

Fiorino, Nadia; Pontarollo, Nicola; Ricciuti, Roberto

Working Paper

A Note on Detecting Dividing Lines in Turnout: Spatial Dependence and Heterogeneity in the 2012 US Presidential Election

CESifo Working Paper, No. 8167

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Fiorino, Nadia; Pontarollo, Nicola; Ricciuti, Roberto (2020) : A Note on Detecting Dividing Lines in Turnout: Spatial Dependence and Heterogeneity in the 2012 US Presidential Election, CESifo Working Paper, No. 8167, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/216563>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

A Note on Detecting Dividing Lines in Turnout: Spatial Dependence and Heterogeneity in the 2012 US Presidential Election

Nadia Fiorino, Nicola Pontarollo, Roberto Ricciuti

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

A Note on Detecting Dividing Lines in Turnout: Spatial Dependence and Heterogeneity in the 2012 US Presidential Election

Abstract

US voters have been moving apart in the last twenty years. This paper analyzes how their voting participation has partitioned by looking at US counties in the 2012 Presidential elections. To tackle this question, we propose a methodology that jointly addresses spatial autocorrelation of the dependent variable and splits the sample according to the non-linearity in the covariates. We find that in different groups of US counties, obtained through a spatial lag regression tree procedure, some variables have different statistical significance (or lack of it), and sometimes different signs. This heterogeneity – which is a manifestation of the complexity of the political behavior - is obfuscated by traditional methods that extrapolate a single average relationship between the variables.

JEL-Codes: D720, C140, C210.

Keywords: turnout, spatial dependence, heterogeneity.

Nadia Fiorino
University of L'Aquila
L'Aquila / Italy
nadia.fiorino@univaq.it

Nicola Pontarollo
University of Brescia
Brescia / Italy
nicola.pontarollo@unibs.it

Roberto Ricciuti
University of Verona
Verona / Italy
roberto.ricciuti@univr.it

1. Introduction

Elections are a central feature of democracy, and scholars have long tried to identify and explain the variation in electoral participation.¹ While a significant strand of literature has shown that the neighborhood and contextual factors may influence voting (among others, Cutts and Fieldhouse 2009), scholarly efforts to explain the electoral behavior of individuals as the result of a *complex* and *multidimensional* process that occurs in space generally do not occupy central locations in the current empirical analyses. Yet, the political and academic debate suggests that the electorate is becoming increasingly polarized. Polarization across new divides such as education (Moretti 2012) or income (McCarty et al. 2016) has become an established feature of both the society and the political system of the US, flanking the more traditional class and racial boundaries (Abramowitz 2014).

Using data on electoral participation in US counties in 2012 Presidential elections, this work investigates the parcellation of the electorate into several homogenous sub-groups that are scattered around the country focusing on the ‘joint effect’ of spatial dependence of voter turnout (i.e. the ‘neighborhood effect’) and the non-linearity effects of its determinants (i.e. the effects producing a sort of ‘polarization’ of voters along the determinants of electoral behavior). Our main contribution is methodological. We extend a classical regression tree, which is not *per se* a spatial technique, taking into account the spatial autocorrelation in the dependent variable. This allows us to deal with the non-homogeneity in regression coefficients avoiding the potential distortionary effects of spatial dependence.

A regression tree approach, initiated by Morgan and Sonquist (1963), is a recursive algorithm that checks for a discontinuity in each covariate, splitting the data into subgroups accordingly, by minimizing the sum of the squared residuals (Breiman et al. 1984). This non-parametric technique searches for an unknown number of sample splits to create subsets of data of growing homogeneity.

¹ For a survey, see Geys (2006).

In this fashion, it allows to identify multiple regimes from a set of control variables.² However, when using geo-referenced data, spatial autocorrelation in the residuals may occur, leading the regression tree - and more generally a regression analysis - to biased and/or inefficient results (Anselin 1988). To overcome this problem, we combine the regression tree approach with spatial autocorrelation in the form of a spatial autoregressive parameter. We build on this methodology and identify the determinants of voter turnout for each homogeneous subgroup of counties, removing issues related to spatial dependence.³ The results confirm a complex multiregime picture of determinants of electoral participation in US counties in 2012 Presidential elections.

2. Literature review

Existing theories and evidence on voting behavior have emphasized that electoral participation is mainly affected by socio-economic, cultural and politico-institutional environments as well as by demographic characteristics. Many studies have also recognized the importance of space and geographical context in explaining individual decisions in electoral participation (Pattie et al. 1995). This literature, which can be summarized in the idea that ‘people who talk together, vote together’ (Pattie and Johnston 2000), is grounded on the theory of political interaction. It largely focuses on neighborhood processes and mainly emphasizes the role of social and household interactions (Fieldhouse and Cutts 2012). Some of these studies have used spatial econometric techniques (e.g., Cutts and Webber 2010), some other traditional nonspatial estimation methods.

Several scholars have also explored the geography of US electoral behavior using a spatial econometric methodology to probe the existence of spatial dependence. Kim et al. (2003) test the hypothesis of reward–punishment and issue–priority voting behavior in the Presidential elections from 1988 to 2000, using a spatial lag model in a Bayesian framework and find significant spatial

² Within this framework, a segmented parametric model is fitted by computing a tree in which every leaf is associated with a fitted model such a linear regression. The benefits from this approach are: the recursive partitioning allows for modeling of nonlinear relationships and automated detection of interactions among the explanatory variables; the statistical formulation of the algorithm ensures the validity of interpretations drawn from the resulting model.

³ Subgroups are not determined by spatial heterogeneity but from non-linearity in explicative variables.

dependence. Tam Cho and Rudolph (2008) investigate the spatial structure of political participation in the United States using a spatial lag model of political participation that restricts the analysis to 5,000 individuals across thirty-two cities and eighteen states, covering every region of the nation. Their findings show that spatial proximity influences voting participation and that the spatial structure of participation is consistent with a contagion process that occurs independently from citizens' involvement in social networks. Lacombe et al. (2014) focus on the potential role that spillovers and spatial components may exert in explaining voter turnout in the 2004 Presidential election. Exploiting advances in Bayesian computation they compare the normal a-spatial linear model, the spatial lag model, the spatial Durbin model, and the spatial Durbin error model and show that the spatial Durbin error model is the most appropriate empirical model. Overall, they find the existence of direct and indirect effects of the set of explanatory variables on voter turnout. Furthermore, several variables traditionally proved to affect voter turnout (i.e. per capita income and the county unemployment rate) are not associated with electoral participation at the county level.

Overall these contributions, also those considering the spatial structure of the data, assume that the relationship between voter turnout and the explanatory variables is linear. This assumption is clearly restrictive when the analysis focuses on local voters and local politics (see Calvo and Escobar 2003) and more in general on electoral processes that control for a variety of variables across a country like the US, which is largely heterogeneous and increasingly polarized in terms of demographic and economic characteristics (Glaser and Ward 2006; Moretti 2012).

Bringing together the relevance of both context and heterogeneity is the goal of this paper.

3. Methodology and data

We start from the spatial lag (SAR) model (Anselin 1988):

$$y = \rho W y + X\beta + \varepsilon \tag{1}$$

where y is the vector corresponding to the independent variables, $\mathbf{W}$ is a $N \times N$ spatial weight matrix, where N is the number of observations, X the matrix of independent variables, β the vector of coefficients and ε the vector of i.i.d. residuals. The parameter ρ , finally, is the spatial lag coefficient, which varies between the minimum and maximum value of the eigenvalue extracted by $\mathbf{W}$, typically around -1 and 1, and measures the strength of the spatial dependence.

To estimate a regression tree in combination with a spatial lag model, eq. (1) is initially estimated, as customary, through a Maximum Likelihood estimator, and then rewritten as:

$$y^* = X\beta + \varepsilon \quad (2)$$

where $y^* = y - \rho\mathbf{W}y$, is the spatially filtered dependent variable. In our case, instead of a global linear regression where there is a single predictive formula holding over the entire data-space, the estimation of eq. (2) is based on a regression tree that performs a recursive partitioning on the model subsampling in correspondence to nonlinearities in the parameters. The algorithm stops when it finds no additional significant non-linearities (at the 5% level) or the sample is smaller than a given threshold. In each terminal node of the regression tree, thus, the initial model is estimated linearly. Parameters are estimated in each (sub)sample by ordinary least squares (OLS) and for each step, additional ‘partitioning’ variables are employed with respect to which parameter stability is first assessed and, if any is found, the best split in subsamples is selected by minimizing the residual sum of squares of the partitioned model.

The empirical analysis employs data on the 2012 Presidential elections at the county level. Theoretical and empirical research have shown the relevance of socio-economic, demographic, cultural and politico-institutional factors in explaining voter turnout at both the individual as well as aggregate level. We build upon the major contributions of this literature and estimate the following regression:

$$Turnout_{c,s} = \rho\mathbf{W}Turnout_{c,s} + \alpha_c + \alpha_3\mathbf{EDU}_{c,s} + \alpha_4\mathbf{ECO}_{c,s} + \alpha_5\mathbf{SOCIOPOL}_{c,s} + \varepsilon_{c,s} + \eta_s + v_s \quad (3)$$

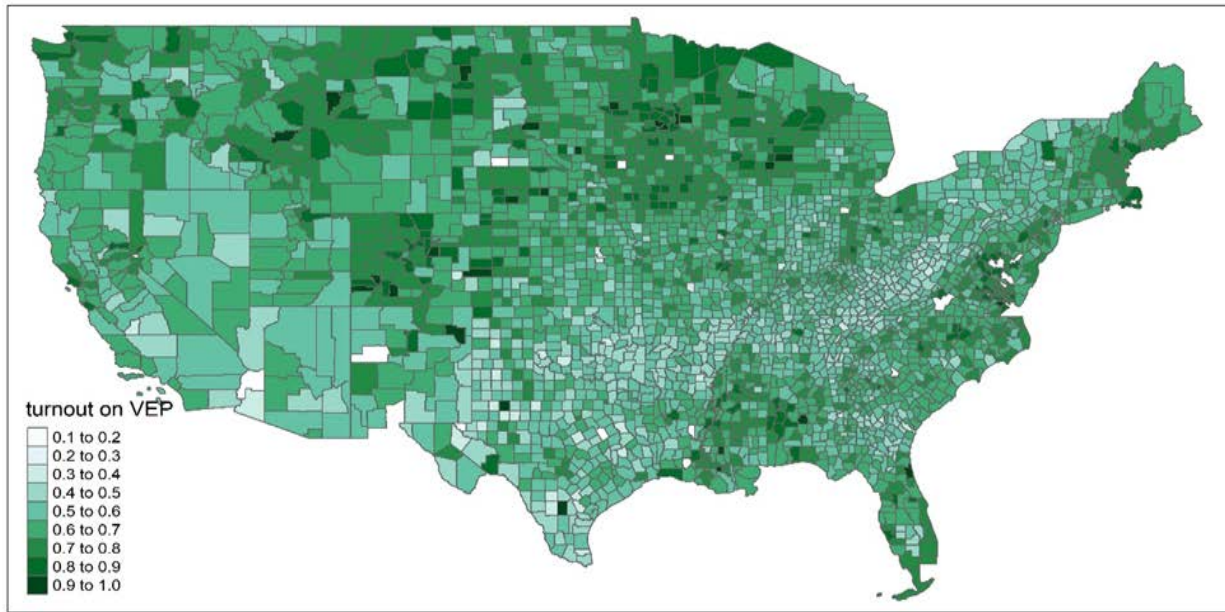
where c is the number of counties. Finally, ε_c is the usual i.i.d. residuals term.

The variables can be described as follows:

- 1) *Turnout*, the dependent variable, is the share of the county voting eligible population (VEP) which is registered and legally empowered to cast its vote in Presidential elections. Using VEP instead of VAP (voting age population) to measure turnout corrects for the number of ineligible felons and non-citizen residents. Nevertheless, the use of these two different types of data is controversial as both measurement methods have their biases (Holbrook and Heidbreder, 2010). We, therefore, estimate the models also using the citizen voting age population (CVAP). The results are presented in Appendix 3.
- 2) The vectors **EDU** and **ECO** include variables aiming at capturing the role of education (*Less than graduate* and *University education*) and of the economic system (*Unemployment* and *Household income*) respectively.
- 3) The vector **SOCIOPOL** contains several socio-demographic variables (*Congregations*, *Urban population*, *Hispanics*, *Blacks*, and *Veterans*). *Blacks* and *Hispanics* indicate the percentage of a county's population of the respective racial category.

The description, the summary statistics and the source of all the data are reported in Table A1.1 and A1.2 in the Appendix. Table A1.3 displays the correlation matrix of the variable involved in the analysis. Figure 1 shows the distribution of turnout across US counties.

Figure 1: voter turnout in the 2012 US Presidential elections across counties



Source: uselectionatlas.org

4. Results

Table 1 details the results of our model for the eleven terminal nodes that share common drivers of voting behavior. We observe that the spatial lag coefficient is statistically significant (therefore the ‘neighborhood effect’ is in play) and, contrary to the standard regression tree model (Table A2.1), no spatial autocorrelation is found in the residuals, guaranteeing unbiased results with respect to spatial autocorrelation. The spatial lag regression tree is preferred to the standard regression tree model also in terms of the Akaike Information Criterion (AIC). Among the spatial lag regression trees, the one with a contiguity matrix based on 9 nearest neighbors is the best in terms of AIC. The significant spatial lag parameter in voting behavior is emphasized by Lacombe et al. (2014) and might be due to the spatial structure of participation, primarily driven by a diffusion process or, alternatively, can be attributed to the spatial structure embedded within the independent variables. This is likely the result of a number of factors such as labor and capital mobility and a common culture or social involvement and social networking (McClurg 2003). Coherently with Tam Cho and Rudolph (2008), we find that political participation is geographically clustered, i.e. citizens’

political participation is influenced by neighbor citizens' political participation, and this is valid even after controlling for additional factors.

The most relevant non-linearities are found for median household income, the root or father node (the cut is at 10,557 dollars), followed by black population (the cut is 2%) and subsequently by urban population (91.1%) and university graduates (43.7%) (see fig. 3). The fact that, given the opportunity to split the sample by several variables, the spatial lag regression tree shows a preference for income splits suggests that income dominates the share of the black, the less than graduate and the urban population as the main identifier for multiple regimes. Economic conditions of voters, as well as race and ethnicity, are the deepest social and cultural divisions within the American electorate and their impact on voter turnout is a persistent finding in the literature (Fraga 2015; McCarty et al. 2016). Nonetheless, within the eleven subsets of counties, the results emphasize different sign and significance of the determinants of turnout.

