Housing in St Andrews

2023/24 Martinmas Semester Progress Report 2

By:

Adam Gutman Carlsen, Matt Chivers, Laurie Dewar, Emily Moreland, Freya Pharis, Jimmy Plumleigh, Kieran Pirie, Sarah Vallet, Eric Xicheng Wang, Finn Watson

Submitted: 24th November 2023

TABLE OF CONTENTS

We would like to extend our sincere gratitude to our supervisors, Dr Sumedh Dalwai and Dr Luc Bridet, for their invaluable guidance and mentorship this semester.

Abstract

This paper examines housing market dynamics in St Andrews, Scotland. Using a hedonic pricing regression on a limited sample of rental properties, we determine that occupancy, location, and number of bathrooms are statistically significant hedonic attributes. We also evaluate the impact of Fife Council's 2019 'Overprovision Policy' aimed at restricting growth in 'Houses in Multiple Occupation' (HMO). These properties, rented to three or more unrelated individuals, must be licenced under Scottish law. Using an isoelastic supply and demand counterfactual model, we decompose the 39% increase in market average student rent between 2018 and 2022 into demand and supply shocks. Our decomposition attributes 36 percentage points (pp) to demand factors (1,230 extra students seeking private accommodation) and 3pp to supply factors (122 fewer HMO-licenced beds). Finally, using Fife Council's records of HMO licences, we estimate a linear probability model to investigate differential attrition of licences by occupancy and location. Unfortunately, our database is corrupted, so results are preliminary and unreliable while we await an accurate dataset. We find no statistically significant evidence of differential attrition between 2015 and 2022.

Introduction

St Andrews is a university town situated in Fife, Scotland, with an estimated population of 18,410[1]. The local housing market is characterised by several unique attributes, notably a substantial and expanding student population, and regulatory restrictions imposed by Fife Council on properties leased to three or more unrelated individuals, known as 'Houses in Multiple Occupation' (HMO).

Since 2018, the University's student population has risen by 1,485 (16.5%), reaching a total of 10,468 students in 2023[2]. This increase has prompted residents to argue that student housing growth is outpricing and displacing local families[3]. Fife Council reacted to residents' concerns in 2019 by restricting the growth of HMO-licenced properties that comprise most of the student housing stock. Fife Council reacted to residents' concerns in 2019 by restricting the growth of HMO-licenced properties that comprise most of the student housing stock. Summarily, the policy introduced a supply constraint to the market. The University has argued against the policy, claiming it increases student rental prices and displaces students to nearby towns. The issue remains one of the main points of 'town and gown' contention between the institution and the local community.

We are unaware of any quantitative academic literature examining the housing dynamics of university towns or the impact of the HMO Overprovision Policy on the St Andrews housing market. Furthermore, to the best of our knowledge, this is the first study into the use of HMO licencing to manipulate housing market dynamics. As such, our research contributes to the field of applied real estate economics and offers objective insight into the controversial debate on housing policy in St Andrews.

The research team consists of ten students from the University of St Andrews taking part in the Vertically Integrated Project programme – an initiative to bring academics and students together to tackle real-world issues through academic research. This semester, the team was organised into three sub-teams, namely 'Register', 'Modelling' and 'Outreach'.

The 'Register' research sub-team focused on secondary data analysis. We conducted a literature review of documentation related to HMO licencing and the Overprovision Policy to build research hypotheses. We analysed the HMO Public Register – a database published by Fife Council documenting HMO-licenced properties – to identify supply-side trends. Unfortunately, our database is corrupted, so results are preliminary and unreliable while we await an accurate dataset. We gathered descriptive statistics to illustrate the change in HMO supply at the aggregate level and based on property characteristics. Following a hypothesis that landlords differentially sacrifice HMO licences based on property characteristics, we conducted a series of linear probability regressions using occupancy and location variables. Finally, we constructed a hedonic pricing model to identify property-level factors that determine rental prices in St Andrews.

The 'Modelling' research sub-team focused on analysing rental price changes since the implementation of the HMO Overprovision Policy in 2019. We looked to predict how rent would behave in response to changes in housing supply and student demand for private housing. We built a supply and demand model in Python using the assumption that supply is perfectly inelastic, and demand is price elastic. We selected Ayton House, a private student accommodation in St Andrews, to approximate the price setting choices of a profit-seeking body. The Ayton House monthly rent in 2018 was used to calibrate the model that used the shocks to predict

[1] National Records of Scotland (2022) "Mid-2020 Population Estimates Scotland"
[2] University of St Andrews (2023) Freedom of information request relating to student population between 2010/11 and 2023/24.
[3] St Andrews QV (2018) "St Andrews HMO problem; where we are and how we got there".

a change in rent for a period after the policy had come into effect. We compared our predicted rent for 2022 with the actual price of the hall to evaluate the predictive ability of our model.

The 'Outreach' sub-team focused on effectively communicating our research findings to a general audience in an easily digestible and accessible manner. Our responsibilities involved running the project website, running the project Instagram account, and creating media for the VIP Conference. We have produced four informative Instagram posts this semester. Furthermore, we produced a poster, script, and video for the VIP Conference synthesising our findings into a presentable format. Additionally, we updated the website to make it more intuitive and add more current information.

This paper is structured by sub-team with sections allocated to Register, Modelling, and Outreach. We present a discussion on data, methodology, and results for Register and Modelling, followed by a broad commentary on lessons learned throughout the semester and efforts by the Outreach team to communicate our findings effectively.

HMO Overprovision Policy Overview

HMO Licencing Overview

In 2000, a mandatory licencing scheme for 'Houses in Multiple Occupation' (HMOs) was instituted under the Civic Government (Scotland) Act 1982[4]. As per this legislation, an HMO licence is mandatory for any rental accommodation that is:

1) Occupied by three or more persons from three or more families:
2) Occupied as a sole or main residence; and,
3) Providing shared basic amenities

The scheme was introduced following several fatal fires and aims to safeguard the interests of tenants by enforcing stringent safety standards within rented properties. HMO licences are initially valid for three years, with the option for renewal thereafter. Failure to comply with HMO licencing requirements constitutes a criminal offence carrying a maximum penalty of £50,000.

The legislation states that the local authority must consider location, condition, amenities, safety, and the potential for public disturbances when evaluating a licence application. Furthermore, Section 131A of the Housing (Scotland) Act 2006[5] provides that the Council may refuse to grant a licence if it deems that there is (or, as a result of granting the licence, would be) an overprovision of HMO properties in the locality. However, the legislation does not provide a specific definition for 'overprovision' or 'locality'.

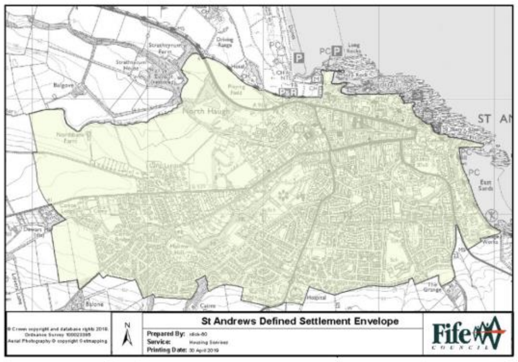
St Andrews Overprovision Policy

On 30th August 2018, Fife Council introduced an authority-wide HMO Overprovision Policy following a consultation and review process. The Overprovision Policy[6] was implemented in St Andrews on 11th April 2019, allowing no further growth in the number of HMOs. At the time, there were 1,219 HMOs in Fife, with a substantial majority (86%) located in St Andrews. HMO-licenced properties accounted for 15% of the St Andrews housing stock, accommodating a total capacity of 6,994 occupants.

The policy was applied to the defined settlement envelope of St Andrews, which is presented in Figure A.

[4] Scottish Government (2012) "Licensing of houses in multiple occupation: statutory guidance for Scottish local authorities".
[5] Scottish Government (2006), "Housing (Scotland) Act 2006".
[6] Fife Council (2021) "Houses in Multiple Occupation (HMO) Policy Statement".

Figure A: Defined Settlement Envelope of St Andrews



The policy was introduced following concerns from residents that the expansion of student housing, primarily composed of HMOs, was outpricing locals and contributing to the studentification of the town[7]. The Council's unconventional approach of using HMO licencing to restrict the supply of rental housing has raised tensions among stakeholders, including the University, the student community, and the town's residents. There are several activist groups advocating for their respective stakeholders, including the Campaign for Affordable Student Housing (CASH) and the Confederation of St Andrews Residents' Associations.

The policy did not impact properties with an existing licence in force; however, it would apply in the event that a pre-existing licence was to expire. Therefore, landlords who wish to retain an HMO licence must submit a renewal application before the expiration of their current licence. By law, the local authority must provide a decision within 12 months from the application date, although the process typically takes 4-6 months.

The policy does not state that Fife Council will reject all new applications. Instead, Fife Council claims that each application is assessed on its own merits, irrespective of the broader policy limiting HMO expansion. However, the onus is on the applicant to persuade Fife Council that the application should be granted. The default presumption is that any licence application within the Overprovision Policy boundary should only be granted if the applicant effectively demonstrates why an exception should be made.

[7] St Andrews QV (2018) "St Andrews HMO problem; where we are and how we got there".

2023 Policy Review

In mid-2021, Fife Council renewed the Overprovision Policy for an additional three years[8]. The Council conducted a review and noted a decrease of 17 active licences between the policy's introduction and February 2023[9]. Based on this observation, the Council opted to renew the policy, introducing more flexibility within the "no growth" position to account for "an agreed exemption in relation to student displacement". Specifically, the Council allocated 15 new licences to privately owned properties managed by the University of St Andrews, accommodating approximately 45 to 60 additional students.

Recent Policy Developments: Fee Structure

Since 2006, the application fee for a new HMO licence was £1,300, while the renewal cost was fixed at £500.[10] These fees applied irrespective of the number of occupants in the property. However, in 2019, Fife Council argued that the existing fees did not cover the cost of administering HMO-related services and thus required subsidies from the Council's General Fund. Subsequently, the fee structure was revised from 1st June 2022 as outlined in Figure B.

Figure B: New HMO Application Fee Structure (from June 2022)

Licensed Occupants	HMO Fee
3-5	£1,628
6-10	£1,903
11-20	£2,288
21-50	£2,563
51-100	£2,838
101-200	£3,542
201-300	£4,884
301-400	£6,248
401-500	£6,952
501+	£8,294

[8] Fife Council (2021) Houses in Multiple Occupation (HMO) Overprovision Policy Statement"

[9] Fife Council (2023), "Cabinet Committee"
[10] Fife Council (2022), "HMO Charging Structure"

HMO Register: Data

The Housing (Scotland) Act 2006[11] Part 5 requires Fife Council to maintain a publicly accessible register of HMO-licenced properties in Fife. This publication discloses information relating to licenced properties, including their location, the number of licenced occupants, date of licence issuance and expiration, and applicant names. The online document is updated quarterly and encompasses licences dating back five years. Previously, a comprehensive register dating back to 2005 was available online. However, this file was removed at the beginning of 2023. An example observation from the register is given below.

F02857/18	University Of St Andrews	W5R18	Andrew Melville Hall North Haugh St Andrews Fife KY16 9SU	Licence Expired	08/05/2018	07/09/2018	10/09/2021	**279**	CON

An observation from the Fife Council HMO Public Register (Q2 2023)[12]

The HMO Public Register was identified as a key secondary data source for identifying longitudinal licencing patterns and the supply-side impacts of the HMO Overprovision Policy.

During the previous semester, the team attempted to convert the public register from PDF format into an Excel spreadsheet but encountered difficulties due to inconsistent data entry and formatting. Furthermore, the public register is a licence-level data set with the same properties appearing multiple times according to the number of licences they have held historically. Consequently, the spreadsheets inherited by the incumbent team contained errors owing to file conversion, data cleaning, and attempts to convert the register from licence-level to property-level, as outlined in the Candlemas Semester 2022/23 Progress Report[13]. Upon studying the spreadsheets, the current team identified issues including the inclusion of HMOs outside of St Andrews, untraceable manual changes, and duplicate property observations.

Existing register datasets

The team inherited several spreadsheets containing unreliable conversions of the register, including "HMO Full Data Set With Kategorical Variable.xlsx" , a property-level data set with binary outputs for each quarter between Q1 2005 and Q4 2025 indicating whether the property had an active licence. The data cells are populated either a "1", signifying that the property had an active licence during the quarter, "0" indicating the absence of an active licence, or "ERROR" denoting that the licence was either conditionally granted or granted under another stipulation. The team interpreted "1" and "ERROR" as indicating the presence of an active licence, as the conditions of any such licence are irrelevant. An example observation is given below.

HMO Address		Occupants		Date Issued	Q1 2005	Q2 2005	Q3 2005	Q4 2005	Q1 2006	Q2 2006
Westburn HouseWest Burn LaneSt AndrewsFifeKY16 9TP		4	1	2/17/2003	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE

An observation from "HMO Full Data Set WIth Kategorical Variable.xlsx"

The data set contains 185,640 binary licence status observations relating to 2,210 properties in the period January 2005 to December 2025.

[11] Scottish Government (2006), "Housing (Scotland) Act 2006
[12] Fife Council (2023) "HMO Public Register – Q2 2023". Available at: https://www.fife.gov.uk/__data/assets/pdf_file/0017/470303/Q2-2023-HMO-Public-Register.pdf
[13] Vemru, A. et al. (2022*) "*HMO Caps in St Andrews 2022 Martinmas Semester Progress Report".

This data set was used to obtain descriptive statistics and run selective attrition regressions. A new 'Postcode' variable was created by extracting the postcode from the 'HMO Address' variable using the string position of 'KY16' to indicate the beginning of the postcode. We added another binary variable ('conservation') to indicate whether the property is within the St Andrews conservation area, as defined by having a postcode beginning "KY16 9". Later, a continuous variable was created for the distance from the property's postcode to the town centre of St Andrews, defined as Tesco Express at 138-140 Market Street. This variable was created by merging the master spreadsheet with a secondary data set containing every postcode within St Andrews (n=420) and the associated walking distance to Tesco (measured in kilometres) according to Google Maps. This data was gathered in the 2021/22 Candlemas Semester and is detailed in the relevant progress report[14]. The Stata code used to create these variables is outlined in Appendix A.

A separate data set, "Full Q.22 merged with Q1.23.xlsx", presents the same data by licence rather than by property. This version contains additional columns, including licence reference number, applicant name, ward, issuance and expiration dates, and the type of application decision ("GRANT", "CON", "DEEM", etc.). The definition of each application decision type is outlined in the Candlemas Semester 2022/23 Progress Report [15]. An example observation is given below.

App Ref Number	Applicant Name/s	WARD	HMO Address	License Status	Date of Ap	Date Issued	Expire Date	Tot Occs	Decision
F02167/17	David Durie	NW19	60 Auld Burn Park St Andrews KY16 8JD	Licence Expired	01.02.2017	14.06.2017	15.02.2020	3	GRANT

An observation from "Full Q.22 merged with Q1.23.xlsx"

A previous team requested a spreadsheet version of the register from the HMO Licencing Team at Fife Council and received "HMO Public List 02 03 2023.xlsx". This data set encompasses 2018 to early 2023 and contains all licences issued during this period. Given its five-year time span, this data set is inadequate for analysing long-term trends. Additionally, the public list does not offer a complete picture of all active licences. While it includes licences issued from 2018 onwards, it overlooks those already active but not issued during this period. This exclusion results in an incomplete representation of the total active licences. However, unlike the other data sets, this file can be considered accurate due to the absence of file conversion and data cleaning errors. An example observation for a single licence is given below.

App Ref Numbe	Applicant Name/s	WARD	HMO Address	Agent Name
F02934/18	Mr Andrew Tennant	W5R18	6 Largo RoadSt AndrewsFifeKY16 8RW	
	Mrs Samantha Tennant	W5R18	6 Largo RoadSt AndrewsFifeKY16 8RW	
		W5R18	6 Largo RoadSt AndrewsFifeKY16 8RW	University Of St Andrews Student Accommodation Services

License Status	Date of App	Date Issued	Expire Date	Tot Occs	Decision
Licence Expired	28/08/2018	01/11/2018	14/08/2021	5	GRANT
Licence Expired	28/08/2018	01/11/2018	14/08/2021	5	GRANT
Licence Expired	28/08/2018	01/11/2018	14/08/2021	5	GRANT

An observation from "HMO Public List 02 03 2023.xlsx"

[14] Alexander, J. et al. (2022) *"HMO CAPS IN ST ANDREWS PROGRESS REPORT 2 CANDLEMAS SEMESTER 2021/22"*
Available at: https://vip.wp.st-andrews.ac.uk/files/2023/03/April-2022-Progress-Report.pdf

[15] Chaney, A. et al. (2023), *"HMO Caps in St Andrews 2023 Candlemas Semester Progress Report"*
Available at: https://vip.wp.st-andrews.ac.uk/files/2023/11/HMO_VIP_2022_23_S2_ProgressReport.pdf

New register datasets

While the public list data is the most accurate as it came directly from the Council, it contains numerous formatting anomalies that hinder its usefulness in data analysis. As shown in the observation above, each licence observation might have duplicate rows solely to convey additional details about applicant and agent names. Consequently, we conducted file cleaning to produce usable *.dta files in both long and wide-form format at the property level.

The first of the new cleaned files, detailed in Week9_ListDataOptimal_FW.do (Appendix B), is Cleaned2018ListDataWideForm.dta. This file consolidates all data from various licences held by a property into a single row, using a wide form format to communicate licence activity across time. Following a similar structure as seen in the "Kategorical" data set, this wide form format employs binary indicators for each month. In this setup, a value of "1" indicates a licence is active in the specified month and year, whereas "0" denotes inactivity. Below, we provide the same observation as before, presenting all licence-related information associated with a property condensed into a single row.

