CHAPTER 5

Analysis of ‘Number of Permits’ with Maximum Likelihood

5.1 Introduction

The analysis in the previous chapter indicated that it is necessary to go beyond the estimations that can be performed with a standard estimation package. This chapter describes two models that examine the effect of the two-rate tax on the number of permits, and which seek to address the criticism that one can raise against the preliminary analyses. Both models use a negative binomial distribution with the more general mean-variance relationship that was introduced in Chapter 3; the implementation of this new relationship is explained in the next two sections. Section 5.4 describes a model that uses ‘density’, ‘income’, ‘population change’ and the tax differential as independent variables, but introduces them in more general functional forms than it could be done with LIMDEP. The model in Section 5.5 incorporates several insights that were gained through the work with the model in Section 5.4. First, it is a fixed effects model that uses dummy variables for each municipality in the data set. Second, it implements the idea that the tax differential will have no effect on construction if the economic condition of a municipality is too depressed. Third, it is estimated with an even more general version of the variance-mean relationship than the earlier model, and it uses data between 1972 and 1994, while the analysis in Section 5.4 used data only until 1993.

5.2 The relationship between then mean and the variance

Section 3.4 motivated a special relationship between the mean and the variance in the estimation of the negative binomial distribution. From the regression function (see equation 5.4 below) one obtains the expected value μ of the dependent variable; this expected value is used to calculate the coefficient of variation (*CV*), from which the variance can be determined as $\sigma^2 = CV^2 \cdot \mu^2$. The two parameters of the negative binomial distribution are then calculated from the mean and the variance. The crucial relationship is between *CV* and the mean, which is formulated as the sum of the minimum *CV* and an added curve with a single maximum and two tails.

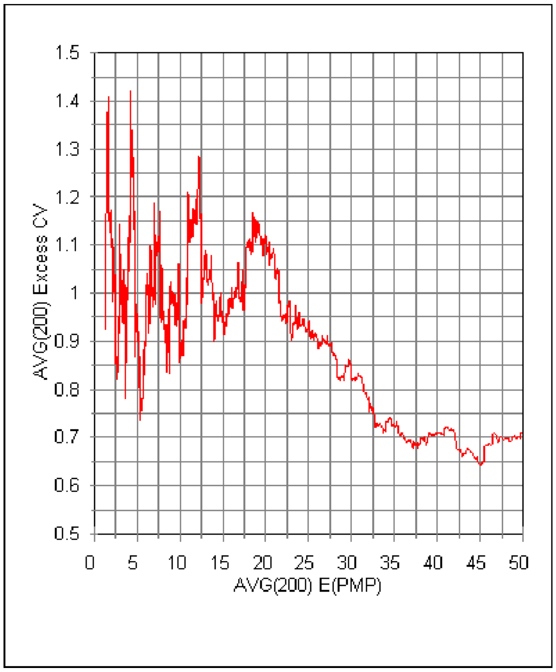

Figure 5.1 Relationship between the actual excess CV and m.

Which added curve is most appropriate for this data? Figure 5.1 shows the relationship between the actual excess *CV* above the minimum *CV*, and the expected number of permits for all 2,689 observations of nonresidential construction of whole units, averaged over 200 observations. To calculate the excess *CV*, each observation is assigned an approximated 'standard deviation' of $|PMP\text{-}E(PMP)|$. The coefficient of variation is then calculated as $|PMP\text{-}E(PMP)|/\,E(PMP)$. Subtracting the minimum *CV* yields the 'excess' *CV*. For low values of μ this 'excess' *CV* increases, while it decreases as μ becomes large. It seems to be sensible to approximate this relationship either by a normal curve with 3 parameters, or by a function motivated by the *t*-distribution with 4 parameters, which is similar to the normal curve, but can have a different kurtosis, that is, tails of different thickness. These functions are general enough to support a wide variety of 'excess' variation; if for example the excess variation of the data does not show a maximum, then the maximum point can be put far enough to the left so that only the right tail of the function is used.

The relationship is now formulated as follows: let $\mu_{i,t}$ be the expected number of permits for municipality *i* in year *t*, so that $CV_{i,t}$ can be determined according to

$$CV_{i,t} = \begin{cases} \sqrt{\dfrac{1-\mu_{i,t}}{\mu_{i,t}}} + \gamma_1 \, e^{\gamma_3 (\ln\mu_{i,t}-\gamma_2)^2} & \text{if } \mu_{i,t} \le \frac{1}{2} \\ \dfrac{1}{2\mu_{i,t}} + \gamma_1 \, e^{\gamma_3 (\ln\mu_{i,t}-\gamma_2)^2} & \text{if } \mu_{i,t} \ge \frac{1}{2} \end{cases} \tag{5.1}$$

with 3 parameters if the normal curve is added, and according to

$$CV_{i,t} = \begin{cases} \sqrt{\dfrac{1-\mu_{i,t}}{\mu_{i,t}}} + \gamma_1 \left(1 + \dfrac{\gamma_4(\ln\mu_{i,t}-\gamma_3)^2}{\gamma_2}\right)^{-\frac{\gamma_4+1}{2}} & \text{if } \mu_{i,t} \le \frac{1}{2} \\ \dfrac{1}{2\mu_{i,t}} + \gamma_1 \left(1 + \dfrac{\gamma_4(\ln\mu_{i,t}-\gamma_3)^2}{\gamma_2}\right)^{-\frac{\gamma_4+1}{2}} & \text{if } \mu_{i,t} \ge \frac{1}{2} \end{cases} \tag{5.2}$$

with 4 parameters if the more general curve is added.[149] The first term in both equations represents the minimum *CV* as determined in Chapter 3, and the second term is the added function. The parameter γ_1 determines the height and γ_2 determines the location of the maximum, while γ_3 is a measure of dispersion. γ_4 determines the thickness of the tails; as γ_4 approaches zero, the function approaches the normal curve. The model in Section 5.4 uses the normal curve, while the model in Section 5.5 uses the more general curve.[150]

5.3 Correction for serial correlation

Both models in this chapter correct for serial correlation in the data by including a lagged residual among the explanatory variables. If data are available for two consecutive years, this method is straightforward to implement. But not every observation has a predecessor: the first year has none at all, and depending on missing data, some observations do not have a direct predecessor, but only one which was observed more than one year ago. If these observations are to be kept in the analysis, then it is necessary to increase the estimate of their variance, because the explanatory power of the lagged residual ought to decrease as the distance to the previous observation increases.[151] The estimate of the variance σ^2 is given by $\sigma_{i,t}^{\ 2} = CV_{i,t}^{\ 2} \cdot \mu_{i,t}^{\ 2}$, which means that the variance will increase with estimated $CV_{i,t}$.

In Section 5.4 the variance of all observations without a direct predecessor is increased by multiplying the excess $CV_{i,t}$ in equation 5.1 by $(1 + \nu_1)$, where $\nu_1 > 0$ is a parameter to be estimated.

In Section 5.5 the excess *CV* of the first observation (which does not have any predecessor) is increased by $(1 + \nu_1)$ as well, but the serial correlation is handled in a more specific way. The residual that is used as explanatory variable is the residual from the previous observation; if the direct predecessor is missing, then the residual that was calculated 2 periods ago is used; if that is missing, too, then the residual that was calculated 3 periods ago is used, and so on. The excess $CV_{i,t}$ is multiplied by

$$\left(1 + \nu_1 \left(1 - \nu_2^{\,n_{i,t}}\right)\right) \tag{5.3}$$

---

[149] As the expected number of permits, $\mu_{i,t}$, is always larger than 0, it is always possible to calculate its logarithm.

[150] Equation 5.2 describes a wider variety of functions than equation 5.1, but the possibility of using it did not occur to me when I estimated the model in Section 5.4.

[151] Alternatively all observations without a direct predecessor could have been eliminated from the analysis, which would have led to a large reduction of the available data.

where $0 \leq \nu_2 \leq 1$, and $n_{i,t}$ is the number of years that are missing between the current observation and its predecessor. If no year is missing, $CV_{i,t}$ is not changed; as the number of missing years becomes larger, the correction approaches the maximum correction ν_1 for the first period.

