Are Swedish housing markets duration dependent?

Bachelor thesis in Economics, NEG300 15HEC. Fall semester 2017
Institutionen för nationalekonomi med statistik
School of Business, Economics and Law, University of Gothenburg

Sweden's rapidly growing housing sector, and the long-last boom it's experienced even in a global financial crisis, has been the topic for several recent studies. In this thesis, we examine house price data from 279 Swedish municipalities during the years 1981-2016 to test for duration dependence. We include municipality-specific growth in real income and population, and a national level real mortgage rate. To test for duration dependence, we use a linear probability model as our benchmark regression, and expand on this using a logit specification. The estimations on the contraction phases are inconsistent, and we can therefore not conclude that contraction phases exhibit duration dependence. The estimations on expansion phases indicate that duration dependence is present.

Jennie Östlund
Jesper Östlund
Thesis supervisor: Inge van den Bijgaart

Contents

1 INTRODUCTION

The housing market played a major part in the build-up, unfolding and aftermath of the financial crisis of 2007. National house prices in OECD economies saw a huge rise in the years leading up to the Great Recession and, for most countries, a subsequent fall of large proportions. The Swedish housing market, along with those of Belgium and Norway, are exceptions. They stand out among the OECD countries as having experienced only a modest and short-lived downturn in comparison (Sørensen 2013). Prices have since trended upwards, making new highs.

The synchronization of financial cycles, within and between countries, has increased significantly since 2007-08 (Claessens et al. 2011); the credit and housing markets stand out as being highly synchronized at the national level. In a time of historically cheap credit and high level of private debt in Sweden, the housing market has been given extra attention in recent years, both in the academic literature as well as in the media. Turk (2015), Giordani et al. (2015) and Flam (2016) are just a few trying to explain the current situation in Sweden's rapidly growing housing sector. In particular, household's debt-to-income ratio in has been a cause for concern (IMF, 2012).

The importance of the housing sector to the Swedish economy, therefore, cannot be overstated. We hope to expand on this debate by looking specifically at up- and downturns in local house prices, as opposed to the more conventional econometric approaches involving, for example, determinants of house prices or house price dynamics. By closely looking at the characteristics of these expansion and contraction phases, we aim to explain the relationship between the length of a phase and the probability of it ending.

Duration analysis examines the relationship of time to an event, conditional on how long a subject has been observed without the event taking place. A presence of duration dependence in Swedish house prices could suggest, for example, that the longer an upturn has been going on, the more likely it is to end – or vice versa. The rationale for choosing duration analysis is, simply put, to examine the relationship between time and housing price cycles.

In this thesis, we study house sales price data from 279 municipalities during the years 1981-2016. From this data we identify the house market cycles per municipality, and by segmenting these into expansion and contraction phases, we aim to test these for duration dependence. In other words: can we, by looking at the duration of a current phase, determine whether this phase is likely to end?

In section 2 we will give a brief overview of the literature on duration dependence; in section 3 we present our data; in section 4 we form our empirical model and describe our method; in section 5 we present and discuss our estimates, and in section 6 any conclusions and implications for further research.

2 RELATED LITERATURE: DURATION DEPENDENCE

Duration analysis[1] is an econometric method that originates from survival analysis, a branch of statistics where the survival time of a given subject is the variable of interest; estimating the survival time of patients that receive a particular drug, for example (Wooldridge, 2010, pp. 983). Lancaster (1979) was one of the first to apply survival analysis to the econometric literature. He used a duration model in trying to describe the duration distribution of unemployment, using data on unemployed men returning to work. Historically, the length of unemployment, wars or marriages are typical durations that have been used in econometric applications of survival analysis (Ohn et al., 2004).

The econometric literature on duration dependence has since grown into a vast body. The applicability of duration analysis spans all the economic disciplines but is, perhaps, most often encountered in the analysis of business cycles, such as equity and/or credit. This is, partly, a result of the increased incentives for doing such analysis; being able to better predict ends of business booms (or busts) could have potentially great value to policy makers. One of the most cited papers dealing with financial cycles is Claessens et al. (2011). Analyzing credit, house price, and equity price data in 21 advanced economies, they cover over 40 years of financial cycles. They try explaining the main features of financial cycles. In particular, how up- and downturns differ, in both duration and amplitude, comparing between equity, credit and housing. They also look at synchronization of different cycles, both at the national and international level. At the core of their research is the question when and how chocks in one market severely influences the other(s). Their findings suggest that synchronization of financial cycles is large both within and across countries, and has been growing over time. Furthermore, they argue that globally synchronized busts tend to exacerbate subsequent recessions.

Cunningham and Kolet (2011) are, to the best of our knowledge, the first to apply duration analysis on housing market data. They use survival models to test for duration dependence in American and Canadian house price indices. They allow for changes in macroeconomic fundamentals by including a range of covariates[2]. Their findings suggest that there is a positive duration dependence in US housing upturns, whereas contraction phases do not exhibit any

[1] Sometimes, depending on the context, referred to as *event history analysis, reliability or failure time analysis, transition analysis, hazard rate analysis* or *survival analysis* (Gujarati, 2015)
[2] US/Canadian Nominal 30/5-Year Conventional Mortgage Rate, Nominal Policy Interest Rate, Nominal Personal Income, Population, Rent (for US only).

duration dependence. In other words, the probability of exiting an expansion phase depend positively on the time spent in that upturn, whilst the probability of exiting a downward phase is independent of the time spent in that phase.

Bracke (2011) use a similar approach when testing forty years of house price data for 19 OECD economies in his country-level analysis. He finds that expansion phases are longer than contraction phases and that upturns show positive duration dependence. This is consistent with the findings of Cunningham and Kolet (2011).

Bénétrix et al. (2012) focus specifically on housing slumps, i.e. the end of contraction phases. Using 40 years of house price data for 19 OECD countries, they find that both upturns and downturns exhibit duration dependence. In their analysis, they show that downturns are more severe if preceded by a longer-than-average boom in house prices.

Our thesis is similar to Cunningham and Kolet (2011) in that it concerns yearly averages for house prices at the local level, while Bracke (2011) and Bénétrix et al. (2012) uses quarterly averages of national data.

3 DATA DESCRIPTION AND SUMMARY STATISTICS

Our data consists of annual average sales price for houses[3] per Swedish municipality for the time period 1981-2016. Of the 290 current municipalities in Sweden, 11 were formed during the sample period and are therefore not included in the sample (see Appendix A). The prices are reported at nominal level, and we have therefore adjusted them for inflation using national-level CPI (CPI 1980=100). Prices and CPI are from Statistics Sweden (SCB).

In order to analyse the duration of up- and downturns, respectively, we need to identify cycles as well as turning points in the local house price series.

The literature on the identification of *business* cycles is extensive and procedures for determining them are well-defined. Burns and Mitchell (1946) pioneered what can be seen of as a graphical approach to analyzing ups and downs in the market. *The classical cycle*, later became the "standard" approach to this strand of econometric analysis, which focuses on pattern recognition. The NBER, for example, make use of this procedure when determining peaks and troughs in U.S. business data. These dates, and additional descriptive statistics, are publicly available, even though their exact method is not published. In regard to the housing market, no such consensus on how to identify cycles is available so an appropriate method must be carefully chosen. [4]

The method we have chosen originates from the work of Bry and Boschan (1971), who modified the *classical cycle* explained above, to better define housing cycles. Their method is known as the BB-algorithm. Harding and Pagan (2002) later modified this method to better fit quarterly data. The modification is known as the BBQ-algorithm and has been widely used in international literature.[5]

---

[3] Definition of "houses", in this context, is: "*One- or two-dwelling buildings means detached one- and two-dwelling buildings as well as semi-detached, row and linked buildings (excluding buildings for seasonal and secondary use*/*Småhus avser friliggande en- och tvåbostadshus samt par-, rad- och kedjehus (exklusive fritidshus*))(SCB).

