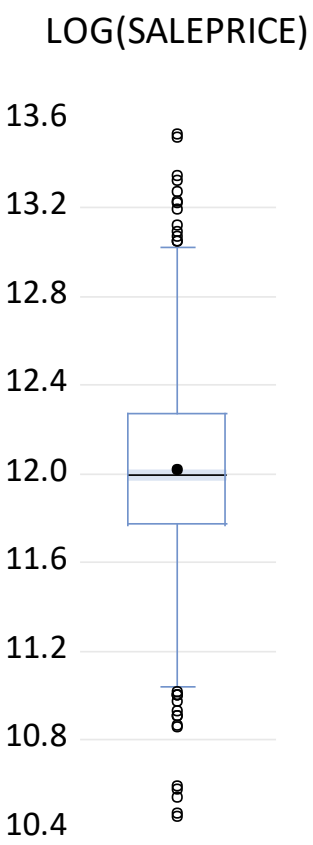
After estimating various equations in our attempt to find the determinants of the log sale prices using TOTALBSMTSF and BEDROOMABVGR series as explanatory variables, we arrived at the following model:

Dependent Variable: LOG(SALEPRICE)
Method: Least Squares
Date: 02/15/23 Time: 18:04
Sample: 1 1460 IF TOTALBSMTSF>0
Included observations: 1423
Huber-White-Hinkley (HC1) heteroskedasticity consistent standard errors and covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.56609	0.166749	63.36531	0.0000
TOTALBSMTSF	0.001444	8.52E-05	16.95561	0.0000
TOTALBSMTSF*BEDROOMABVGR	-0.000185	2.87E-05	-6.458093	0.0000
TOTALBSMTSF^2	-1.14E-07	1.69E-08	-6.778128	0.0000
BEDROOMABVGR=1	0.172852	0.136882	1.262778	0.2069
BEDROOMABVGR=2	0.346183	0.151187	2.289759	0.0222
BEDROOMABVGR=3	0.677476	0.171883	3.941498	0.0001
BEDROOMABVGR=4	1.035048	0.192749	5.369935	0.0000
BEDROOMABVGR=5	1.074292	0.230446	4.661784	0.0000
BEDROOMABVGR=6	1.070602	0.293147	3.652095	0.0003
BEDROOMABVGR=8	1.932345	0.354961	5.443830	0.0000

R-squared	0.476192	Mean dependent var	12.03691
Adjusted R-squared	0.472482	S.D. dependent var	0.393816
S.E. of regression	0.286031	Akaike info criterion	0.342264
Sum squared resid	115.5206	Schwarz criterion	0.382928
Log likelihood	-232.5206	Hannan-Quinn criter.	0.357453
F-statistic	128.3643	Durbin-Watson stat	1.981950
Prob(F-statistic)	0.000000	Wald F-statistic	167.9265
Prob(Wald F-statistic)	0.000000		

