espacial, fueron los idóneos para validar los resultados del modelo SAR 2011 bajo estos requerimientos. Con respecto a los coeficientes de las variables que son significativas en el modelo SAR 2011 con la variable dependiente definida con datos del Censo 2011 se concluye lo siguiente: por cada incremento de 10 puntos porcentuales en la tasa de la cantidad de becas el porcentaje de deserción disminuye en 0,06 puntos porcentuales; por cada 10 000 colones adicionales en el monto promedio de becas el porcentaje de deserción disminuye en 6,5 puntos porcentuales; por cada incremento de 10 kilowatts/hora en el consumo promedio de electricidad el porcentaje de deserción disminuye en 0,54 puntos porcentuales; por cada incremento de un colegio el porcentaje de deserción disminuye en 0,45 puntos porcentuales; y, por último, por cada incremento de 10 puntos porcentuales en la tasa total de profesores el porcentaje de deserción disminuye en 0,13 puntos porcentuales, manteniendo las demás variables constantes aplicable para cada una de las anteriores interpretaciones (Cuadro 24).

Nótese que con esta nueva operacionalización de la variable dependiente porcentaje de deserción colegial, obtenida a partir del Censo 2011 se encuentra la asociación esperada con las variables referidas al programa social *Avancemos* de manera que sus coeficientes son negativos y significativamente distintos de cero. También, la correlación espacial en los residuos del modelo SAR fue validada con un nivel de significancia del 5%.

6. 2. 2. 1 Comportamiento de variables y supuestos del modelo SAR 2011 con variable dependiente del Censo 2011

El porcentaje de deserción según el censo presenta valores centralizados y pocos valores extremos. Además, se aprecia que el comportamiento de la variable según el histograma puede ser considerado como normal. Esto se valida con la prueba Kolmogorov-Smirnoff, cuya hipótesis nula es que los datos tienen una distribución normal; el valor de la probabilidad asociada obtenido fue de 0,226, por lo tanto, se puede suponer que el porcentaje de deserción según el censo tiene distribución normal (Gráfico 26).

Gráfico 26. Histograma de la variable porcentaje de deserción según el censo, 2011

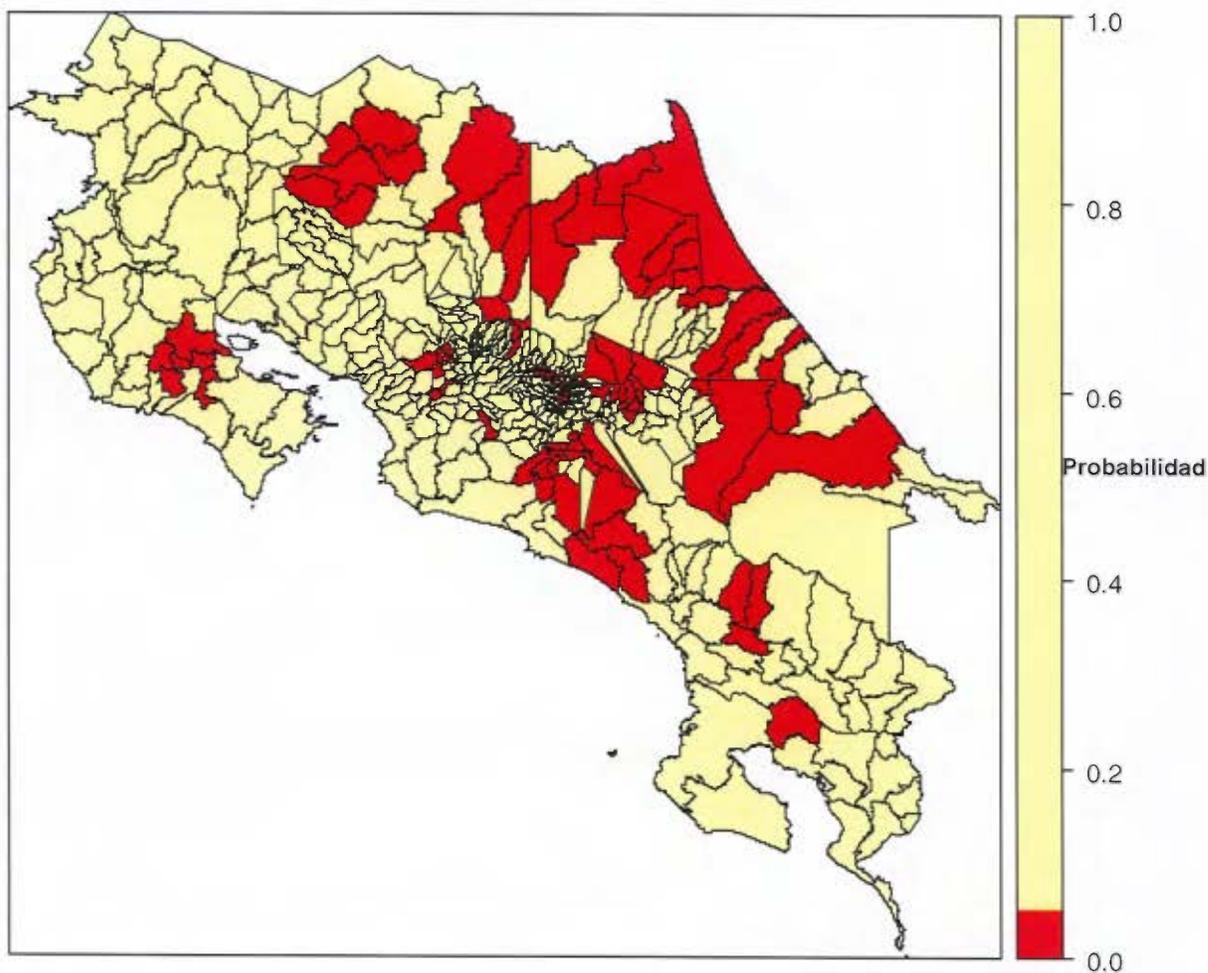


Fuente: Elaboración propia con base en estadísticas del Centro Centroamericano de Población (CCP), 2011.

También, se calcularon los índices de autocorrelación I de Moran para comprobar si existe o no autocorrelación en el patrón espacial del porcentaje de deserción construido a partir del Censo 2011 bajo el supuesto de vecinos con frontera en común. La I de Moran bajo el supuesto de normalidad y de asignación aleatoria obtuvo un estadístico de 13,2 en ambos casos y a su vez resultando significativos, lo cual indica que existe autocorrelación en el patrón espacial del porcentaje de deserción según el censo, es decir, se confirma el rechazo de la hipótesis de no autocorrelación, por lo que se concluye que hay evidencia estadística para afirmar que existe autocorrelación espacial entre los distritos de Costa Rica en cuanto al porcentaje de deserción construido con datos censales.

Como anteriormente se concluyó que hay evidencia estadística para afirmar que existe autocorrelación espacial, con la autocorrelación local se puede identificar algunas zonas geográficas o distritos que presenten este tipo de problema. En el mapa de Costa Rica del Índice de Moran a nivel local, las zonas de color rojo indican las aglomeraciones de una forma más clara y éstas se ubican en distintos lugares del país, lo cual señala una autocorrelación espacial en esos grupos de regiones visualizadas en su mayoría en la parte Norte y Atlántica, y en menor medida en la parte Sur y Oeste del territorio nacional.

Mapa 19. Índice de Moran a nivel local del porcentaje de deserción según el censo, 2011



Fuente: Elaboración propia, 2011.

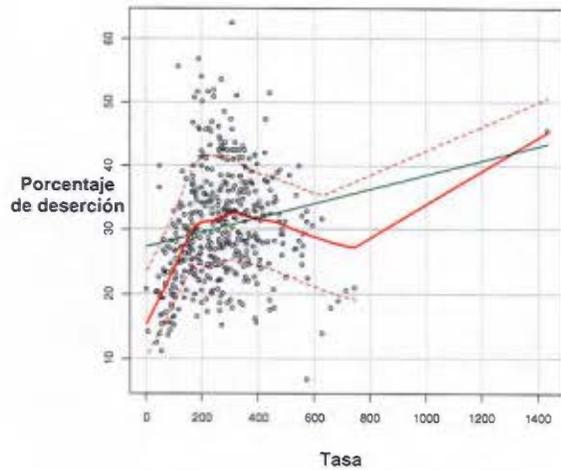
Analizando de forma visual la dispersión existente entre cada una de las variables independientes y el porcentaje de deserción según el censo se identifica la relación que tienen las variables entre sí (Gráfico 27). La tasa de becas y el porcentaje de deserción según el censo mantienen un comportamiento estable agrupándose al inicio del gráfico. Algo similar sucede con el monto promedio de becas y el porcentaje de deserción según el censo con la diferencia de que los

valores se agrupan al final del gráfico debido al pequeño intervalo en que varía el monto promedio de becas.

