Parametric and non-parametric analysis of tax changes

Parametric and Nonparametric Analysis of Tax Changes

James Bugden[1], Iain Fraser[1,2],
Jeffrey S. Racine[3], and Robert Waschik[1]

[1]*La Trobe University, Melbourne, Australia,*
[2]*University of Kent, UK and*
[3]*McMaster University, Hamilton, Ontario, Canada*

Abstract

In this paper we examine the net effect of several major tax changes in Australia on residential property prices. Specifically, we consider the announcement and introduction effects that resulted from several policy changes including the introduction of the Goods and Services Tax (GST) and the accompanying First Home Owner Grant (FHOG). Using a large data set of residential property sales in Melbourne, Australia, between 1992 and 2002 we estimate various models using parametric and nonparametric methods. While our parametric models suggest that the tax policy changes appear to have a *statistically* significant impact on house prices, no *economically* significant impact is detected by our nonparametric models, nor (upon closer inspection) by the parametric models themselves. Given the enormity of the sample size, this provides a telling example of the fundamental difference between statistical and economic significance and its implications for detecting government policy effectiveness.

J.E.L. Classification Codes: R21, C14
Keywords: Tax changes, housing, nonparametric methods, hedonic model

* We would like to thank Landata of the Department of Sustainability and the Environment, Victoria, Australia for providing much of the data used in this study. Fraser and Waschik gratefully acknowledge support from the Faculty of Law and Management Small ARC Grant #104203. Racine gratefully acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC:www.nserc.ca), the Social Sciences and Humanities Research Council of Canada (SSHRC:www.sshrc.ca), and the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca).

1 Introduction

Like many countries Australia experienced rapid growth in residential property values from the mid 1990s. A major concern that accompanied this growth was the affordability of property, especially for first time home buyers. In the midst of this increase in property prices, in 2000 the Australian Commonwealth Government introduced significant changes to the tax system that resulted in the introduction of the Goods and Services Tax (GST). The introduction of the GST and its ensuing effects on economic conditions in general and the housing marking in particular prompted both State and Commonwealth governments to introduce the First Home Owner Grant (FHOG). The FHOG provided a grant of $7,000 (all currencies are in Australian dollars) to people buying their first home, and an additional $7000 grant for the construction or purchase of a new home. The objective of this paper is to examine the impact on housing prices of the announcement and implementation of these tax changes.

As Leung (2003) notes there are two obvious reasons why property is taxed. First, the value of the housing stock is large, and second taxes cannot typically be avoided as property is durable and immobile. But many tax regimes favour home ownership which in turn leads to resource distortions in the economy and net welfare losses. So why do we then observe tax regimes that favour home ownership? Two reasons are frequently cited. First, preferential taxation of housing is politically sensible, as homeowners vote, so politicians tend to favour homeowners with lower property taxes. Second, there are positive externalities associated with the social benefits of house ownership (Glaeser and Sacerdote, 2000).

To date a number of papers have examined various impacts of the tax changes on housing in Australia (e.g., Freebairn, 1999, Productivity Commission, 2004, Guest, 2005, and Wood *et al.*, 2006). For example, Freebairn (1999) uses a partial equilibrium model, Guest (2005) uses a life cycle model, and Wood *et al.* (2006) employ a micro-simulation model. Within this literature a number of issues have been examined with respect to the objectives of the FHOG and GST. For example, several studies estimated the impact of the GST on the price of a new home. Forecasts of the increase in new house prices due to imposition of the GST range from 5.1% (Freebairn 1999) to 5.9% (PC 2004). In terms of the FHOG most research assessed the extent to which it was capitalized into existing house prices such that the benefit which first time buyers expected to derive from the FHOG did not materialize. For example, Freebairn (1999) argued that because of the relatively inelastic supply of housing, the grant was capitalized into higher prices. In contrast, Wood *et al.* (2006) assumed that the housing market was subject to an infinitely elastic supply of housing, so that the introduction of the GST and FHOG had no effect on house prices. Therefore, it is difficult to *a priori* know or predict from the literature how the various tax changes introduced in Australia impacted house prices. In addition, within the literature there has been no consideration or analysis of the response of households to the announcement or the timing of the actual policy changes. We know from the wider economic literature that policy announcements are important. For example, there is a large literature examining the impact of monetary policy announcements eg., Demiralp and Jorda (2004) and Bernanke and Kuttner (2005). We consider this issue in our analysis.

In this paper to improve our understanding of the impact of the tax changes we take a different approach by employing a unique data set that contains details of all residential property sales (i.e., more than 750,000 observations) in metropolitan Melbourne between 1992 and 2002. The size of this data set and the fact that it constitutes not a sample but the population of sales provides us with a justification for employing nonparametric methods. As is typical of data on property sales, the data are mixed in type, hence are ideally suited to the nonparametric regression methods of Racine and Li (2004). Another benefit of the nonparametric approach is that we are estimating a hedonic model that cannot be derived precisely from economic theory. By its very nature the appropriate specification is ultimately an empirical issue. We also note that as data sets increase in size the expected performance of nonparametric methods increases significantly (Kuminoff *et al.,* 2009).

Despite the magnitude of our data set we need to acknowledge several limitations. Specifically, while our data set includes complete location descriptors for each sale and a descriptor on *land use code* which identifies the building as an apartment, townhouse, detached or semi-detached house, it does not contain other descriptive characteristics about the houses sold, such as interior dimensions, or the number of bedrooms or bathrooms. While this data limitation will reduce the explanatory power of our analysis, it should not affect the ability of the model to explain the effect on house prices of the various tax policies in which we are interested. The GST is applied to all houses independent of their characteristics. Also, while the FHOG is only available to first home buyers, the omission of other variables describing characteristics of houses will only lead to problems if first home buyers purchase houses with particular characteristics. Since the FHOG is available to all first home buyers regardless of the home purchased, this is unlikely to be a problem.