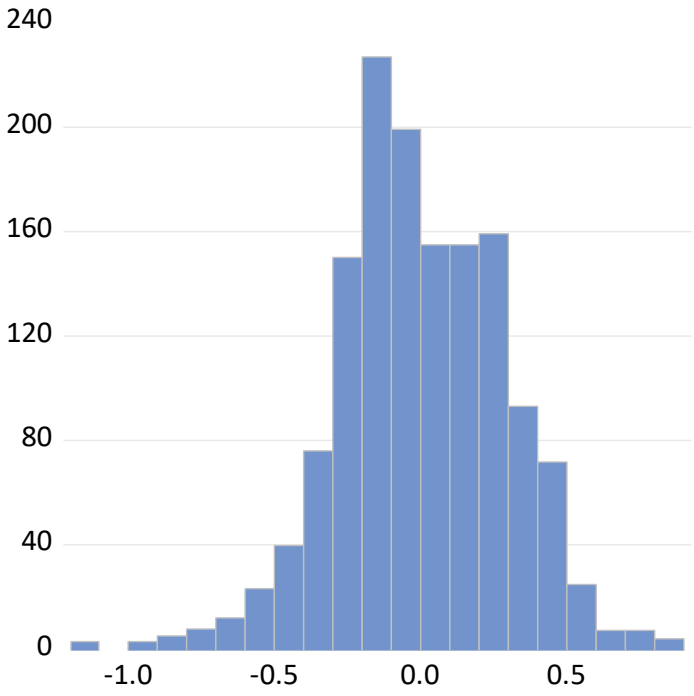


Trying to explain the log sale prices using log of the total basement footage prompted us to restrict our sample only to observations for which positive footage was given. Thus, we removed the 37 houses with a reported basement of 0 square feet from our sample to allow for the log estimation comparison 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.

Ramsey RESET Test
Equation: UNTITLED
Omitted Variables: Squares of fitted values
Specification: LOG(SALEPRICE) C TOTALBSMTSF @EXPAND(BEDROOMABVGR,@DROP(0))

	Value	df	Probability
t-statistic	11.26279	1413	0.0000
F-statistic	126.8504	(1, 1413)	0.0000
Likelihood ratio	122.3356	1	0.0000

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.



Series: Residuals
Sample 1 1460 IF TOTALBSMTSF>0
Observations 1423

Mean	2.84e-15
Median	-0.018864
Maximum	0.897786
Minimum	-1.188925
Std. Dev.	0.285023
Skewness	-0.179213
Kurtosis	3.618707
Jarque-Bera	30.31386
Probability	0.000000

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

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	6.759581	Prob. F(10,1412)	0.0000
Obs*R-squared	65.01022	Prob. Chi-Square(10)	0.0000
Scaled explained SS	83.81044	Prob. Chi-Square(10)	0.0000

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, but it is decreasing in both the 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. Overall, the results are in line with our intuition.

The big data selection methods allowed us to produce the following model:

Dependent Variable: LOG(SALEPRICE)
Method: Variable Selection
Date: 02/26/23 Time: 15:44
Sample: 1 1460 IF TOTALBSMTSF>0
Included observations: 1348
Number of always included regressors: 1
Number of search regressors: 180
Selection method: Stepwise forwards
Stopping criterion: t-stat forwards/backwards = 2/2
Note: final equation sample is larger than selection sample (rejected regressors contain missing values)