5.4 A non-linear model
5.4.1 Setup of the model

In addition to its limited distributional assumptions and its neglect of serial correlation, the analysis with LIMDEP suffered from further shortcomings. First, the models did not restrict the coefficients of ‘Population’ and ‘Months reported’ to be equal to 1. Second, the analysis assumed an exponential relationship between the expected number of permits $\mu_{i,t}$ and the explanatory variables of the form

$$\mu_{i,t} = (pop_{i,t})^{\beta_2} \cdot (m_{i,t})^{\beta_3} \cdot e^{\beta_1 + \beta_4 D_{i,t} + \beta_5 I_{i,t} + \beta_6 \dot{P}_{i,t} + \beta_7 YD_t + \beta_8 TAX_{i,t}} \tag{5.4}$$

where i is the index for the municipality, and t represents the year. But it is not obvious why for example density should be introduced as $\exp(D_{i,t})$; even though this relationship yields a model which can easily be evaluated, a different functional form might be more appropriate, and the coefficients which are estimated with its linear approximation might be distorted. It would clearly be possible to introduce any variable as a polynomial of higher order, or to introduce two variables as a multiplicative combination of each other, but economic theory does not suggest any particular form, and it will be very tedious to discover the ‘optimal’ form with an unguided ‘trial and error’ method of testing different functional relationships and discarding the ones which yield a lower loglikelihood.

The model in this section attempts to introduce the three explanatory variables with more appropriate functional forms. In addition it takes account of serial correlation. The regression equation of the model is given by

$$\mu_{i,t} = f_D(D_{i,t}) \cdot f_I(I_{i,t}) \cdot f_P(\dot{P}_{i,t}) \cdot f_T(ATD_{i,t}) \cdot f_r(r_{i,t-1}) \cdot YD_t \cdot pop_{i,t} \cdot mon_{i,t}\ , \tag{5.5}$$

where $D_{i,t}$ is popultion density, $I_{i,t}$ is income relative to average income, $P_{i,t}$ is the population change, $ATD_{i,t}$ is the adjusted tax differential, $r_{i,t-j}$ is a term that takes serial correlation into account by including the residual that was observed $j > 1$ periods ago, YD_t is a yearly dummy variable, $pop_{i,t}$ is population size and $mon_{i,t}$ is the number of months reported.[152] As the coefficients of 'Population' and 'Months reported' are restricted to 1, no functions are estimated for them.

If one plans to examine functional forms other than polynomials, it becomes very important to find a way to judge the 'fit' of any proposed relationship. A starting point of this search can be a cross plot of the independent and the dependent variable. This cross plot indicates the general relationship between the two variables, and shows if a linear function might be a sufficient approximation. But clearly it is not a very precise indicator of the correct functional form, because it does not enable one to compare two only slightly different functional forms with each other. The comparison of the loglikelihoods of two competing functional forms indicates which of the functions represents the data more closely, but does not show if the better of the two can still be improved.

A tool which helps to decide how to improve the functional form is a cross plot of the independent variable and the residual of the regression. If all other variables except the currently examined variable are included in the regression, the plot of the variable versus the residual shows the values of the independent variable for which the observed number of permits is either over- or underpredicted.[153] The right functional form should correct for these mispredictions, and yield a plot which shows an average residual of 1 and no systematic pattern.[154] But it turns out that for the available data the residual is affected too much by other effects to yield very reliable information.

Another indicator for the right functional form is a plot of the cumulative distribution function (CDF). The CDF of a continuous random variable x determines the probability that x is less than or equal to a certain value α. It is defined as the integral of the probability distribution of x between $-\infty$ and α. As $\alpha \rightarrow -\infty$ the CDF converges to 0, and as $\alpha \rightarrow \infty$ the CDF converges to 1. The expected value of the CDF is ½, and if a plot of the variable, sorted by size, versus the CDF systematically shows values different from ½, then the functional form is most likely incorrect. But this technique also proved to be too unreliable to be of much help.

[152] One point value of every functional form needs to be fixed, for example the vertical intercept, and the yearly dummies are used as scaling coefficients.

[153] In order to filter out some noise, it proved to be more informative to examine the plot of the variable value which is averaged over 200 observations versus the residual which is also averaged over 200 observation.

[154] Note that the expected residual of a multiplicative relationship is not 0 but 1.

A good indicator of potential explanatory power and of the correct functional form turns out to be a graph that shows the relationship between the independent variable and the first derivative of the loglikelihood of each observation with respect to the vertical intercept parameter.[155] If the derivative is positive, then a higher intercept would improve the prediction of this observation, and if it is negative, then this observation ought to have a lower intercept. If the plot of this first derivative shows a systematic pattern over all observations, then the use of a different functional form which takes the suggested change into account will most likely improve the sum of the single loglikelihoods.[156] Once the optimal parameter values of the correct functional form have been found, or if the variable has no explanatory power at all, the mean value of the first derivative will be zero, and its variance over all observations should be small.

Once the general relationship between the dependent and the independent variable has been established, but the exact functional form that best described that relationship has not been found yet, it is helpful to guess a certain function and to determine the first derivative of the loglikelihood with respect to every parameter of this function. This shows over which part of the independent variable a change in the guessed function can improve the loglikelihood, and will lead to a better guess of the correct function. For example, assume that the guessed function is exponential, and that the plot of the average derivative of the exponent with respect to the intercept shows zeros for small parameter values, but becomes negative for large values. Most likely the use of a logistic function, which has a similar slope for small values, but flattens out when the independent variable becomes large, will improve the result.

In the following analysis the function of each variable has a certain vertical intercept; multiplication of all functions with each other as indicated by equation 5.5 yields the intercept for the whole relationship between the dependent and the independent variables. The yearly dummy variables are then used as scaling parameters for this intercept, and the derivatives of the independent variables are taken with respect to these dummy variables. It would have been possible to parameterize all functions to yield a common intercept of 1, so that the total intercept would have been 1 as well, but this would have been merely an aesthetic improvement, and would not have changed the result of the analysis.

---

[155] It is very tedious to determine and program the algebraic derivative of the loglikelihood function for changing functional forms, and it is much easier to calculate a numerical derivative. For the calculation I used Ridders' (1982) algorithm, which is described for the program language *C* in Press *et al* (1995). This algorithm is very precise; a comparison between the algebraic and the numerical derivative of equation 5.5 yielded values that were identical up to the 6th digit.

[156] As the variance of the derivatives will be high, it is again better to use the moving-average of the derivatives to get a more stable indicator. An average of 100 or 200 derivatives is usually enough to show a pattern without too much variation.

The following section describes the search for the correct functional forms for the data set of nonresidential construction of whole units, and Section 5.4.3 examines the data set of residential construction of whole units.

5.4.2 Examination of nonresidential construction of whole units:

A The functional form for 'density':

Population density ought to have a negative influence on building activities. The lower the density, the fewer existing structures usually have to be torn down to build a new building, and the cheaper construction becomes. This is confirmed by the analysis in Section 4.7, which yielded a significant coefficient for density of -1.0107.

Figure 5.2.a shows the relationship between density and the first derivative of the yearly dummy variables when population density is excluded from the analysis, while all other variables are included with their final functional forms as indicated in Table 5.1. The loglikelihood is -4,441.49; the relationship is negative and seems to be linear, which resulted in a first guess of the function as

$$f_D\,(D_{i,t}) \;=\; 1 \;+\; \delta \,\cdot\, D_{i,t}\;. \tag{5.6}$$

The linear function improved the loglikelihood to -4,377.45. The test statistic of the likelihood ratio test is 128.08, and the critical value of the chi-square distribution with 1 degree of freedom at 99 percent confidence is 6.634. The inclusion of density therefore yields a significant improvement in the loglikelihood.[157]

But from Figure 5.2.b it becomes clear that the linear function does not completely eliminate the underlying trend. Places with low density are systematically underpredicted, and places with high density are systematically overpredicted. In order to elevate the function for high density municipalities, the next attempt to find the right function was made with a hyperbola, which had the horizontal axis as one limit, and therefore needed only two parameters (in addition to the (restricted) slope parameter of the hyperbola branch):

$$f_D\,(D_{i,t}) \;=\; \frac{(1 \,+\, \delta_1\; D_{i,t}) \;+\; \sqrt{(1 \,+\, \delta_1\; D_{i,t})^2 \;+\; 4\;\delta_2}}{2}\;. \tag{5.7}$$

[157] The likelihood ratio test is described in Section 3.4. Note that the functional form without density is a restricted form of, and therefore nested in the examined functions, so that a likelihood ratio test is meaningful.