El consumo promedio de electricidad y el porcentaje de deserción según el censo mantienen un comportamiento estable y sus valores se extienden a lo largo y ancho del gráfico. La cantidad de colegios y el porcentaje de deserción según el censo en su mayoría agrupan valores al inicio del gráfico debido a la naturaleza de la variable.

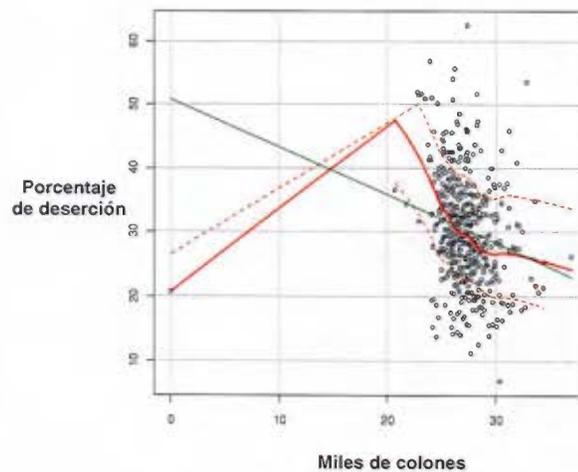
La tasa total de profesores y el porcentaje de deserción según el censo presentan valores al inicio del gráfico en un rango de variación pequeño, similar a lo que ocurre con el monto promedio de becas y el porcentaje de deserción según el censo. Por último, la población en edad colegial (logaritmo natural) y el porcentaje de deserción según el censo distribuyen sus valores de forma muy estable a lo largo y ancho del gráfico.

Gráfico 27. Diagramas de dispersión del porcentaje de deserción según el censo en el 2011 por cada variable predictora, por distritos de Costa Rica.



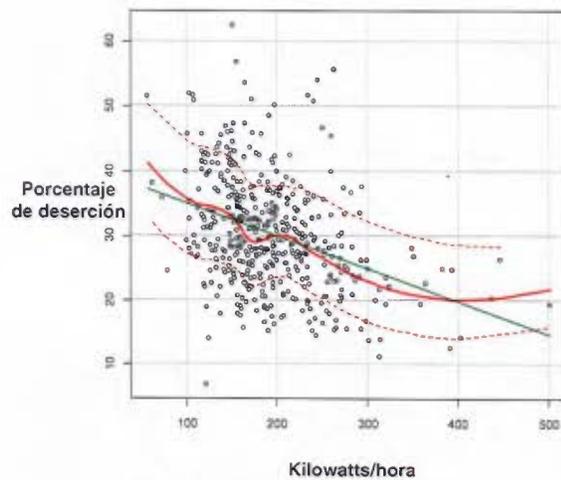
Tasa cantidad de becas

(r de Pearson = 0,1865)



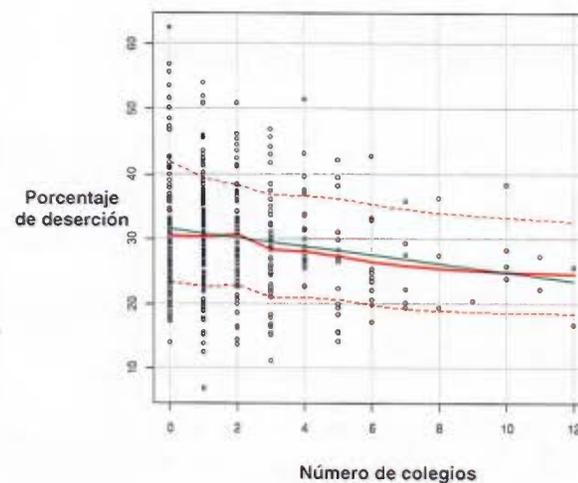
Monto promedio de becas

(r de Pearson = -0,2135)



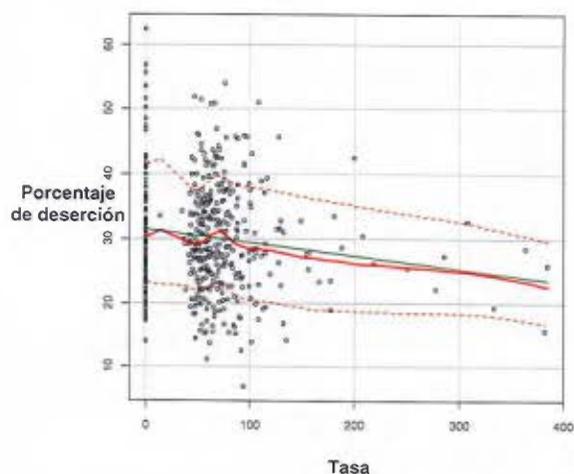
Consumo promedio de electricidad

(r de Pearson = -0,3544)



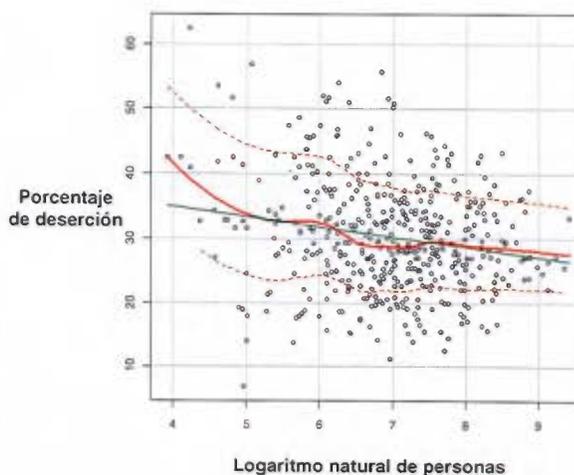
Cantidad de colegios

(r de Pearson = -0,1658)



Tasa total profesores

(r de Pearson = -0,1280)



Población en edad colegial (ln)

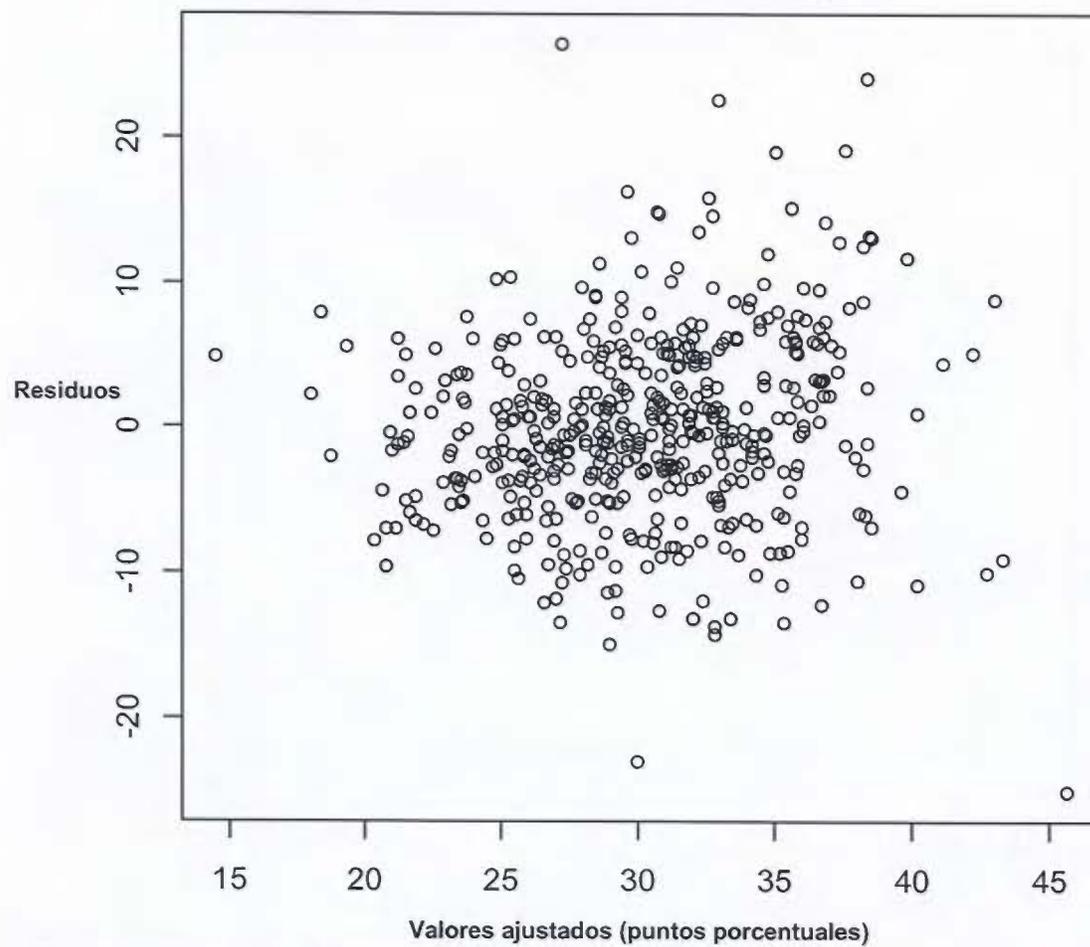
(r de Pearson = -0,1865)

Fuente: Elaboración propia, 2011.

