L2: Event-History Models

Duration Analyis

Prof Gwilym Pryce

AQIM Training

June 2006

Introduction

- Suppose we want to construct a model that explains intervals of time:
  - I.e. where the dependent variable is a duration:
    - Eg Duration of unemployment
- Q/ Any other examples?

Examples of Duration Variables:

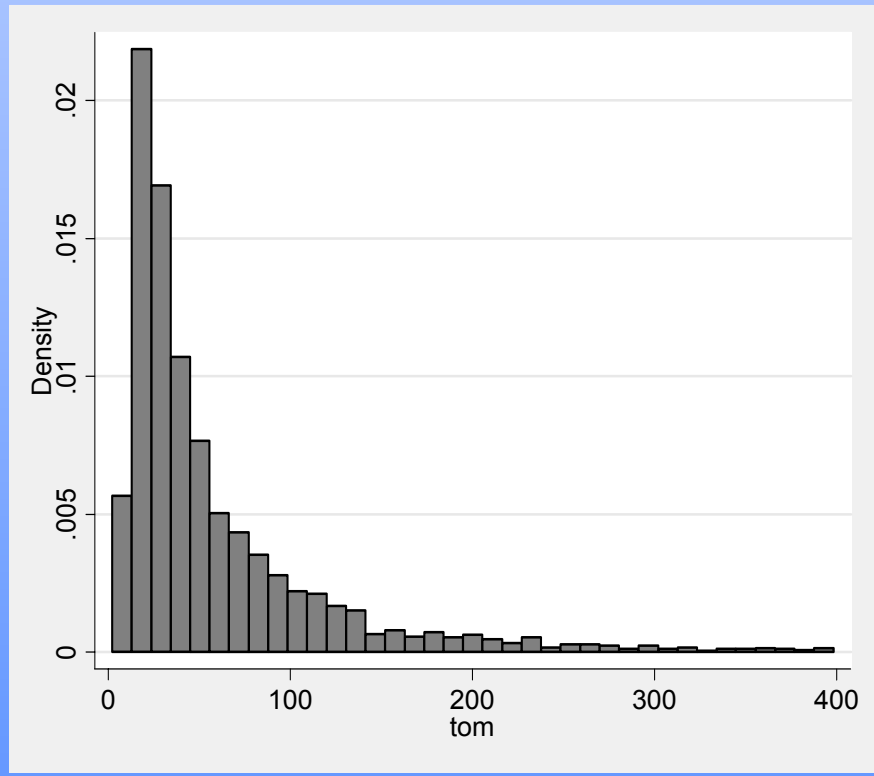
- Time on the market
- Duration of residence
- Length of tenure
- Time-to-default
- Duration in education
- Duration in the rental sector
- Number of years elected
- Days left before you die
- Age before lose virginity
- Age before smoke first cigarette
- Age before first try drugs
- Length of time a company exists before becoming a PLC
- Time to Demutualisation
- Time before contract bird flu
- Length of time in foster care
- Length of stay at a particular address
- Time taken to complete your PhD…

- Strictly speaking we cannot use OLS to model variables like these.
- Q/ Why?

- Can't really use OLS to model durations because:
  - Lower bound of zero
  - Lumpiness of time intervals
  - Non-normal
  - Censoring
  - Duration dependence

E.g. Time-on-the-Market

- Notice:
  - Lower bound of 0
  - Potential lumpiness (days)
  - Very non-normal
  - What about properties that never sell?
    - "Censoring"
  - Duration dependence:
    - Does the longer a property is on the market affect the probability of sale in a given period?



