After estimating various equations in our attempt to find the determinants of the log sale prices using TOTALBSMTSF and BEDROOMABVGR series as explanatory variables, we arrived at the following model:



Trying to explain the log sale prices using log of the total basement footage prompted us to restrict our sample only to observations for which positive footage was given. Thus, we removed the 37 houses with a reported basement of 0 square feet from our sample to allow for the log estimation comparison with the model using the level of footage (note that we still have a large sample left). Further improvement was achieved by expanding the bedroom series into several dummies with zero bedrooms in the basement as the benchmark.

The RESET test rejected the null of linearity, therefore we added the interaction term as well as the square of the basement footage to remedy for misspecification.



The Jarque-Bera statistic led us to rejecting the hypothesis of normally distributed residuals and the White test rejected their homoskedasticity. Thus, the equation was therefore augmented with heteroskedasticity consistent standard errors.

Using the restricted sample, the model without the log transformation slightly outperformed the log one due to its relatively higher R-squared and lower information criteria.

To summarize, our model explains approximately 47 per cent of the log sale price variation—the F test results in the overall significance of the explanatory variables. Except for the dummy indicating one bedroom at the basement level, all variables are also individually significant at the 5% level. The estimated effect of total basement footage is evaluated to have a positive impact on log sale prices which is decreasing both total basement footage and in the number of bedrooms at the basement level. Expanding the series bedroomabvgr led us to estimate that houses with the same basement footage are expected to cost more with each additional bedroom at the basement level, with the only exception being that the effect of bedroomabvgr=5 is estimated to be greater than bedroomabvgr=6. But overall, the results are in line with our intuition.



Characteristics:

* LotArea: 8500
* OverallQual: 7
* OverallCond: 5
* YearBuilt: 2003
* TotalBsmtSF: 1000
* GrLivArea: 1700
* FullBath: 2
* BedroomAbvGr: 3
* GarageCars: 2

point prediction: 165 648.3$

95% CI: 93 500$ - 294 000$









Point estimate: 248 462.47$

95% CI: 191 000$ - 324 000$