Overall our analysis reveals that using a parametric specification indicates that the tax changes were statistically significant. In contrast we show with our nonparametric method that the tax change did not induce any variation in house prices. This allows us to conclude that the FHOG effectively served the government's objective of compensating first-time homeowners for the effects of the introduction of the GST, while not having any appreciable effect on house prices. Thus, our analysis highlights the important contribution made by using nonparametric estimation methods, where appropriate. In fact, these results ultimately provide an important illustration of the hazards of significance tests in the presence of very large data sets.

The structure of the paper is as follows. In Section 2 we briefly describe the various tax changes that impacted the Australian housing market. This is followed by a detailed description of the data used in the paper. Section 4 describes the hedonic model developed and the nonparametric methods employed. Results are presented in Section 5, and Section 6 concludes.

2 Tax Changes and the Australian Housing Market

By the late 1990's it was apparent that the Australian taxation system required fundamental reform. The GST passed Parliament on 29 June 1999, which we consider to be the formal announcement of the tax change. The GST was actually introduced in July 2000 as part of a major tax reform package agreed by the Commonwealth Government and all State Governments. The 10% GST was applied on the supply of almost all goods and services. It had an immediate impact on the housing sector as it covered building materials and services used to build new houses. In addition, the tax also applied to a number of important property management activities as well as selling and conveyancing expenses. The impact of the GST on housing activity was significant with estimates of at least $5 billion raised in tax receipts in 2002-03 (Productivity Commission, 2004).

To counteract the effect of the introduction of the GST on the price of new homes, the government introduced the FHOG. The grant was intended to compensate first-time buyers due to the increase in costs resulting from the introduction of the GST. Details of the FHOG were announced on 1 January 2000 and it was available from July 2000, with eligible applicants entitled to a one-off $7,000 payment. The FHOG was available to first-time purchasers of a new or existing home. In addition, from March 2001 the Commonwealth Government provided an additional $7,000 Commonwealth Additional Grant (CAG) to first-home buyers for the purchase of a new, previously unoccupied house, or a new build. The CAG was reduced from $7,000 to $3,000 June 2002, but the original $7,000 FHOG remained in place for eligible buyers. The FHOG is not means tested. Since it was introduced, the FHOG added \$4.3 billion into the market with over 550,000 grants awarded.

Another important tax in the Australian housing market was the Capital Gains Tax (CGT). The Productivity Commission (2004) note the importance of changes to the CGT in 1999. Prior to September 1999, the CGT was levied on the full estimated capital gains upon sale of an asset. After this date, only 50% of capital gains was subject to CGT increasing investment in residential property for renting.

3 Data

The primary dataset used in this study is the Property Sales Information (PRISM) database from the Victorian Department of Sustainability and Environment. It records all property sales in Greater Melbourne on a daily basis from January 1992 to August 2002, for a total of 950,317 records. For sales after 1 July 2000, price includes the GST. The dataset is neither time-series nor panel data. Every day has multiple sales and many properties have multiple sale dates. Each record in the PRISM dataset includes information on the date of sale, the selling price, and variables describing the exact location of the property. In terms of location we employ use a categorical measure for Municipality. There are 31 of these within the greater Melbourne area. In addition, we also assess location by employing the distance of a suburb from the Melbourne Central Business District (CBD). For each residential property sale, distance was calculated as the distance between the suburb/postcode in which a sale takes place and the Melbourne CBD. *Ceteris paribus* we would expect that homes closer to the CBD would be more expensive, though the relationship between house prices and distance will not necessarily be linear. Finally, we include the mean taxable income associated with the suburb in which a sale takes place.

The PRISM dataset provided one more variable of interest: LUC (Land Use Code). The LUC classifies each property by type, whether residential, rural, commercial, or industrial. Only a small minority of property transactions in the PRISM dataset do not involve residential properties, and of course, because of the focus of this study on the effects of the GST and FHOG (the latter of which only applies to residential properties), we only consider property transactions involving residential properties. The set of LUCs drawn from the PRISM dataset that we use are given in Table 1.

Since house prices have been shown to vary inversely with real mortgage rates we also include the Reserve Bank of Australia's standard variable housing loan rate as an explanatory

Table 1: Description of Residential Land Use Codes

LUC Number	Description	% in PRISM
LUC 10	House (new - detached)	0.5
LUC 11	House (previously occupied)	56.3
LUC 12	Terrace (attached house)	0.3
LUC 13	Dual Occupancy	0.2
LUC 14	Flat/unit/apartment (Multi-storey)	20.8
LUC 15	Townhouse (unit)	2.4
LUC 16	Flat/unit/apartment (retirement)	0.2

variable. House prices, mean taxable income and the mortgage rate were converted from nominal to real using quarterly CPI for Melbourne. As a consequence, despite having daily nominal prices and monthly interest rates, data precision is quarterly.

We note that whilst employing a quarterly time frame with our data reduces data frequency, we assume that any price effects that we might observe as a result of the tax changes will be long lived (ie., greater than a quarter). For example, the reduction in the CGT remained in effect throughout the remainder of the sample time period. As a result we would expect the tax policy changes considered to cause a step change in prices, not a sudden price spike.

Finally, in order to isolate the effects of the announcement and implementation of the relevant tax policies described in Section 2, dummy variables have been included to correspond to changes in the CGT, the GST and the FHOG. A summary of all of the data used in the analysis is given in Table 2 - detailed descriptions of data sources and transformations are provided in the Appendix.

Table 2: Variables Used in Model Estimation

Dependent Variable		Mean	Std Dev
rprice	Real house price ($ Australia)	157471	129669
Continuous Independent Variables			
rincome	Annual real mean taxable income by postcode ($ Australia)	28252	6419
r	Real mortgage rate (Australian)	6.32	2.22
dist	Distance to Melbourne GPO (Kms)	20.7	15.3
t	Quarterly Time trend		
Categorical Independent Variables			
LUC	Land Use Code		
MUN	Municipality		
Dummy Variables		= 1 from ...	
CGT	Dummy variable for Capital Gains Tax Reduction	21/09/1999 − ...	
GSTANN	Dummy variable for announcement of GST	29/06/1999 − 30/06/2000	
FHOGANN	Dummy variable for announcement of FHOG	1/01/2000 − 30/06/2000	
GSTFHOG	Dummy variable for FHOG and GST introduction	1/07/2000 − ...	
CAG	Dummy variable for CAG introduction	9/03/2001 − 30/06/2002	

All data are from PRISM except rincome, sourced from Australian Tax Office Taxation Statistics, and dist, sourced from McGinley datafile (RMIT) and Geoscience Australia, and announcement of CAG from Victorian State Revenue Office.

