

VALUATION OF RESIDENTIAL MORTGAGE DEFAULT AND PREPAYMENT UNDER
STOCHASTIC HOUSE PRICES

by

KONSTANTIN LYUBIMOV

(Under the direction of James B. Kau)

ABSTRACT

This paper applies a valuation model for residential mortgages subject to exogenously specified conditional probabilities of prepayment and default to data on subprime and Alt-A adjustable-rate mortgages. In conformity with the doubly stochastic framework, risk-neutral baseline hazard rates of prepayment and default in the model are dependent on stochastically evolving latent factors, driven by Brownian motions. Hazard processes also incorporate observable variables reflecting evolution of interest rates and house prices. The model is estimated on two subsets of the data: one includes early vintages of non-prime ARMs, the other is comprised of 2005 and later vintages. Results of the calibration of the two sets of model parameters to market prices suggest that whatever changes in the unobserved borrowers' behavior took place in the observation period were relatively soon recognized by the market. Simulations show that for subprime borrowers the effect of house price fluctuations on the probability of default is very significant.

INDEX WORDS: doubly stochastic, default, prepayment, latent factor, nonlinear filtering, mortgage valuation. JEL classification: C15, C41, E44

VALUATION OF RESIDENTIAL MORTGAGE DEFAULT AND PREPAYMENT UNDER
STOCHASTIC HOUSE PRICES

by

KONSTANTIN LYUBIMOV

B.A., Rostov State University of Economics, 1991

M.S., St Petersburg State University of Economics and Finance, 1995

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2010

© 2010

Konstantin Lyubimov

All Rights Reserved

VALUATION OF RESIDENTIAL MORTGAGE DEFAULT AND PREPAYMENT UNDER
STOCHASTIC HOUSE PRICES

by

KONSTANTIN LYUBIMOV

Approved:

Major Professor: James B. Kau

Committee: Donald C. Keenan
Henry J. Munneke

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2010

ACKNOWLEDGMENTS

I am grateful to my supervisor, Professor James Kau, for his continuing support and encouragement during my work on this project. I would like to thank Professor Donald Keenan, whose attention is the reason why many erroneous arguments did not make it to the final version. I would also like to acknowledge constructive comments by Professor Henry Munneke. My understanding of the delicate issues of statistical modeling and model assessment benefited from discussions with Dr Jaxk Reeves (Department of Statistics, UGA). I also owe a debt of gratitude to Professor Edsel Peña (Department of Statistics, University of South Carolina), who graciously shared his expert knowledge of advanced survival models. As usual, all remaining errors are my own. I am thankful to my family for their patience and kindness.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iv
LIST OF FIGURES	vi
LIST OF TABLES	vii
 CHAPTER	
1 INTRODUCTION	1
2 THE PRICING MODEL AND ITS COMPONENTS	5
2.1 AN OVERVIEW OF APPROACHES TO THE VALUATION OF RISKY DEBT	5
2.2 NO-ARBITRAGE PRICING MODEL FOR THE MORTGAGE CONTRACT SUBJECT TO DEFAULT AND PREPAYMENT	15
2.3 MULTIPLICATIVE INTENSITY MODEL WITH COMPETING RISKS OF PREPAYMENT AND DEFAULT	17
2.4 CHARACTERIZATION OF THE CIR TERM STRUCTURE MODEL	23
2.5 EXISTING APPROACHES TO HOUSE PRICE MODELING	26
2.6 AN OVERVIEW OF LITERATURE RELATED TO INTENSITY MODELING	32
3 ESTIMATION OF THE PARAMETERS OF THE MODEL	39
3.1 ESTIMATED PARAMETERS OF THE MULTIPLICATIVE INTENSITY MODELS	39
3.2 ESTIMATION OF AFFINE DIFFUSION-DRIVEN TERM STRUCTURE MODEL	53
3.3 THE HOUSE PRICE PROCESS: THE MODEL AND THE ESTIMATION STRATEGY	59

3.4	PRACTICAL IMPLEMENTATION OF SMC ML ESTIMATION	63
4	CALIBRATION OF THE PRICING MODEL TO THE OBSERVED MARKET PRICES OF RESIDENTIAL MORTGAGES	105
4.1	PRICE OF A MORTGAGE CONTRACT AS A CONDITIONAL EXPECTATION	105
4.2	MONTE CARLO IMPLEMENTATION	108
4.3	APPLICATION OF THE MODEL TO NON-PRIME MORTGAGE MARKET .	111
5	CONCLUSION	119
	BIBLIOGRAPHY	122
	APPENDIX: DRIFT AND DIFFUSION OF THE BOND PROCESS	143

LIST OF FIGURES

1.1	The structure of the reduced-form mortgage valuation model	4
3.1	Mean contract rate, rate on conventional mortgages and T-bill rate during the observation window	75
3.2	Default and prepayment incidence during the observation window	76
3.3	Estimates of the hazards of default for a representative contract under the 'old' regime and under the 'new' regime	82
3.4	Cumulative baseline hazards of default (left) and prepayment (right) esti- mated under the 'old' regime	83
3.5	Cumulative baseline hazards of default (left) and prepayment (right) esti- mated under the 'new' regime	84
3.6	Ratio of the number of foreclosed mortgages to the total number of observed mortgages for an 'old' regime cohort versus a 'new' regime cohort	85
3.7	Monthly series of yields for 5 maturities: January 1970 - December 2008 . .	87
3.8	Simulated paths of yield on 10-year bond under estimated parameters of CIR interest rate process	87
3.9	Smoothed state estimates from Kalman filter, yield on 3-month zero- coupon bond and yield on 10-year zero-coupon bond	89
3.10	Series of smoothed state and response disturbances for the first factor of the CIR term structure	90
3.11	Factor loadings for the estimated parameters of two latent factors	90
3.12	Case-Shiller HPA indices for selected MSAs	91
3.13	Series of relative returns (q) for selected MSAs	92

3.14	Simulated house price appreciation indices for selected MSAs	92
3.15	Estimates of conditional standard deviation from the GARCH models for selected MSAs	93
3.16	The diffusion index for annualized nominal returns across 20 MSAs and the nationwide rate of nominal house price appreciation	94
3.17	Anscombe residuals from default and prepayment models (CIR specifica- tion)	101
3.18	Anscombe residuals from default and prepayment models (log-OU speci- fication)	102
3.19	One-step-ahead predictive density and actual counts of default and pre- payment	103
4.1	Spreads between subprime credit default swaps index (ABX) and LIBOR, thousands basis points	116

LIST OF TABLES

3.1	Multiplicative intensity models: variable definitions	69
3.2	Number of originated and terminated loans by the quarter of origination (the 'old' regime)	70
3.3	Number of originated and terminated loans by the quarter of origination (the 'new' regime)	71
3.4	Summary statistics for model variables and major loan characteristics . . .	72
3.5	Summary statistics for model variables and major loan characteristics for the subset of defaulted loans	73
3.6	Summary statistics for model variables and major loan characteristics for the subset of loans that were prepaid	74
3.7	Summary statistics for the fixed-rate model variables	76
3.8	OLS estimates of the model for the FRM rate	77
3.9	Test statistics for the presence of the calendar-time stratum effects	77
3.10	Comparison of out-of-sample fit for the 'old' and the 'new' models of default	78
3.11	Estimates of the multiplicative intensity default model for the 'old' regime	78
3.12	Estimates of the multiplicative intensity prepayment model for the 'old' regime	79
3.13	Estimates of the multiplicative intensity default model for the 'new' regime	80
3.14	Estimates of the multiplicative intensity prepayment model for the 'new' regime	81
3.15	Estimates of the (empirical) cumulative hazard of default	86
3.16	Summary statistics for the yields	86

3.17	Estimates of the parameters of the affine term structure model	88
3.18	Pearson correlation coefficients between HPA indices	88
3.19	Estimates of the relative housing returns (q) process	95
3.20	Estimates of the relative housing returns (q) process	96
3.21	Estimated parameter values of the GARCH models of housing returns . . .	97
3.22	Estimated parameter values of the GARCH models of housing returns (continued)	98
3.23	ML estimation of the latent termination processes.	99
3.24	ML estimation of the latent termination processes.	100
3.25	Pearson GOF statistic for different model specifications	104
4.1	Calibrated risk adjustments, liquidity and expected loss parameters for for the 'old' regime model (estimated on Q1 1997- Q2 2004 data)	115
4.2	Calibrated risk adjustments, liquidity and expected loss parameters for the 'new' regime model (estimated on post-2004 data)	117
4.3	Predicted hazard of default under the 'old' and the 'new' regime	118

Disclaimer

Part of this material was based upon work partially supported by the National Science Foundation under Grant DMS-0635449 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. The author bears sole responsibility for all errors and shortcomings.

CHAPTER 1

INTRODUCTION

This study proposes a reduced-form model for valuation of residential mortgages, which accounts for uncertainty in housing prices. Historically, reduced-form modeling evolved as an answer to certain deficiencies in the structural approach. In the seminal works by Black and Scholes (1973) and Merton (1974), a tractable solution to the problem of asset valuation was developed, where the issuer of these claims is subject to default risk. In structural models, default probability depends on the ‘fundamentals’, such as market capitalization of the firm and its capital structure.

In a basic structural model, the firm’s equity is a call option on the assets of the firm with a strike price equal to the value of the debt that the firm owes. Under the assumption of the frictionless market (zero bankruptcy costs) the risky debt pricing reduces to pricing the European option. Modigliani and Miller’s (1958) irrelevance of capital structure is built into this approach. The value of the firm’s assets is by definition equal to the value of the debt plus the value of the equity. If the value of the firm’s assets at the date when the debt matures is less than the face value of the latter, then the firm defaults on its debt. Black and Cox (1976) further generalized this model by assuming that default can occur not only at a maturity date but at any point in time, when the asset value reaches some (sufficiently low) threshold value or exogenous default boundary.

Subsequent research extended the framework of Black-Scholes and Merton in several directions: Longstaff and Schwartz (1995) introduced stochastic interest rates, thus allowing for two (possibly correlated) sources of uncertainty. Geske (1977) suggested a

compound option model allowing for different maturities of the debt. The contribution of Leland and Toft (1996) incorporated bankruptcy costs and taxes in the model.

Applications of the structural credit-modeling approach to the mortgage valuation have been proposed by, for example, Titman and Torous (1989) as well as Kau, Keenan, Muller, and Epperson (1992). Mortgage termination by default in this framework is analyzed as rational behavior by a profit-maximizing economic agent. This agent defaults if the value of the property falls below the current value of the remaining payments on the mortgage contract adjusted by the value of the prepayment option and option to default later in the future. Subsequent work (e.g., Stanton (1995)) attempted to reconcile predictions of structural models with empirically observed behavior of many borrowers who failed to prepay when interest rates dropped. Alternative explanation of the ‘suboptimal’ behavior (Longstaff (2003)) presumes an rational borrower who is subject to credit constraints.

Although some structural models are applied in industry (Moody’s KMV is one such example) and have clear economic interpretation, empirical research has shown (e.g. Jones, Mason, and Rosenfeld (1984)) that in applications the Black-Scholes-Merton model tends to underestimate observed credit spreads (that is, spreads between return on risky debt and ‘riskless’ debt). Another deficiency is that usually the underlying asset values (firm assets or property values) are not easily observable. Moreover, if the asset value process is modeled as a pure diffusion, then for the borrower whose assets are in the vicinity of the default boundary, the probability of defaulting the next point in time is technically very close to one: that is, default is predictable. Some of the more recent research (e.g., Collin-Dufresne and Goldstein (2001)) introduces extensions that alleviate the most obvious discrepancies with real-world behavior, such as model-predicted spreads that decrease with maturity. However, often the resulting structural model turns out to be fairly complicated due to the modifications that need to be made.

A reduced-form approach emerged as an alternative to the structural approach in 1990s. Artzner and Delbaen (1995), Jarrow and Turnbull (1995), Lando (1998) and Duffie and Singleton (1999), among others, consider default to be an unpredictable event; it is defined as the first arrival time t of a non-explosive counting process with intensity λ . Hence, the default process is specified exogenously. In many applications, arrivals are assumed to be governed by the Poisson law, whereby the intensity λ can vary stochastically over time. The models of this type are also referred to as intensity-based models and they are exceedingly popular in the literature on reduced-form credit modeling. Oftentimes modelers employ doubly stochastic Poisson (or Cox) process. Intensity models were applied to pricing of corporate debt (Duffie (1999)), interest rate spreads (Duffie and Singleton (1997)), sovereign debt (Duffie, Pedersen, and Singleton (2003)) and various types of credit derivatives (Houweling and Vorst (2005)), but not until recently to mortgage pricing.

My goal is to develop a reduced-form model extending the framework of Kau, Keenan, and Smurov (2004), which appears to be the first application of intensity-based models for valuation of residential mortgages. I seek to incorporate additional stochastically evolving factors, other than interest rates, that are relevant to the decision-making process of the borrowers. Of particular interest is the price of collateral. This extension is a step towards augmenting reduced-form models by some elements of structural approach. The scheme of the implementation is presented in the figure 1.1.

The data from an extensive nationwide panel of mortgages are used to estimate empirical models of mortgage terminations. The parameters of these latter are then used to infer parameters of the (latent) stochastic baseline processes of default and prepayment, which are assumed to be driven by Brownian motions. The value of the mortgage is determined by evaluating the conditional expectation of the mortgage payoff given the risks that the contract might be terminated due to default or prepayment prior to maturity date. This expectation is conditioned on dynamics of five stochastic processes: latent factors for both

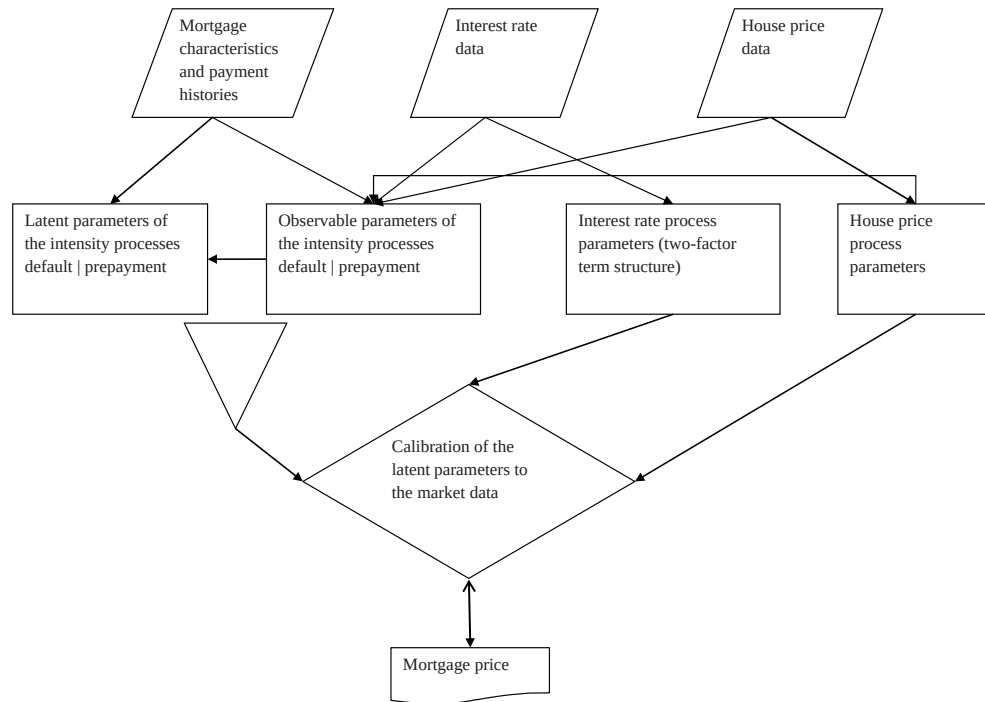


Figure 1.1
The structure of the reduced-form mortgage valuation model

the default and prepayment intensity, two factors in the term structure of interest rates, and a stochastic house price process. Model parameters are calibrated to the observed market prices of mortgages. The complete model is applied to the analysis of the major increase in defaults in non-prime segment of the mortgage market.

The dissertation is organized as follows: first two Sections of Chapter 2 contain a short literature review and discusses the pricing framework, Section 3 reviews the estimation technique and examines duration models estimates, Section 4 is devoted to the

approaches to estimating the term structure of interest rates, Section 5 discusses the modeling of house prices as a stochastic process, Section 6 reviews techniques for estimation of the latent default and prepayment processes' parameters. Chapter 3 reports the results of the estimation of these components of the valuation model. Chapter 4 discusses the calibration procedure and the application of this model to uncovering the main factors in the surge in defaults in U.S. non-prime mortgage sector. Chapter 5 concludes.

CHAPTER 2

THE PRICING MODEL AND ITS COMPONENTS

2.1 AN OVERVIEW OF APPROACHES TO THE VALUATION OF RISKY DEBT

In this section I would briefly discuss the extant literature related to modeling of credit events and mortgage termination in particular. Historically, the structural (option-theoretic) approach to the valuation of the risky debt (Black and Scholes (1973), Merton (1974)) preceded the reduced form models .

Early structural models implied an absence of frictions in the market. In the option-theoretic framework of Black-Scholes-Merton the process for the market value of underlying assets H (assets of the firm or, say, market price of a residential property) was assumed to be log-normal under the risk-neutral measure $\mathbb{Q}$, that is

$$\frac{dH_t}{H_t} = (r - \pi)dt + \sigma dW_t \quad (2.1)$$

where r was the risk-free rate of return, π was the cash payout rate (service flow in case of housing), σ was the instantaneous volatility of the standard Brownian motion W_t . Technical pre-requisites¹ include a measure space $(\Omega, \mathcal{F}, \mathbb{Q})$ equipped with a filtration $\mathbb{F} \equiv (\mathcal{F}_t)_{0 \leq t \leq T^*}$, which supports the value process H . Assets are financed with debt D with maturity $T \leq T^*$ ($T^* > 0$ denotes the time horizon) and notional amount F .

Assume that default can only occur at maturity, when the value of the underlying (real property) is less than F , $H_T < F$. Then, the value of the debt (for the property which is financed by both debt and equity) at the time of default is at most equal to the value of

¹See, for example, Bingham and Kiesel (2004) whose exposition I follow in my discussion of the pricing model.

the property $D_T = H_T$. Creditors' payoff is given by $D_T = F - \max\{F - H_T, 0\}$. This is the payoff of the (European) put option on the value of the underlying asset. The price of the zero-coupon debt with a face value D is the difference between payment of F discounted at the risk-free rate and a (European) put option on the value of the underlying

$$P_D(t, T) = Fe^{-r(T-t)} - P_E(H_t, F) \quad (2.2)$$

where the put option value P_E is given by the Black-Scholes formula

$$P_E(H_t, F) = -H_t e^{-\pi(T-t)} \Phi(-d_1(H_t, T-t)) + Fe^{-r(T-t)} \Phi(-d_2(H_t, T-t)) \quad (2.3)$$

with

$$d_1 = \frac{\log(H_t/F) + (r - \pi + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \quad (2.4)$$

and

$$d_2 = \frac{\log(H_t/F) + (r - \pi + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} - \sigma\sqrt{T-t} \quad (2.5)$$

and where Φ is the cumulative density of the standard normal distribution. This model yields unambiguous predictions regarding the timing of default, provides clear linkage to the economic fundamentals and possesses a closed-form solution. The asset process can rarely be observed, but in the case of corporate bonds, asset values can be inferred from the observable share prices. However, researchers who tested Merton model (e.g., Jones, Mason, and Rosenfeld (1984)) using time series analysis of prices of defaultable coupon (callable) bonds, were for the most part unable to obtain the levels of yield spreads corresponding to those observed in the marketplace. The original model was later extended to incorporate the possibility of default occurring not only at maturity, but at any moment between origination and maturity dates. For example, default may occur the first time the value of the underlying asset decreases below the value of some constant default boundary F . In this case, the time of default τ is the first passage time of the diffusion representing the asset process through the default-triggering boundary: $\tau \equiv \inf\{t > 0 : H_t < F\}$, where, in general, suitable barrier process f_t can substitute for F . Black and

Cox (1976) were the first to introduce this extension to the original Merton model and since then many variants of the first-passage model were suggested. The probabilities of default in this setting are higher than those in the original Merton model, however, for most realistic continuous-time asset and boundary processes closed-form solutions for first passage times are not available. One prominent example is the model of Leland and Toft (1996), who allowed for the possibility of continuous issuance of new debt by the firm. Proceeds from new equity offerings can be used to service the debt in case firm's earnings (which are functions of the assets) are insufficient to cover the dividends. Therefore, the Leland-Toft model implies endogenous default boundary. The study of Eom *et al* (Eom, Helwege, and Huang 2004) indicated that, unlike the Merton model, the Leland-Toft model tends to overpredict yield spreads for a broad range of maturities and issuers.

Yet another shortcoming of the above structural models is the consequence of the assumption that asset values follow a diffusion process. This would lead to vanishingly low instantaneous default probability of a borrower in good standing on a short-maturity debt, and, hence predicted yield spreads on such debt would be near zero. Zhou (2004) suggested augmenting the diffusion representing asset values by a Poisson jump process. Estimation of models in which asset values are driven by Levy processes (with or without time change), is an area of much research effort (see Wu (2008) for a recent review), however, still it is a non-trivial problem. An alternative device to achieve more plausible values of short-term spreads is to allow for measurement error in asset values (Duffie and Lando 2001).

It is noteworthy that yield spreads generated by earlier structural models decline with increasing maturity. This counterfactual implication of these models is the consequence of the specification for the value of the underlying: if the asset process is log-normal with positive drift, then, as time passes, the value of the asset will tend to drift away from any deterministic default boundary, when the capital structure is static. This issue was addressed, among others, by Collin-Dufresne and Goldstein (2001), who built

upon the work of Longstaff and Schwartz (1995). Colin-Dufresne and Goldberg suggested that rather than rolling over the same notional amount of debt, debtors might attempt to maintain some target leverage ratio. This leverage ratio can be modeled as a mean-reverting process (e.g. Ornstein-Uhlenbeck). The aforementioned study of Eom, Helwege, and Huang (2004) included an empirical test of the Colin-Dufresne-Goldberg model: it performed better than both the Merton and the Longstaff-Schwartz models with respect to obligations of the borrowers with good credit rating, but for relatively risky debt predicted spreads that were too high.