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
C	4.548114	0.478976	9.495489	0.0000
OVERALLQUAL	0.055420	0.004106	13.49587	0.0000
GRLIVAREA	0.000289	1.08E-05	26.67400	0.0000
YEARBUILT	0.002181	0.000214	10.21137	0.0000
OVERALLCOND	0.042240	0.003550	11.89807	0.0000
TOTALBSMTSF	0.000361	2.35E-05	15.38790	0.0000
TOTALBSMTSF^2	-1.02E-07	5.23E-09	-19.55966	0.0000
BSMTFINSF1	6.41E-05	9.99E-06	6.418091	0.0000
LOTAREA	2.13E-06	3.30E-07	6.460303	0.0000
CONDITION2="PosN"	-0.733682	0.080417	-9.123459	0.0000
NEIGHBORHOOD="Crawfor"	0.117916	0.016904	6.975529	0.0000
GARAGEAREA	6.99E-05	3.11E-05	2.244833	0.0249
KITCHENABVGR=1	0.057844	0.020421	2.832535	0.0047
SALECONDITION="Partial"	0.106092	0.015595	6.802858	0.0000
SALECONDITION="Normal"	0.054383	0.010728	5.069154	0.0000
HEATINGQC="TA"	-0.023066	0.007593	-3.037597	0.0024
EXTERIOR1ST="BrkComm"	-0.557822	0.105489	-5.287943	0.0000
MSZONING="RL"	0.378804	0.044997	8.418361	0.0000
BSMTEXPOSURE="Gd"	0.057557	0.011837	4.862522	0.0000
MSZONING="FV"	0.375857	0.052417	7.170473	0.0000
NEIGHBORHOOD="NridgHt"	0.117586	0.015308	7.681151	0.0000
NEIGHBORHOOD="BrkSide"	0.052147	0.017077	3.053603	0.0023
NEIGHBORHOOD="StoneBr"	0.128496	0.023076	5.568419	0.0000
MSSUBCLASS	-0.000391	9.47E-05	-4.130417	0.0000
MSZONING="RM"	0.309921	0.043669	7.096976	0.0000
MSZONING="RH"	0.365689	0.054268	6.738602	0.0000
SCREENPORCH	0.000215	5.10E-05	4.205392	0.0000
CONDITION1="Norm"	0.047814	0.009736	4.910811	0.0000
NEIGHBORHOOD="Edwards"	-0.067264	0.014124	-4.762458	0.0000
EXTERIOR1ST="BrkFace"	0.065321	0.017205	3.796602	0.0002
YEARREMODADD	0.000571	0.000231	2.469593	0.0137
BSMTFULLBATH=1	0.020552	0.007639	2.690272	0.0072
BEDROOMABVGR=5	-0.065977	0.026980	-2.445434	0.0146
NEIGHBORHOOD="NAmes"	-0.029918	0.009581	-3.122582	0.0018
CENTRALAIR="Y"	0.068325	0.016466	4.149538	0.0000
BSMTEXPOSURE="No"	-0.020455	0.007301	-2.801683	0.0052
FIREPLACES=1	0.025372	0.007145	3.550884	0.0004
FIREPLACES=2	0.039124	0.012915	3.029214	0.0025
NEIGHBORHOOD="NoRidge"	0.048598	0.019114	2.542543	0.0111
NEIGHBORHOOD="MeadowV"	-0.095998	0.033248	-2.887316	0.0040
EXTERIOR1ST="HdBoard"	-0.026850	0.008883	-3.022542	0.0026
CONDITION1="PosN"	0.064569	0.027529	2.345510	0.0192
NEIGHBORHOOD="Somerst"	0.056422	0.025066	2.250884	0.0246
BSMTCOND="Po"	-0.309344	0.107614	-2.874579	0.0041
GARAGECOND="Fa"	-0.057816	0.019745	-2.928096	0.0035
ROOFMATL="WdShngl"	0.129563	0.044853	2.888603	0.0039
LOTCONFIG="CulDSac"	0.031320	0.012024	2.604792	0.0093
EXTERIOR1ST="Plywood"	-0.033109	0.012494	-2.650042	0.0081
OPENPORCHSF	0.000123	4.89E-05	2.510862	0.0122
GARAGECARS=4	0.173945	0.051056	3.406962	0.0007
CONDITION1="RRAe"	-0.076396	0.034836	-2.193045	0.0285
GARAGECARS=3	0.081059	0.020217	4.009451	0.0001
GARAGECARS=2	0.031451	0.010597	2.967991	0.0031
POOLAREA	0.000197	7.34E-05	2.687505	0.0073
BSMTFINTYPE1="GLQ"	0.022030	0.008811	2.500230	0.0125
SALECONDITION="AdjLand"	0.233502	0.106318	2.196270	0.0283
LOWQUALFINSF	-0.000161	7.38E-05	-2.182850	0.0292
NEIGHBORHOOD="IDOTRR"	0.056188	0.024772	2.268183	0.0235
MOSOLD=5	0.020566	0.008364	2.458855	0.0141
STREET="Pave"	0.115869	0.052269	2.216757	0.0268
BEDROOMABVGR=1	-0.051166	0.018690	-2.737651	0.0063
TOTALBSMTSF*BEDROOMABVGR	-1.01E-05	4.45E-06	-2.275434	0.0230

R-squared	0.927869	Mean dependent var	12.06538
Adjusted R-squared	0.924448	S.D. dependent var	0.376864
S.E. of regression	0.103588	Akaike info criterion	-1.651892
Sum squared resid	13.79933	Schwarz criterion	-1.412430
Log likelihood	1175.375	Hannan-Quinn criter.	-1.562207
F-statistic	271.1926	Durbin-Watson stat	1.931152
Prob(F-statistic)	0.000000		