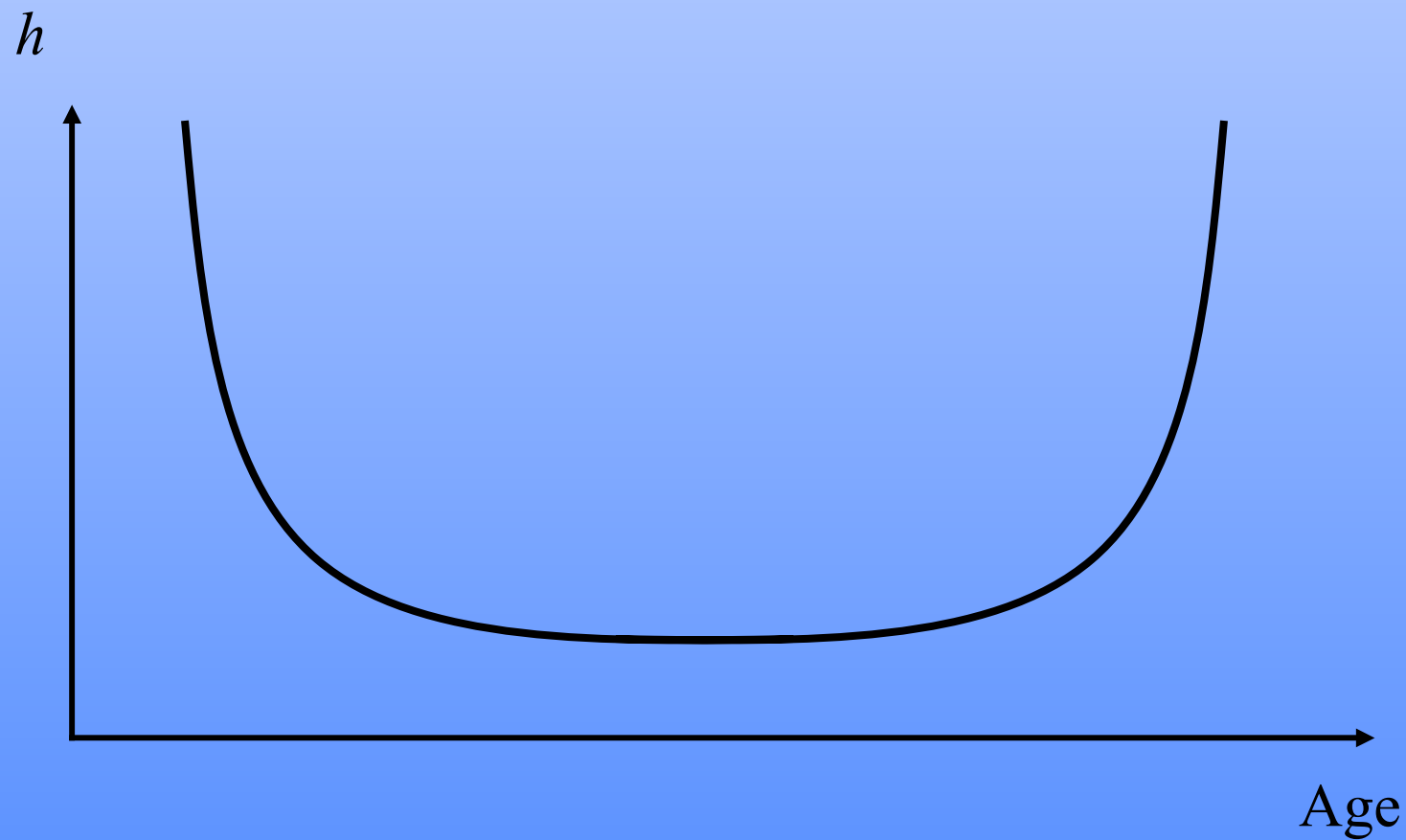
Duration dependence:

- We have just asked:
  - "Does the longer a property is on the market affect the *probability* of sale in a given period?"
- Most of the models used to analyse duration use a related concept to probability:
  - *"Hazard"*
- So we would re-write this question as:
  - "Does the longer a property is on the market affect the *hazard* of sale in a given period?"

What are *Hazards*?

- Hazard function measures the rate at which risk of failure is accumulated
  - "Failure" here is the event of a house selling
- Developed in the bio-stats field
  - "Hazard" is positively related to probabilities, but varies from 0 to infinity.
    - I.e. no upper bound at 1.
- The most famous hazard function is the "bathtub" hazard of human mortality…

Hazard Function for Mortality:

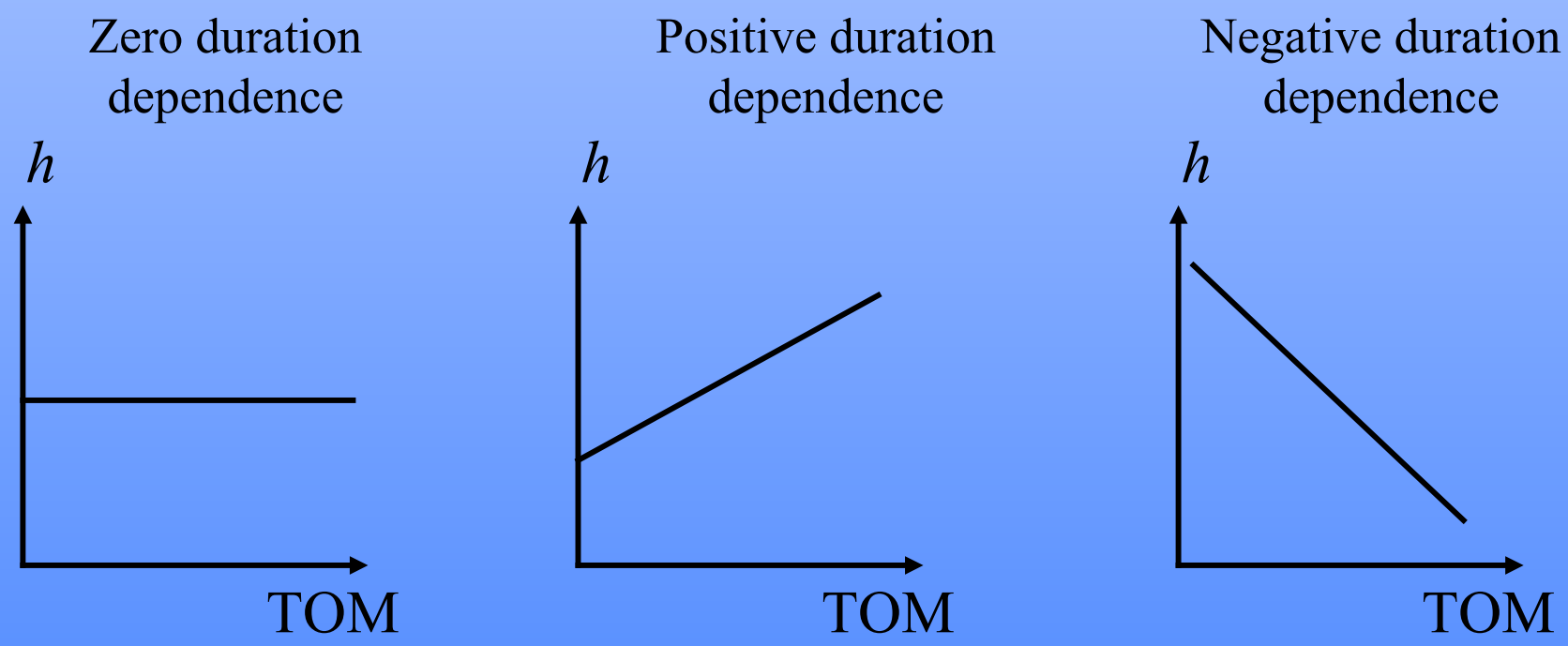


- Very strong duration dependence associated with mortality:
  - How long you have been alive profoundly affects your chances of dying in any particular day.
  - Hazard of death:
    - High for ‘analysis time’ = 0 to 5 years
    - Low for analysis time = 25 to 45 years
    - High for analysis time = 70+ years