HMOName3	Postcode_...	WARD	TotOccs	ApplicantNames	ApplicantName2
6 Largo Road	KY16 8RW	W5R18	5	Mr Andrew Tennant	Mrs Samantha Tennant

AgentName	AppRefNumber	AppRefNumber2	AppRefNumber3
University Of St Andrews Student Accommodation Services	F02855/18	F04182/21	F02934/18

DateofApp	DateIssued	ExpireDate	DateofApp2	DateIssued2	ExpireDate2	DateofApp3	DateIssued3	ExpireDate3
21may2018	14aug2018	14aug2021	13aug2021	05jan2023	13aug2024	28aug2018	01nov2018	14aug2021

LicenseStatus	Decision	LicenseStatus2	Decisio...	LicenseStatus3	Decision3
Licence Supersed...	GRANT	Licence Issued	CON	Licence Expired	GRANT

Y201801	Y201802	Y201803	Y201804	Y201805	Y201806	Y201807	Y201808	Y201809	Y201810
0	0	0	0	0	0	0	1	1	1

An observation from "Cleaned2018ListDataWideForm.dta"

Next, we generated Cleaned2018ListDataLongForm.dta which similarly consolidates all licence information at the property level. This data file differs from the aforementioned data set as it is structured in long form, wherein each row represents a property at a specific month and year. In this format, a row exists for every property during each time period from 2018 onwards. We constructed the data in both long and wide formats to enable future groups to approach data analysis in a more flexible way. Below, we present the same observation as shown in the Cleaned2018ListDataWideForm.dta example, this time showcased in long form.

Year	Month	HMOName3	Postcode_ori...	WARD	TotOccs	ApplicantNames	ApplicantName2	AgentName
2017	1	6 Largo Road	KY16 8RW	W5R18	5	Mr Andrew Tennant	Mrs Samantha Tennant	University Of St Andrews Student Accommodation Services
2017	2	6 Largo Road	KY16 8RW	W5R18	5	Mr Andrew Tennant	Mrs Samantha Tennant	University Of St Andrews Student Accommodation Services
2017	3	6 Largo Road	KY16 8RW	W5R18	5	Mr Andrew Tennant	Mrs Samantha Tennant	University Of St Andrews Student Accommodation Services
2017	4	6 Largo Road	KY16 8RW	W5R18	5	Mr Andrew Tennant	Mrs Samantha Tennant	University Of St Andrews Student Accommodation Services

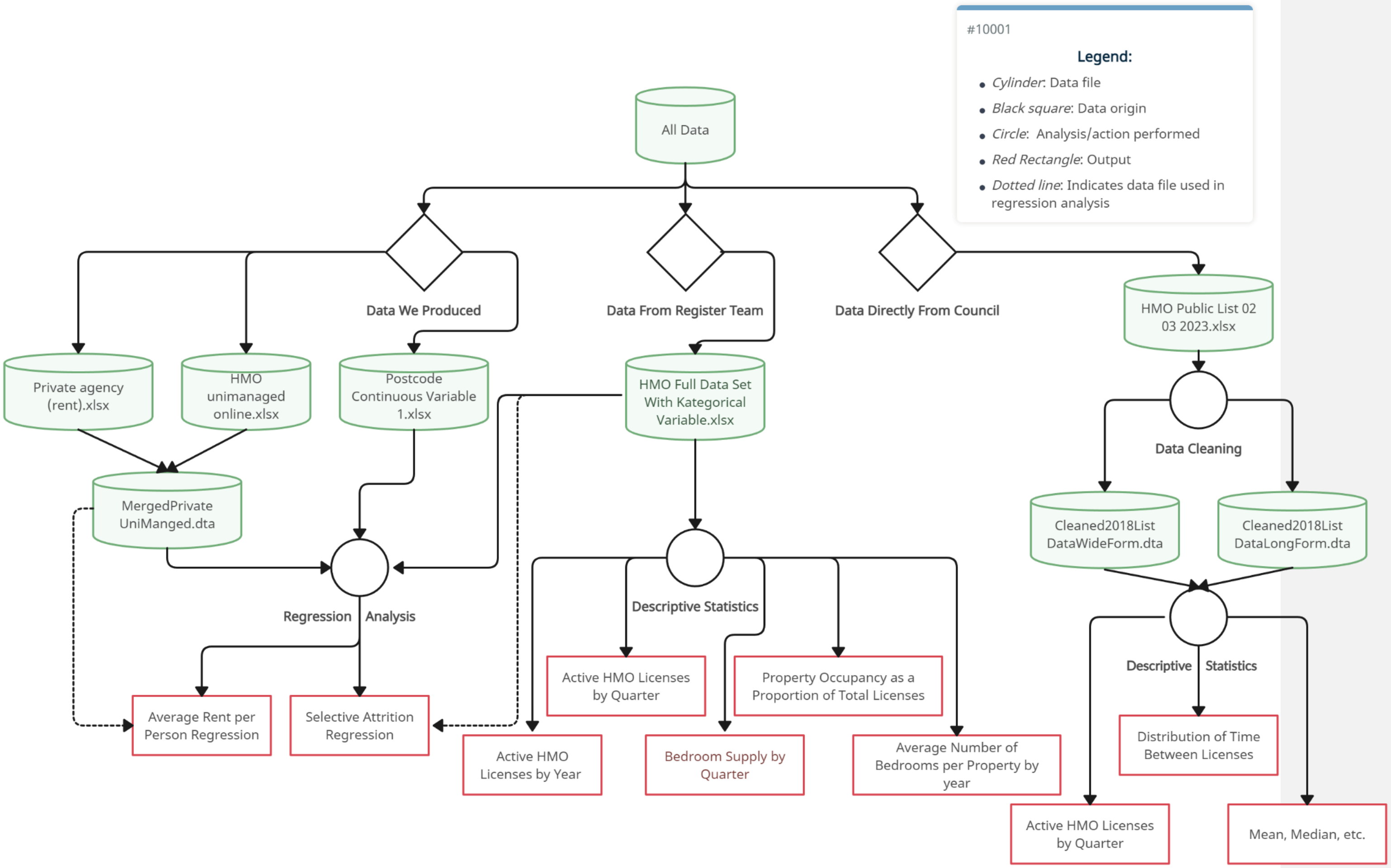
AppRefNumber	AppRefNu...	AppRefNu...	DateofApp	DateIssued	ExpireDate	DateofApp2	DateIssued2	ExpireDate2	DateofApp3	DateIssue...	ExpireDate3	LicenseStatus
F02855/18	F02934/18	F04182/21	21may2018	14aug2018	14aug2021	28aug2018	01nov2018	14aug2021	13aug2021	05jan2023	13aug2024	Licence Supersed...
F02855/18	F02934/18	F04182/21	21may2018	14aug2018	14aug2021	28aug2018	01nov2018	14aug2021	13aug2021	05jan2023	13aug2024	Licence Supersed...
F02855/18	F02934/18	F04182/21	21may2018	14aug2018	14aug2021	28aug2018	01nov2018	14aug2021	13aug2021	05jan2023	13aug2024	Licence Supersed...
F02855/18	F02934/18	F04182/21	21may2018	14aug2018	14aug2021	28aug2018	01nov2018	14aug2021	13aug2021	05jan2023	13aug2024	Licence Supersed...

Decision	LicenseStatus2	Decisio...	LicenseStatus3	Decisio...	Active
GRANT	Licence Expired	GRANT	Licence Issued	CON	0
GRANT	Licence Expired	GRANT	Licence Issued	CON	0
GRANT	Licence Expired	GRANT	Licence Issued	CON	0
GRANT	Licence Expired	GRANT	Licence Issued	CON	0

An observation from "Cleaned2018ListDataLongForm.dta"

Organisational Flowchart

The flowchart provided below delineates the ancestry and relationships among all pertinent register outputs and data files documented in this report. The chart begins by describing the data files' origin, distinguishing between those developed by our team, previous register teams, and directly acquired from Fife Council. Green cylinders represent these data files. The red rectangles indicate the outputs featured in this progress report, along with the data file they originated from. The analyses are illustrated as circles, differentiating them as regression analyses, descriptive statistics, or data cleaning. Finally, dotted arrows are used to explicitly illustrate the data used in each regression analysis.


#10001
Legend:
Cylinder: Data file
Black square: Data origin
Circle: Analysis/action performed
Red Rectangle: Output
Dotted line: Indicates data file used in regression analysis
All Data
Data We Produced
Data From Register Team
Data Directly From Council
Private agency (rent).xlsx
HMO unimanaged online.xlsx
Postcode Continuous Variable 1.xlsx
HMO Full Data Set With Kategorical Variable.xlsx
HMO Public List 02 03 2023.xlsx
Data Cleaning
MergedPrivate UniManged.dta
Cleaned2018List DataWideForm.dta
Cleaned2018List DataLongForm.dta
Regression Analysis
Descriptive Statistics
Descriptive Statistics
Average Rent per Person Regression
Selective Attrition Regression
Active HMO Licenses by Quarter
Property Occupancy as a Proportion of Total Licenses
Active HMO Licenses by Year
Bedroom Supply by Quarter
Average Number of Bedrooms per Property by year
Distribution of Time Between Licenses
Active HMO Licenses by Quarter
Mean, Median, etc.

Fife Council Outreach

After recognising the limitations of our existing data, we took the initiative to approach Fife Council with a two-fold objective: to acquire the Public Register in spreadsheet format dating back to 2013 and to establish a collaborative partnership that would facilitate access to additional data in the future. We emailed Fife Council on 27th September 2023, requesting a meeting to discuss our research goals and possible avenues for collaboration. A response was received from Tracey Drummond (Business Change Manager – Private Sector) and a meeting was scheduled for 20th October 2023.

During the meeting, the research team was represented by Dr Luc Bridet, Dr Sumedh Dalwai, Kieran Pirie, Sarah Vallet, and Jimmy Plumleigh. We discussed avenues for collaboration focusing on data sharing and studying mutually interesting outcomes. The Council received the research project positively and a follow-up data request spreadsheet was sent on 23rd October.

As of 21st November 2023, the Council has provided us with a web-scraped data set of rental prices in Fife (n=679) spanning from 2021 to 2023. Additionally, they have shared links to other information sources that may be helpful for our research. The Council is currently collating the 2013-Present HMO Public Register, HMO application objections data, and data relating to reports of unlicenced HMOs in St Andrews. We expect to receive these data sets in advance of the Candlemas Semester.

Rental Price Data

Towards the end of the semester, the team pursued a secondary workstream to manually web-scrape rental price information. The aim was to construct a basic hedonic price regression model that could potentially serve as a foundation for future semesters. We collected data concerning properties managed by the University (privately owned accommodation rented out by the University on behalf of the property owner) and properties leased through private estate agents.

The information was gathered by manually scraping the University's property rental website, resulting in the creation of 'HMO unimanaged Online.xlsx'. This file contains all 3+ bed properties available for rent during the 2023-24 academic year (n=23). For each property, we collected the monthly rent, occupancy, postcode, energy performance certificate (EPC) rating, number of bathrooms, and HMO registration number.

Following the methodology described in the 'existing register data sets' section, we generated a distance to town variable. This was achieved by cross-referencing postcodes with a secondary data set containing walking distances from all St Andrews postcodes to the town centre, according to Google maps.

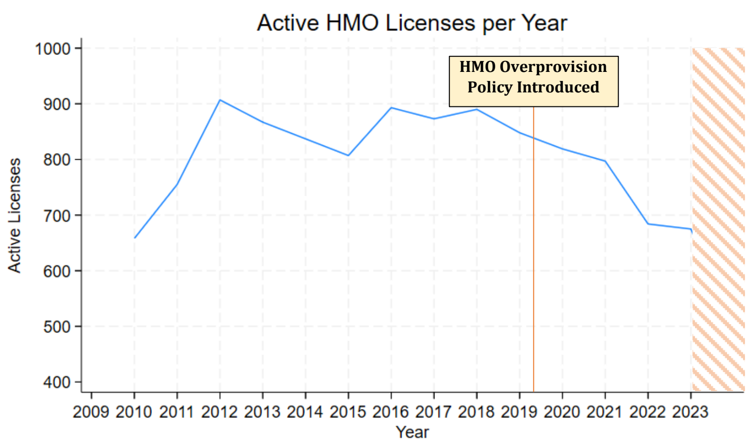
Similarly, we obtained 'Private agency (rent) .xlsx' by manually scraping data from the website of Lawson and Thompson – a private estate agent in St Andrews. We collected information on all properties listed as available for students (n=53). The same variables were collected as above, with the addition of 'number of living rooms and 'deposit'.

Subsequently, these two data sets were merged to produce 'MergedPrivateUniManage.dta' – a consolidated data set of 76 properties available for rent either through Lawson and Thompson or the University of St Andrews as of November 2023.

Descriptive Statistics

We gathered descriptive statistics to understand how the supply of HMO properties and HMO-licenced bedrooms has evolved over time. Specifically, we were interested in understanding the impact of the Overprovision Policy since 2019.

We began by plotting active HMO licences by year to identify trends in the supply of HMO properties before and after the introduction of the Overprovision Policy, as shown in Figure 1.



year	value
2010	658
2011	755
2012	907
2013	867
2014	837
2015	807
2016	893
2017	873
2018	890
2019	848
2020	819
2021	797
2022	684
2023	675
2024	505

Figure 1: Active HMO Licences by Year (Stata code in Appendix C)

Between 2010 to 2012, the number of active licences increased by 38% from 658 to 907. However, from 2012 to 2015, the number of active licences steadily declined to 800 before rebounding to 893 in 2016. This was followed by a relatively stable number of active licences between 2016 and 2019, ranging from 848 to 890.

The introduction of the Overprovision Policy in April 2019 was followed by a marked decline (31%) in the number of active licences from 848 to 648 in 2022. This returned the supply of HMO properties to a level not seen since mid-2010.

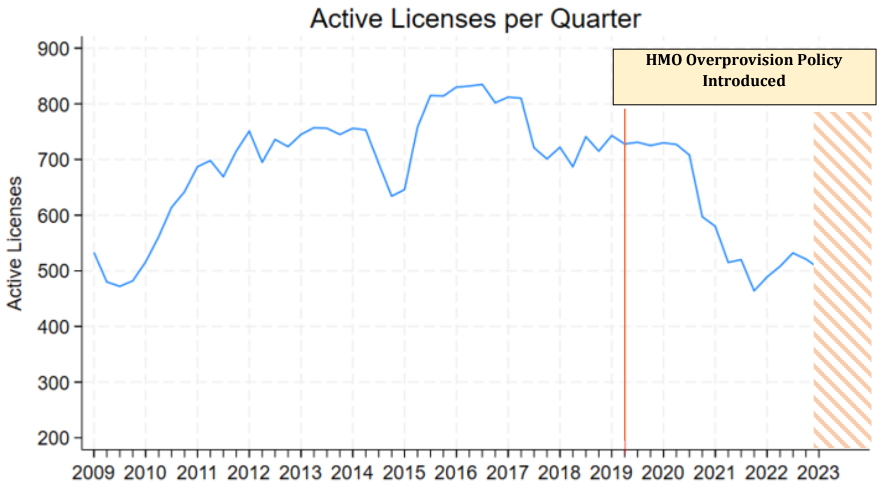


Figure 2: Active HMO Licences by Quarter (Stata code in Appendix D)

Figure 2 illustrates a quarterly version of the same graph and offers additional insight into HMO property supply trends. The range in active licences in the quarterly dataset (464 to 835) is greater than the range in active licences in the annual dataset (658 to 907) due to the methodology used to calculate annual active licences. An HMO property was considered active in a given year if it held an active licence in any of the four quarters. Therefore, in cases where a licence expired mid-year, the annual data misrepresents the actual supply. For example, in 2019, we calculated 848 active licences while there were only 743, 728, 731 and 725 active licences in each quarter, respectively. Consequently, the quarterly data is a better representation of reality.

The data shows a sudden decline in the number of active licences after Q2 2020. The plateau immediately following the policy announcement from Q2 2019 to Q2 2020 is expected as licences were not actively revoked but allowed to voluntarily expire so it is logical that a drop does not immediately occur. Between Q1 2018 and Q1 2023, the count of active licences decreased by 30%.

Although there remained a general upward trend in the number of active HMO licences leading up to the introduction of the Overprovision Policy, the peak of quarterly active HMO licences occurred in Q3 2016 at 835, contrasting with the peak in the yearly data, which was observed at 907 in 2012.

While the Overprovision Policy directly restricted the supply of HMO *properties*, the effect of the policy on the supply of HMO *bedrooms* is more relevant for policy impact evaluation. As such, Figure 3 illustrates the evolution of aggregate HMO-licenced occupancy from 2010 to 2022.

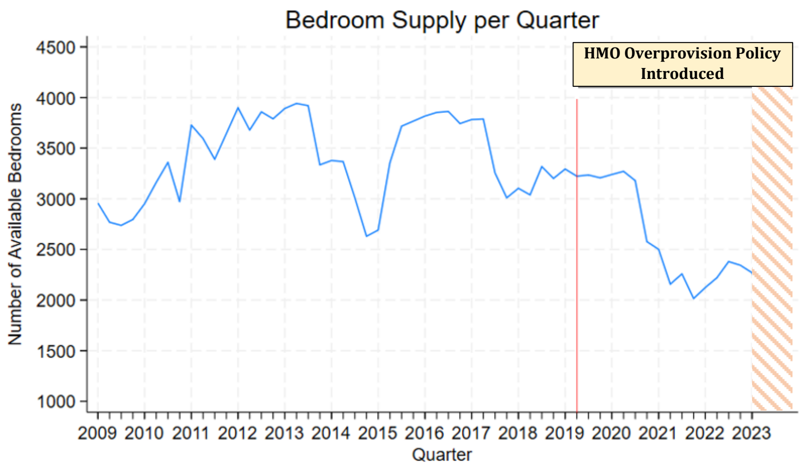


Figure 3: Bedroom Supply by Quarter (Stata code in Appendix E)

Figure 3 highlights that bedroom supply fluctuated significantly over the time period, ranging from 3863 beds in Q3 2016 to 2015 beds in Q4 2021. While there was an evident upward trend in the supply of HMO properties before 2019, any such trend is significantly moderated when considering bedroom supply. This may suggest that the average occupancy of HMO properties decreased to counteract the growth in HMO licences. The significant decline observed in 2014 was primarily due to the loss of a licence for Agnes Blackadder Hall (ABH), which accounted for 561 occupants. Following the introduction of the Overprovision Policy, bedroom supply decreased by 27% from 3223 beds in Q2 2019 to 2344 beds in Q4 2022. The supply of bedrooms in Q3 2021 is lower than any other point in the data set.

