

Assessment code:

ECN219

Registration no:

190146551

Title:

What is the impact of rail infrastructure in Kent, with special focus on the impact of HS1? – Group coursework

Word count:

5748

*Please enter the **assessment code**, your **student registration number**, the **assignment title** (and **question number** if applicable), and **word count** above.*

The word count does not include this coversheet, but please check the assessment instructions from the module leader if the word count includes all diagrams, appendices and references

By submitting this assessment to be marked and graded within the Department of Economics you are confirming that it is all your own work, except where specifically stated, and you have read and understand the University's rules relating to the use of unfair means. The University uses a range of strategies to detect unfair means, including Turnitin, which helps detect plagiarism.

Please tick the optional statements if appropriate.

☒ I consent to the Department using this essay for educational purposes in Economics modules by reproducing it as an example of a real student essay for other students to see and to work from, providing all identifiers including my student registration number are removed.

☐ I wish to indicate that I am a student with dyslexia or other written communication difficulties.

Enter your submission on the following page.

Please retain double line spacing to facilitate the online marking and feedback process. Please leave margins at 2cm and font as 12 point Arial.

*Please submit in **Word** or **PDF** format and check your file is under the **100MB** size limit.*

What is the impact of rail infrastructure in Kent, with special focus on the impact of HS1?

Abstract

Previous research has studied the impact on rail infrastructure on a change in house prices. These reports were focused on other countries, such as: (Dorantes, et al., 2011) on new public transport infrastructure and house prices in Madrid and (Bohman & Nilsson, 2016) on the impact of regional commuter trains on property values in Sweden.

The reports looking at the London area focused on other aspects of the economy, such as an increase in labour productivity. We have found no previous research in the last thirty years linking commuting times to London with house prices in Kent.

Our study focuses on studying train stations in Kent, with focus on the impact of HS1 rail, and the relationship between commuting time to London and housing prices in Kent. We hypothesis that as commuting time to London decreases house prices will increase.

We used HM Land Registry data, the Trainline, MSOAs and LSOAs, as well as information from ORR and HS1 to calculate a many variable used to make four models.

We had three main findings: house prices and time to travel to London have a negative linear relationship, proximity of a house to a train station can considerably affect the house price, construction of HS1 had a positive impact on house prices.

Our results show that increased commuting time has negative on house prices in areas surrounding London. Indicating government investment in more rail infrastructure can encourage migration out of London, reducing inequality and bringing better economic growth.

Introduction

High-quality infrastructure with high coverage is crucial for improving the national economic development and national happiness shown by an index. When Trevithick

developed the first rail train in 1804, no one thought that the railway system would become the economic lifeline of the world powers such as Britain and China. Nor would it have been imagined that the London underground would open in 1863 ([TFL 2020](#)), making the Metropolis bustling with underground traffic.

With the improvement of the railway system, the UK achieved rapid urbanisation with many moving from rural towns and villages in the 19th century to cities, especially London. London, one of the world's financial centres, is full of opportunities and wealth. Thus, people are more inclined to live in London or cities around London to take advantage of these opportunities. However, London's high housing prices are likely to be prohibitive for ordinary families wanting to move there. House prices there are almost double that of surrounding cities.

We want to explore the relationship between infrastructure and house prices. However, the classification of infrastructure is broad, so we narrowed our outlook to the impact of rail infrastructure on house prices in Kent, with a focus on the construction of the HS1 line. The house prices in London have been continuously rising, pricing people out of London. This makes alternative locations outside of London such as Kent more attractive. 46.5% of the population who immigrated to Kent in 2019 were from London. In Kent, people can still look for London job opportunities or maintain their job in London. As many in Kent work in London, we assume one of the most important impacts on house prices will be the ease of access to London. Thus, Kent's house price has been chosen as the research object. We have chosen to focus on HS1 as it will make it easier to infer causality rather than just correlation between rail infrastructure and an increase in house prices.

The research question of our topic is "What is the impact of rail infrastructure in Kent, with special focus on the impact of HS1?". And the hypothesis is that the reduction in commuting time from Kent to London will cause house prices to rise. Because commuting

to London more easily is a big selling point for a Kent house, as most people are willing to spend more money to save time travelling to London.

Reading previous literature, we found that many papers on infrastructure are broad and generalised. These essays tend to be from the perspective of macroeconomics to analyse the pros and cons of infrastructure for the overall economy. Few papers focus on one infrastructure project. We could find few papers discussing a specific infrastructure and analysing its impact on housing prices, particularly in the UK and there was very little study into the region of Kent. Thus, the things discussed in this paper will develop on the previous papers to focus on a small geographical area of Kent.