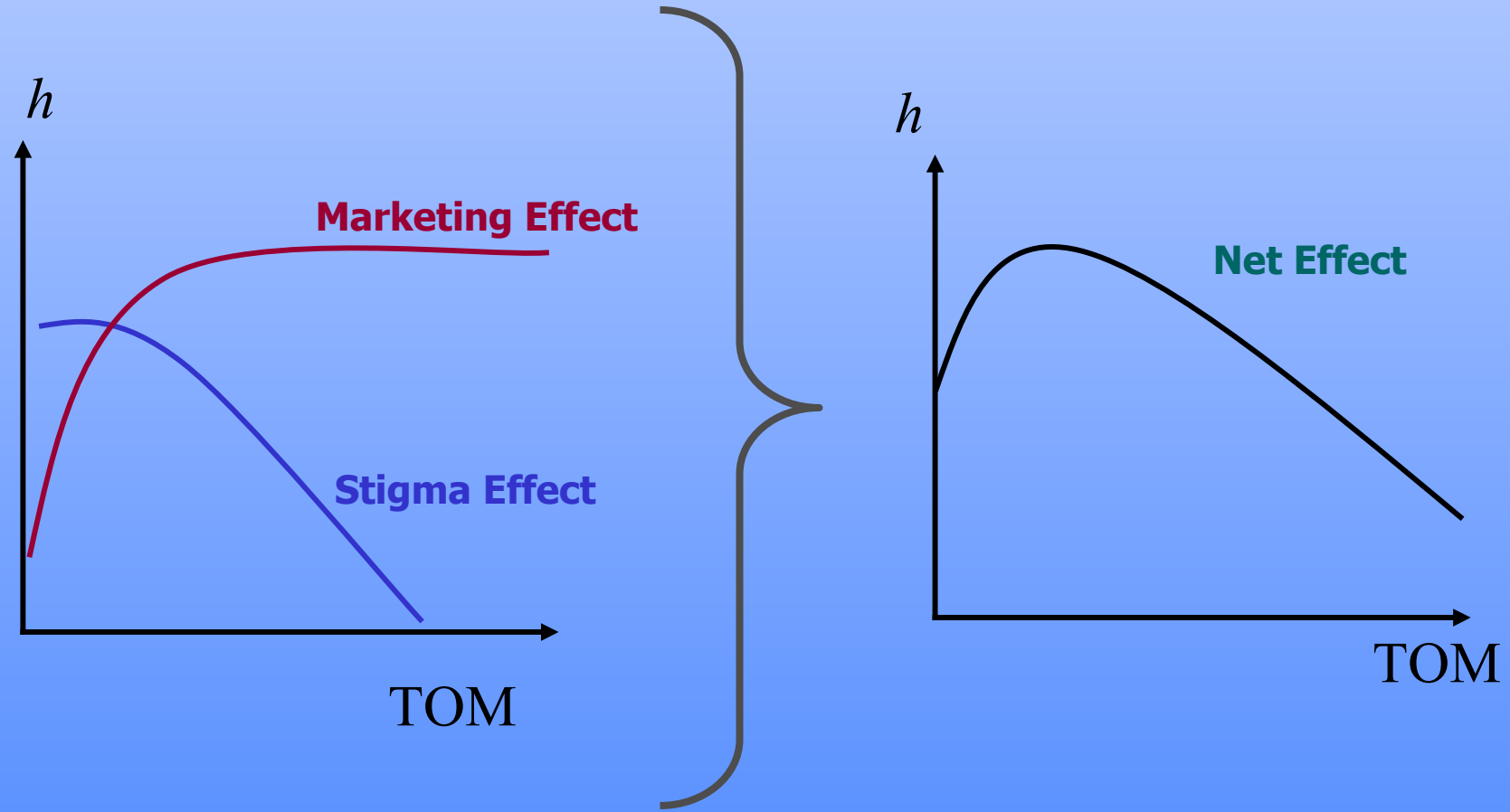
Q/ What will the House Sale hazard function look like?

- Q/ … Is there “Duration Dependence”?