4 Hedonic Models and Nonparametric Methods

Hedonic models typically express the price p of a good i, using an equation like (1).

$$p_i = f(x_i, z_i) \tag{1}$$

Explanatory variables are typically grouped into two vectors. z comprises discrete variables including dummies or categorical variables like whether the housing unit is an apartment or a detached house, while x captures continuous or non-dichotomous variables such as interest rates and incomes.

Among the most important problems identified in estimating an equation like (1) are (i) the lack of guidance on the appropriate functional form for the function $f(\cdot)$, and (ii) mechanisms used to deal with estimation in the presence of categorical explanatory variables z_i.

The simplest solution to (i) involves OLS or maximum likelihood estimation of a linear version:

$$p_i = x_i\beta + z_i\delta + u_i \tag{2}$$

where u_i is an error term assumed to have zero mean and finite variance. It is important to note that this specification typically allows only the model's intercept to shift across the categorical variables. In principle, we should allow all parameters in the model to vary over all realizations of the categorical variables. In the econometric literature it has been shown how estimation of (1) using nonparametric estimation techniques which do not impose a specific functional form upon some elements of $f(\cdot)$ can yield models that are closer to the true data generating process than common parametric specifications found in the literature, particularly when employing a large dataset.

Nonparametric methods are one approach to local modeling (see Racine and Ullah (2006) and Li and Racine (2007) for details). The attractiveness of nonparametric methods is that they allow the data to determine the appropriate model for the data being examined. Using the generalized kernel estimation methods introduced by Li and Racine (2004), we are able to incorporate all of our discrete data as well as the continuous data in a fully nonparametric regression framework. Unlike fully parametric and semiparametric models, this fully nonparametric estimator does not restrict the nature of the relationship among house prices and any of the regressors, whether categorical or continuous. By implementing the nonparametric methods of Racine and Li (2004), not only can our approach accommodate all data types, but we are more likely to capture the inherent non-linearities.

The key feature of nonparametric methods is that they employ local averaging, where the averaging is achieved by the use of kernel functions (a 'kernel' is simply a weighting function). That is, they compute a consistent estimate of a conditional mean by locally averaging dependent variable values that are "close" in terms of their regressors. Specifically, we can define our nonparametric regression model as follows:

$$p_i = h(x_i, z_i) + u_i, \tag{3}$$

where $x_i = (x_{1i}, \ldots, x_{Si})$ and $z_i = (z_{1i}, \ldots, z_{Ri})$ are the set of continuous and discrete variables employed in the analysis. We impose no *a priori* functional form on $h(\cdot)$ allowing it to be flexible. Finally, we assume that u_i has mean zero and variance $var(u_i|x_i) = \sigma^2(x_i)$.

To estimate $h(\cdot)$ we employ the kernel methods developed by Li and Racine (2004). We begin by deriving the kernel function for our continuous variables. Assume that $k(\cdot)$ is the univariate kernel function such that the product kernel can be expressed as:

$$K(x_i, x, h) = \prod_{s=1}^{S} h_s^{-1} k((x_{si} - x_s)/h_s) \tag{4}$$

where x_{si} and x_s are the sth components of x_i and x, and h_s is the smoothing parameter for each continuous variable. To implement our kernel function we employ a standard normal kernel function:

$$k(v) = e^{-v^{2}/2}/\sqrt{2\pi} \tag{5}$$

Next we summarize the kernel function for a discrete regressor z_r introduced by Li and Racine (2004). The kernel function is defined as:

$$l(z_{ri}, z_r, \lambda_r) = 1 \text{ if } z_{ri} = z_r, \text{ and } \lambda_r \text{ otherwise}, \tag{6}$$

where z_{ri} and z_r are rth element.

Next we define the smoothing parameter λ_r which is assumed to lie in the unit interval $[0, 1]$. When $\lambda_r = 0$, the kernel function becomes an indicator function, and when $\lambda_r = 1$, the kernel function is a constant for all values of the discrete variables. This effectively pools all data with respect to this variable.

Following Li and Racine (2004) we can then write the product kernel for our discrete variables as:

$$L(z_i, z, \lambda) \;=\; \prod_{r=1}^{R} l(z_{ri}, z_r, \lambda_r). \tag{7}$$

Given equations (4) and (7) we can employ a local linear kernel estimator to estimate $h(x_i, z_i)$.

The local linear method is based on the following minimization problem:

$$\min_{\{a,b\}} \sum_{i=1}^{n} \left(p_i - a - (x_i - x)'b\right)^2 K(x_i, x, h) \times L(z_i, z, \lambda). \tag{8}$$

The local linear estimates of the conditional mean and derivative at the point (x, z) are given by $\hat{h}(x, z) = \hat{a}$ and $\hat{\beta}(x, z) = \hat{b}$, respectively, where $\hat{a}$ and $\hat{b}$ are the solutions to (8). See Li and Racine (2004) for details.

An important aspect of the estimation process is the choice we make regarding the smoothing parameters (h, λ). In this paper we select the appropriate bandwidths for our discrete and continuous variables by employing the Expected Kullback-Leibler Cross Validation method of Hurvich, Simonoff and Tsai (1998). We explicitly examine the smoothing property of this estimation method and how it affects our results in Section 5.2.

By employing cross validation for both discrete and continuous variables, if any variable is found to be stochastically unrelated to p_i then it is 'smoothed out', i.e., removed automatically from the resulting estimate. This is an important feature of the estimation process that makes this form of nonparametric estimation more efficient than other forms of nonparametric estimation (Li and Racine, 2004).