[4] There are a few methods available when trying to filter out cycles in economic data. For yearly aggregate data, Neftci (1984) and Cashin and McDermott (2002) use the calculus rule, $S_t = 1(\Delta y_t > 0)$. In IMF World Economic Outlook [year] *Growth Resuming, Dangers Remain: Ch. 3 Dealing With Household Debt*, a similar method is used when analyzing yearly data. This method, however, sometimes renders turning points that lie within a short period of time, e.g. one year. Since we want to impose restrictions, such as minimum cycle length of two years, the BBQ-algorithm is, in our view, a better alternative.

[5] Artis, Marcellino and Proietti (2004) and Claessens et al. (2011), for example.

The algorithm identifies, using a "rolling window" of j years, a point (at time t) as a local peak if:

$$(y_{t-j}, \dots, y_{t-1}) < \ y_t > (y_{t-j}, \dots, y_{t-1})$$

Conversely, a local trough is identified at time t if:

$$\left(y_{t-j}, \dots, y_{t-1}\right) > \ y_t < (y_{t-j}, \dots, y_{t-1})$$

The algorithm alternates local maxima with local minima such that a trough must be followed by a peak, and vice versa (Bracke, 2011).

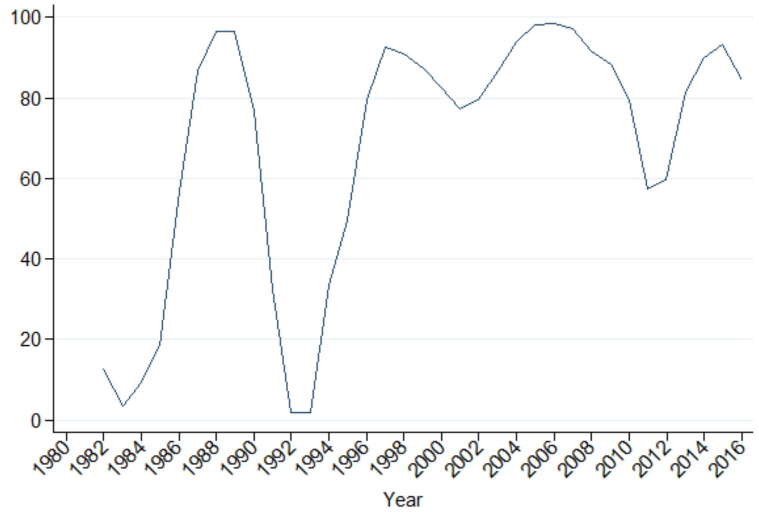
We use a rolling window of one year (j=1). Additionally, following the results of Cunningham et al. (2011) we impose a minimum phase requirement of two years. That is, a peak or trough can be no shorter than two years.

Each municipality experience both up- and downturns during the sample period, so there will be multiple observations per municipality in our data.

3.1 DESCRIPTIVE STATISTICS, PHASES

In figure I, the number of municipalities in expansion phase each year is shown. The mostly high (>80%) and low percentages shows that price cycles are synchronised across Sweden.

Figure I: Percentage of municipalities in expansion phase, 1982-2016



The results are similar to those presented by previous studies. Despite turning points not being identical, the similarity suggests that our settings for the BBQ-algorithm are viable (Table I).

Table I: Turning points, previous studies

	Bracke (2011)	**Bénétrix et al. (2012)**
TROUGH	1985Q4	1986Q3
PEAK	1990Q1	
TROUGH	1996Q1	1996Q1
TROUGH		2009Q2

Table II shows summary statistics of the phases, grouped per metropolitan area. We have used Statistics Sweden’s definition of metropolitan area (see Appendix A).

Table II: Descriptive statistics for phase duration (mean, standard deviation, min, max)

Expansions

Area	N	Mean	Std. Dev.	Min	Max
Göteborg	15	8.33	4.84	5	17
Malmö	21	11.33	5.15	4	17
Stockholm	28	9.14	5.94	4	22
_NA	483	7.02	4.63	2	22
All	**547**	**7.34**	**4.81**	**2**	**22**

Contractions

Area	N	Mean	Std. Dev.	Min	Max
Göteborg	16	3.31	1.30	2	5
Malmö	21	3.00	1.00	2	5
Stockholm	28	3.21	1.34	2	7
_NA	474	3.94	1.90	2	15
All	**539**	**3.85**	**1.85**	**2**	**15**

Note: only full phases, i.e. when both entry and exit in a phase is observed, is included in this table.

The contraction phases are on average 3.5 years shorter than the expansion phases (3.85 vs 7.34 years). The mean durations also tell us that the three main metropolitan areas have shorter contraction phases and longer expansion phases, on average.

Bracke (2011) reports two downturns for Sweden, the first with a duration of 21 quarters, and the second of 17 quarters, equalling 4.75 years of duration on average. For expansions, he reports three phases with a duration of 25, 24 and 56 quarters, corresponding to 8.75 years of duration on average.

Bénétrix et al. (2012) identifies similar phases; four contractions of 6, 27, 23 and 6 quarters of a year, corresponding to an average duration of 3.88 years; three expansions of 31, 13, and 45 quarters of a year, corresponding to an average duration of 7.42 years.

As a side note, Cunningham & Kolet (2011) report 5.8 years of mean duration for expansion phases, and 4.4 years for contraction phases, for US cities. These characteristics are similar to those that we have found in the Swedish housing markets.

Table III: Summary of previous studies, mean duration of phases				
Authour(s)	**Sample period**	**Data type**	**Expansions**	**Contraction**
Bracke (2011)	1970-2010	National-level, quarterly	8.75	4.75
Bénétrix et al. (2012)	1970-2010	National-level, quarterly	7.42	3.88
Our findings	*1981-2016*	*Local level, annual*	*7.34*	*3.85*
Cunningham & Kolet (2011) **Note:** US	1975-2005	National, annual	5.80	4.40

Our results are well in line with above mentioned studies. Bracke (2011) reports slightly longer phases, probably due to the facts that (1) he uses quarterly data and (2) different settings for the BBQ-algorithm. He sets the rolling window to 6 quarters and the minimum phase length to 6 quarters, whereas we use one year as rolling window and two years for the minimum phase length.

3.2 COVARIATES

We have included population growth and growth in average income as covariates as we expect them to affect house price cycles. We expect that an increase in population and/or income would increase demand of houses and subsequently prices. We have also included mortgage rate, as we expect a lower rate would decrease the cost of buying a house and therefore increase demand.

The covariates consist of three municipality-specific variables; growth in real house prices, population growth and growth in average income. The growth in real house prices is calculated from the house prices also used to define phases. The population growth rate is calculated from yearly averages per municipality.[6] The income growth rate is calculated from the average income measured on a yearly basis per capita for all over 16 years. The income is adjusted for inflation using national level CPI (1980=100). The municipal specific covariates are provided by Statistics Sweden (SCB).

We have also included the real annual average mortgage rate after tax. The rate is on national level and it is provided by the National Board of Housing, Building and Planning. Table IV shows summary statistics for covariates.

Table IV: Summary Statistics for Covariates

Variable		Sample period	N	Mean	Std. Dev.	Min	Max
Growth in real house prices, %[a)]	Gpriceadjprice	1982-2016	9765	2.46	10.20	-43.24	96.45
Population growth %	gpopadjpop	1982-2016	9765	0.11	1.00	-4.17	8.21
Growth in average income, % [a)]	gincadjinc	1992-2015	6696	1.81	1.53	-4.12	9.47
Mortgage rate, after tax [b)]	realmortagerate	1991-2012	6138	3.00	1.86	-0.56	6.72

a) Nominal values, deflated by CPI (1980=100)

b) National level.