Figure 2: Spatial lag regression tree (W = 9 k-near neighbours)

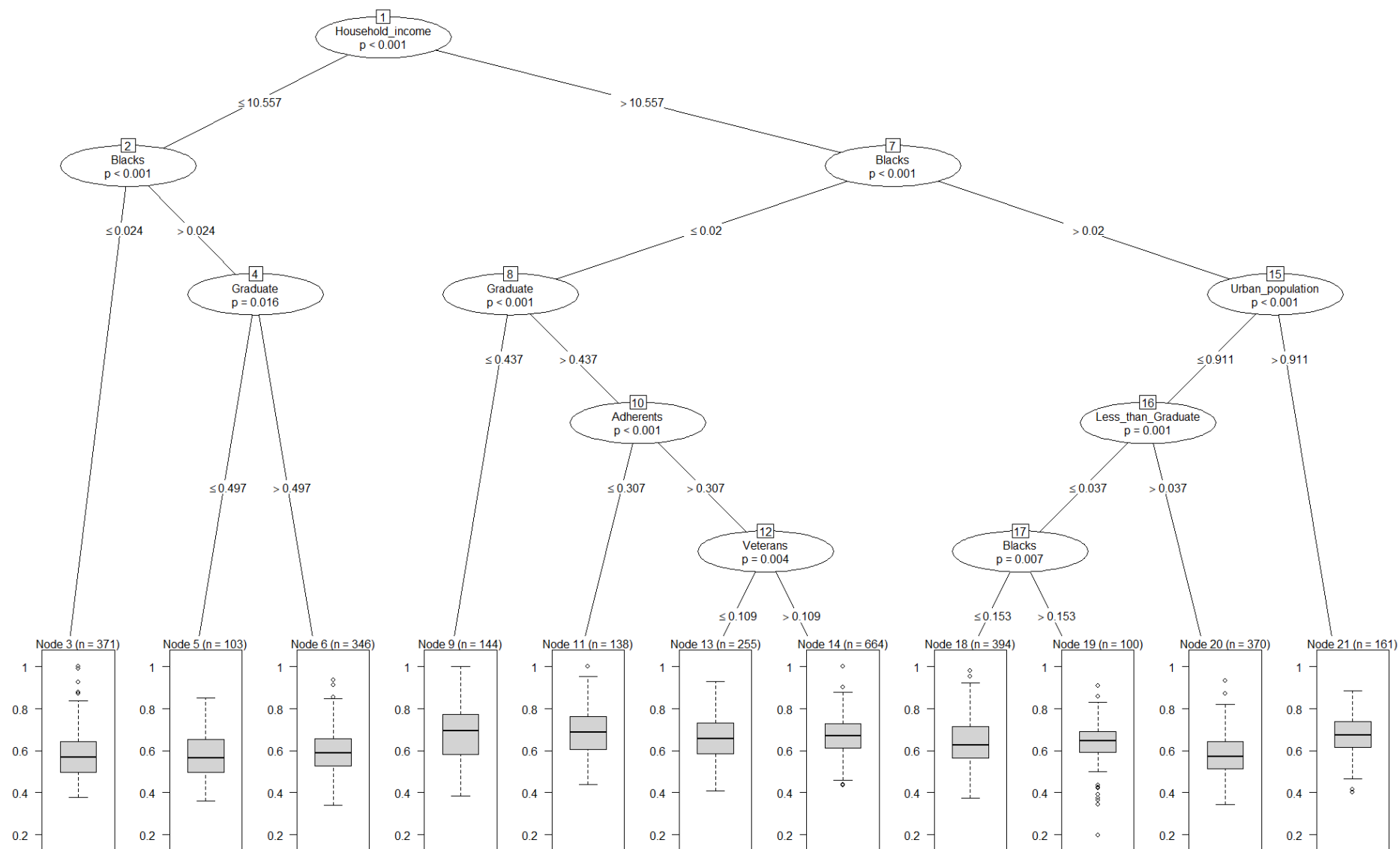


Table 1: Spatial lag regression tree estimates (W = 9 k-near neighbours)

	node 3	node 5	node 6	node 9	node 11	node 13	node 14	node 18	node 19	node 20	node 21
Intercept	-0.9781*** (0.3538)	-1.3067* (0.7410)	0.6338 (0.4103)	-3.0059*** (0.3461)	-0.4830 (0.5506)	-1.2662*** (0.3738)	-1.0620*** (0.2749)	-2.3314*** (0.2464)	-3.0977*** (0.5489)	-0.9952*** (0.3735)	0.1361 (0.4716)
Congregations	0.0252 (0.0184)	0.0415 (0.0440)	0.0565** (0.0239)	-0.0054 (0.0334)	-0.1337 (0.1194)	-0.0159 (0.0285)	0.0609*** (0.0170)	0.0940*** (0.0299)	0.2298*** (0.0657)	0.0924*** (0.0237)	-0.1474*** (0.0538)
Blacks	-2.0733*** (0.5461)	0.1265*** (0.0426)	0.1748*** (0.0225)	-2.1318* (1.2705)	-5.4754*** (1.4141)	1.6377* (0.8661)	-0.9636* (0.5708)	0.0634 (0.0922)	0.1031 (0.0768)	0.1323*** (0.0330)	-0.0162 (0.0575)
Hispanics	0.1153*** (0.0204)	-0.1936*** (0.0688)	-0.1416* (0.0768)	0.1284*** (0.0453)	0.1313 (0.1099)	-0.0590 (0.0430)	-0.0428 (0.0381)	-0.4669*** (0.0821)	-0.6868*** (0.2323)	-0.2353*** (0.0359)	-0.0741 (0.0748)
Household income	0.1198*** (0.0336)	0.1421** (0.0719)	0.0096 (0.0374)	0.3092*** (0.0304)	0.0829* (0.0471)	0.1546*** (0.0311)	0.1242*** (0.0235)	0.2384*** (0.0209)	0.2973*** (0.0474)	0.1464*** (0.0323)	0.0280 (0.0348)
Less than graduate	-0.1171 (0.1313)	0.6719* (0.3805)	-0.8257*** (0.2448)	-0.8966*** (0.2229)	-0.8457* (0.4836)	-0.5949** (0.2342)	-0.7397*** (0.1995)	-0.7456 (0.5828)	1.4295 (1.3272)	-0.0133 (0.2455)	-1.7201*** (0.6512)
University education	0.0013 (0.0696)	0.0517 (0.1688)	-0.9790*** (0.1463)	-0.1245 (0.1496)	-0.2352 (0.1602)	0.0654 (0.1219)	-0.0918 (0.0739)	-0.1031 (0.0698)	0.2213 (0.1544)	-0.6723*** (0.0935)	-0.1189 (0.1414)
Urban population	-0.1264*** (0.0174)	-0.1193*** (0.0318)	-0.1601*** (0.0184)	-0.0922*** (0.0250)	-0.0974*** (0.0284)	-0.0757*** (0.0178)	-0.0541*** (0.0119)	0.0476** (0.0187)	0.0023 (0.0351)	-0.0921*** (0.0157)	0.1007 (0.2183)
Veterans	0.5390*** (0.1349)	-0.2055 (0.3868)	0.7354*** (0.1968)	1.3029*** (0.3065)	1.2261*** (0.2331)	-0.3697 (0.4118)	0.8594*** (0.1590)	0.2622* (0.1547)	-0.7010*** (0.2500)	0.2530 (0.1732)	0.1551 (0.3035)
Unemployment	-0.6300*** (0.0849)	0.4599* (0.2453)	0.0196 (0.1329)	-0.4125 (0.2971)	-0.2461 (0.2198)	-0.8937*** (0.1769)	-0.2682** (0.1237)	0.3351 (0.2059)	0.0699 (0.5019)	-0.1085 (0.1709)	-0.6213 (0.4656)
Rho	0.5465*** (0.0173)										
AIC	-7855										
LM test autocorr. (p-val)	< 0.001										

Notes: *** p<0.01, ** p<0.05, *p<0.10. Standard errors in parenthesis.

Node 3 includes 371 counties with median household income lower than 10,557 dollars and a black population lower than 2.4%. In this node Hispanics, household income and the veterans increase voter turnout while Blacks, urban population and the percentage of the unemployed have a negative impact on voter turnout.

Node 5 (n=103) includes a subset of counties with the same income characteristics as before, a black population higher than 2.4% and a percentage of people with less than graduate education lower than 49.7%. In this subgroup the higher the Blacks, the household income, the less than graduate and unemployed people, the higher voter turnout. The results also indicate that the population living in urban areas has a negative impact on electoral participation.

Node 6 (n=346) differs from the previous node for a percentage of people with less than graduate education higher than 49.7%. In this subset, the results show that the percentage of people affiliated to a congregation, the Blacks and the Veterans increase voter turnout while Hispanics, urban population, less than graduate or with university education are negatively associated with voter turnout.

Node 9 (n=144) encompasses a group of cases of median household income higher than 10,557 dollars, black population lower than 2% and a percentage of less than graduate lower than 43.7%. In this terminal group as the Hispanics, the household income and the Veterans increase also voter turnout increases.

Node 11 (n=138) includes a subset of counties with the same income characteristics as before, a black population lower than 2%, a percentage of people less than graduate higher than 43.7% and a percentage of people affiliated to a congregation lower than 30.7. In this node, median household income and the Veterans increase electoral participation while Blacks, urban population and less than graduate and university graduate do not.

In Node 13 (n=255) are included counties with median household income higher than 10,557 dollars, black population lower than 2%, a percentage of people with less than graduate education higher than 43.7%, a percentage of people affiliated to a congregation higher than 30.7 and

Veterans lower than 10.9%. In this node the higher the Blacks and the median household income, the higher voter turnout. The results also indicate that the urban population, less than graduate, university graduate and the unemployed negatively impact on electoral participation.

Node 14 (n=664) shares the same characteristics of Node 13 but the percentage of Veterans is higher than 10.9%. In this subset people affiliated to a congregation as well as the Veterans increase voter turnout. Household income also has a positive impact, while Blacks, people living in urban areas, people not graduated and people unemployed decrease turnout.

Node 18 (n=394) takes in counties with median household income higher than 10,557 dollars, black population higher than 2%, the percentage of urban population lower than 91.1%, people with less than graduate education lower than 3.7% and black population lower than 15.3%. It is worth noting that in this (and in the following) nodes black population enters again, showing a second non-linearity in this variable. In this terminal group, people adherent to congregations, the Veterans, the urban population and finally the median household income increase voter turnout. The results also show that Hispanics negatively affect voter turnout.

Node 19 (n=100) shares the same characteristics of Node 18 with the black population higher than 15.3%. In this subset of counties, the adherents to a congregation and the median household income increase voter turnout while the Hispanics and the Veterans decrease participation.

Node 20 (n=370), besides the first three cuts as before, has people with less than graduate education higher than 3.7%. In this subset, the Hispanics and population with less than graduate education and the population living in urban areas reduce the participation. Adherents and Blacks are more likely to turn out to vote. Also, household income and voting age positively impact on turnout.

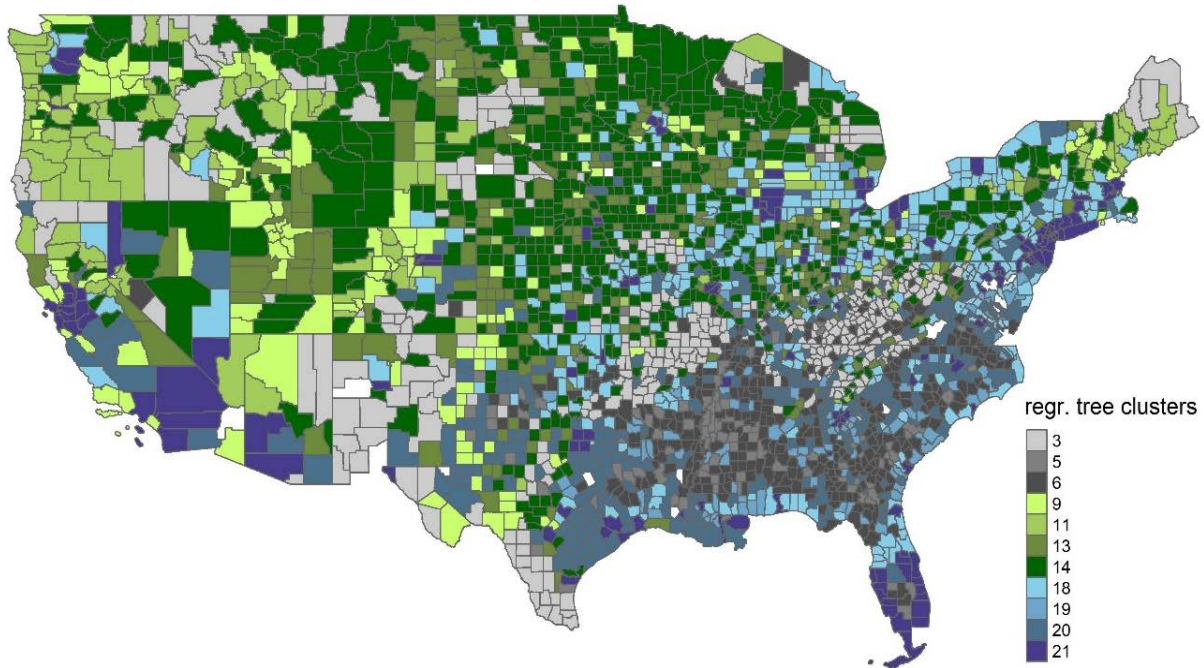
Finally, Node 21 (n=161) consists of counties that after the first two cuts that we have previously seen, have the percentage of the urban population higher than 91.1%. In this terminal group people participating in a congregation and population with less than graduate education reduce the voter turnout.

Our results can be interpreted in two ways: first, by looking at the intermediate nodes and underlining commonalities and differences of the terminal nodes that have one (or more) common departing point(s), given by similar colors; second, by looking at the geographical patterns of terminal nodes across the US, therefore eyeing at the dissimilarities among colors. The first point of view emphasizes the variables involved in the estimations, the second the outcome of the splitting process.

Overall our results show that throughout the sub-groups the effects of the explanatory variables on electoral participation are not uniform and vary across contexts. Indeed, the magnitude and the sign of the coefficients of the variables the analysis includes are partly different across the sub-groups.

Fig. 3 displays the homogeneous sub-groups of counties that share similar features in the determinants of their voting behavior. Colors are based on the most relevant non-linearities, i.e. the ones for median household income and black population. Counties depicted in greyscale are the ones with household income below 10,557 dollars, while the ones in green and in blue-scale above 10,557 dollars. Furthermore, the counties in green-scale have a Black population lower than 2%, and the counties in blue-scale higher than 2%. The map highlights a clustering of different subgroups in several areas of the country, which in turn is a sign of polarization. The group of counties with household income below 10,557 dollars is mainly located in the south-eastern part of the country, with the exception of the coast, where there are richer counties with a Black population higher than 2%. In the area with vertices Kansas to Michigan and from northern California to the State of Washington there are counties with household income above 10,557 dollars and less than 2% of Black people. A few interesting subsets of counties emerge. The counties located on the southern coasts and on the South-East of the country belong to Node 3, while there are large swathes of green (Node 14) in most of the central states. Texas and Florida are divided into two groups: Node 3 and Node 20 (which are very far apart in the figure) for the Lone Star State and Node 18 and Node 21 for the Sunshine State. Node 3 is also recurrent in the Appalachian and mid-West. Finally, there is a group of countries in subgroup 21 belonging to California and New York.