Having noted the decline in active licences and available bedrooms following the introduction of the Overprovision Policy, we sought to identify specific factors that might have contributed to the loss of licences for certain properties. The team considered the location and licenced occupancy of the property in this regard. We theorised that the cost of losing an HMO licence for a property with a large occupancy (e.g., 8 occupants) might be greater than a property with a smaller occupancy (e.g., 3 occupants) because properties with larger occupancies could prove less adaptable for alternative uses or suitable for conversion into non-HMO rentals for single families. Similarly, we hypothesised that centrally located properties might command a higher rent premium from students compared to local families. Consequently, the potential cost incurred from losing a licence for a property in the town centre, assuming its conversion into a non-HMO single family rental, could be greater than that of a property located on the outskirts of St Andrews.

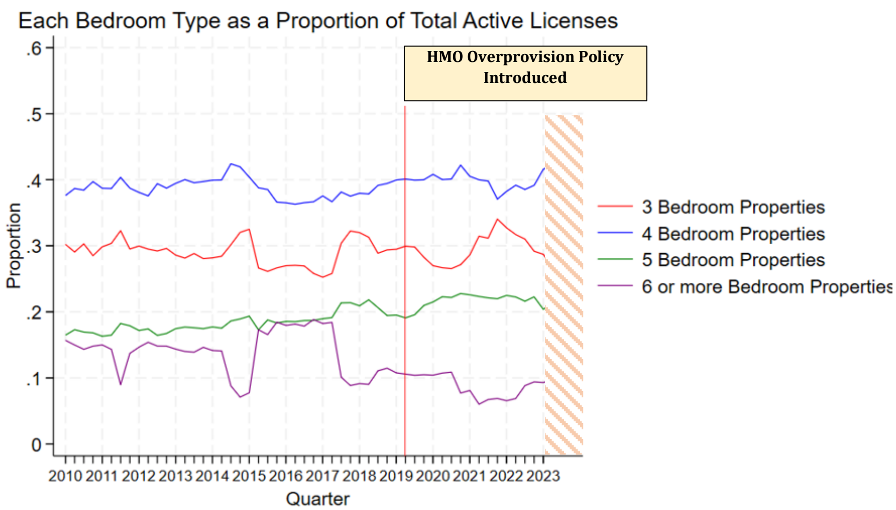


Figure 4: Property Occupancy as a Proportion of Total Licences (Stata code in Appendix F)

Figure 4 splits licenced properties by occupancy and plots each occupancy group as a proportion of total active licences from 2010 to 2023. The graph exhibits broadly stable trends, with each group's proportion remaining within a narrow range of around +/- 0.05 of their original value. Four-bedroom properties comprise the majority (c. 40%) of active HMO licences in the period, followed by three-bedroom (c. 30%), five-bedroom (c. 20%), and 6+ bedroom (c. 10%) properties. The graph suggests a slight proportional decline in three-bedroom properties after mid-2021 and a proportional rise in the number of four-bedroom properties, which may offer weak support to the hypothesis that three-bedroom properties are disproportionally losing their HMO licences.

However, on an aggregate level, the average number of bedrooms per property has remained almost constant since the introduction of the Overprovision Policy, as shown in Figure 5.

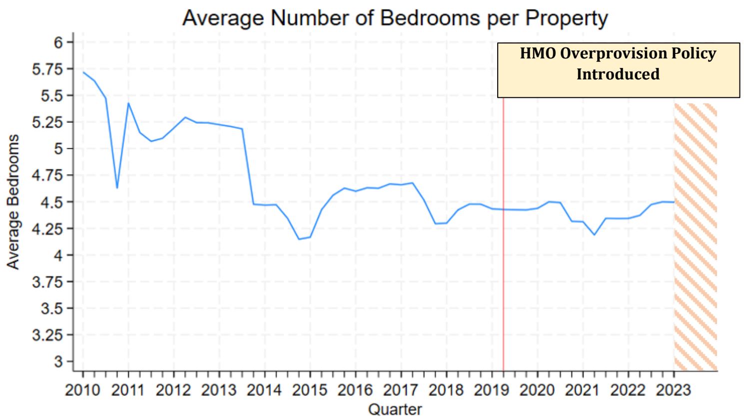


Figure 5: Average Number of Bedrooms per Property by year (Stata code in Appendix G)

Figure 5 highlights that the average occupancy of HMO properties has remained largely stable since 2015 at around 4.3 occupants. This contrasts with a clear downward trend between 2010 and 2015, with the average number of bedrooms per property decreasing from 5.7 in Q1 2010 to 4.1 in Q4 2014. While the stable trend after 2019 does not lend support to the selective attrition hypothesis, it is essential to acknowledge that the average (mean) serves as a limited and potentially unrepresentative statistic for capturing trends in sub-groups, as equal positive and negative deviations from the mean can offset each other. By contrast, the proportion graph discussed previously offers a more compelling visualisation of changes in the licenced occupancy of active HMO licences.

The team considered the location of HMO-licenced properties relative to the centre of town. Figure 6 highlights that the average HMO licenced property is 1.1 km away from the centre of town by walking, with 75% of properties located within 1.55km. Of the properties, 53% were within the St Andrews conservation area.

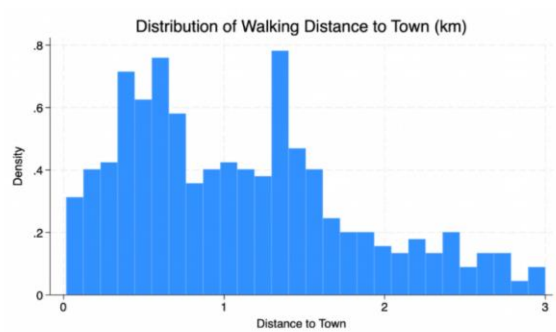


Figure 6: Distribution of Walking Distance to Town (Stata code in Appendix H)

Finally, we investigated licence renewal trends over time, specifically examining the "renewal gap" between licences. We define the "renewal gap" as the duration between the expiration of an existing licence and the issuance date of the renewal licence, indicating the period a property remains unlicenced between licences. We completed this analysis using data from the 'HMO Public List 02 03 2023.xlsx' data set and 'Cleaned2018ListDataLongForm.dta' as these were the only data sets containing the expiration and issue dates for all property licences.

Figure 7 indicates that the median renewal gap between licences is around 400 days. If accurate, this would suggest that many properties used for HMO letting either drop out of the market while awaiting licence renewal or continue to be rented out illegally.

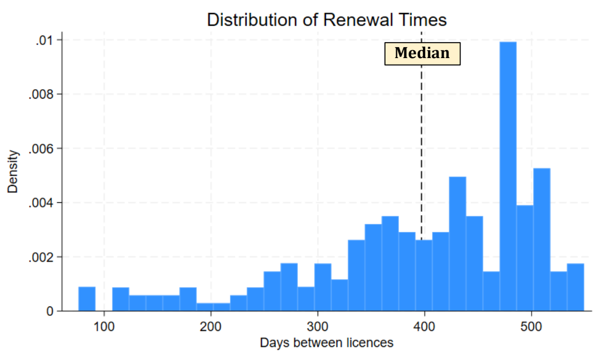


Figure 7: Distribution of Renewal Gap ([Appendix I]())

HMO Register: Methods

Descriptive Statistics

The descriptive statistics were obtained using Stata. The data files and *.do files were stored in a OneDrive folder to facilitate collaboration between sub-team members. As previously mentioned in the data section, the data sets used may be deemed unreliable; therefore, the descriptive statistics presented should be regarded solely as 'proof of concept'.

We generated several scalars to obtain the descriptive statistics. The total number of active licences in each quarter scalar ('total_Q[Quarter][Year]') was generated by sequentially iterating through all quarters and aggregating observations equal to "1" or "ERROR". The total number of active occupants in each quarter scalar ('total_occupants_Q[Quarter][Year]') was similarly obtained by iterating through all quarters and aggregating property occupancy for each active licence observation. The corresponding annual scalars ('total_[Year]', 'total_occupants_[Year]') were generated by aggregating licences and occupants, respectively, if any quarter in a given calendar year had an active licence. The associated Stata code is given in Appendix J.

Selective Attrition Regression

In our descriptive statistics section, we outlined a hypothesis that landlords may have responded to the Overprovision Policy by differentially sacrificing licences based on property occupancy and location. We postulated that the cost associated with losing an HMO licence might increase with property occupancy due to the challenges larger properties present in terms of adaptability for alternative uses or conversion into non-HMO rentals suitable for single families.

Furthermore, we speculated that the cost of forfeiting a licence could decrease with distance from the town centre, as students might be willing to pay a higher rent premium for centrally located accommodations compared to local families. Consequently, assuming the alternative involves converting the property into a non-HMO rental for a single family, the potential loss in rental income from losing a licence for a property situated in the town centre might be greater than that of a property located on the outskirts of St Andrews. It is important to note that the Overprovision Policy did not revoke existing licences and has no negative bearing on the prospect of renewal applications. Therefore, the loss of licences solely reflects choices made by property owners.

We formally tested this hypothesis of "selective attrition" using a linear probability model (LPM) – a type of binary response model. The general form of the model is given by (1)[16].

$$P(\, y = 1 \mid x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{1}$$

The model can be used to estimate the probability of an event occurring as a linear function of explanatory variables. For example, β_1 is interpreted as the change in the probability of success given a one-unit increase in x_1, where success is defined as $y = 1$.

[16] Wooldridge, J. (2010), *"Econometric Analysis of Cross Section and Panel Data",* Second Edition: The MIT Press

We created several binary variables to facilitate the regression analysis. Firstly, we restricted the regression to property observations which held an active licence in Q1 2015 (n=646). We then generated a binary variable called "active_both" to indicate whether the property held an active licence in both Q1 2015 and Q1 2022 (n=291). The time period was chosen to encompass a minimum of three years of observations before and after the introduction of the Overprovision Policy.

A second binary variable, "lost_licence", was formed as the complement of "active_both". The variable was assigned "1" if a property lost its licence during the specified period, while "active_both" was designated as "1" if a property retained its licence during the period. We used these variables interchangeably as the dependent variable, with a preference for "lost_licence" due to its marginally more intuitive interpretation (1 indicating a property lost its licence, 0 indicating otherwise). Swapping the dependent variables did not affect the regression outputs other than altering the signs of the independent variable coefficients. The associated Stata code is given in Appendix K.

We created binary property occupancy variables called 'Bedroom4', 'Bedroom5' and 'Bedroom6plus' indicating whether a property had 4, 5, or 6+ licenced occupants, respectively. A 'Bedroom3' variable was not created for three-bed properties as it would induce perfect collinearity among the independent variables. Therefore, the baseline constant coefficient relates to a 3-bed property.

A 'Bedroom2' and 'Bedroom1' variable would not be relevant as HMO licences are only required for properties with three or more occupants. We decided to create the 'Bedroom6plus' variable because there were very few observations of properties with 6 or more licenced occupants (3.56% of total observations), as demonstrated in the table below. Therefore, it made sense to create a grouped binary variable.

. tab Occupants if Active2015 == 1

Occupants	Freq.	Percent	Cum.
3	210	32.51	32.51
4	261	40.40	72.91
5	125	19.35	92.26
6	27	4.18	96.44
7	6	0.93	97.37
8	6	0.93	98.30
9	2	0.31	98.61
10	3	0.46	99.07
11	1	0.15	99.23
13	1	0.15	99.38
14	1	0.15	99.54
15	1	0.15	99.69
20	2	0.31	100.00
Total	646	100.00	

Additionally, we created a "conservation" variable which was assigned the value "1" if the property falls within the Central St Andrews conservation area, identified by postcodes beginning with "KY16 9". Out of the 646 observations, 356 (55%) were situated within the conservation area, forming a substantial sample size. Our selection of the conservation area definition was guided by anecdotal evidence suggesting that properties north of the Kinness Burn are perceived as particularly central and desirable from a student standpoint thus are most likely to incorporate a student-specific 'central location' rent premium.

However, we later noted that any rental price premium for the location of a property is likely continuous with distance from the town centre. Therefore, as described in the data section, a continuous variable was created

Commented [FW1]: Quantify perhaps? (3.56% of observations 6+plus)

Commented [KP2R1]: Great point. Amended as suggested containing the distance from the property's postcode to the town centre of St Andrews, defined as Tesco Express at 138-140 Market Street.

To illustrate the interpretation of our model, consider example regression (2).

$$lost\_license = \beta_0 + \beta_1 \boldsymbol{Bedroom4} + \beta_2 \boldsymbol{Bedroom5} + \beta_3 \boldsymbol{Bedroom6plus} + \beta_4 \boldsymbol{Conservation} \quad (2)$$

The coefficient on Bedroom4 is interpreted as the probability of a property with four licenced occupants outside the conservation area losing its HMO licence between Q1 2015 and Q1 2022, relative to a three-bed property outside the conservation area (in other words, when all other explanatory variables equal zero). The intercept is interpreted as the probability of a 3-bed property outside the conservation area with an active licence in 2015 losing it by 2022.

Hedonic Pricing Regression

As described in the '<u>rental price data</u>' section, we gathered a small data set of rental prices in St Andrews with the aim of constructing a basic hedonic regression model that could potentially serve as a foundation for work in future semesters.

Hedonic pricing regression models specify a relationship between rent and property characteristics. The independent variables, known as 'hedonic attributes', typically fall into three categories: property-specific, location-based, and neighbourhood-related[17]. Unfortunately, we were restricted to the variables obtained through the web-scraping process.

We used an ordinary least squares (OLS) hedonic regression model, as denoted by equation (3). Within this model, the dependent variable is monthly rent, and the hedonic attributes are number of occupants, walking distance to the town centre, number of bathrooms, and a dummy variable for each EPC rating excluding 'G' which had only one observation. We excluded 'EPC_G' from the regression to prevent perfect collinearity among the EPC independent variables.

$$Rent = \beta_0 + \beta_1 \boldsymbol{Occupants} + \beta_2 \boldsymbol{Distance} + \beta_3 \boldsymbol{Bathrooms} + \beta_4 \boldsymbol{EPC\_C} + \beta_5 \boldsymbol{EPC\_D} + \beta_6 \boldsymbol{EPC\_E} + \beta_7 \boldsymbol{EPC\_F} \quad (3)$$

The 'Occupants' variable is a justifiable hedonic attribute as larger occupancy properties typically command a higher rent. Similarly, a high EPC rating indicates high energy efficiency which may reduce utility bills, thus being a desirable attribute for consumers. Finally, 'Bathrooms' is an intuitive attribute as bathrooms are a desirable property feature that typically commands a higher rent.

[17] Micallef, B., Ellul, R. and Debono, N. (2021) 'A hedonic assessment of the relative importance of structural, locational and neighbourhood factors on advertised rents in Malta', *International Journal of Housing Markets and Analysis*, 15(1), pp. 203-230.

The sampling methodology used to collect the rental price data for our hedonic pricing regression is potentially problematic. Gathering rent data from a single private estate agent outside the primary season for marketing student properties may not offer a comprehensive representation of the student rental housing market in St Andrews. Consequently, the OLS estimators may be biased because the observations were not gathered independently. Hence, future teams may wish to consider improving the data set used for the hedonic price regression. The relevant Stata code is given in Appendix L.

HMO Register: Results

Selective Attrition

The team ran a series of selective attrition regressions using the model specified in the 'HMO Register: Methods' section. Figure 7 presents the output of the most basic regression containing only occupancy variables over the time period 2015 to 2022.

```
. regress lost_license Bedroom4 Bedroom5 Bedroom6plus if Active2015 == 1

      Source |       SS           df       MS      Number of obs   =       646
-------------+----------------------------------   F(3, 642)       =      0.89
       Model |  .665871081         3  .221957027   Prob > F        =    0.4435
    Residual |   159.24899       642  .248051386   R-squared       =    0.0042
-------------+----------------------------------   Adj R-squared   =   -0.0005
       Total |  159.914861       645  .247930017   Root MSE        =    .49805

------------------------------------------------------------------------------
lost_license | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    Bedroom4 |   .0642583   .0461691     1.39   0.164    -.0264023     .154919
    Bedroom5 |   .0137143   .0562637     0.24   0.808    -.0967689    .1241975
Bedroom6plus |   .0857143   .0783723     1.09   0.275    -.0681828    .2396114
       _cons |   .5142857   .0343685    14.96   0.000     .4467974    .5817741
------------------------------------------------------------------------------
```

Figure 7: Basic Selective Attrition Regression Output

The regression had 646 observations encompassing all licences that were active in 2015. All the coefficients were positive, indicating that four-bedroom, five-bedroom and 6+ bedroom properties were comparatively more likely to lose their licence over the given period than three-bed properties. However, none of the occupancy variables were significant at any of the conventional significance levels and the R-squared of 0.0042 indicates that the model has negligible explanatory power. This suggests the absence of a meaningful relationship between property occupancy and the probability of a property losing an HMO licence. Furthermore, it refutes the selective attrition hypothesis relating to the differential sacrifice of HMO licences based on property occupancy.

A second regression was run with the binary 'conservation' variable and generated the output presented in Figure 8.

. regress lost_license Bedroom4 Bedroom5 Bedroom6plus conservation if Active2015 == 1

Source	SS	df	MS			
				Number of obs	=	646
				F(4, 641)	=	1.46
Model	1.43955995	4	.359889987	Prob > F	=	0.2142
Residual	158.475301	641	.247231358	R-squared	=	0.0090
				Adj R-squared	=	0.0028
Total	159.914861	645	.247930017	Root MSE	=	.49722

lost_license	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Bedroom4	.0573765	.0462566	1.24	0.215	-.0334563	.1482093
Bedroom5	.0150492	.0561757	0.27	0.789	-.0952615	.1253599
Bedroom6plus	.0912543	.0783053	1.17	0.244	-.0625117	.2450203
conservation	-.0700843	.0396177	-1.77	0.077	-.1478804	.0077119
_cons	.5550013	.0413162	13.43	0.000	.4738699	.6361328

Figure 8: Selective Attrition Regression Output with Conservation Variable

By adding the conservation variable, the R-squared of the regression increased marginally but remained negligible. The conservation variable was statistically significant at a 10% significance level, and its coefficient is interpreted as 'a three-bedroom property in the conservation area was 7% less likely to lose its HMO licence over the 2015 to 2022 period than a three-bedroom property located outside the conservation area'. This suggests that the location of a property may be a contributing factor to selective attrition and that there is indeed a higher cost associated with losing a licence in the centre of two compared to the outskirts.