Nota: Se presentan los valores del coeficiente de correlación r de Pearson debajo de cada variable.

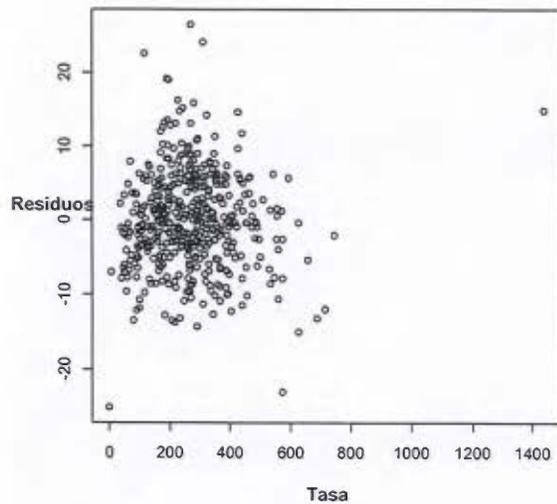
Además, es necesario analizar los supuestos de los residuos del modelo SAR 2011 con porcentaje de deserción según el censo y sus correspondientes valores ajustados de forma gráfica. En el caso óptimo, es necesario que los puntos del gráfico no presenten ningún patrón para afirmar que los residuos son aleatorios con varianza constante. Esto parece ser lo que ocurre ya que los valores se distribuyen de forma aleatoria a lo largo y ancho del gráfico (Gráficos 28 y 29). Para validar el supuesto de normalidad, el gráfico de cuantiles normales debería ser una diagonal bien marcada por los puntos. En ese caso, prácticamente los puntos están alineados en forma diagonal, por lo tanto de forma visual se cumple el supuesto de normalidad (Gráfico 30).

Gráfico 28. Valores ajustados vs. Residuos del modelo SAR 2011 con porcentaje de deserción según el censo.



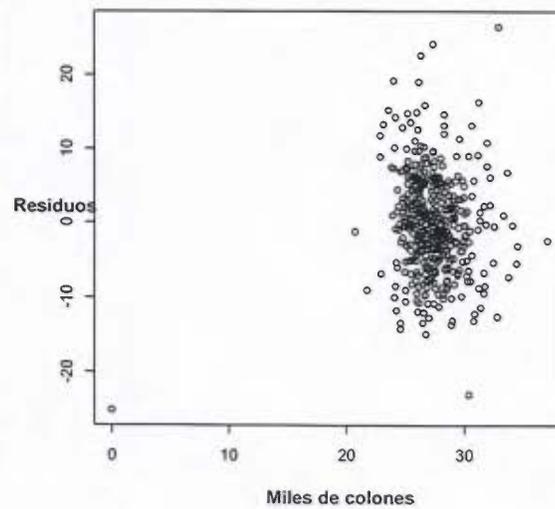
Fuente: Elaboración propia, 2011.

Gráfico 29. Variables predictoras vs. Residuos del modelo SAR 2011.



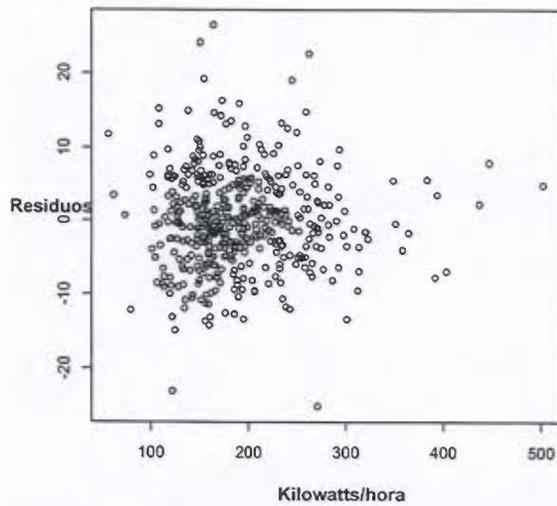
Tasa cantidad de becas

(r de Pearson = 0,0157)



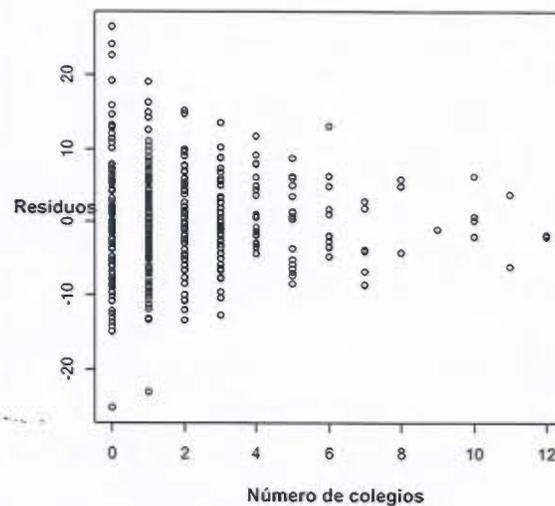
Monto promedio de becas

(r de Pearson = -0,0133)



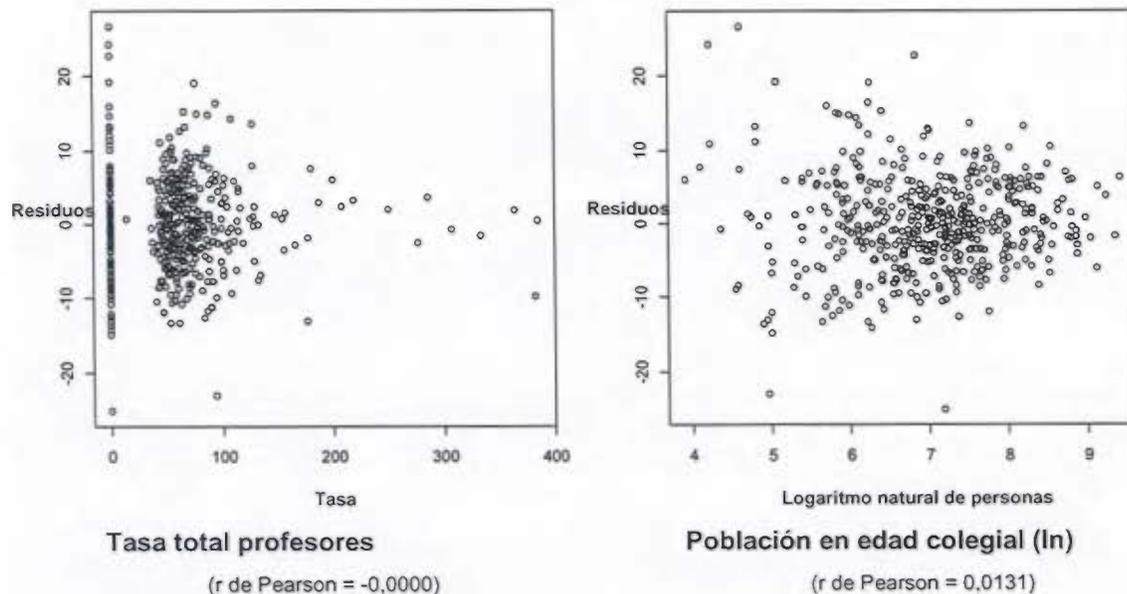
Consumo promedio de electricidad

(r de Pearson = 0,0014)



Cantidad de colegios

(r de Pearson = 0,0141)