Three regression models were brought into STATA calculations to explore Kent's house prices and the time it takes to reach the closest train station and hence the time it takes to get to London. As well as the impact of proximity to a train station and to calculate the impact of HS1 on Kent's house prices. The multiple regression analysis on time taken to travel to London via rail and the impact of HS1 in Kent will obtain more accurate results. The result confirmed the hypothesis that as the time from Kent to London decreases, the house price will get higher and higher. They have an inverse relationship.

According to the result, it proves indirectly that the construction of the railway system will increase the economic level of the city, as prices rise more wealthy individuals will live there and increase economic activity. Therefore, it will be a good choice for the government to increase expenditures on railway construction from the surrounding counties of London to the city of London. First of all, this can solve the problem of excessive valuation of London real estate prices. London houses have become a luxury that ordinary people cannot match, rather than a "house" in a popular sense. Secondly, the population density of London can be evacuated to solve the problem of urban overcrowding. However, this policy will not only improve the situation in London. From a

macro perspective, more railway construction will bring more jobs to a wider range of people and the unemployment rate will decrease. This may be particularly attractive to the current government who have stated their aim to “level up the UK” ([Financial Times 2020](#)). The increase in housing prices due to railway construction will also provide the government with higher taxes, through stamp duty which could cover part of the deficit caused by public funding of rail infrastructure.

In the process of confirming our results through data analysis, we did encounter some limitations. In the process of modelling and calculation, we did not have sufficient resources to achieve a significantly high r-squared. It also underestimated the commuting time for families to the station. But these problems have been improved by refining calculations and model data. For future research, we would recommend researching income inequality in London and its surrounding areas, they can analyse our research when classifying house prices into categories such as size and housing type.

Literature Review

Some reports that we researched studied the impact on rail infrastructure on a change in house prices, but they were not specific to our area of focus and instead reports were modelled around cities in other countries. There has been no previous research to assess the impact of commuter rail on the change in house prices in the Kent region.

Many reports researched the effects that the rail infrastructure in and around London had focused on other aspects of the economy such as an increase in labour productivity in the capital and a geographical shift in the labour supply. More specifically, the research in these reports focused more on how the availability of commuter rail helped employees in counties surrounding London, such as Kent, travel into the capital at a decreased time, which contributed to higher productivity seen in firms. Additionally, some studies correlated accessibility to commuter rail with a geographical shift in labour. These studies elaborated

on how commuter rail had a strong impact in local areas resulting in employees relocating to a house closer to a station to reap its benefits.

These reports helped contribute to our motivation of this study as well as give us sufficient evidence to compose a hypothesis. However, we had seen no previous study within the last thirty years linking commuting times to London with house prices in Kent.

When researching similar reports for this study, it was necessary to refine our research down to aspects of literature that were most significant to our study. Our findings from these papers studied the impacts of HS1, commuting time and transport innovations on house prices, migration of labour and labour productivity.

One paper closely linking to our study analysed the relationship between new public transport infrastructure and house prices in Madrid ([Dorantes, et al., 2011](#)). This study looked at how the newly built Metro Line 12, otherwise known as Metrosur, in Madrid, Spain, influenced housing prices in this region. The methodology in this report explained how hedonic pricing was categorised into three subcategories and used in a linear regression model to assess the impact of the Metro Line 12 on housing prices in Madrid. The subcategory that was the most significant to our study was 'Accessibility Characteristics,' such as transport and other services. This hedonic house price function was expressed in their linear regression model which found that better access to Metro Line 12 led to increased house prices.

The impact of regional commuter trains on property values in Sweden ([Bohman & Nilsson, 2016](#)) was another paper researched for this report. The methodology and findings in this report were a significant contribution to our motivation. The report focuses on how single-family households of a varying household income and house price were affected by regional commuter trains. The income levels of each household were measured and split into percentiles. The lowest percentile represented those of the lower middle class and

those in the highest percentile represented those of the highest class. The findings suggested that single-family households in the lowest quantile benefitted the most from having access to public transport. The findings also suggested that for the richest quantile, there was a little, even negligible, effect of availability to public transport when choosing a house. This was perhaps because the lower quantile studied in this report were more likely to commute into a city to work to provide for their income. This is significant to our motivation in our report as we can study just how important the railway system is to the labour market who commute from Kent into London.