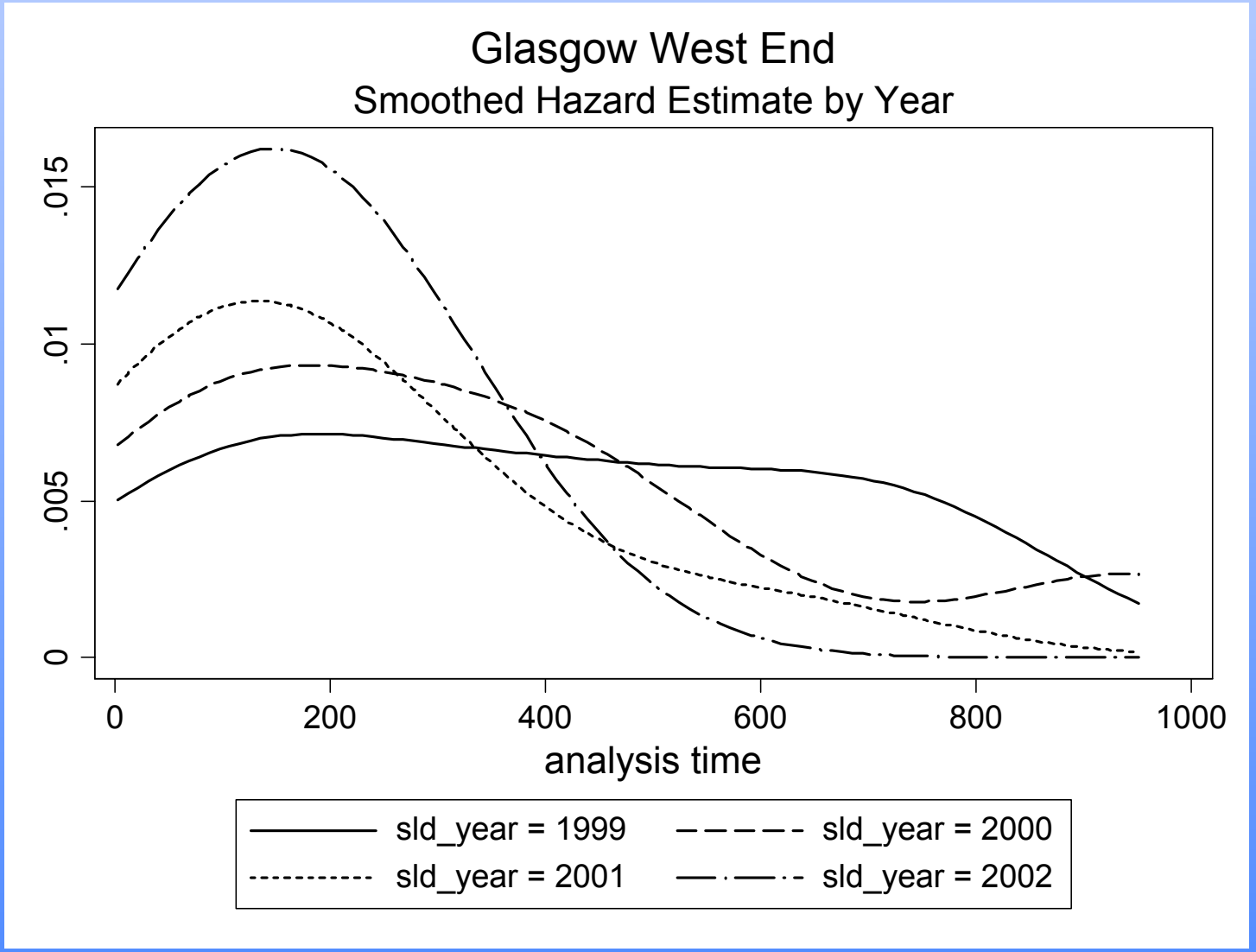
– *If a property has been on the market for a prolonged period, perhaps this will be a signal of poor quality and deter bidders.*



Perhaps there will be an Interplay of Marketing & Stigma Effects:



Glasgow West End by Year



Modelling Duration Variables

- If there is no duration dependence, we can use the simplest model called the Exponential Model
- If there is monotonic duration dependence
    - Either always increasing or always decreasing hazard rate
  - We can use the Weibull method
- If there is non-monotonic duration dependence
    - Rises then falls, or falls then rises
  - We can use the log-logistic regression method
- If there are undulations:
    - Use semi-parametric approach like Cox Proportional Hazards Regression

- Hazard rate is not “proportional” over time
  - The shifts are not parallel
- If the shifts were proportional,
  - We could use “proportional hazards” regression
    - Cox semi parametric.
- If the shifts are not proportional between subsets,
  - Then we may be able to deal with this by stratifying between subsets
- E.g. Procyclical pattern to Duration Dependence

  $\Rightarrow$Need to stratify by year

Alternatives to Regression Analysis:

- As the above example illustrates, regression analysis can get complicated when the dependent variable is a measure of duration
- There is quite a lot you can do without using regression analysis however:
  - Life Tables
  - Hazard curve estimation

Life-Tables:

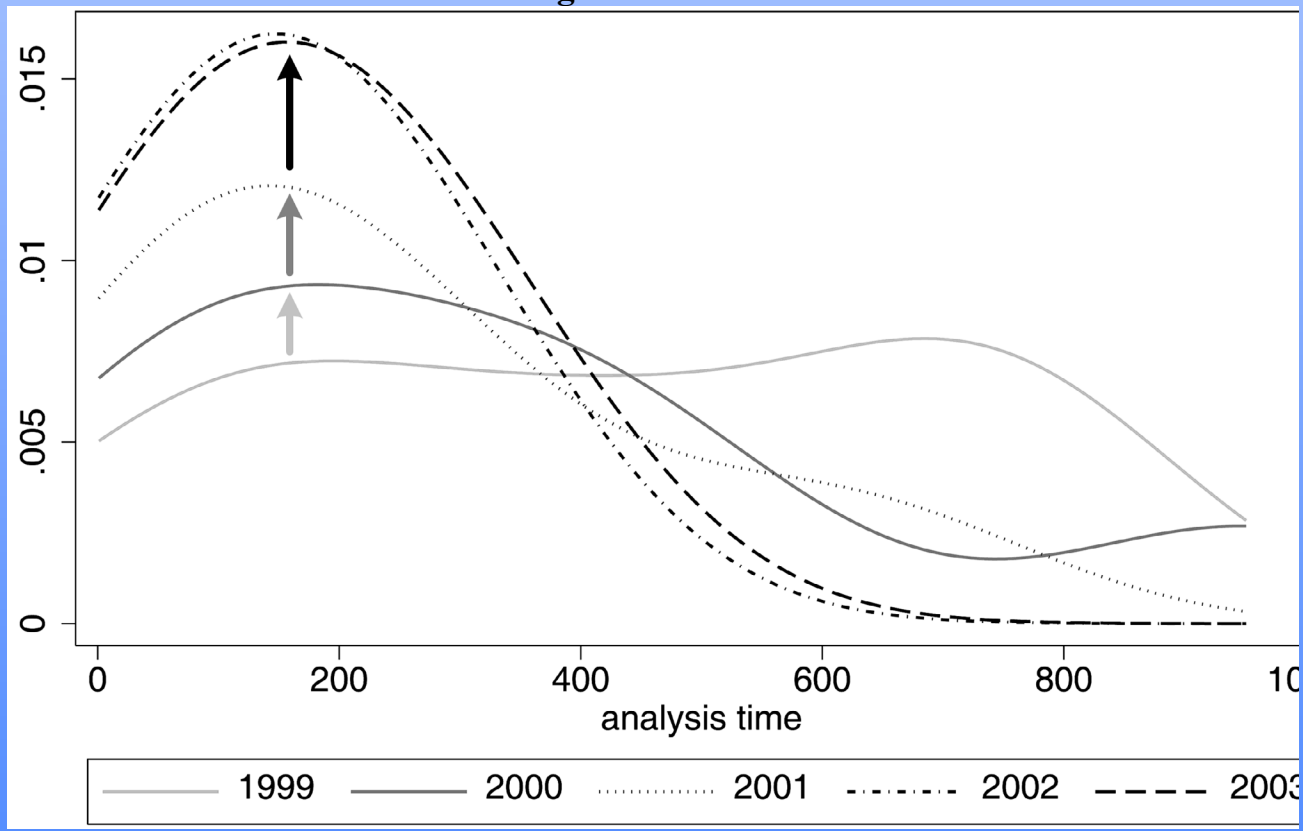
		Strathclyde (n = 27,623)			
TOM Interval (Days)		**Cum. Failure**	**Hazard**	**95% CI**	
0	14	9.83%	0.0074	0.0071	0.0077
14	30	40.88%	0.0260	0.0255	0.0265
30	60	62.85%	0.0152	0.0148	0.0156
60	90	73.47%	0.0111	0.0107	0.0115
90	120	79.68%	0.0088	0.0084	0.0092
120	150	83.77%	0.0075	0.007	0.0079
150	180	86.83%	0.0069	0.0065	0.0074
180	210	88.98%	0.0059	0.0055	0.0064
210	240	90.95%	0.0065	0.006	0.0071
240	270	92.43%	0.0059	0.0054	0.0065
270	300	93.53%	0.0052	0.0046	0.0058
300	330	94.61%	0.0061	0.0054	0.0067
330	365	95.61%	0.0058	0.0051	0.0065
365	730	99.34%	0.0041	0.0039	0.0042
730	1095	99.79%	0.0029	0.0024	0.0033