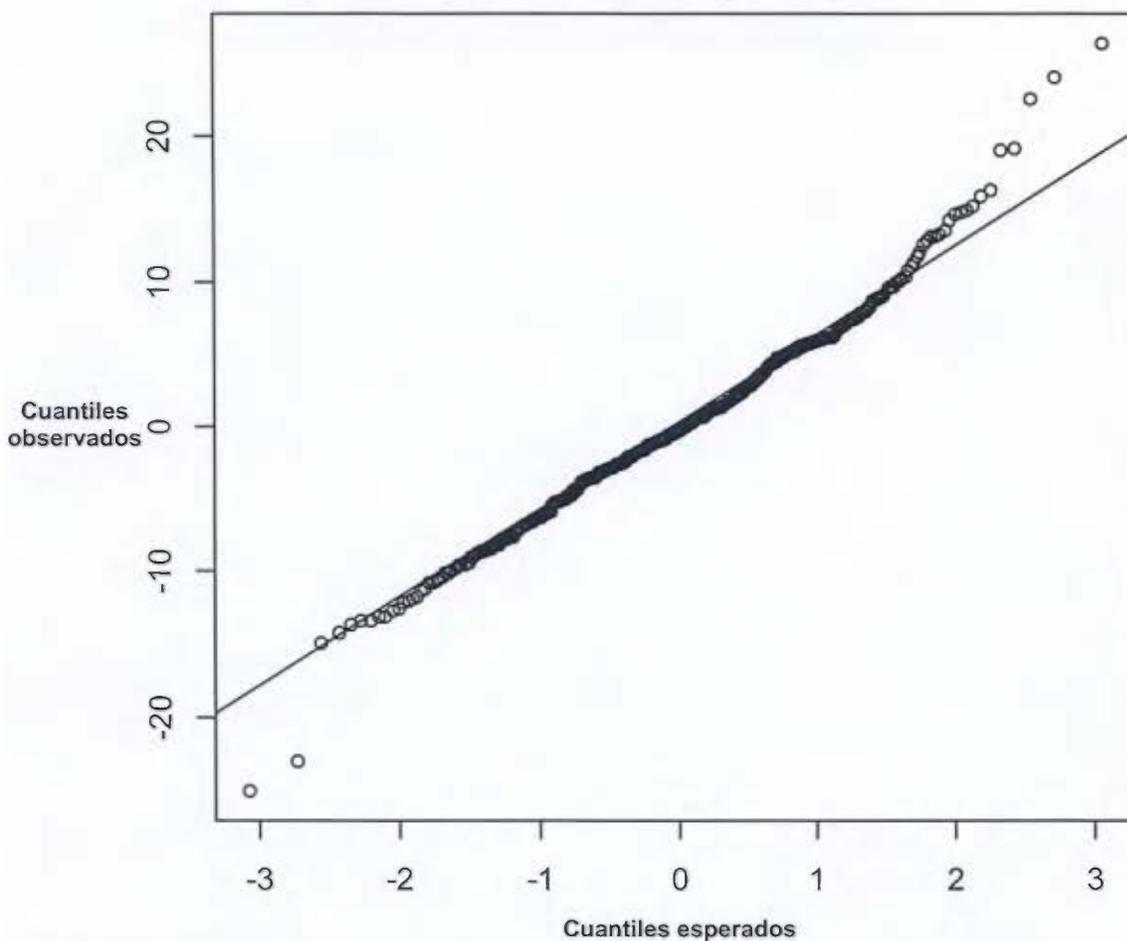


Fuente: Elaboración propia, 2011.

Nota: Se presentan los valores del coeficiente de correlación r de Pearson debajo de cada variable.

Además, el supuesto de normalidad de los residuos fue corroborado mediante la prueba Kolmogorov-Smirnoff cuya hipótesis nula es que los datos tienen una distribución normal. Según los resultados obtenidos, el valor de probabilidad asociada es de 0,5071, por lo tanto, con un nivel de significancia del 5% existe evidencia estadística para no rechazar la hipótesis nula, concluyendo que los residuos tienen una distribución normal (Gráfico 30).

Gráfico 30. Gráfico de Normalidad de los residuos del modelo SAR 2011 con porcentaje de deserción según el censo.



Fuente: Elaboración propia, 2011.

De igual forma, se analizó el supuesto de normalidad para cada una de las variables del modelo SAR 2011 con porcentaje de deserción según el censo, mediante la prueba Kolmogorov-Smirnoff cuya hipótesis nula indica que los datos tienen una distribución normal. Los resultados para todas las variables en cuestión arrojaron que existe evidencia estadística para rechazar la hipótesis nula, por lo tanto, los datos de cada una de las variables no poseen una distribución normal tomando en cuenta dicha prueba estadística, excepto para las variables porcentaje de deserción según el censo y población en edad colegial (logaritmo

natural) cuyos valores de probabilidad asociada fueron de 0,226 y 0,200, respectivamente, por lo que con un nivel de significancia del 5% existe evidencia estadística para no rechazar la hipótesis nula, de manera que se puede afirmar que los datos de ambas variables tienen una distribución normal.

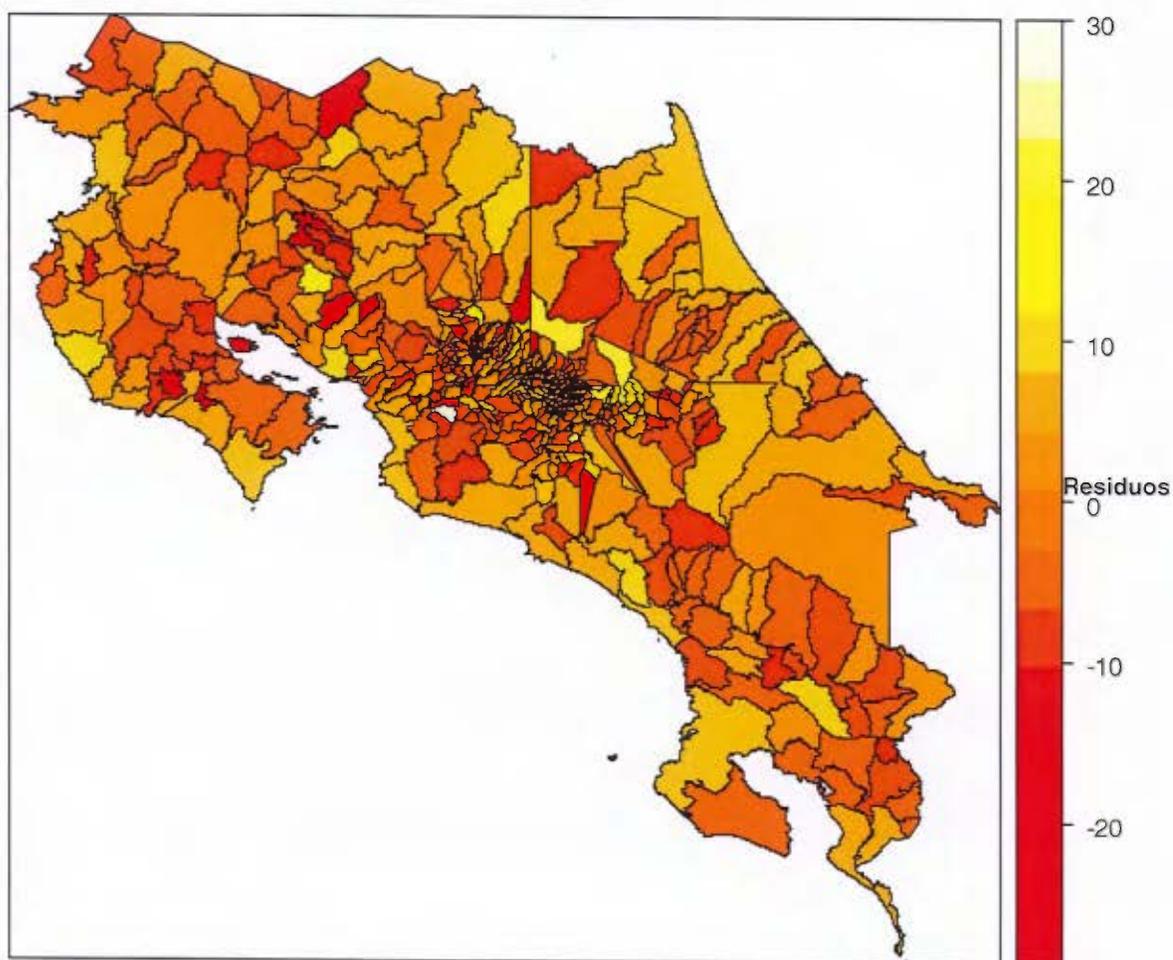
Seguidamente, para este análisis previo del modelo SAR 2011 con porcentaje de deserción según el censo, se calculó la I de Moran para cada una de las variables correspondientes. Los resultados señalan que existe evidencia estadística para afirmar que hay autocorrelación espacial entre los distritos de Costa Rica para cada una de las variables analizadas.