Research under HS1 was conducted in a 2009 report which analysed the economic benefits of High Speed 1 ([Arter, 2009](#)). This report studied the change in demand for stations operating under HS1 such as Ebbsfleet and Ashford International caused by a change in commuting times. This report oversaw an estimated journey time savings of 10 - 40% compared to existing services. Additionally, the report estimated a total of £3.8 billion transport benefits. This report has found that the construction of HS1 has allowed more accessibility from different areas in Kent into the city centre of London. This report is useful at explaining our HS1 variable in our linear regression, as we expect this form of rail infrastructure to have a significant boost in housing prices in hs1 accessible areas. This is due to the reduced commuting times into London as well as the expected £62m - £360m rise in additional earnings for commuters living in areas surrounding the HS1 stations.

We have chosen to research the impact of time taken to commute to London on average house prices in different areas within the Kent region. We expect to observe that a quicker commuting time will lead to an increase in house prices, defining commuting time as the combined factors of proximity to a train station and the time taken commuting on a rail network into London. Our hypothesis will focus on studying a variety of train stations in Kent, with a special focus on the HS1 rail as one of our variables. We felt that this was a

relevant and significant topic as there had been no previous reports explaining the relationship between commuting time to London and housing prices in Kent.

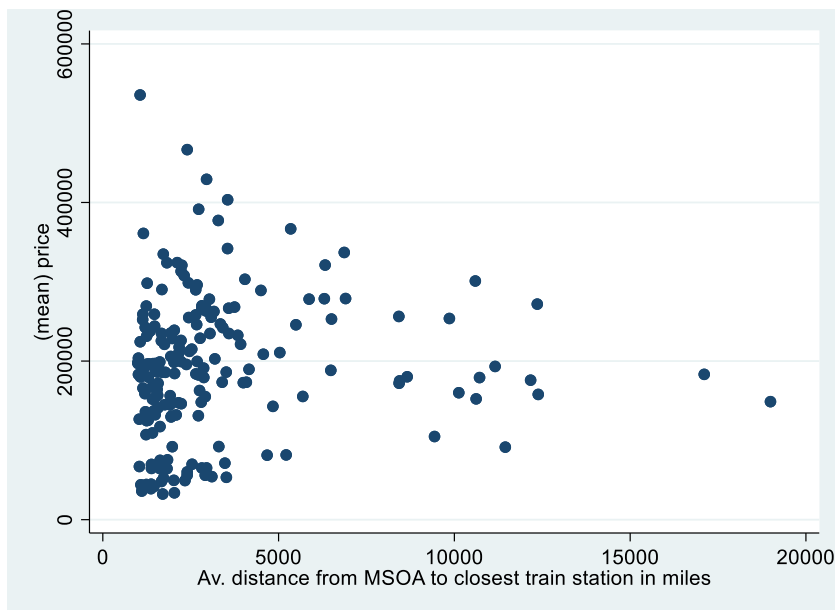
Data Description

We used three main sources of data: HM Land Registry for housing information, the Trainline for average train times and middle layer super output areas (MSOA) and lower layer super output areas (LSOA) data for distance from a house to the closest train station.

HM Land Registry registers the ownership of property. Registering all housing transactions in England since 1995 collecting information such as postcode, time, price, property type, etc. HM Land Registry is a non-ministerial department of the UK Government ([HM Land Registry 2021](#)). HM Land Registry receives a budget from HM Treasury which is balanced by the fees the Land Registry charges. The department is politically independent meaning the data is unlikely to be tampered with and can largely be trusted.

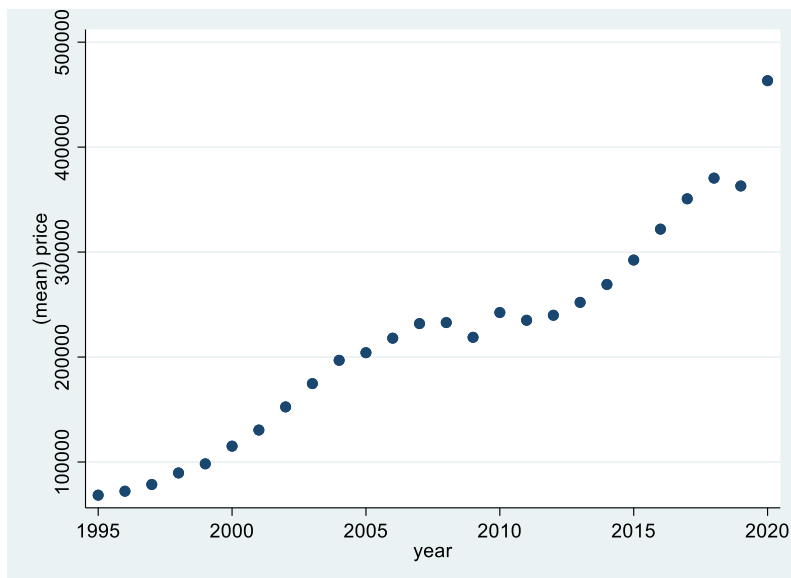
Variables from HM Land Registry: **Price** in £s, **Postcode**, **Property type** (detached, semi-detached, terraced flat, other), **new_building** (newly built property or not), **estate_type** (freehold/leasehold), **region**, **year**, **month**, **id** (identification number)

Figure 1



(figure 1) demonstrates that the further the distance from a train station, the lower the house price is. There is an inverse relationship.