Evolution of structural models applied to mortgages and mortgage-backed securities followed the path outlined above, broadly speaking. The earliest models, e.g. Cunningham and Hendershott (1984), were cast in Merton's framework, typically with a single source of uncertainty (either the value of housing asset or the interest rate) with the only reason for premature termination of the contract being either prepayment or default (as in the cited article). Dunn and McConnell (1981) used interest rate evolution as the factor triggering financially optimal prepayment of the mortgage in their valuation model for mortgage pass-throughs. They also considered exogenous prepayment (due to the personal circumstances of a particular borrower) which, unlike endogenous prepayment, tended to have positive effect on the price of the contract. Such 'suboptimal' prepayment was introduced as a Poisson process whose mean rate of arrival was determined by an empirical prepayment model. However, except for commercial mortgages with prepayment penalties effectively prohibiting prepayment (the model for the latter with two sources of uncertainty - term structure and house prices - was developed in Titman and Torous (1989)), termination may take form either of prepayment or of default. In the structural model borrower chooses the less costly of the two depending on whether the face value of the loan F is greater or less than the current value of the collateral H . Although default on the fully insured mortgage may on the surface look like prepayment to the lender, it occurs in a quite different environment and is seldom to the lender's

advantage. Incorporation of both default and prepayment in a single model was accomplished by Kau, Keenan, Muller, and Epperson (1992), Schwartz and Torous (1992) as well as Leung and Sirmans (1990) (the latter used discrete-time setting). Downing, Jaffee, and Wallace (2010) analyze pool-level data using the geometric Brownian motion assumption for house price process, parameters of which they calibrate, along with a one-factor Cox, Ingersoll & Ross (CIR) (see Cox, Ingersoll, and Ross (1985)) term structure. They apply the PDE approach to solve for optimal default and prepayment boundaries and derive the values of default and prepayment hazards as well as the values of a 'background' hazard², which is assumed to be separate from the former two.

Models with both default and prepayment typically treat mortgage rates as an exogenous factor. Another strand of the structural literature (some authors term it "mortgage-rate based", as in Pliska (2006)) regards mortgage rate as an endogenous factor in the model. Stanton and Wallace (1998), Longstaff (2003), Dunn and Spatt (2005) focus on prepayment and analyze the influence of such factors as transaction costs, borrowers' mobility or credit constraints on the dynamics of the latter. Pliska (2006) and Goncharov (2006) provide theoretical results as to the existence of the solution to the problem of finding mortgage rate implied by a riskless yield curve and a given pattern of prepayment behavior for fixed-rate default-free mortgages (the latter paper deals exclusively with 'suboptimal' behavior). They compare structural and reduced-form approaches observing (Goncharov 2006) that obtaining similar result for the reduced-form setting would be a challenging task.

The present discussion of structural or option-based approach and its applications to mortgage valuation in particular has been brief and necessarily subjective. Many important contributions were not mentioned for the sake of brevity. It appears that the main drawbacks of structural models mentioned above, such as limited possibilities of incorporating available information about borrowers and economy or inability to

²The 'background' hazard function characterizes the conditional probability of termination for nonfinancial reasons, see Downing, Jaffee, and Wallace (2010, Section 3.3.1).

match market levels of yields, manifest themselves in mortgage applications as well as in corporate credit risk applications. This often hinders the ability of structural model to explain observed patterns of mortgage terminations, in particular prepayment³. The better performance of reduced-form models in explaining market spreads is among the reasons of the popularity of the latter. Seminal papers establishing the pricing framework in which the process for contract termination is determined exogenously include, among others, Jarrow and Turnbull (1995), Artzner and Delbaen (1995) Lando (1998), Madan and Unal (1998), Duffie and Singleton (1999). In these models the date of termination (say, by default) of the debt contract is not foreseeable.

Specifically, the termination time τ is a non-negative random variable with $Prob(\tau < \infty) = 1$, $Prob(\tau = 0) = 0$. In the Cox (doubly stochastic Poisson) process framework, this random variable is exponentially distributed. Let Y be a d -dimensional stochastic process (state process) on the space $(\Omega, \mathcal{F}, \mathbb{G}, \mathbb{Q})$ where $\mathbb{G} = (\mathcal{G}_t)_{0 \leq t \leq T^*}$, $\mathcal{G}_t \subset \mathcal{F}$ and the jump process $J_t = (\mathbf{1}_{[\tau \leq t]})$ is termination process, which generates the filtration $\mathbb{J} = (\mathcal{J}_t)_{0 \leq t \leq T^*}$, $\mathcal{J}_t = \sigma(\{\tau \leq u\} : u \leq t)$. T^* is finite time horizon in the model, such that $T^* > 0$, $T \leq T^*$. Augmented σ -algebra $\mathcal{F}_t = \mathcal{G}_t \vee \mathcal{J}_t = \sigma(\mathcal{G}_t, \mathcal{J}_t)$ is the maximum of these two; $\mathbb{F}$ is the corresponding filtration. Under appropriate technical conditions intensity can be represented for some function $\lambda : \mathbb{R}^d \mapsto \mathbb{R}_+$ as a function of the state vector $\lambda_t = \lambda(Y_t)$. τ is the time of the first jump of a Poisson process with intensity parameter λ^4 .

Let, further, a *short-rate* process $r_t = r(Y_t)$ be defined for some function $r : \mathbb{R}^d \mapsto \mathbb{R}$. Let $V^{(s)}(t, T)$ denote time t price of the contract that will survive till maturity T with certainty (e.g. default-free bond), then $V^{(s)}(t, T)$ is simply $V^{(s)}(t, T) = \exp\left(-\int_t^T r(u)du\right)$. Survival probability (which is formally defined for the general case in the next section) at any time period t given the realization of the intensity process $\lambda(Y_t)$ up to t is an expectation over

³It is acknowledged that nonfinancial considerations are often prominent in the decision to prepay a mortgage.

⁴See e.g. Cox and Isham (1980) for the concise discussion of the various generalization of the Poisson process. Daley and Vere-Jones (2003) is a widely cited reference on the theory of such processes.

all possible paths of the risk-neutral intensity process $\lambda_{\mathbb{Q}}(Y_t)$ ⁵:

$$\mathbb{Q}(\tau > t) = \mathbb{E}^{\mathbb{Q}} \exp \left\{ - \int_0^t \lambda(u) du \right\} \quad (2.6)$$

The price $V^{(d)}(\cdot)$ of a defaultable contract with the face value equal to unity and the payoff (assumed to be received at maturity) in case of default equal to $V(\tau)$, is given as the expected present value of payoff under all possible outcomes:

$$V^{(d)}(0, T) = \mathbb{E}^{\mathbb{Q}} \left(\exp \left\{ - \int_0^T r(u) du \right\} (\mathbf{1}_{[\tau > T]} + V(\tau) \mathbf{1}_{[\tau \leq T]}) \right) \quad (2.7)$$

or, alternatively, as the price of a default-free contract less the expected present value of the loss due to default

$$V^{(d)}(0, T) = V^{(s)}(0, T) - \mathbb{E}^{\mathbb{Q}} \left(e^{-\int_0^T r(u) du} (1 - V(\tau)) \mathbf{1}_{[\tau \leq T]} \right) = \quad (2.8)$$

$$V^{(s)}(0, T) - \mathbb{E}^{\mathbb{Q}} \left(e^{-\int_0^T r(u) du} \int_0^T (1 - V(u)) \lambda(u) e^{-\int_0^u \lambda(v) dv} du \right) \quad (2.9)$$

The question of mapping from the risk-neutral measure $\mathbb{Q}$ to the physical probability measure $\mathbb{P}$ (correspondingly, from the risk-neutral intensity $\lambda_{\mathbb{Q}}$ to the real-world intensity $\lambda_{\mathbb{P}}$) is an empirical one. As Singleton (2006) points out, not only concurrent levels of $\lambda_{\mathbb{Q}}$ and $\lambda_{\mathbb{P}}$ may differ, but also the characteristics of the processes, in general, may be quite divergent (they may not be, for example, both of a same type, e.g., diffusions).

There are several assumptions about the recovery process which are common in the reduced-form models. One particular recovery formulation is often called recovery of treasury (RT) (Kijima 2003), since it can be viewed as replacement of the bond in default with a treasury bond of a reduced value with the same maturity; an alternative term for it is the fractional recovery of no-default value. This recovery assumption is natural for the long-term bonds, holders of which are not protected against default by an acceleration clause. Another recovery formulation – often termed the (fractional) recovery of face value (RFV) (see, for instance, Singleton (2006)) – is closely related to the legal practice. Under this assumption, creditor receives at default smaller of the fraction of the face value

⁵See, e.g., Lando (2004, Ch. 5).

that remains after liquidation of the assets, or the face value of the debt. This assumption is most appropriate for the case of mortgage foreclosure and for covenant-protected corporate bond defaults. Unlike RFV, RT implies that holders of coupon bonds are entitled to some fraction of future coupon values, too. Under (fractional) recovery of market value (RMV) formulation, an investor receives at the time of default a fraction of the market value of the debt as of the moment just prior to the default. This convention has an advantage of having clear economic interpretation as the loss in value associated with default.

Although reduced-form models do not provide links between economic fundamentals characterizing an issuer and the probability of premature termination, it is, however, possible to introduce into the model state variables, which may or may not have straightforward economic interpretation. The general formula for price process of a defaultable claim in reduced-form framework with state variable vector Y , $\mathcal{G}_T$ -measurable payoff process V and $\mathbb{G}$ -predictable recovery process X , which determines the rate of payoff, is given below. It encompasses all aforementioned recovery assumptions⁶:

$$V^{(d)}(t, T) = \mathbf{1}_{[\tau > t]} \mathbb{E}^{\mathbb{Q}} \left(\int_{(t, T]} \exp \left\{ - \int_t^u r(Y_v) + \lambda(Y_v) dv \right\} \lambda(Y_u) X_u du \mid \mathcal{G}_t \right) + \mathbf{1}_{[\tau > t]} \mathbb{E}^{\mathbb{Q}} \left(\exp \left\{ - \int_t^T r(Y_v) + \lambda(Y_v) dv \right\} V \mid \mathcal{G}_t \right) \quad (2.10)$$

In many cases, state variables employed in reduced-form model do not admit clear structural interpretation. For example, in the study of corporate spreads (Duffee 1999), dynamics of prices of default-free zero-coupon bonds depends on two uncorrelated state variables, whereas instantaneous credit spread is the function of three independent state variables (two aforementioned term structure state variables and an additional spread-specific one). All three processes are latent. Some researchers attempt to bring to the reduced-form models certain 'flavor' of a structural ones by introducing such state variables as, for example, leverage ratio of the firm (Bakshi, Madan, and Zhang 2006) but,

⁶See, e.g. Bingham and Kiesel (2004, Proposition 9.4.1).

for the most part, underlying state variables are not easily interpretable in terms of economic fundamentals. That did not, however, preclude widespread use of reduced-form models in bond pricing, where for many types of contracts there is only one possibility of premature termination of debt contract – namely, by default.

This is not the case for callable bonds, residential mortgages and derivatives backed by these contracts. Here the second way of termination – termination by prepayment – plays an important role. For mortgages prepayment is not a rare event and, while bulk of prepayments are to the lender's disadvantage, there is significant number of 'non-financial' prepayments, which thereby partially compensate losses associated with financially optimal prepayments. Importantly, prepayment may act as a competing alternative to default and this has to be reflected in the estimation. This was acknowledged in Kau, Keenan, and Smurov (2004), where intensities of default and prepayment were modeled as mean-reverting uncorrelated scalar diffusion processes. Term structure (modeled as two factor Cox-Ingersoll-Ross one) entered the model via time varying spread between contract rate at origination ⁷ and actual yield on 10-year Treasury bond. Pricing procedure amounted to Monte Carlo integration of discounted payoffs that incorporated simulated paths of interest rate and latent prepayment and default factors. Calibration of such parameters as multiplicative and additive risk adjustments, liquidity adjustment and loss rate in case of default (under the RFV assumption) to the market data was required in order to apply risk-neutral valuation procedure. Implementation of such forward-looking pricing procedure allowed authors to decompose the value of a representative mortgage into components reflecting market assessment of risk according to the observed prepayment and default behavior of borrowers.

Most of other contributors in this field limited their task to modeling only one of termination processes. DeGiorgi and Burkhard (2006) estimated conditional intensity of default including macroeconomic, locational variables, time since origination, seasonal

⁷The model of Kau, Keenan, and Smurov (2004) was applied to valuation of fixed-rate mortgages.

dummies and mortgage rate in the set of covariates. They used Swiss data on more than 70000 individual fixed-rate and adjustable-rate mortgages observed in 1994-2000. The log-additive model for conditional default intensity was estimated non-parametrically (smoothing splines were used to describe the contribution of each predictor). Kolbe and Zagst (2008) used doubly stochastic framework to model prepayment. They chose Ornstein-Uhlenbeck dynamics for their instantaneous short rate process. For the baseline hazard process, along with latent baseline factor, they included the second state variable - quarterly GDP growth rate - which entered additively into the drift term of the OU diffusion. Kolbe and Zagst estimated parameters of their model using pool-level GNMA data. They employed nonlinear function of spread (between the weighted-average coupon (WAC) of the mortgage pool and the 10-year treasury par yield) and linear and cubic terms in burnout variable as the covariates in the empirical estimation of the baseline. Their tests failed to reject one-factor parameterization of baseline hazard, but they reported that the two-factor model for baseline prepayment provided somewhat better fit. Liao, Tsai, and Chiang (2008) developed a reduced-form model for mortgage valuation with yields explicit solutions; in their model the hazard rates of prepayment and default are assumed to be the linear functions of the respective baseline hazards and other relevant state variables. The set of state variables includes interest rates house price process and household income process. Using geometric Brownian motion specification for the latter two, extended Vasicek model for the interest rate and the assumption of a deterministic loss rate they perform the sensitivity analysis of price to the volatilities of state variables, correlations between them and other parameters of their model.

Outside of the field of real-estate finance, but in closely related work Jarrow, Li, Liu, and Wu (2006) consider valuation of callable bonds: in their study call option on a bond is modeled via intensity process, thus the model of Jarrow *et al.* much like the model of Kau, Keenan, and Smurov (2004) has four stochastic processes: two for the term structure, one

for default intensity and one for call intensity; ML estimation of the model parameters is based application of extended Kalman filter.

Extant literature on empirical reduced-form valuation⁸ of mortgages and MBS is relatively scant, but there is an ongoing research effort stimulated by the popularity of reduced-form models in the credit risk literature.

2.2 NO-ARBITRAGE PRICING MODEL FOR THE MORTGAGE CONTRACT SUBJECT TO DEFAULT AND PREPAYMENT

General pricing formula given by (2.9) should be reconciled with the fact that termination by default is observed only on the payment dates, i.e. in monthly intervals, rather than completely at random. For the purpose of estimation convenience in our model it is assumed that prepayment also follows this pattern. The formulation of the latent termination process in the present work follows Belanger, Shreve, and Wong (2004), who generalize the notion of a hazard process to the situations when the event of interest (e. g. default) occurs only at certain dates (payment dates). One may think of a non-decreasing, $\{\mathcal{F}_t\}$ -predictable process Γ_t ⁹ which has right continuous paths with left limits (RCLL), $\Gamma_0 = 0$ a.s.. The continuous Brownian filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T^*}$ can be thought of as flow of publicly available information about the borrower. Probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ also supports positive random variable Ξ which is independent of $\mathcal{F}_{T^*}$. Termination time τ is

⁸By 'reduced-form' valuation, I mean here reduced-form models with stochastic state variables. More general rubric, sometimes called empirical or 'econometric' approach, includes not only the aforementioned intensity-based, reduced-form models but also simpler statistical models without stochastic dynamics of state variables. Reduced-form models in this broader sense have been applied to the pricing of mortgage contracts and their derivatives for quite a while. In industry such models have been found to be particularly suitable for modeling prepayment (Patruno (1994)). Schwartz and Torous (1989) is one of the early examples of published work on empirical mortgage-backed securities valuation. Deng, Quigley, and Van Order (2000) provides an example of an elaborate hazard model developed for the assessment of the risk of mortgage termination. Since this literature is abundant, I only touch on some of the contributions in this field while discussing model specification.

⁹Some authors, e.g. Belanger, Shreve, and Wong (2004), reserve the term *hazard process* for the process Λ_t defined below, others use this term for Γ_t , while Λ_t is referred to as *martingale hazard process*.

defined as $\tau = \inf \{t \in [0, T^*]; \Xi \leq \Gamma(t)\}$. Survival process S_t can be defined via cumulative distribution function¹⁰ of Γ as

$$S(t) = 1 - F(\Gamma_t) = \mathbb{Q}(\tau > t | \mathcal{F}_{T^*}) = \mathbb{Q}(\tau > t | \mathcal{F}_t) \quad 0 \leq t \leq T^* \quad (2.11)$$

(Martingale) hazard process Λ_t is also non-decreasing RCLL process¹¹ defined as

$$\Lambda_t \equiv - \int_{(0,t]} \mathbf{1}_{[S_{u-} > 0]} \frac{dS_u}{S_{u-}} \quad 0 \leq t \leq T^* \quad (2.12)$$

$$= \sum_{0 < t_n \leq t} \lambda_{t_n} \quad (2.13)$$

where $S_{u-} \equiv \lim_{v \uparrow u} S_v$ and $S_{0-} \equiv 1$, $t_n : \mathbf{1}_{[\tau > t_n]} = 1$.¹² Next, consider the *compensator* process $A_t^\tau \equiv \Lambda(t \wedge \tau)$ of the jump process $J_t = \mathbf{1}_{[\tau \leq t]}$. This process represents cumulative conditional likelihood of default or *default trend* (respectively, *prepayment trend*). Doob-Meyer decomposition of the counting process N_t (which is a pure jump process with jumps of size one) yields $M_t = N_t - A_t$ where M_t is a (local) martingale. Intuitively, the compensator for the counting process is a foreseeable part of the process based on observations prior to time t : $dA_t = \mathbb{E}(dN_t | \mathcal{F}_{t-})$, and dM_t is the part that cannot be foreseen. Realizations of the two hazard processes are observed discretely and they are parameterized (at the n^{th} payment date) as:

$$\lambda_{t_n, \mathbb{P}}^{(\ell)} = \exp(\mathbf{z}(t_n) \boldsymbol{\beta}^{(\ell)}) l_{t_n, \mathbb{P}}^{(\ell)} \quad \ell = def, prep \quad (2.14)$$

where $l_t^{(\ell)}$ are latent baseline hazard process driven by independent Brownian motions. Estimation of the parameters $\boldsymbol{\beta}$ of the multiplicative intensity model given the values of observed variables $\mathbf{z}$ is the subject of the next chapter. Importantly, the Brownian motions in (2.14) admit physical probability measure $\mathbb{P}$. Using standard no-arbitrage argument, the price of a mortgage contract¹³ can be represented as (cf Belanger, Shreve, and Wong

¹⁰ $F(\xi) = \mathbb{Q}(\Xi \leq \xi)$ and using the fact that $\{\tau \leq t\}$ is equivalent to $\{\Xi \leq \Gamma_t\}$.

¹¹ Martingale hazard Λ_t coincides with hazard Γ_t iff c.d.f of termination time τ is continuous, cf Bielecki and Rutkowski (2002), Proposition 4.5.1. Then Λ_t is continuous.

¹² For $F(\xi) = 1 - \exp(\xi)$ survival $S_t > 0 \forall t$ and indicators are redundant. On random interval $[0, \tau)$ $\Gamma_t = -\log(S_t)$ and $S_t = \exp(-\Lambda_t^c) \prod_{0 < s \leq t} (1 - \Delta \Lambda_s)$ where Λ^c is continuous part of Λ and $\Delta \Lambda_s = \Lambda_s - \Lambda_{s-}$.

¹³ Under a martingale measure $\mathbb{Q}$ (such that M_t is $\mathbb{F}$ -martingale) and using the assumption of RFV.

(2004), Proposition 3.2):

$$V(t, T) = \mathbf{1}_{[\tau > t]} \mathbb{E}^{\mathbb{Q}} \left(\delta C \int_{(t, T]} e^{-\int_t^u (r_s + \lambda_s) ds} \lambda_u du | \mathcal{G}_t \right) + \mathbf{1}_{[\tau \leq t]} \mathbb{E}^{\mathbb{Q}} \left(e^{-\int_t^T (r_s + \lambda_s) ds} C | \mathcal{G}_t \right) \quad (2.15)$$

where δ is the constant recovery rate, C is contractually specified payoff of the mortgage contract and filtrations $\mathbb{F}$ and $\mathbb{G}$ are defined above (see (2.10)). Martingale hazard process Λ_t is an aggregate one over the two kinds of termination, since termination time $\tau = \tau^{(d)} \wedge \tau^{(p)}$ (cf Lemma 7.1.2. in Bielecki and Rutkowski (2002)).

$$\Lambda_t = \sum_{0 < t_n \leq t} \lambda_{t_n}^{(d)} + \sum_{0 < t_n \leq t} \lambda_{t_n}^{(p)} \quad (2.16)$$

After taking into account adjustments for taxes and illiquidity (Duffie and Singleton 1999), the pricing formula for value of mortgage at the time of origination¹⁴ can be written as:

$$V(0, T) = \mathbb{E}^{\mathbb{Q}} \left(\delta C \int_0^T e^{-\int_0^u ((1-\text{tax})r_s + \text{lq} + \lambda_s) ds} \lambda_u du | \mathcal{G}_t \right) + \mathbb{E}^{\mathbb{Q}} \left(e^{-\int_0^T ((1-\text{tax})r_s + \text{lq} + \lambda_s) ds} C | \mathcal{G}_t \right) \quad (2.17)$$

Next four sections are devoted to the presentation of the approaches to specification and parameter estimation of the components of the formula (2.17). I begin with review of model specification and parameter estimation of β in (2.14).

2.3 MULTIPLICATIVE INTENSITY MODEL WITH COMPETING RISKS OF PREPAYMENT AND DEFAULT

Let T represent the length of time (duration) that individual mortgagor holds the mortgage contract. It is naturally to think of T as a continuous random variable in which case the survivor function (2.11) is simply

$$S(t) = 1 - F(t) \quad (2.18)$$

¹⁴The value at origination should equal amount of loaned funds to prevent arbitrage.

where $F(t)$ is the cumulative distribution function¹⁵ of T . The hazard function is defined as the instantaneous probability of a failure (termination of the contract either by prepayment or by default) conditional on survival:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{Pr(t \leq T < t + \Delta t) | T \geq t}{\Delta t} = \frac{f(t)}{S(t)} \quad (2.19)$$

where $f(t) = \frac{dF(t)}{dt}$ is the density function¹⁶ of T assuming strictly positive values. However, a rational borrower is expected to abrogate the contract only on payment dates, since this allows such borrower to enjoy an extra period of unpaid housing services. Assuming that hazard within the interval between successive payment dates is constant, the discrete version of hazard function obtains as

$$\lambda_j^{dis}(t) = \frac{f_j^{dis}}{f_j^{dis} + f_{j+1}^{dis} + f_{j+2}^{dis} + \dots} = \frac{f_j^{dis}(t)}{S_j^{dis}(t)} \quad (2.20)$$

where the density f_j^{dis} in the discrete case can be represented as

$$\begin{aligned} f_j^{dis} &= \sum_{i \in I} f_i^{dis} \mathbf{1}\{a_j \geq a_i\} - F^{dis}(a_{j-}) = \\ F^{dis}(a_j) - F^{dis}(a_{j-}) &= F^{dis}(a_j) - F^{dis}(a_{j-1}) = \\ S^{dis}(a_j) - S^{dis}(a_{j+1}) \end{aligned} \quad (2.21)$$

In the foregoing equation survivor function $S^{dis}(t)$ is defined as

$$S^{dis}(t) = \sum_{j \in J} f_j^{dis} \mathbf{1}\{t \leq a_j\} = \sum_{[j|a_j \geq t]} f_j^{dis} \quad (2.22)$$

with usual condition on probability masses: $\sum_{j \in J} f_j^{dis} = 1$. Duration variable T takes discrete values $0 \leq a_j \leq a_{j+1}$ and $a_{j-} = \lim_{t \rightarrow j} a_t$. Hence discrete-time hazard function can be written as $\lambda_j^{dis} = \frac{f_j^{dis}(t_j)}{S_j^{dis}(t_{j-})}$. Discrete-time survivor function can be expressed via discrete-time hazard function as

$$S^{dis}(t) = \prod_{j|t_j \leq t} (1 - \lambda_j^{dis}) \quad (2.23)$$

¹⁵Alternative definition of survivor function preferred by some authors is $S(t) = Pr(T > t)$; in the discrete case one should be careful to distinguish between the two.