Y, finalmente, para comprobar si el parámetro lambda capturó toda la variación espacial, se estudiaron los residuos de los modelos SAR 2011 con porcentaje de deserción según el censo, utilizando la prueba I de Moran con Simulaciones Monte Carlo. Se realizaron 1 000 simulaciones y el valor de la probabilidad obtenido para el modelo SAR 2011 con porcentaje de deserción según el censo es de 0,781 resultando ser no significativo con un nivel de significancia del 5%, por lo tanto, se concluye que no hay ninguna estructura espacial en los residuos, en otras palabras, los residuos no dan indicios de autocorrelación global.

Esto puede ser corroborado observando el mapa coroplético de los residuos del modelo SAR 2011 con porcentaje de deserción según el censo, en el cual, no se visualizan aglomeraciones globales porque los colores están distribuidos aleatoriamente. Esto significa que para este modelo, el inconveniente presente en los anteriores modelos propuestos fue solucionado en términos de la inexistencia de aglomeraciones globales ya que en las distintas zonas del mapa los colores son diferentes.

Es así como en esta sección se validaron los supuestos del modelo SAR tomando en cuenta la aproximación del porcentaje de deserción a partir de los datos del Censo 2011. Es importante indicar que los resultados obtenidos fueron los esperados y por lo tanto, en la siguiente sección se va a establecer el modelo SAR final escogido con todos los detalles correspondientes.

Mapa 20. Coroplético de los Residuos del modelo SAR 2011 con porcentaje de deserción según el censo.



Fuente: Elaboración propia, 2011.

6. 2. 2 Resultados de los modelos SAR 2011 con variable dependiente definida con datos del Censo 2011

A continuación, se detallan los resultados obtenidos del modelo SAR 2011 con el porcentaje de deserción según el censo a nivel distrital. Todos los coeficientes de las variables incluidas en este modelo resultaron ser significativos con un nivel de significancia del 5%, excepto la tasa de becas y la población en edad colegial (logaritmo natural) que son significativos pero con un nivel de significancia del 10% (Modelo A). No obstante, todas las variables independientes predicen la deserción y en la dirección que se esperaba, es decir, cuanto mayor sean cada una de las variables independientes, el porcentaje de deserción disminuye. Además, se hizo el ejercicio de calcular el modelo sin tomar en cuenta la variable monto promedio de becas en miles de colones (Modelo B) y de esta forma, el coeficiente de la variable tasa de becas resultó ser significativo al 5%, esto puede explicarse debido a la alta correlación existente entre la variable tasa de becas y el monto promedio de becas en miles de colones, es por ello que, en ese caso, el coeficiente de la tasa de becas no resultó significativo con un nivel de significancia del 5% en el Modelo A (Cuadro 25).

También, sería importante investigar en un futuro la razón por la que el coeficiente de la variable población en edad colegial (logaritmo natural) resultó ser no significativo con un nivel de significancia del 5%, experimentando con distintas variables y otros datos. Además, estos resultados difieren de los expuestos por Brenes (2014) debido a que los utilizados en esta investigación corresponden a registros administrativos y no completamente obtenidos del Censo 2011.

Como ejemplo de interpretación se tiene que el monto promedio de becas varía aproximadamente entre 21 000 y 37 000 colones, entonces por cada 10 000

colones adicionales de beca el porcentaje de deserción disminuye en 6,5 puntos porcentuales en el Modelo A, manteniendo las demás variables constantes. Por su parte, cuanto mayor es la tasa de la cantidad de becas menor es el porcentaje de deserción en el Modelo B, es decir, por cada incremento de 10 puntos porcentuales en la tasa de la cantidad de becas el porcentaje de deserción disminuye en 0,06 puntos porcentuales, manteniendo las demás variables constantes (Cuadro 25).

Además, el valor de lambda del modelo SAR 2011 con porcentaje de deserción según el censo es de 0,609 y es significativo con un nivel de significancia del 5%, lo que indica que sí existe autocorrelación espacial en los residuos y a su vez, justifica el hecho de utilizar un modelo SAR y no un modelo de regresión por mínimos cuadrados ordinarios. Por lo tanto, el problema que existía en los modelos SAR anteriores es que los datos administrativos del MEP escondían una autocorrelación espacial existente (Cuadro 25).

Cuadro 25. Modelos SAR 2011 con porcentaje de deserción según el censo.

Variables	Modelo A			Modelo B		
	Coef.	Error est.	Prob.	Coef.	Error est.	Prob.
Constante	65,201	5,015	*	49,044	3,701	*
Tasa de becas	-0,005	0,003		-0,006	0,003	*
Monto promedio de becas	-0,653	0,140	*			
Consumo promedio de electric.	-0,049	0,008	*	-0,054	0,009	*
Población en edad colegial (ln)	-0,797	0,430		-0,845	0,439	
Cantidad de colegios	-0,387	0,190	*	-0,454	0,194	*
Tasa total de profesores	-0,012	0,006	*	-0,013	0,006	*
<i>Lambda</i>	<i>0,609</i>	<i>0,047</i>	<i>*</i>	<i>0,617</i>	<i>0,047</i>	<i>*</i>

*Significativo al 5%.

Fuente: Elaboración propia, 2011.

Para efectos de comparar el modelo SAR 2011 resultante con otros tipos de modelos de regresión, se decidió calcular un modelo SAR 2011 a nivel cantonal y otro modelo de regresión lineal por mínimos cuadrados ordinarios, conservando la misma cantidad de variables en ambos modelos. Con respecto al modelo SAR

2011 a nivel cantonal, los resultados indicaron que el valor de lambda es de 0,680 siendo significativo, lo cual señala que existe autocorrelación espacial en los residuos. Sin embargo, es importante señalar que ninguno de los coeficientes de las variables es significativo con un nivel de significancia del 5%, por lo tanto en este modelo, las variables no logran explicar la asociación estadística con el porcentaje de deserción, aunque el signo de los coeficientes es negativo como es esperado en este caso, excepto para la variable población en edad colegial (logaritmo natural) y los errores estándar de las variables tienen que ser necesariamente mayores en el modelo SAR 2011 cantonal que en el modelo SAR 2011 distrital, porque el tamaño de muestra es menor. Además, el análisis cantonal no tiene suficiente potencia estadística porque son solamente 81 cantones y entre ellos existe heterogeneidad tomando en cuenta tanto el porcentaje de deserción colegial como las variables propias del programa social *Avancemos*. El hecho de que el modelo SAR 2011 distrital tenga mayor potencia estadística por su mayor tamaño de muestra evidencia la ventaja de haber realizado el análisis a nivel de distritos (Cuadro 26).

Cuadro 26. Modelos SAR 2011 con porcentaje de deserción según el censo por cantones.

Variables	Modelo A			Modelo B		
	Coef.	Error est.	Prob.	Coef.	Error est.	Prob.
Constante	32,778	11,070	*	30,151	8,120	*
Tasa de becas	-0,003	0,008		-0,003	0,008	
Monto promedio de becas	-0,090	0,259				
Consumo promedio de electric.	-0,027	0,017		-0,028	0,016	
Población en edad colegial (ln)	1,102	1,065		1,193	1,037	
Cantidad de colegios	-0,065	0,057		-0,067	0,057	
Tasa total de profesores	-0,008	0,020		-0,009	0,020	
<i>Lambda</i>	<i>0,680</i>	<i>0,119</i>	<i>*</i>	<i>0,677</i>	<i>0,119</i>	<i>*</i>

*Significativo al 5%.

Fuente: Elaboración propia, 2011.