Fig. 3 Groups identified by regression tree analysis



Source: original elaborations

5. Conclusions

We have proposed a combined methodology to deal with heterogeneity in regression coefficients and spatial dependence in the analysis of turnout in the 2012 US Presidential elections. We find that turnout in a given county is function of the turnout in the surrounding ones and that, in different groups of US counties, obtained through a spatial lag regression tree procedure, some variables have different statistical significance (or lack of it), and sometimes different signs, revealing multiple regimes of voting behavior mainly driven from household income, the root node. This non-homogeneity in regression coefficients, which in the spatial lag regression tree are unbiased by spatial dependence, is obfuscated by traditional methods that extrapolate a single average relationship between the variables.

Further work may add a temporal dimension to shed light on the degree of persistence of the groups, as well as a geographically detailed analysis of the individual determinants of electoral behavior.

References

- Abramowitz, A.I. (2015). The New American Electorate, in Thurber J.A and Yoshinaka, A (eds.) *American Gridlock: The Sources, Character, and Impact of Political Polarization*, New York, Cambridge University Press: 19-45.
- Anselin L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., (1984). *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA.
- Calvo, E., Escobar, M. (2003). The Local Voter: A Geographically Weighted Approach to Ecological Inference. *American Journal of Political Science* 47 (1): 189-204.
- Cutts, D., Fieldhouse, E. (2009). What Small Spatial Scales are Relevant as Electoral Contexts for Individual Voters? The Importance of the Household on Turnout at the 2001 General Election. *American Journal of Political Science* 53 (3): 726-39.
- Cutts, D., Webber, D. (2010). Voting Patterns, Party Spending and Relative Location in England and Wales. *Regional Studies* 44 (6): 735-60.
- Fieldhouse, E., Cutts, D. (2012). The Companion Effect: Household and Local Context and the Turnout of Young People. *The Journal of Politics* 74 (3): 856-869.
- Fraga, B.L. (2016). Candidates or Districts? Reevaluating the Role of Race in Voter Turnout. *American Journal of Political Science* 60 (1): 97-122.
- Glaser E.L., Ward, B.A. (2006). Myths and Realities of American Political Geography. *Journal of Economic Perspectives* 20 (2): 119–144.
- Holbrook, T., Heidbreder, B. (2010). Does Measurement Matter? The Case of VAP and VEP in Models of Voter Turnout in the United States. *State Politics & Policy Quarterly* 10 (2): 157-179.

- Kim, J, Euel E., Ding-Ming W. (2003). A Spatial Analysis of County-level Outcomes in US Presidential Elections, 1988-2000. *Electoral Studies* (22): 741-61.
- Lacombe, D., Holloway, G., Shaughnessy, T. (2014). Bayesian Estimation of the Spatial Durbin Error Model with an Application to Voter Turnout in the 2004 Presidential Election. *International Regional Science Review* 37 (3): 298-327.
- McCarty, N., Poole, K.T., Rosenthal, H. (2016). *Polarized America: The Dance of Ideology and Unequal Riches*. MIT Press: Cambridge, MA.
- McClurg S.D. (2003), Social Networks and Political Participation: The Role of Social Interaction in Explaining Political Participation, *Political Research Quarterly* 56: 449-464.
- Moretti (2012). *The New Geography of Jobs*. Houghton Mifflin Harcourt, Boston, New York.
- Morgan, J.N., and Sonquist, J.A. (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association*, 58: 415-434.
- Pattie, C.J., Johnston, R.J. (2000). People Who Talk Together Vote Together: An Exploration of Contextual Effects in Great Britain. *Annals of the Association of American Geographers* 90 (1): 41-66.
- Pattie, C.J., Johnston, R.J., Fieldhouse, E.A. (1995). Winning the Local Vote: The Effectiveness of Constituency Campaign Spending in Great Britain, 1983–1992. *American Political Science Review* 89 (4): 969-983.
- Tam Cho, W.K., Rudolph, T.J. (2008). Emanating Political Participation: Untangling the Spatial Structure behind Participation. *British Journal of Political Science* 38 (2): 273-289.

APPENDIX

Appendix 1 provides variables' definitions and sources, descriptive statistics, and a correlation matrix.

Figure A2.1 and Table A2.1 in **Appendix 2** give the results for the standard regression tree model, without the spatial lag. Comparison of the AIC statistics claims that the spatial lag regression tree outperforms the standard model. Moreover, Appendix 2 provides robustness checks when using voting eligible population (VEP) as dependent variable for several structures of the **W** matrix. The results for the parameter stability tests (Hjort and Koning, 2002, Zeileis, 2005) in the terminal nodes, i.e. in the final iteration of the estimation procedure are shown in Table A2.10. There are not splits at the 5% significance level, which means that no additional non-linearities are present. As the minimum size of the group has been set to 100, the test is not carried out for subsamples with a lower size.

Hjort NL, Koning A (2002). "Tests for Constancy of Model Parameters over Time." *Nonparametric Statistics*, 14, 113–132

Zeileis A (2005). "A Unified Approach to Structural Change Tests Based on ML Scores, F Statistics, and OLS Residuals." *Econometric Reviews*, 24, 445–466.

Appendix 3 presents the regression trees results when the dependent variable is CVAP (citizen voting age population) by using several specifications of the **W** matrix. As above, Table A3.10 reports the stability tests, which share the above results.

Table A1.1: Variables, description and sources

Variable	Description	Source
VEP	Turnout based on Voting Eligible Population	uselectionatlas.org; http://www.electproject.org/ and Special Tabulation From the 2006-2010 5-Year American Community Survey
CVAP	Turnout based on Citizen Voting Age Population	uselectionatlas.org; Special Tabulation From the 2006-2010 5-Year American Community Survey
Urban population	% of population living in urban areas	Census 2010
Veterans	% of veterans over population	Census 2010
Unemployment	% of unemployed people	Census 2010
Congregations	% of people affiliated to a congregation in every U.S. county	Census 2010
Household income	Log of the median household income	Census 2010
University education	% of people 25 years and over with 4 or more years of college (until 1980) or bachelor's degree or higher (since 1990)	Census 2010
Less than graduate	% of people 25 years and over with less than 9th grade	Census 2010
Hispanics	% of Hispanic people	Census 2010
Blacks	% of black people	Census 2010

Table A1.2: Descriptive statistics

	Min.	Mean	Std.dev	Max.
CVAP	0.1747	0.5861	0.0962	1
VEP	0.1971	0.6307	0.1085	1
Urban population	0	0.4068	0.3094	1
Veterans	0.0275	0.1127	0.0271	0.3149
Unemployment	0	0.0691	0.031	0.361
Congregations	0.0307	0.5118	0.1724	1
Household income	9.885	10.693	0.2384	11.712
Less than graduate	0.0000	0.0423	0.0264	0.2102
University education	0.1865	0.5191	0.0679	0.7791
Hispanics	0	0.0830	0.1326	0.957
Blacks	0	0.0874	0.1435	0.857

Table A1.3: correlation matrix

	CVAP	VEP	Urban population	Veterans	Unemployment	Congregations	Household income	Less than Graduate	University education	Hispanics	Blacks
CVAP	1										
VEP	0.9594	1									
Urban population	-0.0553	-0.0515	1								
Veterans	0.1623	0.1586	-0.1938	1							
Unemployment	-0.2884	-0.2679	0.0397	-0.1108	1						
Congregations	0.066	0.0551	-0.0401	-0.1317	-0.2423	1					
Household income	0.4241	0.4095	0.4029	0.0316	-0.4403	-0.0323	1				
Less than graduate	-0.4118	-0.398	-0.1727	-0.3552	0.2155	0.1047	-0.4852	1			
University education	-0.0876	-0.0814	-0.5664	0.4785	0.0483	-0.0266	-0.3997	-0.0427	1		
Hispanics	-0.2252	-0.2215	0.2829	-0.2484	-0.0024	0.1147	0.0338	0.4793	-0.3705	1	
Blacks	-0.01	-0.0025	0.0897	-0.2641	0.4237	-0.0069	-0.2999	0.1578	-0.0684	-0.1001	1

APPENDIX 2 - RESULTS ON VEP

Figure A2.1: standard regression tree

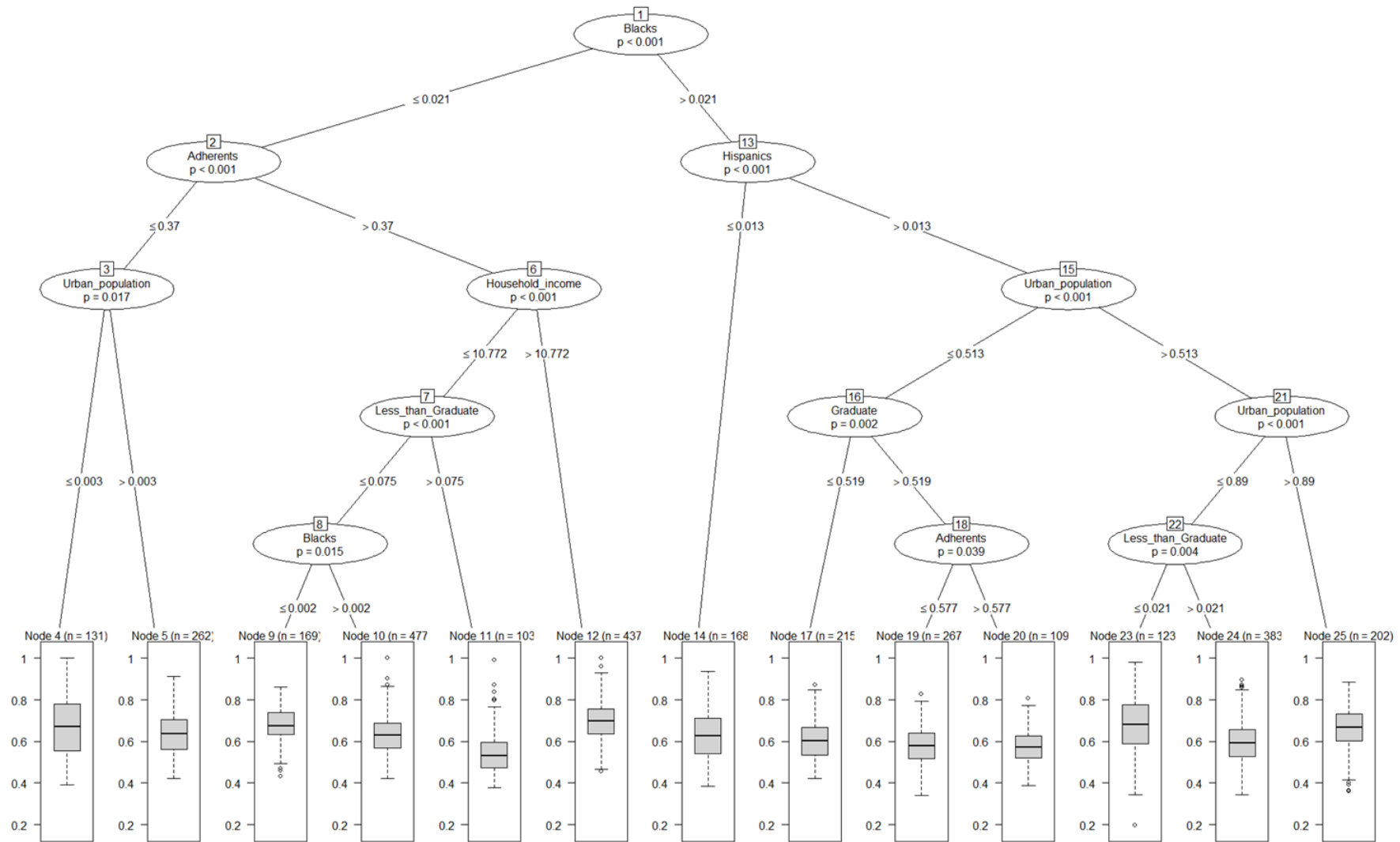


Table A2.1: standard regression tree

[illegible]

Figure A2.2: spatial lag regression tree ($W = 5$ k-near neighbours)

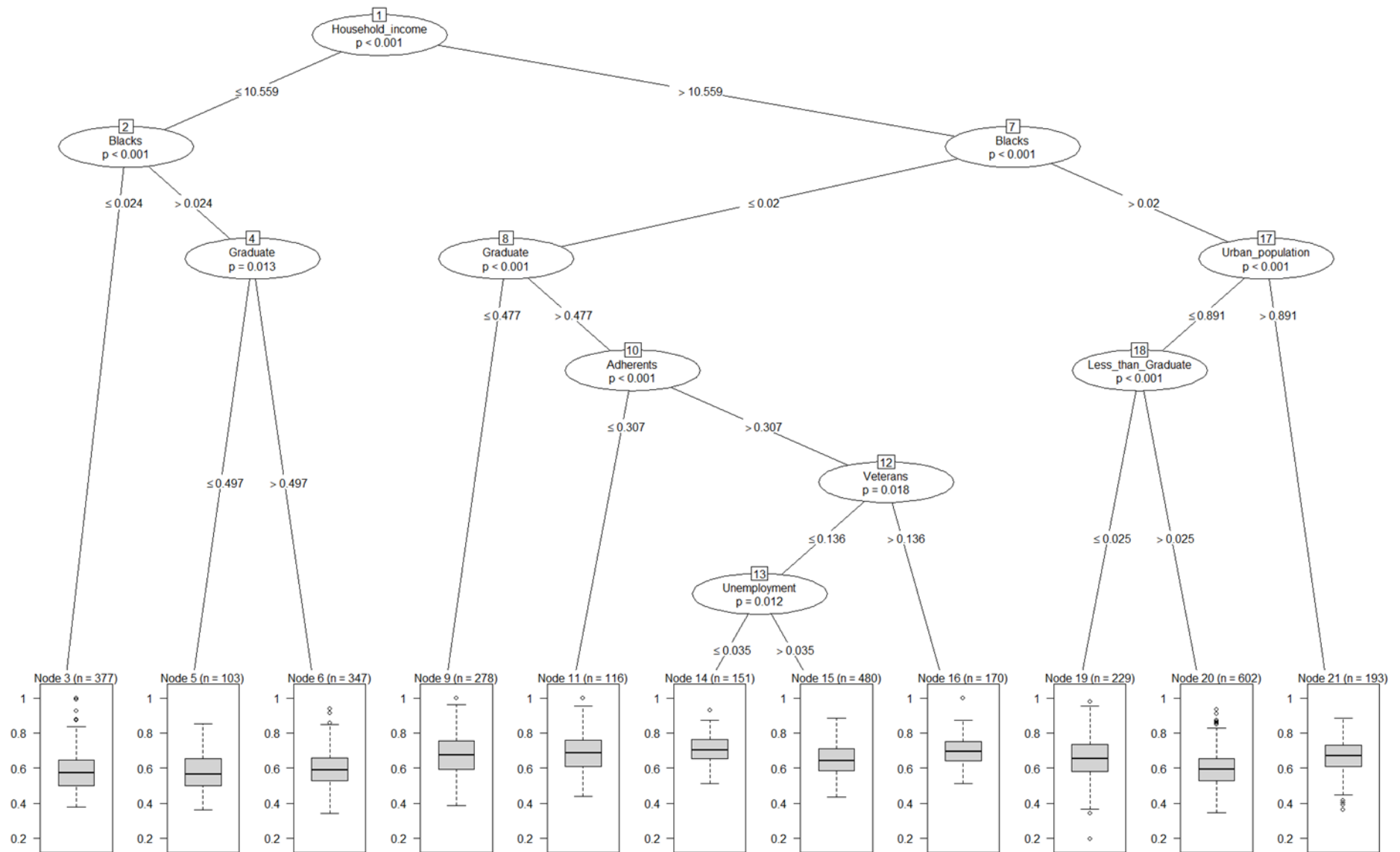


Table A2.2: spatial lag regression tree ($\mathbf{W} = 5$ k-near neighbours)[illegible]