¹⁶Cf specialization to the discrete case in (2.13).

While continuous survivor function is monotonically non-increasing on the interval from 1 to 0 (assuming that eventually all subjects will fail, i.e. assuming non-defective distribution of the duration variable), the discrete analog is a decreasing step function with steps occurring at event times t_j (payment dates). Cumulative hazard function for discrete-time models takes the form $\Lambda^{dis}(t) = \sum_{j|t_j \leq t} \lambda_j^{dis}$ familiar by (2.13)¹⁷. Generally speaking, the distribution of T can have both discrete and continuous components, in which case cumulative hazard function consists of discrete and continuous terms¹⁸:

$$\Lambda(t) = \sum_{j|t_j \leq t} \lambda_j^{dis} + \int_0^t \lambda(u) du. \quad (2.24)$$

As it is common with duration data, the data set under consideration is subject to censoring. First, not all the mortgages have prepaid, defaulted or have been completely paid off during the observation period. That means that our model have to account for right censoring of the duration data. Secondly, payment histories for the mortgages of the earliest vintages starts several years after the origination date, therefore, there is a problem of left censoring for the part of the data set. An assumption that censoring process (denoted by $Y(t) : Y_i(t) = 1$ if individual i is in the risk set at time t_- and 0 otherwise) is independent of the data-generating process for duration simplify matters. This standard assumption doesn't look implausible in our case. Since the parameters of the distribution of Y do not provide additional information about the parameters of the distribution of T , there is no need to model the censoring mechanism explicitly. Even if distribution of censoring process may be related to the data-generating process that generates the regressors $\mathbf{z}$, still the non-informative property allows one to concentrate on modeling of duration as a function of covariates and baseline hazard, but not the censoring process. More specifically, let the $\mathbb{Y} : (\mathcal{Y}_t)_{\{0 \leq t \leq T^*\}}$ be the filtration generated jointly by the stochastic data-generating processes of covariates and censoring: $\mathcal{Y}_t = \sigma(\mathbf{z}_i, Y_i(s), i \in N : s \leq t)$. Let

¹⁷In continuous time summation over $y_j s$ is replaced by the integral over the range of t . In this latter case is can be easily seen that cumulative hazard equals negative of $\log(S(t))$.

¹⁸Cf footnote commenting (2.13) in the previous section.

$\mathbb{C}$ denote the filtration generated by $\mathbb{Y}$ and counting process $N(t)$ so that $\mathcal{C} = \mathcal{Y} \vee \mathcal{N}$. Then for the case of two competing risks with corresponding filtrations $\mathbb{C}^{def}$ and $\mathbb{C}^{pre}$, if one assumes independency between competing risks, duration will be determined by the information in the minimal filtration generated by the sigma-algebra: $\mathcal{C} = \mathcal{C}^{def} \vee \mathcal{C}^{pre}$. Assuming that covariates enter the hazard multiplicatively, i.e. hazard function has the following form

$$\lambda(t|\mathbf{z}(t), \beta) = l(t|\mathbf{z}(t), \beta) \exp(\mathbf{z}'(t)\beta) \quad (2.25)$$

it has been shown (Cox and Oakes (1984), Kalbfleisch and Prentice (2002)) that factoring complete likelihood into parts, one of which depends only on β and one – both on β and baseline hazard $l(t|\cdot)$, is permissible. This simplifies inference on β , since now the latter can be based only on the partial likelihood, which is a function of β only. For the case of two competing risks of default and prepayment, denoted by corresponding subscripts, general form of the partial likelihood is

$$L(\beta_d, \beta_p) = \prod_{i=1}^{k_d} \frac{\exp\{\mathbf{z}'_{d,i}(t_{d,i})\beta_d\}}{\sum_{s \in R(t_{d,i})} \exp\{\mathbf{z}'_s(t_{d,i})\beta_d\}} \cdot \prod_{i=1}^{k_p} \frac{\exp\{\mathbf{z}'_{p,i}(t_{p,i})\beta_p\}}{\sum_{s \in R(t_{p,i})} \exp\{\mathbf{z}'_s(t_{p,i})\beta_p\}} = L_d(\beta_d) \cdot L_p(\beta_p) \quad (2.26)$$

where $R(t_{\ell,i}) = \{s : t_s \geq t_{\ell,i}\}$ is set of $i = 1, \dots, I$ subjects (contracts) which are at risk of ℓ^{th} type of termination at time $t_{\ell,i}$, $t_{\ell,1} < \dots < t_{\ell,k_\ell}$ are times of type ℓ termination¹⁹. Thus a likelihood function for two independent competing risks factors out into two separate likelihood functions for each of the risks. I will suppress reference to a particular risk in what follows and leave only subscripts denoting subject or set of the subjects to simplify notation. Maximum partial likelihood estimator of β obtains by setting the score function U defined as

$$U(\beta) = \frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{i=1}^N \delta_i \left(\mathbf{z}_i(t_i) - \frac{\sum_{s \in R(t_i)} \mathbf{z}_s(t_i) \exp\{\mathbf{z}'_s(t_i)\beta\}}{\sum_{s \in R(t_i)} \exp\{\mathbf{z}'_s(t_i)\beta\}} \right) \quad (2.27)$$

¹⁹Cf Kalbfleisch and Prentice (2002, p.255, (8.13)). It is assumed that only a single termination at a time may happen. For the various statistical methods of dealing with tied failure times see Andersen, Borgan, Gill, and Keiding (1993) or Aalen, Borgan, and Gjessing (2008).

equal to 0 by usual maximum likelihood arguments. Here δ_i is indicator taking value of 0 if observation i was censored, N is sample size. Although partial likelihood estimator $\hat{\beta}$ is neither a marginal nor a conditional likelihood estimator but rather a product of limited information likelihood procedure, this estimator is consistent for β and asymptotically normal for large samples. For a discrete-time model (2.25) takes the form

$$\lambda(t|\mathbf{z}(t), \beta) = l(t|\mathbf{z}(t), \beta) \exp(\mathbf{z}'(t_{a-1})\beta) \quad (2.28)$$

for $t \in [t_{a-1}, t_a]$ if covariate values remain constant within this interval. Note that there are no restrictions on the behavior of baseline hazard (e.g. constancy) within the interval. Thus we can obtain estimates for the realizations of continuous-time baseline intensity process at the endpoints of one-month interval. Kalbfleisch and Prentice (2002) show that estimator (2.27) is still unbiased, consistent and asymptotically normal for the discrete case whereas the unbiased estimate of variance is given by

$$V_u(\beta) = \sum_{j=1}^k \delta_j \mathcal{V}(\beta, t_j) - \sum_{j=1}^k \sum_{s \in R(t_j)} (\mathbf{z}_s(t_j) - \mathcal{E}(\beta, t_j))^{\otimes 2} \exp[2 \mathbf{z}'_s(t_j)\beta] \hat{\alpha}_0(\beta, t_j) \quad (2.29)$$

$$\text{where } \hat{\alpha}_0(\beta, t_j) = \frac{\delta_j(\delta_j - 1)}{\left(\sum_{s \in R(t_j)} \exp[\mathbf{z}'_s(t_j)\beta] \right)^2 - \sum_{s \in R(t_j)} \exp[2 \mathbf{z}'_s(t_j)\beta]},$$

$$\mathcal{V}(\beta, t_j) = \sum_{i=1}^N (\mathbf{z}_i(t_j) - \mathcal{E}(\beta, t_j)) (\mathbf{z}_i(t_j) - \mathcal{E}(\beta, t_j))' p_i(t_j)$$

$$\mathcal{E}(\beta, t_j) = \sum_{i=1}^N \mathbf{z}_i(t_j) p_i(t_j),$$

$$p_i(\beta, a_j) = \frac{\exp[\mathbf{z}'_i(a_j)\beta]}{\sum_{s \in R(a_j)} \exp[\mathbf{z}'_s(a_j)\beta]}.$$

This estimate is used to form sandwich-type estimator of asymptotic covariance of $(\hat{\beta} - \beta) : \left(\sum_{j=1}^k \delta_j \mathcal{V}(\hat{\beta}, t_j) \right)^{-1} V_u(\hat{\beta}) \left(\sum_{j=1}^k \delta_j \mathcal{V}(\hat{\beta}, t_j) \right)^{-1}$.

It has been assumed throughout that dependence between the risks of prepayment and default is only through common set of covariates entering both of the hazards. Kau, Keenan, Muller, and Epperson (1995), Deng, Quigley, and Van Order (2000) among others

pointed out to the importance of modeling prepayment and default jointly in the valuation of residential mortgages. Dependence between risks can be introduced in different ways. One popular choice is to parameterize the transformation of duration variable for a particular risk $-\log(\Lambda_0(t; \alpha))$ as $\mathbf{z}'(t)\beta + \nu + \epsilon$, where parameter ν can be either common between different risks or ν_j may be correlated. However, this modeling scheme comes at significant additional computational cost. The likelihood for the model with correlated competing risks and discrete data is no longer simply a product of the two individual likelihoods. Additional assumptions about the behavior of hazard or timing of the event of interest within the interval between the observations are necessary to achieve the identification. It is often assumed that either hazard rate is constant within the interval or density is constant, although other choices have been also analyzed in the literature²⁰. Typically in the above parameterization parameter ν represents unobserved heterogeneity in the population with regard to the certain type of hazard. Latent durations T^p, T^d are assumed to be conditionally independent, given these unobserved heterogeneity components. In such *mixed proportional hazard model* (ν^p, ν^d) may be modeled via continuous multivariate joint distribution, or the mixture can be discrete. In this latter case the likelihood contribution of an individual observation is

$$L^*(\beta|\nu) = \sum_{k=1}^K \pi_k L_k(\beta|\nu_{prep}, \nu_{def}) \quad (2.30)$$

where π_k is probability associated with the mass point k , $\sum_{k=1}^K \pi_k = 1$ and $L_k(\beta|\nu_{prep}, \nu_{def})$ is likelihood contribution for particular combination of mass points. That is the approach which was advocated by Heckman and Singer (1984), among others. Abbring (2001) and others analyzed the conditions for identification of the model (see, e.g. Abbring and Van Den Berg (2007) for a discussion of necessary conditions). These conditions are not too

²⁰Jenkins (2005) provides a survey of different approaches to modeling mixed hazard with correlated risks.

restrictive, however the estimator, which is the solution to the optimization problem

$$\arg \max_{\beta, \nu, \pi} \left\{ \sum_{i=1}^N \log(L(y_i | \mathbf{z}_i; \beta, \nu, \pi)) \right\} \quad (2.31)$$

is notorious for numerical problems arising in the estimation in the extensive data set. In my attempts to estimate the mixed proportional hazard model with different number of support points for heterogeneity distribution I found that often the distribution of heterogeneity was estimated to be degenerate, i.e. over 99% of the probability is concentrated in a single masspoint (the results of MPH estimation are not reported here). Potentially feasible approach for modeling competing risks dependency via parametric copula, however, is unlikely to suit the particular case of default and prepayment, since the correlation between these risks is negative.

2.4 CHARACTERIZATION OF THE CIR TERM STRUCTURE MODEL

A definition of short (spot) rate seems a natural way to start the discussion of the term structure. Denote by B a *money account* the process with continuous sample path and finite variation, such that $B_0 = 1$. The short rate r_t is an adapted process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with almost all paths integrable on $[0, T]$, such that it solves the differential equation $dB_t = r_t B_t dt$, where $B_t = B_t(\omega)$ for almost all $\omega \in \Omega$, or, equivalently:

$$B_t = \exp \left(\int_0^t r_u du \right), \quad \forall t \in [0, T] \quad (2.32)$$

In the arbitrage-free market prices at time t of zero-coupon bonds maturing at time T are given by:

$$p(t, T) = \mathbb{E}_t^{\mathbb{Q}} \left(\exp^{-\int_t^T r_u du} \right) \quad (2.33)$$

and the expectation above (see, e.g. Björk (2004)) implies the dynamics of the short rate under the measure $\mathbb{Q}$ of the following form

$$dr_s = (\mu - \alpha\sigma)ds + \sigma dW_s \quad (2.34)$$

$$r_t = r$$

where α , μ and σ depend on time and short rate. The process α is termed *market price of risk* and is defined as $\alpha_t = \frac{\rho_t^T - r_t}{\sigma_t^T}$, where ρ_t^T is the local return at time t on the bond with maturity T , σ_t^T is the local volatility and the difference $\rho_t^T - r_t$ represents the risk premium on the bond. The measure $\mathbb{Q}$ is such that for any maturity $T^* \in [0, T]$ the relative zero-coupon bond price process $\frac{p(t, T^*)}{B_t}$, $\forall t \in [0, T^*]$ is a martingale, therefore it is termed (spot) *martingale measure*. One obtains risk-neutral expectation hypothesis or local expectation hypothesis if expectation with respect to martingale measure in (2.33) is replaced with expectation with respect to objective measure $\mathbb{P}$. The term 'local expectation' can be interpreted so that the bond price at any moment s is equal to the expected value (under actual probability $\mathbb{P}$) of the bond price in the next period (while the length of that period tends to 0) discounted at spot rate at s . The family of bond prices (which are assumed to be arbitrage free relative to r) is often referred to as *term structure*. One approach to modeling term structure is to assume that short rate follows Itô process, or, more specifically, a diffusion process of the form:

$$dr(t) = \mu(t, r(t)) dt + \sigma(t, r(t)) d\bar{W}(t) \quad (2.35)$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ in 2.35 are the same functions as in 2.34, however, the underlying probability measure is now the objective probability $\mathbb{P}$. Measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent probability measures on a filtrations of a Brownian motion. It is known (Duffie (2001, Appendix A)) that if $\mathbb{Q}$ is equivalent to $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F}_T)$, then probability measure $\mathbb{Q}$ is a Radon-Nikodym derivative:

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp \left(\int_0^T \alpha_u dW_u - \frac{1}{2} \int_0^T |\alpha_u|^2 du \right) = \eta_T, \quad \mathbb{P}\text{-a.s.} \quad (2.36)$$

where random variable η_T is a.s. strictly positive and its $\mathbb{P}$ -expected value integrates to 1. In other words,

$$dW_t = d\bar{W}_t + \int_0^t \alpha_u du \quad (2.37)$$

It follows that different choices of market price of risk α imply different martingale measures $\mathbb{Q}$. However, one should execute caution in the choice of α , so that the latter will be compatible with aggregate level of risk aversion in the market.

In general the drift and volatility functions $\mu(\cdot)$ and $\sigma(\cdot)$ can be arbitrarily complex. However, for the term structure model to be tractable, certain restrictions are often imposed. A widely used class of models is one in which the price of a zero-coupon bond can be represented as:

$$p(t, T) = \exp [A(t, T) - B(t, T) r] \quad (2.38)$$

for some deterministic functions A and B . In this case the model is said to belong to *affine* or *exponential affine* class. In affine models (see, e.g., Duffie and Kan (1996)) the dynamics for the short rate, given by SDE 2.35 imply the following form of drift and diffusion :

$$\begin{aligned} \mu(t, r) &= a(t) r + b(t) \\ \sigma(t, r) &= \sqrt{d(t) r + e(t)} \end{aligned} \quad (2.39)$$

where $a(\cdot)$, $b(\cdot)$, $d(\cdot)$ and $e(\cdot)$ are deterministic functions satisfying the following differential equations:

$$\begin{cases} \frac{\partial}{\partial t} B(t, T) + a(t) B(t, T) - \frac{1}{2} d(t) B^2(t, T) = -1 \\ B(T, T) = 0 \end{cases} \quad (2.40)$$

and

$$\begin{cases} \frac{\partial}{\partial t} A(t, T) = b(t) B(t, T) - \frac{1}{2} e(t) B^2(t, T) \\ A(T, T) = 0 \end{cases} \quad (2.41)$$

(cf Björk (2004, (22.24-22.25))), for which there are closed-form solutions available in several cases. Cox, Ingersoll, and Ross (1985) developed the general equilibrium model in

which the dynamics of short rate follows the following SDE:

$$dr(t) = \kappa (\theta - r(t)) dt + \sigma \sqrt{r(t)} d\bar{W}(t) \quad (2.42)$$

under the martingale measure $\mathbb{Q}$. A condition $\kappa\theta > \frac{\sigma^2}{2}$ ensures that the process stays non-negative. The change of measure is given by (cf Brigo and Mercurio (2006, p. 65):

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp \left(-\frac{1}{2} \int_0^t \alpha^2 r(u) du + \int_0^t \alpha \sqrt{r(u)} dW(u) \right) \quad (2.43)$$

Hence market price of risk exhibits increased volatility in the periods when interest rates are more volatile. Now, denoting $A_0(t, T) = \log(A(t, T))$, the term structure of zero-coupon bonds in the CIR model is given by :

$$\begin{aligned} A_0(t, T) &= \left[\frac{2h \exp((\kappa + h)(T - t)/2)}{2h + (\kappa + h)\{\exp((T - t)h) - 1\}} \right]^{2\kappa\theta/\sigma^2} \\ B(t, T) &= \frac{2(\exp((T - t)h) - 1)}{2h + (\kappa + h)\{\exp((T - t)h) - 1\}} \\ h &= \sqrt{\kappa^2 + 2\sigma^2} \end{aligned} \quad (2.44)$$

under the risk-neutral probability. Price of a zero-coupon-bond obtains as :

$$p(t, T) = A_0(t, T) \exp(-B(t, T)r(t)) \quad (2.45)$$

Chen and Scott (1992) suggested multi-factor extension of CIR model, where each of several independent latent factors y_j follows the square-root diffusion 2.42, and short rate is the sum of these factors: $r = \sum_{j=1}^J y_j$. Under these assumptions the term structure equations (under the objective probability) have the same form as above (now including market price of risk factor):

$$\begin{aligned} A_{0j}(t, T) &= \left[\frac{2h_j \exp((\kappa_j + h_j + \alpha_j)(T - t)/2)}{2h_j + (\kappa_j + h_j + \alpha_j)\{\exp((T - t)h_j) - 1\}} \right]^{2\kappa_j\theta_j/\sigma_j^2} \\ B_j(t, T) &= \frac{2(\exp((T - t)h_j) - 1)}{2h_j + (\kappa_j + \alpha_j + h_j)\{\exp((T - t)h_j) - 1\}} \\ h_j &= \sqrt{(\kappa_j + \alpha_j)^2 + 2\sigma_j^2} \end{aligned} \quad (2.46)$$

and bond price is given by(cf Cox, Ingersoll, and Ross (1985, (23))):

$$p(t, T) = \left(\prod_{j=1}^J A_{0j}(t, T) \exp \left(- \sum_{j=1}^J B_j(t, T) y_j(t) \right) \right) \quad (2.47)$$

In the next section I discuss the estimation of the two-factor version of this model using the data set of yields on zero-coupon bonds of different maturities.

2.5 EXISTING APPROACHES TO HOUSE PRICE MODELING

Several strands of literature are relevant to the analysis of the properties of the house process. In particular, many recent partial equilibrium models of portfolio choice, e.g., (Flavin and Yamashita (2002) or Cocco (2005)) have incorporated real estate asset. Often-times these studies are cast in one-period framework (while many allow for a bequest motivation). Return on holding housing may be represented as a sum of an expected return and a stochastic component. In an intertemporal utility model with housing good (H) and a numeraire good (C), lifetime utility of a representative agent (household) is given by:

$$U \equiv \int_0^\infty \exp(\beta t) u(H(t), C(t)) dt \quad (2.48)$$

where β is a discount rate, with a budget constraint

$$g(t) X(t) + S(t) + C(t) = (1 - \tau) Y_r(t) + (1 - \tau) r(t) F(t) dt \quad (2.49)$$

where $g(t)$ is the price of a unit of housing stock, $X(t)$ is the amount of units purchased, $S(t)$ are (net) real savings, $Y_r(t)$ is the real income of the household, τ is the marginal tax rate of the household, $r(t)$ is market interest rate. Additional constraint

$$\frac{\partial H(t)}{\partial t} = X(t) - \delta H(t) \quad (2.50)$$

where δ is the depreciation rate, describes the dynamics of housing stock owned by a household. Household in the model is allowed to invest in the financial asset as well as in the housing. The last constraint:

$$\frac{\partial F(t)}{\partial t} = S(t) - \pi F(t) \quad (2.51)$$

where π is the rate of inflation and $F(t)$ is the value of net assets other than housing, governs the dynamics of financial assets. If, as is often assumed, credit rationing with shadow price $\xi(t)$ is present, then usual first-order conditions and absence of arbitrage imply (see Meen (2001)):

$$g(t) = s(t) / \left((1 - \tau) r(t) - \pi + \delta - \frac{\partial g}{\partial t} \frac{1}{g(t)} + \frac{\xi(t)}{\frac{\partial u}{\partial C}} \right) \quad (2.52)$$

where $s(t)$ is the imputed rent (price of housing services). That is, market efficiency requires that return on housing asset has to be equal to the after-tax return on alternative financial asset. The exposition above relied on the assumption of minimal market frictions.

Life-cycle portfolio allocation problem is often related to the optimal mortgage choice (see, e.g. Campbell and Cocco (2003)) which is, in turn, related to the dynamics of interest rates and house prices. A recent study by de Jong, Driessen, and Van Hemert (2007) include five possible sources of uncertainty, namely expected and unexpected inflation, real interest rate, house prices and stock prices. Drift term in the equation for nominal house price is represented as nominal interest rate less imputed rent (which is assumed to be constant) adjusted for composite market price of risk for all factors. Shock is the sum of shocks to stock prices, real interest rates and inflation (which are correlated) and idiosyncratic component. This idiosyncratic component (orthogonal to the other sources of risk) is not priced in the model. In the estimation part de Jong, Driessen, and Van Hemert (2007) first obtain parameters for interest rate and inflation processes by quasi-ML, then estimate market price of risk for interest rate and expected inflation by matching yields on bonds with different maturities and finally calibrate the parameters of the remaining state variable processes to the data (the data used for house price dynamics calibration are annual returns calculated using OFHEO repeated-sales index for US cities from 1980 to 2003). In particular, de Jong, Driessen, and Van Hemert (2007) report the magnitude of the standard deviation for idiosyncratic risk, which individual homeowners are subject to, to be in the range 6%-12% (*cf* our estimates of house price volatility in table 3.4). This

paper is one of relatively few studies in which parameters of the stochastic housing process were estimated. In many cases²¹ authors obtain the parameter estimates for house price process (or commercial property price process for the former paper) by calibration.