During the sample period, house prices have increased, on average, by 2.4 percentages per year. The covariates are more informative when measured by expansion and contraction phase, respectively (Table V).

[6] Municipalities that were divided during the sample period will see a large decrease in population size the year of division. We have therefore adjusted the population size with the decrease to clear the number to only include organic growth (see appendix).

Table V: Summary Statistics for Covariates, per phase

Expansions

Variable		**Sample period**	**N**	**Mean**	**Std. Dev.**	**Min**	**Max**
Growth in real house prices, %[a)]	gpriceadjprice	1982-2016	4008	6,79	8,64	-24,36	79,88
Population growth %	gpopadjpop	1982-2016	4008	0,06	1,04	-4,17	8,21
Growth in average income, % [a)]	gincadjinc	1992-2015	2688	2,08	1,46	-2,93	9,47
Mortage rate, after tax [b)]	realmortagerate	1991-2012	2673	2,96	1,65	-0,56	6,72

Contractions

Variable		**Sample period**	**N**	**Mean**	**Std. Dev.**	**Min**	**Max**
Growth in real house prices, %[a)]	gpriceadjprice	1982-2016	2074	-5,93	8,70	-43,24	51,04
Population growth %	gpopadjpop	1982-2016	2074	-0,09	1,01	-3,18	5,75
Growth in average income, % [a)]	gincadjinc	1992-2015	1668	0,95	1,61	-4,12	6,69
Mortgage rate, after tax [b)]	realmortagerate	1991-2012	1778	3,39	2,36	-0,56	6,72

a) Nominal values, deflated by CPI (1980=100)

b) National level.

Note: Only full phases, i.e. when both entry and exit in a phase is observed, is included in this table.
As phases are calculated with the restriction of at least a 2-year consecutive price decrease/increase, expansion(contraction) phases may include observations where the yearly growth in house prices is negative(positive).

The average yearly growth in house prices are 6.79 percentages in expansion phases and -5.93 percentages in contraction phases. The average population growth is positive in expansion phases, and negative in contraction phases, which is in line with general economic theory: the higher the demand, the higher price. However, the differences are relatively small, and we should be careful to draw any conclusions.

The mortgage rate is on average higher in contraction phases, 3.39 percentages as compared to 2.96 percentages in an expansion phase.

The average income growth is approximately 1 percentage point less in a contraction phase compared to expansion phase, 0.95 versus 2.05 percentages.

We have also included dummy variables, D1 and D2, for different time periods to control for unobserved time varying effects, for example policy changes.

Group	Years	Characteristics
Reference group	1981-1991	Fixed exchange rate
D1	1992-2007	Floating exchange rate was introduced 1992
D2	2008-2016	Real estate tax abolished; global financial crisis

The first cut-off point marks a historical change in the Swedish currency regime; Riksbanken abolished the fixed exchange rate and the let Swedish Krona float. This had significant impact on the economy as well as the housing market; more importantly, it changed forever the mechanisms through which mortgage rates were set (Turk 2015). The second cut-off point is 2008 when the start of the global financial crisis was well under way. The real estate tax was also abolished in 2008.

4 EMPIRICAL MODEL

To test for duration dependence, we will use a regression model suggested by Cunningham and Kolet (2011). Expansion and contraction phases will be tested separately as we predict they are behaving differently. Contraction phases are on average 3.5 years shorter than the expansion phases, which may mean that they have a stronger relation to duration and/or other covariates.

Our dependent variable is y_{it} a binary variable that takes value 1 if the municipality is in expansion/contraction phase and 0 if the state exits the phase in that time period. Exit is defined as the last year of a phase, i.e. the year during which the peak/trough of the phase is. The dependent variable can be interpreted as the likelihood of staying in the phase.

We estimate:

$$\Pr(1 - hazard) = \Pr(Y_{it} = 1)$$

More specifically:

$$Y_{it} = \alpha_i + \beta_1 duration_{t-1} + \beta X + \mu_{it} \quad (1)$$

Where X is a vector of covariates, and μ_i is the error term. $duration_{t-1}$ is the duration of the current phase at time t-1.

In our first specification, we estimate using a linear probability model (LPM). The linear probability model has its limitation, even if it can give a good indication of relationships. It assumes a linear relationship between the probability of the dependent variable and the covariates, meaning that the partial effect for any variable will be constant. Furthermore, since we estimate using OLS we have no restriction on the predicted probability, which theoretically could be over 1 or under 0. A logistic regression model is therefore preferable when estimating probabilities (Gujarati, 2015, pp. 172).

In the logit model, the probabilities are assumed to follow the logistic probability distribution, and is written:

$$Pr(y_{it} = 1) = P_i = \frac{1}{1 + e^{-z_i}}$$

Where $Z_i = \beta X_i + \mu_i$

$\mathbf{X}_i$ is a vector of covariates and μ_i is the error term.

$$Pr(y_{it} = 0) = 1 - P_i = \frac{1}{1 + e^{z_i}}$$

With this probability function, the probabilities range between 0 and 1, for any values of the $\mathbf{X}_i$ (between -∞ to ∞). The estimation of the model is executed by transforming the probabilities to an odds ratio (OR), i.e. by simply dividing the probabilities:

$$\frac{P_i}{1 - P_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i}$$

And thereafter taking the natural log of the odds ratio:

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = \ln(e^{Z_i}) = Z_i = \beta\chi + \mu_i$$

Hence, in the logit model, we are estimating the log of the odds ratio. It allows for the parameters to be linear, even if the relationship between the explanatory variables and the probability is not. (Gujrati, 2015, Chapter 8.3)

The estimated coefficient in a logit model gives the predicted change in the log likelihood for a one-unit change in the explanatory variable. We are also reporting a pseudo-R^2, comparable with the R^2 in a linear regression model, and the goodness of fit of the model. The goodness of fit measures how good the model is of predicting correctly by comparing predicted values with the actual observations. Any value over 50% is usually considered good (higher than pure chance) (Wooldridge, 2010, pp. 573). In our sample we have a majority of observations where the current phase continues, i.e. yt=1. We are therefor also interested in how well the model predicts the least likely outcome, being yt=0, and are reporting the sensitivity and specificity. Sensitivity is the proportion of positive observations (P(Y=1)) predicted as such by the model. Conversely, the specificity is the proportion of negative observations (P(Y=0)) predicted as such by the model.

We have excluded all observations that are part of incomplete phases, i.e. when the entry or exit of a phase is outside of the sample period, as suggested by Ohn et al.(2004). In survival analysis, these are called censored observations and there are various techniques to account for it. Given our relatively large sample size, we do not expect the excluding of the censored observations to pose a major problem in the estimation.

Since we have several observations per municipality over time, and covariates that do not vary on municipality level (mortgage rate), we can expect the observations to be correlated within the clusters, i.e. municipality. We risk overstating the precision of the estimates and to correct for this we use clustered standard errors (Wooldridge, 2002, pp. 455).

Since we are using logit, a nonlinear functional form, the marginal effects of the covariates are not given by their individual coefficients. Rather, a function of the coefficients renders their marginal effects on the probability of an event occurring7.

[7] The change in probability is rendered by the partial derivative of the expression for Pr(Y=1). Generally, for the logit function, this expression is $\beta_i e^{X\beta}(1 + e^{X\beta})^{-2}$, which varies as the values of the covariates vary (Kennedy, 2008).

5 ESTIMATION RESULTS

The results from our estimations using LPM are presented in table VI. The first row contains estimations for variable lag_duration, which is the duration of the current phase up until time=t. The second row contains the estimations for the lagged duration squared.