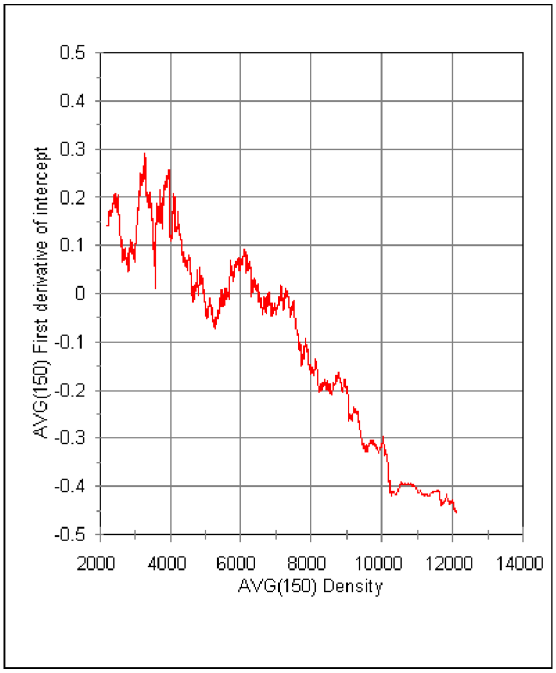

Figure 5.2.a Density and the first derivative of the intercept when density is excluded.

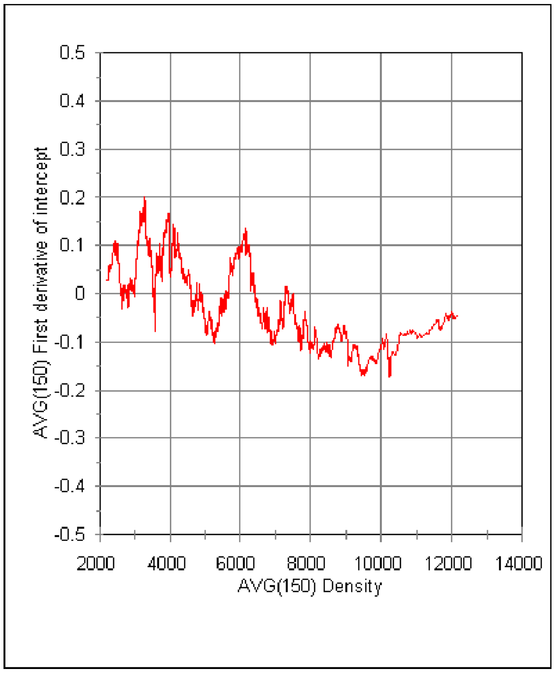

Figure 5.2.b Density and the first derivative of the intercept once density is included as a linear function.

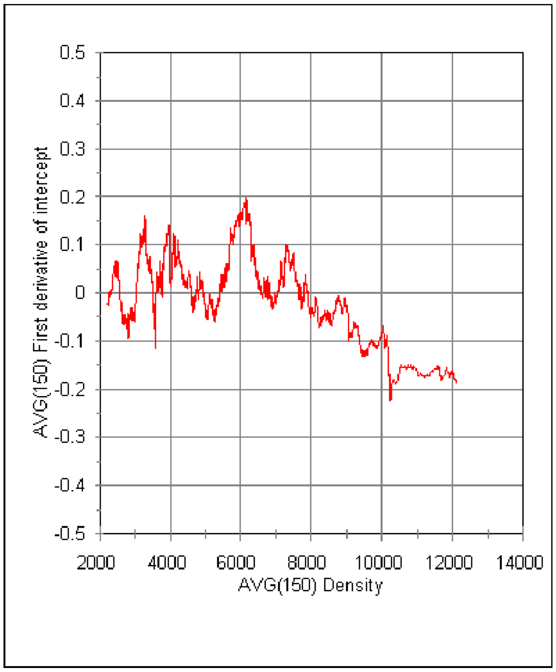

Figure 5.2.c Density and the first derivative of the intercept once density is included as a hyperbola with one branch equal to the horizontal axis.

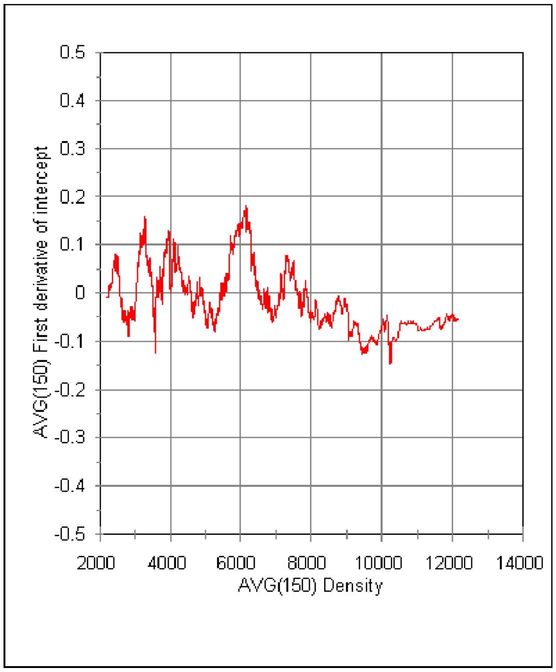

Figure 5.2.d Density and the first derivative of the intercept once density is included as a hyperbola with two positively sloped branches.

The parameter δ_1 represents the asymptotic slope of the unconstrained branch of the hyperbola, and δ_2 indi-cates the distance of the hyperbola from the point where both asymptotes intersect. This function improves the loglikelihood to 4,364.03, but as shown by Figure 5.2.c the trend has not been eliminated completely. The true functional form needs still to be lower for high density municipalities, so the next attempt was made with a hyperbola whose two branches were normalized so that they intersect at a vertical value of 1. This function, which uses 4 parameters, is given by

$$f_D\,(D_{i,t}) = \frac{A \cdot B + \sqrt{(A-B)^2 + 4\,\delta_4}}{2}$$

where (5.8)

$$A = 1 + \delta_1\,(\,D_{i,t} - \delta_3\,)$$
$$B = 1 + \delta_2\,(\,D_{i,t} - \delta_3\,)\;.$$

The parameters δ_1 and δ_2 represent the asymptotic slopes of the two hyperbola branches, δ_3 indicates the density at which the two asymptotes intersect, and δ_4 indicates the distance of the hyperbola from the intersection point of the two asymptotes. The loglikelihood is 4,360.84, which is an improvement of 3.19. This increase is too small to justify the two additional parameters on the 99 percent level. A further guess of the true function is an exponential decay with two parameters:

$$f_D\,(D_{i,t}) = 1 + \delta_1\,e^{-\delta_2\,D_{i,t}}\;, \tag{5.9}$$

which yields a loglikelihood of 4,360.88, and is a significant improvement over equation 5.7 because no additional parameters are used. Still the plot of the first derivatives in Figure 5.2.d shows an overprediction for high density municipalities. The true functional form seems to be rather parabolic; usage of the function

$$f_D\,(D_{i,t}) = 1 + \delta_1\,(\,D_{i,t} - \delta_2\,)^2 \tag{5.10}$$

results in a loglikelihood of -4,360.41, which is slightly better than the loglikelihood obtained by the exponential function, and it removes most of the trend in the first derivatives as shown in Figure 5.3.

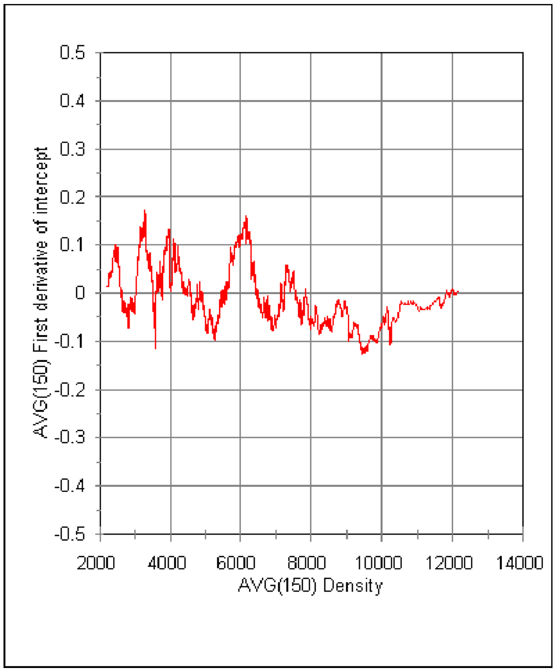

Figure 5.3 Density and the first derivative of the intercepts once density is included as a parabola.