In a concluding regression, we replaced the binary conservation variable with the continuous variable 'Distance to Town.' The output is given in Figure 9. In this revised regression, the R-squared value decreased, and none of the variables were statistically significant at conventional levels. Moreover, the regression F-test yielded a p-value of 0.93, signifying that the overall model is not statistically significant.

. regress lost_license Bedroom4 Bedroom5 Bedroom6plus DistancetoTown if Active2015 == 1

Source	SS	df	MS			
				Number of obs	=	584
				F(4, 579)	=	0.20
Model	.201795645	4	.050448911	Prob > F	=	0.9381
Residual	145.791355	579	.251798541	R-squared	=	0.0014
				Adj R-squared	=	-0.0055
Total	145.993151	583	.250417068	Root MSE	=	.5018

lost_license	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Bedroom4	.0370502	.0488752	0.76	0.449	-.0589441	.1330446
Bedroom5	.0196308	.0585148	0.34	0.737	-.0952964	.1345579
Bedroom6plus	.0122124	.0871586	0.14	0.889	-.1589732	.1833979
DistancetoTown	-.0226068	.0420939	-0.54	0.591	-.1052822	.0600687
_cons	.5004418	.0459862	10.88	0.000	.4101217	.5907619

Figure 9: Selective Attrition Regression Output with Distance to Town Variable

Overall, the regressions do not lend support to our selective attrition hypothesis. While the unreliability of the underlying data may be a contributing factor, we believe the results are strong enough to preliminarily reject our null hypothesis that HMO licences were differentially sacrificed based on property occupancy and location between 2015 and 2022.

Hedonic Price Regression

We ran a hedonic pricing regression using the model specified in the '[HMO Register: Methods]' section. The output is presented in Figure 10 below.

```
. reg Rent Occupants Distance BathroomToilets EPC_C EPC_D EPC_E EPC_F

      Source |       SS           df       MS      Number of obs   =        76
-------------+----------------------------------   F(7, 68)        =     28.13
       Model |  51681276.9         7  7383039.55   Prob > F        =    0.0000
    Residual |  17846636.6        68  262450.538   R-squared       =    0.7433
-------------+----------------------------------   Adj R-squared   =    0.7169
       Total |  69527913.4        75  927038.846   Root MSE        =     512.3

-----------------------------------------------------------------------------
           Rent | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
----------------+------------------------------------------------------------
      Occupants |   754.6377   77.56714     9.73   0.000     599.8549    909.4205
       Distance |  -475.2009   100.6473    -4.72   0.000    -676.0395   -274.3623
BathroomToilets |   174.4239   67.49519     2.58   0.012      39.7393    309.1084
          EPC_C |   119.4393     318.78     0.37   0.709    -516.6763    755.5548
          EPC_D |    148.374   313.4244     0.47   0.637    -477.0547    773.8027
          EPC_E |  -105.8823   331.2693    -0.32   0.750    -766.9198    555.1553
          EPC_F |   42.23198   397.2128     0.11   0.916    -750.3938    834.8577
          _cons |  -319.4527   453.9876    -0.70   0.484    -1225.371    586.4654
-----------------------------------------------------------------------------
```

Figure 10: Hedonic Pricing Regression Output

The regression yielded a high R-squared of 0.7433 and an adjusted R-squared of 0.7133, indicating that the model explains the 74% of the variation in rental prices. The coefficient on the 'Occupants' hedonic attribute (754.6) was found to be statistically significant at the 1% significance level. This implies that, all other factors remaining constant, a one-person increase in a property's occupancy results in an average rent increase of £754.6 per month. The 'Distance' variable coefficient (-475.2) was also significant at the 1% significance level. This suggests that on average and holding all else equal, rent decreases by £475.2 per month for each kilometre farther away the property is from the town centre.

The coefficient on the number of bathrooms variable (174.4) was significant at the 10% significance level. This suggests that a one-unit increase in the number of bathrooms within a property increases average rent by £174.4 per month, holding all other factors constant.

Notably, the EPC dummy variables were insignificant at the conventional levels, suggesting that energy efficiency does not effectively explain variation in rental prices. Additionally, we noted the presence of minor collinearity between these variables, as demonstrated by the correlation matrix and VIF score outputs below.

```
. corr Occupants Distance BathroomToilets EPC_C EPC_D EPC_E EPC_F
(obs=76)

             | Occupa~s Distance Bathro~s    EPC_C    EPC_D    EPC_E    EPC_F
-------------+---------------------------------------------------------------
   Occupants |   1.0000
    Distance |  -0.1221   1.0000
BathroomTo~s |   0.3896  -0.2878   1.0000
       EPC_C |   0.0566   0.0473   0.1180   1.0000
       EPC_D |  -0.0366   0.0447  -0.2081  -0.5669   1.0000
       EPC_E |  -0.0932  -0.0788   0.0690  -0.2715  -0.4309   1.0000
       EPC_F |   0.0930  -0.0910  -0.0000  -0.1409  -0.2236  -0.1071   1.0000
```

```
. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
       EPC_D |      7.09    0.141004
       EPC_C |      5.71    0.175251
       EPC_E |      4.51    0.221929
       EPC_F |      2.28    0.438957
BathroomTo~s |      1.35    0.738596
   Occupants |      1.21    0.826519
    Distance |      1.12    0.895955
-------------+----------------------
    Mean VIF |      3.32
```

We tested for heteroskedasticity using the White Test (p-value: 0.2495) and failed to reject the null hypothesis, concluding that the model does not suffer from unrestricted heteroskedasticity. Additionally, we ran a Shapiro-Wilk test on the model residuals (p-value: 0.000) and rejected the null hypothesis that the error term is normally distributed. This implies that our standard errors may not be robust, thereby impacting the reliability of our t-test inferences.

We failed to reject the null hypothesis of Ramsey's RESET test (p-value: 0.3710), suggesting that we have chosen an appropriate functional form. However, there is extensive literature on the correct choice of functional form for hedonic price regressions, with many authors pointing towards using log form or including Box-Cox transformations to achieve the optimal specification[18]. Therefore, despite the RESET result, future teams may wish to consider justifying the functional form as a first step to improving this model.

[18] Huh, S. and Kwak, S. (1997) 'The Choice of Functional Form and Variables in the Hedonic Price Model in Seoul', *Urban Studies*, 34(7), pp. 967-1149.

HMO Register: Discussion

Selective Attrition

We were unable to prove our null hypothesis that landlords differentially sacrifice licences based on property occupancy and location between 2015 and 2022. Notably, the Adj-R^2 of the regressions, which captures the proportion of variance in the dependent variable that is predicted by the independent variables adjusted for the number of predictors, ranged from -0.0055 to 0.0028 indicating that our model explains variance in the data extremely poorly. Furthermore, all regressions yielded statistically insignificant coefficients on the explanatory variables suggesting that selective attrition is not a meaningful characteristic of the St Andrews rental market. Alternatively, there may be more relevant variables or a more suitable functional form that generates statistically significant results.

Our findings contradict those obtained by the team in the Martinmas semester of 2022/23[19], who conducted an identical linear probability model on selective attrition from 2015 to 2022. Notably, our regression coefficients yielded no statistical significance. Conversely, the previous team identified statistically significant '5-bedroom' and '6-bedroom' coefficients, suggesting both are more likely to lose their licence than a 3-bed property, holding all else equal. The previous team's regression output is presented below.

```
Call:
lm(formula = active_both ~ as.factor(bedroom) + conservation,
    data = hmo_both, subset = active_2015 == 1)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5821 -0.4138 -0.2844  0.4557  0.7156

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.56176    0.03498  16.058  < 2e-16 ***
as.factor(bedroom)4 -0.01751    0.04021  -0.435    0.663
as.factor(bedroom)5 -0.16825    0.03881  -4.335 1.58e-05 ***
as.factor(bedroom)6 -0.27737    0.05461  -5.079 4.41e-07 ***
conservationTRUE     0.02030    0.02914   0.697    0.486
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4906 on 1172 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.03868,   Adjusted R-squared:  0.0354
F-statistic: 11.79 on 4 and 1172 DF,  p-value: 2.157e-09
```

Figure 11: Selective Linear Regression Result from the Martinmas semester 2022/2023

We suspect the difference is due to us using different data than the previous team. Although, as previously discussed in the data section, our dataset is flawed, the unknown dataset of the previous semester is equally dubious due to using a combination of unknown R code and manual Excel manipulation.

[19] Vemuru et al (2022), *"HMO Caps in St Andrews 2023 Martinmas Semester Progress Report"*

This is not to say that data reliability is not a significant issue with regards to our selective attrition results. The unreliability of our underlying data casts significant doubt on all register analysis presented in this progress report. The primary register data file used throughout the progress report deviates significantly from officially reported statistics from Fife Council. For example, the Council claimed there were 1,029 active HMO licences in 2022[20] whereas our data claims there was 684 active licences across 2022. This demonstrates the inaccuracy and limits associated with our available data sets. This is very likely due to either inaccuracies in the publicly available register, or with the data conversion methodology used in previous semesters. In this way, our data, and by extension our regression results, likely do not reflect the reality of the housing market in St Andrews.

Future teams may benefit from employing a logit or probit model to gain more nuanced insights into selective attrition. A significant shortcoming of linear probability models is the assumption of linearity between the dependent and independent variables. By contrast, logit and probit models can effectively capture non-linear relationships and interactions[21].

Hedonic Regression

The hedonic regression yielded statistically significant coefficients for 'Occupants', 'Distance', and 'Bathrooms'. Subsequent econometric tests indicated that we selected an appropriate functional form, and the model does not exhibit unrestricted heteroskedasticity. However, the EPC variables proved insignificant, suggesting their exclusion from future regressions as they do not meaningfully explain rental price variation. This outcome aligns with existing literature emphasizing structural, neighbourhood, and environmental characteristics as the most pertinent hedonic attributes[22].

We believe there is considerable scope for improving the hedonic pricing regression by integrating additional hedonic attributes based on literature review. For example, distance to public transport, distance to the local market, and proximity to green spaces[23]. Furthermore, the 'Distance' variable could be refined by calculating the distance from the property address to the town centre, rather than the centre of the property's postcode.

Expanding the data set applied to the regression either through web-scraping or primary data collection could further enhance the model's validity. The current sample size (n=76) is relatively small and likely suffers from selection bias, as noted in the data section.

General Reflections

[20] Thomson, L. (2023), "*Agenda & Papers for Meeting of Cabinet Committee of 4th May 2023.pdf*"

[21] Wooldridge, J.M. (2009). *Introductory econometrics : a modern approach*. Fourth Edition ed. Mason (Oh): South-Western, Cengage Learning. pp 575-580

[22] Owusu-Ansah A (2013), *"A review of hedonic pricing models in housing research"*

[23] CFI Team (2023) *"Hedonic Regression Method"*

Future analyses could pivot towards examining licence renewal trends. The Council's Overprovision Policy states that properties licenced prior to the policy are 'grandfathered' in, permitting them to renew their licences without restrictions, provided they adhere to the tenant safety standards. Yet, specific details surrounding the licence renewal process remain unclear. We have identified several questions, including: to what degree do grandfathered properties let their licences lapse? When a licence lapses, is it typically replaced by a new property? What conditions render these grandfathered properties ineligible for renewal? These questions may be addressed by identifying the "renewal threshold" – a defined period after which a grandfathered property must comply with the stricter criteria allegedly applicable to new property applications. Investigating these aspects and validating the Council's claims about licence renewals will offer deeper insights into factors contributing to licence attrition and the impact of the HMO Overprovision Policy.

Finally, our decision to exclusively use Stata for data analysis was highly rewarding. Firstly, the nature of *.do files facilitates auditability of our results and ensures transparency relating to our methodology. Secondly, using a common software allowed us to integrate different analyses seamlessly and collaborate through a shared OneDrive folder.

Modelling: Data

Data

Though data has been chosen to fit most closely within the academic years 2018/19 and 2022/23 this will be stylised throughout as the calendar years of 2018 and 2022 for convenience. Looking to build on previous modelling, our data was taken from the conclusions of previous progress reports, we believed that this would allow the modelling to start on a solid foundation. We utilized Ayton House data to depict the pricing dynamicsof a profit seeking organisation in the St Andrews private housing rental market. Though the halls offer a range of accommodation we chose to use the 4-bedroom data as we wanted to study the prices of HMO properties (accommodation with three or more than residents), and this is the most common HMO arrangement within St Andrews.[24] Though previous analysis has taken a mean of different accommodation options, the team has seen all the Ayton properties (with more than 2 bedrooms) have a relatively similar price variance (when compared to the price variance within model). Therefore, one metric was deemed to be sufficiently illustrative. The supply of 6802 bedrooms was calculated by the lost HMO sub team in the previous progress report.[25] The HMO provision for 2018 is assumed to be 122 greater than the 2022 value as this is the change in HMO provisions since the policy change, according to data provided by Fife Council. The price elasticity of demand (hereafter elasticity) was derived from an average of different Scottish cities. The elasticities for 2018 and 2022 of 0.67 is also an assumption.[26]

	2018 (or pre-policy value)	**2022**	**Notes**
Total Student number	8983	10468	(FTE) Full time equivalent
Student halls number	3912	4167	According to data from St Andrews University
Calculated student private market number	5071	6031	The difference between total student number and student halls number, because we assumed that all students not living in halls of residence are in demand of private housing
HMO Provision per academic year	6802	6894	According to data from Fife Council
Total Ayton House 4 bed ensuite rent (51 weeks) (£)	8304	10302	According to data from Ayton House
Monthly Rent (£)	692	859	Total rent divided by 12 rounded to the nearest pound
Elasticity	0.67		Kenneth Gibb and Daniel Mackay, The Demand for Housing in Scotland: New Estimates from the Scottish House Condition Survey, May 2001, p.13

Modelling: Methods

[24] Anne Kathyrn Chaney et al., *HMO Caps in St Andrews 2023 Candlemas Semester Progress Report*, (St Andrews, 2023), p. 55.
[25] Chaney et al. *HMO Caps 2023,* p. 17.
[26] Kenneth Gibb and Daniel Mackay, The Demand for Housing in Scotland: New Estimates from the Scottish House Condition Survey, May 2001, p.13

This model visualises the effects of demand and supply changes in a housing market, using an isoelastic demand function and a perfectly inelastic supply curve, to analyse the impact of student population changes and housing licence variations in St Andrews.

The code is written in Python and utilizes libraries NumPy, Math and Matplotlib for mathematical computations and data visualisation. The code is as follows:

```
import numpy as np
import matplotlib.pyplot as plt
import math

# Define a function to represent isoelastic demand, which depends on price (p), a
constant (a), and elasticity (elas)
def isoelastic_demand(p, a, elas):
    return a * p ** (-elas)
```

The isoelastic demand function is selected for its ability to represent different types of demand responsiveness through the elasticity parameter. This allows for flexible analysis of how demand changes in response to price variations, which is crucial in understanding housing market dynamics.

The supply curve is represented by a vertical line (meaning it is perfectly inelastic); the quantity supplied ('q_supply') remains constant and equal to the initial 'quant' value. Since we have assumed that HMO housing supply in St Andrews is fixed and does not depend on price. The code calculates 'q_supply_values' for the same ranges of prices as used for the demand curve.

```
# Define a function for vertical supply, which is independent of price and constant at
a given quantity (quant)
def vertical_supply(p, quant):
    return quant

# Function to calculate and return the 'x1' parameter, derived from quantity, price,
and elasticity in the given params dictionary
def calculate_x1(params):
    params['x1'] = params['quant'] / (params['price'] ** -params['elas'])
    return params['x1']

# Set initial parameters representing some economic scenario
original_params = {
    'quant': 6924,     # 2018 quantity
    'price': 692,      # 2018 price according to Ayton House
    'elas': 0.67,      # Assumed elasticity
    'x1': None          # Scalar value
}

# Update the original parameters with the calculated 'x1' value
original_params['x1'] = calculate_x1(original_params)

# Dictionary to hold student number data for different years
students_not_in_halls_number = {
    '2018': 5071,
    '2022': 6031,
}

# Function to simulate the effect of a demand shock, adjusting the parameters
accordingly
def counterfactual_demand_shock(params, students_not_in_halls_number):
   demand_shock = students_not_in_halls_number['2022'] /
students_not_in_halls_number['2018']
    a = params['x1']
    a *= demand_shock
    new_equilibrium_price = (params['quant'] / a) ** (-1 / params['elas'])
    return new_equilibrium_price, a
```

To assess the impact of demand shocks, such as an increase in student population, we adjust the demand curve parameters accordingly. The function demonstrates how external factors, like changes in student numbers, can influence the housing market, providing insight into potential future scenarios.

```
# Calculate new equilibrium price and adjusted constant 'a' after demand shock
new_equilibrium_price, new_a = counterfactual_demand_shock(original_params,
students_not_in_halls_number)

# Update parameters with new equilibrium price
params = original_params.copy()
params['price'] = new_equilibrium_price
params['x1'] = params['quant'] / (params['price'] ** -params['elas'])

# Modify quantity to reflect a counterfactual scenario
params['quant'] = 6802 # According to Fife Council we have lost 122 HMO licences since
2018
params['price'] = (params['quant'] / params['x1']) ** (-1 / params['elas'])
```

The model reflects real-world changes by adjusting the quantity of available housing following the loss of HMO licences due to the Overprovision policy. The adjustment, based on data from Fife Council, ensures our analysis remains representative to the actual housing market conditions.

```
# Check that the calculated price matches an expected value
# Assert round(params['price']) == 896, f"Expected price to be approximately £859, but
got £{params['price']}"
```

The model extracts the corresponding population value from 'student_population' and calculates the equilibrium price using the formula:

$$price = \left(\frac{quant}{a * population}\right)^{-\frac{1}{elas}}$$

```
# Calculate 'a' for original and updated parameters
original_a = original_params['quant'] / (original_params['price'] ** -
original_params['elas'])
updated_a = params['quant'] / (params['price'] ** -params['elas'])
```

Generating a range of price values using NumPy's 'linspace' function allows us to plot demand curves over a wide spectrum of prices. This broad range helps in understanding how demand responds not just at current market prices, but also under significantly different market conditions.