Table VI: Test for duration dependence, real house price cycles, Sweden 1981-2016

	(1) Expansion	(2) Expansion	(3) Contraction	(4) Contraction
Lag duration	-0.00796***	0.0120**	0.0181***	0.00478
	(-10.10)	(3.06)	(11.11)	(0.70)
Lag duration, squared		-0.00131***		0.000881*
		(-5.03)		(2.37)
_cons	0.906***	0.858***	0.674***	0.701***
	(137.65)	(86.12)	(61.18)	(34.31)
N	4008	4008	2074	2074
R^2	0.01	0.01	0.02	0.02

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
All models are estimated using LPM (Stata command *regress*). Standard errors are clustered by municipality.

In model 1 the estimations show a positive duration dependence. An increased duration is expected to decrease the likelihood of staying in an expansion phase. However, the coefficient is small; a one-year increase in duration is expected to decrease the relative probability of remaining in the phase with 0.008.

In model 2, we added a squared term for the lagged duration, to see if the duration dependence shows non-linearity. Bénétrix et al. (2012) founds a non-linear relationship between the probability of exiting a phase and its duration. The coefficient of the lagged duration changed to opposite sign, and the squared term is negative. The estimation suggests a positive relationship in the beginning of a phase, i.e. an increase in duration increases the likelihood of remaining in that phase, but as the phase continues this relationship will eventually reverse[8].

[8] This happens when: year x 0.01200 = $year^2$ x -0.00131 <=> year = 9

For contraction phases, model 3 and 4, the estimations suggest a negative duration dependence. The likelihood of staying in a contraction phase, is expected to increase when the duration increases, i.e, the longer the contraction phase has been going on, the higher the likelihood of the phase continuing. When we add the squared term, the coefficient of the lagged duration becomes insignificant.

Since the LPM does not restrict the dependent variable (probability of exiting a phase) between zero and unity, we move forward using a logit specification. The results are reported in table V for expansion phases and Table VI for contraction phases. We report the coefficients, the Wald Chi^2, the pseudo R^2, the log likelihood, and the goodness of fit of the model.

Table VII: Test for duration dependence, real house price cycles, Sweden 1981-2016. Expansions						
		(5)	**(6)**	**(7)**	**(8)**	**(9)**
Duration t-1		-0.0618***	-0.176***	.-0.135***	-0.0678***	-0.0506***
		(-9.81)	(-8.17)	(-8.43)	(-3.36)	(-4.07)
Population Growth			0.555***		0.419***	
			(7.24)		(5.52)	
Growth in real income			0.243***		0.0383	
			(3.96)		(0.58)	
Growth in real house prices			-0.0726***		-0.0865***	
			(-9.71)		(-10.02)	
Real mortage rate after tax			-0.123*		-0.176**	
			(-2.10)		(-3.23)	
Decade 1992-2007					2.170***	1.717***
dummy variable					(9.53)	(14.11)
Decade 2008-2016					0	-0.0519
dummy variable					(.)	(-0.42)
_cons		2.201***	4.288***	3.305***	2.639***	1.513***
		(35.42)	(13.12)	(18.08)	(9.36)	(28.98)
N		4008	2583	2583	2583	4008
Pseudo R2		0.010	0.151	0.0491	0.217	0.101
Wald chi2		96.19	212.25	71.08	259,94	288,68
LL		-1580,71	-668,72	-748,63	-616,41	-1436,66
Goodness of fit		86.35%	90.75%	90.90%	91.06%	86.35%
	Sensitivity	*100%*	*99.19%*	*100%*	*98.68%*	*100%*
	Specificity	*0%*	*6.38%*	*0%*	*14.89%*	*0%*

z statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
Note: All models are estimated using a logit model (command *logit* in Stata). The standard errors are clustered by municipality. Goodness of fit is classified using *estat* classification. For more test regarding goodness of fit, se Appendix C

Table VIII: Test for duration dependence, real house price cycles, Sweden 1981-2016. Contractions

		(10)	(11)	(12)	(13)	(14)
Duration t-1		0.126***	0.171***	0.121***	0.166***	0.143***
		(8.66)	(11.46)	(8.46)	(11.02)	(10.80)
Population Growth			-0.228***		-0.259***	
			(-3.94)		(-4.20)	
Growth in real income			-0.366***		-0.391***	
			(-8.05)		(-8.23)	
Growth in real house prices			0.0743***		0.0729***	
			(11.02)		(10.89)	
Real mortage rate after tax			0.244***		0.293***	
			(9.93)		(8.87)	
Decade 1992-2007					-0.381**	-1.505***
dummy variable					(-2.75)	(-11.49)
Decade 2008-2016					0	-2.136***
dummy variable					(.)	(-15.24)
_cons		0.632***	0.220	0.499***	0.362*	2.006***
		(9.13)	(1.59)	(7.01)	(2.54)	(14.45)
N		2074	1591	1591	1591	2074
Pseudo R2		0.020	0.109	0.022	0.111	0.076
Wald chi2		74.92	217.66	71.62	213.34	288.87
LL		-1164.53	-849.00	-932,13	-847.06	-1098.14
Goodness of fit		74.01%	75.05%	71.34%	74.92%	74.01%
	Sensitivity	*100%*	*94.10%*	*100%*	*93.83%*	*100%*
	Specificity	*0%*	*27.63%*	*0%*	*27.85%*	*0%*

z statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
Note: All models are estimated using a logit model (command *logit* in Stata). The standard errors are clustered by municipality. Goodness of fit is classified using *estat* classification. For more test regarding goodness of fit, se Appendix C

5.1 EXPANSION PHASES

In **model 6** we test for duration dependence with only the lagged duration as an explanatory variable. The coefficient is negative, meaning that the likelihood of staying in an expansion phase is expected to decrease the longer the phase continues.

In **model 6** we have included the covariates population growth, growth in real income, growth in real house prices and real mortgage rate (after tax). The sample period is restricted by mortgage rate and growth in real income, as we only have observations for these covariates between 1992-2012. The coefficient of lag duration is still negative but increases. Growth in population and income are expected to have a positive effect on the likelihood of staying in an expansion phase. Conversely, growth in real house prices and an increase in real mortgage rate are expected to affect the likelihood of staying in that phase negatively.

In **model 7** we test with only lag duration as an explanatory variable, and restrict the sample period to 1992-2012 to see if the change in the estimates of lag duration in model 5 and 6 are mostly due to sample period being smaller or due to the additional covariates. We see a small change in the estimate of lag duration (between model 6 and 7) which suggests that the effect of lag duration on the expected probability of continuing a phase, varies over time.

In **model 8** we have included the decade dummies. D2 is omitted because of collinearity, as this model only includes observations between 1992-2012. The reference group is therefore observations between 2007 and 2012. Growth in real income becomes insignificant, probably due to the fact that increases in income during the period 1992-2007 is largely caught by the time dummy D1. The estimate of lag duration is still negative but has decreased in magnitude, suggesting that much of the affect is caught by the decade dummy D1.

