[Re] A New Approach to Estimating the Production Function for Housing

Aniruddha Dutta, Mansi Rathod, Mayur Gupta, Pradeep Prabhakar, Sreeja Vishaly Manohar

23 February, 2020

A reference implementation of:

Epple, Dennis, Brett Gordon, and Holger Sieg. 2010. *“A New Approach to Estimating the Production Function for Housing.”* American Economic Review, 100 (3): 905-24.DOI: 10.1257/aer.100.3.905

# Introduction

The purpose of the paper is to do a replication of the paper “A New Approach to Estimating the Production Function for Housing” (Epple, Gordon, and Sieg [2010](#ref-epple2010new)) that estimates the production function for housing. Based on two factors - the observed variation in land prices and housing values per unit of land, the original paper has provided an algorithm to identify the housing supply function per unit of land. This in turn is used to derive an estimate of the underlying production function. The approach yielded plausible estimates for the price elasticity of housing supply per unit of land, based on data from Allegheny County in Pennsylvania. In the paper,we propose a reference implementation by adding two additional models to replicate an estimation function that relates land price and home value per unit land and produce the corresponding tables and plots.Estimating housing production functions is challenging, as the quantity and price per unit of the housing services are not observed by the econometrician. Replicating the study helps us to understand how the underlying production function is estimated by treating prices and quantities of housing services as latent variables, without relying on strong functional form assumptions.

# Methods

The current methodology provided by the author for estimating house price supply function uses parametric models. The author in the paper aims to establish a function which relates land price and home value per unit land . The parametric functions used for the purpose is OLS with different variations in it. The paper has used multiple transformation on data to produce linear, log linear and polynomial linear models to best identify the relation.

Since, the function forms the root of further analysis and calculation, it is imperative that we replicate the study of these models to test the robustness and the assumptions behind the models. We decided to split the method of implementation in two direction. The first method is to use a different model for replication and the other is changing the loss function used in the model.

For the first case of different model, we decided use Generalized Linear Model with different distribution families. Using different families allows us to verify the condition of normality of error in the models used by authors. We noticed that the gaussian distribution family with log independent and dependent variable produced a line that fitted the best with the data and the corresponding coefficients were very similar to the log linear model used by the author.

In the case of different loss function, we wrote a function code to implement log linear regression with gradient descent loss function. The gradient descent loss function provides a more flexible approach because of the presence of hyperparameters *learning rate* and *number of iteration*. The author of the paper has not explored this methodology, therefore we had no information about the value of hyperparameters *learning rate* and *number of iteration*, therefore we ran a simulation study to determine the hyperparameters that delivered values of coefficients very similar to that of the models used by the author

The paper uses linear function in estimating the coefficients of alpha, beta and gamma in the HNIP estimator. The linear model is realized using Ordinary Least Squares (OLS) method. The estimator is replicated by replacing the OLS based linear function with a bayesian regression function. Bayesian linear regression is used because it is more flexible to further model development and can directly model posteriors of derived/calculated quantities. “Bolstad” package provides a bayes.lm function which has been used to complete the replication. Upon analysis, we can conclude that the replicated estimator produced coefficients that are similiar to the actual paper.

The paper uses Ordinary Differential Equations for calculating the supply functions, . For this package ‘odesolve’ has been used in the base paper. Package ‘odesolve’ has been removed from the CRAN repository. The more recent package ‘deSolve’ completely supersedes odesolve and we have installed the same.

## Data

The dataset used by the paper (Epple, Gordon, and Sieg [2010](#ref-epple2010new)) comes from Allegheny County Web site (Pennsylvania state) maintained by the Office of Property assessments, which is based on the appraisals conducted by Sabre Systems and uses land value appraisals as well as property appraisals. Using geocoding, property assignment to travel zones and subsetting only for functioning residential properties, the count came down to 358,677 properties. Of these properties,the authors used a subsample of housing units that were built after 1995 for the estimation procedure which was 6362 houses. The data captures some important metrics for the houses such as the price of land, value per unit land, plot area, travel time and its geo-location (latitude and longitude). We will be using the same data for our replication in this paper. The descriptive statistics of the dataset with 6362 residential properties is shown below.