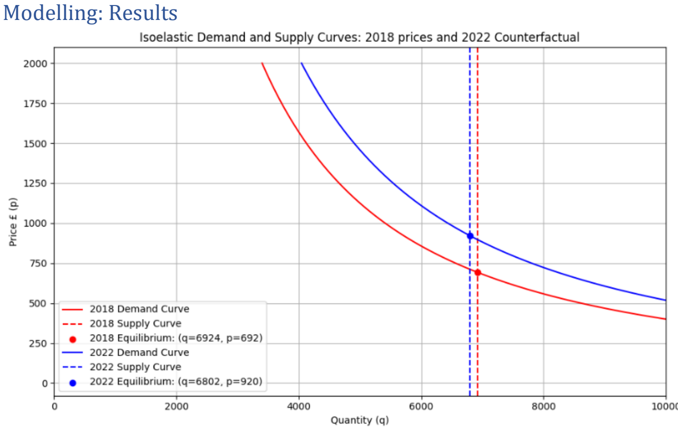
```
# Generate a range of price values for plotting
p_values = np.linspace(0, 2000, 100)

# Compute the demand quantities for original and updated scenarios across different
prices
q_demand_values_original = [isoelastic_demand(p, original_a, original_params['elas'])
for p in p_values]
q_demand_values_updated = [isoelastic_demand(p, updated_a, params['elas']) for p in
p_values]

# Begin plotting the demand and supply curves
plt.figure(figsize=(10, 6))
```

The descriptive titles and labels in the plot are to provide context and clarity.

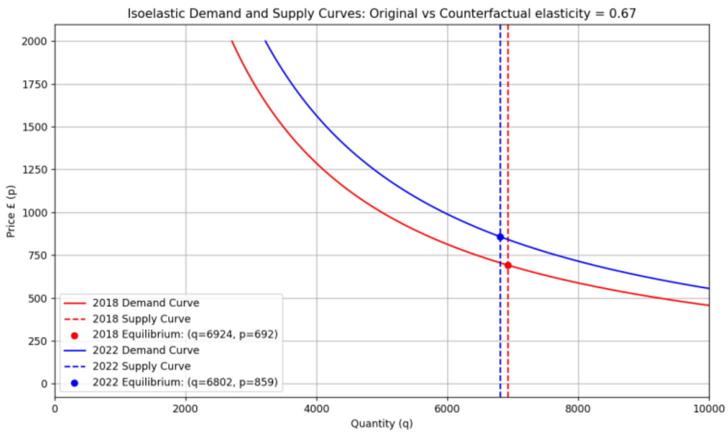
```
# Plot the 2018 demand curve and mark the equilibrium point
plt.plot(q_demand_values_original, p_values, label='2018 Demand Curve', color='red')
plt.axvline(x=original_params['quant'], color='red', linestyle='--', label='2018
Supply Curve')
```

```python
plt.scatter(original_params['quant'], original_params['price'], color='red',
label=f'2018 Equilibrium: (q={round(original_params["quant"])},
p={round(original_params["price"])})')

# Plot the 2022 demand curve and mark the new equilibrium point
plt.plot(q_demand_values_updated, p_values, label='2022 Demand Curve', color='blue')
plt.axvline(x=params['quant'], color='blue', linestyle='--', label='2022 Supply
Curve')
plt.scatter(params['quant'], params['price'], color='blue', label=f'2022 Equilibrium:
(q={round(params["quant"])}, p={round(params["price"])})')

# Setting labels, title, and grid for the plot
plt.xlabel('Quantity (q)')
plt.ylabel('Price £ (p)')
plt.title('Isoelastic Demand and Supply Curves: 2018 prices and 2022 Counterfactual.')
plt.legend()
plt.grid(True)
plt.xlim(0, 10000)

# Finalize and display the plot
plt.tight_layout()
plt.show()
```

Modelling: Results

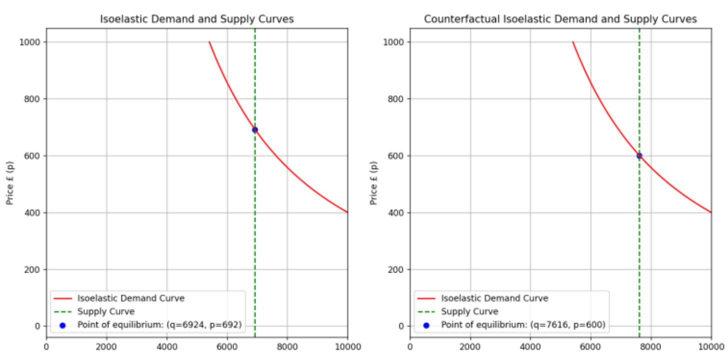


Based on the initial price of £692 and assuming perfectly inelastic supply with a price elasticity of demand at 0.67, our model predicts a price equilibrium of £920 in 2022. This implies a 39% price increase in the St Andrews rental market four-year period. Yet this model slightly overestimates Ayton House's actual rent for that year, which was £859. This can be broken down into a supply shock, which, holding other factors constant, would have generated a more moderate increase to £711, whilst a demand shock would have led to a monthly rent of £896. This disparity in the magnitude of these two shocks makes sense given the difference in their absolute values: A decrease in 122 beds compared with a 960 increase in students not in halls in the St Andrews private rental market. Thus, it is unsurprising the model suggests that demand is a much more powerful explanatory factor in the price increase than the fall in supply. Whilst a demand shock alone would have caused a 36% increase in prices; a supply shock alone would have only caused an 3% increase (see Appendix H). From this model we can theorise that the HMO cap has been a weaker factor as compared to the growth in student numbers in causing the increased housing prices, with our assumed elasticity of 0.67.

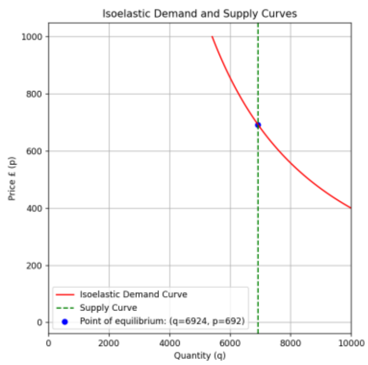


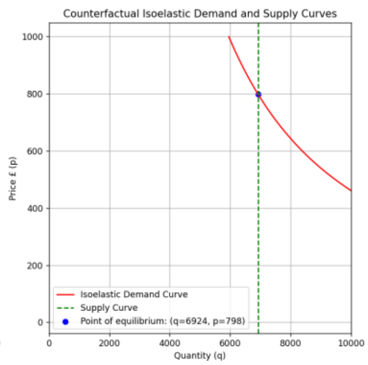
We have also found that when changing our elasticity assumption to 0.885 instead of 0.67, the price corresponds to the Ayton House price exactly. The elasticity of 0.885 suggests a relatively more elastic demand. For every 1% increase in price, the quantity demanded decreases by approximately 0.885%. An elasticity value closer to one indicates a more responsive change in demand with respect to price changes. This means in practical terms, consumers are somewhat sensitive to price changes. An elasticity of 0.67, which was the elasticity we used, indicates a less elastic demand compared to an elasticity of 0.885, suggesting that consumers are less responsive to price changes than in the case of the elasticity mentioned previously. Products with such elasticity are usually considered necessities or have fewer substitutes available, which is the case for the St Andrews housing market.

The graph below depicts the effect of a supply-side shock, highlighting the effect of a 10% increase in supply, which would lead to a decrease in price of £92, holding everything else constant.



This graph below highlights a demand-side shock, an increase in demand of 10% will lead to an increase in price of £106, holding everything else constant.





Modelling: Discussion

As previously outlined, our model may overestimate rent in St Andrews, and we attribute this to several factors: our elasticity assumption, an exaggerated supply shock, and an inflated demand shock.

Firstly, our elasticity assumption was derived as a mean from estimates of cities that are generally more self-contained. Smaller towns like St Andrews likely have more options in proximity (such as Cupar and Guardbridge), absorbing some of the demand. Adjusting for a higher elasticity, which considers the same supply and demand shocks, suggests that the equilibrium price is closer to £856. This aligns with the rate charged by Ayton House owners for that year (refer to Appendix L for the graph). Although we assumed a perfectly inelastic supply curve for simplicity, it's important to note that housing supply can vary slightly year-to-year, and this simplification might limit the model's applicability in dynamic housing markets.

Secondly, the model overestimates the reduction of rooms, failing to consider the ability of the non-HMO private market for two-bedroom student properties in St Andrews to absorb growing demand. Our model assumes that the failure to acquire HMO licenses leads to a total loss of all bedspaces, which is not representative of reality. In practice, larger houses may be put on the market for two people, resulting in a smaller decrease in supply.

Finally, our measure of St Andrews student private demand is a crude calculation of the total student population (undergraduates and postgraduates) minus university halls residents. This oversimplification arises from the lack of data on the actual student demand for private housing in St Andrews and neighbouring towns, assuming that all students not living in halls demand private housing specifically in St Andrews. This does not align with reality, as some students living in private housing may opt for accommodation in Dundee, Guardbridge, Leuchars, and other nearby towns for reasons such as cheaper rent and larger living spaces. Subsequent actions for this module might involve collecting data on the number of commuting students. This would provide a more precise estimate for those looking for housing within St Andrews. As prices rise, it is likely that more individuals are priced out of the St Andrews market, making our current measure an increasingly inappropriate proxy as the difference between our estimate and the true population parameter grows.

Outreach

Initially starting as the Survey team, we switched gears to become the Outreach team of the VIP. The switch better suited the skills of the team and become of greater significance once the VIP conference was announced.

As an Outreach team, our purpose is to take results and data from the register and modelling teams and synthesise this into easily digestible content for our audience, who likely doesn't have the technical knowledge or understanding to accurately interpret technical data produced from this project. This audience includes anyone who is affected by or has an interest in our project and its results, such as, St Andrews students, faculty, renters, landlords, homeowners, locals, etc. Our content is multifaceted, broken down into three categories: Instagram posts, website, and conference media.

Rather than to lead our audience towards a certain opinion or conclusion, our goal is to present unbiased results to equip our audience with information to form their own opinions.

On evaluating the VIPs public image, we realised that the data, which was difficult to read was going unnoticed by students who we deemed to be the target audience. This was because they would benefit from learning about the issue as it affects them so greatly. To make the data more accessible, we decided to revamp both the website and the Instagram page.

Instagram

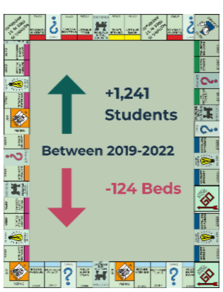
One of our main objectives this semester as the outreach team was running the project's Instagram account. The HMO and rent crises affect students, and this would be essential in inculcating awareness amongst the student body. The previous posts were solely focussed on encouraging people to fill in the survey and nothing has been posted since the closure of the survey. As such, the Instagram was out of date and did not represent the current work and findings of the VIP.

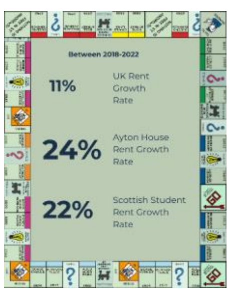
We also found that the previous posts were not very engaging; they did not illustrate the aims of the project or fully communicate what the project is about, which is perhaps some of the most vital information to get the relevant audience interested in our findings. The posts were also not the most visually appealing as there were some technical problems with viewing them (when looking at the account's profile the tops and the bottoms of posts would get cut off) and the colours used in them were not eye-catching.

To address these issues, we created a new graphic, bright, and colourful theme using Canva. For most of the semester, we were limited to only making draft posts with the formatting and themes we would use on published posts, as we were waiting to receive usable and accurate data from the register and modelling teams. To avoid misunderstanding of raw "actual" data versus proposed/ models, we came up with the idea of using different colours within our overall colour scheme to represent the varying post types: blue for a post featuring raw data, pink for a post explaining the definition of something, yellow for a post featuring a model and orange for a post featuring an update. The Canva link to these designs can be found here.

It was quickly realised that the type and quantity of content that can be produced would be limited. Our team could not simply cut and paste anything at random from past PR's. There is a burden to ensure that what the outreach team represents in its outputs is presentable. For example, a lot of the past register data is known to be inaccurate, so using any of this data to give an exact value on, for instance, the percent decrease in licenced HMO properties, would be misrepresentative. Outreach also must be careful to distinguish the difference between raw data and a model. To present a model as factual raw data to an audience with limited technical knowledge of this project would be misleading. Other constraints on Instagram content include post size - there is a limited space to present content. This content must be distilled into compact, easy to read information. Instagram content is typically consumed on small smartphone screens, so the information displayed has to be limited. Most Instagram captions are very short for this same reason, as well as for the fact that it is difficult to hold the attention of our audience on Instagram, as many viewers likely won't read a caption of any length before scrolling away. Providing a link to our website, which has more information, in both the captions of our Instagram posts and the Instagram account bio is an effective way to keep text condensed while delivering more information to viewers.

Following the conference, we created new post drafts using Canva. These were created with the "Saintopoly" theme from the poster. However, these did not get posted on the account since the data used to create these numbers is known to be inaccurate.





We have created four Instagram posts that have been posted. Two of these posts are infographics, one is the 'Saintopoly' board, and the final post is the 'Saintopoly' video. The first post answers the question "What is an HMO?", with a swipe explaining what HMO Licences are.

The second features the finalised version of the monopoly board.

The third is an "About the Project" post which contains three slides: background, project goals, and potential outcomes.

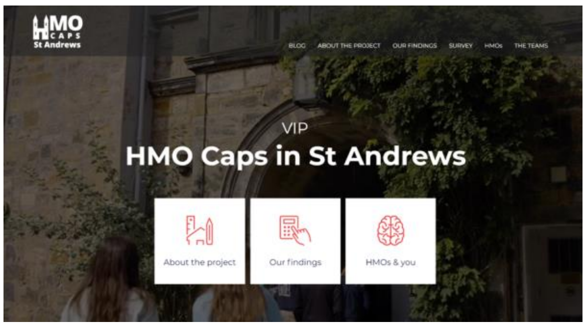
The fourth is the final version of the 'Saintopoly' video.

As touched on above, one of the greatest challenges for our outreach team is creating content and sharing information without being involved in the research. This has been a challenge for past outreach teams and will likely be a challenge for any future outreach teams. A top priority of the outreach team is to only present data that is known to be completely factual and accurate, and to present it in a clear, concise way that does not misrepresent the project's findings or does not represent a model as data or does not misrepresent an estimate as a fact. The best way for future outreach teams to ensure this priority is met is to do checks. If someone creates a draft Instagram post, have another outreach team member check that information, and have team members who created that data and who did the research confirm the post is accurate. Communication between the outreach team and the research teams is vital.

We also updated the Instagram with our rebranded project name "Housing St Andrews." We created a new logo for the profile picture, updated the bio to encompass the wider scope of our project, and changed the username to Housing St Andrews.

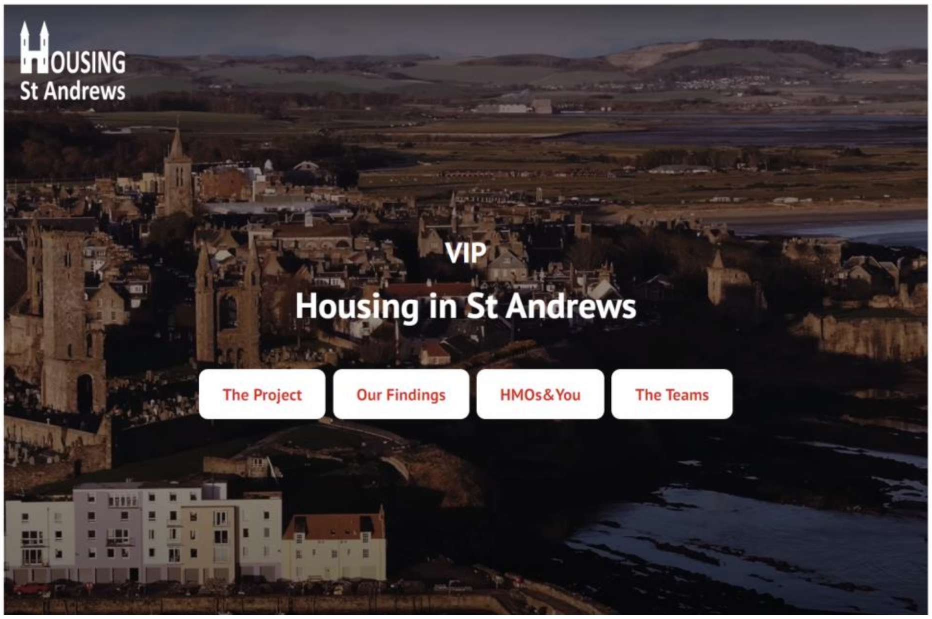
Website

The VIP had a website from last semester, but the website was quite unintuitive, as it was hard to navigate and read. As such, we decided to re-design the website, making it more intuitive and useable, while being easy to expand and iterate upon by future teams. Initially our goal was to use a software called "Wixsite", as it offers a sleek user experience.



However, due to difficulties with university integration and uncertainties regarding hosting, we moved away from that plan. Instead, we used WordPress to revamp the site, as the university already uses this service for their website hosting. The new website is a lot more intuitive to use compared to the old and offers easy navigation and access to find out more about Housing in St Andrews VIP and its findings.