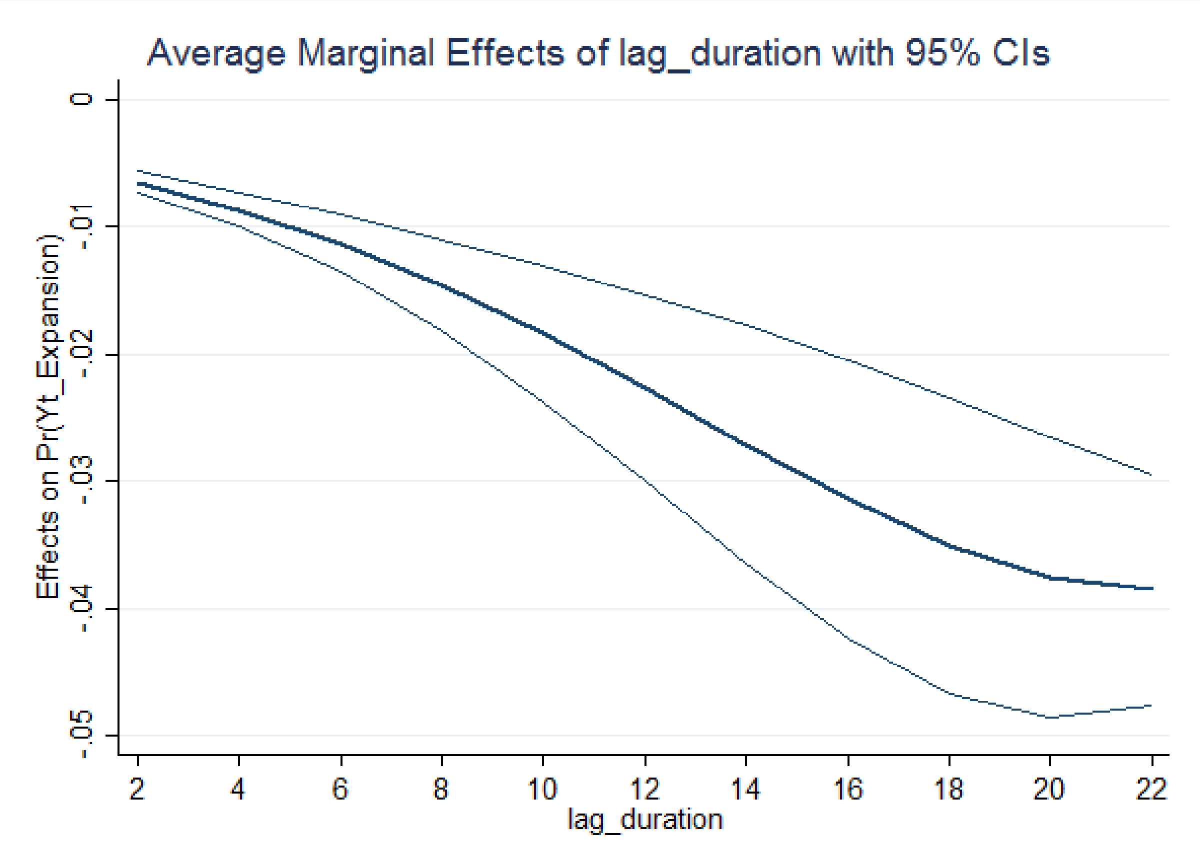
In **model 8** we have excluded all but the dummy variables and lag duration. Lag_duration and D1 are significant, but D2 is not. This suggests that the effects of lag duration is not different between the time periods 1982-1991 and 2008-2016.

The model with the best predictive power is model 8, it predicts 91 percentages correctly; with a sensitivity of 99 percent and specificity of 15 percent. This means that the model predicts the most likely outcome, being P(Y=1), correctly in 99 percent of the observations, and the least likely outcome P(Y=0) correctly in 15 percent of the observations.

Model 6 has also a strong predictive power, with a total correctly predicted observations of 91 percent, and a sensitivity of 99 percent and specificity of 6 percent. Non of the other models have any power when it comes to predicting the least likely outcome.

5.2 *EXPANSION PHASES: MARGINAL EFFECTS*

Figure I1: Average marginal effects of lag_duration using model 6



Since this is a non-linear model, the estimated impact of a change in lag_duration is different for every point of duration. In figure II, the average marginal effects of duration on the probability of leaving is graphed. The marginal effects is estimated using model 6 and by putting the other covariates to their means. The marginal effect of one additional year of duration increases as the duration of the phase increases. That is, the longer a municipality has been in an upturn, the more impact one additional year of expansion has on the probability of house prices declining in the subsequent years (i.e. the price cycle turning).

5.3 CONTRACTION PHASES

The estimations on contraction phases is stated in table VI. In **model 10** we include only one variable, namely the lag duration dummy. It is significant at the 1 % level, indicating that the length of the current contraction phase is indeed associated with the probability of the phase ending. The coefficient is positive, suggesting that a prolonged contraction phase increases the likelihood of remaining in that phase.

In **model 11,** all four covariates are included, all of which are statistically significant. The lag duration is still positive with a slightly higher coefficient. Growth in population and real income have negative signs, suggesting that they are negatively correlated with the likelihood of staying in a contraction phase. Conversely, growth in real house prices have the opposite sign (as in the corresponding *model 6* for expansions). The pseudo-R^2 are higher than before, suggesting that this specification better explains the variation in the dependent variable.

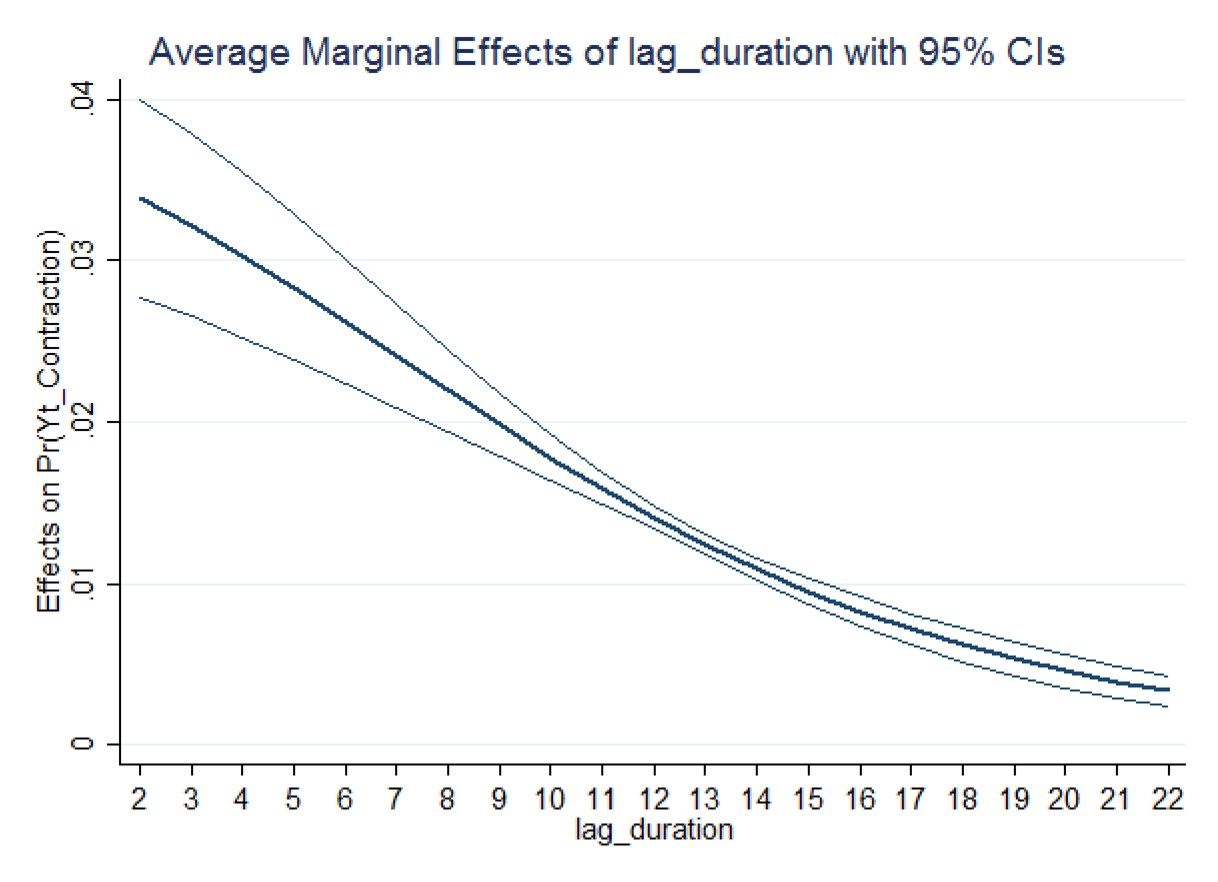
In **model 12**, we have only included the variable lag duration but restricted the sample to the time perion 1992-2012. The coefficient of the lag duration is only slightly smaller than that of model 10 and 11, suggesting that the effect of duration on the probability of exiting a phase is not time variant.