Finally, given the above nonparametric model, we will present our regression results using partial regression plots. A 'partial regression plot' is a 2D plot of the dependent variable versus one regressor when all other regressors are held constant at their respective medians/modes. These results allow us to present the multivariate regression function via a series of bivariate plots. This is the same approach that has been adopted by Maasoumi *et al.* (2007), among others.

5 Results

We now present our results for both parametric and nonparametric regression specifications. Our dependent variable in all models estimated is (the log of) the Real Sale Price (in $ Australian). All model specifications take logs of all continuous variables. Section 5.1 presents results of parametric model specifications, and nonparametric results follow in Section 5.2.

5.1 Parametric Specification

Our estimation procedure starts with a naïve OLS model followed by a flexible parametric specification that includes all regressors, treating time, LUC, CGT, GSTANN, FHOGANN, GSTFHOG, CAG, and municipality as unordered factors, and modelling distance, the interest rate and income as continuous variables. In the flexible specification we allow for interaction between all discrete and continuous variables, and allow for higher order polynomials in the continuous variables.

5.1.1 Naïve OLS

Results for a naïve OLS model employing the data set excluding the various categorical are presented in Table 3. This analysis allows us to to see what inferences we might draw from this data regarding the various tax changes we are examining. Given the sample size it is of little surprise to find that all the variables are statistically significant. The variables in general tend to confirm *a priori* expectations regarding sign and magnitude. For example, the income elasticity (1.43) is elastic and similar to that found in Abelson *et al.* (2005) (1.41). As expected, interest rates are found to be negatively related to price (interest elasticity of -0.003, compared to -0.04 in Abelson *et al.*). The parameter estimates on the time variables reflect the property market cycle over the sample period. Time is initially negatively related to price, but positively related to price after 1996. The parameter estimates on the distance variables show that distance is negatively related to price up to a distance from the center of at least 60 kms, beyond which prices rise with distance. This result is consistent with the Melbourne housing market as there is a boundary beyond which sub-division of land for dense new housing developments is no longer allowed. At this point you frequently observe larger houses with land, or hobby farms and lifestyle properties.

Next we consider the announcement effect of the various tax changes. The GST announcement effect is positively related to price but the FHOG announcement effect is negatively

Table 3: OLS Results Full Data Set

Variable	Coefficient	Standard Error	P Value
Intercept	-2.597143	.0347751	0.000
lnIncome	1.430696	.0033406	0.000
Interest Rate	-.0030843	.0003619	0.000
Time	-.0223099	.0003128	0.000
Time Squared	.0006	9.34e-06	0.000
Distance	-.0067478	.0001146	0.000
Distance Squared	.0000501	1.78e-06	0.000
GSTANN	.0159164	.0038813	0.000
FHOGANN	-.0432193	.0035621	0.000
CGT	.0411454	.0042469	0.000
GSTFHOG	-.0505108	.005624	0.000
CAG	.0084242	.002812	0.003
Adjusted R^2	0.3412		
Sample Size	745,466		

related to price. These results can be explained by the fact that the introduction of the GST should act to bring purchase decisions forward (see Wood *et al.,* 2006), putting upward pressure on house prices. On the other hand, the announcement of the FHOG should delay a house purchase by buyers who can gain this financial incentive, putting downward pressure on house prices. These results would appear to indicate that announcements of policies like the GST and FHOG can have a substantial impact on the market.

Finally, we consider the implementation of the various taxes. The parameter estimate for GSTFHOG shows that the introduction of the GST and the FHOG caused prices to fall. This is consistent with the observation that the announcement of these policies caused buyers to bring forward their purchase decision. The impact of the changes to the CGT are positive since they significantly reduced the tax burden on property investors. Lastly, the CAG had a positive effect on house prices. About 20% of the value of the grant is capitalized into higher house prices.

5.1.2 Flexible Parametric Model

We now consider a flexible parametric model to allow for interaction among all the continuous and discrete variables, the forms of non-linear interaction that theory would suggest are important for reduced form hedonic models. We selected our preferred model by employing the Schwarz-Bayes Information Criterion (BIC) as the basis for stepwise regression. The results of this estimation procedure yielded a BIC optimal model that has 234 parameters, is highly nonlinear in variables, and involves interaction among a number of the discrete and continuous variables. The optimal model specification involves non-linear terms on distance and income, and interaction terms on Municipality and distance (the complete model specification is given in the Appendix).

Summary statistics from this model are:

- Residual standard error: 0.4083 on 745228 degrees of freedom
- Multiple R^2: 0.4746, Adjusted R^2: 0.4744
- F-statistic: 2840 on 237 and 745228 degrees of freedom, p-value: $<$ 2.2e-16

The characteristics of the BIC optimal model suggest that the naïve OLS model suffers from mis-specification and omitted variables bias. The overall power of the flexible specification, as indicated by the adjusted R^2 is significantly higher than that of the basic model for the full data set. This is partly to be expected because of the use of additional variables (ie, LUC and Municipality) and the inclusion of many non-linear interactions.

Because of the high dimensional nature of the model estimated we summarize the preferred model with a range of partial regression plots (partial mean surfaces) along with bootstrapped error bounds, presented in Figures 1 and 2. Our method of constructing the partial regression plots is as follows:

1. Fit the full model via OLS
2. Generate predictions for the model in which all regressors are held constant at their median/modal values except the regressor appearing on the horizontal axis of the respective plot
3. For the continuous variables, this yields a scatter plot with the estimate given by a solid line and the error bounds given by the upper and lower bounds respectively
4. For the discrete variables, this yields a scatterplot with a series of dots arranged vertically. The middle dot represents the predicted value of log(price) for a given level

of the discrete variable, and the upper and lower dots represent upper and lower 95% error bounds respectively.

Clearly, the flexible parametric model is superior to the naïve OLS specifications presented in Section 5.1.1, implying that it is important to include the higher-order and interaction terms in the model. Furthermore, the inclusion of the two categorical variables, LUC and Municipality, should reduce the problem of omitted variable bias. Parmeter and Pope (2009) note that the inclusion of spatial variables significantly reduces problems of omitted variable bias.