B The functional form of 'income relative to average income'

Relative income can be expected to have a negative influence on nonresidential building activity in a community. Communities with a high relative income are more likely to serve as residential municipalities whose inhabitants work in a neighbor town; rich communities usually try to prevent firms from settling within their borders. This is also suggested by the analysis in Section 4.7, which yielded a coefficient on income of -0.19583, although it was only insignificantly different form zero. Figure 5.4.a on the following page shows the first derivative of the intercept dummies when relative income is excluded. The loglikelihood is 4,386.55, so the first guess is again the linear relationship. The linear function with one parameter

$$f_I\ (I_{i,t}) = 1 + \iota_1\ I_{i,t} \qquad (5.11)$$

yields a loglikelihood of 4,366.29, but as Figure 5.4.b shows, the correct function is flatter for high levels of relative income. Figure 5.4.c shows the derivatives if a parabolic relationship (as in equation 5.7) is assumed. The loglikelihood improves to 4,360.41. Even though the graph still shows some systematic variation, I did not find a function which could reduce this variation and pass a likelihood ratio test, so that the parabolic function seems to be the best.

C The functional form of 'population change':

Municipalities that had a substantial increase in their population are more likely to build new structures than places that lost many of their inhabitants in recent years. But this additional construction which is caused by the population change, ought to take place primarily in the residential sector; for nonresidential structures it is likely that economic prosperity and therefore increased construction activities attract more people, so that the causation may be reversed. To keep the possibility of reverse causation as small as possible, one should examine population change in previous periods instead of current population change. This finding is confirmed in the analysis in Section 4.7, where population change had a positive but insignificant influence for nonresidential construction, and a positive and significant influence for residential construction. But as Figure 5.5 shows, when lagged population

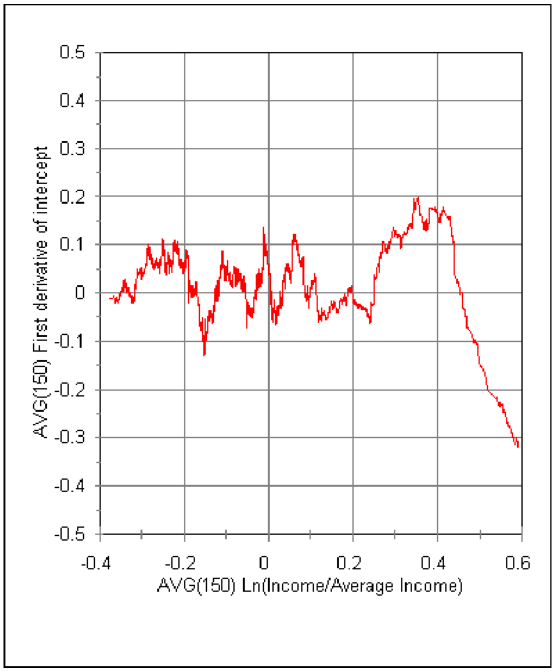

Figure 5.4.a Relative income and the first derivative of the intercept when relative income is excluded.

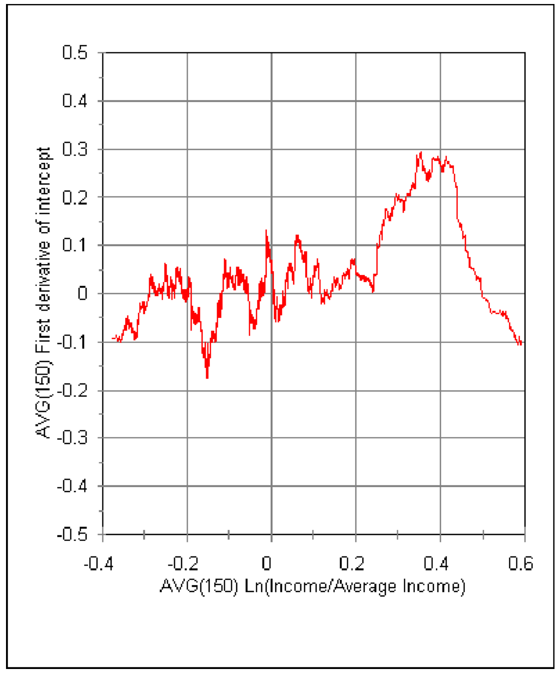

Figure 5.4.b Relative income and the first derivative of the intercept once relative income is included as a linear function.

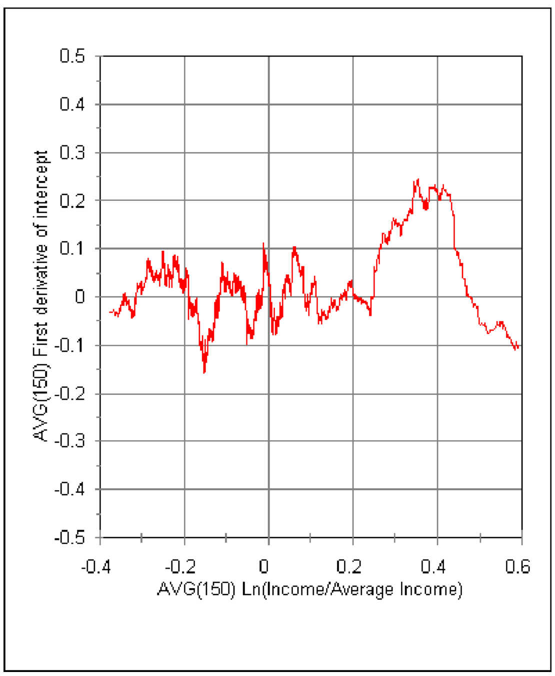

Figure 5.4.c Relative income and the first derivative of the intercept once relative income is included as a parabola.

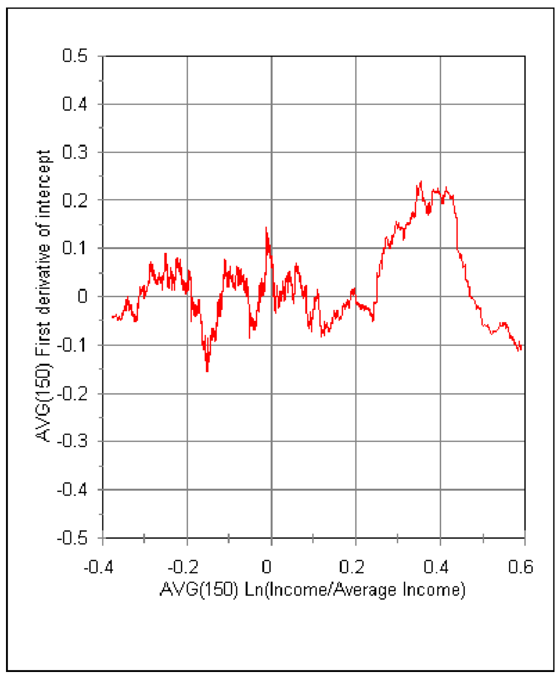

Figure 5.4.d Relative income and the first derivative of the intercept once relative income is included as a combination of a parabola and a trigonometric function.

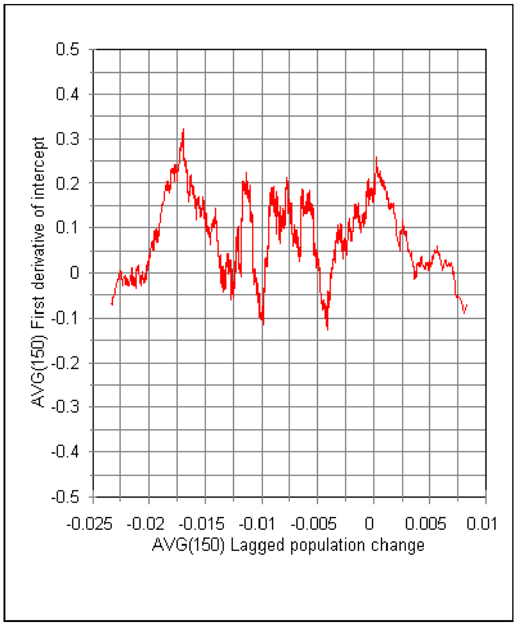

Figure 5.5 Lagged population change and the derivative of the intercept when lagged population change is excluded from the analysis.

change is excluded from the analysis it does not seem to have any noticeable trend or impact on the intercept dummies, except some apparently random noise. Using a linear relationship between PMP and lagged population change improves the loglikelihood to 4,359.57, which is not enough to pass the likelihood-ratio test; the variable should not be included in the further analysis.