Figure A2.3: spatial lag regression tree ($W = 6$ k-near neighbours)

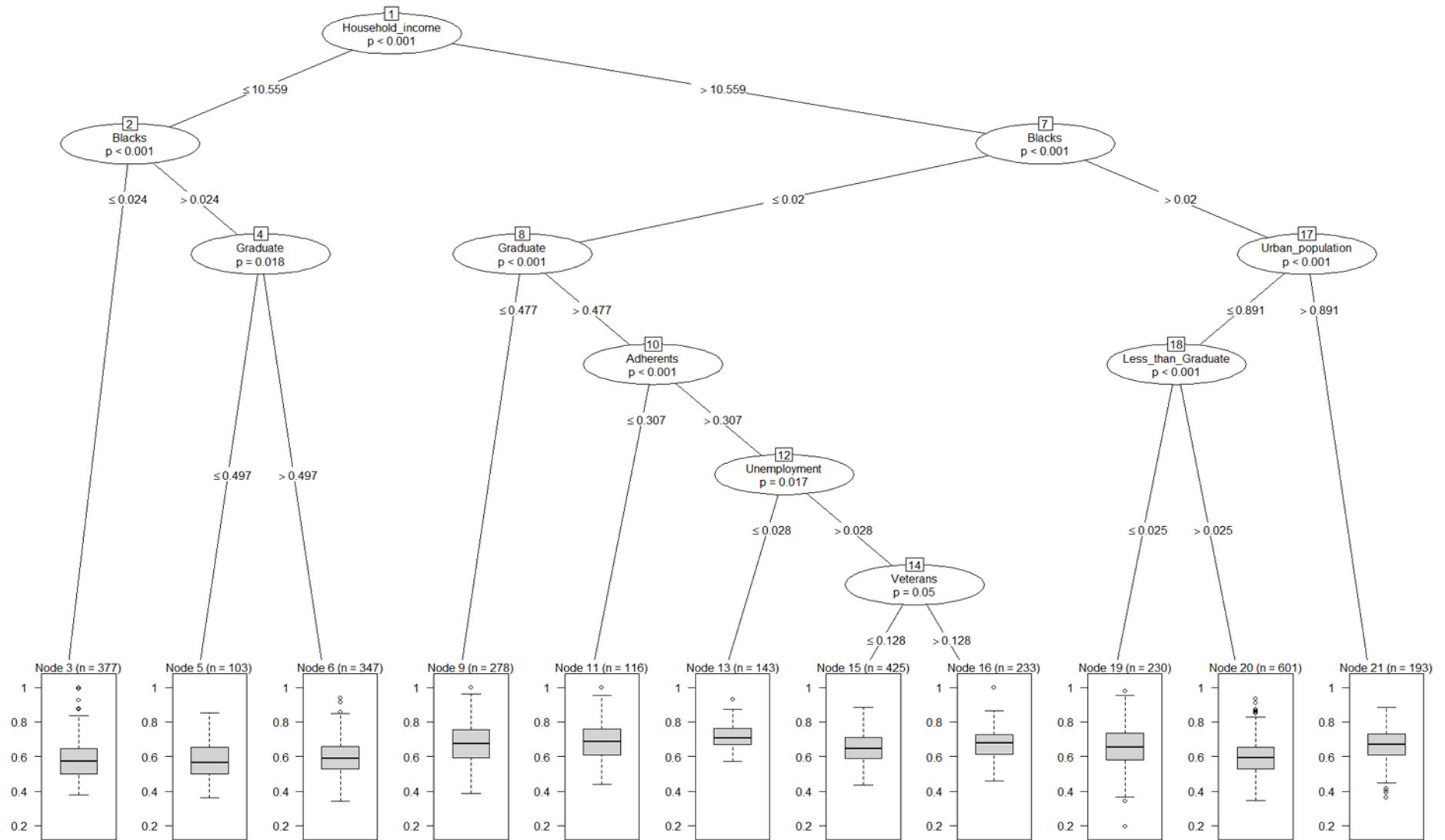


Table A2.3: spatial lag regression tree ($\mathbf{W} = 6$ k-near neighbours)[illegible]

Figure A2.4: spatial lag regression tree ($W = 7$ k-near neighbours)

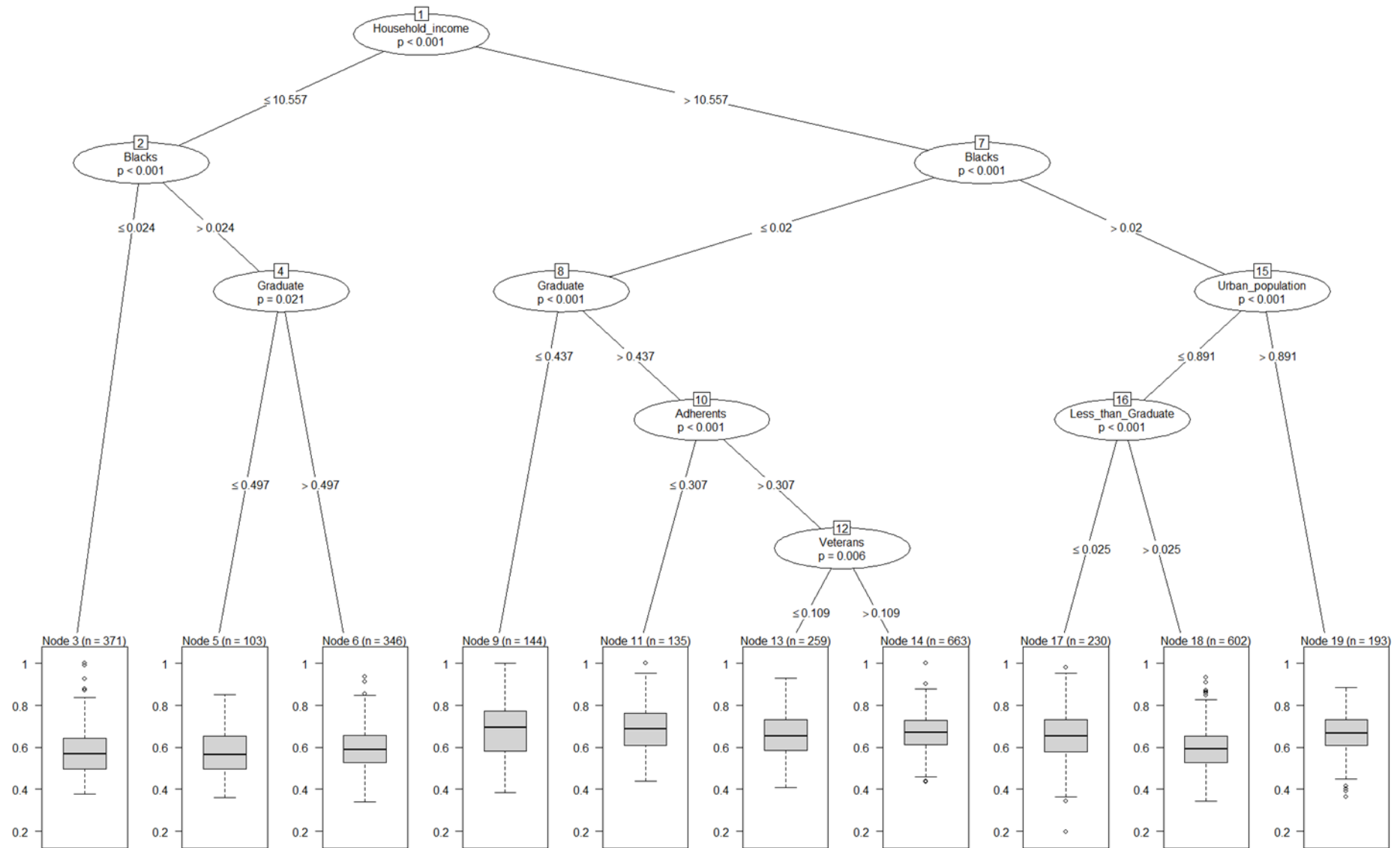


Table A2.4: spatial lag regression tree ($\mathbf{W} = 7$ k-near neighbours)[illegible]

Figure A2.5: spatial lag regression tree ($W = 8$ k-near neighbours)

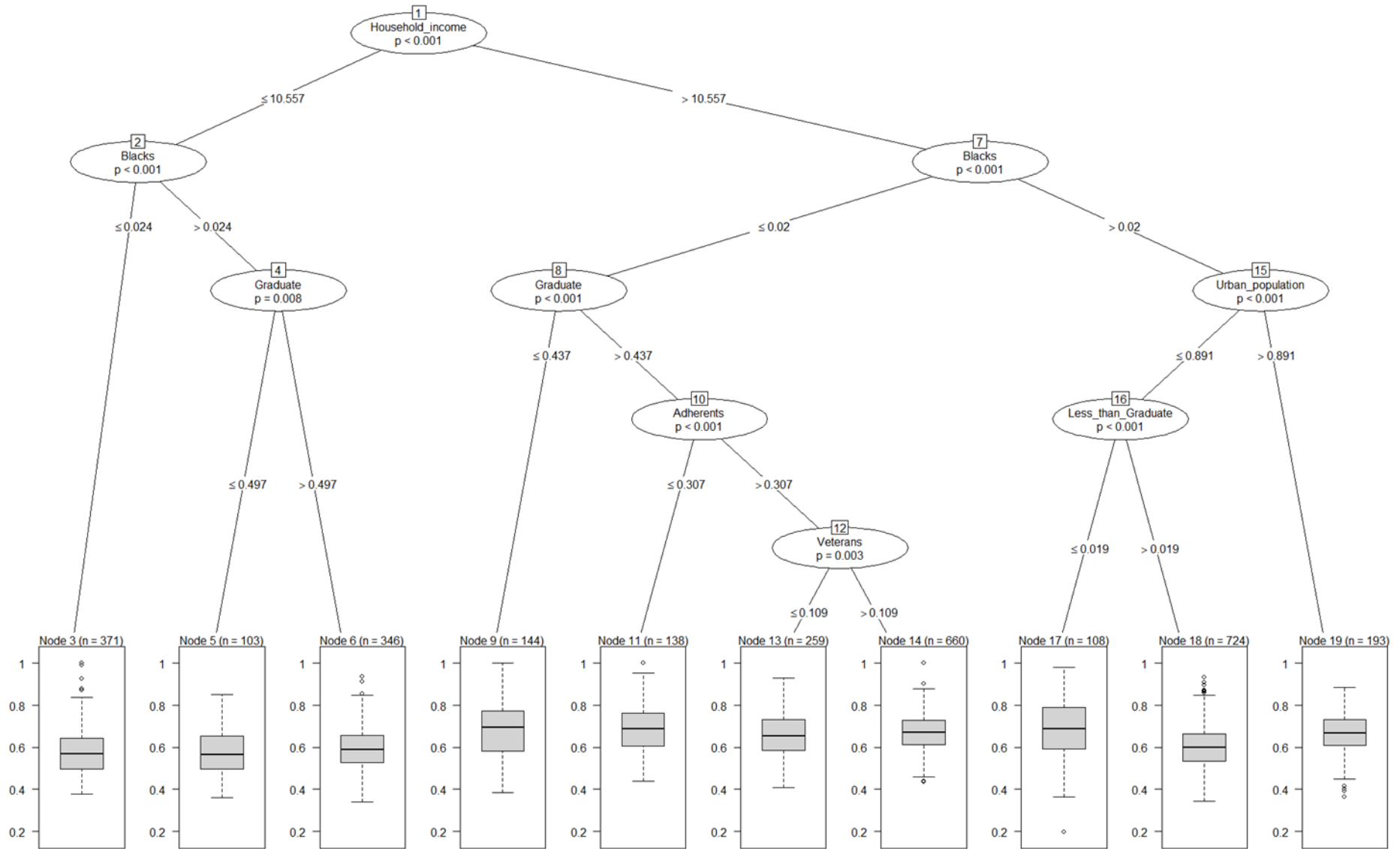


Table A2.5: spatial lag regression tree ($\mathbf{W} = 8$ k-near neighbours)[illegible]

Figure A2.6: spatial lag regression tree ($W = 10$ k-near neighbours)

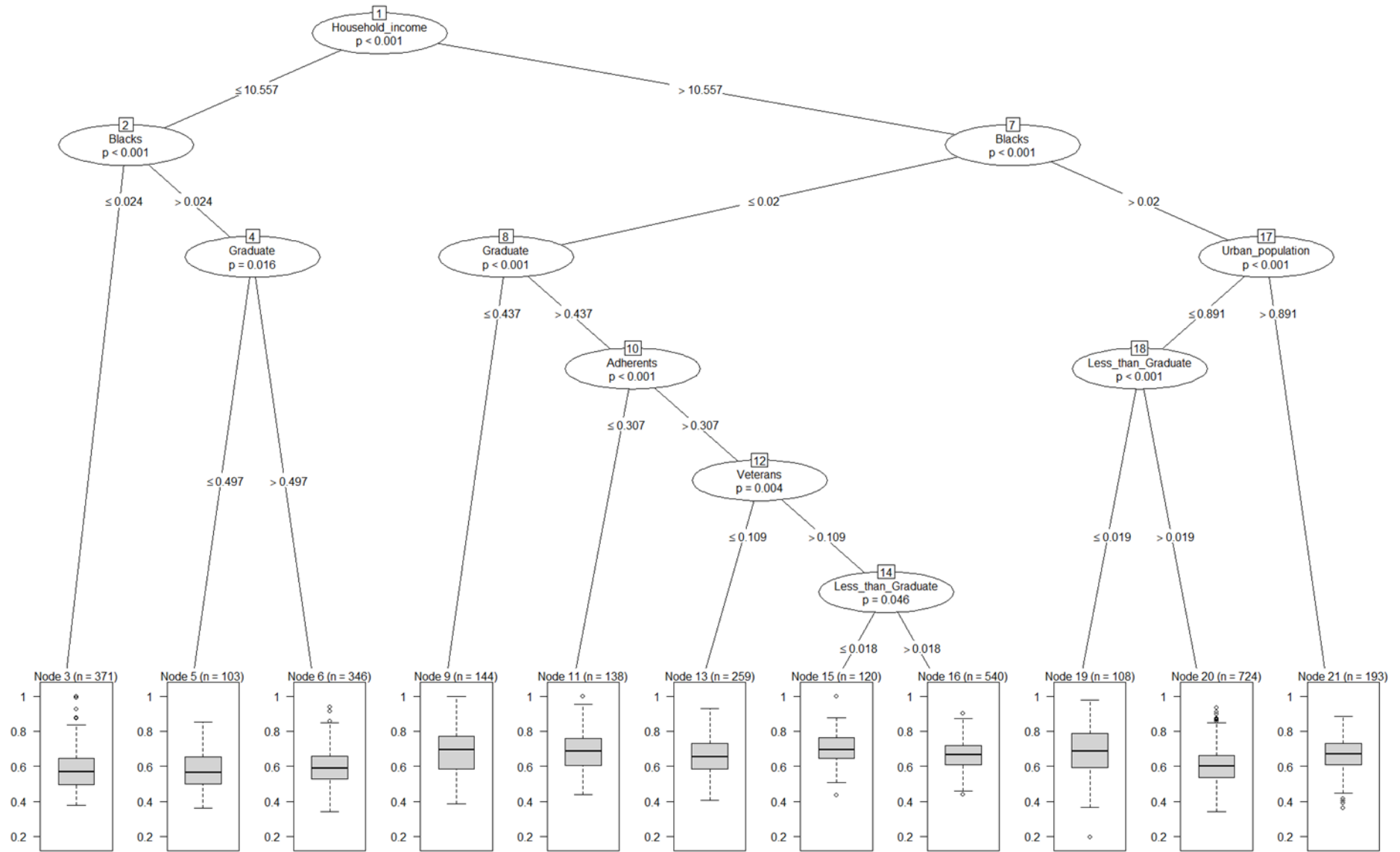


Table A2.6: spatial lag regression tree ($\mathbf{W} = 10$ k-near neighbours)[illegible]