Models set in a general equilibrium framework such as Ortalo-Magne and Rady (2006) often relax single-generation assumption but do not allow for multiple assets. Lately some researchers attempted to bring real estate to consumption-based asset-pricing models as in Piazzesi, Schneider, and Tuzel (2007). Piazzesi *et al* modeled the share of housing consumption in the aggregate consumption as highly persistent and heteroscedastic variable and suggest that to a large extent volatility in consumption is due to a *composition* risk (changes in asset prices relative to changes in expenditure shares). Composition risk is higher in bad times; higher perceived risk induces more precautionary savings so that discount factor decreases and bond prices fall by less for a unit change in expenditure ratio than stock prices do. Thus changes in expenditure share on housing are viewed as an additional factor in Capital Asset Pricing Model. However, housing process is not modeled explicitly, rather the model is calibrated to data on aggregate housing expenditure in National Income and Product Accounts (NIPA). Among the challenges that researchers working in the general equilibrium framework face, is the need to reconcile high volatility of residential investment, high volatility of house prices and regularities of the business cycles. Data from a large sample of OECD countries (Goodhart and Hofmann 2007) suggests that the *change* in real house prices is a good predictor for turning points in business cycle. Although the owner-occupancy rate differs substantially across these countries, still even for the country with lowest level of home ownership more than 40% of households live in their own house. Therefore, it is highly likely that house prices have a direct effect on economic activity. Due to sluggish supply response house prices tend to amplify the effect of macroeconomic shocks. Apart from supply inelasticity the other reason for this amplification is that so much housing wealth is used as a collateral. Consequently, the

²¹E.g. Titman and Torous (1989) or more recent empirical paper on valuation of MBS Downing, Jaffee, and Wallace (2010).

failure of borrowers to repay their debts during cyclical downturns leads to increased pressure on financial intermediaries and further suppresses asset prices in general. Modeling of house prices within the framework of dynamic general equilibrium models poses significant challenge. It is not surprising, then, that most empirical house price studies are cast in reduced-form setting. In reduced-form models housing price equation is derived by equating pre-specified supply and demand equations. Alternatively, housing price equation can be obtained by inversion of a demand function. An empirical tool of choice for many researchers is vector autoregression (VAR)²². For example Meen (2001) presents the following specification of error correction form for the house price model (2.52):

$$\Delta \log(g)_t = \gamma_1 \Delta \log(g)_{t-1} + \gamma_2 \Delta \log(X)_t + \gamma_3 (\log(g) - \gamma_4 \log(X))_{t-1} + u \quad (2.53)$$

where $X = [Y_{rpc}, W, H, HH]'$, Y_{rpc} is real per-capita income, W is real wealth, HH is the number of households. Price of a unit of housing $g(\cdot)$ is a function of income, wealth, housing stock, population, mortgage stock (M) and the expected nominal price of housing (P).

$$\log(g) = \alpha_1 \log(Y_{rpc}) + \alpha_2 \log(W) + \alpha_3 \log(HH) + \alpha_4 \log(H) + \alpha_5 \log(M) + \alpha_6 r + \alpha_7 \frac{\partial \log(P)}{\partial t} + \epsilon \quad (2.54)$$

One interpretation for a parameter γ_1 in this specification is that of a 'bubble-builder' and γ_3 – a 'bubble-burster'. Iacovello (2000) estimates VECM with real GDP in place of Y , real house price index, short-term nominal interest rate, real money aggregate and consumer inflation. He finds empirical support for his hypothesis about the presence of cointegration between GDP and real house prices in six European countries in the varying sample periods (late seventies to late nineties). The null hypothesis of a unit root in a house price series is rejected for 2 countries out of 6. Cross-country analysis suggests that countries with low transaction costs, high rate of ownership (one can infer

²²Or vector error correction model (VECM) if macroeconomic series used in the estimation of the model are non-stationary and cointegrated, which is often the case.

that these are the countries with more developed institutional structure of the housing market) display greater sensitivity of house prices to monetary shocks. Another approach in this line of research is to condition housing demand directly on consumers' expenditure rather than the determinants of expenditure. Dynamic model of Pain and Westaway (1997) contains ratio of housing assets to consumer expenditure, real house prices and rate of ownership in the housing market as well as time dummies to capture the policy-change shocks. The results of the estimation on the UK data suggest that this parsimonious model (employed at the time of publication by Bank of England) performs as well as more elaborate error-correction models. However, the two distinctive features of the housing market are localization and spatial dependencies, whereas standard error correction model restricts parameters to be the same across the spatial dimension. This may be overly restrictive, especially for the case when the data used in the estimation is already aggregated at the regional level. The latent-class model of van Dijk, Frances, Paap, and van Dijk (2007) allows for the different parameters across groups of regions (latent classes). Not surprisingly, house prices within the compact country (the Netherlands) are found to be cointegrated across regions, so the author constrain time trends to be not too divergent, at the same time allowing the speed of convergence towards the equilibrium to be different for different regions. ML estimation via EM algorithm suggests that limited number of latent classes (two), which differ with regard to the response to macroeconomic shocks, is sufficient to capture the essential features of the dynamics of regional prices. An alternative approach pursued by Holly, Pesaran, and Yamagata (2006) is to allow model parameters to vary across regions without any clustering. They apply panel error correction model to state-level (OFHEO) repeat-sale indices. In this model the logarithm of real price of housing is the linear function of real disposable income and real interest rate. The authors represent error term as a linear function of unobserved common effects and idiosyncratic component for each state (these latter are allowed to be spatially correlated). The evidence in Holly, Pesaran, and Yamagata (2006) with regard to

the presence of unit root in the panel of statewide house price indices is mixed, whereas the cointegration relationship between house price index in real terms and real incomes is found to be significant after accounting for heterogeneity and spatial correlation. Holly, Pesaran, and Yamagata (2006) who used series up to 2003 suggested that there was little evidence of house price bubble in all but limited number of states (Pacific division and some Northeastern states).

Although many empirical studies focused on the properties of the first moment of the price process, the volatility of house prices received increasing attention lately. Capozza, Hendershott, and Mack (2004) found that prices across MSA and across time periods exhibit heterogeneity with regard to shocks in income, population and construction costs. Miller and Peng (2006) were able to obtain the estimates of the house price volatility for about 12% of 277 MSA, for which series of house price index were available and found evidence favorable to the hypothesis of asymmetric response²³ of volatility to shocks to GMP and appreciation rate of housing. Dolde and Tirtiroglu (2002) find that realized house price returns (as measured by Freddie Mac CMHPI for the four US Census regions) respond to volatility shocks and this response takes the sign opposite to that of the volatility shock. These authors also conclude that most of the volatility shifts in the observed period (1975-1993) were on a regional rather than national scale²⁴. Crawford and Fratantoni (2003) compare forecasting performance of univariate time-series models of house prices accounting for persistent volatility versus performance of models that do not include this factor and found that former usually is better. These authors have found strong evidence in favor of the presence of autoregressive conditional heteroscedasticity in 60% of the cases (three out of five states).

In the present study we have adopted a modeling approach that takes into account regional differences in house price appreciation rates. Our model of house prices is a

²³Larger and more protracted effect of negative shocks as indicated by impulse response analysis of VAR model.

²⁴See figure 3.16 for the graphical representation of the comovements in the regional house prices.

reduced-form continuous-time one depending on a regional-specific latent factor with diffusion dynamics. Fluctuations of interest rates (bond prices) play the role of a common factor. We have chosen the specification so that we are able to obtain closed-form solutions for the house price volatility.

2.6 AN OVERVIEW OF LITERATURE RELATED TO INTENSITY MODELING

2.6.1 SEQUENTIAL MONTE CARLO APPROACH TO LIKELIHOOD APPROXIMATION

Extant research abounds with different sorts of implementations of Sequential Monte Carlo (SMC) techniques (a.k.a. *particle filtering*). In what follows I will briefly review some of the theoretical concepts related to Sequential Monte Carlo methods in general before going into the details of the particular implementation.

In the generalized state-space framework an econometrician seeks to infer parameters of the unobserved state $\mathbf{x}$ (with the parameter vector $\boldsymbol{\theta}$). The model is described by:

$$y_t = H_t(\mathbf{x}_t, \mathbf{v}_t) \quad (2.55)$$

A (discrete) time counter t in the foregoing observation equation runs from 1 to T . $\{Y_n\}_{n \geq 0}$ is $\mathbb{R}^{n_y}$ -valued stochastic process taking values in $\mathcal{Y}$ while unknown state $\{X_n\}_{n \geq 0}$ is $\mathbb{R}^{n_x}$ -valued stochastic process taking values in $\mathcal{X}$. Hereafter, with some abuse of notation $\mathbf{x}_t$ will stand either for random variable or for its realization. The initial density of the process is denoted by $X_0 \sim \nu$. The evolution of the state vector $\mathbf{x}$ is described by the following equation:

$$\mathbf{x}_t = G_t(\mathbf{x}_{t-1}, \mathbf{u}_t) \quad (2.56)$$

where state disturbances $\mathbf{u}_t$ are serially independent, $\mathbf{u}_t$ and $\mathbf{v}_t$ are mutually independent and functions G_t and H_t governing the evolution of the state and observation processes can be non-linear. In principle observations can depend on the history of state process and observational history, but often they are assumed to be independent conditionally on hidden states; also, often the state process is taken to be Markovian with the transition

density $X_n|X_{n-1} = x \sim f_\theta(\cdot|x)$. Consequently, the models of this class are sometimes alternatively termed hidden Markov models (HMM). The state value at a given point in time in this case is independent of the previous history²⁵ of the state process and past observations

$$p(x_t|x_{1:t-1}, y_{1:t-1}) = p(x_t|x_{t-1}) \quad (2.57)$$

The quantities of primary interest (if the task is parameter estimation) are typically confined to likelihood, predictive and filtering densities. For likelihood-based parametric inference one needs to evaluate the *likelihood*:

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{x}|\boldsymbol{\theta})p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})d\mathbf{x} \quad (2.58)$$

To model the evolution of the dynamic system one is interested in learning the *predictive* density:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \quad (2.59)$$

This distribution can be derived by marginalizing the joint posterior:

$$p(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}) = \frac{p(\mathbf{y}_{1:t-1}|\mathbf{x}_{0:t-1})p(\mathbf{x}_{0:t-1})}{\int p(\mathbf{y}_{1:t-1}|\mathbf{x}_{0:t-1})p(\mathbf{x}_{0:t-1})d\mathbf{x}_{0:t-1}} \quad (2.60)$$

This latter satisfies the recursion:

$$p(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}) = p(\mathbf{x}_{0:t-2}|\mathbf{y}_{1:t-2}) \frac{p(\mathbf{y}_{t-1}|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{x}_{t-2})}{p(\mathbf{y}_{t-1}|\mathbf{y}_{1:t-2})} \quad (2.61)$$

The filtering distribution obtains as:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t} \quad (2.62)$$

However, direct computation of many quantities of interest such as the normalizing constant in the denominator of (2.61), the filtering density $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ or expectations with respect to the posterior $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ involves evaluation of integrals of high dimension and is often infeasible. An exception is the Gaussian case with linear observation and transition equations (Kalman filter). In non-linear non-Gaussian setting SMC (particle filters)

²⁵For the first-order Markov models the history beyond the preceding time period.

often prove indispensable. Particle methods rely on discrete approximation of distributions of interest by a weighted empirical distribution of a large set of M independent and identically distributed samples (particles). More formally, $\{\hat{\mathbf{x}}_{0:t}^i, w_{0:t}^i\}_{i=1}^M$ can be thought of as a random measure that characterizes the posterior density $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$. Particle weights w^i are normalized so that at each t they sum up to 1. Therefore, the particle set provides a discrete weighted approximation to the posterior filtering density:

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \approx \sum_{i=1}^M w_{0:t}^i \delta(\hat{x}_{0:t}^i - x_{0:t}) \quad (2.63)$$

where δ is the Dirac delta function. This discrete distribution is zero at any point x except for the points under the support of the particles. Nevertheless, estimation of the expectations of the functions of a hidden state can be performed using this discrete approximation as:

$$\int f(\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) d\mathbf{x}_{0:t} \approx \sum_{i=1}^M w_{0:t}^i f(\mathbf{x}_{0:t}^i) \quad (2.64)$$

Initially SMC was developed as a sequential version of sampling importance resampling (SIR) of Smith and Gelfand (1992). The problem with using (2.64) is in updating the weights w_t^i given the weights in the previous period w_{t-1}^i . This goal can be achieved using sequential importance sampling (SIS). Taking the state transition density as a proposal density (denote this proposal density by q) for the time periods $1, \dots, t$ and denoting by $\nu(\mathbf{x}_0)$ proposal for the initial distribution of the hidden state we can construct proposal recursively as:

$$q_t(\mathbf{x}_{0:t}) = q_{t-1}(\mathbf{x}_{0:t-1}) p(\mathbf{x}_t|\mathbf{x}_{t-1}) \quad (2.65)$$

It can be observed that all the information about the i^{th} path up to time t is contained in the particle $\hat{x}_{t-1}^i$ so storing the preceding $(\hat{x}_{t-2}^i, \hat{x}_{t-3}^i)$ values of the particles for the purpose of updating is not necessary. For the target density which we denote by π ($\pi_t(\cdot) = p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$), analogous decomposition yields:

$$\pi_t(\mathbf{x}_{0:t}) = \frac{p(\mathbf{x}_{0:t}, \mathbf{y}_{1:t})}{p(\mathbf{y}_{1:t})} = \frac{p(\mathbf{y}_{0:t-1})}{p(\mathbf{y}_{0:t})} \pi_{t-1}(\mathbf{x}_{0:t-1}) p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{y}_t|\mathbf{x}_t) \quad (2.66)$$

where we made use of (2.61). Ratio of the latter to the former provides the way to update the (unnormalized) weights:

$$\begin{aligned}\tilde{w}_t^i &= \tilde{w}_t^i(\hat{\mathbf{x}}_{0:t}^i) = \frac{\pi_{t-1}(\hat{\mathbf{x}}_{0:t-1}^i)}{q_{t-1}(\hat{\mathbf{x}}_{0:t-1}^i | \mathbf{y}_{1:t-1})} \frac{p(\hat{\mathbf{x}}_t^i, \mathbf{y}_t | \hat{\mathbf{x}}_{t-1}^i)}{q_t(\hat{\mathbf{x}}_t^i | \hat{\mathbf{x}}_{t-1}^i, \mathbf{y}_{1:t})} \\ &= w_{t-1}^i \frac{p(\mathbf{y}_t | \hat{\mathbf{x}}_t^i) p(\hat{\mathbf{x}}_t^i | \hat{\mathbf{x}}_{t-1}^i)}{q_t(\hat{\mathbf{x}}_t^i | \hat{\mathbf{x}}_{t-1}^i, \mathbf{y}_{1:t})}\end{aligned}\quad (2.67)$$

The problem with weights when SIS is applied to long enough paths is that of weight degeneracy: no matter how large M is, eventually all the probability mass concentrates in a few particles. This led to the introduction (Gordon, Salmond, and Smith (1993)) of a resampling step into the algorithm, during which step the particles are redistributed so as to provide even coverage of a posterior. The algorithm for a generic SIR particle filter (with a multinomial resampling step as in Gordon, Salmond, and Smith (1993)) can be written as:

Algorithm 1 (SIR with multinomial resampling).

Step 1 Draw samples $\hat{\mathbf{x}}^1, \dots, \hat{\mathbf{x}}^M$ from proposal (instrumental) distribution q .

Step 2 Calculate unnormalized weights $\tilde{w}^i = \frac{\pi(\hat{\mathbf{x}}^i)}{q(\hat{\mathbf{x}}^i)}$.

Step 3 Calculate normalized weights $w^i = \frac{\tilde{w}^i}{\sum_{i=1}^M \tilde{w}^i}$

Step 4 Draw N discrete r.v. $(I^1, \dots, I^N)$ with $\Pr(I^1 = j) = w^j$ conditionally independently, given $\hat{\mathbf{x}}^1, \dots, \hat{\mathbf{x}}^M, j = 1, \dots, M$.

Step 5 Set $\hat{\mathbf{x}}_\star^i = \hat{\mathbf{x}}^{I^i}$.²⁶ This yields an updated sample from the target π with weights equal to $\frac{1}{N}$.

This algorithm propagates particles with high importance weights and eliminates those with low weights by drawing N samples from the set of the old particles with

²⁶Subscript $\star$ stands for update.

probabilities proportional to the importance ratios. However, since multinomial resampling in steps 4-5 above introduces non-negligible noise (variance in the number of replicates of the old particles), modifications of a generic particle filter include other sampling schemes (residual resampling, systematic resampling) that alleviate the extra noise problem. Resampling step, while mitigating the effect of weight degeneracy, at the same time leads to the thinning of the initial sample. Since the initial values of the path are simulated only once resampling reduces the number of unique particles at each successive iteration. This depletion develops sooner when the prior (transition density in case of basic SIR filter) is far from the likelihood. Unfortunately, with importance weights given by recursive formulae such as (2.66) the variance of weights inevitably increases over time. If one wants to preserve the initial sample, he should have increased the size of the 'particle cloud' exponentially over time. Various improvements have been devised to mitigate this effect, for example Gilks and Berzuini (2001) suggest replenishing the set of particles via occasional Markov Chain moves. In case of multiple models or unknown hyperparameters it might be beneficial if particles can 'jump' to a new point in the space of parameters; this so-called evolution step is achieved by sampling from Markov transition kernel of invariant distribution $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ using either Gibbs sampler or Metropolis-Hastings method. Resample-MCMC move algorithm allows in addition to drawing state samples $\mathbf{x}_t^i$ at time t to modify the values of the paths up to L steps back thus 'rejuvenating' particle swarm. An alternative to Resample-MCMC move known as block sampling goes further and instead of drawing single state value attempts to sample L values along the path simultaneously at time t . Doucet, Briers, and Senecal (2006) argue that this strategy can significantly decrease the number of resampling steps and reduce the variance of importance weights.

We have chosen the strategy developed by Pitt and Shephard (1999) who improved on generic SIR filter by incorporating the information contained in the new state value in the incremental importance weight in (2.67). M. Pitt and N. Shephard cast this method in

the framework of auxiliary variables and hence the term Auxiliary Particle Filter (APF). Later it was shown that APF can be viewed as a sequential Monte Carlo method with the target distribution including information contained in the *predictive* likelihood $\tilde{p}(\mathbf{y}_{t+1}|\mathbf{x}_t)$:

$$\pi_t(\mathbf{x}_{0:t}) = p(\mathbf{x}_{0:t}, \mathbf{y}_{1:t})\tilde{p}(\mathbf{y}_{t+1}|\mathbf{x}_t) \quad (2.68)$$

and importance distribution of the form

$$q_t(\mathbf{x}_t|\mathbf{x}_{1:t-1}) = q(\mathbf{x}_t|\mathbf{y}_t, \mathbf{x}_{t-1}) \quad (2.69)$$

The following auxiliary particle filter (APF) algorithm was implemented for the estimation of the unobserved states (latent intensities):

Algorithm 2 (ASIR with smooth resampling).

At time $n=1$

- Draw samples from $q(\hat{x}_1^i|y_1)$, $i = 1, \dots, M$.
- Calculate unnormalized importance weights $\tilde{w}_1(\hat{x}_1^i) = \frac{\nu(\hat{x}_1^i)g(y_1|x_1^i)}{q(\hat{x}_1^i|y_1)}$
($g(\cdot) \sim \text{Poi}$ - observational density).
- Calculate normalized importance weights $\tilde{W}_1^i \sim \tilde{w}_1(\hat{x}_1^i)$.

At time $n \geq 2$

- Calculate normalized first-stage weights $W_{n-1}(\hat{x}_{n-1}^i) \propto \tilde{W}_{n-1}(\hat{x}_{n-1}^i) \times \tilde{p}(y_n|\hat{x}_{n-1}^i)$.
- Resample pairs $\{W_{n-1}^i, \hat{x}_{n-1}^i\}$ to obtain $\{\frac{1}{M}, \bar{x}_{n-1}^i\}$.
- Draw samples from $q(\hat{x}_n^i|y_n, \bar{x}_{n-1}^i)$ and set $\hat{x}_{1:n}^i \leftarrow (\bar{x}_{n-1}^i, \hat{x}_n^i)$.
- Calculate unnormalized second-stage weights as $\tilde{w}_n(\hat{x}_{n-1:n}^i) = \frac{g(y_n|\hat{x}_n^i)f(\hat{x}_n^i|\hat{x}_{n-1}^i)}{\tilde{p}(y_n|\hat{x}_{n-1}^i)q(\hat{x}_n^i|\hat{x}_{n-1}^i)}$
($f(\cdot)$ - transition density of the states).
- Calculate normalized second-stage weights $\tilde{W}_n^i(\hat{x}_n^i) \propto \tilde{w}_n(\hat{x}_{n-1:n}^i)$.

Early on it was suggested to use some likely value γ associated with the transition density to approximate empirical filtering density as $g(y_{t+1}|\gamma_{t+1}^i) f(\hat{x}_{t+1}|\hat{x}_t^i)$, (Pitt and Shephard 2001, p. 278) . More recently it has been pointed out that this approximation may lead to the estimates with large or even unbounded variance (Doucet and Johansen 2008, p. 25) . We have opted to choose the following form for the predictive likelihood $\tilde{p}$:

$$\tilde{p}(y_{t+1}|\hat{x}_t) \propto \int \tilde{g}(y_{t+1}|\hat{x}_{t+1}) f(\hat{x}_{t+1}|\hat{x}_t) d\hat{x}_{t+1} \quad (2.70)$$

It is important to ensure that $\tilde{p}(y_{t+1}|\hat{x}_t)$ be more diffuse than $p(y_{t+1}|\hat{x}_t)$, therefore the approximate likelihood density $\tilde{g}(\cdot)$ should have 'fatter' tails than $g(\cdot)$.

It has been shown (Chopin 2004) that Central Limit Theorem holds for estimators obtained by SMC methods. The asymptotic variance for the mean estimator of an arbitrary function $h(\cdot)$ in a state space of dimension s is of the order $O(n^{\frac{s}{2}})$ for the multinomial sampling scheme²⁷. For the properties of approximate MLE we refer to the exposition in Cappé, Moulines, and Rydén (2005, p. 443-468), who are able to prove that under standard conditions MLE for generalized state-space models (HMM) possess usual properties (consistency, efficiency and asymptotic normality).

Chapter 3 deals with parameter estimation of the components of the pricing model. I begin with multiplicative intensity (hazard) parameter estimation and proceed in the order of the Sections in the present Chapter.

²⁷ n here is the number of observations (length of the series).