A continuación se estima el modelo SAR 2011 con datos distritales, pero usando los estimadores mínimo-cuadráticos, para comparar los resultados. Los

coeficientes de las variables siguen siendo negativos tal y como se espera, no obstante, solamente los coeficientes de las variables monto promedio de becas y consumo promedio de electricidad resultaron ser significativos con un nivel de significancia del 5%, además se observa que los errores estándar son similares tomando en cuenta lo obtenido en el modelo SAR 2011 con porcentaje de deserción según el censo a nivel distrital. También es importante señalar que el coeficiente de determinación R^2 es relativamente bajo, repercutiendo en el porcentaje de variabilidad explicado por la variable porcentaje de deserción según el censo (Cuadro 27). Por lo tanto, dado que con el modelo SAR 2011 todos los coeficientes de las variables resultaron ser significativos con un nivel de significancia del 5% (excepto el coeficiente de la variable población en edad colegial con logaritmo natural) y el valor de lambda también es significativo, estos resultados refuerzan la utilización de un modelo de regresión espacial debido a la existencia de autocorrelación espacial en la estructura de datos distritales de Costa Rica en el 2011. Aun así, se reconoce que, si se hubiera usado la estimación por mínimos cuadrados en lugar de la estimación del SAR, se hubiera llegado a la misma conclusión sustantiva: la variable de aplicación de *Avancemos* monto promedio de becas en miles de colones está asociada significativamente con una reducción en el porcentaje de deserción colegial.

Cuadro 27. Modelos lineales 2011 con porcentaje de deserción según el censo.

Variables	Modelo A			Modelo B		
	Coef.	Error est.	Prob.	Coef.	Error est.	Prob.
Constante	65,304	5,565	*	44,611	3,878	*
Tasa de becas	-0,002	0,003		-0,004	0,003	
Monto promedio de becas	-0,755	0,149	*			
Consumo promedio de electric.	-0,049	0,008	*	-0,052	0,008	*
Población en edad colegial (ln)	-0,502	0,479		-0,302	0,490	
Cantidad de colegios	-0,242	0,224		-0,326	0,229	
Tasa total de profesores	-0,013	0,007		-0,013	0,007	
R^2	R^2 Modelo A = 0,192			R^2 Modelo B = 0,147		

*Significativo al 5%.

Fuente: Elaboración propia, 2011.

Recapitulando lo obtenido al inicio de este capítulo, se recuerda que los modelos SAR con la variable dependiente de datos administrativos provocaban que no fuese identificada la asociación significativa entre el programa social *Avancemos* y la deserción colegial, pero cuando la variable dependiente proviene de datos censales sí se logra identificar una asociación significativa entre el programa social *Avancemos* y la deserción colegial. También, antes el parámetro lambda era no significativo con la deserción proveniente de datos administrativos, es decir, los modelos sugerían que no existía autocorrelación espacial y ahora con la inclusión de la variable dependiente de datos censales el lambda resulta ser significativo con un valor muy alto cercano a 0,62 y por lo tanto sí existe una autocorrelación espacial y la misma estaba siendo ocultada u obviada sencillamente por usar datos administrativos ya que el inconveniente radica en la inexistencia de deserción en aquellos distritos que no poseen colegio, siendo esto erróneo debido a que los datos administrativos del MEP están centrados en el distrito donde está ubicado el colegio y no en el distrito de residencia del estudiante y esto precisamente estaba ocultando el proceso de autocorrelación espacial.

7. Conclusiones

7. 1. Conclusiones referidas a la utilización del modelo SAR con los datos distritales de Costa Rica

En este capítulo se contrastan los principales hallazgos encontrados durante el desarrollo y análisis de la investigación con literatura conocida sobre el tema en cuestión. En relación con el modelo SAR aplicado a los datos distritales costarricenses y al análisis realizado durante esta investigación es importante señalar que si otros investigadores están aplicando un modelo SAR utilizando datos administrativos y encuentran un valor de lambda pequeño, y no es significativo, esto puede deberse a que el tamaño de muestra de la base compuesta por las unidades geográficas no tenga la suficiente potencia estadística para que dicho lambda sea significativo (se requeriría hipotéticamente que Costa Rica tuviera más distritos). Este planteamiento fue obtenido mediante el análisis de las simulaciones tomando en cuenta distintos niveles que determinan la autocorrelación espacial.

Además, con el análisis de las simulaciones queda comprobada la robustez de los supuestos del modelo SAR ante distintos niveles de autocorrelación espacial y con diferentes matrices de pesos. También, es válido indicar que los estimadores del SAR no indican tener sesgo hacia una dirección determinada (las estimaciones estuvieron centradas en el valor esperado) y mantienen una distribución aproximadamente normal, aun cuando la distribución de los errores no sea normal porque las estimaciones obtenidas son similares a los parámetros teóricos propuestos. Este ejercicio fue importante porque se logró detectar que

para valores de λ bajos (en su mayoría menores a 0,23) la potencia de la prueba es baja y esto hace que el λ no sea significativamente distinto de cero en una proporción alta de la muestra, lo cual sugiere que no se podría detectar autocorrelación espacial con una significancia del 5%.

Por lo tanto, para afirmar la existencia de autocorrelación espacial con datos distritales de Costa Rica, es necesario obtener un valor de λ mayor a 0,36 para que la potencia de la prueba sea aceptable, es decir, superior al 70% aproximadamente, sin importar el tipo de matriz de pesos que sea utilizada. Aunado a lo anterior, cabe señalar que el λ obtenido en el modelo SAR 2011 con el porcentaje de deserción según el censo es de 0,61 siendo un valor que tiene una potencia de prueba alta y significativamente distinto de cero, por lo tanto es válido afirmar la existencia de autocorrelación espacial específicamente con la estructura de datos distritales de Costa Rica.

En el modelo SAR, los coeficientes de las variables que están relacionadas directamente con el programa social *Avancemos* (la tasa de la cantidad de becas y el monto promedio de becas en miles de colones) tienen una asociación significativa con el porcentaje de deserción construido a partir de datos censales. En esa misma línea, los coeficientes de las variables del consumo promedio de electricidad, la cantidad de colegios y la tasa total de profesores también resultaron ser significativos en el modelo SAR. La única variable cuyo coeficiente no indica asociación significativa con el porcentaje de deserción según el censo corresponde a la población en edad colegial. El riesgo de utilizar este rango tan amplio de edades es que puede subestimar la asociación. Sin embargo, se escogió porque esa es la población meta del programa social *Avancemos*.

Investigaciones como la realizada por Rosero (2006) refuerzan la importancia de usar datos administrativos en análisis que involucren datos geográficos a nivel distrital para encontrar resultados estadísticamente significativos. En el caso

específico de esta investigación, el modelo SAR logró demostrar la autocorrelación espacial presente en la estructura distrital del país hasta que el porcentaje de deserción fue estimado mediante variables censales. Con registros administrativos, no se encontró autocorrelación espacial, ya que el sistema educativo costarricense sugiere que la autocorrelación existe porque estudiantes que viven en un distrito pueden asistir al colegio de distritos vecinos.

También, se sugiere a las instituciones públicas involucradas en la asignación de programas sociales a hogares con escasos recursos utilizar modelos SAR para medir asociaciones entre variables de interés y así puedan identificar la existencia o no de autocorrelación espacial que permita valorar el otorgamiento de ayudas en diversas regiones del territorio nacional. Por otra parte, si se quiere realizar una investigación de evaluación de impacto de programas, debe ser tomado en cuenta el marco conceptual para el análisis de impacto, los indicadores y diseños para la evaluación de impacto, los modelos de análisis estadístico y las metodologías específicas de interés. Con estos requerimientos en la evaluación de impacto se puede conocer si realmente hubo un impacto después de la implementación del programa.

7. 2. Conclusiones referidas a la implementación del programa social Avancemos y la deserción colegial

Por otro lado, con respecto a la asociación estadística entre el programa social *Avancemos* y la deserción colegial, en esta investigación fueron tomados en cuenta datos administrativos y censales durante el periodo de estudio para analizar los diferentes modelos SAR propuestos. Desde sus inicios, el programa social *Avancemos* ha sido de gran ayuda en aquellas familias de escasos recursos socioeconómicos para mantener a los estudiantes dentro del sistema educativo de secundaria. Precisamente, ese es el objetivo principal del programa,

por ello es importante valorar su ejecución para mejorar la problemática de la deserción colegial presente en nuestro país. De hecho, Trejos (2012b) manifiesta que las becas han repercutido en el aumento entre los asistentes a la secundaria, pasando entre los años 2007 y 2011 del 17,2% al 39,4%, respectivamente. Y, de acuerdo con los resultados obtenidos en esta investigación, la tasa de la cantidad de becas ha contribuido a disminuir el porcentaje de deserción, lo que significa que por cada incremento de 10 puntos porcentuales en la tasa de la cantidad de becas el porcentaje de deserción disminuye en 0,06 puntos porcentuales, manteniendo las demás variables constantes, de manera que el programa social *Avancemos* ha ayudado a mantener a los estudiantes dentro del sistema educativo.