Selection Summary	
Number of selected regressors:	61
Added OVERALLQUAL	
Added GRLIVAREA	
Added YEARBUILT	
Added OVERALLCOND	
Added _1STFLRSF	
Added TOTALBSMTSF^2	
Added TOTALBSMTSF	
Removed _1STFLRSF	
Added BSMTFINSF1	
Added LOTAREA	
Added CONDITION2="PosN"	
Added NEIGHBORHOOD="Crawfor"	
Added GARAGEAREA	
Added KITCHENABVGR=1	
Added SALECONDITION="Partial"	
Added SALECONDITION="Normal"	
Added HEATINGQC="TA"	
Added EXTERIOR1ST="BrkComm"	
Added MSZONING="RL"	
Added BSMTEXPOSURE="Gd"	
Added MSZONING="FV"	
Added NEIGHBORHOOD="NridgHt"	
Added NEIGHBORHOOD="BrkSide"	
Added NEIGHBORHOOD="StoneBr"	
Added MSSUBCLASS	
Added MSZONING="RM"	
Added MSZONING="RH"	
Added SCREENPORCH	
Added CONDITION1="Norm"	
Added NEIGHBORHOOD="Edwards"	
Added EXTERIOR1ST="BrkFace"	
Added YEARREMODADD	
Added BSMTFULLBATH=1	
Added BEDROOMABVGR=5	
Added NEIGHBORHOOD="NAmes"	
Added CENTRALAIR="Y"	
Added BSMTEXPOSURE="No"	
Added FIREPLACES=1	
Added FIREPLACES=2	
Added NEIGHBORHOOD="NoRidge"	
Added NEIGHBORHOOD="MeadowV"	
Added EXTERIOR1ST="HdBoard"	
Added CONDITION1="PosN"	
Added NEIGHBORHOOD="Somerst"	
Added BSMTCOND="Po"	
Added GARAGECOND="Fa"	
Added ROOFMATL="WdShngl"	
Added LOTCONFIG="CulDSac"	
Added EXTERIOR1ST="Plywood"	
Added OPENPORCHSF	
Added GARAGECARS=4	
Added CONDITION1="RRAe"	
Added GARAGECARS=3	
Added GARAGECARS=2	
Added POOLAREA	
Added BSMTFINTYPE1="GLQ"	
Added SALECONDITION="AdjLand"	
Added LOWQUALFINSF	
Added NEIGHBORHOOD="IDOTRR"	
Added MOSOLD=5	
Added STREET="Pave"	
Added BEDROOMABVGR=1	
Added TOTALBSMTSF*BEDROOMABVGR	

*Note: p-values and subsequent tests do not account for variable selection.

Even though the model includes too many variables for it to be easily justified by economic theory, we chose this one as it had the highest adjusted R squared and lowest information criteria compared to other selection methods.

To elaborate on our selection process, we considered all the EViews VARSEL options: uni-directional, stepwise, swapwise, combinatorial, GETS and Lasso. The output of the lastly mentioned method is given below for illustration.