CHAPTER 3

ESTIMATION OF THE PARAMETERS OF THE MODEL

3.1 ESTIMATED PARAMETERS OF THE MULTIPLICATIVE INTENSITY MODELS

3.1.1 MORTGAGE DATA

The empirical analysis is based on the data provided by Black Box Logic, LLC that maintains Bond and Loan Information System (BLIS) database. The data provider claims that their complete data set contains information about 6000 pools of non-agency securitized mortgages. We used part of the data set containing information about adjustable-rate mortgages originated during the 1997-2009 period and observed from 2000 through the first half of 2009. Since our main subject of interest are non-prime loans, but there is no indication in the data whether the mortgage is subprime, Alt-A or prime non-conforming, we have chosen to include in the analysis only mortgages with FICO score at origination of 720 or less¹. We include only 30-year loans that are secured by 1-4 family houses or condos. Our analysis puts high emphasis on the measure of borrower's equity, therefore we consider only first liens for which combined loan-to-value ratio is 80% or greater and is equal to LTV at origination (this reduces total number of observations by about 1/3 compared to the sample of loans with LTV upwards of 40%). As an additional guard against the present of 'silent seconds' we search for the second liens that were originated

¹This may not coincide with labeling a contract as subprime or Alt-A by the secondary market but there should be significant degree of overlap. For example Gerardi, Shapiro, and Willen (2008) observe that only 10% of the subprime mortgages in their analysis had LTV higher than 720. See also Mayer, Pence, and Sherlund (2010) who tabulate median FICO score at origination for a variety of non-prime mortgage types and various vintages; their estimates for the Alt-A category lie in the range 694-708.

within 3 months from the date of origination of the first-lien loan which combined loan-to-value ratio is equal to the sum of LTV at origination of the first-lien loan and LTV at origination of the second lien loan. Although this procedure produced few matches, we excluded the suspect first-lien loans from further analysis. We further restrict our analysis to the loans that were originated in the 20 largest MSAs that are represented in the Case-Shiller house price appreciation (HPA) index. We removed all loans which original balances were less than \$20,000 in year 2000 \$ (roughly 15% of the median house price) as these properties are unlikely to be representative of the more 'conventional' set of properties. We also winsorize (remove top 0.5% of) our data with respect to the loan size, which results in the largest loan in our sample being about 1.8 million dollars.

The records for each contract contain, in particular, the following information: the year and month of origination and termination (if the latter occurred within the observation window), the indicator for the type of termination (prepayment or default), the contractual amount of loan at origination and outstanding loan balance at the observation date, loan-to-value ratio at origination, initial contract interest rate, value of index rate at the time of origination, date of the first adjustment of the interest rate, contractually specified margin, lifetime cap on the contract rate, periodic cap on the contract rate, lifetime floor on the contract rate, contract rate at the observation date, indicator of the type of documentation furnished by the borrower (full documentation or less than full documentation), purpose of the loan (purchase or refinancing), indicator of whether mortgaged property is primary or non-primary residence, indicator of whether the mortgage contract includes prepayment penalty, zip code of the area where the property is located. For the contracts with initial teaser rate (i.e. rate lower than the sum of the index rate and margin) the length of the period for which the teaser is effective is specified. Given the indicator of the presence of the teaser and the calculated length of time until first rate adjustment, the approximate breakdown of contracts by type looks as follows: about a quarter of the loans can not be identified (most often due to the missing values of the variables con-

taining date of first interest rate adjustment); out of remaining loans share of the so-called '2/28' ARMs is almost 60%. '3/27' ARMs constitute about 7%, 'regular' ARMs with initial fixed period extending for 3, 5, 7 or 10 years constitute the rest². Summary statistics for the main loan characteristics are reported in the table 3.4. The breakdown by the time of origination and by the time of termination is shown in the tables 3.2 and 3.3. Most of the contracts were originated at the peak of the subprime boom – in 2005-2006 (about 70% of all contracts in the sample). Figure 3.2 presents the series of default and prepayment in calendar time. One can observe that default incidence in our sample has been steadily rising since early 2006 and reached its absolute peak in early 2008. Prepayment peaked at about the same time and was rapidly declining since then. Unlike in prepayments, there was second, smaller peak in defaults in early 2009 after a drop in late 2008, which can possibly be attributed to either lagged effect on the housing and labor markets of the acute phase of crisis in Fall 2008 or, perhaps, to some end-of-the-fiscal year accounting considerations. It doesn't seem that we are observing effect of re-setting the rates on '2/28' mortgages to the non-teasered level since comparison of figure 3.2 and figure 3.1 does not reveal any sharp decline in the mean contract rate in the sample prior to its local peak in the summer of 2007, which coincided with liquidity crisis on the market for subprime-backed securities. On the other hand, casual examination of figure 3.12 suggests that for many markets cumulative amount of house price depreciation by the beginning of 2009 reached 25% -30% or more compared to peak home price levels. For a typical mortgage from our sample (originated in 2005-2006) would result in negative equity in excess of 10%-15%. In such equity position temporary loss of labor income of one (or more than one) of the members of the household may lead to the decision to abrogate the contract.

Figure 3.1 depicts dynamics of the initial contract rate during the observation window and its relation to the key interest rates. Table 3.4 in some sense complements figure 3.1 providing information about such components of the contract rate as margin and teaser

²ARMs with fixed 5-year period seem to be the most popular, comprising about a half of all the loans in the 'regular' ARMs category.

(on average across the time dimension). In turn, tables 3.5 and 3.6 expand the information in table 3.4 allowing for comparison of the characteristics of the loans that prepaid versus loans that were foreclosed. As expected, the latter have higher average contract rate (74 bp difference on average) and somewhat higher original LTV (85.94% vs 85.37%). The credit scores of the two subsamples of borrowers differ substantially (614.6 vs 634.2). An average borrower, who defaulted on his mortgage, was willing to accept higher margin (5.59 percentage points vs 5.2 pp), higher lifetime cap (13.41% vs 12.51%) and noticeably higher lifetime floor (7.58% vs 6.96%) on the contract rate. There was higher share of loans secured by non-primary residence (8% vs 6%) among those in default. Perhaps due to the fact that Alt-A loans are more likely to be less than fully documented, there were less loans with incomplete documentation among foreclosed mortgages than among prepaid mortgages (62% vs 74%). No pronounced difference with regard to the loan size between the two groups can be detected. These observational differences may serve as indirect evidence for the ability of lenders to (some degree) screen the borrowers with regard to their potential riskiness.

3.1.2 SPECIFICATION OF THE MODEL

There is vast and growing empirical literature studying mortgage default and prepayment. For all the variety of approaches there seems to exist general agreement as to the factors proven to have strong influence on the probability of a premature termination of a residential mortgage. One of the major causes of prepayment is refinancing: refinancing incentives for borrowers emerge when rates on the feasible set of contracts drop below their existing contractual rates. To capture this factor, some measure of the difference in rates should be developed, and there is substantial latitude in existing approaches. One natural specification is to simply use the difference between the original contract rate and the current (benchmark) rate as in Schwartz and Torous (1989), or the difference as a percentage of the latter (e.g. Ambrose and Sanders (2003)) or of the former (Cunningham

and Capone (1990), among others). Those arguing in favor of the ratio often claim that the incentive to refinance is dependent on the level of rates as well as on the difference. Some, for example, Pavlov (2001), Richard and Roll (1989) use ratio of initial and current rates or variation thereof, as in Richard and Roll (1989), others add non-linear terms in the primary variable, e.g. Huang and Ondrich (2002), Ciochetti, Deng, Gao, and Yao (2002). It is not uncommon to include discounting, as in using market price of the loan (present value of mortgage payments at prevailing current market rate), e.g. Clapp, Deng, and An (2006) or some function thereof (value of prepayment option as in Deng, Quigley, and Van Order (2000)). It should be noted that the majority of loans in our sample are '2/28' or '3/27' ARMs that were initially conceived as products for borrowers, whose credit was impaired, and who presumably would qualify for a more conventional loans after they would have been current on their obligations for the period of 2 to 3 years. This motivated the specification where interest rate spread was defined as the difference between the current contract rate on the existing contract and the contract rate on a hypothetical fixed-rate mortgage with balance equal to the outstanding balance on the ARM: $SprToFRM = (r_{ARM,t} - r_{FRM,t}) / r_{FRM,t}$. The model for the projected FRM rate is estimated using observed rates on the pooled subsample of fixed-rate mortgages originated during the same time period (1997-2008) within the same 20 MSAs (see table 3.7 for the summary of these data). We model the potential interest rate on the FRM that former ARM borrower might be offered as a linear function of the loan-to-value ratio at origination ($OrigLTVRatio$), FICO score of the borrower ($FicoScoreOrig$), loan size ($LoanSizeOrig$)³ and the term structure at the time of refinancing. Because the latter is so important, we use three variables to capture the effect of term structure on the mortgage rate: $Yield10Y$ is the current yield on the 10-year Treasury note which is widely regarded as a benchmark rate for 30-year FRMs, $Yield10YSq$ is the quadratic term intended to capture possible nonlinearities and $SlopeTS$ is the slope of the term structure curve approximated as

³It should be stressed that loan size at the time of origination of *FRM* is equal to the outstanding mortgage balance on *ARM*.

the difference in 10-year and 1-year yields⁴. Other characteristics of the borrower/loan which are included in the model are dichotomous variable for the type of documentation (*LoDocLoan*), interaction of this variable with the loan size (*SizeOrigLoDoc*) and dichotomous variable indicating whether collateral is not a primary residence for the borrower (*InvestLoan*). The values of the *LoDocLoan* and *InvestLoan* are assumed to be the same at the time of hypothetical refinancing as they were for the original ARM. We also include MSA fixed effects in the model (estimates of the fixed effects, which are significant at 5% level for all but 2 MSAs, are not reported in the table 3.8 where the results of the estimation are tabulated). Given the concerns about possible presence of the unreported second loans and ubiquity of piggyback lending in the observation period, we added a dichotomous variable *LTV80Orig*, which takes value of 1 if LTV of original ARM loan was exactly 80%, taking this as indication of the possible presence of a piggyback loan. Indeed, the estimate of the coefficient of *LTV80Orig* is found to be positively related to the interest rate, above and beyond the positive effect of *OrigLTVRatio*. We calculate *SprToFRM* using observed values of interest-rate related variables at each observation date, outstanding mortgage balance for each loan in the sample at each observation date, loan and borrower characteristics such as *InvestLoan*, which are assumed to be static, and the estimated parameters (reported in table 3.8). Higher value of *SprToFRM* is expected to increase the intensity of prepayment.

It was observed in the literature that dollar-measured refinancing incentive is proportional to the size of the loan (e.g. Bennett, Peach, and Peristiani (2001)). Also, transaction costs of refinancing will be relatively lower for larger loans, and borrowers with higher loan balances are, perhaps, more often approached by mortgage brokers. Whether sensitivity of such borrowers to the incentive is the same as the one of borrowers with smaller

⁴See table 3.7 for the sample statistics of the variables used for the estimation of fixed-rate model. Some of the definitions for the variables that are specific to this model are also provided in the footnotes to that table, while the variables that are common to fixed-rate and intensity models are defined in table 3.1.

loans is less clear, however, an expected sign of the coefficient on the variable *LoanSizeOrig*⁵ is positive in the prepayment model.

FICO score is universally used measure of borrower's credit quality. Unfortunately, this is a static metric in our data, so the quality of this measure may deteriorate over time. I include *FicoScoreOrig* into the hazard model for prepayment expecting to find general positive relationship between the credit rating of the borrower and probability of prepayment. Since so many contracts in our data have initial low 'teaser' rate and are liable to possible steep increases in the contract rate after the expiration of the teaser, it is likely that borrowers will consider prepaying the mortgage around the time of the expiration of the teaser. First variable that is intended to capture the incentive due to teaser is *TeaserI*, a dynamic covariate proportional to the extent of initial reduction of the contract rate; its value decreases over the effective period of teaser and thus it can be expected to negatively affect the hazard of prepayment. The static variable, *TeaserM*, equal to the magnitude of the initial reduction of the contract rate, on the contrary, should be related positively to the probability of prepayment, as it is likely to be highly correlated with the magnitude of increase in mortgage payment after teaser expiration. Finally, the dichotomous variable *D24Mon* (which takes value of 1 around the time of expiration of teaser for the most populous category of '2/28' ARMs) is intended to capture jump-type increase in prepayment above and beyond effects represented by *TeaserI* and *TeaserM*.

Other dynamic variables that are relevant for prepayment decision of the borrower include the current contract rate (*CurrContRt*) which, *ceteris paribus*, should be positively affecting the prepayment intensity, and the squared term in contract rate (*CurrContRtSq*) that accounts for nonlinearities in borrower's response to changing interest rate environment. Seasonality is generally considered to be an important explanatory variable in the models of mortgage prepayment, as relocation is generally tilted towards summer months. To capture this tendency in a parsimonious way, I add a sinusoidal trend in cal-

⁵Table 3.1 provides definition of this and other variables in the intensity model.

endar time with period equal to 1 year, maximum in June and minimum in December (variable *SineSeason*). Option-theoretic mortgage models predict that prepayment (and default) decisions are influenced by expectations of future paths of state variables. To account for the variability of house prices we include the dynamic covariate reflecting historical variability of the house prices in a given MSA (*HPriceVol*). According to option-theoretic models, this covariate should be positively associated with the likelihood of prepayment.

As additional static controls we use *LTV80Orig*, *PrepPenaltyLoan* – indicator of the presence of prepayment penalty, *InvestLoan* – indicator of status of the collateral and *LoDocLoan* – indicator of less-than-full documentation. The first two variables are expected to have negative effect on prepayment (recall that *LTV80Orig* is a rough proxy for possible presence of the second loan secured by the same property), whereas the effect of the latter variables is not clear *ex ante*. It might be that borrowers who do not provide proof of income and/or assets have more problems when they try to refinance; on the other hand, if there is strong correlation between incomplete documentation and Alt-A status, then these loans may be no harder to refinance than true subprime loans with full documentation. Therefore, the effect of *LoDocLoan* is left for empirical investigation.

Finally, the ability of the borrower to refinance is inversely related to the value of equity in the property. We construct the measure of loan-to-value ratio at the time of observation (*CurrentLtv*) as fraction, which numerator contains outstanding balance on the loan at the time of observation and denominator represents estimate of house value obtained by applying MSA-level house price appreciation index to the initial value of the collateral⁶ (original loan balance / original loan-to-value ratio). Borrowers with more equity in the house (in other words, those with lower *CurrentLtv*) presumably may qualify for a new loan easier, and thus the expected sign of the coefficient is negative. Addition-

⁶This is fairly common approaches in the literature: see, e.g. Firestone, Van Order, and Zorn (2007) who used Freddie Mac MSA-level price index. Deng, Quigley, and Van Order (2000) used estimated weighted repeat sales index (also on the MSA level) to adjust house values.

ally, rising house prices increase an incentive to take the equity out of the property, which is another reason to expect negative sign.

The same set of covariates is used in the multiplicative intensity model of default. In the latter model the measure of borrower's equity position, *CurrentLtv*, is of primary importance⁷. Its direct effect is positive⁸. Credit scores at origination are assumed to be inversely related to default probabilities. Dynamic covariates related to interest rate environment, such as *CurrContRt* and *SprToFRM* are expected to have positive signs while *TeaserI* should have negative impact. *HPriceVol* is expected to be positively related to the probability of default by the same argument as in the case of prepayment. The estimates of static covariates such as *TeaserM* and *LoDocLoan* are likely to be positive. If default costs are heterogeneous, in all likelihood these are less for the owners of the property which is not the primary residence, hence expected positive sign for *InvestLoan*. Priors for other static covariates and for the *SineSeason* are less clear-cut, but one may expect that, for example, for *PrepPenaltyLoan* second-order effect will call for a negative sign. However, if prepayment penalty is more often accepted by generally higher risk borrowers, then this covariate will have positive sign. I prefer not to take stand with respect to the effect of these covariates.

As it happens, there might be other relevant factors which were not included in the duration model; some of these may not be observable. One may conjecture that borrowers' heterogeneity with respect to their propensity to prepay is potentially a significant factor. As a robustness check, I augmented the model specification by variables repre-

⁷See, e.g. Case and Shiller (1996) or Yang, Buist, and Megbolugbe (1998) for an empirical evidence that measure of home equity is one of the most significant factors that affect the decision of a homeowner to default.

⁸So far the discussion was limited to direct, or first-order effects of the covariates. However, since default and prepayment are acting as competing risks, there are also second-order effects. For example if *SineSeason* increases probability of prepayment, its effect on probability of default is negative, *ceteris paribus*. However, if default also exhibit certain seasonal pattern then one may prefer to remain agnostic about the overall effect of *SineSeason* on default. In general, second-order effect of a covariate may either reinforce its primary effect if this covariate affects the default and prepayment oppositely (as in the case of *CurrentLtv*), or second-order extent may somewhat mitigate the first-order effect if the effect of a covariate on default and prepayment is the same.

senting various types of deterministic time trends in default and prepayment. These were found to be either non-significant or marginally significant and not materially improving AIC and BIC.

Importantly, model specification includes the stratification of the sample by the time of origination. As opposed to typical reduced-form modeling of mortgage duration, in the present framework the estimation of the effects of observable variables on the hazard of default and prepayment is not the end in itself, but merely the first stage in the estimation of the model, where terminations are assumed to follow doubly-stochastic Poisson process. The estimates of the multiplicative intensity model parameters allow one to evaluate realizations of stochastic baseline hazard functions l_t^ℓ (see (2.14)) of default and prepayment. These are realizations at event times (in the 'mortgage time' domain, i.e. on a discrete set taking values from 1 to 360). Having obtained the sample of the realizations, one can proceed to estimate the underlying continuous-time stochastic processes (which is the topic of Chapter 6). In order to reconcile stochastic (more precisely, doubly-stochastic) nature of termination with the above estimation strategy, Kau, Keenan, and Smurov (2004) proposed to stratify the data according to the origination period. I follow their approach and assume that contracts originated within the same calendar quarter share the common baseline; thus I obtain 43 pairs⁹ of sample paths of default and prepayment. These paths are of unequal length (first sample corresponds to loans originated in Q1 1997 that were observed throughout 2001-2009, whereas last sample – loans originated in Q1 2008 – was observed for little more than a year). While it is in principle possible to stratify by months instead of by quarter, it has proven impractical¹⁰. It appears that given

⁹Reader is reminded that originations span 1997 - first quarter of 2008 and that second half of 2004 is excluded from the estimation.

¹⁰An argument in favor of finer stratification enters in conflict with data considerations: the early strata are usually sparse and consequently the estimates obtained from the risk sets with but a few observations are subject to huge variation. A. Smurov in his unpublished dissertation (Smurov 2004) supports this point (cf footnote 15 on p. 28); he also cites the statistical evidence that suggests superior fit of a model with quarterly stratification versus models stratified over longer time periods.

existing two-stage estimation scheme quarterly stratification provides reasonable compromise between precision of the estimates and data limitations. I apply statistical test for stratification proposed by Ridder & Tunali (Ridder and Tunali 1999), where, the test statistic

$$C = (\hat{\beta}_s - \hat{\beta}_u)' [\mathbf{V}(\hat{\beta}_s - \hat{\beta}_u)]^{-1} (\hat{\beta}_s - \hat{\beta}_u) \quad (3.1)$$

is asymptotically $\chi^2(p)$ distributed under the null hypothesis (p is the number of estimated parameters, the s subscript denotes the stratified model, the u denotes the unstratified one, $\mathbf{V}(\cdot)$ is the covariance matrix). In all cases the null hypothesis that the true model is the unstratified one is rejected at 1% significance level (see table 3.9).

3.1.3 TESTABLE HYPOTHESIS AND ESTIMATION RESULTS

My initial intent was to estimate the model using the complete data set. However, there seems to exist a widely accepted point of view, according to which screening of the non-prime borrowers by primary lenders had become more lax at some point during the last decade and these lax standards were in place for some time, before they were tightened in the wake of liquidity crunch of the second half of 2007¹¹. This change in the lending practices might have resulted in the observationally different groups of borrowers: one group, who obtained their mortgages under what we hereafter call 'old' regime might have, for example, higher average FICO score than the group who were extended loans under the 'new' regime. Since FICO score usually is found to be a significant factor in reduced-form model of default, one may expect higher estimated default hazard for the latter group. However, naive comparison of default hazards in the first 36 months in 'mortgage time' of, for example, cohort originated in Q1 1997 and cohort originated in Q1 2005 (supposing that the former has higher average FICO) wouldn't allow one to distinguish the effect of observable characteristics from the effect of unobservables. If one could somehow observe

¹¹See, among others, Mayer, Pence, and Sherlund (2010), Dell'Ariccia, Igan, and Laeven (2008) for analysis of dynamics of various aspects of the lending standards.

the behavior of 1997 cohort in the first three years of the term of the mortgage, had this cohort been placed in the economic environment of 2005-2007, then one might decide whether the two cohorts were intrinsically different (that is different in their hidden properties), or they were just two random draws from the same population that happened to have some variation in some observable features, such as credit score.

Thus our working hypothesis is that there was some unseen shift in the behavior (or, equivalently, hidden characteristics¹²) of the subprime and Alt-A borrowers has occurred at some point during our observation period. This shift might have been responsible (at least, partly) for the huge surge in non-prime default during 2007-2009. I pursue two goals in my empirical analysis:

- Detect whether there were two distinct regimes in 1997-2009 and whether the regime change can be attributed to the change in the intrinsic nature of non-prime borrowers.
- If the unseen nature of the borrowers had changed during the observation period, how much of the increase in default could be associated with that change and how much – with other factors (change in observable characteristics, worsening economic environment).

Instead of estimating the default and prepayment models for the complete panel, I divide the data into 2 subsamples: one that corresponds to the 'old' regime and one that corresponds to the 'new' regime¹³. Initially I used mortgages originated in 1997-2003 for the estimation of the intensity models of termination under the 'old' regime and those

¹²This equivalence is based on assumption that borrowers' behavior is the product of their hidden characteristics and so is effectively set at the time of origination, though the consequences of their inherent behaviors, as regards mortgage termination, may only be seen much later).

¹³Figure 3.6 illustrates the difference between the average default rates in the first 3 years of the contract term for the 2 cohorts belonging to different regimes.

originated in 2006-2008 – for the models of the 'new' regime¹⁴ (the subsample of mortgages originated in 2004-2005 was excluded). To begin with, I estimated the (unstratified) default model for these two subsamples and obtained the initial parameter vectors $\tilde{\beta}_{old}^d$ and $\tilde{\beta}_{new}^d$. I conducted Wald test for the equality of the old and new parameter estimates and found that null hypothesis of equality was soundly rejected. Next, I estimated proportional hazard model of default on the subsamples of mortgages originated during calendar months of 2004-2005 period. For each calendar month I estimated two models: one with parameters restricted to be equal to $\tilde{\beta}_{old}^d$ and one with parameters set at $\tilde{\beta}_{new}^d$. I then compared goodness-of-fit of the two models¹⁵ and found that the first ('old') model steadily dominated the second one until August 2004. In September 2004 the 'new' model fit the data marginally better, while beginning from October 2004 the dominance of the 'new' model became statistically significant at 5% level and remained so until December 2005. Having had established Fall 2004 to be the time of the potential regime change, I then re-estimated the unstratified models of default using subsamples of mortgages originated during Q1 1997 - Q2 2004, and Q1 2005 - Q1 2008, for the $\tilde{\beta}_{old}^d$ and $\tilde{\beta}_{new}^d$ respectively. I repeated the procedure described above for the monthly subsamples of mortgages originated in July 2004 - December 2004 and reached essentially the same conclusions (results are reported in the table 3.10).