While the relationships revealed by the partial regression plots in Figures 1 and 2 suggest that the non-linearities only have a different effect on the relationship between time and house prices, it is important to emphasize that these partial regression plots are all evaluated at the median/mode of the variables not appearing on the axis. Also, these partial regression plots do not illustrate any interaction effects.

Finally, as was the case with the naïve OLS model, both the CGT and the GST announcement had a positive effect on house prices. However, the BIC optimal model found no explanatory power associated with the date of implementation of the GST and FHOG. The increase in the flexibility of the model to capture important non-linearities shows that a key policy result from the initial analysis no longer holds.

5.2 Nonparametric Results

We now employ the local linear nonparametric regression model of Li and Racine (2004). We assume that price is an unknown function of CGT, CAG, GSTANN, FHOGANN, GSTFHOG, LUC, Municipality, time, distance, interest rates and income, where all continuous variables are in natural logarithms. We treat CGT, CAG, GSTANN, FHOGANN, GSTFHOG, LUC, and Municipality as nominal categorical variables, time as an ordered discrete variable, and distance, interest rates, and income as continuous. We use the method of Hurvich, Simonoff and Tsai (1998) for bandwidth selection. See Li and Racine (2004) for the properties of this method. Partial regression plots comparable to those in Figures 1 and 2 appear in Figure 3.

We find that LUC, Municipality, time, distance, interest rates and income all have explanatory power. But of the various tax variables which are the focus of our study, none of the tax change dates, have any explanatory power. This result is due to the fact that cross-validated bandwidth selection has smoothed out these variables by assigning the upper

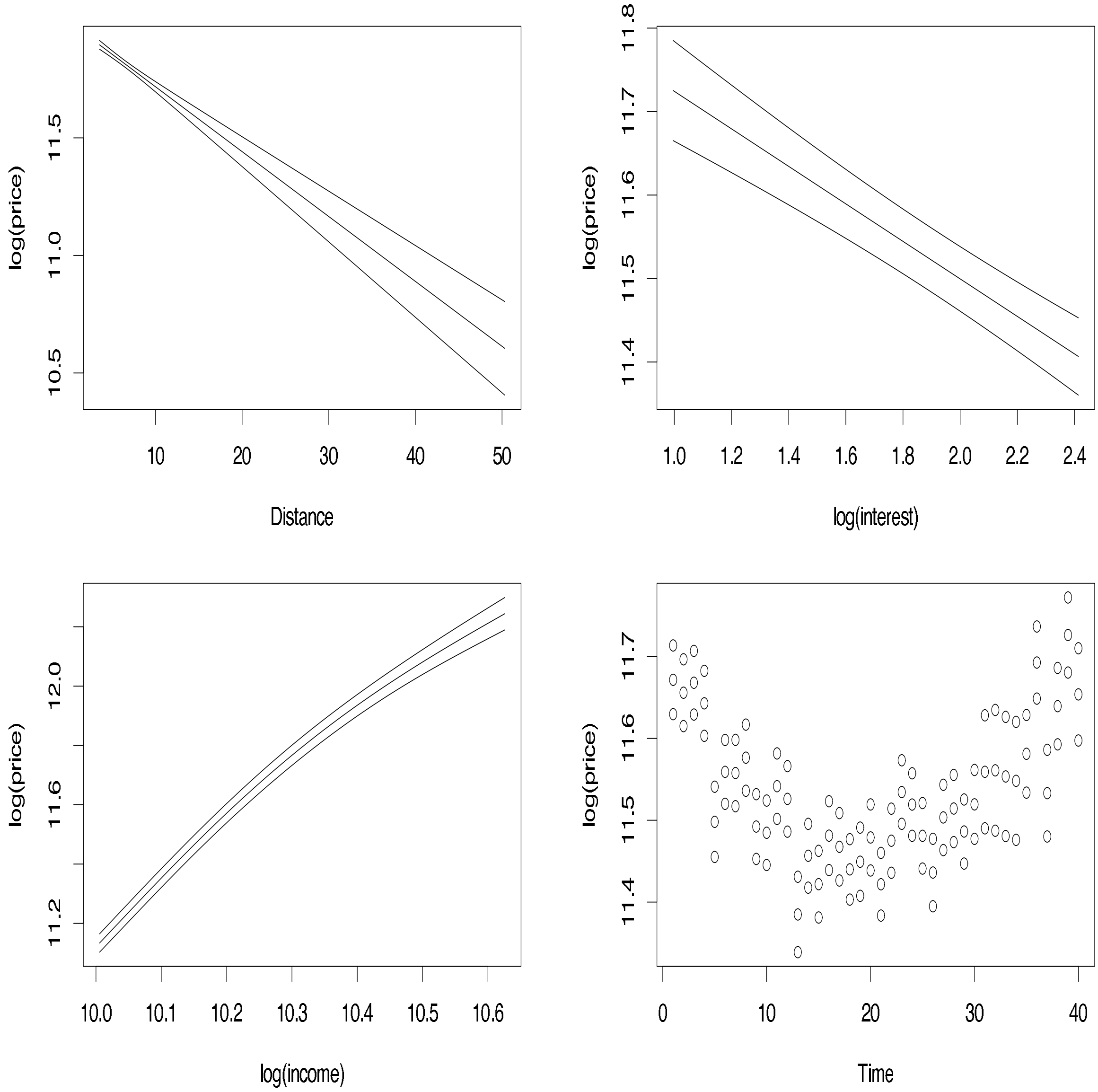

Figure 1: BIC-optimal partial regression plots.