Figure A2.7: spatial lag regression tree (W = queen)

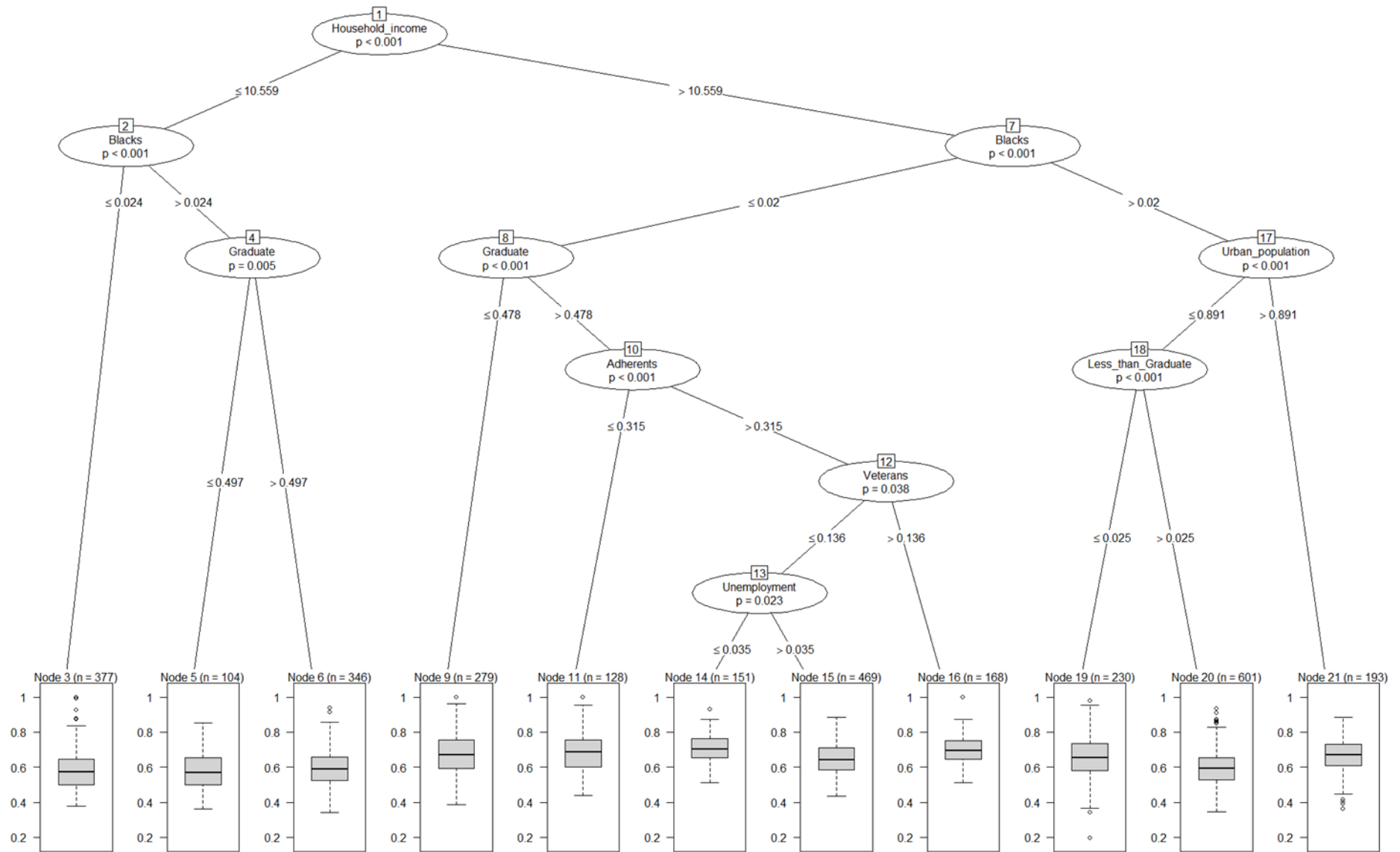


Table A2.7: spatial lag regression tree ($\mathbf{W}$ = queen)

Figure A2.8: spatial lag regression tree (W = exponential decay)

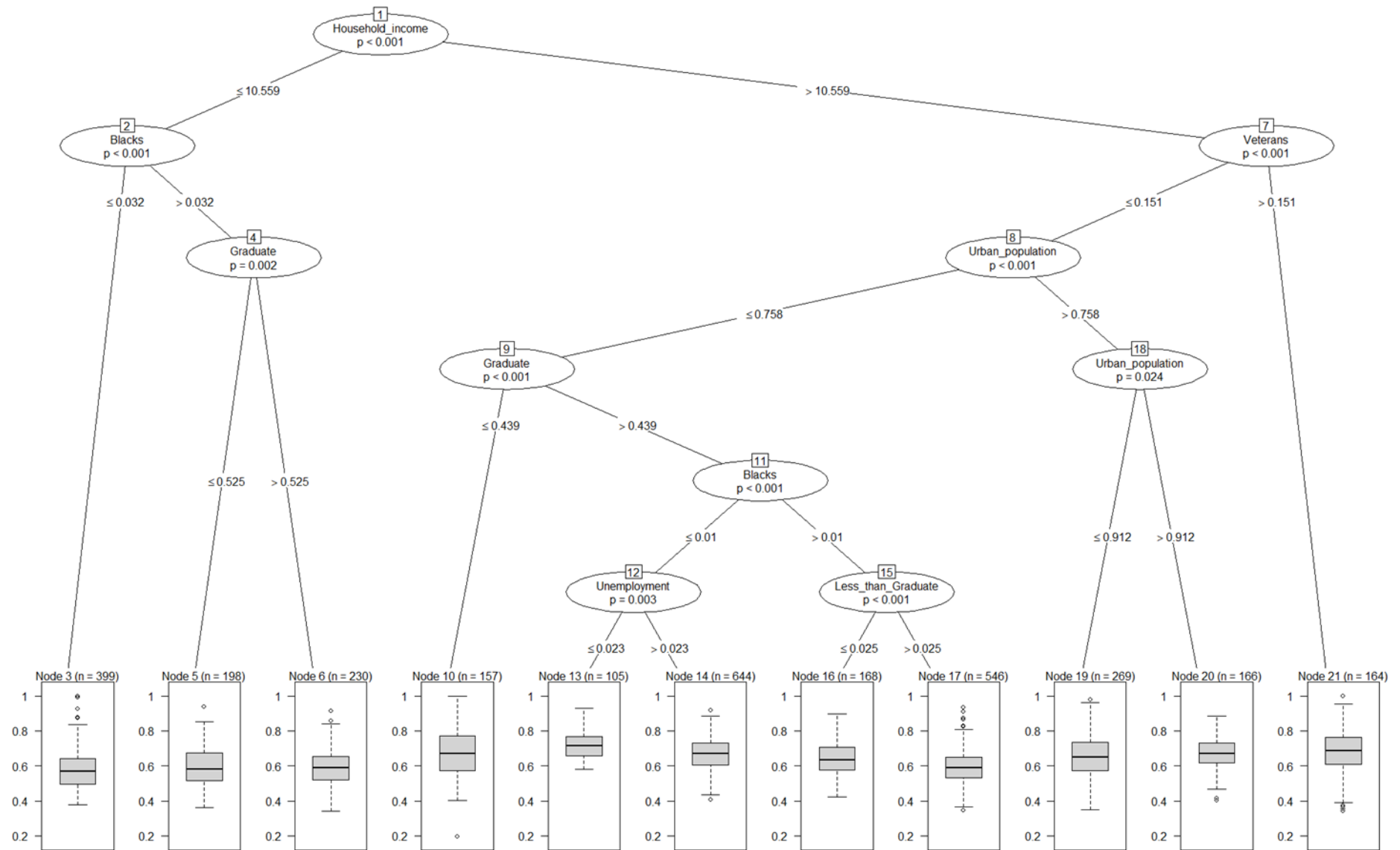


Table A2.8: spatial lag regression tree ($\mathbf{W}$ = exponential decay)[illegible]

Figure A2.9: spatial lag regression tree (W = exponential decay, cut-off = 200 km)

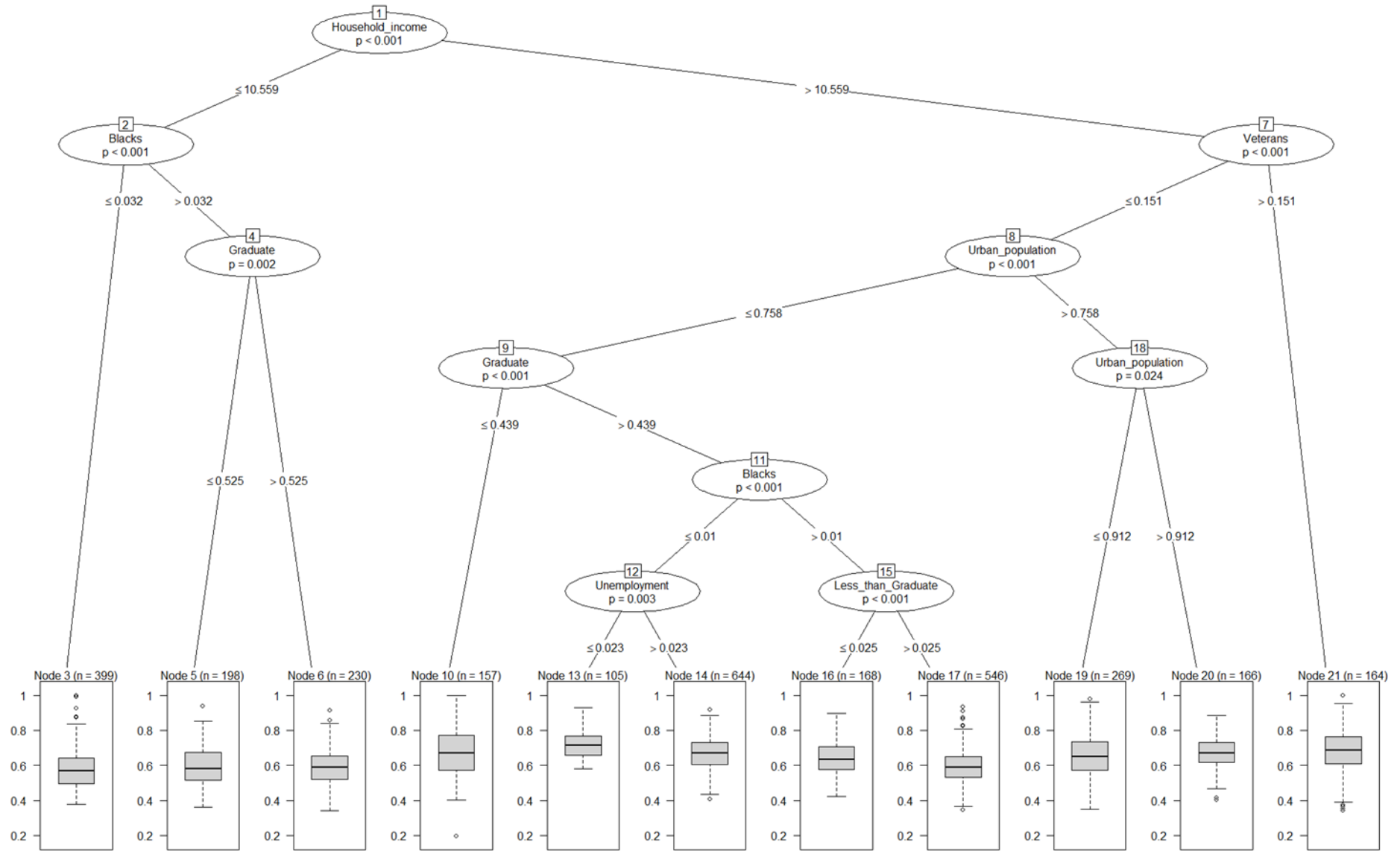


Table A2.9: spatial lag regression tree ($\mathbf{W}$ = exponential decay, cut-off = 200 km)[illegible]

Table A2.10: Structural change test (p-value)

		Urban_pop.	Veterans	Unemployment	Adherents	Household_income	Less_than_Graduate	Pct univ.	Hispanics	Blacks	Urban_pop.
OLS	Node 4										
	Node 5	0.8307	0.2183	0.9999	0.9253	0.8903	0.9998		0.4552	0.9856	0.0835
	Node 9										
	Node 10	0.9998	0.6893	0.9992	0.9998	0.4750	0.9998		0.6647	0.9973	0.9999
	Node 11										
	Node 12	0.9790	0.9124	0.9935	0.0620	0.8836	0.6352		0.3588	0.3323	0.9966
	Node 14										
	Node 17	1	1	1	0.9939	1	1		0.3430	1	0.6860
	Node 19	0.9386	0.1602	1	0.9999	1	0.9999		1	1	0.3485
	Node 20										
	Node 23										
	Node 24	0.9880	0.7586	0.0799	0.8615	0.9023	0.3482		0.1011	0.9997	0.9041
K=5	Node 25	0.7422	0.9999	0.9921	0.7906	0.8023	0.9635		0.9722	0.9537	0.9887
	Node 3	0.6099	0.9971	0.9325	0.9128	1	0.9565		0.3632	0.9994	0.9997
	Node 6	0.9810	0.8332	0.9983	0.5445	0.9982	0.9847		0.8678	0.0516	0.4457
	Node 9	0.0572	1	0.9937	0.7535	0.9998	0.9800		0.9991	0.9352	0.9369
	Node 11										
	Node 13	0.0504	0.9761	0.0122	0.0638	0.9995	0.1605		0.9998	0.7965	0.9170
	statistic	10.8812	19.4310	14.3389	16.5202	15.4782	20.1477		12.7718	27.5760	16.7932
	Node 15	1	0.9573	0.9999	0.9986	0.9998	0.9268		1	0.26501	0.9978
	Node 19										
	Node 19	0.9131	0.9132	1	0.9972	0.6097	0.9805		0.9942	0.9562	0.9713
K=6	Node 20	0.7283	0.9860	0.9966	0.7868	0.9498	0.6965		0.7608	0.9101	0.9963
	Node 3	0.6051	0.9989	0.9173	0.9106	1	0.9959		0.3396	0.9977	0.9997
	Node 5										
	Node 6	0.9846	0.9122	0.9991	0.3146	0.9968	0.9669		0.8406	0.0844	0.5473
	Node 9	0.0865	0.9999	0.9924	0.8050	0.9994	0.9909		0.9982	0.9823	0.9526