In **model 13** the decade dummies D1 and D2 are added, the latter of which is omitted due to collinearity. Correlation analysis tells us that D2 is highly correlated with lag duration, probably due to the short-lived correction in the Swedish housing market in the first years of the financial crisis.

In **model 1** we have excluded all four macroeconomic covariates. All time dummy variables are statistically significant at the 1 % level. Compared to the corresponding expansion *model 8* above, the decade dummy D2 is now statistically significant.

5.4 CONTRACTION PHASES: MARGINAL EFFECTS

Figure II2: Average marginal effects of lag_duration using model 11



In figure II, the marginal effects of lag duration on the probability of the continuing of a current contraction phase is graphed. The marginal effect is decreasing with duration, suggesting that the longer a contraction phase has endured, the less is the impact of one extra year of duration on the probability remaining in the contraction phase.

6 CONCLUSION

From our estimations we can conclude that house price cycles appear to have a positive duration dependence for expansion phases. The longer the phase, the less the likelihood of continuing that phase. These findings are consistent with those of Cunningham & Kolet (2011) on the US housing market, as well as Bracke's (2011) studies of the Swedish housing market.

Both population and real income growth are expected to increase the probability of continuing an expansion phase. Conversely, increase in real mortgage rate and house price growth, are expected to decrease the probability. The hazard rate (see Appendix B) confirms this intuition. The distribution is skewed to the right and so the rate at which municipalities exit upturns seem to be related to its length. The additional goodness of fit tests seen in Appendix C tells us that models 5 and 6 are significant, further validating this conclusion.

The estimations for contraction phases suggests a negative duration dependence, with a decreasing effect, meaning that the longer a contraction phase has endured, the less likely is it to end. The hazard rate of downturns confirms this. They are not as conclusive as in upturns and the distribution isn't as clearly Even though models 9 and 10 indicate robust estimation (see Appendix C), the overall results from contractions are inconclusive.

One of the limitations of our study, is the short sample period of covariates, specifically real income and mortgage rates that are only observed from 1992. This effectively leads to a smaller sample size. If these covariates could be observed for a longer time period, the estimations might differ.

Given the limited availability of free data on the housing market, our analysis does have inherent shortcomings; an estimation including, for example, user owned flats (bostadsrätter) would make more sense since both markets operate under, mostly, the same mechanisms. Furthermore, the measures of income could be refined. Having household disposable income on a municipality level would, most likely, increase the viability of the analysis and provide a more comprehensive picture of the local housing markets in Sweden.

7 ACKNOWLEDGEMENTS

We would like to thank our thesis supervisor Inge van den Bijgaart for her insightful help and support. We are also grateful to Kent Olsson for all his help and enthusiasm.

We would also like to thank Marie Rosberg at Boverket, Martin Verhage and Niclas Sjölund at Statistics Sweden; for providing us with valuable data.

8 REFERENCES

8.1 PRINTED

Bénétrix, A.S., Eichengreen, B., O'Rourke, K.H., (2012). *How housing slumps end*. Economic Policy 27 (72), pp. 647–692.

Bracke, Philippe, (2011) *How Long Do Housing Cycles Last? A Duration Analysis for 19 OECD Countries*, IMF Working Papers, 11/231., pp. 1-29

Cameron, A. and Trivedi, P. (2005). *Microeconometrics: Methods and applications*. New York, N.Y.: Cambridge University Press., pp. 473-473

Claessens, Stijn, M. Ayhan Kose, and Marco E. Terrones (2011), *Financial Cycles: What? How? When?* IMF Working Paper 11/76 (Washington: International Monetary Fund).

Cunningham, Rose and Kolet, Ilan (2011); *Housing market cycles and duration dependence in the United States and Canada*, Applied Economics, 43:5, pp. 569-86

Giordani, P., A. Grodecka, S. Kwan, P. Morales, Dilan Ölcer and Erik Spector (2015), *Asset valuation and financial stability*, Economic Commentaries, No. 15, Sveriges Riksbank

Englund, Peter (2011), *Swedish house prices in an international perspective*, the Riksbank's commission of inquiry into risks on the Swedish housing market, Sveriges Riksbank.

Flam, Harry (2016), *Har vi en bostadsbubbla?(Are we in a housing bubble?)* Ekonomisk Debatt, 4/2016, Swedish Economics Association

Gujarati, Damodar N. (2015). *Econometrics by Example*. Second ed., pp. 344

Kennedy, P. (2008). *A guide to econometrics* (6.th ed.). Oxford: Blackwell., pp. 248

Lancaster, Tony (1979). *Econometric Methods for the Duration of Unemployment*, Econometrica, vol. 47, no. 4, pp. 939–956

Ohn Jonathan, Larry Taylor, Adrian Pagan (2004), *Testing for Duration Dependence in Economic Cycles*, Econometrics Journal, Vol. 7, No. 2, pp. 528–49

Sørensen, Peter Birch (2013), *The Swedish Housing Market: Trends and Risks*, Report to the Swedish Fiscal Policy Council, 2013/5, Swedish Fiscal Policy Council.

Turk, Rima A., (2015), *House prices and household debt interactions in Sweden*, IMF Working Paper No. 15/276.

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.,

Wooldridge, J. M.(2003) *Introductory econometrics : a modern approach*. Australia Cincinnati, Ohio: South-Western College Pub.

The IMF (2012), *World Economic Outlook (WEO): Growth Resuming, Dangers Remain*, Chapter 3: Dealing with Household Debt, April 2012

8.2 DATA SOURCES

SCB

- **Average income, per municipality**

  Sammanräknad förvärvsinkomst, medelinkomst för boende i Sverige den 31/12, tkr efter region, kön, ålder, inkomstklass och år.

  Internal reference code: HE0110K1

  Contact person: petter.lundberg@scb.se

  Tabellinnehåll: Medelinkomst, tkr **Region**: Kommun/alla **Kön**: totalt **Ålder**: totalt 16+ **Inkomstklass**: totalt **År**: 1991-2015

  [2017-12-19]

- **Population size, per municipality**

  Folkmängden efter region, tabellinnehåll och år

  Internal reference code: BE0101N1

  Contact person: tomas.johansson@scb.se

  [2017-12-19]

- **Real house prices per municipality**

  Köpeskilling för småhus, medelvärde i tkr efter region, fastighetstyp och år

  Internal reference code: BO0501C2

  Contact person: martin.verhage@scb.se

  [2017-12-19]

- **KPI 1980=100**

  Internal reference code: 000000KL

  Contact person: angelica.arellano@scb.se

  [2017-12-19]

Boverket

Real mortgage rates, after tax. 1991-2012

Sent by e-mail from Marie Rosberg at Boverket 2017-11-20

APPENDIX A: MUNICIPALITY SPECIFICATIONS

A.1 MUNICIPALITY DIVISION: CHANGES DURING SAMPLE PERIOD

Municipalites formed during sample period and previously part of.[9]

Formed	Id	Name	Previously part of:	
1983-01-01	0128	Salem	0127	Botkyrka
1983-01-01	2418	Malå	2417	Norsjö
1983-01-01	1445	Essunga	1470	Vara
1983-01-01	0117	Österåker	0187	Vaxholm
1983-01-01	2403	Bjurholm	2460	Vännäs
1992-01-01	0461	Gnesta	0480	Nyköping
1992-01-01	0488	Trosa	0480	Nyköping
1995-01-01	1443	Bollebygd	1490	Borås
1995-01-01	1814	Lekeberg	1880	Örebro
1999-01-01	0140	Nykvarn	0181	Södertälje
2003-01-01	0330	Knivsta	0380	Uppsala

Municipalities formed during the sample period is not included in the sample.