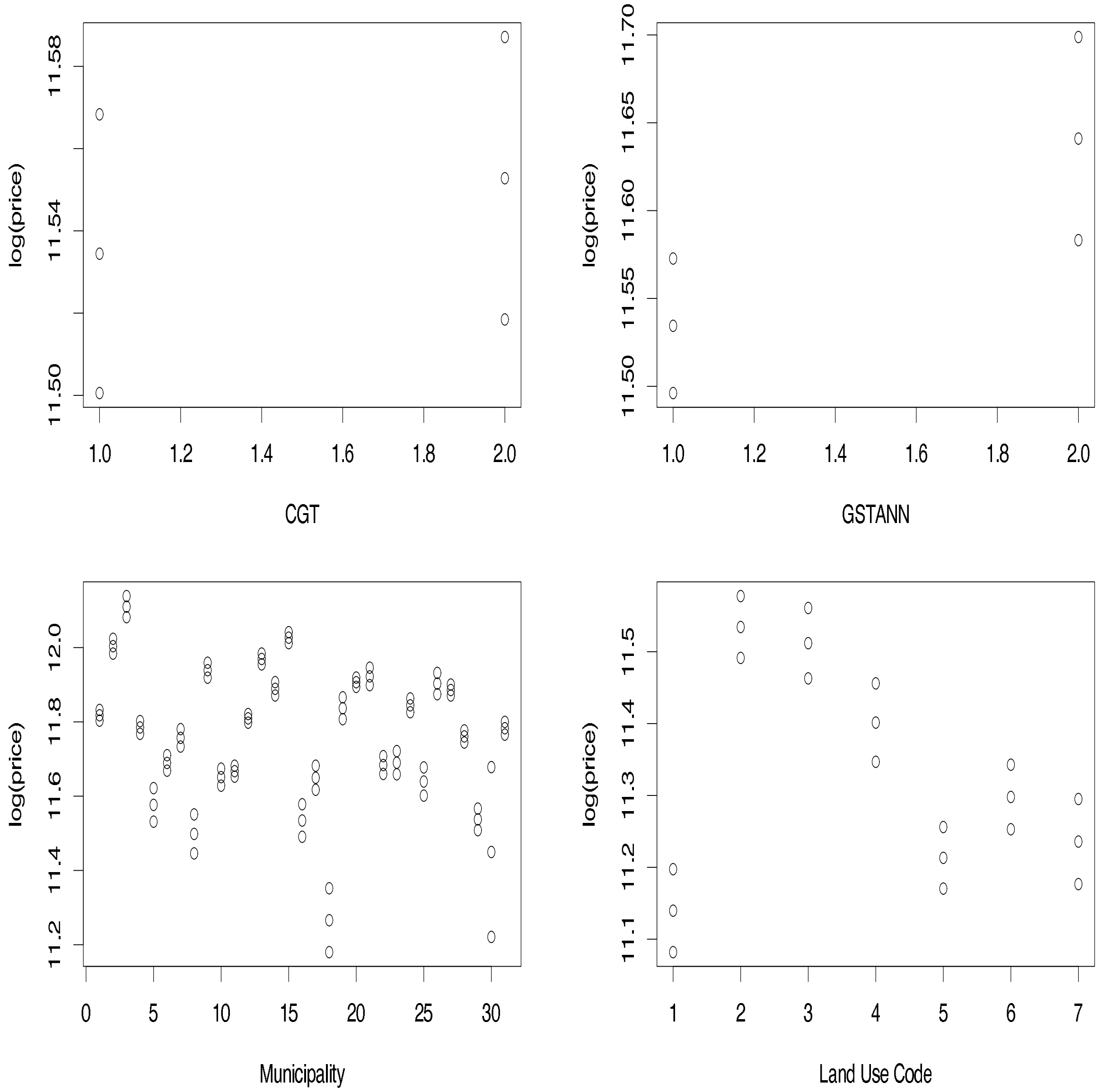

Figure 2: BIC-optimal partial regression plots. For the covariate Municipality, the integers on the horizontal axis refer to town names in alphabetical order, i.e., 31 Levels: 1=Banyule, 2=Bayside, 3=Boroondara, 4=Brimbank, 5=Cardinia, 6=Casey . . . 31=Yarra Ranges. For the covariate LUC, the integers correspond to the codes 10-16, i.e., 1=10 (New detached house), 2=11 (House previously occupied),. . . , 7=16 (retirement flat/unit/apartment). For the covariate CGT the integers correspond to 0/1, i.e., 1=0 (No reduction), 2=1 (reduction). For the covariate GSTANN the integers correspond to 0/1, i.e., 1=0 (Not announcement), 2=1 (Announcement period).