K=7	Node 11									
	Node 13									
	Node 15	1	0.9117	1	0.9952	0.9941	0.9951	1	0.1375	0.9999
	Node 16	1	0.9998	1	0.6438	0.7369	0.9969	0.9993	1	1
	Node 19	0.8097	0.9648	1	0.9956	0.7068	0.9746	0.9920	0.9737	0.9628
	Node 20	0.5389	0.9756	0.9960	0.8074	0.9898	0.7623	0.7463	0.6696	0.9990
	Node 21									
	Node 3	0.5900	0.9973	0.9048	0.8872	1	0.9951	0.2902	0.9944	0.9998
	Node 5									
	Node 6	0.9854	0.9398	0.9988	0.3152	0.9990	0.9890	0.9326	0.1271	0.5082
K=8	Node 9									
	Node 11									
	Node 13	0.9997	1	0.9970	0.9466	0.9995	1	1	0.9193	0.9819
	Node 14	0.1569	0.2120	0.0733	0.9966	0.8107	0.1047	1	0.5758	1
	Node 17	0.8072	0.9719	1	0.9972	0.7213	0.9779	0.9946	0.9612	0.9692
	Node 18	0.4831	0.9332	0.9993	0.8239	0.9975	0.7388	0.6672	0.4794	0.9998
	Node 3	0.6482	0.9984	0.9044	0.9396	1	0.9983	0.2748	0.9905	0.9998
	Node 5									
	Node 6	0.9852	0.8812	0.9997	0.2161	0.9956	0.9838	0.8083	0.1286	0.3227
	Node 9									
K=9	Node 11									
	Node 13	0.9999	1	0.9947	0.9195	0.9996	1	1	0.8855	0.9898
	Node 14	0.1354	0.2398	0.1264	0.9986	0.7587	0.1391	1	0.5379	0.9999
	Node 17									
	Node 18	0.1529	0.5088	0.1773	0.9814	0.9998	0.1202	0.2409	0.0807	0.9971
	Node 19									
	Node 3	0.5487	0.9986	0.8781	0.9683	1	0.9925	0.2768	0.9890	0.9998
	Node 5									
	Node 6	0.9862	0.7958	0.9999	0.2176	0.9817	0.9788	0.8917	0.1555	0.2460
	Node 11									
	Node 13	0.9999	1	0.9964	0.9753	0.9999	1	1	0.9281	0.9663
	Node 14	0.0798	0.2174	0.1485	0.9948	0.8822	0.1157	1	0.6175	1
	Node 18	0.9999	0.7459	0.5595	0.9998	0.8854	0.9996	0.7610	0.0561	0.8501

queen	Node 17	1	1	0.9999	0.9996	0.9976	1	1	0.9996	0.2668
	Node 18									
	Node 19									
	Node 3	0.5108	0.9965	0.9153	0.9515	1	0.9119	0.4106	0.9959	0.9996
	Node 5									
	Node 6	0.9867	0.9334	0.9996	0.2981	0.9959	0.9985	0.8995	0.0884	0.5476
	Node 9	0.0964	1	0.9784	0.8547	0.9990	0.9993	0.9980	0.9708	0.9249
	Node 10	2.62E-04	2.71E-05	0.0022	8.53E-08	0.9231	0.0117	0.2596	0.7776	0.0094
	Node 11									
	Node 14									
	Node 15	1	0.9949	0.9999	0.9998	0.9988	0.9914	0.9995	0.0725	0.9998
	Node 16									
	Node 19	0.7420	0.9540	0.9999	0.9991	0.7166	0.9752	0.9919	0.9783	0.9902
	Node 20									
	Node 21	0.5864	0.9639	0.9995	0.8182	0.9975	0.8190	0.6905	0.7280	0.9997

Note: test based on Zeileis A. (2006), Implementing a Class of Structural Change Tests: An Econometric Computing Approach. *Computational Statistics & Data Analysis*, 50, 2987–3008. Minimum size of the terminal nodes set at 100 observations. Test not performed on terminal nodes with less than 100 observations.

APPENDIX 3 - RESULTS ON CVAP

Figure A3.1: standard regression tree

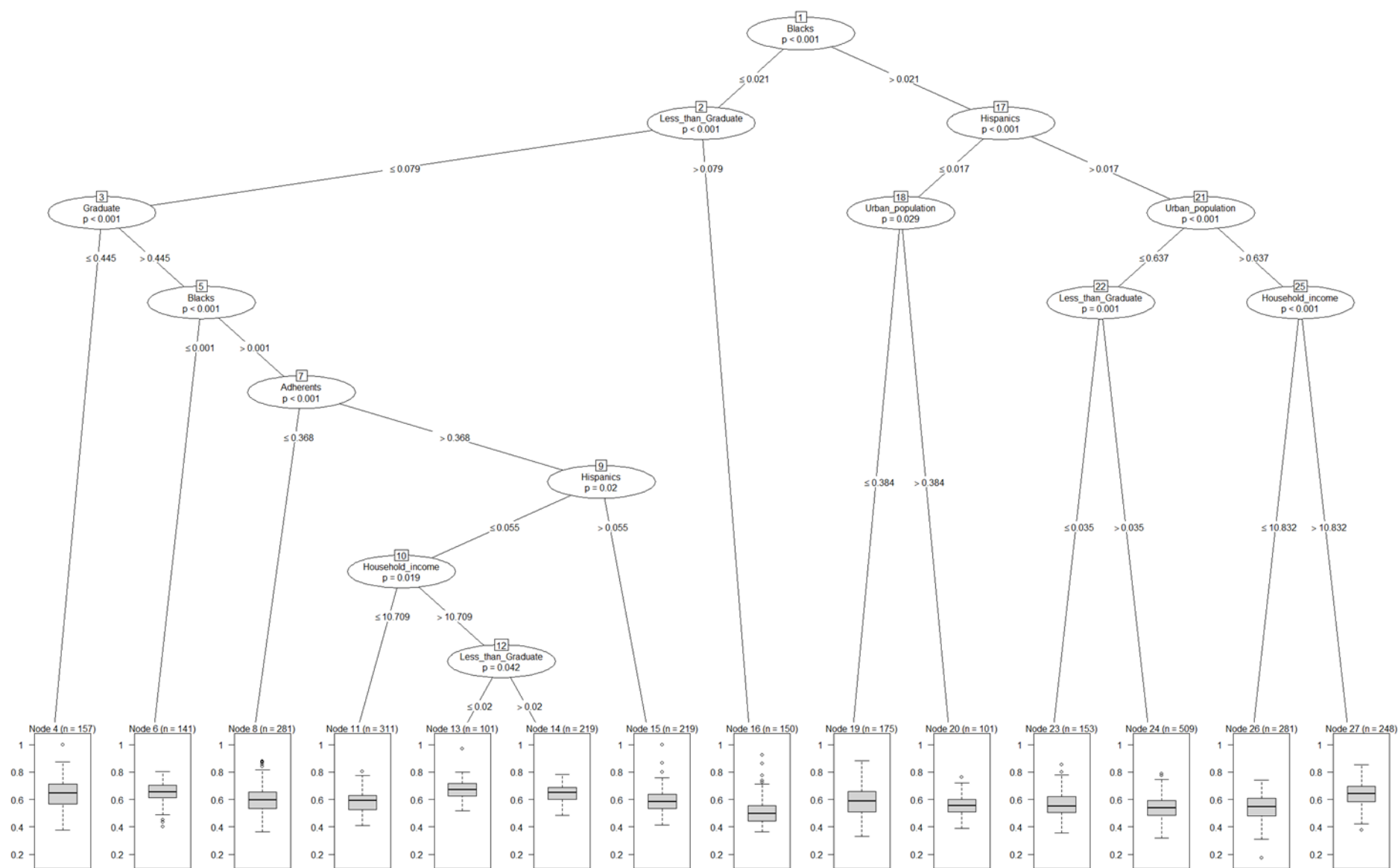


Table A3.1: Standard regression tree

[illegible]

Figure A3.2: spatial lag regression tree ($\mathbf{W}$ = 5 k-near neighbours)

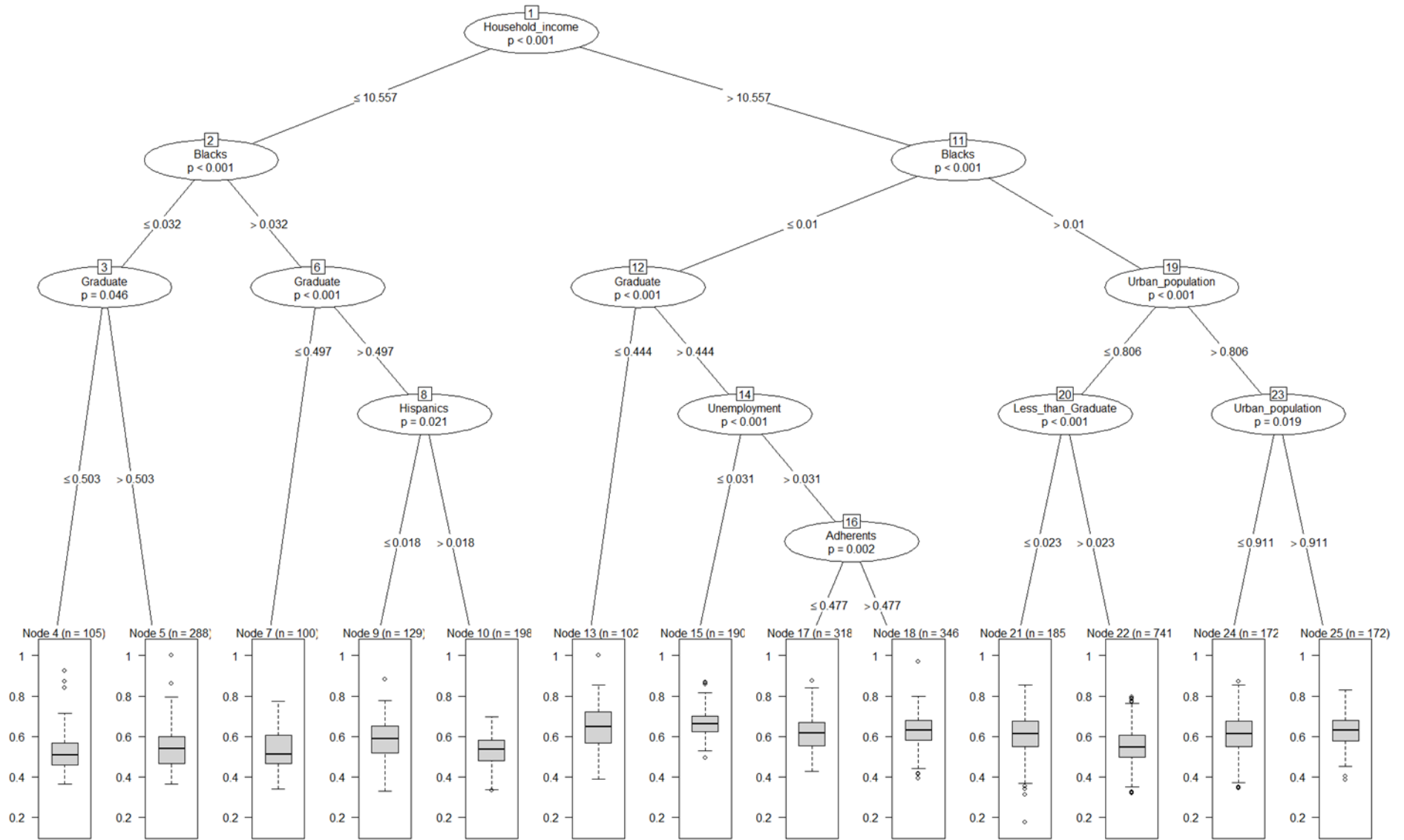


Table A3.2: spatial lag regression tree ($\mathbf{W} = 5$ k-near neighbours)[illegible]

Figure A3.3: spatial lag regression tree ($W = 6$ k-near neighbours)

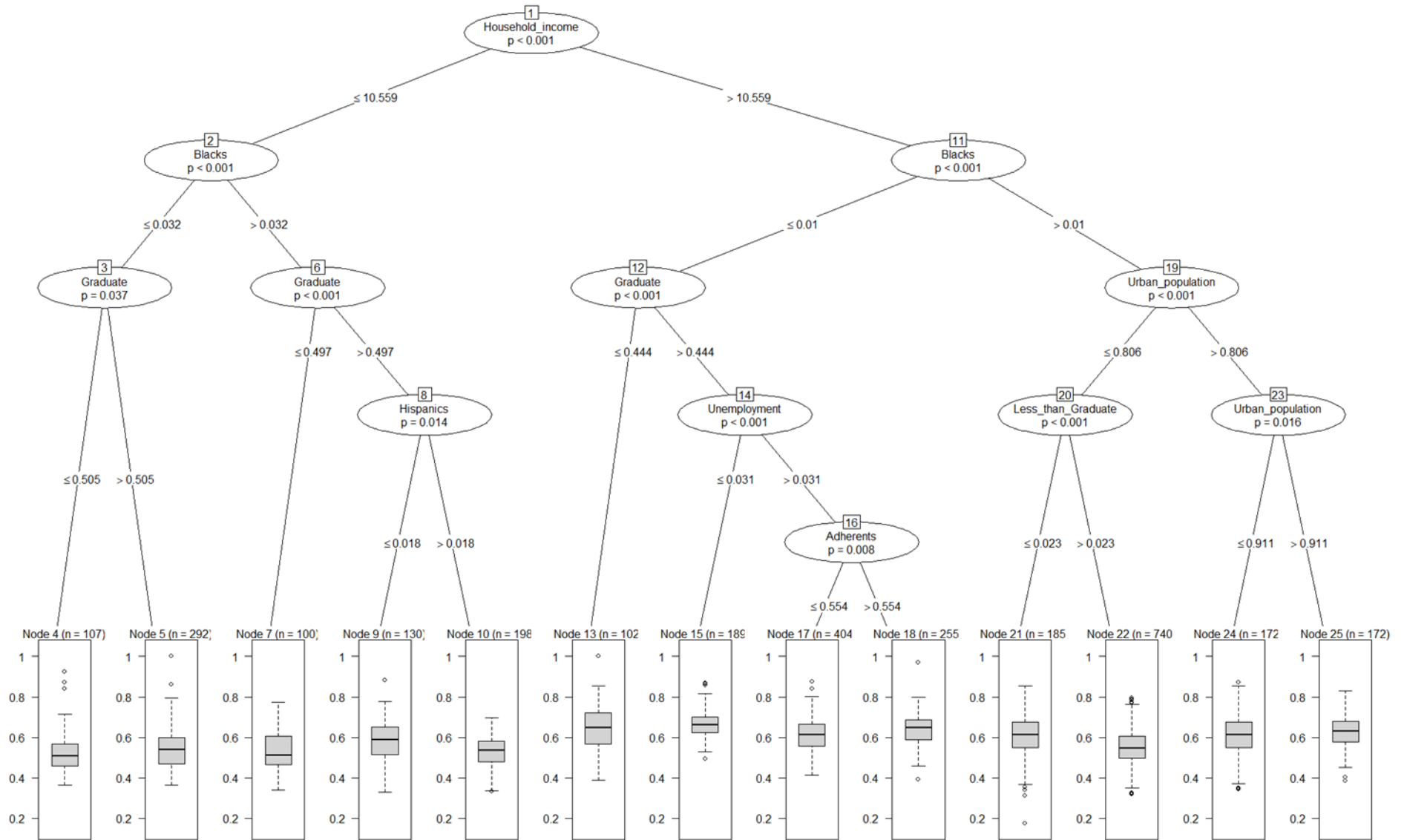


Table A3.3: spatial lag regression tree ($\mathbf{W} = 6$ k-near neighbours)[illegible]