Dependent Variable: LOG(SALEPRICE)
Method: Variable Selection
Date: 02/25/23 Time: 15:33
Sample: 1 1460
Included observations: 1423
Number of always included regressors: 41
No search regressors
Selection method: Lasso
Lambda at minimum AIC: NA
Huber-White-Hinkley (HC1) heteroskedasticity consistent standard errors and covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.850735	0.740982	6.546361	0.0000
TOTALBSMTSF^2	-1.03E-07	4.65E-09	-22.06037	0.0000
BEDROOMABVGR=1	-0.040210	0.023352	-1.721890	0.0853
BEDROOMABVGR=3	0.053458	0.013942	3.834365	0.0001
BEDROOMABVGR=4	0.086241	0.028451	3.031195	0.0025
BEDROOMABVGR=5	0.047465	0.051331	0.924691	0.3553
BEDROOMABVGR=6	0.107856	0.074947	1.439098	0.1503
TOTALBSMTSF*BEDROOMABVGR	-4.43E-05	9.72E-06	-4.554507	0.0000
LOTAREA	2.63E-06	4.84E-07	5.434233	0.0000
OVERALLQUAL	0.060714	0.004891	12.41324	0.0000
OVERALLCOND	0.063222	0.003917	16.13904	0.0000
YEARBUILT	0.002850	0.000364	7.827759	0.0000
TOTALBSMTSF	0.000540	3.51E-05	15.36765	0.0000
GRLIVAREA	0.000284	2.26E-05	12.55591	0.0000
GARAGECARS	0.056512	0.007818	7.228093	0.0000
NEIGHBORHOOD="Blueste"	-0.097278	0.020248	-4.804317	0.0000
NEIGHBORHOOD="BrDale"	-0.150588	0.028197	-5.340647	0.0000
NEIGHBORHOOD="BrkSide"	0.008454	0.033253	0.254239	0.7993
NEIGHBORHOOD="ClearCr"	0.074787	0.030983	2.413766	0.0159
NEIGHBORHOOD="CollgCr"	0.029998	0.014264	2.103044	0.0356
NEIGHBORHOOD="Crawfor"	0.130214	0.028531	4.563943	0.0000
NEIGHBORHOOD="Edwards"	-0.056609	0.035923	-1.575847	0.1153
NEIGHBORHOOD="Gilbert"	0.033877	0.016536	2.048725	0.0407
NEIGHBORHOOD="IDOTRR"	-0.128856	0.050955	-2.528812	0.0116
NEIGHBORHOOD="MeadowV"	-0.140536	0.029902	-4.699872	0.0000
NEIGHBORHOOD="Mitchel"	-0.025551	0.023753	-1.075704	0.2822
NEIGHBORHOOD="NAmes"	-0.000172	0.021521	-0.007990	0.9936
NEIGHBORHOOD="NoRidge"	0.109734	0.025695	4.270669	0.0000
NEIGHBORHOOD="NPkVill"	-0.064604	0.021731	-2.972855	0.0030
NEIGHBORHOOD="NridgHt"	0.115876	0.024712	4.689049	0.0000
NEIGHBORHOOD="NWAmes"	-0.026447	0.020268	-1.304883	0.1921
NEIGHBORHOOD="OldTown"	-0.090898	0.031162	-2.916962	0.0036
NEIGHBORHOOD="Sawyer"	-0.007727	0.025525	-0.302720	0.7621
NEIGHBORHOOD="Somerst"	0.069680	0.018555	3.755221	0.0002
NEIGHBORHOOD="StoneBr"	0.144373	0.034405	4.196278	0.0000
NEIGHBORHOOD="SWISU"	0.010096	0.035183	0.286964	0.7742
NEIGHBORHOOD="Timber"	0.039547	0.022467	1.760199	0.0786
NEIGHBORHOOD="Veenker"	0.091703	0.039467	2.323530	0.0203
BSMTQUAL="Fa"	-0.083714	0.035300	-2.371487	0.0179
BSMTQUAL="Gd"	-0.084019	0.017443	-4.816865	0.0000
BSMTQUAL="TA"	-0.093802	0.020799	-4.509958	0.0000