Estimated parameters of the default intensity models for the two regimes are reported in the tables 3.11 and 3.13. Straightforward comparison of the parameter estimates sug-

¹⁴Extant literature provides some indirect evidence as to the timing of possible shift which was brought about by rapidly deteriorating underwriting standards (see, e.g. Demyanyk and Van Hemert (2009)). It seems to be a wide-held belief that by 2006 lending practices had substantially deteriorated.

¹⁵The intuition behind this procedure is that any changes in unobserved characteristics will be for the most part absorbed in the baseline. The estimates $\tilde{\beta}_{old}^d$ correspond to the 'average' baseline hazard under the hypothetic 'old' regime, while $\tilde{\beta}_{new}^d$ – to that under the 'new' regime (as we have more or less firm priors that characteristics of the 'new' regime already established by 2006). Thus, by comparing the fit of models where $\hat{\beta}^d$ are set to either 'old' or 'new' estimates we are effectively comparing whether the shape of the baseline for a given month corresponds more to the estimated 'old' baseline hazard or to the estimated 'new' baseline hazard.

gests that the 'new' regime is likely to be characterized by increased riskiness. For example, parameter coefficients at *CurrentLtv* and *Ltv80Orig* are greater under the 'new' regime, estimate of the effect of the house price volatility becomes significant. Sensitivity to the current contract rate is also estimated to be greater, as are the coefficients at *LoDocLoan* and *InvestLoan*. Graphical representation of the default hazards under the 'old' regime and under the 'new' regime presented by figure 3.3 confirms the above intuition. Default hazard is estimated for a randomly picked contract using actual values of dynamic covariates observed in 2004-2008 and Nelson-Aalen estimates of baseline hazards from the unstratified default intensity models for the two regimes. One can observe that the increase in the default hazard around duration=24 months is much more steep under the 'new' regime, which can be attributed to the increased proportion of '2/28' ARMs. Importantly, the behavior under the two regimes diverge at durations greater than 3 years. The hazard under the second regime does not exhibit signs of decline, unlike the 'old' regime hazard. To quantify these intuition, I perform the following experiment: I calculate the hazard of default for a randomly chosen contract separately for each regime, using empirical estimates of the baseline hazards of default and prepayment obtained from non-stratified model for the respective regime. The results of the estimation are reported in Panel A of table 3.15. It appears that borrowers in the 'new' regime are substantially more risky judging by the value of the cumulative hazard (increase to 24% compared to 11% in the 'old' regime). However, if the values of the house prices are held constant, the increase in default hazard is much less pronounced (15% instead of 24% for the actual covariate values) as the results in Panel B of table 3.15 indicate. Thus, increase in the default incidence due to unobservable shift in borrowers' behavior (regime change) seems to be of the secondary importance compared to the effect of the changes in the economic environment (house prices).

Parameter estimates for prepayment models (tables 3.12 and 3.14) are largely as expected. As with default models, one can observe increased sensitivity to contempora-

neous LTV and current contract rate under the second regime. The estimate of the credit score parameter has also become larger in magnitude. As far as teaser effects, the dynamic covariate (*TeaserI*) estimate is smaller, but the effect of magnitude of teaser (*TeaserM*) in prepayment is more pronounced under the second regime. The estimate of *D24Mon* is also smaller under the second regime, which can be explained by inability of borrowers to refinance loans with 2-year teaser period in the unfavorable house price environment and increased incidence of default for these loans. The estimate of the *LoDocLoan* parameter has dropped from 1.35 to 0.37 (corresponding decrease in hazard ratio from 3.86 to 1.45) under the new regime, while the estimate of *RefiLoan* has increased from 0.10 to 0.25 (increase in hazard ratio by 17 percentage points); these effects may be due to seizing up of the subprime market and to reversal of the trend in underwriting in the recent years, which made refinancing into another subprime loan very problematic. Thus recent refinancings are likely to be into conventional or FHA loans, either of which requires borrower to provide documentation about his income and assets. One unexpected sign is that of *HPriceVol* estimate under the 'new' regime. This negative sign may either indicate problems with volatility models specification or reflect expectations of the protracted decline in house prices that hinder household mobility.

Visual examination of the cumulative baseline hazards of default and prepayment (figures 3.4 and 3.5) suggests that these are indeed sufficiently different – not only across the two regimes but within the same regime as well. This graphical evidence seems to support, at least indirectly, treatment of the quarterly baselines as realizations of the stochastic baseline default and prepayment processes, which will be the subject of the analysis in one of the subsequent sections.

3.2 ESTIMATION OF AFFINE DIFFUSION-DRIVEN TERM STRUCTURE MODEL

Several approaches to the estimation of the term structure have been developed to date and this is still an area of an ongoing research. In earlier studies the unobserved factor

(factors) were often approximated by observable variables. This approach was pursued, e.g., by Chan, Karolyi, Longstaff, and Sanders (1992) who used a yield on one-month Treasury bill as a proxy for an unobserved spot rate and derived moment conditions for a GMM estimation of diffusion parameters from the Euler discretization of the SDE 2.42. Titman and Torous (1989) were among the researchers, who employed a set of cross-sectional restrictions implied by the term structure to estimate the parameters of an underlying stochastic process. Their technique relied on minimization of the sum of squared deviations between model and market prices for a set of Treasuries for each date in the observation window for which the latter prices were quoted. Yet another approach, represented for example by Gibbons and Ramaswamy (1993), also relies on GMM for estimation. Conditional moment restrictions are based on a stationary density of the state (they used returns on T-bills of different maturities, which were assumed to be observed with errors). Researchers who preferred ML methods for estimation of diffusion parameters typically had to deal with the situation in which the number of observed maturities N exceeds the number of latent factors in the term structure. Different solutions have been proposed. Some, for example Chen and Scott (1992), posited that J bonds are priced exactly, while remaining $N - J$ are priced by the model up to an additive error (which in the cited paper was assumed to follow first-order autoregressive process). Typically errors are assumed to be independent of state process. Others, like Pearson and Sun (1994), used only J yields, thus circumventing the problem. The above estimation strategy, however, rests on the implicit assumption that state variable remains affine under $\mathbb{Q}$, which, in turn, imposes certain limitations on the functional form for the market price of risk process. Simulated maximum likelihood, on the other hand does not require such assumptions. However, this approach (for example, Brandt and Santa-Clara (2002)) is computationally demanding even for today's abundant computing power. Simulated maximum likelihood as implemented in Brandt and Santa-Clara (2002) is based on substituting Euler discretization transition density (essentially Gaussian) for

the unknown (in general case) transition density of diffusion. It entails constructing large number of sample paths of the discretized continuous-time process in order to achieve sufficient degree of accuracy in the Monte Carlo approximation of the true transition density. This approach is related to Bayesian methods (see Johannes and Polson (2005) for a relatively recent review of this literature, which is growing rapidly as of time of writing), as shown in Elerian, Chib, and Shephard (2001).

Bayesian methods (most prominently, MCMC) represent another way to deal with the problem of parameter estimation for continuous-time models. Durham and Gallant (2002) suggest a way to reduce the computational burden of simulated maximum likelihood using importance sampling. Several other approaches include approximating transition density by Hermite polynomials and subsequent ML estimation of the parameters (Aït-Sahalia 2002) moment-based approach (efficient method of moments), which uses score function from the auxiliary model obtained by Hermite polynomial approximation to construct a GMM-type criterion function and martingale estimating functions (see, for example, Kessler and Sørensen (1999)).

Even this short and incomplete overview of methods designed for the purpose of continuous-time models parameter estimation gives some impression of the amount of research effort that was devoted to solving this problem. For a good survey of literature related to term structure estimation see, e.g., Dai and Singleton (2003). For the task at hand I chose the approximate maximum likelihood based on state-space representation and filtering of unobservable states as, for example, in Chen and Scott (2003) or Duan and Simonato (1999). In this framework all N yields are assumed to be priced with error. The observation equation is 2.47, in the vector form it can be re-written as:

$$P_t(Y_t; \Psi, T) = A_0(\Psi, T) \exp(-B(\Psi, T) Y_t) \quad (3.2)$$

where model parameters are packed in column vector Ψ and subscript now stands for the time of the observation. Instantaneous time t rate r_t is the limit of the negative log of a

bond price as its maturity approaches 0:

$$r_t(Y_t; \Psi) = -\lim_{T \rightarrow 0} \frac{\log P_t(Y_t; \Psi, T)}{T} \quad (3.3)$$

and continuously compounded yield-to-maturity of a zero-coupon bond is accordingly:

$$R_t(Y_t; \Psi, T) = -\frac{\log P_t(Y_t; \Psi, T)}{T} \quad (3.4)$$

Under the assumption of the presence of observational errors, the expression for the yield takes the following form (*cf* Duan and Simonato (1999, (5,6))):

$$R_t(Y_t; \Psi, T) = -\frac{\log A(\Psi, T)}{T} + \frac{\log B(\Psi, T)}{T} Y_t + \epsilon_t \quad (3.5)$$

where ϵ_t is i.i.d. $N(0, \sigma_\epsilon)$. Yields of bonds of different maturities can be stacked:

$$\begin{bmatrix} R_t(Y_t; \Psi, T_1) \\ \vdots \\ R_t(Y_t; \Psi, T_N) \end{bmatrix} = \begin{bmatrix} -\log A(\Psi, T_1)/T_1 \\ \vdots \\ -\log A(\Psi, T_N)/T_N \end{bmatrix} Y_t + \begin{bmatrix} (B(\Psi, T_1)/T_1) \\ \vdots \\ B(\Psi, T_N)/T_N \end{bmatrix} \begin{bmatrix} \epsilon_{T,1} \\ \vdots \\ \epsilon_{T,N} \end{bmatrix} \quad (3.6)$$

The transition equation for the state-space model is based on the discretization of the state processes. Let h be the length of a discretization step and denote conditional mean of a (vector) state process by $m(Y_t; \Psi, h)$ and conditional variance by $\Phi(Y_t; \Psi, h)$. The transition equation for a discretized process is then given by:

$$Y_{t+1} = m(Y_t; \Psi, h) + [\Phi(Y_t; \Psi, h)]^{\frac{1}{2}} \eta_{t+1} \quad (3.7)$$

where η_{t+1} is the vector of i.i.d $N(0, 1)$ disturbances and $\Phi^{\frac{1}{2}}$ stands for Cholesky decomposition of conditional variance matrix Φ . In the Gaussian case first two moments characterize the transition distribution completely, but for non-Gaussian Cox-Ingersoll-Ross model the transition distribution is approximated therefore the Kalman filter is only quasi-optimal. Hence the parameter estimation is based on quasi-maximum likelihood. Parameter estimates are approximately asymptotically normal:

$$\sqrt{\mathcal{T}}(\hat{\Psi}_{\mathcal{T}} - \Psi_0) = N(0, \hat{F}_{\mathcal{T}}^{-1} \hat{G}_{\mathcal{T}} \hat{F}_{\mathcal{T}}^{-1}) \quad (3.8)$$

where $\mathcal{T}$ is the number of periods (observations in the series), and matrices $\hat{F}_{\mathcal{T}}$ and $\hat{G}_{\mathcal{T}}$ are given by the following expressions:

$$\hat{F}_{\mathcal{T}} = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} f_t(\hat{\Psi}_{\mathcal{T}}; \mathbf{R}_t) \quad (3.9)$$

where $\mathbf{R}_t = [R_t(Y_t; \Psi, T_1), \dots, R_t(Y_t; \Psi, T_N)]'$ and

$$\hat{G}_{\mathcal{T}} = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \frac{\partial \log l_t(\hat{\Psi}_{\mathcal{T}}; \mathbf{R}_t)'}{\partial \Psi} \frac{\partial \log l_t(\hat{\Psi}_{\mathcal{T}}; \mathbf{R}_t)}{\partial \Psi} \quad (3.10)$$

l is approximated likelihood function and function f depends on conditional mean μ_t and conditional variance Ω_t , which are output of the Kalman filter:

$$f_t(\hat{\Psi}_{\mathcal{T}}; \mathbf{R}_t) = \frac{\partial \mu_t'}{\partial \Psi} \Omega_t^{-1} \frac{\partial \mu_t}{\partial \Psi} + \frac{1}{2} \frac{\partial \Omega_t'}{\partial \Psi} (\Omega_t^{-1} \otimes \Omega_t^{-1}) \frac{\partial \Omega_t}{\partial \Psi} \quad (3.11)$$

Estimation of the model was carried out with N (number of monthly time series of zero-coupon yields) equal to five: yields on 3-month, 6-month, 12 month T-bills, 5-year and 10-year T-bonds. Series data were obtained from CRSP (Fama-Bliss Discount Bond files) and J. Huston McCulloch website at Ohio State¹⁶. Estimation period was taken to be January 1970 to December 2008 (468 observations). Figure 3.7 provides a graphical display of the series of yields, whereas descriptive statistics are presented in table 3.16. The range of the interest rates in the analyzed period spanned the interval from 1.5% to more than 16% with shorter yields being more volatile and their empirical distribution deviating farther from normality. The results of the estimation are summarized in table 3.17. Parameters of the first latent factor are estimated more precisely, than that of the second factor, however, even the imprecise estimates of the latter indicate that stationarity might be an issue for the second factor. Duan and Simonato (1999) and Chen and Scott (2003) reported similar imprecise estimates for the mean-reverting and long-run mean parameters of the second latent factor. The long-run mean of the first factor (θ_1) is estimated

¹⁶<http://www.econ.ohio-state.edu/jhm/ts/ts.html>. 10-year yield series were added to Fama-Bliss data since it is commonly believed that rates on long T-bonds and swaps provide better benchmark for conventional mortgage rates than shorter instruments' rates.

to be slightly less than 3%, which is close to the values reported in the most empirical studies investigating multi-factor time-homogeneous CIR model of the term structure. Estimated standard deviations of pricing errors for all maturities are reasonably small. Half-life ($\log 2/\kappa$) of the shocks to the first factor is estimated to be about 10 months, indicating strong degree of mean reversion¹⁷. Negative sign of the market price of risk α implies positive risk premium. As with other parameters, the precision of the estimate of the market price of risk for the second factor is lower than for the corresponding parameter of the first factor (this likewise applies to the combination $\kappa + \alpha$, which estimate for the second factor is negative). Relatively poor precision for the second factor estimates is probably due to the limitations imposed by use of only first two conditional moments in the filtering of the distribution which is likely asymmetric, as figure 3.10 may suggest, and/or limitations of the chosen term structure that are discussed, for example, in Rebonato (1996). Figure 3.10 presents graphs of smoothed Kalman filter estimates $\hat{\eta}_t$ and $\hat{\epsilon}_t$ of, respectively, state and response innovations. One can observe the histogram of former is almost symmetric, albeit leptokurtic, while the latter are appreciably left skewed. Gaussian response innovations don't seem to be particularly well suited to fit the behavior of rates in the periods of possible regime change (such as early 80s). Latent state variable estimates are rendered in figure 3.9: as expected, the output from the Kalman smoother is fairly close to the observed yield on the shortest maturity¹⁸.

In a way of additional diagnostics of the model, figure 3.11 displays the factor loadings which are calculated from 2.46. In the low-dimensional term structure models the parameter coefficients on the state variables are conventionally interpreted as the ones reflecting the first N principal components of interest rate (e.g. first component - the level factor, second component - the slope factor and so on). One can observe fairly typical patterns: almost flat for the level and convex to the origin - for the slope factor. Finally, figure

¹⁷Half-life can be interpreted as the half of the time which it takes for the process to return to its long-run mean level after the shock.

¹⁸I didn't use available yield on 1 month T-bills as this is widely considered to be more noisy than 3-month yield.

3.8 illustrates several paths of interest rates which were produced by simulation using estimated parameters of the term structure.

In general, results of the estimation of the term structure appear to be consistent with the findings of the extant research. Parameter values will be used further in the simulations of the discount rate (spot rate) in (2.17), projected ARM rates and house values.

3.3 THE HOUSE PRICE PROCESS: THE MODEL AND THE ESTIMATION STRATEGY

3.3.1 THE MODEL OF THE HOUSE PRICE PROCESS

We build our model on the intuition presented in (2.52) accounting, however, for frictions existing in the real world. Specifically, we assume that the ratio of the return on housing asset to the return on alternative financial asset, instead of being time-invariant, tends to some steady state level, but being continually disturbed by random shocks, it evolves as a mean-reverting stochastic process. As far as the choice of financial asset is concerned, our motivation included desire to maintain the inverse relationship between the changes in interest rates and the changes in house prices. More formally, we again consider the bond process $p(\cdot)$ of the previous Chapter (2.47). We further assume that the discount bond with maturity T is continuously rolled over (at each moment a long position in the 'old' bond is being liquidated and a long position in the 'new' bond with the same maturity T is opened). Returns from such investment strategy are given by $\tilde{p}(\cdot) \exp(\int_0^t [R(u; \cdot) - \varphi_b(u)] du)$ where $\varphi_b(u)$ is the price of risk associated with bond process, $R(\cdot)$ is continuously compounded yield on the bond ($\tilde{p}$ is used instead of p to denote the bond that is being continuously reinvested). Returns on investment in housing asset are given by $H(\cdot) \exp(\int_0^t [s - \varphi_H(\cdot)] du)$, where φ_H is the price of risk associated with house price process and s is the flow of services from the housing asset expressed as a percentage of its value. Next, we introduce the 'relative return' process of the form (under

the risk-neutral measure $\mathbb{Q}$):

$$q(t) = \frac{H(t) \exp(\int_0^t [s - \varphi_H(u)] du)}{\tilde{p}(t; \cdot) \exp(\int_0^t [R(u; \cdot) - \varphi_b(u)] du)} \quad (3.12)$$

We posit the mean reverting form for the 'relative return' process under the physical probability measure $\mathbb{P}$:

$$\frac{dq(t)}{q(t)} = \kappa_q(\theta_q - q(t))dt + \sigma_q(\cdot)dW_q^{\mathbb{P}}(t) \quad (3.13)$$

where we assume that the Brownian motion W_q is orthogonal to the Brownian motions driving latent factors y_1, y_2 of the term structure. Under the pricing measure $\mathbb{Q}$, the SDE for the q process incorporates the price of risk for the q process:

$$\frac{dq(t)}{q(t)} = (\kappa_q\theta_q - \kappa_qq(t) - \varphi_q(t))dt + \sigma_q(\cdot)dW_q^{\mathbb{Q}}(t) \quad (3.14)$$

Re-writing (3.12) as $q(t) = \frac{H(t)}{\tilde{p}(t; \cdot)} \exp(\int_0^t [s - \varphi_H(\cdot) - R(u; \cdot) + \varphi_b(u)] du)$, taking logs and differentiating, we obtain:

$$\frac{dq(t)}{q(t)} = \frac{dH(t)}{H(t)} - \frac{d\tilde{p}(t; \cdot)}{\tilde{p}(t; \cdot)} + s - \varphi_H(t) - R(t; \cdot) + \varphi_b(t) \quad (3.15)$$

Given our definition of $\tilde{p}(\cdot)$ the bond process $\frac{dp(t; y_1(t), y_2(t))}{p(t; y_1(t), y_2(t))}$ can be expressed as $\frac{dp(t; y_1(t), y_2(t))}{p(t; y_1(t), y_2(t))} = \frac{d\tilde{p}(t; y_1(t), y_2(t))}{\tilde{p}(t; y_1(t), y_2(t))} + R(t)$. Hence, $\frac{dH(t)}{H(t)}$ under risk-neutral measure can be expressed via $\frac{dq(t)}{q(t)}$

as:

$$\frac{dH(t)}{H(t)} = \frac{dq(t)}{q(t)} + \frac{dp(t; \cdot)}{p(t; \cdot)} - (s - \varphi_H(t) + \varphi_b(t)) \quad (3.16)$$

Equivalently SDE for house process can be written as:

$$\begin{aligned} \frac{dH(t)}{H(t)} + (s - \varphi_H(t)) &= (\mu_q(t) - \varphi_q(t))dt + \sigma_q(t)dW_q^{\mathbb{Q}}(t) + \\ &(\mu_b(t; \cdot) - \varphi_b(t; \cdot))dt + \sigma_b(t; \cdot)dW_b^{\mathbb{Q}}(t) \end{aligned} \quad (3.17)$$

where $\sigma_b(t; \cdot)$ is diffusion function for the bond process¹⁹, $\mu_q(\cdot)$ and $\mu_b(\cdot)$ are drifts of the q and bond processes, respectively, under the real-world measure $\mathbb{P}$. From (3.17) we deduce

¹⁹Since we have 2-factor term structure with independent factors, drift and diffusion of the bond process are functions of two term structure factors. See Appendix for the explicit expressions for drift and diffusion in terms of latent factors of the term structure.

that $\varphi_H(t) = \varphi_q(t) + \varphi_b(t; \cdot)$). Operationally, we assume Cox, Ingersoll and Ross form for the SDE (3.13) and the affine form for φ_q so that square-root diffusion form remains the same under the risk-neutral measure. We calibrate $\varphi_q = \varphi_{01} + \varphi_{11}q$ to the data so as to facilitate empirical estimation of the parameters governing the q process.

The risk adjustment for the bond process is determined by the form of the price of risk processes for the latent factors of the term structure. Since the factors are independent, the price of risk for the bond is the sum of the two terms for the latent factors:

$$\varphi_b = B_1(t, T)\alpha_1 y_1(t) + B_2(t, T)\alpha_2 y_2(t) \quad (3.18)$$

where B_j s are given by (2.47). In the next section the data and the estimation process are described in greater detail.

3.3.2 DATA AND RESULTS OF THE HOUSE PRICE PROCESS ESTIMATION

We use monthly Case-Shiller house price appreciation (HPA) series for the estimation of house process. Figure 3.12 presents graphs of these series for selected regions. Certain shortcomings (such as smoothing of actual price movements in the aggregated indices) associated with estimation of the parameters of the house process from aggregated data have been discussed in the literature (Downing, Stanton, and Wallace (2007)), however, it should be noted that Case-Shiller index is likely to suffer from smoothing less than other aggregated indices (such as 'classic' OFHEO), since it doesn't rely on appraised values. Figure 3.16 and table 3.18 provide some information in regard to degree of spatial correlation for selected series. Diffusion index is simply an indicator which takes the value of 1 if house prices in a given region and a given period increased and 0 if prices decreased. Common seasonal component in house price variation notwithstanding, it is clearly seen that price movements in largest MSAs used to have a lot in common in the observed period, especially in the most recent years. Table 3.18 reports correlation coefficients for HPA rate in 20 MSAs. The strength of correlation typically decreases with the distance

between the MSAs (Detroit was chosen as an example of MSA that deviated from the common trend due to regional specific factors being most pronounced).