Figure A3.4: spatial lag regression tree ($W = 7$ k-near neighbours)

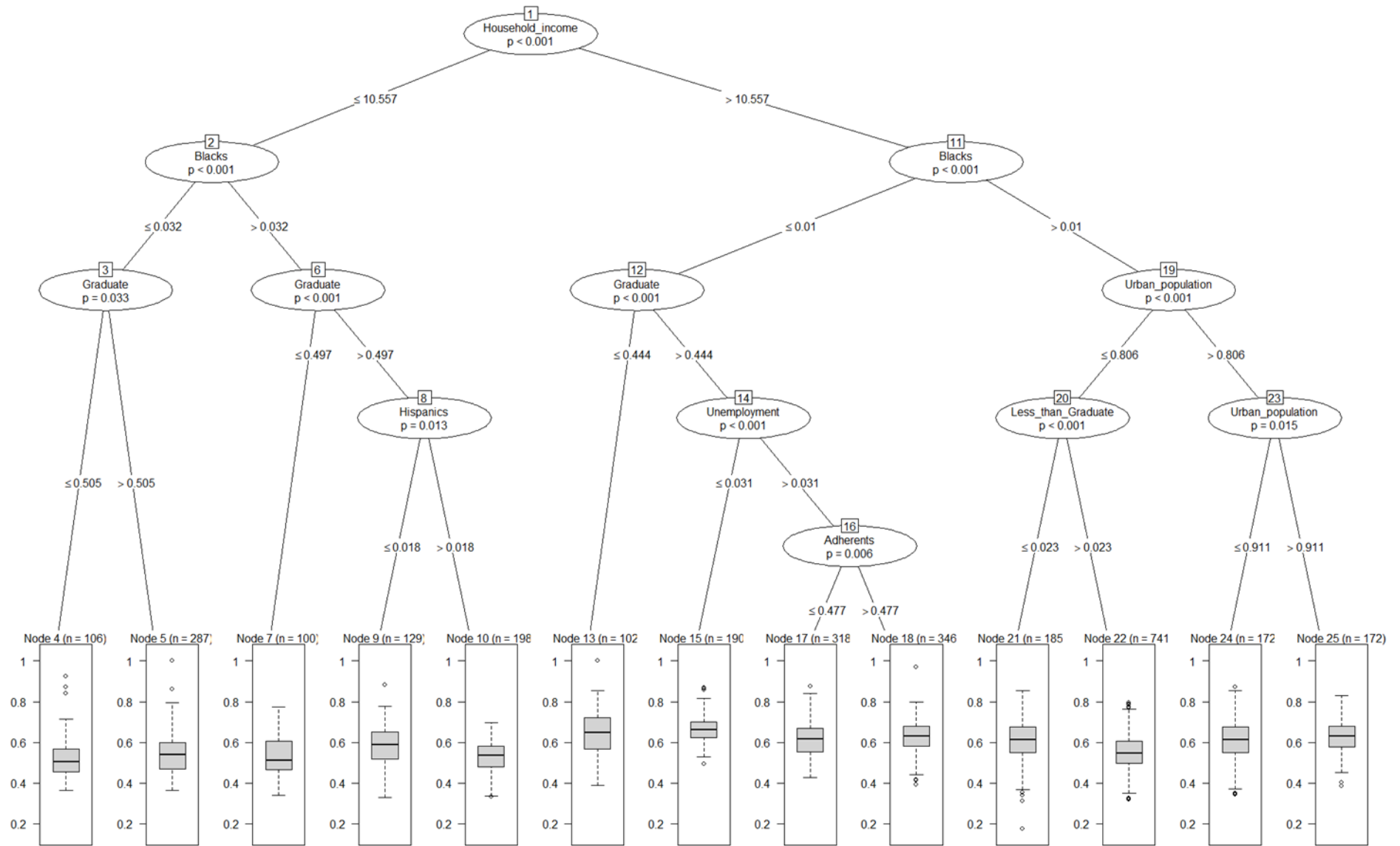


Table A3.4: spatial lag regression tree ($\mathbf{W} = 7$ k-near neighbours)[illegible]

Figure A3.5: spatial lag regression tree ($W = 8$ k-near neighbours)

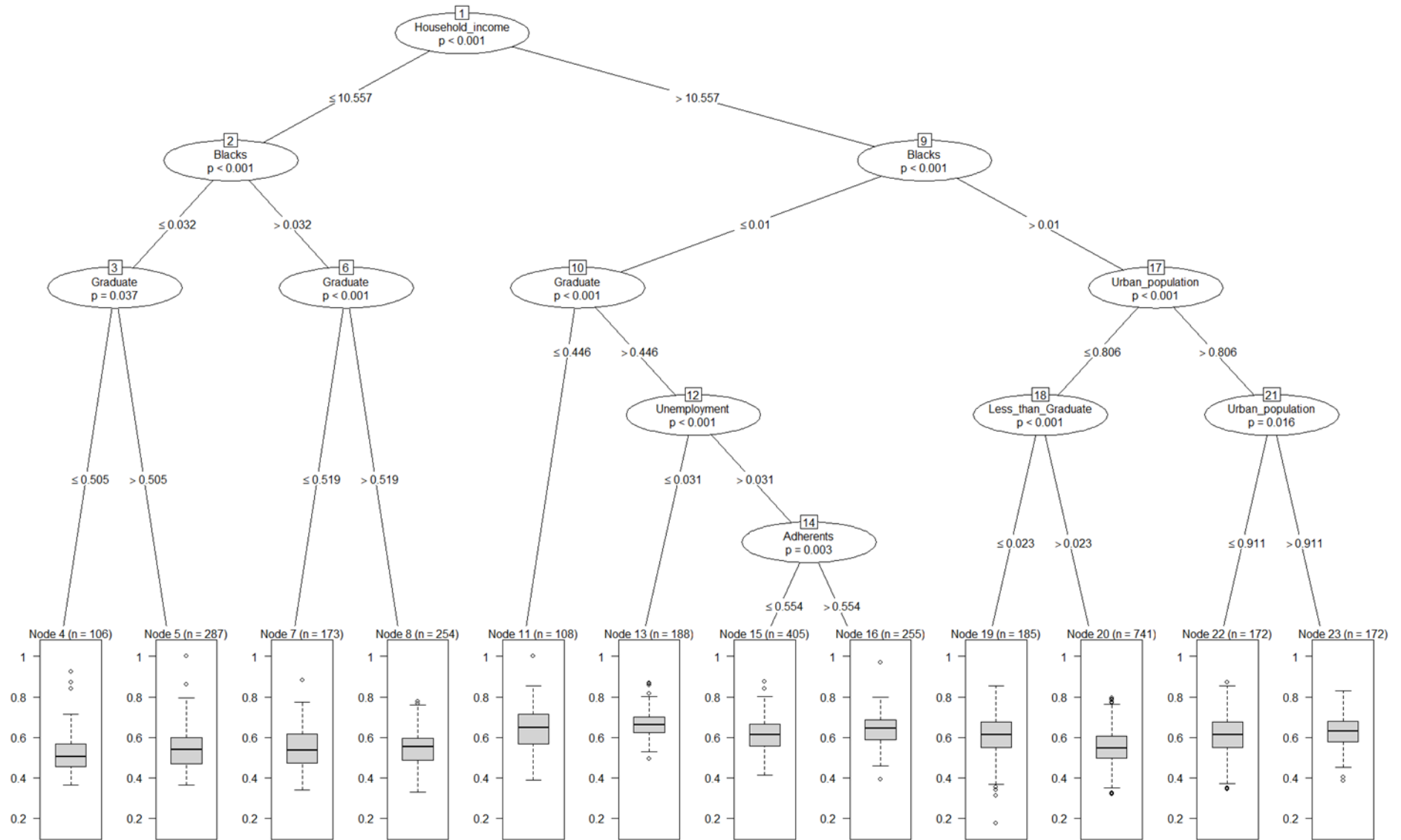


Table A3.5: spatial lag regression tree ($\mathbf{W} = 8$ k-near neighbours)[illegible]

Figure A3.6: spatial lag regression tree ($W = 9$ k-near neighbours)

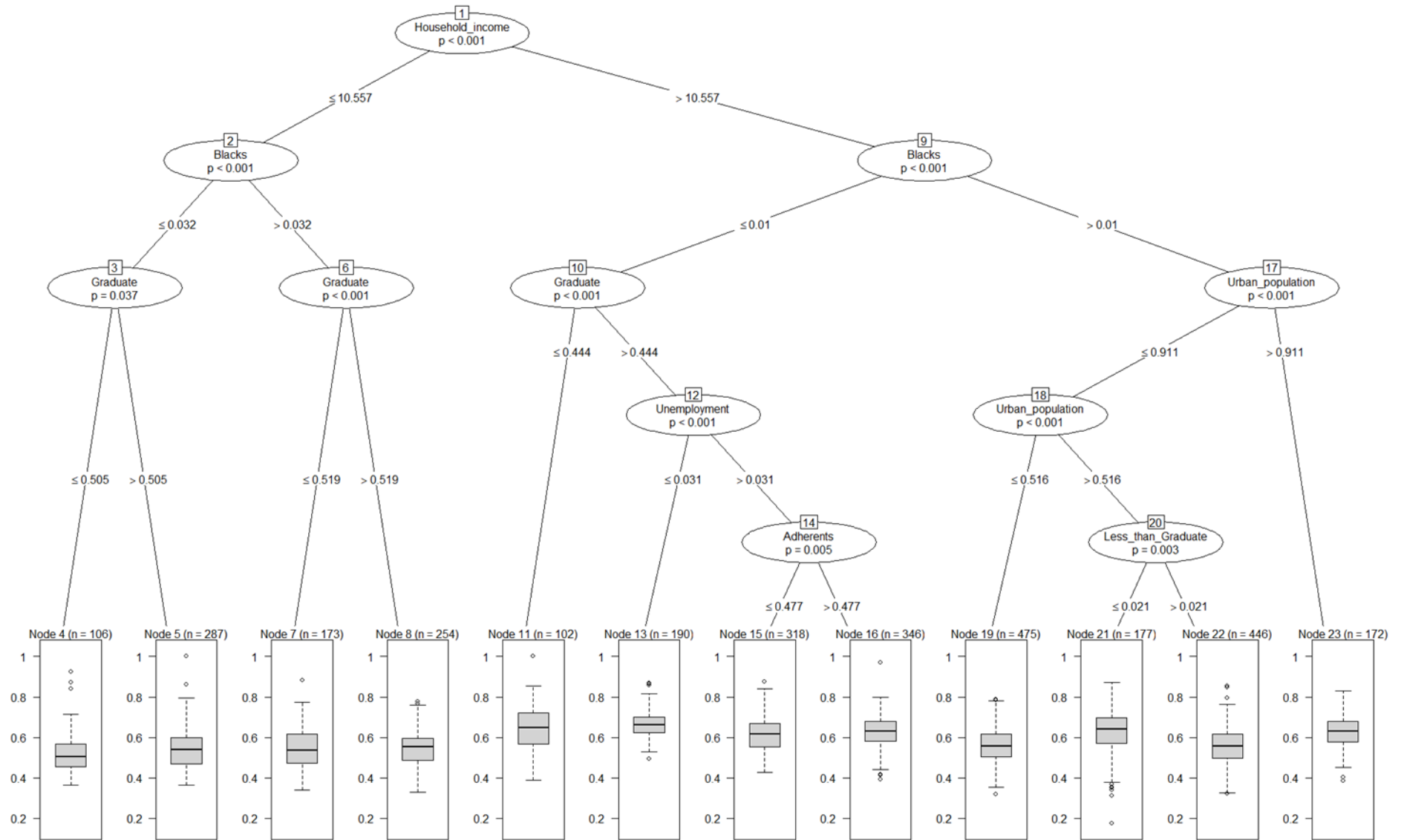


Table A3.6: spatial lag regression tree ($\mathbf{W} = 9$ k-near neighbours)

Figure A3.7: spatial lag regression tree ($W = 10$ k-near neighbours)