Figure 2



(Figure 2) shows that house prices have tended to increase over time with a notable decrease in 2009 coinciding with the great financial crash.

Only property price in Kent were the focus of this study so all properties in a region other than KENT were omitted. We also omitted data about the month the transaction took place instead focusing on yearly changes.

We used the 30 most used train stations in Kent ([Office of Rail and Road 2020](#)) merged with the land registry data making the new variable **distance** (the distance to a train station from a specific postcode) calculated at the MSOA level, using **lsoa11cd** (LSOA) and **msoa11cd** (MSOA) as well as making a new variable indicating the name of the closest train station **closest_station**.

Having distance as a straight-line measurement is not the most useful and realistic way of measuring distance to a train station. If we had more time and resources, we would use road distance to station as most will drive or cycle to their closest station.

The station usage data comes from Office of Rail and Road (ORR) a non-ministerial government department and independent regulator. The data is therefore likely to be unmanipulated and accurate. The data used is outdated being from 2016/2017, as it out of date more recent data or better still an average over time would be more useful.

MSOA data was designed to improve the reporting of small area statistics and are built up from groups of output areas, originally released in 2004 for England and Wales. They are the standard units for presenting local statistical information across National Statistics. (ONS 2011). The fact they are used extensively by the government means that it is likely they are useful and accurate for geographical based queries in the UK.

We used HS1's website ([highspeed1 2020](#)) to find stations that are on the HS1 line. Using this to make the dummy variable **hs_station**, whether the closest station is a HS1 station. A problem with this data is, that is an extremely small sample size with only two HS1 stations in Kent. This means it is difficult to extrapolate any meaningful correlations. The fact HS1 was opened in 2007 was used by the dummy variable **hs1**.

time_in_minutes (train time) uses data from Trainline ([Trainline 2021](#)). Trainline is an independent digital rail and coach ticketing platform in Great Britain. It provides the website services for eight of the 20 UK train operating companies to sell tickets as well as its own brands.

It was difficult to find a dataset for train times so, we manually had to search average weekday time to any London station from each of the 30 Kent stations we chose. This is one reason we only focused on 30 stations due to time constraints. A disadvantage of using average weekday time is it can be skewed by outliers. An example of this is the

station High brooms, where most train journey times were under 1hr 30 but, one train was 3h 18m skewing the average.

Data from Google maps was used to calculate the variables **time_to_station** and used in the variable **time_to_london**. Google maps is one of the most used websites/apps for directions and has many datapoint to accurately predict journey times. However, we did calculate the journey time while in the coronavirus pandemic so times may be higher in “normal” times when there is likely to be more traffic.

Methodology

We calculated the variable **distance** (the average distance in miles between each house and the closest train station) by first, reducing the number of stations to the 30 most used train stations in Kent. Then the distance was calculated at the MSOA level, using **lsoa11cd** (LSOA) and **msoa11cd** (MSOA), in a straight line between each postcode in Kent in the dataset and the closest train station. We also made a new variable **closest_station** which is the name of the closest station out of the 30 chosen.

time_to_station calculated through Google maps using postcode data from the Land registry and **closest_station**. A random number generator was used 8 times to choose 8 different postcodes. The predicted journey time was calculated for these postcodes and then predicted time/**distance** was averaged across the 8 houses to get a constant time/distance ratio to be used for the whole data set, the average distance per minute being 390.92. This ratio was then combined **distance** for the whole dataset to create **time_to_station**. **time_to_station** was added to **time_in_minutes** to create the new variable **time_to_london** to show the total journey time to London from “door to door”.

$$(1) \ln(\text{price}) = b_0 + b_1 \ln(\text{time_to_london}) + m$$

Equation (1) illustrates the fundamental relationship between house prices and the time taken to travel to London, via rail, which will provide us with a framework on which to build on. The purpose of this simple regression model is to understand whether the relationship is positive or negative. It is expected that this model will bias the absolute value, however the sign of the coefficient will provide us with a basic understanding of the correlation between the dependent and independent variable. The range of house prices in our dataset is substantially large, meaning it is a highly skewed variable, therefore we transformed the house price variable by taking the natural logarithm, ensuring that the distribution of the variable is now more 'normal'. As a result, we transformed the **time_to_london** variable into its natural logarithm, this was to simplify the interpretation as the coefficient will now be the estimated percentage change.

$$(2) \ln(\text{price}) = b_0 + \gamma + b_1 \ln(\text{time_to_london}) + b_2(\text{hs_station}) + b_3(\text{hs_station} \times \text{hs1}) + m$$

γ = Fixed effects which consist of the year, ranging from 1996 to 2020, the property type, whether the property is newly built and the estate type.