D The functional form of lagged residuals:

As the estimated relationship is multiplicative, the residual is obtained by dividing the expected PMP into the observed PMP. If the observed PMP is completely explained by the regression, then the residual is equal to 1; any deviation from this value is a sign of either over- or underprediction.

Figure 5.6.a shows the relationships between the residual lagged by one period and the first derivative with respect to the dummy intercept. The relationship is positive: the larger the lagged residual $r_{i,t-1}$, the larger the potential improvement in the loglikelihood when the intercept is increased, but the additional improvement decreases with the size of the lagged residual. In general, if $r_{i,t-1}$ is equal to one, then last year's prediction was exactly equal to the observed value of PMP, and the best estimate for this year's PMP(independent of the influence of the other variables) is the same as it was for last year's PMP, so the function should cross the point (1,1). If $r_{i,t-1}$ is smaller than 1, then last year's predicted PMP was too large,so that (with serial correlation) this year's PMP will most likely be smaller than otherwise predicted; the estimated value should be corrected downwards. The reverse is true for $r_{i,t-1} > 1$; if last year's PMP was larger than predicted, then this year's PMP should be larger as well. The figure suggests an inverse exponential function, which converges to a maximum larger than 1 as the residual becomes large, and uses 2 parameters:

$$f_r(r_{i,t-1}) = \rho_1 - (\rho_1 - 1)\, e^{\frac{-\rho_2 \cdot (r_{i,t-1} - 1)}{\rho_1 - 1}} . \qquad (5.12)$$

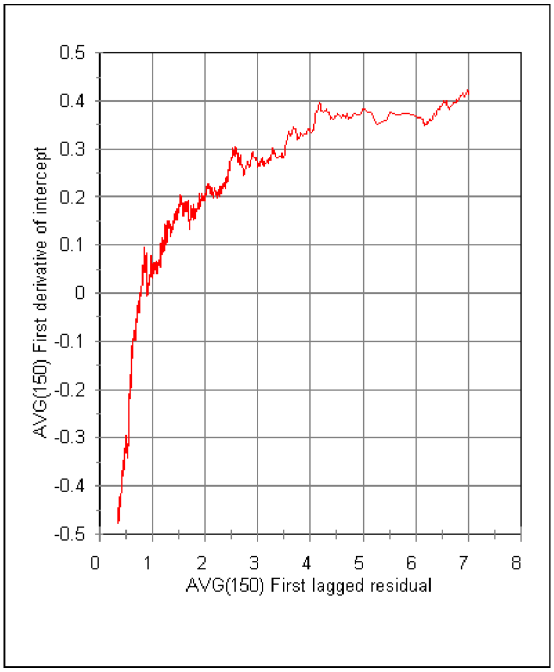

Figure 5.6.a First lagged residual and the first derivative of the intercepts if the first lagged residual is not included in the analysis.

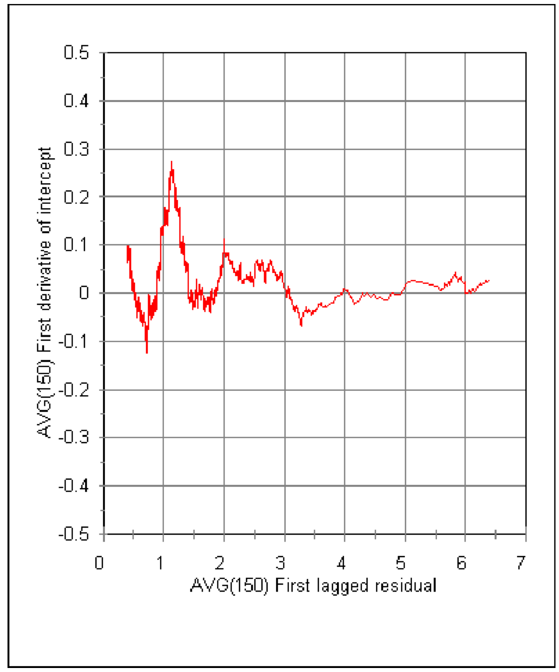

Figure 5.6.b First lagged residual and the first derivative of the intercepts once the first lagged residual is included as an inverse exponential function.

The parameter ρ_1 describes the asymptotic value to which the function converges when the residual becomes large, and ρ_2 describes the slope of the function at the point (1,1). Figure 5.6.b shows the first lagged residual and the first derivative of the dummy variables once the first lagged residual is included in the analysis; the relationship no longer has a trend. As a lagged residual neither exists for the first year nor for each year that follows a year that is missing, it cannot be included as an explanatory variable for those years. As pointed out in section 5.3, this should increase the variance of the estimation; this increase is estimated as an additional parameter ν. The loglikelihood improves from 4,416.37 to 4,360.42, that is, the three additional parameters are significant and pass the likelihood-ratio test.[158]

[158] The critical value for three parameters is 11.345 at 99 percent confidence.

E The functional form of the adjusted tax differential:

As only 15 two-rate tax cities are included in the data set, a graph which shows the first derivative of the intercept as a function of the adjusted tax differential does not have enough different data points to yield any interesting insights. The best guess is to include the tax differential as a linear function with one parameter. The loglikelihood rises by 4.78, which is enough to pass the likelihood ratio test. The value of the coefficient is -0.067, which seems to suggest that the two-rate tax has a negative influence on the number of permits issued.

The analysis is summarized in Table 5.1. Column 4 shows the coefficients and standard errors of the functional forms which fit the data best.[159] Notice that a high standard error in the non-linear model does not necessarily signal that the parameter does not have any influence. For example, the standard error for ι_2 is 0.1837, which means that the coefficient is not significantly different from zero. But in this case it does not mean that income does not have any influence on the loglikelihood, because ι_2 describes the horizontal location of the minimum of the parabola; a location at zero is one possible outcome, and does not change the significance of the variable 'income'. The same interpretation applies to the standard error of γ_3, which stands for the horizontal location of the normal curve that is added to the minimum *CV*. To determine the significance of the functional form of each explanatory variable it is necessary to undertake the likelihood ratio test; column 5 shows for each variable the test statistic and the degrees of freedom used.

The most surprising finding is the significantly negative coefficient of the tax parameter. This suggests that an increase in the adjusted tax differential will lead to a reduction in the number of housing units built. However, it is possible that the negative coefficient is due to certain town-specific effects, which are not captured by the model. Section 5.4 describes a fixed effects model, which uses, among other improvements, a dummy variable for each municipality. The tax coefficient of that model is slightly positive, but not significant.

For this model it is more informative to examine the technical aspects of the analysis: how much was gained by describing the influence of each explanatory variable as a nonlinear function, by taking account of serial correlation, and by using the more general method of describing heteroskedasticity. The last column of Table 5.1 shows the estimation results of a model that uses the same mean-variance relationship as the negative binomial analysis in

[159] The variance of the maximum likelihood estimator can be calculated in various ways (see Green, 1993, pp. 115-116). I used the diagonal of the inverse of the expected second derivatives matrix, which is the variance matrix of the first derivatives vector. This variance matrix is the matrix of the product of the gradient with its transpose, summed over all observations. To invert the matrix I used the program MATLAB.