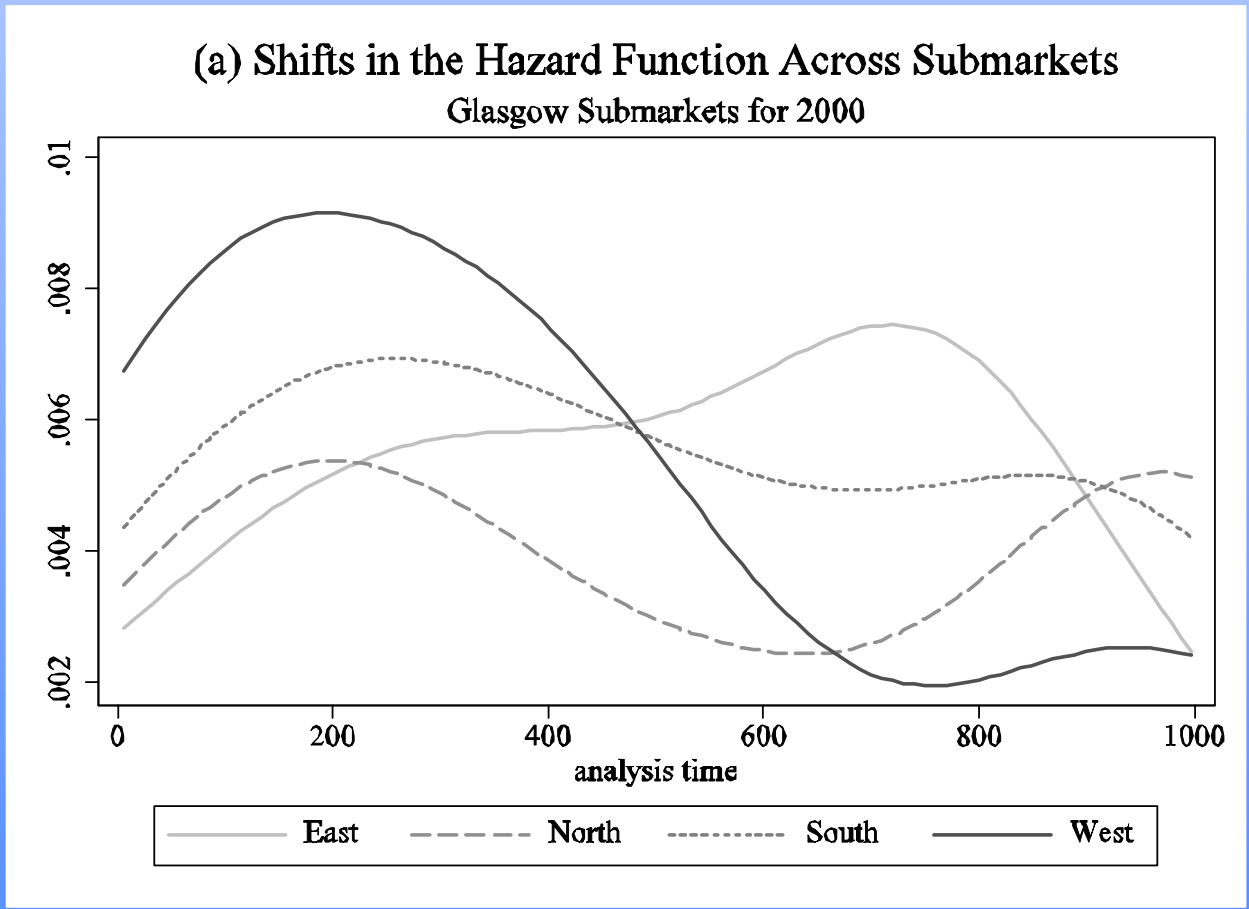
Pryce & Gibb (2006) see Housing Resources page of www.gpryce.co.uk