For the estimates of house service flow we use median house price values from American Community Survey. We divide 'owner's equivalent rent of primary residence' reported by Bureau of Labor Statistics for each of the 20 MSAs by the median house price value for that region. We correct for the maintenance expenses incurred by an owner using estimates of the share of these expenses reported by BLS. At the first stage of the estimation we calibrate the values of φ_{01} and φ_{11} so that resulting empirical series of q appear relatively flat. We use series of yields on zero-coupon bonds with 10-year maturity estimated by J. Huston McCulloch²⁰ to construct empirically observed q series. As a next step, we attempt to estimate the parameters of the q process by maximum likelihood; often the optimization procedure would not converge – in this case the parameters φ_{01} and φ_{11} are re-calibrated and procedure is repeated until convergence is achieved. Results of the maximum-likelihood estimation of parameters κ_q , θ_q , σ_q are reported in tables 3.19 and 3.20. Estimates of the mean-reversion speed parameter κ_q vary in the range from 0.03 to 0.15 which implies the half-life of the process in the range from about 5 months to 23 months. Volatility estimates vary from 0.05 to 0.24 with house prices in the "sandy states" (CA, FL, AZ, NV) exhibiting more variability than house prices in the inland states. Figure 3.14 presents graphs of the mean of a sample from simulated house process for selected regions. Sample paths of the house process (under the risk-neutral measure) are backed out given simulated paths of the q -process and simulated bond prices. Market price of house risk, as it was indicated in the previous section, is derived from the prices of risk for bond and relative return processes.

Importantly, the effect of house price appreciation (depreciation) in the model is transmitted not only through changes in the level of house prices that translate into changes of borrower's equity but also through the variability of house price. This latter effect can be

²⁰<http://www.econ.ohio-state.edu/jhm/ts/ts.html>

though of as primarily influencing borrower's expectation about future course of house prices and hence the value of embedded options to default on the loan or to prepay. There are examples in the extant literature (e.g. Miller and Peng (2006)) when volatility estimates obtained from the GARCH model fit to HPA index were used in the VAR analysis of volatility, home appreciation, personal income dynamics, population, unemployment and gross regional product. Miller and Peng (2006) concluded that Granger causation runs from home appreciation to volatility. In our model house volatility is dependent both on interest rates volatility and idiosyncratic regional factor estimated from the historical data. We include the covariate *HPriceVol*, which comes from univariate GARCH model²¹ as the proxy for historical volatility²² in our hazard-based analysis of conditional probability of mortgage termination. We extract the series of conditional standard deviation for each of the 20 regions from the estimated GARCH model for the respective region²³. A variety of GARCH specifications have been explored: in particular, the hypothesis of asymmetric effects of the volatility shocks on the house price appreciation as captured by models with leverage effects (such as E-GARCH or T-GARCH) did not find strong support in the sample. Two-component model for conditional volatility (conditional variance represented as a sum of a persistent and a short-run components) also was rejected by the data. Final specifications are presented in tables 3.21 and 3.22. ARMA terms in conditional mean were selected based on considerations of parsimony, given the values of Akaike and Schwartz information criteria. In each of the 20 cases GARCH effects are estimated to be highly statistically significant.

In the next section I address modeling of the latent processes that represent a stochastic part of a probability of a mortgage termination²⁴.

²¹Parameter estimates are reported in Tables 3.21 - 3.22.

²²Nelson (1990) was the first to point out the link between certain diffusion processes and ARCH/GARCH processes.

²³Figure 3.15 provides an illustration of the estimated conditional standard deviation.

²⁴Although direct interpretation of the latent factors is somewhat complicated, they may plausibly be considered as reflecting changes in economic conditions in general and / or changes in the social norms and attitudes (e.g. towards personal bankruptcy or breaching the contract) over time.

3.4 PRACTICAL IMPLEMENTATION OF SMC ML ESTIMATION

I assume the following two specifications for both latent intensity of default and latent intensity of prepayment. The first specification is the square root (Cox-Ingersoll-Ross) diffusion for the baseline process l , described by the following SDE under the physical probability measure $\mathbb{P}$:

$$d\mathbf{l}_t = \mathbf{a}^{(1)}(\mathbf{l}_t, t) dt + \mathbf{b}^{(1)}(\mathbf{l}_t, t) d\mathbf{W}_t^{(1)} \quad (3.19)$$

where $\mathbf{l}_t = (l_t^d, l_t^p)'$, $\mathbf{W}_t^{(1)} = (W_t^{(1),d}, W_t^{(1),p})'$

$$\mathbf{a}^{(1)}(\mathbf{l}_t, t) = \begin{pmatrix} \kappa^{(1),d} (\mu^{(1),d} - l_t^d) \\ \kappa^{(1),p} (\mu^{(1),p} - l_t^p) \end{pmatrix}, \quad \mathbf{b}^{(1)}(\mathbf{l}_t, t) = \begin{pmatrix} \sigma^{(1),d} \sqrt{l_t^d} & \rho_{12}^{(1)} \\ \rho_{21}^{(1)} & \sigma^{(1),p} \sqrt{l_t^p} \end{pmatrix}$$

The second specification assumes that natural logarithm of the latent intensity follows mean-reverting Ornstein-Uhlenbeck (OU) process under the objective probability measure. The SDE for the log-intensity $\mathbf{x}_t = \log(\mathbf{l}_t)$ has the following form:

$$d\mathbf{x}_t = \mathbf{a}^{(2)}(\mathbf{x}_t, t) dt + \mathbf{b}^{(2)}(t) d\mathbf{W}_t^{(2)} \quad (3.20)$$

where $\mathbf{x}_t = (x_t^d, x_t^p)'$, $\mathbf{W}_t^{(2)} = (W_t^{(2),d}, W_t^{(2),p})'$

$$\mathbf{a}^{(2)}(\mathbf{x}_t, t) = \begin{pmatrix} \kappa^{(2),d} (\mu^{(2),d} - x_t^d) \\ \kappa^{(2),p} (\mu^{(2),p} - x_t^p) \end{pmatrix}, \quad \mathbf{b}^{(2)}(t) = \begin{pmatrix} \sigma^{(2),d} & \rho_{12}^{(2)} \\ \rho_{21}^{(2)} & \sigma^{(2),p} \end{pmatrix}$$

I proceed to estimate the parameters of these model under the assumption of zero correlation between Brownian motions driving each of the processes. I employ the following state-space representation:

$$P(y_{j,t_n}^\ell | \lambda_{j,t_n}^\ell) = \frac{1}{y_{j,t_n}^\ell!} (\lambda_{j,t_n}^\ell)^{y_{j,t_n}^\ell} \exp(-\lambda_{j,t_n}^\ell). \quad (3.21)$$

$p(\cdot)$ is the probability density for the n^{th} element of the observational vector where $\ell = d, p$. Here y_t denotes the observed number of events (defaults or prepayments) in the j^{th}

stratum in the time period t_n (n^{th} month since origination).

$$\lambda_{j,t_n}^\ell = \sum_{k=1}^{K(n,j)} \lambda_{k,t_n}^\ell = l_{j,t_n}^\ell \cdot \sum_{k=1}^{K(n,j)} \exp(\mathbf{z}_k(t_n)\beta^\ell) \quad (3.22)$$

is the cumulative value of intensity for the j^{th} stratum at time t_n . k is the index for individual contract with a covariate vector $\mathbf{z}$ in a given stratum. p here and in what follows is a generic symbol for a (conditional) probability distribution.

State vector is determined independently of the (3.21); for the former (CIR) specification the transition density of the state is given by:

$$p(l_t^\ell | l_s^\ell) \sim \text{n.c.}\chi^2(\nu, \eta) \quad (3.23)$$

where $\nu = \frac{4\kappa\mu}{\sigma^2}$, $\eta = \frac{4\kappa \exp(-\kappa(t-s))}{\sigma^2(1-\exp(-\kappa(t-s)))} \lambda_0(s)$, where $s < t$ and superscripts denoting the type of termination are omitted for notational convenience. For time-homogeneous CIR process ($\mu(t) = \mu$) the condition $2\kappa\mu > \sigma^2$ is assumed to hold.²⁵ The state equation for the second (log-OU) specification takes the form:

$$p(l_t^\ell | l_s^\ell) \sim \text{LN}(\nu, \eta) \quad (3.24)$$

where ν and η are, respectively, mean and variance of the log-normal distribution with parameters of corresponding normal distribution given by $\nu^* = \lambda_0(s) \exp(-\kappa(t-s)) + \mu(1 - \exp(-\kappa(t-s)))$, $\eta^* = \frac{\sigma^2}{2\kappa} (1 - \exp(-2\kappa(t-s)))$.

Sequential methods offer a possibility to filter out latent state and estimate the parameters in (3.23 or 3.24).

The choice of predictive likelihood relies on t distribution with low degrees of freedom $\tilde{g}(\cdot) \sim t_5(\alpha, \beta)$ as approximation of likelihood when the number of observed counts is significant (greater than 7). Location parameter α and scale parameter β of this t -distribution

²⁵For time-inhomogeneous CIR process if one is willing to accept a *constant dimension* condition, i.e. assume that process obeys $\frac{\kappa\mu(t)}{\sigma^2(t)} = \text{constant}$, the transition density still remains tractable. This latter model is treated for example in Musiela and Rutkowski (2005).

are chosen to equal, respectively, mean and variance of the respective mixed Poisson distribution²⁶. For example, for CIR specification (noncentral χ^2 mixed Poisson) these quantities are, respectively:

$$\alpha^i = \hat{x}_s^i \exp(-\kappa(t-s)) + \mu(1 - \exp(-\kappa(t-s))) \quad (3.25)$$

$$\beta^i = \hat{x}_s^i \frac{\sigma^2}{\kappa} [\exp(-\kappa(t-s)) - \exp(-2\kappa(t-s))] + \frac{\mu\sigma^2}{2\kappa} (1 - \exp(-\kappa(t-s)))^2 + \alpha^i \quad (3.26)$$

for any $s < t$.

For log-OU specification (where state x is log-intensity) the location and scale are given by

$$\alpha^i = \hat{x}_s^i \exp(-\kappa(t-s)) + \mu(1 - \exp(-\kappa(t-s))) \quad (3.27)$$

$$\beta^i = \frac{\sigma^2}{2\kappa} (1 - \exp(-2\kappa(t-s))) + \alpha^i \quad (3.28)$$

Our sample for the estimation consists of 1868 stratum*duration observations (1429 observations for the 'early' model and 439 observations for the 'recent' model). To minimize numerical difficulties arising due to the presence of extremely noisy observations we used only stratum*durations for which number of contract at risk was no less than 30. The number of discrete support points was chosen to be 5,000. We apply off-line procedure for maximum likelihood estimation of parameters of latent intensities using the output of the particle filter. Initial distribution for the particles was chosen to be uniform centered around small positive value (no less than 3 orders of magnitude less than median of empirical Nelson-Aalen estimates²⁷). Since vector of observed counts is sorted in the chronological order of strata, the filter has to be initialized at the beginning of each new stratum. The value of likelihood is approximated as $\hat{p}_\theta(\mathbf{y}_{1:T}) = \hat{p}_\theta(y_1) \prod_{n=2}^T \hat{p}_\theta(y_n | \mathbf{y}_{1:n-1})$. Smooth resampling routine suggested by Pitt (Pitt 2002) is used to alleviate problems caused by discontinuities in empirical CDF. While filter likelihood $\hat{p}_\theta$ is unbiased, the

²⁶Mixing distribution is either log-normal or noncentral chi-square one, which parameters are determined by values of κ , μ and σ in (3.20) or (3.19).

²⁷Experiments with different initial densities, including asymmetric ones, proved that results of estimation are robust to the choice of initial distribution as long its upper decile lies below mean and medial of the empirical distribution of Nelson-Aalen estimates.

approximate log-likelihood is not, therefore we resort to standard bias correction to $\log \hat{p}_\theta(\mathbf{y}_{1:T})$ as in Pitt (2002). We approximate gradient of likelihood numerically by central differences and use combination of Sequential Quadratic Programming and trust-region algorithms (MATLAB Optimization Toolbox) for numerical optimization of parameters. Unfortunately, the estimates of the parameters' standard errors obtained from the numerical approximation of Hessian are likely to be biased downwards²⁸. Tables 3.23 and 3.24 present ML estimates of the parameters of default and prepayment processes under specifications 3.20 and 3.19, respectively. For the former specification one can observe that volatility of the default process is greater under the first regime which is likely due to the fact that there are much more sparse strata and consequently, 0 counts. Hence the increase in the number of events by 5-10 for that period (given our pure diffusion specification, i.e. process without jumps) can be explained only by high volatility. On the contrary, the strata in the second period are all populous so that period-to-period variation of 20-30 defaults or more goes without much notice. The speed of mean reversion is substantially lower in the recent period that indicates that rise in defaults continues longer than it used to in a more quiet environment. The long-run mean and volatility of default process is somewhat lower under the new regime in the BK specification which decrease is more than compensated by elevated combined effect of observables. In the CIR case only the volatility under the new regime is lower than under the old one.

Figures 3.17 and 3.18 suggest that the fit of the log-OU specification for the latent intensities is better than that of the CIR specification. The latter tend to overpredict prepayment under both regimes and default under the new regime (as noted before, this is probably the side effect of elevated volatility that arises as a result of an attempt to fit the diffusion process to the fluctuations of high amplitude²⁹). On the contrary, default under the recent regime tends to be underpredicted. Pattern of the residuals for the log

²⁸Theorem 12.5.7 in Cappé, Moulines, and Rydén (2005) states that observed information at the MLE is a consistent estimator of Fisher information matrix.

²⁹Figure 3.19 illustrates the behavior of mean of one-step ahead prediction density for the two specifications. CIR paths seem to have more problems in capturing the sudden surges in default.

Ornstein-Uhlenbeck specification of the prepayment intensity is closer to the normal: still the highest peaks under both regimes are not captured well, as quantile-quantile plots illustrate. Default process under the 'old' regime appears to fit better (the estimated parameters misfit only periods with highest default intensity) while for the 'new' regime the model tends to overpredict lowest counts (typically associated with the first 3-4 months after origination) and underpredict very high intensities (above 300 defaults in a given month). More formally, null hypothesis of normality of transformed (Anscombe) residuals for both 'old' and 'new' models is rejected by Jarque-Bera test at 5% confidence level. Null hypothesis about the absence of autocorrelation in the transformed residuals for the first 6 lags is not rejected at 5% level for the default model under the 'old' regime by Ljung-Box test, but it is rejected for the prepayment models under both regimes and for the default model under the 'new' regime.

For the general goodness-of-fit test I have implemented the following procedure. I simulated 5,000 series from the process (3.23) or (3.24) with parameters reported in the table 3.24 and the table 3.23 for each stratum³⁰. I further calculated arithmetic mean of these series and weighted mean, with weights produced by the filtering procedure. These means were multiplied by factor $\sum_{k=1}^{K(n,j)} \exp(z' \hat{\beta}^k)$ from (3.22) for the respective stratum(j)*duration(n) combination. In this way, for each stratum*duration cell I obtained two values of expected counts $\hat{y}_{jn}$ and $\tilde{y}_{jn}$. I calculated 'grand' χ^2 Pearson statistic as

$$G_g = \sum_n \sum_j X_{jn}, \quad (3.29)$$

where $X_{jn} = \frac{(y_{jn} - \hat{y}_{jn})^2}{\hat{y}_{jn}}$, and its counterpart with X_{jn} replaced by $\tilde{X}_{jn} = \frac{(y_{jn} - \tilde{y}_{jn})^2}{\tilde{y}_{jn}}$. The procedure was repeated 100 times with different initializations of random number generator to guard against possible biases induced by the Monte Carlo procedure. The results (mean values over 100 Monte Carlo replications) are reported in the table 3.25. They are

³⁰For example, simulated intensity series for the first of 13 strata used in the estimation of the model for the 'new' regime are of length 54, series for the second stratum are of length 51 and so on

in line with the observations made based on informal graphical analysis of the residual plots: the log-Gaussian diffusion specification for both default and prepayment process performs better than the CIR specification. This specification cannot be rejected by the data at conventional significance level in all 4 cases (default and prepayment under both 'old' and 'new' regimes). In the subsequent analysis (calibration of the model) I intend to use only log-OU specification.

Table 3.1
Multiplicative intensity models: variable definitions

Variable name	Definition
<i>CurrentLtv</i>	Estimate of LTV ratio at the time of observation, % (house price at the time of observation is approximated using MSA-level HPA index).
<i>LTV80Orig</i>	1 if LTV at origination is exactly 80%, 0 otherwise.
<i>LoanSizeOrig</i>	Loan amount at origination, in constant (2000) million dollars.
<i>TeaserM</i>	Calculated difference between the regular contract rate and the 'teaser' rate, percentage points.
<i>TeaserI</i>	Interaction of the previous variable with a length of the time remaining until expiration of the 'teaser' rate.
<i>FicoScoreOrig</i>	FICO score of the borrower at origination (divided by 100).
<i>LoDocLoan</i>	Indicator variable that equals 1 for the loan with partial or no verification of borrower's income or assets.
<i>PrepPenaltyLoan</i>	Indicator variable that equals 1 for the loan encumbered by prepayment penalty.
<i>InvestLoan</i>	Indicator variable that equals 1 for the loan secured by the property other than primary residence.
<i>RefiLoan</i>	Indicator variable that equals 1 for the loan taken out for the purpose of refinancing.
<i>CurrContRt</i>	Contract rate at the time of observation, percentage points.
<i>CurrContRtSq</i>	Square of the above.
<i>SprToFRM</i>	Spread between the contract rate and the predicted FRM rate ³¹ (ratio of the difference between the two / predicted FRM rate).
<i>D24Mon</i>	Indicator variable that equals 1 at durations 24, 25, 26 months, 0 otherwise.
<i>SineSeason</i>	Seasonal trend (a sinusoid with maximum in June and minimum in December).
<i>HPriceVol</i>	Estimate of the house price volatility, % (based on the GARCH model for each region)

Table 3.2
Number of originated and terminated loans by the quarter of origination (the 'old' regime)

Year	Quarter	No. originated	No. prep.	No. def.	% prep.	% def.
1997	Q1	20	6	6	0.30	0.30
	Q2	37	14	11	0.38	0.30
	Q3	50	15	24	0.30	0.48
	Q4	65	28	22	0.43	0.34
1998	Q1	48	20	12	0.42	0.25
	Q2	73	31	19	0.42	0.26
	Q3	86	45	16	0.52	0.19
	Q4	37	14	9	0.38	0.24
1999	Q1	41	16	13	0.39	0.32
	Q2	53	17	15	0.32	0.28
	Q3	86	40	25	0.47	0.29
	Q4	98	45	29	0.46	0.30
2000	Q1	113	53	36	0.47	0.32
	Q2	157	86	31	0.55	0.20
	Q3	281	205	51	0.73	0.18
	Q4	266	124	72	0.47	0.27
2001	Q1	188	67	65	0.36	0.35
	Q2	326	113	119	0.35	0.37
	Q3	849	379	327	0.45	0.39
	Q4	813	398	266	0.49	0.33
2002	Q1	757	371	208	0.49	0.27
	Q2	1422	814	344	0.57	0.24
	Q3	3083	2146	556	0.70	0.18
	Q4	2271	1557	356	0.69	0.16
2003	Q1	1795	1078	319	0.60	0.18
	Q2	4697	3188	710	0.68	0.15
	Q3	7383	4854	831	0.66	0.11
	Q4	8694	6334	1212	0.73	0.14
2004	Q1	9679	6904	1297	0.71	0.13
	Q2	19041	13034	2850	0.68	0.15
Total		62509	41996	9851	0.67	0.16

Table 3.3
 Number of originated and terminated loans by the quarter of origination (the 'new' regime)

Year	Quarter	No. originated	No. prep.	No. def.	% prep.	% def.
2005	Q1	26608	15564	6220	0.58	0.23
	Q2	35753	18220	9797	0.51	0.27
	Q3	30017	12491	9718	0.42	0.32
	Q4	24995	8761	8603	0.35	0.34
2006	Q1	20305	5868	8082	0.29	0.40
	Q2	23773	5163	9770	0.22	0.41
	Q3	19280	3475	8031	0.18	0.42
	Q4	18789	2446	7707	0.13	0.41
2007	Q1	13854	1496	5494	0.11	0.40
	Q2	6674	618	2609	0.09	0.39
	Q3	1248	132	357	0.11	0.29
	Q4	161	11	21	0.07	0.13
2008	Q1	61		11		0.19
Total		221518	74245	76420	0.34	0.34

Table 3.4
Summary statistics for model variables and major loan characteristics

Variable	N	Mean	St. dev.	Min	Max
	Panel A.	Static variables.			
<i>OrigLTVRatio^a</i>	284027	85.67	6.60	80	108.89
<i>LTV80Orig</i>	284027	0.49	0.50	0	1
<i>LoanSizeOrig</i>	284027	0.22	0.13	0.02	1.80
<i>TeaserM</i>	284027	1.59	1.75	0	10.01
<i>FicoScoreOrig</i>	284027	630.6	56.55	374	720
<i>OrigContRate^b</i>	284027	7.46	1.62	1.00	18.62
<i>Margin^c</i>	230286	5.14	1.80	1.00	9.99
<i>Ceiling^d</i>	234643	12.76	2.69	2.58	25.34
<i>Floor^e</i>	188682	7.18	1.92	2.51	18
<i>LoDocLoan</i>	284027	0.67	0.47	0	1
<i>PrepPenaltyLoan</i>	284027	0.51	0.50	0	1
<i>InvestLoan</i>	284027	0.07	0.25	0	1
<i>RefiLoan</i>	284027	0.42	0.49	0	1
	Panel B.	Dynamic variables.			
<i>CurrentLtv</i>	6065604	0.91	0.19	0.01	1.54
<i>CurrContRt</i>	6185335	7.58	1.72	1.00	17.38
<i>SprToFRM</i>	6065604	-0.08	0.21	-0.97	1.35
<i>TeaserI</i>	6185335	12.11	27.59	0	443.4
<i>SineSeason</i>	6185335	0.01	0.71	-1	1
<i>HPriceVol</i>	6073575	0.07	0.03	0.01	0.23

^a - LTV at origination, ^b - contract rate at origination, ^c - margin over the index rate, ^d - lifetime cap on the contract rate, ^e - lifetime minimum for the contract rate.