A.2 METROPOLITAN AREAS[10]

Metropolitan areas from Sweden Statistics definition

1970 – 2004

Stockholm:		**Göteborg:**	**Malmö:**
0127 Botkyrka	0184 Solna	1440 Ale	1231 Burlöv
0162 Danderyd	0180 Stockholm	1480 Göteborg	1261 Kävlinge
0125 Ekerö	0183 Sundbyberg	1401 Härryda	1262 Lomma
0136 Haninge	0138 Tyresö	1384 Kungsbacka	1281 Lund
0126 Huddinge	0160 Täby	1482 Kungälv	1280 Malmö
0123 Järfälla	0114 Upplands Väsby	1441 Lerum	1230 Staffanstorp
0186 Lidingö	0139 Upplands-Bro	1481 Mölndal	1263 Svedala
0182 Nacka	0115 Vallentuna	1402 Partille	1287 Trelleborg
0128 Salem		1415 Stenungsund	1233 Vellinge
0191 Sigtuna		1419 Tjörn	
0163 Sollentuna		1407 Öckerö	

Added 2005

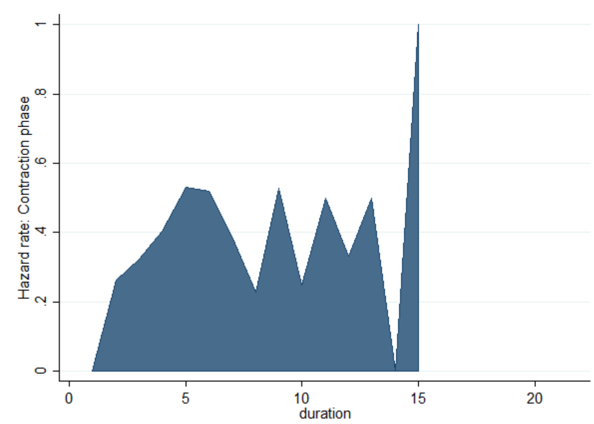
Stockholm:	**Göteborg:**	**Malmö:**
0188 Norrtälje	1489 Alingsås	1267 Höör
0140 Nykvarn	1462 Lilla Edet	1264 Skurup
0192 Nynäshamn		1285 Eslöv
0181 Södertälje		

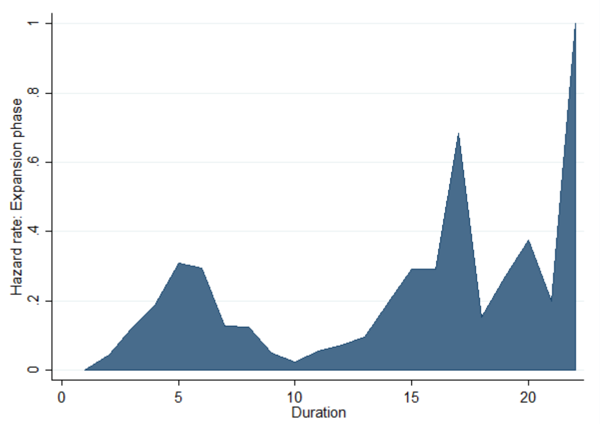
[9] Information provided by SCB as a note in the download file of the housing price data.
[10] https://www.scb.se/Grupp/Hitta_statistik/Regional%20statistik/Indelningar/_Dokument/Storstadsomr.pdf [2017-12-19]

APPENDIX B: THE HAZARD FUNCTION

Let T be a nonnegative random variable denoting the to a failure event. In econometric analysis, the *probability distribution function* $f(t)$ (or cumulative distribution function $F(t) = \Pr(T \leq t)$) is used to describe how the probability to an event is distributed over time. In survival analysis, the *hazard function* $h(t)$is preferred over the pdf merely as a matter of convenience[11].

The hazard function (or *empirical hazard rate*) is simply the total number of municipalities exiting a contraction (or expansion) phase at time *t* divided by the total number of municipalities that were still in a contraction (or expansion) phase at the beginning of *t*. We show the empirical hazard rates for all expansion and contraction phases separately.





Graph A1

The hazard rate for contraction phases

Graph A2

The hazard rate for expansion phases

Even though we aren't using a standard duration model, per se, we thought these graphs would serve as an intuition as to how the likelihood of exiting a phase changes with time (how long the subject has been in a particular phase).

From a graphical point of view, the graphs suggest that the hazard rate for exiting an expansion phase are somewhat skewed to the right. In other words, the municipal housing markets in Sweden seem to exhibit more (less) duration dependence whilst in an expansion (contraction) phase. The corresponding graph for contraction exhibit the opposite, if any, duration dependence; there is a vague distribution of hazard rates skewed to the left.

[11] Cleves, M.A., Gould, W.M., Gutierrez, R.G., (2004). *An Introduction to Survival Analysis Using Stata*. Stata Press, College Station

APPENDIX C: GOODNESS-OF-FIT TESTS

Goodness of fit typically involves describing how well a model fits observed values with the predictied values. In linear models, practitioners often look at, the R^2 , among other statistics. For nonlinear models, the logit in our case, the corresponding measure is the pseudo-R^2. We have included those values in Table V and VI, as mentioned earlier. As is the case with the R^2 for linear models, there are multiple ways of interpreting the pseudo-R^2.

There is no consensus on which approach is the ”best” for determining a binary models adequacy (Cameron and Trivedy, 2005), so one has to look at others, in addition to the pseudo-R^2.

One common approach is to look at the predicted outcomes and compare to the observed values. With a binary dependent variable, one has to translate the predicted values to either a 0 or a 1, in order to do this comparison. Then the expression $\sum_i (y_i - \hat{y})^2$ gives the total number of predictions that were wrong, which would occur when $(y_i - \hat{y})$ is equal to either $(0, 1)$ or $(0, 1)$. There is, though, a drawback to this method. For example, using $\hat{y} = 1$ when $\hat{p} > 0.5$ as a prediction rule could overestimate a models predictive power if, say, almost all observations are $y = 1$.

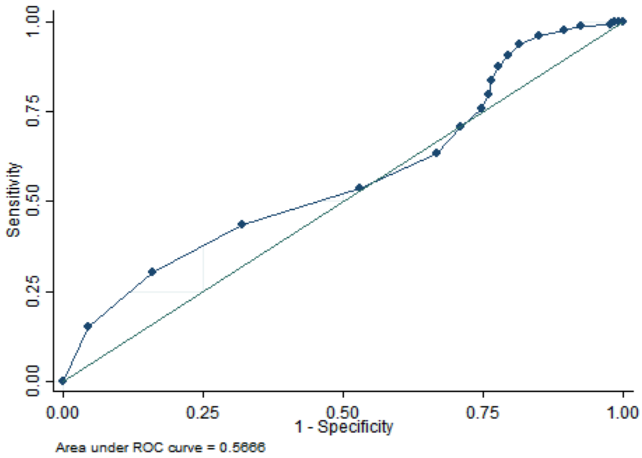
Since we estimate our models for expansion and contraction phases separately, almost all observatitions, in either of the two samples for each model, are $y = 1$. Remember, that for our sample, the change in y_i from unity to zero occurs only when there is a turning point identified by the BBQ-algorithm. In other words, since we estimate contractions and expansions separately, the observations for either of the two samples will contain almost exclusively ones.

Returning to the prediction rule above, we look at different cutoff values (c) for which $\hat{y} = 1$. By letting $\hat{y} = 1$ when $\hat{p} > c$ and plotting the proportion of $y = 1$ correctly specified against the proportion of $y = 0$ incorrectly specified, we will obtain what is called the *receiver operating characteristics* (ROC) surve (Cameron and Trivedy, 2005). The more area under this curve, the better the predictive power the model has. This is done for all ten models separately.