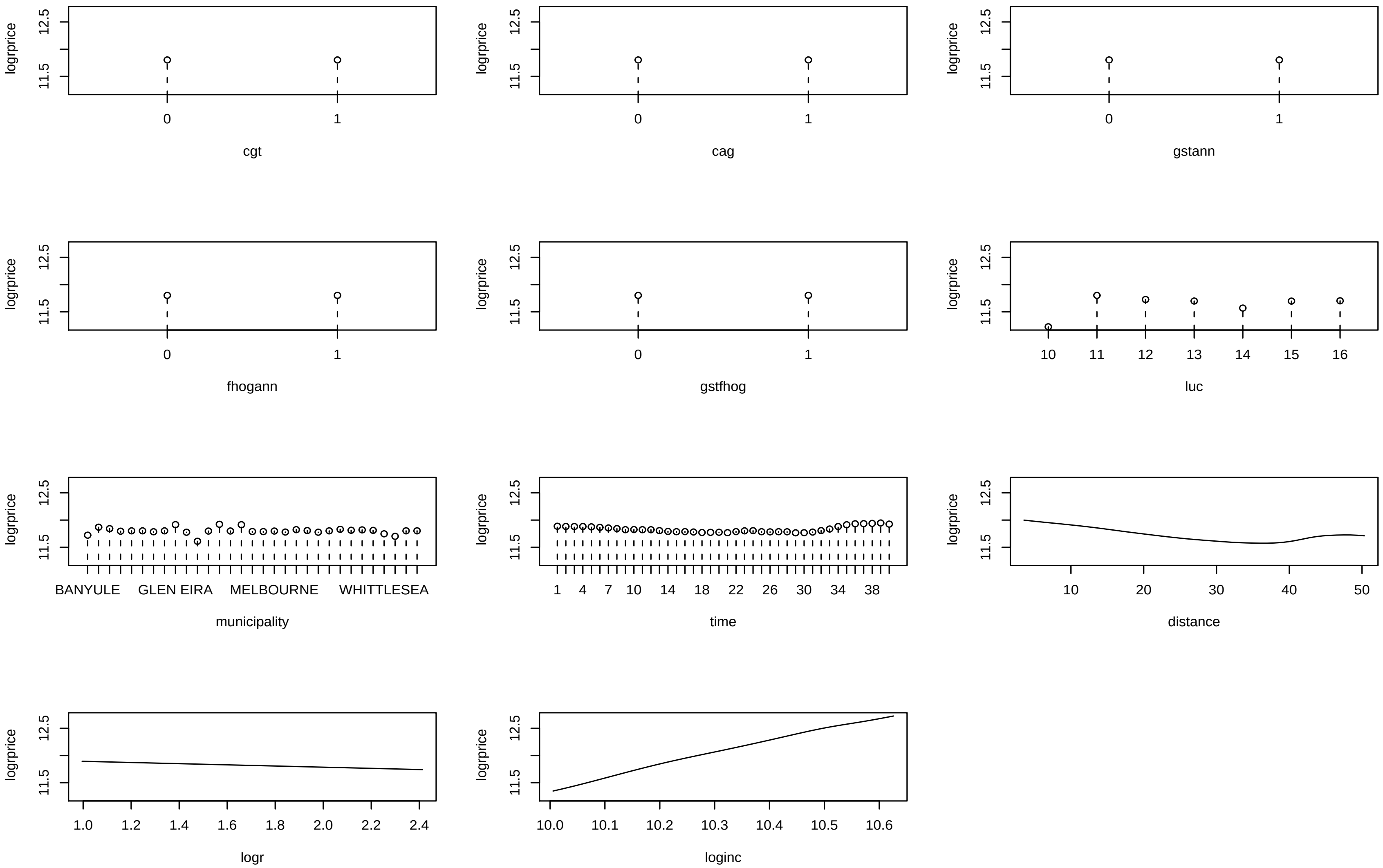

Figure 3: Fully nonparametric specification

bound values to the respective smoothing parameters, which is a property of Li and Racine's method in the presence of irrelevant variables (see Hall, Li and Racine (2007) for details). Overall, the nonparametric model yields an R^2 of 0.528, which implies that the model provides a better description of the data generating process than either of its parametric counterparts (whose respective goodness-of-fits were 0.3412 and 0.4746). Thus, the key finding in these results is that the changes in the tax system and the associated FHOG appear to have no explanatory power. *Prima facie*, this conflicts with the parametric results presented in Section 5.1.

Racine *et al* (2006) propose a test for the significance of categorical variables such as GSTFHOG, etc. in a nonparametric setting. A feature of the test is that if the smoothing parameter hits its upper bound, the variable will be deemed insignificant with probability one asymptotically. Therefore, there is no need to even run the test in case. Each of these variables is deemed insignificant by application of this test.

Given the importance of this result that the changes in the tax system have no explanatory power, we considered how smoothing the trend around the time of these events may have masked certain results. To see if the smoothing property of Li and Racine's method affected any of the results, we re-ran our analysis using a frequency estimator for time. That is, we set its bandwidth to zero which results in period by period analysis with zero smoothing over time. We found our nonparametric results to be qualitatively unchanged, with the sole exception being that the variability of the frequency estimates is substantially higher than than where time is smoothed. However, the findings for the tax variables are unaffected by smoothing/not smoothing time. The complete set of regression plots for this additional analysis where the bandwidth is set to zero to force zero smoothing over time is available from the authors on request.

As noted above, the results from the nonparametric model regarding the changes in the tax system conflict with the parametric results presented in Section 5.1. In order to reconcile this discrepancy, we revisit the naïve parametric specification, dropping the five tax variables (GSTANN, FHOGANN, CGT, GSTFHOG, and CAG) from the model, then re-estimating. This exercise changes the model's R^2 from 0.3412 to 0.3407, which strongly suggests that these variables have no predictive power whatsoever. An increase in explanatory power of the magnitude < 0.0006 (i.e., in the order of six over ten-thousand) is telling, and again brings to mind the fundamental distinction between statistical and economic significance. Though there exists a literature on adjusting the size of tests in large sample settings such as those implied by our dataset, these methods tend to be ignored by econometricians in practice. We leave this as an exercise for the interested reader. We therefore present this exercise solely to support our contention that these variables have no predictive power of interest to policy-makers.

In summary, comparing the parametric and nonparametric results, the much higher R^2 lends strong support to our findings, particularly in light of the enormity of the sample size. Thus, we conclude that the various changes that took place in Australia in terms of the housing market in particular and tax policy more generally do not appear to have had any appreciable economic impact on house prices whatsoever.

6 Conclusion

In this paper we used parametric and nonparametric methods to estimate a hedonic price model describing the behaviour of house prices in Melbourne, Australia over the period 1992-2002. The main purpose to estimating this model is to evaluate the effects of tax policy changes in Australia on house prices. We employed a very large data set comprising all residential property sales in Melbourne over the period, containing both continuous and categorical variables. Nonparametric estimation methods have been shown to be superior to parametric methods with such large data sets, but in the past, could not accommodate categorical variables in a fully nonparametric setting. We employed a recently developed nonparametric statistical method that accommodates both discrete and continuous data.

Parametric estimation results show that the various tax policy changes had a significant effect on the behaviour of house prices. However, nonparametric estimation results suggest that none of the variables which capture the changes in Australian tax policy have any explanatory power. This result implies that the various changes made to the tax system had the desired effect. That is, the effects of the introduction of the GST and the various other tax changes were offset by the introduction of the FHOG. Furthermore, there is no evidence that house prices were affected by the announcement of these policy changes as we might have expected.

Finally, removing the tax policy variables from the parametric estimation did not affect the explanatory power of the regression. We conclude that in the presence of a very large data set, naïve parametric estimation methods can give misleading significance results.

