

MEASURING THE IMPACT OF POLLUTION ON HOUSING PRICES*

PREPARED BY COWGATE CONSULTING GROUP (CCG)

MAY 2021

EXECUTIVE SUMMARY

- The Federal Case against ChemXcom has determined that their plant No 14 was responsible for pollution and degradation of the environment over the period 2008 to 2011.
- CCG was commissioned to address the key question of determining the impact on housing prices that can be attributable to the pollution and degradation of the environment caused by ChemXcom.
- CCG will argue that the initial statistical analysis presented by the plaintiffs in this case is flawed in terms of application and in terms of the methodology used.
- Using the same data, a preferable regression methodology that controls for extra factors that determine housing prices proves that the initial claims for damages are highly exaggerated. Specifically, damages are estimated to be 46% less than those claimed by the plaintiffs.

* This is a purely fictitious report prepared for student use. Any similarity with actual persons or firms is accidental.

INTRODUCTION

The Federal Case against ChemXcom has determined that their plant No 14 was responsible for pollution and degradation of the environment over the period 2008 to 2011. In deliberations on possible damages caused to the residents of the surrounding community, there has been a submission by the plaintiffs that attempts to quantify the impact of the pollution on housing prices. Cowgate Consulting Group (CCG) was commissioned to comment on the statistical aspects of the submission and, in general terms, to address the key question of determining the impact on housing prices that can be attributable to the pollution and degradation of the environment caused by the ChemXcom.

We, CCG, will argue that the initial statistical analysis is flawed in terms of application and in terms of the methodology used. Using the same data, our regression methodology will show that the initial claims for damages are highly exaggerated.

A DIFFERENCE IN MEANS APPROACH

In the submission made by the plaintiffs their case for damages is in part based on a very simple statistical analysis of housing prices. Their approach was to divide the sample into two parts depending on whether or not the houses were located within a five kilometer radius of the chemical plant. The means of house prices in the two subsamples were then compared and the difference attributable to the pollution caused by ChemXcom.

We stress that this is what they *seem* to have done as we are unable to replicate their results. In any case, we dispute the validity of such an approach to the problem. This simple comparison of means does not control for other factors that may affect house prices and failure to account for these omitted factors is likely to bias the results obtained from a simple “difference-in-means” methodology. A more natural approach to modelling the implied multivariate relationship is the use of multiple regression analysis.

This study analyses a sample of 142 house prices based on completed sales during 2010. It is important to stress that these are the data used by the plaintiffs in their submission. Moreover, a reputable market research firm has collected them and we have no reason to question their accuracy.

The data used for this study were:

PRICE = sale price (\$'000)

AREA = area of house (square meters)

SIZE = size of block of land (square meters)

AGE = age of house (years)

NEAR = a dummy variable that is unity if house is within 5km of ChemXcom plant No. 14 and is zero otherwise.

A REGRESSION MODEL

A regression model has been used to analyse the impact of pollution on housing prices. In particular, we assumed a general model of the form:

$$(1) \quad PRICE_i = \alpha + \beta NEAR_i + \gamma X_i + u_i$$

where u is the random disturbance and X is a control variable to be chosen to maximize fit. This model specification has the advantage that it encompasses the plaintiff's simple approach as a special case. If γ is set to zero then the remaining coefficients have a simple interpretation. The mean house price for those houses that are not near to the plant will be represented by α while $\alpha + \beta$ will represent the mean house price for those houses that are near. Thus, the difference in means reported by the plaintiffs can be represented in our framework by the coefficient β . In this regression framework, the estimate of β (given $\gamma = 0$) should be the same as that reported as the difference in means.

When we estimate the simple regression model with only the *NEAR* dummy the resultant ordinary least squares results are given by:

$$(2) \quad \begin{aligned} \hat{PRICE}_i &= 131.9 - 40.0 NEAR_i \\ R^2 &= 0.17 \end{aligned}$$

Unlike the difference in means of \$48,000 reported by the plaintiffs we find a somewhat lower estimate of \$40,000. We are unable to determine how they arrived at their figure, which in itself must cast some doubt on their analysis. But as we have stressed there are even more fundamental difficulties with their approach.

The correlations between *PRICE* and other key determinants of housing prices are given in Table 1. These correlations were as expected. The bigger the house and the bigger the land, the higher the resultant price. Similarly, old houses were worth less than newer houses. Because *AREA* has the highest correlation with *PRICE* it is the best single predictor. On this basis it was the variable chosen to include in equation (1). Other potential explanatory variables are likely to be correlated with *AREA* causing multicollinearity problems if we were to add them to this specification. Consequently our preferred model includes *AREA* together with the *NEAR* dummy.

Table 1: Correlations with Price

	<i>AREA</i>	<i>SIZE</i>	<i>AGE</i>
Correlation	0.66	0.37	-0.60

Estimating this model by ordinary least squares produced the following results:

$$(3) \quad \hat{PRICE}_i = 47.9 - 26.1NEAR_i + 0.32AREA_i$$

$$R^2 = 0.51$$

Coefficient signs were as expected and all coefficients are precisely estimated. The *t*-ratio for the estimated *NEAR* coefficient was -4.32 and for *AREA* 9.76. The 5% critical value for the *t* distribution is 1.96 and hence both coefficients are significantly different from zero.

This model was deemed to be satisfactory because of the high R^2 value and because both explanatory variables are statistically significant. The large *t*-ratio for the estimated *AREA* coefficient indicates the importance of this determinant of housing prices and vindicates our approach. The estimated coefficient on *NEAR* in equation (3) represents our preferred estimate of the impact of pollution on housing prices. This estimate is \$26,100 and thus is 45.6% less than the estimated impact presented by the plaintiffs.

CONCLUSION

The econometric analysis presented here does not support the claims made on behalf of the plaintiffs in terms of the impact of pollution on housing prices in the affected community. Even when their approach is used we find a much smaller estimated impact. Using a preferable regression approach that controls for extra factors that determine housing prices the analysis supports an even smaller estimated impact.