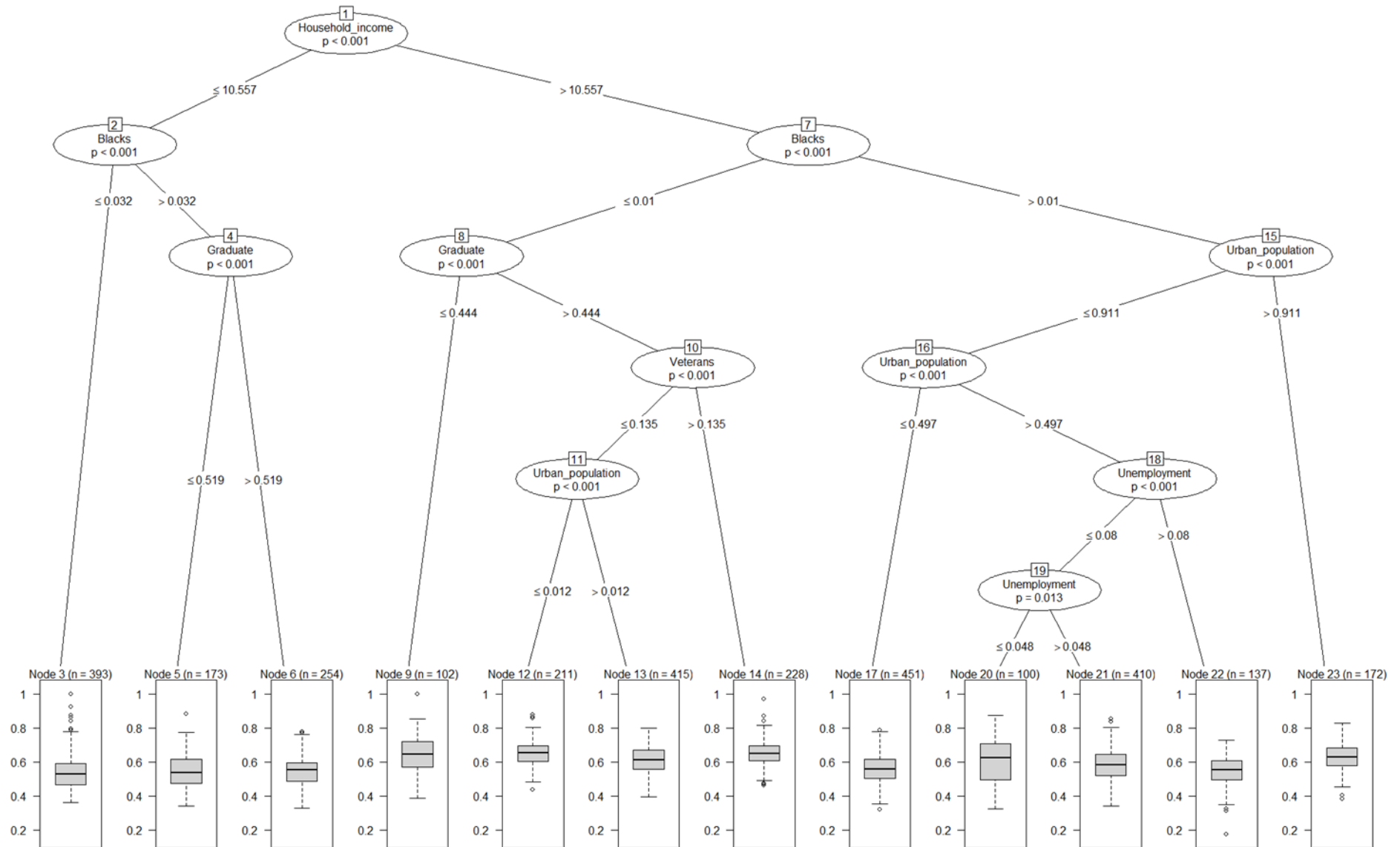


Table A3.7: spatial lag regression tree ($\mathbf{W} = 10$ k-near neighbours)[illegible]

Figure A3.8: spatial lag regression tree ($\mathbf{W}$ = exponential)

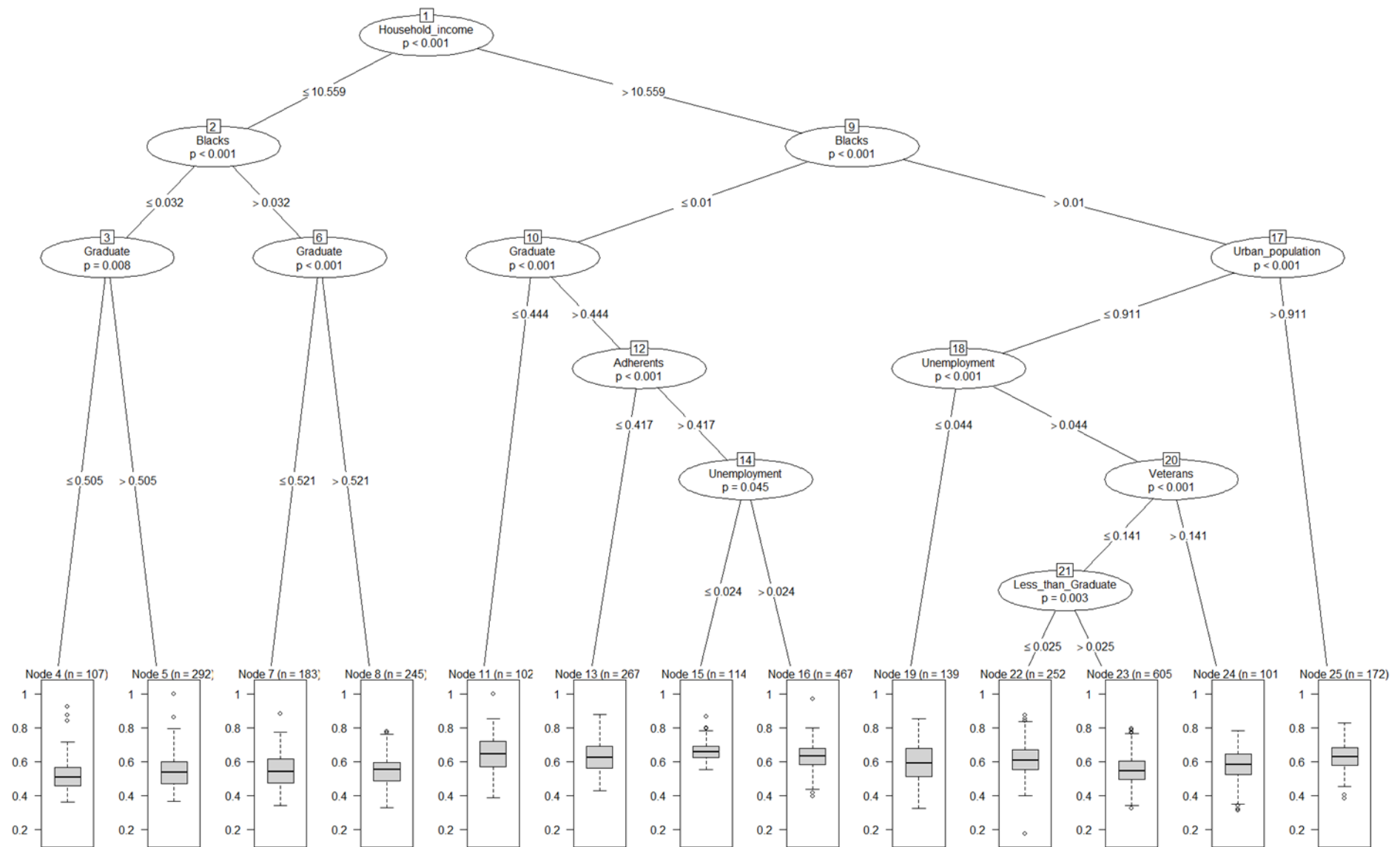


Table A3.8: spatial lag regression tree (exponential)

[illegible]

Figure A3.9: spatial lag regression tree (W = exponential, cut off 200 km)

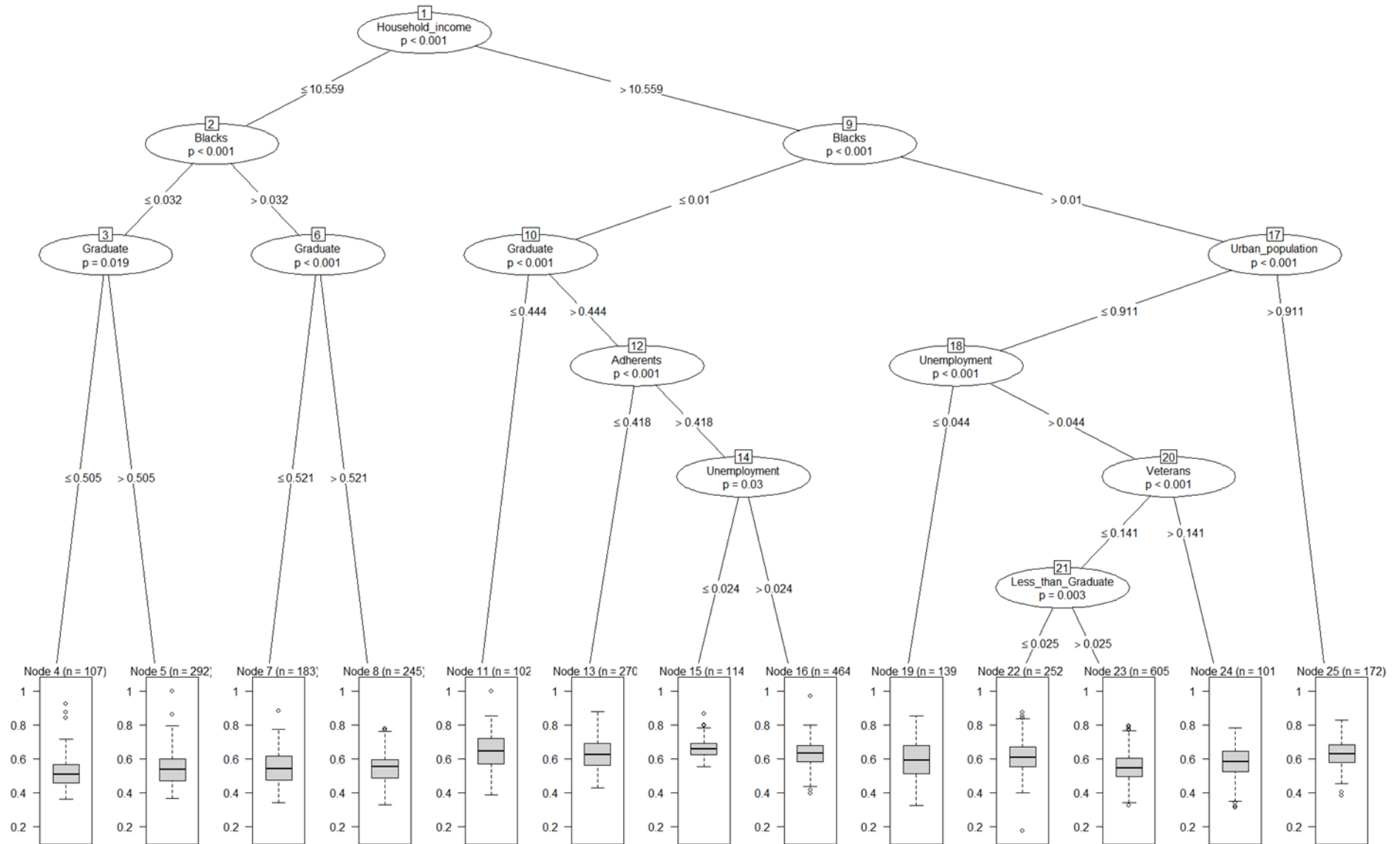


Table A3.9: spatial lag regression tree (exponential, cut off = 200)

[illegible]

Figure A3.10: spatial lag regression tree ($\mathbf{W}$ = queen)

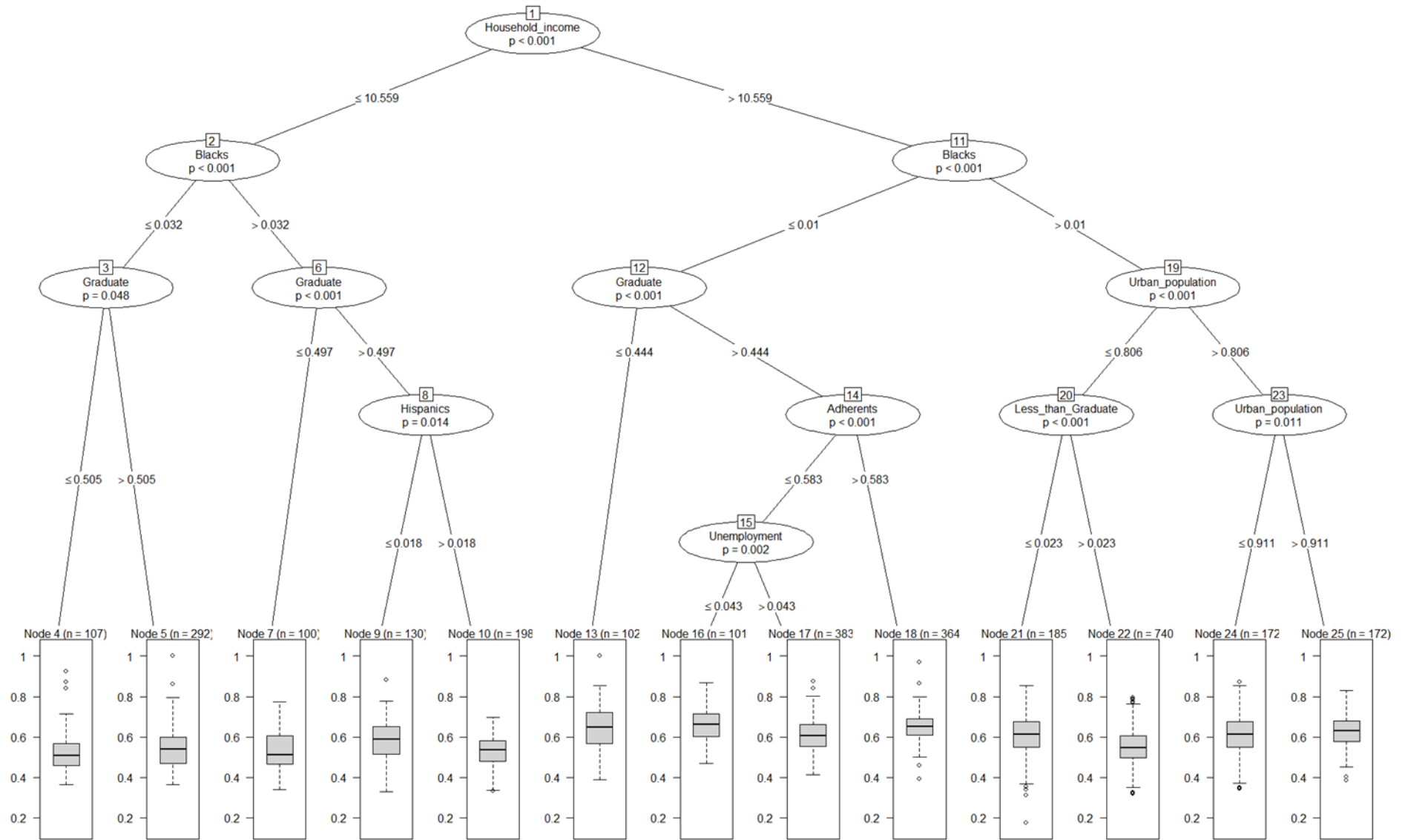


Table A3.10: spatial lag regression tree (queen)

[illegible]

Table A3.10: Structural test (p-value)

[illegible]

Node 22	1	1	0.9149	0.9970	0.9369	0.8661	1	0.9982	0.9999
Node 23	0.5035	0.9999	0.9970	0.9377	0.5270	0.9930	0.9998	0.9999	0.1465
Node 24									
Node 25									

Note: test based on Zeileis A. (2006), Implementing a Class of Structural Change Tests: An Econometric Computing Approach. *Computational Statistics & Data Analysis*, 50, 2987–3008. Minimum size of the terminal nodes set at 100 observations. Test not performed on terminal nodes with less than 100 observations.