Housing in St Andrews



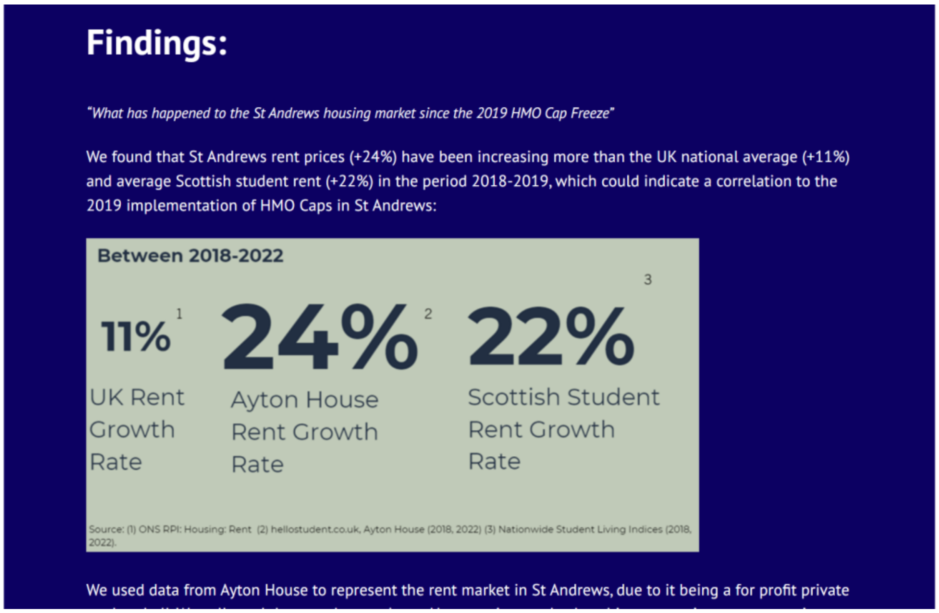
The website is used to present Housing in St Andrews VIP's findings in more detail and explain the project in a digestible way to a general audience. We also host information about HMO caps and the Housing in St Andrews VIP teams one can join. Progress reports are also accessible here. Furthermore, the website is a wonderful way for VIP conference attendees to quickly visit through a QR code and read up on the project and gain more insight than what we can communicate through a short presentation and poster. Additionally, the website serves as infrastructure for probable future survey waves, blog posts, or other ways in which Housing in St Andrews VIP wishes to interact with the wider St Andrews community.

Website Link

An ongoing challenge with the website, is getting the same "background" video for the front page, as we made for the Wix website. The main reason for having a background video on the front page, is to increase engagement and interest in the site through movement. It seems that we are not able to upload any videos to the WordPress. The video can be found here:

Website Background Video

For the conference, we had the "Our Findings" page go into more detail for the findings on our poster, being accessible with the QR code on the poster. This page would expand upon why we used the data that we did, and what it tells us in more detail. The findings page did not include any other register or modelling findings, as they were based on incorrect data. This can be changed in the future.



Conference

As an outreach team, our priority was the announced VIP conference, and we had fun and interactive plans for the same. The conference responsibilities can be categorised into 4 parts, although all are interconnected: Presentation, Script, Video & Poster. Overall, the theme of the conference responsibilities is the famous board game "Monopoly." The goal is to create a St Andrews's "Monopoly" board, using prices as an Index. It is an ideal choice as it both recognizable and relatable to the project at large. With regards to the presentation & poster, they aim to tie in elements from the board game, to follow the theme. The content for the presentation & script will be extracted from the interim PR. We then held follow-up meeting to ensure that there is no misrepresentation or misunderstanding of data.

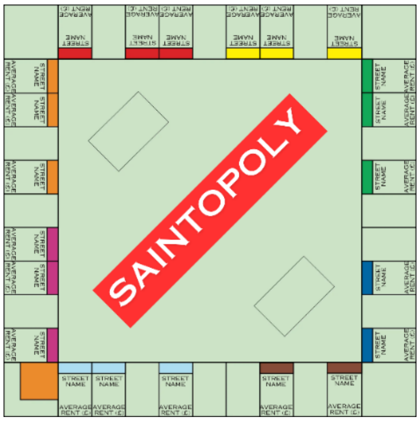
Monopoly Video

The Monopoly concept stemmed from an initial idea for a short video to be displayed at the conference. Monopoly seemed like an accessible theme for our audience as it ties in findings with a recognisable game. Our Monopoly board, however, shall be personalised to St Andrews. It shall be called 'Saintopoly' and all the streets within the board shall be the names of streets within St Andrews.

Whilst the initial concept of the Monopoly board remained the same, figuring out how best to showcase rent prices in this format was a little challenging. At first, the plan had been to showcase how the average rent in St Andrews has increased by having the video start by showing rent prices from the past. The video would then showcase how these prices have increased by showing the current rent prices. However, there was insufficient data that would allow us to make this concept a reality.

When it became clear that this initial idea would not be possible, the idea of using pound signs to represent prices increasing (£ to £££) was our next plan. This idea had the limitation of not being based on any actual data (and thus it did not represent any of our findings).

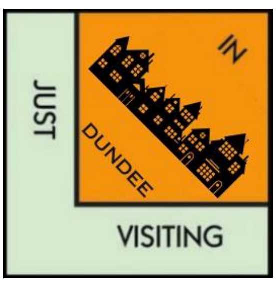
As this second idea did not feel grounded enough in any actual figures, a new idea was created. The plan was then to have the video start with just the street names seen on the 'Saintopoly' board, with the prices appearing on the streets over the course of the video. These averages were based on the responses people gave to the previous team's survey from last semester. The rough draft of the 'Saintopoly' board can be seen below.



Developing on this, a fully filled in version of the 'Saintopoly' board was created. Each 'street' square on the board featured a St Andrews street name and the average rent for that street. These averages were taken from the survey conducted by members of the VIP last academic year. The rents went from lowest to highest, much like in a real monopoly board. The colours of these 'street' squares corresponded with the prices (brown being the cheapest, blue being the most expensive). This allowed us to move forward with creating the 'Saintopoly' video.

The initial version of the 'Saintopoly' video was drafted, however, there were some issues with it. Firstly, the 'Nothing is free in St Andrews!' square was easily missed. It was also pointed out that the Dundee being prison analogy may be taken the wrong way, as it came across as too harsh in the video. This video can be seen [here](#).

As such, for the following version, the frames seen above were changed. The jailer was replaced with the monopoly man, and the man inside the prison, was replaced with a graphic of some houses (the same graphic that can be seen in the Instagram theme). There was also a larger focus put on the 'Nothing is free in St Andrews!' square. Each of these changes helped to soften the video and its message. [Here](#) is the version of the Saintopoly video shown at the conference.



The feedback from the VIP conference indicated that the flow and ending (which featured the changed Dundee joke!) were well received. However, one of the key components of the video, the squares turning blue, left the audience feeling confused. From this, it was clear that the blue squares indicating the most expensive streets on a monopoly board was not a universally known fact. As such, it was decided it would be best to cut this sequence entirely as it was lengthy and did not land. With these changes in mind, the finalised version of the 'Saintopoly' video was created. This new version featured 'chance' cards that outlined some hypothetical scenarios in which certain scenarios affected the average rent prices in St Andrews. These percentages are accurate as they are not based off of any recorded statistics. This finalised version can be viewed: [here](#).

Board Game

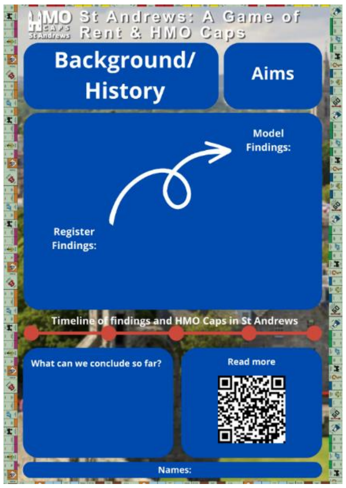
After various revisions corresponding with each change to the 'Saintopoly' video, the finalised 'Saintopoly' board was created. This finalised version can be seen below.

We had intended on having a physical board game at the conference, however, this ended up not happening. Whilst this would have been a fun and interactive addition to our display, this may have been for the best as the layout for the conference (high tables that you stand, not sit at) would have made setting up the board game a little challenging. For future teams, it would be a good idea to get in touch with those running the conference in order to find out what the layout will be, as it may change from year to year.

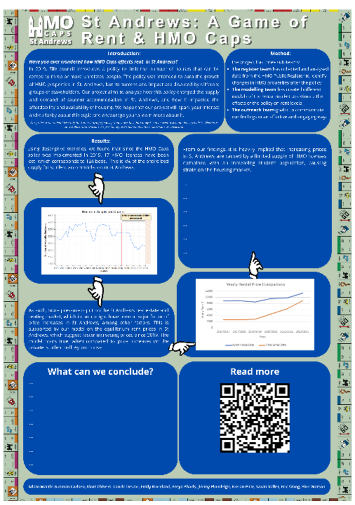
Poster

The poster told a story of increasing prices in St Andrews and the connection to supply and demand, while exploring why the locals might have wanted the HMO-cap, and what it means for students. The poster continues the monopoly theme in style. The findings for the conference and the poster where "new" findings, as we, at the time, could use neither register nor modelling findings, as the dataset they used was found to be incorrect. This made getting everything ready in time challenging.

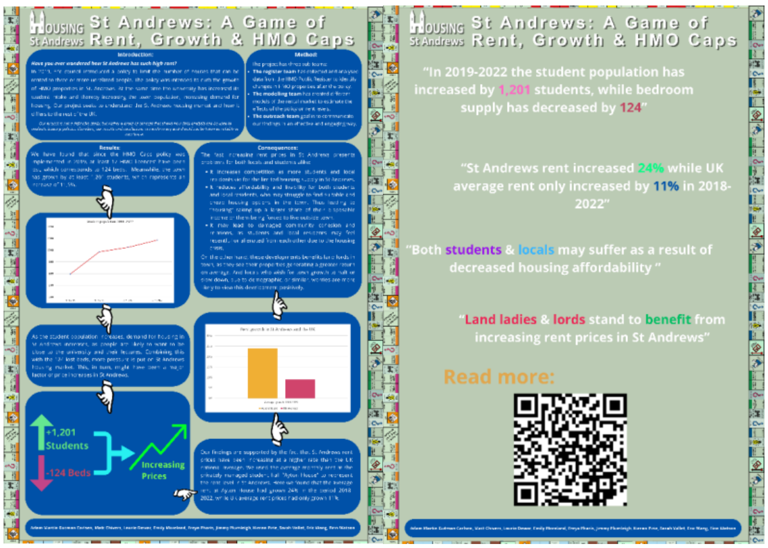
We went through the following iterations:



The first draft included a "monopoly" border to fit in with our theme. At this point it was not clear as to what story or findings we wanted to present. The poster was based on good elements from other former VIP-posters.



The following draft started telling a story about decreasing supply due to HMO licences and increasing prices. It used a graph from the register for bedroom supply, and a graph from modelling for prices. These later had to be replaced as the data was found to be incorrect. We decided to tell this story, as we deemed it easy enough for a general audience to understand while it relates to our project. The background was also changed to be less distracting, and we believe that a solid colour would print better than a picture.



In the following draft, the final story we wanted to tell was almost completed. The poster tells a story about student bedroom supply decreasing in St Andrews, while demand, through a growing student population, is increasing, thus leading to increasing prices. These price increases are then compared to the rest of the UK. It also touches on consequences for both students and locals. Also, as stated, we could not present earlier modelling and register findings based on incorrect data, and we were confident in the legitimacy of the new data for the poster. We had additionally decided on having 2 posters, the 2nd one being a poster to catch attention and also featuring the QR code. This was scrapped for the final poster, as it was deemed unnecessary. Lastly, the poster was rebranded to reflect the new VIP name - Housing in St Andrews.

For the final poster, we sought to make the design simpler and easier to read. New graphs were made that fit the theme, and a graph for rent in St Andrews was added. We restructured the story to focus on prices first, before supply and demand, as conference attendees were more likely to be able to relate to these. We split the consequences section, focusing on why locals might have wanted the HMO cap in the introduction, while discussing student consequences in the end. Additionally, we included Scottish Student Rent Growth Rate to make it more representative. Following the hand-in of the poster design for printing, we found out that the Scottish Student rent growth numbers might not have been entirely accurate, as it was based on a very small and shifting sample size, thus possibly not being very representative. We mentioned this at the conference at our poster stall.

Presentation

With regards to the Presentation, we tied in the "Monopoly" theme a lot more casually. With only two slides allowed, we had to keep it a lot more information focused. For the preparation, we had three meeting amongst members of the three teams. We coordinated data, and then finally came up with the story of an increase in prices, telling the story of supply and demand. The graphs and other data were mainly taking from the poster. It is crucial to note that there were many an unnecessary worry due to a wrong graph being placed through oversight. We now learnt to always ask those responsible to double check the graphs the provide.

Two speakers were delivered the presentation following another meeting to develop a script that accurately presented the data. A run through was done several times over teams. On the Monday of the conference, both the speakers did a run through with final edits, focusing on both timing and clarity. We definitely recommend keeping the presentation under three minutes, with two speakers. A run through should have ideally been done earlier, but the edits were manageable.

As for our own mistakes, it is critical to note that we had to improve both our horizontal & vertical communication, working better as a team. Secondly, time management was an issue too, which delayed feedback. Speaking with supervisors and each other was the way through these problems.

Lessons from the Conference

Although the conference went quite well, there are areas of improvement and lessons relevant to future teams.

Firstly, we wrote a script for interacting with people at our poster booth. We did not use this script. It isn't very natural to interact with people while using a script, and most people approached the poster on their own and then asked questions to engage with us. This script was also based on having a physical "Saintopoly" board at our booth, and using that as a point to draw in observers by asking them "do you want to play Saintopoly?" However, we were not able to create this physical board for the conference, so the main point of the script was lost. As such interactions don't happen in a vacuum, having a script is unnecessary.

As discussed above we also created a video to go along with our poster at the conference. We reached out to the organisers just before the conference to inquire about getting a monitor to display this video at our booth, however we never heard back. Our mid-conference solution was to play this video on a laptop, but we were not able to play it for the entire duration of the conference as we didn't set it up on the laptop right away. For future teams, it is recommended that you make such inquiries well in advance of the conference as communication could be slow.

It is also recommended that future teams inquire well ahead of time regarding the setup of the conference (i.e. poster placement, size, lighting, location). It is possible that these vary year to year and the more information you have prior to showing up to the conference the better prepared you can be.

Another point of note was that we were told there was a limit of two presentation slides. Our first draft of the presentation had five content slides, which we cut down to two. Yet other teams at the conference had more than two slides. It is worth clarifying this well before the conference as well.

One thing that worked very well with the presentation was having two speakers. A change in tone halfway through the presentation helps to hold your audience's attention. Each speaker presented one slide, so the transition from one speaker to another coincided with switching slides. This appeared to be effective and split the load of presenting so that one person doesn't have all the pressure.

The following are the key lessons for the poster at the conference:

We found that few, if any, people scanned the QR code. Perhaps the QR code should have been mentioned at the stall, or it could have been more visible. There is also the possibility that it is not necessary to have, if conference attendees are not interested on interacting with our findings on a deeper level.

We found that having an overall theme, "Saintopoly", was great and attracted people due to its uniqueness, funny elements and the theme being fitting overall.However, we also found that some attendees were distracted by or just focused on the theme and design. They didn't engage much with the findings themselves.

We were unfortunate with the location of our poster, and future VIP's might want to try to influence the poster location, should the conference be held at the same location. Poor lighting made the poster stand less out, and it might be a good idea for future VIPs to include a light source to illuminate the poster, should it be necessary. The poster was situated right next to the entrance of the hall, thus making it easy to miss. Additionally, people were crowded toward the opposite end of the hall where the food and drink were located.

We also found that us and attendees might have blocked the poster from view by standing too close to it, thereby decreasing people coming to view it. Maybe we should have made some kind of perimeter around the poster, to avoid visibility being blocked. Another option could have been to have two posters or a horizontal poster instead to increase the width of our stall. This would also allow for more than one team member to explain and talk about the poster to more people, as there was only room for one with our single, vertical poster. We did, however, find that those engaging with our poster stayed for quite a long time.

We had prepared a script for interacting with attendees, but we found that the poster was used as the main way of engaging with attendees. We only had a small table to use for engaging with attendees. The table was instead used to showcase our "Saintopoly" board game video.

The poster was a great success, as the Housing in St Andrews VIP ended up winning "The Best Research Display" in the Arts category, for the first time.

General Lessons

Workflow & Effective Organisation

Effective organisation emerged as a pivotal skill cultivated and applied throughout the project, facilitating the handling of multiple tasks and goals without introducing unnecessary inefficiency. Key components of effective workflow management are regular team and project-level meetings, well-structured code repositories and file management, and communication with supervisors.

Proper structuring of code repositories and file management was integral to the production and dissemination of code. This helped to streamline the work processes, minimizing confusion about file purpose and location while increasing total productivity.

Communication and Teamwork:

Much of the project's success hinged on cultivating a robust communication framework, encompassing both internal team dynamics and project-level interactions. Frequent meetings allowed team members to harmonize their weekly work, ensuring that teams shared a common understanding of the week's objectives. These meetings also enabled efficient delegation of responsibilities among team members.

The VIP project is a group project, implying that communication is the key. It is important to take the advantage of having others to help you and give you their opinions on your work can only be beneficial. Each person has unique strengths and weaknesses, which means that we can learn from one another.

Finally, engaging in consistent and open communication with supervisors provided indispensable direction, helping teams navigate the expansive research landscape. Such communication enabled teams to stay aligned with the goals of the project and provided guidance in overcoming roadblocks and confusion.

Managing workflow can be a challenging with everyone having different schedules, assignments, and unexpected issues that can arise throughout the week. Though it can be tough to get into the routine of it, the best way to combat this is through constant communication between team members. This way, other members of the team can assist the member in need if they are unable to complete their portion of the tasks for that week.