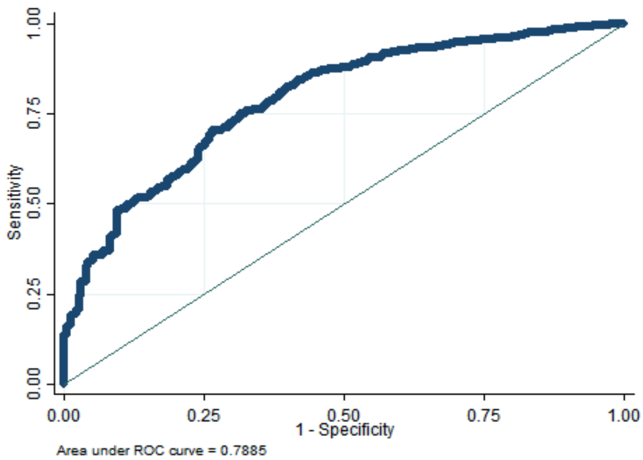
C.1 EXPANSION PHASES

7 är ny och 12

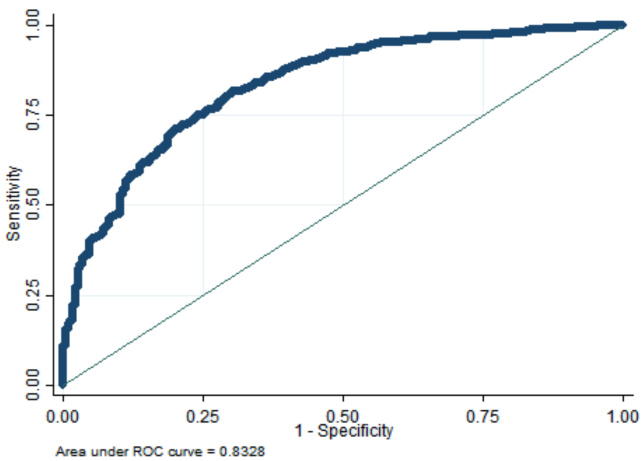
Model	**(1)**	**(2)**	**(3)**	**(4)**
Area under ROC curve	0.5666	0.7885	0.8328	0.7290
Hosmer-Lemeshow test				
Number of observations	4008	2583	2583	4008
number of groups	8	10	10	10
Hosmer-Lemeshow Chi^2	174.05	10.68	7.51	124.59
Prob > Chi^2	0.0000	0.2204	0.4823	0.0000



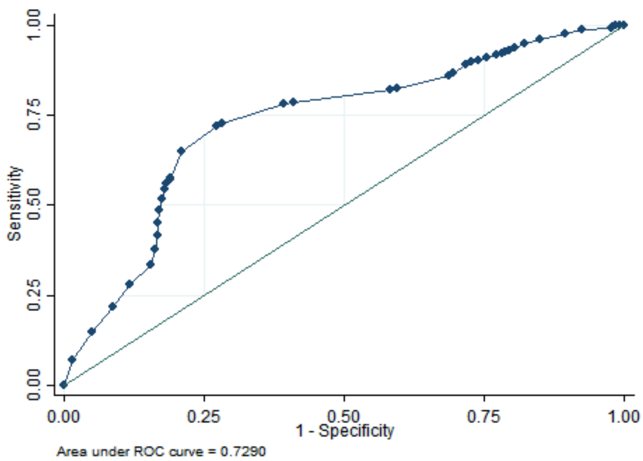
Model 1



Model 2



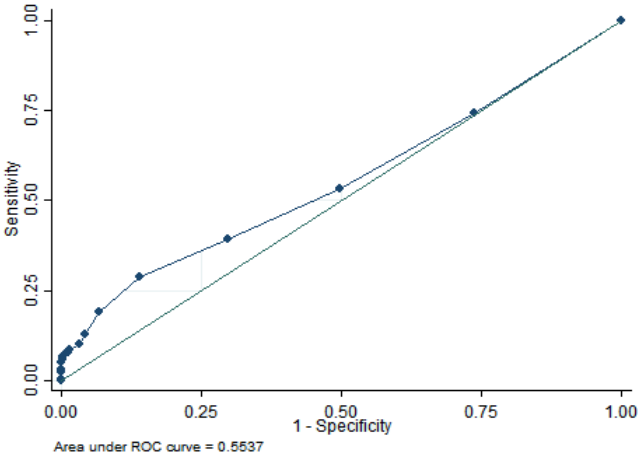
Model 3



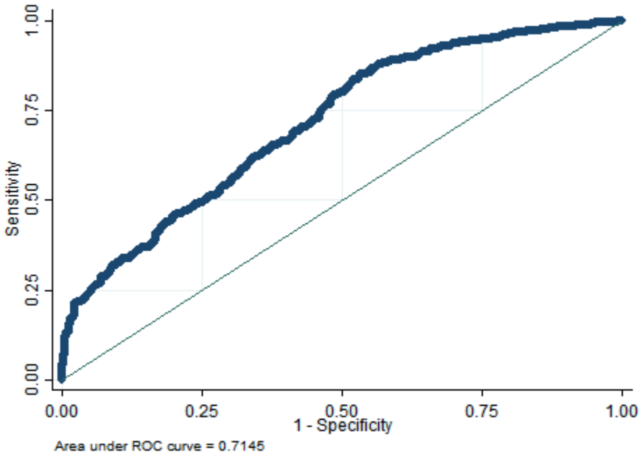
Model 4

C.2 CONTRACTION PHASES

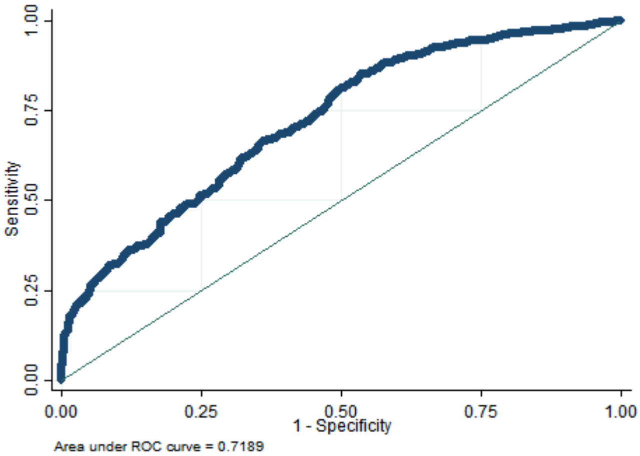
Model	(5)	(6)	(7)	(8)
Area under ROC curve	0.5537	0.7145	0.7189	0.6596
Hosmer-Lemeshow test				
Number of observations	2074	1591	1591	2074
number of groups	7	10	10	9
Hosmer-Lemeshow Chi^2	35.01	36.01	24.71	63.26
Prob > Chi^2	0.0000	0.0000	0.0017	0.0000



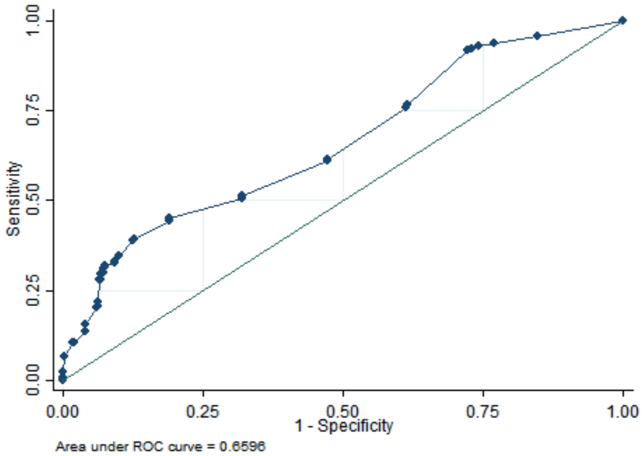
Model 5



Model 6



Model 7



Model 8