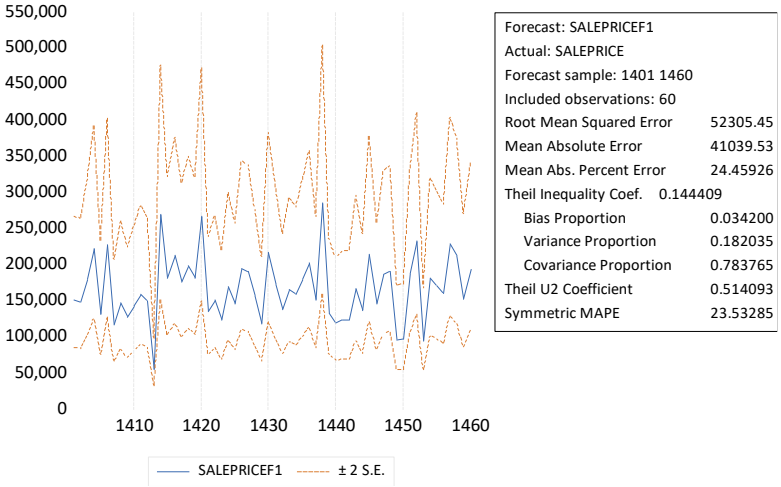
R-squared	0.893986	Mean dependent var	12.03691
Adjusted R-squared	0.890918	S.D. dependent var	0.393816
S.E. of regression	0.130068	Akaike info criterion	-1.213129
Sum squared resid	23.38025	Schwarz criterion	-1.061561
Log likelihood	904.1414	Hannan-Quinn criter.	-1.156516
F-statistic	291.3512	Durbin-Watson stat	1.996329
Prob(F-statistic)	0.000000		

Estimating the equation of both the small OLS model and our preferred big data selection method model on a restricted sample of 1400 observations and forecasting the remaining 60 produced the following results:

Dependent Variable: LOG(SALEPRICE)
Method: Least Squares
Date: 02/26/23 Time: 15:35
Sample: 1 1400 IF TOTALBSMTSF>0
Included observations: 1364
Huber-White-Hinkley (HC1) heteroskedasticity consistent standard errors and covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.55120	0.168704	62.54269	0.0000
TOTALBSMTSF	0.001457	8.60E-05	16.93883	0.0000
TOTALBSMTSF^2	-1.15E-07	1.64E-08	-7.030142	0.0000
TOTALBSMTSF*BEDROOMABVGR	-0.000188	2.89E-05	-6.504244	0.0000
BEDROOMABVGR=1	0.177237	0.137866	1.285577	0.1988
BEDROOMABVGR=2	0.351380	0.152538	2.303554	0.0214
BEDROOMABVGR=3	0.688115	0.173496	3.966183	0.0001
BEDROOMABVGR=4	1.053017	0.194390	5.417028	0.0000
BEDROOMABVGR=5	1.091746	0.232553	4.694603	0.0000
BEDROOMABVGR=6	1.094296	0.295667	3.701116	0.0002
BEDROOMABVGR=8	1.963760	0.358286	5.480981	0.0000

R-squared	0.482186	Mean dependent var	12.03724
Adjusted R-squared	0.478358	S.D. dependent var	0.395414
S.E. of regression	0.285586	Akaike info criterion	0.339488
Sum squared resid	110.3502	Schwarz criterion	0.381570
Log likelihood	-220.5307	Hannan-Quinn criter.	0.355240
F-statistic	125.9905	Durbin-Watson stat	1.969939
Prob(F-statistic)	0.000000	Wald F-statistic	161.3406
Prob(Wald F-statistic)	0.000000		



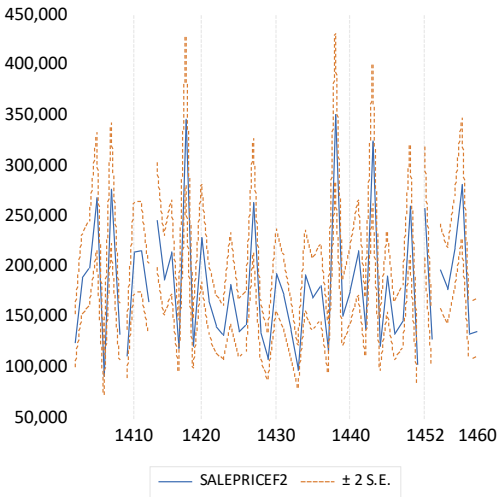
Forecast: SALEPRICEF1	
Actual: SALEPRICE	
Forecast sample: 1401 1460	
Included observations: 60	
Root Mean Squared Error	52305.45
Mean Absolute Error	41039.53
Mean Abs. Percent Error	24.45926
Theil Inequality Coef.	0.144409
Bias Proportion	0.034200
Variance Proportion	0.182035
Covariance Proportion	0.783765
Theil U2 Coefficient	0.514093
Symmetric MAPE	23.53285