Time Management

Last minute work often led to delays in feedback or unnecessary complications. Working at the last hour, when you have several teammates, often led to discrepancies and inefficiency. Setting interim deadlines helped avoid a last-minute scramble to produce work.

Research, Analysis and Quantitative skills:

This project allowed us to have a better understanding of the housing market in St Andrews. This subject often being the centre of discussion it was a great opportunity to analyse and observe the elements influencing that trend. to gain some skills. The different teams used several coding programs for quantitative analysis and find evidence to support our research. This semester helped and gave us an introduction on how to use some tools to come with quantitative analysis and result.

Future Work

Register

Identify additional independent variables in the selective regression

The current regression models are comprised of only two categories of independent variable: bedroom type and distance. In order to improve explanatory power, we plan to introduce new independent variables. Such variables might encompass rent prices or property values as causal factors of licence attrition. Acquiring rent data poses a challenge due to its limited availability; however, other factors like relative house value (as derived from council tax bands) could serve as proxy variables.

Generate additional descriptive statistics

In addition to using the new independent variables, expanding our current descriptive statistics to offer further insights into the effect of the policy on the rental market may be fruitful for future teams. While most of our descriptive statistics are currently grounded on whether a property has an active licence or not, future statistics may illustrate other changes in the housing market such as the evolution of licence renewal applications. The forthcoming Fife Council data is expected to provide more information on licence renewals, facilitating the generation of additional statistics to broaden the scope of how the market has responded to the policy.

Web scraping of archived websites to create historical rent data set

The robustness of current rent indices should be improved for enhanced accuracy in future analysis. Several promising data sources have been found including Lawson and Thompson's historical rent data from 2017 which enables analysis of their rent prices over time. In addition, web archives of the university website may provide valuable sources of historical student housing prices. Future groups would benefit from gathering additional online sources of historical rent price to strengthen price indices. In future analyses, the change and evolution of rent data could yield important findings related to the overprovision policy although distinguishing regulatory-linked inflation and general trends in inflation could be difficult.

Extending the number of observations in the hedonic price regression dataset

This semester we limited our regression to the properties that were posted on Lawson and Thompson and the University website. The team believes that having more observations can be beneficial by improving the robustness of the hedonic regression. Future teams can use other estate agent websites offering private accommodation for students, such as Rollo's and Thorntons, to scrape information for new observations.

Follow-up with the Overprovision Review

The Council announced that they will be allocating an additional 15 new HMO licences for university managed properties to accommodate 45 to 60 additional students. It would potentially prove fruitful for future register teams to follow up on this development to see if the policy revision achieves its objective or not.

Improve functional form and specification of hedonic regression

Research indicates that employing natural logarithms in hedonic regression models can lead to more robust coefficients and simplify interpretation[27]. Among these, the Log-level OLS Model stands out as a widely used and effective approach in hedonic pricing. Consequently, integrating logged rent prices into the model could prove to be beneficial.

Modelling

Implementation of new data:

Implement halls of residence data, student population data and university managed properties data from Freedom of Information into new models (e.g., discrete modelling).

Analysing the trends in Student Population:

Work on hall student population data, discover trends of change in population in specific halls and areas over years.

Outreach:

Website

Results and findings on the website should cover more findings and go more into detail with them than what we were able to do at the conference. The information uploaded here should be easier to digest and understand than the progress report.

Instagram

The focus of the Instagram should continue establishing the 'basics' for understanding the aims of the project before delving more into posting any results. This gives those in charge of the Instagram more time to determine whether the results they are posting are suspect or not and will make the project more easily understandable for a general audience.

Website Management Guide

Following the conference, a guide for following HMO VIP Outreach Teams needs to be created. This should include design language of the website, how to edit with WordPress, how to add new pages and navigation links for them, and suggestions for website content they could create which our team did not prioritise.

[27] Owusu-Ansah A (2013), *"A review of hedonic pricing models in housing research"*

Blog

Low priority future work includes a blog on the website. This blog would serve to complement the Instagram. The idea is to have an accompanying blog post for each Instagram post, where people can read more about the specific content, they were exposed to in the Instagram post.

References

Anderson, H (2018) "*Value of nature implicit in property prices – Hedonic Pricing Method (HPM) methodology note".* Available at: https://www.ons.gov.uk/economy/environmentalaccounts/methodologies/valueofnatureimplicitinpropertypriceshedonicpricingmethodhpmmethodologynote

Alexander, J. et al (2022) *"HMO CAPS IN ST ANDREWS PROGRESS REPORT 2 CANDLEMAS SEMESTER 2021/22"* Available at: https://vip.wp.st-andrews.ac.uk/files/2023/03/April-2022-Progress-Report.pdf

CFI Team (2023) *"Hedonic Regression Method".* Available at: https://corporatefinanceinstitute.com/resources/data-science/hedonic-regression-method/

Chaney, A et al (2023), *"HMO Caps in St Andrews 2023 Candlemas Semester Progress Report"* Available at: https://vip.wp.st-andrews.ac.uk/files/2023/11/HMO_VIP_2022_23_S2_ProgressReport.pdf

Dashe, W. (2023) *St Andrews HMO Cap Creates 'locked bedroom' scenario, Councillor says*. The Saint [online]. Available at: https://www.thesaint.scot/post/st-andrews-hmo-cap-creates-locked-bedrooms-scenario-councillor-says

Fife Council (2022) "HMO Charging Structure". Available at: https://www.fife.gov.uk/__data/assets/pdf_fil0020/359120/HMOFees.01.06.22-1.pdf

Fife Council (2021) "Houses in Multiple Occupation (HMO) Policy Statement". Available at: https://www.fife.gov.uk/__data/assets/word_doc/0027/84519/HMO-Policy-Statement-V1.1.docx

Fife Council (2023), "Cabinet Committee Meeting Minutes". Available at: https://www.fife.gov.uk/__data/assets/pdf_file/0037/473779/Agenda-and-Papers-for-Meeting-of-Cabinet-Committee-of-4-May-2023.pdf

Fife Council (2023) "HMO Public Register – Q2 2023". Available at: https://www.fife.gov.uk/__data/assets/pdf_file/0017/470303/Q2-2023-HMO-Public-Register.pdf

Huh, S. and Kwak, S. (1997) 'The Choice of Functional Form and Variables in the Hedonic Price Model in Seoul', *Urban Studies*, 34(7), pp. 967-1149.

Kenneth Gibb and Daniel Mackay, The Demand for Housing in Scotland: New Estimates from the Scottish House Condition Survey, May 2001, p. 13.

Scottish Government (2012) "Licensing of houses in multiple occupation: statutory guidance for Scottish local authorities". Available at: https://www.gov.scot/publications/licensing-multiple-occupied-housing-statutory-guidance-for-scottish-local-authorities/

Scottish Government (2006) *"Housing (Scotland) Act 2006".* Available at: https://www.legislation.gov.uk/asp/2006/1/contents

St Andrews QV (2018) "St Andrews HMO problem; where we are and how we got there". Available at: https://standrewsqv.org.uk/2018/11/the-st-andrews-hmo-problem-where-we-are-and-how-we-got-there/

The Scottish Government (2012), "LICENCING OF HOUSES IN MULTIPLE OCCUPATION: STATUTORY GUIDANCE FOR SCOTTISH LOCAL AUTHORITIES" Available at: https://www.gov.scot/publications/licensing-multiple-occupied-housing-statutory-guidance-for-scottish-local-authorities/

Micallef, B., Ellul, R. and Debono, N. (2021) 'A hedonic assessment of the relative importance of structural, locational and neighbourhood factors on advertised rents in Malta', *International Journal of Housing Markets and Analysis*, 15(1), pp. 203-230.

National Records of Scotland (2022) "Mid-2020 Population Estimates Scotland". Available at: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/mid-2020

Owusu-Ansah A (2013) *"A review of hedonic pricing models in housing research".* Available at: https://www.researchgate.net/profile/Anthony-Owusu-Ansah-2/publication/287232776_A_review_of_hedonic_pricing_models_in_housing_research/links/5b8f915945851540d1c9dd54/A-review-of-hedonic-pricing-models-in-housing-research.pdf

Thomson, L. (2023) "*Agenda & Papers for Meeting of Cabinet Committee of 4th May 2023".* Available at: https://www.fife.gov.uk/__data/assets/pdf_file/0037/473779/Agenda-and-Papers-for-Meeting-of-Cabinet-Committee-of-4-May-2023.pdf

University of St Andrews (2023) Freedom of information request relating to student population between 2010/11 and 2023/24

Vemru, A. et al. (2022*) "HMO Caps in St Andrews 2022 Martinmas Semester Progress Report",* Available at: https://vip.wp.st-andrews.ac.uk/files/2023/03/October-2022-Progress-Report.pdf

Wooldridge, J. (2010) *"Econometric Analysis of Cross Section and Panel Data",* Second Edition: The MIT Press

Wooldridge, J. (2009) "*Introductory econometrics: a modern approach"*, Fourth Edition. Mason (Oh): South-Western, Cengage Learning.

Appendices

Appendix A: Creation of the postcode and distance to Tesco variables

```
// Continuous distance to Tesco variable generation

import excel "`data'Postcode Continuous Variable.xlsx", firstrow clear
drop Description
gen postcode_merge = substr(Postcode, 6, .)

save Postcode.dta, replace
import excel "`data'\HMO Full Data Set WIth Kategorical Variable.xlsx", sheet("List of
Addresses") cellrange(A2:FR9502) firstrow clear
drop if Occupants==.

* Extract the postcode from the HMOAddress variable
gen Postcode_original = " "
gen KY16_position = strpos(HMOAddress, "KY16")
replace Postcode_original = substr(HMOAddress, KY16_position, 8) if KY16_position > 0
drop KY16_position
gen postcode_merge = substr(Postcode, 6, .)

merge m:1 postcode_merge using "Postcode.dta"
drop if Occupants==.
drop _merge
drop postcode_merge
drop Postcode
gen serial =_n
label variable DistancetoTown "Walking distance, km"
order serial HMOAddress Occupants DistancetoTown Postcode_original DateIssued
save Kategorical_clean.dta, replace
```

Appendix B: Data cleaning and transformation of HMO Public list 02 03 2023.xls

```
gen serial =_n
order serial

// Fill app AppRefNumber --> no empty cells
replace AppRefNumber = AppRefNumber[_n-1] if AppRefNumber[_n] == ""

// Distinguishes between duplicate lines and sort accordingly
gen sort_helper = ApplicantName == ""
sort AppRefNumber sort_helper ApplicantNames
drop if HMOAddress == "HMO Address"
drop sort_helper serial


// Create serial number
g sr_no = _n

// Create application num which distinguishes which rows below to which property
gen application_num = .
replace application_num = 1 in 1
replace application_num = cond(HMOAddress == HMOAddress[_n-1], application_num[_n-1], application_num[_n-1]+1) in 2/l


// Create applicationMax (how many duplicate rows for the license) and applicationwithinno (the specific row within a license)
variables
gsort application_num sr_no
by application_num : gen applicationwithinno  = _n, a(application_num)
by application_num : egen applicationMax  = max(applicationwithinno)

// To consolidate all info on a single row, "position_change" is the number of rows each row needs to go up to be on the first
row in a license
gen position_change = applicationwithinno[_n] - 1

// Reorder for ease of analysis
order AppRefNumber application_num applicationMax applicationwithinno position_change sr_no ApplicantNames WARD HMOAddress
AgentName LicenseStatus DateofApp DateIssued ExpireDate TotOccs Decision

// Generates ApplicantName variables so that all names can be consolidated on a single row
gen ApplicantName2 = ""
gen ApplicantName3 = ""
gen ApplicantName4 = ""

// Loop through all name variables
forval x = 2/4{
    // Assigns appropriate name to the correct row as specified by "position_change"
    replace ApplicantName`x' = cond(`x' == applicationwithinno[_n] & application_num[_n] == application_num[_n-position_change[
_n]], ApplicantNames[_n], "")

    // shift all Names in applicant2 up 1 row, applicant 2 up 2 rows, etc
    replace ApplicantName`x' = ApplicantName`x'[_n+`x'-1] if _n < _N
}

// Consolidate agent names to the same level
replace AgentName = cond(applicationwithinno == applicationMax, AgentName[_n-position_change[_n]], AgentName[_n-position_change
[_n]+1] )
```

```
// Removes duplicate rows. All licenses now consolidate to single row
keep if applicationwithinno == 1

// Creates new address variables (removing things like St. Andrews, Fife, etc) which frequently lead to varied naming for the
same property. This enables us to perform functions to check whether something is the same property in the future
gen Postcode_original = ""
gen KY16_position = strpos(HMOAddress, "KY")
replace Postcode_original = substr(HMOAddress, KY16_position, .) if KY16_position > 0
drop KY16_position
sort Postcode_original

order AppRefNumber ApplicantNames ApplicantName2 ApplicantName3 ApplicantName4 AgentName sr_no application_num applicationMax
applicationwithinno position_change    WARD HMOAddress Postcode_original LicenseStatus DateofApp DateIssued ExpireDate TotOccs
Decision

// Clean HMO Names so that renewed licenses from the same property can be identified
gen postcodelength = strlen(Postcode_original), a(Postcode_original)
gen HMOName1 = substr(HMOAddress, 1, strlen(HMOAddress)-postcodelength), a(HMOAddress)
gen HMOName2 = subinstr(HMOName1, "St Andrews", "", .), a(HMOName1)
gen HMOName3 = subinstr(HMOName2, "Fife", "", .), a(HMOName2)


// Clean data, get rid of variables not be used anymore
drop sr_no application_num applicationwithinno applicationMax position_change

// Sort by HMOName (licenses held by the same property next to each other)
order HMOName3
sort HMOName3
*g sr_no = _n, a(HMOName3)

// Generates license num
gen license_num = ., a(HMOName3)
replace license_num = 1 in 1
replace license_num = cond(HMOName3[_n] == HMOName3[_n-1], license_num[_n-1], license_num[_n-1]+1) in 2/l

// Generates license_num which indicates which # license of a property it is
gsort license_num license_num
by license_num : gen proplicenseno  = _n, a(license_num)
// generates MaxLicense which shows how many licenses the property has
by license_num : egen MaxLicense  = max(proplicenseno)
// Generates position_change: serves same purpose of previous variable of the same name
gen position_change = proplicenseno[_n] - 1

// Beginning of the process to converting from license to property level data

// Generates variables for the details (application date, issue date, expiration date, etc) for each license a property has
forval num = 2/4{
    gen DateofApp`num' = ""
    gen ExpireDate`num' = ""
    gen DateIssued`num' = ""
    gen AppRefNumber`num' = ""
    gen Decision`num' = ""
    gen LicenseStatus`num' = ""
}

// Reoder for ease of viewing
sort HMOName3
order DateofApp DateIssued ExpireDate DateofApp2 DateIssued2 ExpireDate2 DateofApp3 DateIssued3 ExpireDate3 DateofApp4
DateIssued4 ExpireDate4 proplicenseno position_change

// Consolidates DateIssued, DateofApp, ExpireDate variables to their respective variable and move up to single row
forval x = 2/4{
    // If a property has multiple licenses these variables capture the information about their different licenses
    replace DateIssued`x' = cond(`x' == proplicenseno[_n] & license_num[_n] == license_num[_n-position_change[_n]], DateIssued[
_n], "")
    // Ex: if A property has two licenses, DateofApp2 will capture the application date of the second license
    replace DateofApp`x' = cond(`x' == proplicenseno[_n] & license_num[_n] == license_num[_n-position_change[_n]], DateofApp[_n
], "")
    replace ExpireDate`x' = cond(`x' == proplicenseno[_n] & license_num[_n] == license_num[_n-position_change[_n]], ExpireDate[
_n], "")
    replace AppRefNumber`x' = cond(`x' == proplicenseno[_n] & license_num[_n] == license_num[_n-position_change[_n]],
AppRefNumber[_n], "")
    replace Decision`x' = cond(`x' == proplicenseno[_n] & license_num[_n] == license_num[_n-position_change[_n]], Decision[_n],
 "")
    replace LicenseStatus`x' = cond(`x' == proplicenseno[_n] & license_num[_n] == license_num[_n-position_change[_n]],
LicenseStatus[_n], "")

    // shift all info up to single row
    replace DateIssued`x' = DateIssued`x'[_n+`x'-1] if _n < _N
    replace DateofApp`x' = DateofApp`x'[_n+`x'-1] if _n < _N
    replace ExpireDate`x' = ExpireDate`x'[_n+`x'-1] if _n < _N
    replace AppRefNumber`x' = AppRefNumber`x'[_n+`x'-1] if _n < _N
    replace Decision`x' = Decision`x'[_n+`x'-1] if _n < _N
    replace LicenseStatus`x' = LicenseStatus`x'[_n+`x'-1] if _n < _N

}
```

```
// Now, transforming to wideform data
// Turn total occupants variable from string into float variable
destring(TotOccs), replace
summarize TotOccs

// Convert the relevant date variables into Stata date format
gen DateIssued_new = date(DateIssued, "DMY")
gen ExpireDate_new = date(ExpireDate, "DMY")
forval x = 2/4{
    gen ExpireDate`x'_new = date(ExpireDate`x', "DMY")
    gen DateIssued`x'_new = date(DateIssued`x', "DMY")
    gen DateofApp`x'_new = date(DateofApp`x', "DMY")
}

// For each month from 2017 to 2026 it generates a binary variable which = 1 if the license is active or = 0 if it is not
forval year = 2017/2026 {
    forval month = 1/12 {
        if `month' >= 10 {
            gen Y`year'`month' = "15/`month'/`year'"
            gen Y`year'`month'_2 = date(Y`year'`month', "DMY")
            drop Y`year'`month'
            gen Y`year'`month'_3 = 0

            replace Y`year'`month'_3 = 1 if Y`year'`month'_2 >= DateIssued_new
            replace Y`year'`month'_3 = 0 if Y`year'`month'_2 >= ExpireDate_new

            forval license = 2/4{
                replace Y`year'`month'_3 = 1 if Y`year'`month'_2 >= DateIssued`license'_new
                replace Y`year'`month'_3 = 0 if Y`year'`month'_2 >= ExpireDate`license'_new
            }
            drop Y`year'`month'_2
            rename Y`year'`month'_3 Y`year'`month'
        }
        else {
            gen Y`year'0`month' = "15/`month'/`year'"
            gen Y`year'0`month'_2 = date(Y`year'0`month', "DMY")
            drop Y`year'0`month'
            gen Y`year'0`month'_3 = 0

            replace Y`year'0`month'_3 = 1 if Y`year'0`month'_2 >= DateIssued_new
            replace Y`year'0`month'_3 = 0 if Y`year'0`month'_2 >= ExpireDate_new

            forval license = 2/4{
                replace Y`year'0`month'_3 = 1 if Y`year'0`month'_2 >= DateIssued`license'_new
                replace Y`year'0`month'_3 = 0 if Y`year'0`month'_2 >= ExpireDate`license'_new
            }
            drop Y`year'0`month'_2
            rename Y`year'0`month'_3 Y`year'0`month'
        }
    }
}

// Drop non property-level properties
drop if proplicenseno != 1
```

```
// Drop variances which don't serve a purpose anymore
drop proplicenseno position_change license_num MaxLicense HMOName1 HMOName2 postcodelength DateIssued_new ExpireDate_new
ExpireDate_new ExpireDate2_new DateIssued2_new DateofApp2_new ExpireDate3_new DateIssued3_new DateofApp3_new ExpireDate4_new
DateIssued4_new DateofApp4_new

// Order for optimal viewing
order HMOName3 Postcode_original WARD  TotOccs ApplicantNames ApplicantName2 ApplicantName3 ApplicantName4 AgentName
AppRefNumber AppRefNumber2 AppRefNumber3 AppRefNumber4 DateofApp DateIssued ExpireDate DateofApp2 DateIssued2 ExpireDate2
DateofApp3 DateIssued3 ExpireDate3 DateofApp4 DateIssued4 ExpireDate4 LicenseStatus Decision LicenseStatus2 Decision2
LicenseStatus3 Decision3 LicenseStatus4 Decision4

// Save datafile for faster computation in the future: now you can just import it
cd "`data'"
save Cleaned2018ListDataWideForm, replace

// Reshape to long form format
gen id = _n  // Create a unique identifier for each observation
reshape long Y, i(id) j(Time)
drop id  // Drop the temporary identifier variable
rename Y Active
gen Year = int(Time / 100)
gen Month = mod(Time, 100)
drop Time
order Year Month

// Save for faster computation in the future
save Cleaned2018ListDataLongForm, replace
```