Equation (2) will provide us with a more precise estimation of the relationship between House Prices and the time taken to travel to London, via rail. Fixed effects are added to the model to increase the r^2 and in turn the goodness of fit. Furthermore, adding the 'year' variable as a fixed effect will capture the time-series variation as house prices naturally increase overtime. Additionally, economic shocks could perhaps skew the data therefore the 'year' variable is paramount in ensuring accurate results. We interact the **hs_station** variable with **hs1**. **Hs_station** is a dummy variable which determines whether the house was affected by HS1, the metric being whether the **closest_station** is a HS1 station (Ebbsfleet or Ashford). Furthermore, the **hs1** variable is a dummy variable which recognises whether the observation was pre or post the construction of HS1, the construction being in 2009. The interaction of these two variables captures the impact of

HS1 on house prices. Moreover, HS1 provides us with a great opportunity to further test the extent to which train lines impact house prices. This model will bias the HS1 coefficient due to the **time_to_London** variable, therefore equation (3) will test the impact of HS1 in Kent.

$$(3) \ln(\text{price}) = b_0 + \gamma + b_1(\text{prox_traintime}) + b_2(\text{hs_station}) + b_3(\text{hs1}\#\text{hs_station}) + b_4(\text{prox_london}) + b_5(\text{hs1}\#\text{prox_london}) + b_6(\text{hs_station}\#\text{prox_london}) + b_7(\text{hs1}\#\text{hs_station}\#\text{prox_london}) + m$$

γ = Fixed effects which consist of the year, ranging from 1996 to 2020, the property type, whether the property is newly built and the estate type.

Equation (3) will test the impact of HS1 on house prices within Kent. **prox_traintime** is the proximity between the train station and London, determined by the equation, $1 / \text{time_in_minutes}$ and **prox_london** is the proximity between the house and London, formed by the equation, $1 / \text{time_to_london}$. An interaction between **hs1**, **hs_station** and **prox_london** is formed to test the impact of HS1. This interaction captures the entire impact of HS1, it includes whether the observation was pre or post the construction of HS1, whether the house's closest station is a High Speed station and the proximity to London via rail.

$$(4) \ln(\text{price}) = b_0 + \gamma + b_1 \ln(\text{time_to_london}) + m$$

γ = Each year ranging from 1996 to 2020.

Equation (4) is a panel data regression model, panel data being the repeated survey of a sample across a time period. Therefore, the dataset was collapsed to convert the dataset into a set of means, meaning that we can now apply a panel data model to our dataset to further analyse the impact of rail on house prices.

Gauss-Markov assumptions:

Assumption 1: Linear in parameters

The model is linear in parameters such that:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

Each of the models in this section satisfy this assumption.

Assumption 2: Random Sampling

We sourced as much data as we possibly could, which in turn ensures that the sample is random.

Assumption 3: Collinearity

There is collinearity present in the multiple regression models, this is due to the models containing many variables and it is inevitable that some of the variables will be correlated. However, our models having no perfect collinearity is what ensures that this assumption is not violated.

Assumption 4: Zero Conditional Mean

It is unrealistic to assume that the zero conditional mean has not been somewhat violated, there are many factors that affect house prices and we do not have the capabilities to capture every aspect of house prices. However, this creates a risk of endogeneity. There are variables not included, that could impact the variables that are included. For example, the **distance** variable was calculated in a straight line from the house to the train station. Yet, roads and means of transport could influence this variable. There is no realistic solution to this therefore we must interpret the results with caution.

Assumption 5: Homoscedasticity

In the results section, we will test whether heteroskedasticity is present in our regression models. If so, we will alleviate the issue by using Stata to robust the standard errors, ensuring that this assumption is satisfied.

Strengths and Limitations:

Our dataset contains many variables which should lead to our multiple regression models having a relatively high R-squared. This should give us an opportunity to provide a fairly accurate conclusion to our hypothesis and understand the degree to which reductions in commuting time to London impact house prices.

However, we are somewhat limited with how accurate our results can be. This is because the number of factors that can impact house prices are extremely large and beyond our capabilities. Furthermore, to provide a more accurate conclusion, it could perhaps be important to examine all surrounding areas of London rather than just Kent.