Dependent Variable: LOG(SALEPRICE)
Method: Least Squares
Date: 02/26/23 Time: 15:46
Sample: 1 1400 IF TOTALBSMTSF>0
Included observations: 1293

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.351081	0.493708	8.813064	0.0000
OVERALLQUAL	0.056948	0.004197	13.56735	0.0000
GRLIVAREA	0.000288	1.11E-05	26.02281	0.0000
YEARBUILT	0.002152	0.000219	9.834463	0.0000
OVERALLCOND	0.041648	0.003631	11.47066	0.0000
TOTALBSMTSF	0.000361	2.39E-05	15.07218	0.0000
TOTALBSMTSF^2	-1.01E-07	5.31E-09	-19.05047	0.0000
BSMTFINSF1	6.79E-05	1.02E-05	6.634755	0.0000
LOTAREA	2.08E-06	3.31E-07	6.279720	0.0000
CONDITION2="PosN"	-0.733507	0.080605	-9.100057	0.0000
NEIGHBORHOOD="Crawfor"	0.118171	0.017772	6.649279	0.0000
GARAGEAREA	6.99E-05	3.21E-05	2.178973	0.0295
KITCHENABVGR=1	0.059049	0.020680	2.855325	0.0044
SALECONDITION="Partial"	0.107595	0.015893	6.770018	0.0000
SALECONDITION="Normal"	0.058336	0.010982	5.311868	0.0000
HEATINGQC="TA"	-0.021881	0.007787	-2.810018	0.0050
EXTERIOR1ST="BrkComm"	-0.549122	0.105660	-5.197065	0.0000
MSZONING="RL"	0.379742	0.045191	8.403083	0.0000
BSMTEXPOSURE="Gd"	0.056856	0.012108	4.695792	0.0000
MSZONING="FV"	0.386135	0.053209	7.257003	0.0000
NEIGHBORHOOD="NridgHt"	0.111101	0.015491	7.171795	0.0000
NEIGHBORHOOD="BrkSide"	0.051490	0.017602	2.925303	0.0035
NEIGHBORHOOD="StoneBr"	0.125492	0.023204	5.408288	0.0000
MSSUBCLASS	-0.000376	9.63E-05	-3.905397	0.0001
MSZONING="RM"	0.305422	0.043758	6.979790	0.0000
MSZONING="RH"	0.365504	0.054411	6.717405	0.0000
SCREENPORCH	0.000215	5.19E-05	4.139598	0.0000
CONDITION1="Norm"	0.048116	0.009950	4.835647	0.0000
NEIGHBORHOOD="Edwards"	-0.077549	0.014574	-5.321142	0.0000
EXTERIOR1ST="BrkFace"	0.063527	0.017481	3.634004	0.0003
YEARREMODADD	0.000697	0.000237	2.936438	0.0034
BSMTFULLBATH=1	0.019880	0.007828	2.539556	0.0112
BEDROOMABVGR=5	-0.062990	0.027093	-2.324976	0.0202
NEIGHBORHOOD="NAmes"	-0.031375	0.009825	-3.193232	0.0014
CENTRALAIR="Y"	0.070561	0.016856	4.186187	0.0000
BSMTEXPOSURE="No"	-0.018684	0.007456	-2.505946	0.0123
FIREPLACES=1	0.024785	0.007297	3.396641	0.0007
FIREPLACES=2	0.039389	0.013262	2.970170	0.0030
NEIGHBORHOOD="NoRidge"	0.048868	0.019646	2.487433	0.0130
NEIGHBORHOOD="MeadowV"	-0.094120	0.033509	-2.808812	0.0051
EXTERIOR1ST="HdBoard"	-0.029348	0.009086	-3.230147	0.0013
CONDITION1="PosN"	0.064831	0.027636	2.345844	0.0191
NEIGHBORHOOD="Somerst"	0.046855	0.026190	1.789051	0.0739
BSMTCOND="Po"	-0.304276	0.107772	-2.823325	0.0048
GARAGECOND="Fa"	-0.041347	0.020126	-2.054401	0.0401
ROOFMATL="WdShngl"	0.123464	0.044975	2.745172	0.0061
LOTCONFIG="CulDSac"	0.025880	0.012303	2.103582	0.0356
EXTERIOR1ST="Plywood"	-0.034417	0.012827	-2.683148	0.0074
OPENPORCHSF	0.000105	4.97E-05	2.105998	0.0354
GARAGECARS=4	0.170740	0.051325	3.326629	0.0009
CONDITION1="RRAe"	-0.074035	0.034950	-2.118321	0.0343
GARAGECARS=3	0.075785	0.020759	3.650692	0.0003
GARAGECARS=2	0.027958	0.010830	2.581545	0.0100
POOLAREA	5.73E-05	8.50E-05	0.674103	0.5004
BSMTFINTYPE1="GLQ"	0.017771	0.009022	1.969778	0.0491
SALECONDITION="AdjLand"	0.249010	0.106585	2.336261	0.0196
LOWQUALFINSF	-0.000124	7.66E-05	-1.615742	0.1064
NEIGHBORHOOD="IDOTRR"	0.057437	0.024891	2.307569	0.0212
MOSOLD=5	0.019342	0.008636	2.239755	0.0253
STREET="Pave"	0.108600	0.052404	2.072359	0.0384
BEDROOMABVGR=1	-0.045140	0.019319	-2.336611	0.0196
TOTALBSMTSF*BEDROOMABVGR	-9.14E-06	4.52E-06	-2.021938	0.0434