References

Abelson, P., R. Joyeux, G. Milunovich, and D. Chung (2005), “Explaining House Prices in Australia: 1970-2003”, *The Economic Record* 81:Special Issue, s96–s103.

Bernanke, B.S. and K.N. Kuttner (2005), “What Explains the Stock Markets’ Reaction to Federal Reserve Policy?”, *Journal of Finance* 60:3, 1221–1257.

Demiralp, S. and O. Jorda (2004), “The Response of Term Rates to Fed Announcements”, *Journal of Money, Credit and Banking* 36:3, 387–405.

Freebairn, J. (1999), “Tax Reform Proposals and Housing”, *Economic Papers* 18:3, 73–84.

Glaeser, E.L. and B. Sacerdote (2000), “The Social Consequences of Housing”, *Journal of Housing Economics* 9:1, 1–23.

Guest, R.S. (2005), “A Life Cycle Analysis of Housing Affordability Options for First Home Owner-Occupiers in Australia”, *The Economic Record* 81:254, 237–248.

Hall, P., Q. Li, and J. S. Racine (2007), “Nonparametric Estimation of Regression Functions in the Presence of Irrelevant Regressors”, *Review of Economics and Statistics* 89:4, 784-89.

Hurvich, C.M., J.S. Simonoff and C.L. Tsai (1998), “Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion”, *Journal of Royal Statistical Society* B 60, 271–293.

Kuminoff, N.V., C.F. Parmeter and J.C. Pope (2009), “Specification of Hedonic Price Functions: Guidance for Cross-Sectional and Panel Data Applications”, Working Paper No. 2009-02, Department of Agricultural and Applied Economics, Virginia Tech, Blacksburg, VA 24061 (`http://www.aaec.vt.edu./aaec/`).

Leung, C. (2003), “Macroeconomics and Housing: A Review of the Literature”, *Journal of Housing Economics* 13:4, 249-267.

Li, Q. and J.S. Racine (2004), “Cross-Validated Local Linear Nonparametric Regression”, *Statistica Sinica* 14:2, 485–512.

Li, Q. and J.S. Racine (2007). Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton, NJ.

Maasoumi, E., J. Racine and T. Stengos (2007), “Growth and Convergence: A Profile of Distribution Dynamics and Mobility”, *Journal of Econometrics* 136:2, 483–508.

Parmeter, C.F. and J.C. Pope (2009), “Quasi-Experiments and Hedonic Property Value Methods”, Prepared for the Handbook on Experimental Economics and the Environment, J.A. List and M.K. Price (Eds.), Edwards Elgar Publishers.

Productivity Commission (2004), *First Home Ownership* Report No. 28, Melbourne, Australia. (www.pc.gov.au).

Racine, J.S. and Q. Li (2004), "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data", *Journal of Econometrics* 119:1, 99–130.

Racine, J.S. and A. Ullah (2006), *Nonparametric Econometrics*, Palgrave Handbook of Econometrics, Volume 1: Theoretical Econometrics.

Racine, J.S. and J. Hart and Q. Li (2006), "Testing the Significance of Categorical Predictor Variables in Nonparametric Regression Models", *Econometric Reviews* 25:4, 523-44.

Wood, G., R. Watson and P. Flatau (2006), "Microsimulation Modelling of Tenure Choice and Grants to Promote Home Ownership", *The Australian Economic Review* 39:1, 14-34.

Appendix

Data sources and tranformations:

- The distance data was provided by Dr. Francis McGinley, Senior Research Fellow at the Transport Research Centre, RMIT Business, who used a 1997 postcode datafile and calculated crow fly distances in kilometres for postcode centroids to the Melbourne General Post Office (GPO, designated centre of the Melbourne CBD, postcode 3001). For missing postcodes we used Geoscience Australia's "As the Cocky Flies" distance to the Melbourne GPO (37d48'S, 144d57'E), available from `http://www.ga.gov.au/map/names/distance.jsp`.

- Data on mean taxable income by postcode was sourced from the Australian Tax Office website `http://www.ato.gov.au/individuals/pathway.asp?pc=001/005/009/011&mfp=001/002&mnu=5489\#001_005_009_011` and from hardcopies of the Taxation Statistics (Canberra, A.C.T.). There were some postcodes in the PRISM dataset for which there were no corresponding postcodes in any of the Taxation Statistics years. For example, a separate postcode for the Docklands neighbourhood did not exist until a few years ago. PRISM records for which no corresponding postcode existed in the Taxation Statistics were removed from the sample.

- Interest Rates were sourced from the Reserve Bank of Australia series FILRHLVBS from `F05HIST.xls` available at `http://www.rba.gov.au/Statistics/Bulletin/index.html` F Tables - Indicator Lending Rates.

- The CPI was sourced from the Australian Bureau of Statistics (6401.0 Consumer Price Index, Australia TABLE 1B. CPI: All Groups, Index Numbers (Quarter)(a)) (1989-90 base year). For the September 2000 quarter, the ABS recommends removing the effect of the GST, noting that it is "likely to have contributed 2.5 to 3% points to the CPI increase in the September quarter." (ABS Special Article: Measuring the impact of the new tax system on the September Quarter 2000 Consumer Price Index p.2). We subtracted 3% from the quarterly inflation of 3.8%.

Specification of flexibile parametric model:

The optimal specification of the flexible parametric model is summarized as:

$$\begin{aligned}
logrprice &= CGT + CAG + factor(LUC) + factor(municipality) + factor(time)\\
&+ distance + distance^2 + distance^3 + distance^4 + distance^5 + distance^6\\
&+ logr + loginc + loginc^2 + loginc^3 + loginc^4 + loginc^6\\
&+ factor(LUC) \times loginc + factor(municipality) \times loginc\\
&+ factor(municipality) \times distance + factor(time) \times distance\\
&+ factor(LUC) \times distance + factor(LUC) \times logr\\
&+ factor(municipality) \times logr + CGT \times loginc + \epsilon
\end{aligned}$$

where $factor(\cdot)$ indicates that the variable $(\cdot)$ is an (unordered) categorical one (i.e., a dummy variable(s)).