Results

Model 1: Simple Regression Model

STATA Estimated Equation 1:

$$\ln(\text{price}) = 13.34 + -.311 \ln(\text{time_to_london})$$

This model has a sample size of 719,434. As predicted, the linear relationship between house prices and the time it takes to travel to London, via rail, is negative. This complies with recent literature as Debrezion, Pels and Rietveld's paper on the impact of Rail Transport on Real Estate Prices concluded that "if the distance between the location of the real estate and the nearest railway station doubles, the real estate price declines by 1 per cent" ([Debrezion et al, 2010](#)).

Due to the model being log-log, the interpretation is as follows: a 1% increase in the time it takes to get to London via rail, results in a 0.31% decrease in house prices. The relationship may not seem significant; therefore, it may be more fitting to interpret the results as showing that a 10% increase in the time it takes to get London results in a 3.1% decrease in house prices. The P-value also signifies that the coefficient is statistically significant *see table 1*.

However, the R-squared is 0.0301, this is particularly low and demonstrates that the independent variable explains very little of the variance in the dependent variable.

Therefore, it is crucial that we now use a multivariate regression model to achieve a more accurate coefficient.

Model 2: Multiple Regression Model

STATA Estimated Equation 2:

$$\text{Log}(\text{price}) = 13.22 + \gamma + -.396(\text{time_to_london}) + -.181 (\text{hs_station}) + \\ -.0323(\text{hs_station} \times \text{hs1})$$

γ = Fixed effects which consist of the year, ranging from 1996 to 2020, the property type, whether the property is newly built and the estate type.

The multiple regression model contains more explanatory variables across the same observations used in the model (1). Model (2) has an R-squared value of 0.6893, this illustrates that the model explains 68.93% of the variance of the response data around its mean. In model (2), β_1 has a higher absolute value, demonstrating that model (1) placed downward bias on the time to London coefficient. The coefficient on time to London, β_1 , is -.396, meaning that a 10% increase in the time it takes to travel to London via rail results in a 3.9% decrease in house prices. Furthermore, the P-value indicates that the result is statistically significant *see table 2*. In terms of house prices, this figure is significant and

suggests that the geographical location of a house, in respect to a Train Station, can considerably affect the house price. This interpretation is consistent with economic theory, as (Miles and Sefton, 2020) demonstrate that “in a growing economy the speed of travel improvements is crucial to the evolution of land and house prices”. It is worth noting that there are economic disadvantages to house prices rising. The Guardian claimed that “rising house prices were making it increasingly difficult for young people to get on to the property ladder, Aviva said, after average UK house prices had risen more than 50% to £279,000 between 2005 and 2015” (Monaghan, 2016, para. 3).

Model 3: Multiple Regression Model

STATA Estimated Equation 3:

$$\ln(\text{price}) = 11.03 + \gamma + 15.03(\text{prox_traintime}) + .232(\text{hs_station}) + -.0385(\text{hs1\#hs_station}) + 17.56(\text{prox_london}) + -.502(\text{hs1\#prox_london}) + -28.70(\text{hs_station\#prox_london}) + 1.015(\text{hs1\#hs_station\#prox_london}) + \mu$$

γ = Fixed effects which consist of the year, ranging from 1996 to 2020, the property type, whether the property is newly built and the estate type.

This model has an R-squared value of 0.6912 and the P-values display statistical significance *see table 3*. β_7 , the coefficient on the interacted three desirable variables, is positive. This indicates that the construction of HS1 had a positive impact on house prices, illustrating that improvements of infrastructure can increase house prices. This finding concurs with recent literature. Helen Simpson (2019) conducted a test analysing the economic impacts of HS1 in Kent, she found that as a result of HS1, “a 10-minute reduction in travel time in Kent was associated with around 1.5% - 3% higher house prices”.

Some of the variables do not agree with recent literature and our prediction that HS1 will increase house prices. β_3, β_5 and β_7 are all negative, these coefficients are peculiar and could perhaps be skewed by a multitude of factors. For example, the sample size is relatively small due to HS1 only impacting 2 stations in Kent, only 87,341 of the 719,434 observations were affected by HS1.

Model 4: Panel Regression

STATA Estimated Equation 4:

$$\ln(\text{price}) = 12.70 + \gamma + -.359\ln(\text{time_to_london}) + \mu$$

γ = Each year ranging from 1996 to 2020.

The sample size in model (4) is 373,931 and once again, the coefficient for time to London is negative, further supporting our hypothesis, the P-value also shows statistical significance *see table 4*. Exploring the relationship between house prices and the time it takes to travel to London, via rail, has consistently led us to conclude that the longer it takes to travel to London, via rail, the lower the house price, holding everything else constant.