R-squared	0.928459	Mean dependent var	12.06546
Adjusted R-squared	0.924914	S.D. dependent var	0.378384
S.E. of regression	0.103684	Akaike info criterion	-1.648168
Sum squared resid	13.23380	Schwarz criterion	-1.400517
Log likelihood	1127.541	Hannan-Quinn criter.	-1.555224
F-statistic	261.8987	Durbin-Watson stat	1.923209
Prob(F-statistic)	0.000000		

The RMSE of the smaller model was 52305.45$, while the bigger model had RMSE of 19356.74$. When comparing the two, we need to take note of the difference in included observations, which was 60 for the smaller model and 55 for the bigger one due to unavailability of data. However, it is still likely that the big data methods specified model performs better in terms of forecasting as the magnitude of its RMSE is less than half of the small OLS model.



Forecast: SALEPRICEF2	
Actual: SALEPRICE	
Forecast sample: 1401 1460	
Adjusted sample: 1401 1460	
Included observations: 55	
Root Mean Squared Error	19356.74
Mean Absolute Error	13071.35
Mean Abs. Percent Error	7.473611
Theil Inequality Coef.	0.050087
Bias Proportion	0.021383
Variance Proportion	0.016045
Covariance Proportion	0.962571
Theil U2 Coefficient	0.188976
Symmetric MAPE	7.439990

Finally, using the given house characteristics,

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

the big data model produced a

- point prediction of 127 541.9$,
- with 95% CI in the range from 94 100$ to 173 000$,

where the rest of the necessary explanatory variables were added to the observation 1461 based on the most frequent characteristics of the rest of the sample.

For comparison, the small OLS model produced a point prediction 165 648.3$ and 95% CI from 93 500$ to 294 000$, where the width of the interval illustrates that the bigger model produces relatively more certain forecasts.

The boxplot of SalePrice below shows that the forecast is near the mean and median price of the houses sold in our data.