Table 5.1 Nonresidential construction of whole units

	Functional Form		Model I Coefficient	χ^2 (DF)	Model II Coef-ficient
Density	$1 + \delta_1 (D_{i,t} - \delta_2)^2$ (Parabolic function)	δ_1	2.430E-8 (3.580E-8)	(162.16) (2)	2.857E-8 (3.69E-9)
		δ_2	14275.40 (1345.57)		13041.83 (758.9)
Income	$1 + \iota_1 (I_{i,t} - \iota_2)^2$ (Parabolic function)	ι_1	-0.4381 (0.1203)	52.28 (2)	-0.2941 (0.1064)
		ι_2	-0.2869 (0.1837)		-0.6099 (0.3049)
Tax	$1 + \tau\, ATD_{i,t}$ (Linear function)	τ	-0.0665 (0.0098)	9.56 (1)	-0.0654 (0.0181)
Residuals	$\rho_1 - (\rho_1 - 1)\, e^{\frac{-\rho_2(r_{t-1} - 1)}{\rho_1 - 1}}$ (Negative exponential function)	ρ_1	3.1938 (0.8046)	111.92 (2)	2.7244 (0.6678)
		ρ_2	0.2321 (0.0277)		0.2252 (0.0254)
CV	$CV_{min} + (1+\nu)\cdot\gamma_1 e^{-\gamma_2(\ln\mu - \gamma_3)^2}$ (Normal function)	γ_1	0.9554 (0.0250)		1.5239 (0.0464)
		γ_2	-0.0677 (0.0126)		-
		γ_3	-0.1175 (0.2208)		-
Adjustment-factor for *CV*		υ	0.0238 (0.0604)		-
Loglike-lihood			-4360.41		-4406.01

Note: Standard errors are shown in parentheses. Coefficients of yearly dummy variables are not shown.

Chapter 4, but includes the variables according to equation 5.5. Compared to the model in Chapter 4, this model uses six additional parameters to describe density, income, the residuals, and the mean-variance relationship, but saves two parameters by restricting the coefficients of ‘Population’ and ‘Months reported’ to 1. Similar to the model in Chapter 4, density and income are significant, while population change is not significant.[160] The coefficient of the tax parameter was also negative in the earlier model, but it was not significantly different from zero. The loglikelihood of that model was -4467.004; the refinements lowered the loglikelihood by 60.99 at the cost of four additional parameters, which is a significant improvement over the old model.

Comparison of Column 4 and Column 6 in Table 5.1 shows how much is gained if the new mean-variance relationship is used. The loglikelihood of the model in Column 4 is lower by 45.60, which means that the additional 3 parameters in the new mean-variance relationship are a significant improvement.

5.3.2 Residential construction of whole units:

To avoid a lengthy repetition of the procedure of finding the optimal functional forms, Table 5.2 on the following page reports only the results. Density is still significant, and a higher population density leads to a lower number of building permits in a municipality. Contrary to the analysis of nonresidential construction, income is introduced with a logistic function and has a significantly positive influence, as wealthier communities are able to build more houses. The analysis in Chapter 4 already indicated that population change has a positive influence on the decision to construct new residential buildings; population change is introduced with a hyperbolic function (compare with equation 5.6), and the introduction of population change improves the loglikelihood by 47.46. The lagged first residual again improves the loglikelihood noticeably. The positive coefficient on taxes does not pass a likelihood ratio test, and has a high standard error. The results of an analysis with the old mean-variance relationship are shown in Column 6; the loglikelihood of the new analysis is significantly higher than in the model with the old mean-variance relationship.

[160] See the third column of Table 4.5.

Table 5.2 Residential construction of whole units

	Functional Form		Model I Coefficient	χ^2 (DF)	Model II Coefficient
Density	$1 + \delta_1 e^{-\delta_2 D_{i,t}}$ (Exponential function)	δ_1	322.1007 (64.2664)	138.92 (2)	244.34 (75.4323)
		δ_2	9.442E-5 (1.400E-5)		8.564E-5 (2.301E-5)
Income	$(1 + e^{\iota_1 (I_{i,t} - \iota_2)})^{-1}$ (Logistic function)	ι_1	-6.6316 (1.0816)	74.66 (2)	-7.0041 (0.9196)
		ι_2	-0.2989 (0.0201)		-0.3377 (0.0179)
Population change	$\frac{1}{2}(1+\pi_1\dot{P}_{i,t}+\sqrt{(1+\pi_1\dot{P}_{i,t})^2+4\pi_2})$ (Hyperbolic function)	π_1	7.3765 (11.6067)	47.46 (2)	6.4608 (3.9224)
		π_2	1.4403 (4.3982)		2.5081 (2.7835)
Tax	$1 + \tau\, ATD_{i,t}$ (Linear function)	τ	-0.0251 (0.0631)	0.32 (1)	-0.0127 (0.02616)
Residuals	$\rho_1 - (\rho_1 - 1)\, e^{\frac{-\rho_2(r_{t-1} - 1)}{\rho_1 - 1}}$ (Negative exponential function)	ρ_1	1.0808 (0.0346)	259.54 (2)	1.1234 (0.0291)
		ρ_2	0.6474 (0.0392)		0.5450 (0.0305)
CV	$CV_{\min}+(1+\nu)\cdot\gamma_1 e^{-\gamma_2(\ln\mu - \gamma_3)^2}$ (Normal function)	γ_1	4.1468 (2.5684)		28.279 (0.4882)
		γ_2	-0.0087 (0.0045)		-
		γ_3	-9.0400 (6.0829)		-
Adjustment Factor for *CV*		υ	0.1528 (0.0428)		-
Loglikelihood			-13933.28		-14404.79

Note: Standard errors are shown in parentheses. Yearly dummies are not shown.

5.5 A fixed effects model

5.5.1 Setup of the model

The preliminary analysis in Chapter 4 indicated that a fixed effects model that uses a dummy variable for each municipality might yield a better fit of the data; after the introduction of the dummy variables into the regression equation, the independent variables ‘density’, ‘average income’, and ‘population change’ became insignificant. These dummies can be interpreted as the economic *status* S_i of municipality *i*. The status describes the economic condition of a municipality; the lower the status, the worse is this economic condition.[161] As there exists neither a *best* nor a *worst* condition, the status can be any real number.

Clearly, the incentives to construct new buildings in municipalities in depressed economic condition are very low, although even in a municipality in very depressed condition a building permit will be issued, and some construction will take place, once in a while. The number of building permits one can reasonably expect in a given year is therefore still larger than zero. Yet it will hardly be possible to find a general reason for these occasional permits, as they probably depend on unique events in a certain municipality, for example the replacement of a house that burned down in a fire. It is very unlikely that these construction decisions are influenced by tax advantages.

To incorporate the idea of such a ‘threshold’ status, below which all construction is unpredictable without detailed knowledge of the specific economic circumstances peculiar to each decision, the estimated equation must have a kink at this threshold, below which explanatory variables do not have any influence anymore. If the economic conditions are good, then the expected PMP $\mu_{i,t}$ will be determined by $\mu_{i,t}^* = S_i \cdot YD_t \cdot f_T(ATD_{i,t})$. If $\mu_{i,t}^*$ falls below the threshold value θ, then the expected PMP is given by the value of this threshold. To guarantee that the minimum predicted PMP is always positive, the threshold value is calculated as $\theta = e^{\kappa}$, where κ is the parameter to be estimated. This leads to the following equation for the expected PMP:

$$\mu_{i,t} = \begin{cases} \mu^*_{i,t} & if\ \mu^*_{i,t} \geq \theta \\ \theta & if\ \mu^*_{i,t} < \theta\ . \end{cases} \qquad (5.24)$$

[161] The model in Section 5.4 can be reinterpreted as a model which uses the status of a municipality. Instead of measuring the status as a dummy variable, the model in Section 5.4 measured the status as $S_{i,t} = f_D(D_{i,t}) \cdot f_I(I_{i,t}) \cdot f_P(P_{i,t})$.

However, this equation does not allow for a correct estimation of the coefficients. Assume that a municipality has such a low status that throughout the years for which data are available the best prediction about building permits that can be made is always determined by θ. First, it would be impossible to estimate the status of this municipality, because any value of S_i below $\theta/(YD_t \cdot f_T(ATD_{i,t}))$ will not affect the predicted number of permits anymore, and its most likely value cannot be determined. The second and more severe problem is that it will become impossible to estimate the correct value of θ. Any attempt to optimize θ by slightly changing its value would not be able to take account of the possibility that this change in θ will increase the predicted number of PMP for any of the municipalities with undeterminable status in at least one year above this threshold value, so that their impact on θ cannot be taken into consideration.

The solution of these problems is to restrict the status to values between 1 and infinity, where a status of 1 is interpreted as 'minimal' or 'undeterminably bad'. The status is not determined as an absolute value anymore, but it is expressed as a multiple of the minimal status.