Test for Heteroskedasticity:

A Breusch-Pagan test was conducted on models 2 and 3 to test for heteroskedasticity. The chi-squared value on each model respectively was, 4622.85 and 5063.25. These values are notably high and thus satisfy the alternative hypothesis, meaning that the error terms do not have constant variance. Ultimately demonstrating that heteroskedasticity is present. In order to alleviate this issue, we robust the standard errors in each of the models.

Conclusion

To conclude, our hypothesis where we predicted to see a decrease in time spent commuting to London showing an increase in house price was found to be negatively correlated; supported by our regression models. Our three regression models: simple regression, multiple regression and multiple regression biasing HS1 were found to have R squared values of 0.0301, 0.6893 and 0.6912 respectively. These R square values demonstrated a mild to fairly strong correlation. Additionally, our fourth model, using panel regression, also calculated a negative correlation between our independent and dependent variables. These four regressions have given us sufficient evidence to conclude that the time taken to commute into London ultimately impacted house prices.

The information provided by our regressions should be considered when designing future policies regarding house prices and infrastructure in London. Our results show a clear impact that commuting time has on house prices in areas surrounding London. In other words, our results indirectly indicate that more complete railway construction in London and surrounding counties will bring about better economic growth to cities, otherwise house prices will not increase. Therefore, it could be argued that it is beneficial for the government to consider the economic and social benefits of building more railway construction that connects counties surrounding London to the city centre. An example of this is HS1 and its impact on house prices. This form of rail infrastructure has had a positive effect on house prices in Ebbsfleet and Ashford and has shown these areas to be increasingly desirable to commuters. This could have included commuters who previously lived in a location that is closer to the city centre but have recently moved due to improved travel times available outside of London. Since there has been a migration movement in recent years from living in London to moving out of London, the Government investing in more rail infrastructure like HS1 and the Crossrail can help to allow more migration from London to outer areas such as Kent. This can solve a variety of problems such as the imbalance of house prices in the London housing market. If more commuters can migrate

from Greater London to areas outside of London it could eventually disperse the distorted house market prices in London. This in return can help more people to afford housing in these areas. Furthermore, this process can help the Government to tackle any potential issues of overcrowding in the suburban areas of the city. By investing into the transport system in counties surrounding London, the Government can manipulate the flow of migration out of London by making these areas more accessible and desirable, consequently reducing the highly densely populated capital.

Future researchers who may wish to compose a study focusing on income inequality in London and surrounding areas may analyse our research when breaking house prices down into categories such as size and house type. Smaller households with a lower annual income may benefit from new policies that help support their income. Investing in more infrastructure can help give them more access to other surrounding locations in and around London where they may not have had previous access to, for example, if they had no ownership of a car. Offering lower-income households more methods of transportation can in return increase accessibility to more jobs in nearby areas, which can help to reduce income inequality that is prevalent in London.

We faced some limitations when gathering results and data for our hypothesis. One limitation was how we modelled the houses in our initial regression. By modelling hedonically, some internal and external factors that affect a house price were included which were not relevant to our study. For example, factors such as employment rate, percentage of green space, shops and restaurants, leisure facilities, schools and local crime rates, age and size of the house were not excluded in our first regression. Including these factors may have affected the results. This may explain our low R-squared value in our simple regression model. When repeating our hypothesis this time using a multiple regression, holding property type to be fixed, we found our R squared value to be 0.6893 therefore indicating that our independent variable can explain the variation in our

dependent variable. If we were to repeat these tests, to further improve the accuracy of our results, we could model each household to be of a specific type and age, for example, we could model our house to be a semi-detached house of the 1930's period (a common property size and age in Greater London and Kent.) This modelling can help improve our accuracy because our research will be more refined, eliminating the possibility of internal features of a property affecting the overall price of it. Another limitation which occurred during this investigation was calculating the exact distance from a household to their closest train station. Some assumptions were made when calculating the variable distance, including the speed at which commuters can travel from their household to their local train station. Additionally, using linear distance, or MSA, to calculate the distance travelled from the household to the station, we have underestimated the distance and therefore the time taken to commute from the house to the station. To avoid more inaccuracy in the future, we can use google maps to measure multiple distances eg the distance of each road, to mimic a more accurate representation of a commuter travelling to their nearest station.

References

Arter, K., 2009. Economic Impact of High Speed 1, London: Colin Buchanan.

Bohman, H. & Nilsson, D., 2016. The impact of regional commuter trains on property values: Price segments and income, s.l.: Elsevier.