Appendix C: Code generating yearly active licences graph.

```
// Import "Kategorical" Dataset
import excel "HMO Full Data Set WIth Kategorical Variable.xlsx", sheet("List of Addresses")
cellrange(A2:FR9502) firstrow

// Creating yearly active indicators and calculating totals
forval year = 2009/2023 {
    // Generating an indicator for active properties in any quarter of the year
    gen active_`year' = (Q1`year' == "1" | Q2`year' == "1" | Q3`year' == "1" | Q4`year' == "1"
 | Q1`year' == "ERROR" | Q2`year' == "ERROR" | Q3`year' == "ERROR" | Q4`year' == "ERROR")

    // Summing up the indicators to get the total active properties for the year
    egen total_`year' = total(active_`year')
    // redefine as scalar so that data can survive the "clear" function below
    scalar activelic_`year' = total_`year'

}

// Creating a dataset for the graph
clear
set obs 15
gen year = 2009 + _n - 1
gen value = .

// Create counter variable for position
local counter = 1

// Loop to assign activelic_`year' to value for graph creation
forval year = 2009/2023 {
    replace value = activelic_`year' in `counter'
    local counter = `counter' + 1
}

// Create graph
twoway (line value year, xlab(2009(1)2023) xtick(2009(1)2023) ylab(400(100)1000) ytick(400(100
)1000)), title("Active HMO Licenses per Year") xtitle("Year") ytitle("Active Licenses")

// Display the table of yearly totals
list year value
```

Appendix D: Code generating quarterly active licences graph.

```
// Initializing total active licenses for each quarter of each year
forval year = 2009/2023 {
    forval qtr = 1/4 {
        scalar total_Q`qtr'`year' = 0
    }
}

// Counting active licenses
forval property = 1/2210 {
    forval year = 2009/2023 {
        forval qtr = 1/4 {
            if Q`qtr'`year'[`property'] == "1" | Q`qtr'`year'[`property'] == "ERROR" {
                scalar total_Q`qtr'`year' = total_Q`qtr'`year' + 1
            }
        }
    }
}

// Creating a dataset for the graph
clear
set obs 60  // 15 years * 4 quarters
gen quarter_year = .
gen value = .

// Loop through the quarters and set values from the scalar variables
local i = 1
forval year = 2009/2023{
    forval qtr = 1/4{
        replace quarter_year = `year' + (`qtr' - 1) * 0.25 in `i'
        replace value = total_Q`qtr'`year' in `i'
        local i = `i' + 1
    }
}

// Creates the linegraph. Note: tikmarks for each quarter because adding a label for each
quarter would make the graph labels overlap/ugly
twoway (line value quarter_year, xlab(2009(1)2023) xtick(2009(.25)2023) ylab(200(100)900)
ytick(200(100)900)), ///
       title("Active Licenses per Quarter") xtitle("Quarter") ytitle("Active Licenses")
```

Appendix E: Code generating Quarterly bedroom supply graph.

```
// Initializing total active licenses for each quarter of each year
forval year = 2009/2023 {
    forval qtr = 1/4 {
        scalar total_Q`qtr'`year' = 0
    }
}

// Counting active licenses
forval property = 1/2210 {
    forval year = 2009/2023 {
        forval qtr = 1/4 {
            if Q`qtr'`year'[`property'] == "1" | Q`qtr'`year'[`property'] == "ERROR" {
                scalar total_Q`qtr'`year' = total_Q`qtr'`year' + Occupants[`property']
                display total_Q`qtr'`year'
            }
        }
    }
}
// Creating a dataset for the graph
clear
set obs 60  // 15 years * 4 quarters
gen quarter_year = .
gen value = .

// Loop through the quarters and set values from the scalar variables
local i = 1
forval year = 2009/2023{
    forval qtr = 1/4{
        replace quarter_year = `year' + (`qtr' - 1) * 0.25 in `i'
        replace value = total_Q`qtr'`year' in `i'
        local i = `i' + 1
    }
}

// Creates the linegraph. Note: tikmarks for each quarter because adding a label for each
quarter would make the graph labels overlap/ugly
twoway (line value quarter_year, xlab(2009(1)2023) xtick(2009(.25)2023) ylab(1000(500)4500)
ytick(1000(500)4500)), ///
       title("Bedroom Supply per Quarter") xtitle("Quarter") ytitle("Number of Available
Bedrooms")
```

Appendix F: Code generating the graph of each bedroom type as a proportion of total active licences.

```
// Generate variables used
forval year = 2010/2023{
    forval qtr = 1/4{
        scalar numberof3bed_Q`qtr'`year' = 0
        scalar numberof4bed_Q`qtr'`year' = 0
        scalar numberof5bed_Q`qtr'`year' = 0
        scalar numberof6plusbed_Q`qtr'`year' = 0
        scalar numberofall_Q`qtr'`year' = 0
    }
}

// Loop over years and quarters
forval year = 2010/2023 {
    forval quarter = 1/4 {
        scalar numberofall = 0
        // Loop over properties
        forval property = 1/2210 {
            scalar qtr_active = Q`quarter'`year'[`property']

            // If property active
            if qtr_active == "1" | qtr_active == "ERROR" {
                scalar numberofall = numberofall + 1

                // Aggregating # of properties w/ x bedroom count based on capacity
                if Occupants[`property'] == 3 {
                    scalar numberof3bed_Q`quarter'`year' = numberof3bed_Q`quarter'`year' + 1
                }
                if Occupants[`property'] == 4 {
                    scalar numberof4bed_Q`quarter'`year' = numberof4bed_Q`quarter'`year' + 1
                }
                if Occupants[`property'] == 5 {
                    scalar numberof5bed_Q`quarter'`year' = numberof5bed_Q`quarter'`year' + 1
                }
                if Occupants[`property'] >= 6 {
                    scalar numberof6plusbed_Q`quarter'`year' = numberof6plusbed_Q`quarter'`year' + 1
                }
            }
        }
        // Calculating proportions
        scalar prop3bed`quarter'`year' = numberof3bed_Q`quarter'`year'/numberofall
        scalar prop4bed`quarter'`year' = numberof4bed_Q`quarter'`year'/numberofall
        scalar prop5bed`quarter'`year' = numberof5bed_Q`quarter'`year'/numberofall
        scalar prop6plusbed`quarter'`year' = numberof6plusbed_Q`quarter'`year'/numberofall
    }
}
```

```
// Creating a dataset for the graph
clear
set obs 56  // 13 years * 4 quarters
gen quarter_year = .
gen value3 = .
gen value4 = .
gen value5 = .
gen value6plus = .

// Create time and value variables for graph
local i = 1
forval year = 2010/2023{
    forval qtr = 1/4{
        replace quarter_year = `year' + (`qtr' - 1) * 0.25 in `i'
        replace value3 = prop3bed`qtr'`year' in `i'
        replace value4 = prop4bed`qtr'`year' in `i'
        replace value5 = prop5bed`qtr'`year' in `i'
        replace value6plus = prop6plusbed`qtr'`year' in `i'
        local i = `i' + 1
    }
}

// Create graph
twoway (line value3 quarter_year, lc(red) lw(thin)) || ///
       (line value4 quarter_year, lc(blue) lw(thin)) || ///
       (line value5 quarter_year, lc(green) lw(thin)) || ///
       (line value6plus quarter_year, lc(purple) lw(thin)), ///
       legend(label(1 "3 Bedroom Properties") label(2 "4 Bedroom Properties") label(3 "5 Bedroom Properties") label(4
"6 or more Bedroom Properties")) ///
       xlab(2010(1)2023) xtick(2010(.25)2023) ylab(0(.1).6) ytick(0(.1).6) ///
       title("Each Bedroom Type as a Proportion of Total Active Licenses") xtitle("Quarter") ytitle("Proportion")
```

Appendix G: Code generating average bedrooms per property graph

```
// Generate variable used
forval year = 2010/2023{
    forval qtr = 1/4{
        scalar running_total_Q`qtr'`year' = 0
    }
}

// loop over all years and quarter
forval year = 2010/2023{
    forval quarter = 1/4{
        scalar number_active = 0
        //Loop over all rows
        forval property = 1/2210{
            // Logic checking if property is active
            if Q`quarter'`year'[`property'] == "1" | Q`quarter'`year'[`property'] == "ERROR"{
                //Total active licenses
                scalar number_active = number_active + 1
                //Total bedrooms
                scalar running_total_Q`quarter'`year' = running_total_Q`quarter'`year' + Occupants[`property']
            }
        }
        //Calclulate average
        scalar avg_Bed_Q`quarter'`year' = running_total_Q`quarter'`year'/number_active
    }
}

clear
set obs 56  // 13 years * 4 quarters

// create variables for graph use
gen quarter_year = .
gen value = .

local i = 1
// update time variable and assign the avg scalar to the "value" variable
forval year = 2010/2023{
    forval qtr = 1/4{
        replace quarter_year = `year' + (`qtr' - 1) * 0.25 in `i'
        replace value = avg_Bed_Q`qtr'`year' in `i'
        local i = `i' + 1
    }
}

// Create graph
twoway (line value quarter_year, xlab(2010(1)2023) xtick(2010(.25)2023) ylab(3(.25)6) ytick(3(.25)6)), ///
       title("Average Number of Bedrooms per Property") xtitle("Quarter") ytitle("Average Bedrooms")
```

Appendix H: Probability distribution of walking distance to town.

```
/*set working directory*/
cd "`stata_wd'"

/* Import Excel file */
import excel "`data'Postcode Continuous Variable.xlsx", firstrow clear

/*Calculate the descriptive statistics*/
summarize DistancetoTown,
summarize DistancetoTown, detail

/* Create a frequency table with specified intervals*/
tabulate DistancetoTown


/*Create Histogram*/
histogram DistancetoTown, bin(28) freq title("Properties in St. Andrews")
egen DistancetoTown_counts = bin(28)

/*Additional info on histogram*/
title (propertiesinstandrews)
xtitle ("Distance to Town")
ytitle ("Number of Properties")
xlabel(0(1)28)
```

Appendix I: Code generating the distribution of the renewal gap.

```
*Calculate time differences between expirations and next issue dates

* generate float date variables for calculation (they equal like days since jan 1, 1960
or something like that)
gen DateIssued_num = date(DateIssued, "DMY")
gen ExpireDate_num = date(ExpireDate, "DMY")

gen DateIssued2_num = date(DateIssued2, "DMY")
gen ExpireDate2_num = date(ExpireDate2, "DMY")

gen DateIssued3_num = date(DateIssued3, "DMY")
gen ExpireDate3_num = date(ExpireDate3, "DMY")

* sometimes the licenses are not formatted in order (time) so logic is necessary to
differentiate the license order
* Note: after doing this I realized its probably easier to fix the formatting than jump
through all these logic statements

gen diff1 = .
gen diff2 = .


summarize diff1, mean
local mean_diff1 = r(mean)

summarize diff2, mean
local mean_diff2 = r(mean)

local totalmean = (`mean_diff1' + `mean_diff2')/2

display "Average difference across all observations: " ///
        "`totalmean'"
// Average difference across all observations: 140.5204900718209

*histogram diff1 if diff1 > 0 || histogram diff2 if diff2 > 0
cd "`data'"

twoway (histogram diff1 if diff1 > 0, width(2)) ///
       (histogram diff2 if diff2 > 0, width(2))
graph save "RenewalTime_GreaterThanZero", replace



twoway (histogram diff1 if diff1 < 0, width(2)) ///
       (histogram diff2 if diff2 < 0, width(2))
graph save "RenewalTime_LessThanZero", replace
```

Appendix J: Code to create total variables for quarterly active licences and occupants.

```
// Creating yearly active indicators and calculating totals
forval year = 2009/2023 {
    // Generating an indicator for active properties in any quarter of the year
    gen active_`year' = (Q1`year' == "1" | Q2`year' == "1" | Q3`year' == "1" | Q4`year' == "1"
 | Q1`year' == "ERROR" | Q2`year' == "ERROR" | Q3`year' == "ERROR" | Q4`year' == "ERROR")
}
```

```
// Counting active licenses
forval property = 1/2210 {
    forval year = 2009/2023 {
        forval qtr = 1/4 {
            if Q`qtr'`year'[`property'] == "1" | Q`qtr'`year'[`property'] == "ERROR" {
                scalar total_Q`qtr'`year' = total_Q`qtr'`year' + 1
            }
        }
    }
}
```

Appendix K: Code producing active_both and lost_licence

```
gen active_both = (Q12015 == "1" | Q12015 == "ERROR") & (Q12022 == "1" | Q12022 ==
"ERROR")
gen lost_license = 1 - active_both
gen Active2015 = (Q12015 == "1" | Q12015 == "ERROR")
```

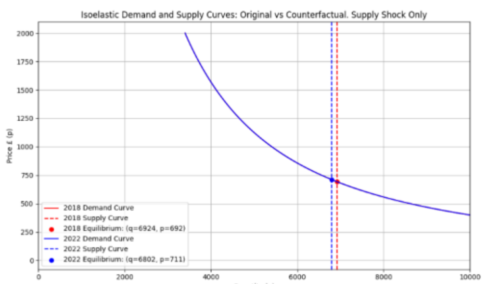
Appendix L: Code merging University managed dataset and Lawson and Thompson dataset

```
// Import & Merge Data Files

// Import and clean the Uni Managed data set, save as .dta
import excel "`data'HMO unimanaged Online .xlsx", firstrow
drop if _n > 23
gen UniManaged = 1
gen Private = 0
drop I J L M N O P K
save UniManaged.dta, replace

// Import and clean the private agency data set, merge with Uni Managed .dta

import excel "C:\Users\stv2\OneDrive - University of St Andrews\Register Team OneDrive\Stata\Data\Private agency (rent) .xlsx", sheet("Sheet1") firstrow clear


/*"C:\Users\kmp21\OneDrive - University of St Andrews\VIP STATA\Stata\Data\Private agency (rent) .xlsx", sheet("Sheet1") firstrow clear*/

gen Private = 1
gen UniManaged = 0
drop O
drop if _n > 53
merge m:m Adresses PostCode Distance EPC Occupants BathroomToilets Rent HMO using "UniManaged.dta"
```

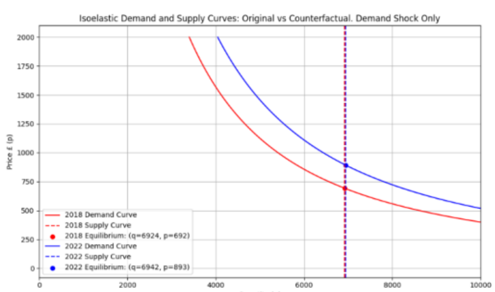
Appendix M: Code for the Hedonic Price Regression

```
/* Creating dummies for the EPC levels */
gen EPC_C = 0
replace EPC_C = 1 if EPC == "C"
gen EPC_D = 0
replace EPC_D = 1 if EPC == "D"
gen EPC_E = 0
replace EPC_E = 1 if EPC == "E"
gen EPC_F = 0
replace EPC_F = 1 if EPC == "F"

 /* Regression of Rent as the dependent Variable and occupants,the number of Bathrooms & Toilets
 and the Dummy EPC (using dummy variables) as independent variables*/

 regress Rent Occupants Distance BathroomToilets EPC_C EPC_D EPC_E EPC_F

 /* Analysis of the correlation */

 corr Occupants Distance BathroomToilets EPC_C EPC_D EPC_E EPC_F

 estat vif
```

Appendix N: **Supply Shock Only Isoelastic Demand**



Appendix O: **Demand Shock Only Isoelastic Demand**



Appendix P: **Isoelastic Demand Greater Absolute Value of Elasticity**