The yearly dummies are rescaled to values in (0,1], which is achieved by dividing the values of all dummies by the value of the highest dummy (that is, the economically best year). Any municipality in bad economic condition can now achieve a maximum predicted value of 1 (calculated as 'status' multiplied by 'yearly dummy') in the best year, while the prediction in all other years is below 1. A predicted value of 1 or lower is therefore equivalent to the minimum prediction θ. A change in θ will now take the impact of all municipalities with status 1 into account, because a lower value of θ will lead to a predicted PMP that is larger than θ in these municipalities in at least one year. For all municipalities that are in better condition, the status dummy will be above 1 in at least one year, so that the yearly dummy and (potentially) the tax variable can influence the prediction. PMP can now be estimated as

$$\mu_{i,t} = \begin{cases} \mu^*_{i,t} \cdot \theta & \textit{if } \mu^* \geq 1 \\ \theta & \textit{if } \mu^* < 1 \ . \end{cases} \qquad (5.25)$$

To determine the expected number of permits, $\mu_{i,t}$ is multiplied by 'Population' and by 'Months reported'.

5.5.2 Serial correlation

The analysis in Section 5.4 revealed that the data is serially correlated, which makes it necessary to include a lagged residual as an explanatory variable. The equation of $\mu_{i,t}^*$ can be rewritten as

$$\mu_{i,t}^* = S_i \cdot YD_t \cdot f_T(ATD_{i,t}) \cdot f_r(r_{i,t-j}) \ , \qquad (5.26)$$

where $f_r(r_{i,t-;})$ is a function of the last observed residual, *j* periods ago. But the use of lagged residuals together with a fixed effects model is problematic: the status dummy for each municipality is an estimate of the average value of PMP, and any information about observed PMP that are either larger or smaller than the average PMP gives an indication of how to correct the average prediction for the remaining years. This correction is not due to serial correlation, but only to the fact that the status dummies are estimated with the same data as all the other coefficients. A potential solution of this problem would be to divide the data set into two parts, and to estimate the dummies from one part and the other coefficients from the remaining data. However, as the number of years of data for each municipality is small, especially the number of years in which cities adopted the two-rate tax, it is doubtful that this effective data reduction would actually improve the analysis.

It seems more promising to include a correction into the functional form of $r(r_{i,t-j})$. If the residual $r_{i,t-1}$ is equal to one, then last year's prediction was exactly equal to the observed number of PMP, and the best estimate for this year's PMP (independent of yearly dummy and tax parameter) is the same as it was for last year's PMP. If $r_{i,t-1}$ is smaller than 1, then last year's predicted PMP was too large, so that (with serial correlation) this year's PMP will most likely be smaller than otherwise predicted; the estimated value should be corrected downwards. But an observed PMP below the average PMP for all years for this municipality also means that the average for the remaining years needs to be corrected upwards, which translates into a higher predicted PMP for this year.[162] The reverse is true for $r_{i,t-1} > 1$; if last year's PMP was larger than predicted, then this year's PMP should be larger as well. On the other hand, if last year's PMP was above the average, then the average of the remaining years will be lower, which should reduce the prediction of this year's PMP. Clearly the two effects may point in opposite directions, and it is necessary to take take account of both of them to measure the serial correlation correctly.[163]

Serial correlation can be corrected in the same manner as in Section 5.4 with an exponential function which crosses the point (1,1), so that, if last year's PMP was predicted correctly, no correction will be made for this year's prediction. In the model in Section 5.4 lagged residuals were included only when they were obtained from the directly preceding observation. But even if the preceding observation is missing, an earlier lagged residual will still explain some serial correlation, as it would have influenced the prediction (and therefore

[162] Obviously it is also possible that last year's PMP was lower than predicted (last year's year dummy might have been very large), but that it was still above the long run average of this municipality; in this case both effects lead to a reduction of this year's expected PMP.

[163] This becomes even clearer if *all* lagged residuals are included into the equation as explanatory variables. Each residual gives an indication of how the status should be adjusted to yield a better prediction of the remaining years; the last year's PMP will then be 'predicted' with certainty.

the residual) of last year's number of permits if this number had been observed. In general, the explanatory power of a residual will decrease if the residual is used to predict serial correlation in years which do not follow immediately. It is therefore necessary to change equation 5.12 by including ν_2^n in the exponent, where *n* is the number of *missing* observations between two subsequent observations. As *n* increases, the exponent becomes smaller and the whole correction approaches 1. A lagged residual that was observed an infinite time ago would have no explanatory power, and would not affect the current prediction of construction activity. But if the explanatory power of the residual is lower, the variance of the estimated parameters should increase; this will be achieved by multiplying the excess coefficient of variation in equation 5.2 by $(1+\nu_1(1-\nu_2^n))$, so that the correction of *CV* approaches the maximum correction for the first year as the number of missing observations increases.

The correction for the change in the average value of PMP is simply a straight, downward sloping line that crosses the point (1,1). If the last observed number of permits was equal to the expected average number of permits, then no correction should take place, because the expected average number of permits for the remaining years would not change. If the last observed number of permits was equal to τ times this average number, where τ is the total number of observed permits for this municipality, so that all observed construction activity took place in a single year, then it should be predicted with certainty that no permit will be observed in the subsequent year. The line should therefore intersect the horizontal line at the sum of all permits that were observed for this municipality. The functional form of the correction is now given by

$$CORR_{i,t} = \left(\frac{\sum_{\tau=1}^{T} permit_{i,\tau} - permit_{i,t-j}}{\sum_{\tau=1}^{T} permit_{i,\tau} - E(permit_{i,t-j})} \right) , \tag{5.27}$$

so that the complete correction for serial correlation between time *t* and time *t-j* can be expressed as

$$f_r(r_{i,t-j}) = \left(\rho_1 - (\rho_1 - 1) \exp\left(\frac{-\rho_2 \cdot \nu_2^{n_{i,t}} \cdot (re_{i,t-j} - 1)}{\rho_1 - 1} \right) \right) \cdot CORR_{i,t} . \tag{5.28}$$

5.5.3 The different models and their estimates

The adjusted tax differential can be introduced in different ways. It is possible that the level of this year's tax differential has an impact on construction, but it is also possible that a tax impact comes not only come from the current value of ATD, but rather from the change, or even the lagged change in ATD from one year to the next. The following analysis focuses on both possibilities. Model I measures only the level effect by using only the value of the tax differential, so that $f_T(ATD_{i,t})$ is given by

$$f_T\,(ATD_{i,t}) \;=\; \exp(\;\tau\;\; ATD_{i,t}\;)\;. \tag{5.29}$$

Model II uses the value of the tax differential, the change, and the lagged change in the tax differential, which results in

$$f_T\,(ATD_{i,t}) \;=\; \exp(\;\tau_1\;\; ATD_{i,t}\;+\;\tau_2\;\;\Delta ATD_{i,t}\;+\;\tau_3\;\;\Delta ATD_{i,t-1}\;)\;. \tag{5.30}$$

Model III serves as a control model for the impact of the tax differential, as it is estimated without any tax effect (so that $f_T(ATD_{i,t}) = 1 \;\forall\; i,t$).[164] Models I to III are estimated with the new relationship between the mean and the coefficient of variation which was introduced in Chapter 3. Model IV is used as a control model for this relationship, because it is estimated with the usually used relationship $CV = [\;(1\text{-}\delta_{i,t}\;\mu_{i,t})/\mu_{i,t}\;]^{.5}$.

The coefficients and standard deviations of the examination of the 4 data sets are shown in Tables 5.3 and 5.4. The results are fairly consistent. Correcting for serial correlation always yields a significant improvement. The tax coefficients do not increase the loglikelihood of any data set by enough to pass a likelihood ratio test. This is supported by the standard deviations, which are not significant for all 4 data sets. Yet although the coefficients of the tax differential are not significant in any data set with the new mean-variance relationship, the data set for additions and alterations to nonresidential buildings (the last column in Table 5.4) shows a significantly positive coefficient for the level of the tax differential, and a significantly negative effect for the change of the tax differential when the old mean-variance relationship is used.

One might wonder if the separate examination of construction of whole units and construction of additions and alterations is justified, because the separate examination uses almost twice as many parameters. To test this issue I combined the two data sets of nonresidential construction into a single data set with 219 municipalities, and introduced two dummy vari-

[164] As the standard deviations can only be estimated from the *asymptotic* covariance matrix, it is helpful to use an additional likelihood ratio test of the tax coefficient to gain additional confidence about its significance.