<https://journals.sagepub.com/doi/10.1177/0042098010371395>

Debrezion, G. et al., (2010). The Impact of Rail Transport on Real Estate Prices: An Empirical Analysis of the Dutch Housing Market [Online]. Sage Journals. [viewed 18 January] Available from: <https://journals.sagepub.com/doi/10.1177/0042098010371395>

Dorantes, L. M., Paez, A. & Vassallo, J. M., 2011. Analysis of House Prices to Assess Economic Impacts of New Public Transport Infrastructure, Madrid: National Academy of Sciences.

Giles, C., (2020). Levelling up: how wide are the UK's regional inequalities. Financial Times. [online]. 2 March. [Viewed 03 January 2021] Available from: <https://www.ft.com/content/c9db4c66-5971-11ea-a528-dd0f971febbc>

highspeed1., (2020). Our stations [online]. highspeed1. [viewed 28 December 2020]. Available from: <https://highspeed1.co.uk/our-stations>

HM Land Registry., (2021). About us [online]. HM Land Registry. [viewed 04 January 2021]. Available from: <https://www.gov.uk/government/organisations/land-registry/about>

Miles and Sefton., (2020). House Prices and Growth with Fixed Land Supply [online]. The Economic Journal. [Viewed 18 January 2021]. Available from: <https://academic.oup.com/ej/advance-article-abstract/doi/10.1093/ej/ueaa088/5870104?redirectedFrom=fulltext>

Monaghan., (2020). House Prices to force extra 1m young people to live with parents [online]. The Guardian. [Viewed 17 January 2021]. Available from: <https://www.theguardian.com/business/2016/may/27/house-prices-force-1m-young-people-live-with-parents>

Office of Rail and Road., (2020). table-1415 time series of passenger entries and exits and interchanges by station [Data set]. ORR data portal. [Accessed 04 January 2021]. Available from: <https://dataportal.orr.gov.uk/statistics/usage/estimates-of-station-usage>

ONS., (2011). 6. Super Output Area (SOA) [online]. ONS. [viewed 05 January 2021]. Available from: <https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography>

Simpson, H., (2019). The economic impacts of HS1 in Kent [online]. unpublished. [Viewed 02 January 2021] Available from: https://drive.google.com/file/d/1dX7fwrHENT-2y7Dd_OhnaCd-lcaRxbXi/view?usp=sharing

Trainline., (2021). Trains from Tonbridge to London [online]. Trainline. [viewed 06 January 2021]. Available from: <https://www.thetrainline.com/train-times/tonbridge-to-london>

TFL., (2020). London Underground [online]. Transport for London. [Viewed 05 January 2021]. Available from: <https://tfl.gov.uk/corporate/about-tfl/culture-and-heritage/londons-transport-a-history/london-underground>

Appendix

Table 1: Simple Regression Model

Variables	Ln (House price)
Ln (Time to London)	-0.311*** (0.00208)
Constant	13.34*** (0.00925)
Observations	719,434
R-squared	0.030

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2: Multiple Regression Model

Variables	(1) ln (House Price)
Property: Semi-Detached	-0.548*** (0.00804)
Property: Terraced	-0.535*** (0.0200)
Property: Flat	-0.484*** (0.00131)
Property: Other	-0.703*** (0.00130)
Newly Built Property	0.189*** (0.00152)
Estate: Freehold	-0.484*** (0.00803)
Estate: Leasehold	0.00907 (0.154)
ln (Time to London)	-0.396*** (0.00125)
HS Station	-0.181*** (0.00175)
HS1#HS Station	-0.0323*** (0.00294)
Constant	13.22*** (0.00631)
Observations	719,434
R-squared	0.689

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Multiple Regression Model

Variables	(1) ln (House Price)
Property: Semi-Detached	-0.533*** (0.00425)
Property: Terraced	-0.531*** (0.00502)
Property: Flat	-0.479*** (0.00138)
Property: Other	-0.694*** (0.00135)
Newly Built Property	0.199*** (0.00164)
Estate: Freehold	-0.486*** (0.00406)
Estate: Leasehold	0.00723 (0.170)
Proximity Train time	15.03*** (0.237)
HS Station	0.232*** (0.00399)
HS1#HS Station	-0.0385*** (0.00616)
Proximity London	17.56*** (0.298)
HS1#Proximity London	-0.502** (0.218)
HS Station#Proximity London	-28.70*** (0.214)
HS1#HS Station#Proximity London	1.015*** (0.322)
Constant	11.03*** (0.00330)
Observations	719,434
R-squared	0.691

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4: Panel Regression

Variables	(1) ln (House Price)
Ln (Time to London)	-0.359*** (0.00734)
Constant	12.70*** (0.0327)
Observations	373,931
Number of postcodes	40,331

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1