De manera general se puede afirmar que sí existe una asociación estadística entre el programa social *Avancemos* y el porcentaje de deserción colegial cuando esta última es construida a partir de datos censales porque hay datos para cada distrito de estudio. Por el contrario, el porcentaje de deserción medido con datos administrativos se refiere al distrito donde está ubicado el colegio; mientras que las variables relacionadas con el programa social *Avancemos* se refieren al distrito de residencia del estudiante por consiguiente esto oculta el proceso de autocorrelación espacial entre los distritos.

Con respecto al programa social *Avancemos*, Trejos (2012a) destaca que dicho programa ha ayudado a aumentar la presencia de estudiantes en el colegio, dato similar al obtenido en esta investigación porque el modelo SAR indica que conforme aumenta la cantidad de becas, el porcentaje de deserción disminuye, prediciendo la deserción en la dirección que se esperaba. También, Sánchez *et al.* (1984) y Brenes (1995) señalan en otras cosas que la deserción puede estar relacionada con el nivel económico de las personas que tienen a sus hijos dentro del sistema educativo, esto porque en algunos casos el estudiante debe salir del sistema educativo para incorporarse al mercado laboral aunque las opciones

reales sean poco remuneradas y artesanales. Por lo tanto, el ingreso fue estimado mediante la incorporación de la variable consumo promedio de electricidad presente en el modelo SAR como variable *proxy* del ingreso y los resultados arrojan que conforme aumenta el consumo promedio de electricidad, el porcentaje de deserción disminuye. No obstante, en esta investigación el coeficiente de la variable población en edad colegial resultó ser no significativo con un nivel de significancia del 5% y este resultado es diferente al encontrado por Brenes (2014), una de las razones puede ser que en esta investigación algunas de las variables son tomadas de registros administrativos y no totalmente de datos del Censo 2011.

Por otra parte, Collado (2004) recalcó que existe un patrón geográfico en la deserción escolar en ciertos lugares del país. El Programa Estado de la Nación (2003) encontró algunas zonas donde la deserción y la repitencia es más frecuente, así como Gonzaga *et al.* (2008) identificaron ciertas provincias con valores de deserción mayor al promedio nacional. Asimismo, el Programa Estado de la Nación (2008) mostró que el programa social *Avancemos* es prioridad en aquellas zonas geográficas que posean mayores niveles de pobreza, siendo para el 2007, la provincia de Puntarenas la que obtuvo un mayor porcentaje de beneficiarios con el 30%, seguida de Guanacaste con el 22,7%. Y, el Programa Estado de la Nación (2013) expresa que para el 2011 existe un predominio de beneficiarios de las becas de *Avancemos* ubicados en las zonas más alejadas del centro del país. Esto viene a reforzar la autocorrelación espacial encontrada en esta investigación tomando en cuenta la estructura espacial de los distritos de Costa Rica, obteniéndose aglomeraciones en la parte Norte y Atlántica, y en menor medida en la parte Sur y Oeste del territorio nacional.

Y, para finalizar, se recomienda a los investigadores profundizar en la aplicación de los modelos SAR en otros niveles geográficos de nuestro país como por ejemplo el cantonal para verificar si existe o no autocorrelación espacial

tomando en cuenta esa estructura de datos. Además, seleccionar algunos programas sociales que no hayan sido analizados desde el punto de vista espacial para aportar diversas alternativas con distintas variables de interés y así detectar la asociación existente en el caso que aplique.

8. Referencias bibliográficas

Acevedo, I. y Velásquez, E. (2008). *Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales*. Ecos de Economía N°. 27. Medellín, Colombia.

Aguilar, L.F. (2009). *Marco para el análisis de políticas públicas*. En F. Martínez Navarro y V. Garza Cantú, *Política Pública y democracia en América Latina*. Miguel Ángel Porrúa, México.

Alfaro, M., Gamboa, A., Jiménez, S., Martín, J., Ramírez, A. y Vargas, M. (2008). *Construcción del perfil profesional docente de séptimo año: respuesta a una necesidad actual*. Revista Educare. Vol. XII, N°. 2, pp. 31-45.

Anselin, L. y Kelejian, H. (1997). *Testing for spatial error autocorrelation in the presence of endogenous regressors*. Regional Research Institute and Department of Economics. West Virginia University, USA.

Anselin, L. (1988). *Spatial econometrics: methods and models*. *Studies in Operational Regional Science*. Kluwer Academic Publishers. USA.

Anselin, L. (2001). *Spatial econometrics. A companion to theoretical econometrics*. Edited by Badi H. Baltagi. Blackwell Publishing. Oxford, United Kingdom.

Anselin, L. (2003). *Spatial Externalities, Spatial Multipliers, and Spatial Econometrics*. *International Regional Science Review* 26, 2: 153-166. Sage Publications. USA.

- Arce, G. (s.f.). *Políticas educativas y pobreza en Costa Rica. En las políticas de combate a la pobreza en Costa Rica*. Editores Víctor Hugo Céspedes y Ronulfo Jiménez. Academia de Centroamérica, obra sin publicar.
- Arguedas, I. y Jiménez, F. (2008). *Factores que promueven la permanencia de estudiantes en la educación secundaria*. Instituto de Investigación en Educación. San José, Costa Rica.
- Aroca, P. (2000). *Econometría espacial: Una Herramienta Para el Análisis de la Economía Regional*. IDEAR. Universidad Católica del Norte. Chile.
- Borges, A., Patta, B., Cornélio, S. (2009). *Changing the focus: a spatial econometrics assessment of the convergence process in Minas Gerais (1991-2000)*. VI World Conference – Spatial Econometrics Association. Salvador, Brasil.
- Brenes, G. (2014). *Factores Contextuales de la Comunidad Asociados con Repitencia, Deserción y Reprobación en las Escuelas y Colegios de Costa Rica*. Centro Centroamericano de Población (CCP). Universidad de Costa Rica. San José, Costa Rica.
- Brenes, W. (1995). *Determinantes de la deserción y repetición escolar en primero y segundo ciclo: una experiencia de investigación participativa basada en el enfoque de riesgo en escuelas y comunidades de la subregión educativa de Puriscal*. Fondo de las Naciones Unidas para la Infancia (Unicef). San José, Costa Rica.

- Céspedes, V. y Jiménez, R. (2010). *Hacia el fortalecimiento de las políticas de combate a la pobreza en Costa Rica*. Academia de Centroamérica. San José, Costa Rica.
- CGR. (2011). *Informe sobre el resultado de las acciones públicas implementadas en el Ministerio de Educación Pública para mejorar la cobertura en educación secundaria*. Contraloría General de la República. San José, Costa Rica.
- Cliff, A. y Ord, K. (1972). *Testing for Spatial Autocorrelation Among Regression Residuals*. Geographical Analysis Vol. 4, Issue 3. United Kingdom.
- Cliff, A. y Ord, K. (1975). *Model Building and the Analysis of Spatial Pattern in Human Geography*. Royal Statistical Society and the Institute of British Geographers. United Kingdom.
- Cliff, A. y Ord, K. (1981). *Spatial Processes: Models and Applications*. Pion Limited. London, United Kingdom.
- Collado, A. (2004). *Distribución espacial de la fecundidad adolescente en la Gran Área Metropolitana de Costa Rica*. Costa Rica a la luz del Censo del 2000, ed. L. Rosero, 29-62. Centro Centroamericano de Población, Universidad de Costa Rica, Programa Estado de la Nación, Instituto Nacional de Estadística y Censos. San José, Costa Rica.
- Coss, R. (2003). *Simulación: un enfoque práctico*. Editorial Limusa, S.A. de C.V. Grupo Noriega Editores. México, D. F.
- Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Inc. New York, USA.