Table 5.3 Residential construction of whole units and additions and alterations

		Residential construction of whole units				Residential additions and alterations			
		Model I	Model II	Model III	Model IV	Model I	Model II	Model III	Model IV
Tax:	τ	0.0682 (0.0714)	-0.0074 (0.0934)	-	0.0007 (0.0314)	-0.0591 (0.1139)	0.0397 (0.1623)	-	0.0251 (0.0591)
	τ_2	-	0.0505 (0.1693)	-	-0.0601 (0.1424)	-	-0.0373 (0.2663)	-	0.0807 (0.0624)
	τ_3	-	0.3213 (0.2188)	-	0.2768 (0.0732)	-	-0.0184 (0.2122)	-	0.0617 (0.0550)
Threshold:	θ	-5.0061 (0.1623)	-5.0061 (0.1616)	-5.0053 (0.1621)	-4.9423 (0.1824)	-4.8546 (0.2876)	-4.8546 (0.2874)	-4.8546 (0.2874)	-5.0938 (0.5162)
Serial Correlation:	ρ_1	0.0361 (0.0314)	0.0360 (0.0316)	0.0361 (0.0320)	0.0409 (0.0307)	0.0515 (0.0150)	0.0516 (0.0150)	0.0516 (0.0151)	0.2057 (0.0399)
	ρ_2	0.0834 (0.0436)	0.0837 (0.0422)	0.0837 (0.0444)	0.0778 (0.0297)	0.1376 (0.0257)	0.1376 (0.0257)	0.1376 (0.0259)	0.2803 (0.0214)
CV:	γ	1.5052 (0.0960)	1.5067 (0.0968)	1.5055 (0.0973)	26.3013 (0.4850)	0.6582 (0.0103)	0.6582 (0.0103)	0.6582 (0.0103)	63.1104 (0.9646)
	γ_2	20.2087 (6.6212)	20.2320 (6.6462)	20.4253 (6.6739)	-	-9.6456 (1.8207)	-9.6433 (1.8181)	-9.6315 (1.8146)	-
	γ_3	-0.4683 (1.1025)	-0.4780 (1.1081)	-0.4846 (1.1251)	-	4.5367 (0.0744)	4.5359 (0.0744)	4.5384 (0.0745)	-
	γ_4	-0.08938 (0.1560)	-0.08718 (0.1573)	-0.0864 (0.1615)	-	0.1448 (0.1095)	0.1451 (0.1094)	0.1444 (0.1092)	-
	υ_1	0.8739 (0.1250)	0.8768 (0.1231)	0.8717 (0.1274)	0.1404 (1.4083)	0.1154 (0.2307)	0.1166 (0.2302)	0.0732 (0.2382)	0.0001 (0.8500)
	υ_2	0.2578 (0.0760)	0.2583 (0.0758)	0.2541 (0.0755)	-	0.4253 (0.0560)	0.4258 (0.0560)	0.4235 (0.0557)	-
Loglikelihood		-14093.21	-14091.47	14093.74	-14422.48	-13361.81	-13361.77	-13362.27	-14147.61

Note: Standard errors are shown in parentheses. Coefficients of yearly dummies and status dummies are not shown.

Table 5.4 Nonresidential construction of whole units and additions and alterations

		Nonresidential construction of whole units				Nonresidential additions and alterations			
		Model I	Model II	Model III	Model IV	Model I	Model II	Model III	Model IV
Tax:	τ	0.0582 (0.1140)	0.0079 (0.1580)	-	0.0294 (0.1183)	0.0706 (0.1121)	0.1404 (0.1127)	-	0.2213 (0.0488)
	τ_2	-	0.1172 (0.2867)	-	0.0991 (0.1906)	-	-0.2467 (0.2090)	-	-0.1128 (0.0492)
	τ_3	-	0.0503 (0.2556)	-	0.0501 (0.2112)	-	-0.0982 (0.1621)	-	0.0281 (0.0412)
Threshold:	θ	-6.0657 (0.1603)	-6.0627 (0.1607)	-6.0568 (0.1608)	-5.9357 (0.1961)	-8.9841 (25.2846)	-8.9841 (17.9284)	-8.9841 (30.5434)	-9.3063 (1.0028)
Serial Correlation:	ρ_1	0.1525 (0.1866)	0.1530 (0.1870)	0.1495 (0.1838)	0.1643 (0.1672)	0.4908 (0.1223)	0.4909 (0.1219)	0.4908 (0.1215)	0.4176 (0.0831)
	ρ_2	0.0803 (0.0288)	0.0802 (0.0289)	0.0809 (0.0282)	0.0648 (0.0302)	0.2943 (0.0134)	0.2943 (0.0132)	0.2943 (0.0134)	0.2652 (0.0130)
CV:	γ	0.7731 (0.0266)	0.7726 (0.0266)	0.7731 (0.0266)	1.2214 (0.0504)	1.2722 (0.1031)	1.2610 (0.1034)	1.2773 (0.0783)	15.7258 (0.3293)
	γ_2	4.7764 (0.9241)	4.7742 (0.9136)	4.7803 (0.9272)	-	540.6885 (1038.26)	633.162 (1600.91)	639.625 (1658.23)	-
	γ_3	0.7363 (0.1340)	0.7327 (0.1343)	0.7392 (0.1340)	-	-0.1339 (0.6115)	-0.0701 (0.8621)	-0.1736 (0.6345)	-
	γ_4	-0.01975 (0.2975)	-0.01978 (0.2928)	-0.01972 (0.2990)	-	77.9122 (420.80)	92.4258 (484.05)	91.228 (423.23)	-
	υ_1	0.0011 (3.2727)	0.0023 (3.1969)	0.0002 (3.2570)	0.46873 (1.8026)	0.0004 (0.6577)	0.0003 (0.6497)	0.0003 (0.6543)	0.8188 (0.1564)
	υ_2	0.0952 (0.0939)	0.1008 (0.0946)	0.0955 (0.0941)	-	0.4115 (0.0715)	0.4085 (0.0713)	0.4083 (0.0710)	-
Loglikelihood		-4315.60	-4315.11	-4316.08	-4371.54	-8489.98	-8488.69	-8490.21	-8784.11

Note: Standard errors are shown in parentheses. Coefficients of yearly dummies and status dummies are not shown.

ables to capture the shift in the level and the slope of the regression line for additions and alterations. The analysis with the combined data set had therefore 246 parameters, and resulted in a loglikelihood of -13,556. The separate analysis that is shown in Table 5.4 yielded a loglikelihood of -8,487 with 242 parameters for nonresidential whole units, and a loglikelihood of -4,315 with 244 parameters for nonresidential additions and alterations.[165] The sum of both loglikelihoods is -12,802 for 486 parameters. The improvement in the loglikelihood of 754 passes the likelihood ratio test, so that the use of 240 additional parameters and the separate examination of the two data sets is justified.

The new mean-variance relationship of Model II results in a significantly higher loglikelihood than Model IV for all 4 data sets; the improvement for the two data sets of additions and alterations is much higher than for the data sets of construction of whole units. The parameter γ_2, which can be interpreted as a measure of the range of the added curve between the two inflection points, is between 4.96 and 20.42 for the first 3 data sets, while it is between 540.7 and 639.6 for the last data set. This indicates two possibilities: first, the data in the 4 data sets are created by different mechanisms, and one should not be surprised to find vastly different results. Second, it is possible that the added curve does not measure the true heteroskedasticity, but that is influenced by large outliers. The outliers in the first three data sets could be large enough that it becomes possible to improve the loglikelihood by locating a relatively narrow hump around these outliers, while the fourth data set either does not have any large outliers, or the outliers are too dispersed to be described by a single maximum. However, this raises the general question of what constitutes ‘true outliers’, because it is difficult to decide if a few observations must be excluded from the sample if they show a drastically different behavior than most of the other observations. Every data set in the analysis has a high variation among the observed number of permits within many municipalities, and without detailed information about the single municipalities it is not possible to determine ‘true outliers’ in most cases. If these observations are not excluded, then it is necessary to try to explain them as well as possible.

---

[165] The analysis of additions and alterations used two municipalities that were excluded from the data set of whole units because these municipalities reported their construction of whole units for only one year between 1980 and 1994, which would have led to difficulties with the estimation of the status dummy.