Table 1 - Descriptive statistics of Residential data  
===================================================================  
Statistic Mean Median St. Dev. Min Max   
-------------------------------------------------------------------  
Value per unit of land, v 21.44 14.29 26.91 0.15 366.62   
Price of land, p 3.32 2.28 3.86 0.05 41.75   
Lot area (sq. ft.) 26,756.10 15,507 52,196.96 540 1,207,483  
Travel time (minutes) 29.12 30 9.47 1 59   
-------------------------------------------------------------------

# Results

The original paper had estimated the r(v) function which characterizes the relationship between the price of land and the value of housing per unit of land. Using OLS for log-linear, linear, quadratic and cubic models, it reports the estimates of the equilibrium locus. We have replicated the same results using gaussian log-linear model and gradient descent log linear model. Results are enclosed in Table 1. For both of these two models, they fit the main features of the residential data substantially well and their performance is similar to the earlier four models. All the p-values were calculated using heteroskedasticity-robust standard errors and they were significant at 1% level. For effective and comprehensive comparisons of all the six models used, we have also enclosed a plot of all our six models. Please refer to figure 1. As iterated, we see r(v) functions form all the six models yielding similarly shaped slopes. We were able to establish the monotonicity condition of $r(\v)$ to be satisfied for the two replication models as well under all the polynomial estimation cases for the range of values of v observed in the data.

From the base paper, the log linear models establishes the change in value of land (variable: v) with unit change in land price (variable: pland). It calculates of slope to be $log(\v)$ as 0.909 and the intercept coefficient as -1.605. Upon replicating the same intentions through Generalized Log Linear Model, we get the slope to be 0.9186 and the intercept as -1.63. the slope and intercept coefficient from Gradient Descent Log Linear Model are 0.9184 and -1.6311 respectively. Thus, we observe similar performance in calculating the housing per unit of land from all of them.

The r(v) function replicated using bayesian regression is similear to the one estimated in the original paper using OL based Linear Regression. The original paper produced a mean of 0.192999, -0.001682 and 0.000012 and standard deviation of 0.014704, 0.014704 and 0.000005 for V, V^2 and V^3 coeffiecients respectively. The replicated bayesian model produced a mean of 0.192481,-0.001661 and 0.000012, and standard deviation of 0.014845, 0.000579 and 9 5.0e-06 respectively. this proves that the replicated bayesian model produced outputs similar to the original paper.

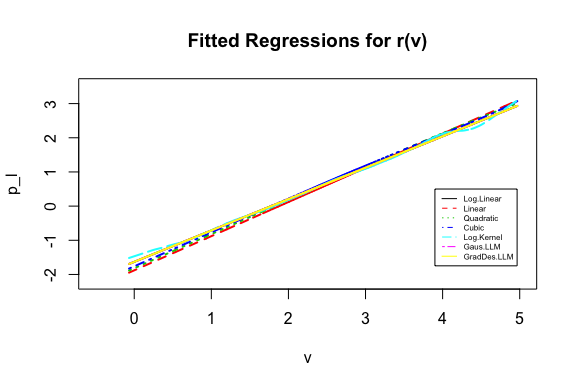


Figure 1: Rplot\_with\_AllModel

# Conclusion

We have discussed in the paper (Epple, Gordon, and Sieg [2010](#ref-epple2010new)) about how to estimate the production function for housing. In order to do, we have worked on calculating the estimation functions and supply functions. We encountered bit blockage in calculating the ordinary differential equations since (Epple, Gordon, and Sieg [2010](#ref-epple2010new)) calculated it using the now-obsolete ‘odesolve’ package. We were successful in replacing its dependencies with the ‘deSolve’ package. Since price and quantities for house are rarely observed individually, therefore we treated this two metrics as latent variables. Thus, we were able to calculate the production function without resorting to strong functional form assumptions. Of the three main functions: estimation, supply and production, we have replicated the first two functions with different linear model and loss functions. Compared to the four models used in (Epple, Gordon, and Sieg [2010](#ref-epple2010new)) our Gaussian Generalized Log Linear Model and Gradient Descent Log Linear Model had similar performance. The result is illustrated in Table 1. The main insight behind the approach is that the observed variation in land prices and housing values per unit of land is sufficient to identify the housing supply function per unit of land. Given the supply function per unit of land it is straightforward to recover the underlying production function. The production functions for housing plays an important role in conducting applied general equilibrium policy analysis. Many urban policies—such as school voucher programs, property tax reforms, urban development policies are likely to affect the demand for housing and residential sorting patterns.

# References

1.@epple2010new

Epple, Dennis, Brett Gordon, and Holger Sieg. 2010. “A New Approach to Estimating the Production Function for Housing.” *American Economic Review* 100 (3): 905–24.