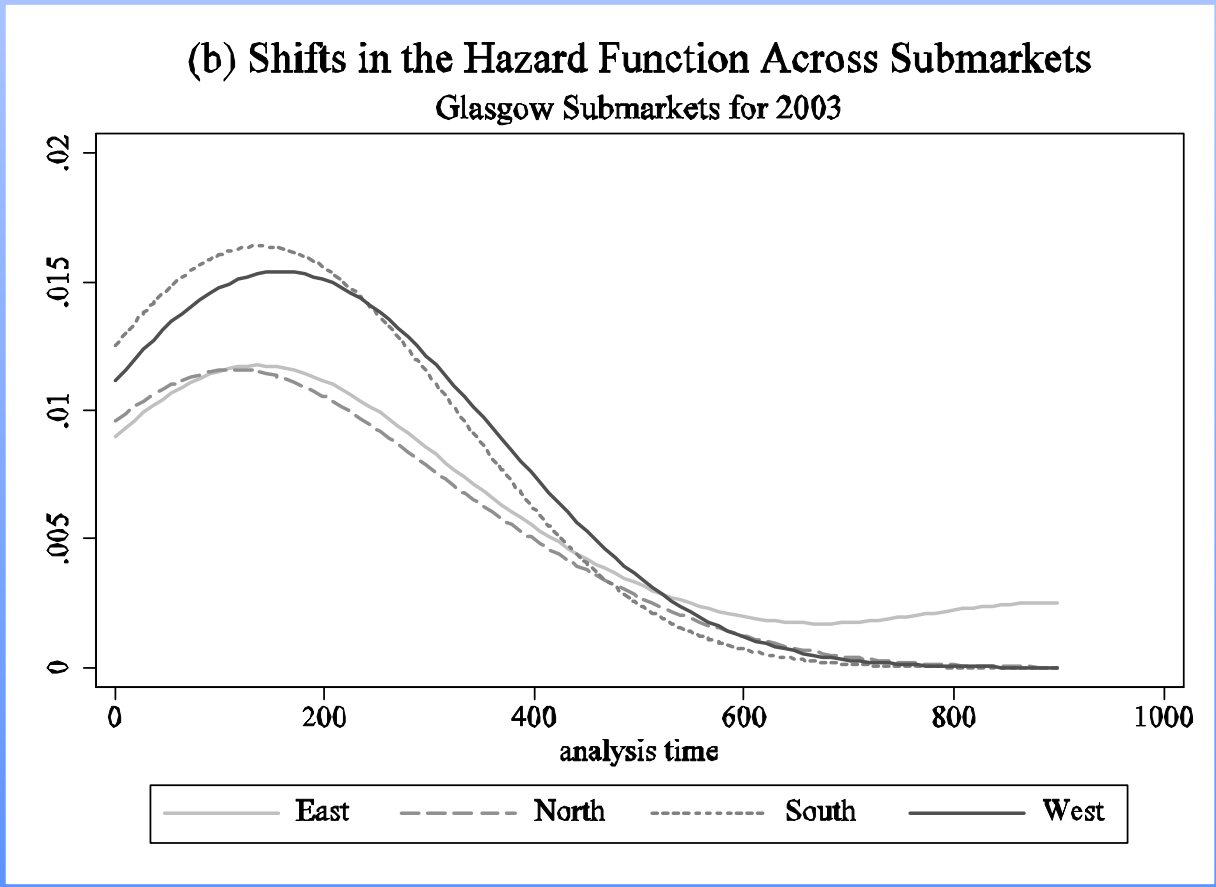
Hazard curve estimation:

Figure 2: Metamorphosis of the Hazard Function During a Boom

West End of Glasgow for Each Year 1999-2003




(a) Shifts in the Hazard Function Across Submarkets
Glasgow Submarkets for 2000
.002
.004
.006
.008
.01
0
200
400
600
800
1000
analysis time
East
North
South
West


(b) Shifts in the Hazard Function Across Submarkets
Glasgow Submarkets for 2003
.02
.015
.01
.005
0
0
200
400
600
800
1000
analysis time
East
North
South
West

Software Packages:

- I would strongly recommend that you use Stata rather than SPSS:
  - Excellent graphics facilities
    - Kernel-density hazard estimation (smooth curves)
  - Easier to interpret output
  - Excellent documentation
  - Programs written by one of the leading experts
    - (Prof Cox, Durham)
  - Excellent diagnostic facilities
  - Wide range of estimation types

Reading:

- Kennedy, P., ch.16, particularly 16.4.
- Cleves, M. A., Gould, W.W. and R.G. Gutierrez (2002) "*An introduction to Survival Analysis Using Stata*", Stata Press, Texas.
- Stata Manual on "*Survival Analysis and Epidemiological Tables*"
- Box-Steffensmeir & Jones "Event History Modelling"
  - Geared towards social scientists.
- Example Journal Paper:
  - Pryce, G. and Gibb, K. (2006) 'Submarket Dynamics of Time-to-Sale', *Real Estate Economics*, forthcoming.