- Ding, Y. y Fotheringham, S. (1992). *The integration of spatial analysis and GIS*. Comput., Environ and Urban Systems, Vol. 16. pp. 3-19. USA.
- Dormann, C. *et al.* (2007). *Methods to account for spatial autocorrelation in the analysis of species distributional data: a review*. Ecography Vol. 30. Journal Compilation.
- Epperson, B. y Li, T. (1996). *Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics*. Proc. Natl. Acad. Sci. Vol. 93, pp. 10528-10532. USA.
- Fernández, A. y Del Valle, R. (2013). *Desigualdad educativa en Costa Rica: la brecha entre estudiantes de colegios públicos y privados. Análisis con los resultados de la evaluación internacional PISA*. Comisión Económica para América Latina y el Caribe (Cepal). Revista Cepal N°. 111. Santiago de Chile.
- Fotheringham, S. y Rogerson, P. (1993). *GIS and spatial analytical problems*. INT. J. Geographical Information Systems. Vol.7, N°. 1, pp. 3-19.
- Furtado, B. y van Oort, F. (2011). *Neighborhood weight matrix in a spatial-quantile modeling environment: Evidence from Brazil*. European Regional Science Association. Vienna, Austria.
- Galiani, S. (2006). *Políticas sociales: instituciones, información y conocimiento*. División de Desarrollo Social. Comisión Económica para América Latina y el Caribe (Cepal). Santiago de Chile.
- Gentle, J. (1982). *Encyclopedia of statistical sciences*. Willey. New York, USA.

- Giraldo, R. (2011). *Estadística Espacial*. Departamento de Estadística, Universidad Nacional de Colombia. Bogotá, Colombia.
- Gonzaga, W., Valverde, M., Herra, M. y Solís, M. (2008). *Doscientos días anuales de clase efectiva en el Sistema Educativo Costarricense: Un Análisis Situacional*. Instituto de Investigación en Educación, Ministerio de Educación Pública. San José, Costa Rica.
- Griffith, D. (2005). *Spatial Autocorrelation*. Department of Geography, Syracuse University. USA.
- Hernández, D. y Barrantes, H. (2002). *Deserción escolar en séptimo año: Factores que inciden en la deserción escolar de estudiantes de séptimo año*. Ministerio de Educación Pública. San José, Costa Rica.
- INEC. (2006). *Estrategia Nacional de Desarrollo Estadístico, Costa Rica: Hacia el fortalecimiento y desarrollo del Sistema de Estadística Nacional*. Instituto Nacional de Estadística y Censos. San José, Costa Rica.
- Jiménez, W. y Gaete, M. (2010). *Abandono (deserción) escolar en la enseñanza secundaria en Costa Rica, 2009-2010*. Ministerio de Educación Pública. San José, Costa Rica.
- Kelejian, H. y Prucha, I. (1997). *Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem*. *International Regional Science Review* 20, 1 & 2: 103-111. USA.
- Kelejian, H. y Prucha, I. (1998). *A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive*

Disturbances. Kluwer Academic Publishers. Journal of Real Estate Finance and Economics, Vol. 17:1, 99-121. Boston, USA.

Kelley, R. y Barry, R. (1997). *Quick Computation of Spatial Autoregressive Estimators*. Geographical Analysis. Vol. 29, N°. 3. The Ohio State University. Ohio, USA.

Krämer, W. (2003). *Finite sample power of Cliff-Ord-type test for spatial disturbance correlation in linear regression*. Journal of Statistical Planning and Inference 128 (2005) 489-496. Dortmund, Germany.

Kutner, M., Nachtsheim, C., Neter, J. y Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill/Irwin. Fifth Edition. New York, USA.

Lauridsen, J. y Kosfeld, R. (2004). *A Wald Test for Spatial Nonstationary*. Estudios de Economía Aplicada, 22(3), pp. 475-486.

Lee, L. (1999). *Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models*. Department of Economics. The Ohio State University. Ohio, USA.

Mideplan. (2007). *Plan Nacional de Desarrollo "Jorge Manuel Dengo Obregón" 2006-2010*. Ministerio de Planificación Nacional y Política Económica. San José, Costa Rica.

Mideplan. (2010). *Plan Nacional de Desarrollo "María Teresa Obregón Zamora" 2011-2014*. Ministerio de Planificación Nacional y Política Económica. San José, Costa Rica.

- Mideplan. (2013). *Índice de Desarrollo Social*. Ministerio de Planificación Nacional y Política Económica. San José, Costa Rica.
- Mivah. (2007). *Avancemos. Construyamos el futuro compartiendo responsabilidades*. Ministerio de Vivienda y Asentamientos Humanos, Rectoría Sector Social y Lucha contra la Pobreza. San José, Costa Rica.
- Montgomery, D. (2004). *Diseño y análisis de experimentos*. Limusa Wiley. Segunda Edición. Distrito Federal, México.
- Moran, P. (1948). *The Interpretation of Statistical Maps*. Journal of the Royal Statistical Society. Vol. 10, N°. 2, pp. 243-251. London, United Kingdom.
- Patuelli, R., Griffith, D., Tiefelsdorf, M. y Nijkamp, P. (2006). *The Use of Spatial Filtering Techniques: The Spatial and Space-Time Structure of German Unemployment Data*. Tinbergen Institute Discussion Paper N°. 06-049/3. The Netherlands and USA.
- Programa Estado de la Nación. (2003). *Décimo Informe Estado de la Nación en Desarrollo Humano Sostenible*. San José, Costa Rica.
- Programa Estado de la Nación. (2008). *Segundo Informe Estado de la Educación*. San José, Costa Rica.
- Programa Estado de la Nación. (2011). *Tercer Informe Estado de la Educación*. San José, Costa Rica.
- Programa Estado de la Nación. (2013). *Cuarto Informe Estado de la Educación*. San José, Costa Rica.

- PNUD-Costa Rica y UCR. (2011). *Atlas del Desarrollo Humano Cantonal de Costa Rica*. Programa de las Naciones Unidas para el Desarrollo-Costa Rica, Universidad de Costa Rica-Escuela de Estadística. San José, Costa Rica.
- Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (Preal). (2003). *Deserción escolar: Un problema urgente que hay que abordar*. Serie Políticas, Año 5, N°. 14. Formas y Reformas de la Educación, Santiago, Chile.
- Quesada, J., Masís, D., Barahona, M., Meza, T., Cuevas, R. y Rhenán, J. (1999). *Costa Rica Contemporánea: raíces del Estado de la Nación*. Editorial Universidad de Costa Rica. San José, Costa Rica.
- Rosero, L. (2006). *Uso de las estadísticas vitales de mortalidad para evaluar el impacto de la reforma del sector de la salud en las localidades de Costa Rica*. Notas de Población N°. 89 Cepal. Centro Centroamericano de Población, Universidad de Costa Rica. San José. Costa Rica.
- Ruiz, A. (2006). *Universalización de la educación secundaria y reforma educativa*. Editorial de la Universidad de Costa Rica-CONARE. San José, Costa Rica.
- Sánchez, M. y Sauma, P. (2011). *Vulnerabilidad económica externa, protección social y pobreza en América Latina*. Cepal, Flacso, UN/DESA. Santiago, Chile.
- Sánchez, Z., Ugalde, J., Vargas, W y Venegas, P. (1984). *Las incidencias de la crisis socioeconómica en la deserción estudiantil del sistema educativo costarricense*. Ministerio de Educación Pública. San José, Costa Rica.
- Subirats, J., Knoepfel, P., Larrue, C. y Varone, F. (2008). *Análisis y gestión de políticas públicas*. Ariel Ciencia Política. Barcelona, España.

- Trejos, J. D. (2008). *Impacto de los programas sociales selectivos sobre la desigualdad y pobreza*. Decimocuarto Informe Estado de la Nación en Desarrollo Humano Sostenible. Informe Final. San José, Costa Rica.
- Trejos, J. D. (2009). *La inversión social durante la recesión del 2009*. Documento propio. San José, Costa Rica.
- Trejos, J. D. (2012a). *Pobreza, Desigualdad de Oportunidades y Políticas Públicas en América Latina*. Editor responsable: Olaf Jacob. Coeditores: Marcela Peticara y María Lucía Rodríguez. Fundación Konrad Adenauer (KAS). Rio de Janeiro, Brasil.
- Trejos, J. D. (2012b). *La inversión social en el 2011: evolución en un contexto de lenta recuperación económica y crisis fiscal*. Decimoctavo Informe Estado de la Nación en Desarrollo Humano Sostenible. Informe Final. San José, Costa Rica.
- Vargas, W. (1987). *La Deserción Educativa en los Cantones de Mayor Pobreza Rural Costarricense*. Tesis de Maestría en Sociología. Universidad de Costa Rica. San José, Costa Rica.
- Viton, P. (2010). *Notes on Spatial Econometric Models*. City and Regional Planning. The Ohio State University. USA.