Table 3.5

Summary statistics for model variables and major loan characteristics for the subset of defaulted loans

Variable	N	Mean	St. dev.	Min	Max
	Panel A.	Static variables.			
<i>OrigLTVRatio</i>	86271	85.94	6.71	80	108.89
<i>LTV80Orig</i>	86271	0.47	0.50	0	1
<i>LoanSizeOrig</i>	86271	0.21	0.12	0.02	1.64
<i>TeaserM</i>	86271	1.65	1.79	0	10.01
<i>FicoScoreOrig</i>	86271	614.6	60.62	374	720
<i>OrigContRate</i>	86271	7.92	1.53	1.00	18.62
<i>Margin</i>	73388	5.59	1.55	1.00	9.99
<i>Ceiling</i>	75537	13.41	2.54	2.58	25.34
<i>Floor</i>	66318	7.58	1.73	2.51	16.90
<i>LoDocLoan</i>	86271	0.62	0.48	0	1
<i>PrepPenaltyLoan</i>	86271	0.56	0.50	0	1
<i>InvestLoan</i>	86271	0.08	0.27	0	1
<i>RefiLoan</i>	86271	0.40	0.49	0	1
	Panel B.	Dynamic variables.			
<i>CurrentLtv</i>	1811732	0.93	0.18	0.03	1.54
<i>CurrContRt</i>	1813812	8.10	1.62	1.00	17.38
<i>SprToFRM</i>	1811732	-0.06	0.21	-0.97	1.35
<i>TeaserI</i>	1813812	11.73	27.19	0	421.3
<i>SineSeason</i>	1813812	0.00	0.71	-1	1
<i>HPriceVol</i>	1813812	0.07	0.03	0.02	0.23

Table 3.6
Summary statistics for model variables and major loan characteristics for the subset
of loans that were prepaid

Variable	N	Mean	St. dev.	Min	Max
	Panel A.	Static variables.			
<i>OrigLTVRatio</i>	116241	85.37	6.27	80	107.2
<i>LTV80Orig</i>	116241	0.48	0.50	0	1
<i>LoanSizeOrig</i>	116241	0.21	0.12	0.02	1.75
<i>TeaserM</i>	116241	1.53	1.74	0	9.70
<i>FicoScoreOrig</i>	116241	634.2	52.36	414	720
<i>OrigContRate</i>	116241	7.18	1.48	1	16.43
<i>Margin</i>	84139	5.20	1.78	1.00	9.99
<i>Ceiling</i>	86137	12.51	2.63	2.7	25
<i>Floor</i>	71684	6.96	1.84	2.51	17.83
<i>LoDocLoan</i>	116241	0.74	0.44	0	1
<i>PrepPenaltyLoan</i>	116241	0.51	0.50	0	1
<i>InvestLoan</i>	116241	0.06	0.23	0	1
<i>RefiLoan</i>	116241	0.45	0.50	0	1
	Panel B.	Dynamic variables.			
<i>CurrentLtv</i>	1722826	0.79	0.12	0.00	1.54
<i>CurrContRt</i>	1722902	7.18	1.60	1.05	16.38
<i>SprToFRM</i>	1722826	-0.04	0.18	-0.55	1.15
<i>TeaserI</i>	1722902	12.16	25.59	0	354.5
<i>SineSeason</i>	1722902	0.00	0.71	-1	1
<i>HPriceVol</i>	1722902	0.06	0.02	0.02	0.23

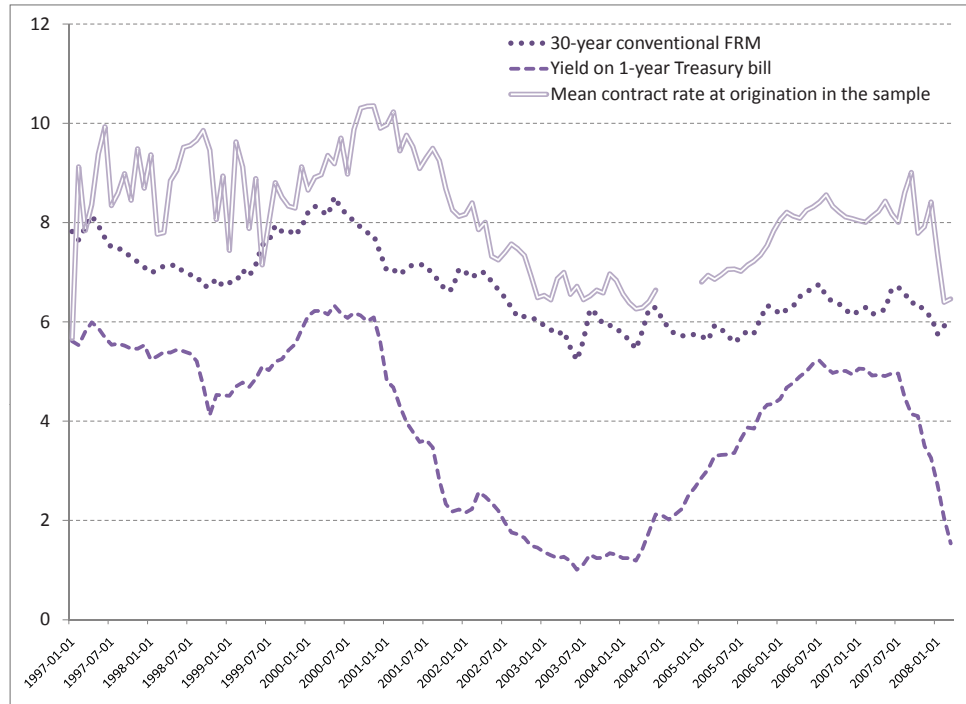


Figure 3.1

Mean contract rate, rate on conventional mortgages and T-bill rate during the observation window

Source for conventional mortgage rate and 1-year T-bill yield:

<http://research.stlouisfed.org/fred2/>. Mean mortgage rate in the sample is calculated by month of loan origination.

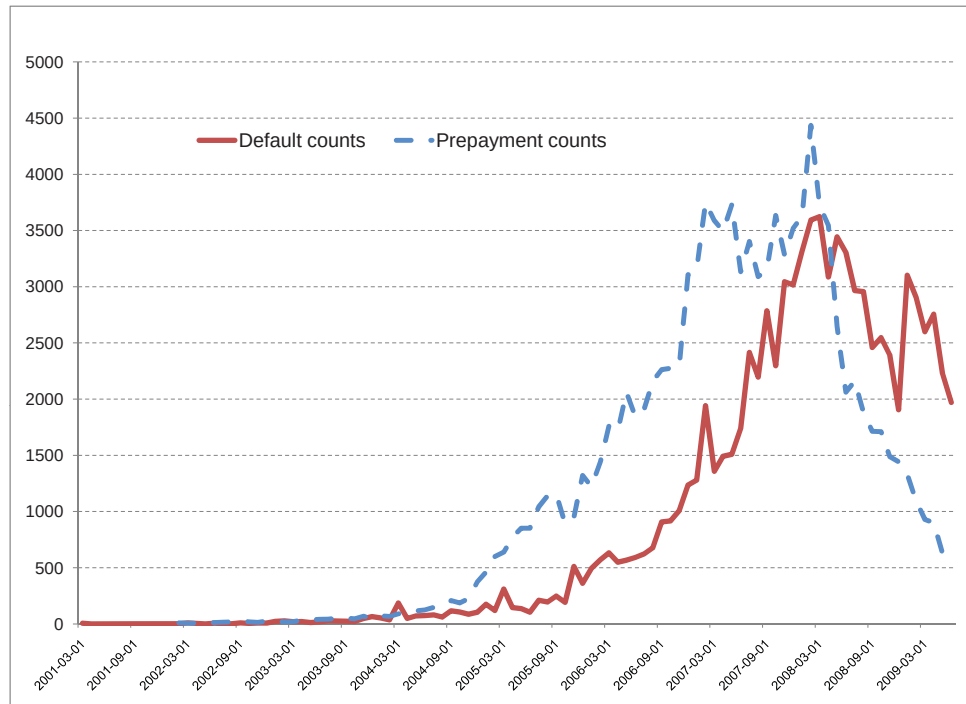


Figure 3.2
Default and prepayment incidence during the observation window

Table 3.7
Summary statistics for the fixed-rate model variables

Variable	N	Mean	St. dev.	Min	Max
<i>OrigLTVRatio</i>	82582	0.87	0.07	0.8	1.09
<i>FicoScoreOrig</i>	82582	6.33	0.56	3.45	7.20
<i>LoanSizeOrig</i>	82582	0.20	0.13	0.02	1.73
<i>OrigContRate</i>	82582	7.77	1.54	2.00	18.38
<i>Yield10Y^a</i>	82582	4.46	0.41	3.33	6.89
<i>SlopeTS^b</i>	82582	1.17	1.22	-0.41	3.14
<i>LoDocLoan</i>	82582	0.70	0.46	0	1
<i>InvestLoan</i>	82582	0.12	0.32	0	1
<i>LTV80Orig</i>	82582	0.37	0.48	0	1
<i>SizeOrigLoDoc^c</i>	82582	0.14	0.14	0	1.73

^a - yield on the 10-year Treasury note, ^b - difference between the 10-year and 1-year Treasury yields, ^c - interaction between *LoanSizeOrig* and *LoDocLoan*.

Table 3.8
OLS estimates of the model for the FRM rate

Variable	Estimate	Std. error	t-stat.	p-value
<i>Intercept</i>	0.6031	0.3133	1.93	0.0542
<i>OrigLTVRatio</i>	11.642	0.1221	95.39	<.0001
<i>LTV80Orig</i>	0.5954	0.0146	40.89	<.0001
<i>FicoScoreOrig</i>	-0.7293	0.0088	-83.05	<.0001
<i>LoanSizeOrig</i>	-4.3616	0.0720	-60.62	<.0001
<i>SizeOrigLoDoc</i>	1.2135	0.0766	15.85	<.0001
<i>Yield10Y</i>	0.1085	0.1260	2.84	0.0045
<i>Yield10YSq^a</i>	0.0368	0.0137	2.68	0.0074
<i>SlopeTS</i>	-0.3024	0.0038	-79.43	<.0001
<i>LoDocLoan</i>	-0.0080	0.0186	-0.43	0.6681
<i>InvestLoan</i>	0.2904	0.0126	23.08	<.0001
Number of obs.	82,582			
Adjusted R ²	0.5008			

Dependent variable – *OrigContRate*. Heteroscedasticity consistent standard errors and *t*-statistics are reported. Estimates of MSA fixed effects are not reported.

^a– square of *Yield10Y*.

Table 3.9
Test statistics for the presence of the calendar-time stratum effects

Statistic	Prep-t ('old')	Prep-t ('new')	Default ('old')	Default ('new')
<i>C</i> -stat.	3,805.8 (<0.01)	1,787.5 (<0.01)	314.8 (<0.01)	832.9 (<0.01)

Test statistics $C = (\hat{\beta}_s - \hat{\beta}_u)'[V(\hat{\beta}_s - \hat{\beta}_u)]^{-1}(\hat{\beta}_s - \hat{\beta}_u)$ is asymptotically distributed as χ^2 (16) (see Ridder and Tunali (1999)). P-values are reported in parenthesis below the *C*-stat. value.

Table 3.10

Comparison of out-of-sample fit for the 'old' and the 'new' models of default

Calendar month	'Old' model fit ($-2 \log L$)	'New' model fit ($-2 \log L$)	Difference	No. obs.
Jul 2005	27,968	28,009	-41	244,169
Aug 2005	21,745	21,771	-26	188,589
Sep 2005	25,015	25,051	36 ^a	195,398
Oct 2005	38,786	38,612	174 ^a	213,463
Nov 2005	28,201	28,127	74 ^a	216,560
Dec 2005	41,264	41,188	76 ^a	267,016

^a likelihood ratio test in favor of superior fit of the estimated 'new' model is significant at 5% level. Estimated parameters $\tilde{\beta}_{old}^d$ of the 'old' model are obtained by fitting unstratified proportional hazard model to the data on mortgages originated in Q1 1997 - Q2 2004, estimated parameters $\tilde{\beta}_{new}^d$ of the 'new' model – by fitting model to the data originated in Q1 2005 - Q1 2008.

Table 3.11

Estimates of the multiplicative intensity default model for the 'old' regime

Variable	Estimate	Std. error	p-value	Hazard ratio
<i>Ltv80Orig</i>	0.084	0.029	0.003	1.087
<i>LoanSizeOrig</i>	-0.701	0.140	<0.001	0.496
<i>TeaserM</i>	0.085	0.009	<0.001	1.088
<i>FicoScoreOrig</i>	-0.690	0.025	<0.001	0.501
<i>LoDocLoan</i>	0.104	0.023	<0.001	1.110
<i>PrepPenaltyLoan</i>	0.030	0.022	0.169	1.030
<i>InvestLoan</i>	0.196	0.040	<0.001	1.217
<i>RefiLoan</i>	-0.083	0.021	<0.001	0.920
<i>D24Mon</i>	0.167	0.097	0.087	1.181
<i>CurrentLtv</i>	3.375	0.137	<0.001	29.23
<i>HPriceVol</i>	0.485	0.440	0.271	1.623
<i>SprToFRM</i>	2.183	0.141	<0.001	8.873
<i>CurrContRt</i>	0.468	0.043	<0.001	1.597
<i>CurrContRtSq</i>	-0.030	0.002	<0.001	0.971
<i>TeaserI</i>	-0.008	0.002	<0.001	0.992
<i>SineSeason</i>	-0.087	0.034	0.011	0.917

Log likelihood -72,706. Number of observations 1,413,052.

Table 3.12

Estimates of the multiplicative intensity prepayment model for the 'old' regime

Variable	Estimate	Std. error	<i>p</i> -value	Hazard ratio
<i>Ltv80Orig</i>	-0.299	0.014	<.0001	0.741
<i>LoanSizeOrig</i>	0.145	0.056	0.010	1.156
<i>TeaserM</i>	0.040	0.005	<.0001	1.041
<i>FicoScoreOrig</i>	0.067	0.015	<.0001	1.070
<i>LoDocLoan</i>	1.350	0.017	<.0001	3.858
<i>PrepPenaltyLoan</i>	-0.200	0.011	<.0001	0.819
<i>InvestLoan</i>	-0.234	0.022	<.0001	0.792
<i>RefiLoan</i>	0.104	0.010	<.0001	1.110
<i>D24Mon</i>	0.663	0.032	<.0001	1.941
<i>CurrentLtv</i>	-1.754	0.053	<.0001	0.173
<i>HPriceVol</i>	4.339	0.224	<.0001	76.67
<i>SprToFRM</i>	1.583	0.081	<.0001	4.870
<i>CurrContRt</i>	0.555	0.024	<.0001	1.741
<i>CurrContRtSq</i>	-0.035	0.001	<.0001	0.966
<i>TeaserI</i>	-0.026	0.001	<.0001	0.974
<i>SineSeason</i>	0.043	0.017	0.012	1.044

Log likelihood -327,166. Number of observations 1,413,052.

Table 3.13

Estimates of the multiplicative intensity default model for the 'new' regime

Variable	Estimate	Std. error	<i>p</i> -value	Hazard ratio
<i>Ltv80Orig</i>	0.413	0.010	<0.001	1.511
<i>LoanSizeOrig</i>	0.621	0.034	<0.001	1.860
<i>TeaserM</i>	0.023	0.002	<0.001	1.023
<i>FicoScoreOrig</i>	-0.575	0.008	<0.001	0.563
<i>LoDocLoan</i>	0.286	0.008	<0.001	1.331
<i>PrepPenaltyLoan</i>	0.038	0.008	<0.001	1.039
<i>InvestLoan</i>	0.294	0.013	<0.001	1.342
<i>RefiLoan</i>	-0.211	0.008	<0.001	0.810
<i>D24Mon</i>	-0.089	0.029	0.002	0.915
<i>CurrentLtv</i>	4.247	0.088	<0.001	69.89
<i>HPriceVol</i>	0.852	0.108	<0.001	2.344
<i>SprToFRM</i>	2.071	0.075	<0.001	7.930
<i>CurrContRt</i>	0.775	0.023	<0.001	2.171
<i>CurrContRtSq</i>	-0.043	0.001	<0.001	0.957
<i>TeaserI</i>	-0.001	0.000	<0.001	0.999
<i>SineSeason</i>	-0.011	0.012	0.354	0.989

Log likelihood -724,575. Number of observations 4,652,552.

Table 3.14

Estimates of the multiplicative intensity prepayment model for the 'new' regime

Variable	Estimate	Std. error	<i>p</i> -value	Hazard ratio
<i>Ltv80Orig</i>	-0.483	0.011	<.0001	0.617
<i>LoanSizeOrig</i>	0.338	0.039	<.0001	1.402
<i>TeaserM</i>	0.089	0.002	<.0001	1.093
<i>FicoScoreOrig</i>	0.213	0.010	<.0001	1.237
<i>LoDocLoan</i>	0.371	0.009	<.0001	1.450
<i>PrepPenaltyLoan</i>	-0.133	0.008	<.0001	0.875
<i>InvestLoan</i>	-0.252	0.016	<.0001	0.777
<i>RefiLoan</i>	0.246	0.008	<.0001	1.279
<i>D24Mon</i>	0.370	0.028	<.0001	1.448
<i>CurrentLtv</i>	-2.700	0.060	<.0001	0.067
<i>HPriceVol</i>	-1.465	0.187	<.0001	0.231
<i>SprToFRM</i>	1.463	0.062	<.0001	4.321
<i>CurrContRt</i>	0.788	0.026	<.0001	2.200
<i>CurrContRtSq</i>	-0.042	0.001	<.0001	0.959
<i>TeaserI</i>	-0.005	0.000	<.0001	0.995
<i>SineSeason</i>	0.085	0.013	<.0001	1.089

Log likelihood -705,635. Number of observations 4,652,552.

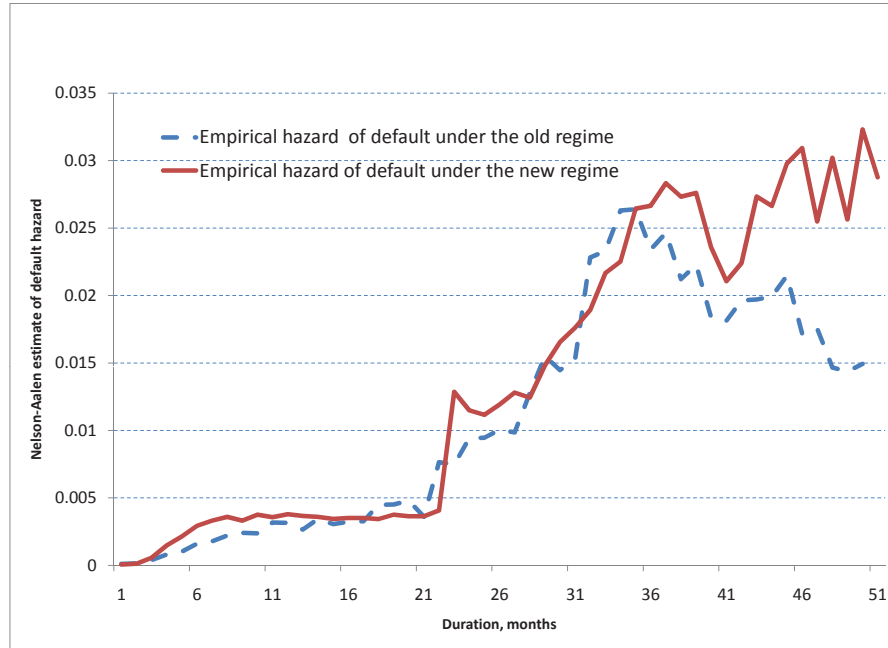


Figure 3.3

Estimates of the hazards of default for a representative contract under the 'old' regime and under the 'new' regime

Empirical estimates of the monthly hazard under the two regimes. The Nelson-Aalen of baseline default hazard for the respective unstratified model were used in the calculations. Contract was originated in Atlanta in January 2004: loan size \$181,155, LTV 80%, FICO 673, margin 5.7 p.p., lifetime floor 6.25%, lifetime cap 12.25%, teaser 2.375 p.p. Contract was originated for the purpose of refinancing, full documentation was provided, prepayment penalty existed.

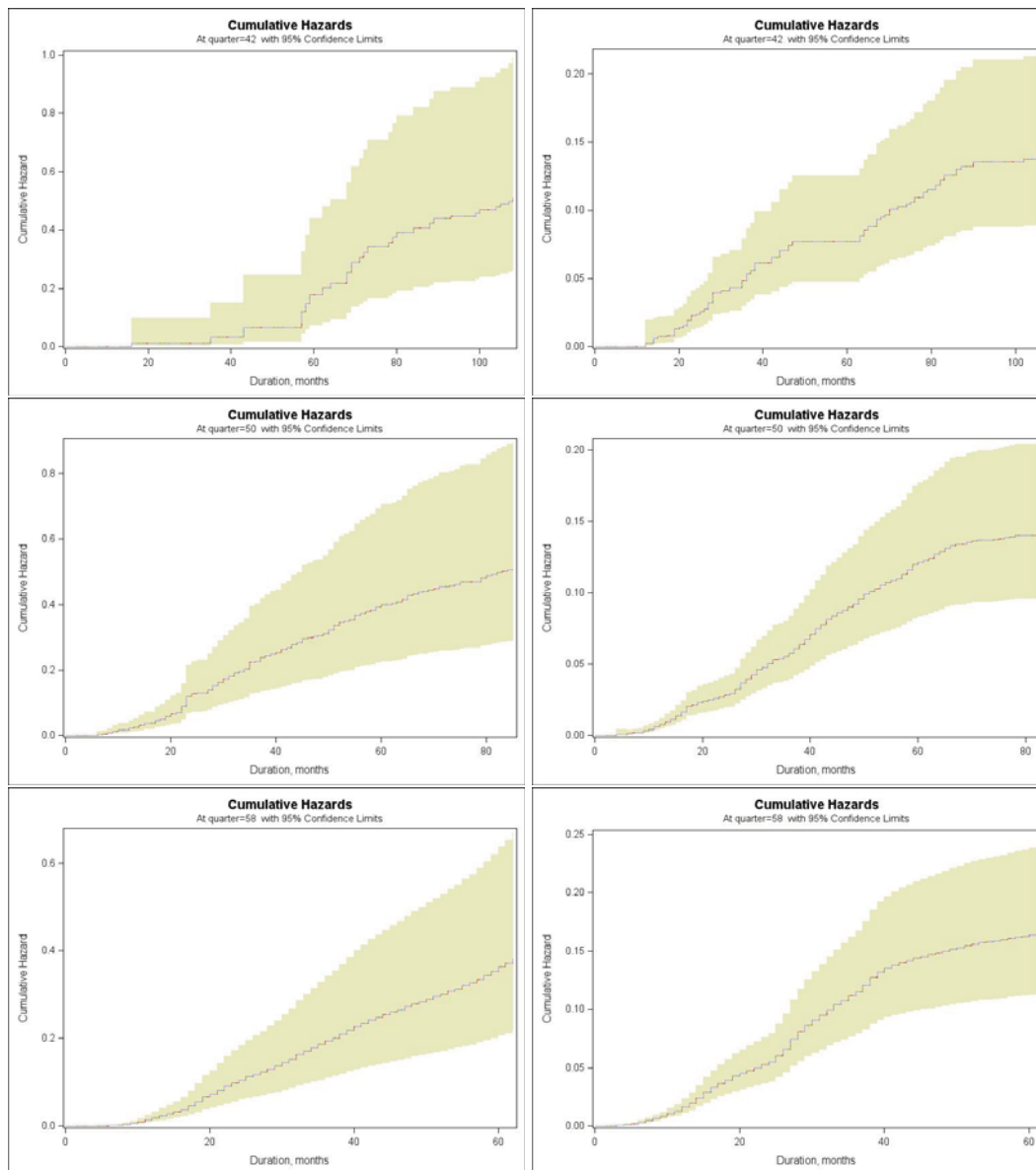


Figure 3.4

Cumulative baseline hazards of default (left) and prepayment (right) estimated under the 'old' regime

Nelson-Aalen estimates of cumulative baseline hazard (covariate values are set to 0) are depicted for the following strata: first row – Q2 2000, second row – Q2 2002, third row – Q2 2004.

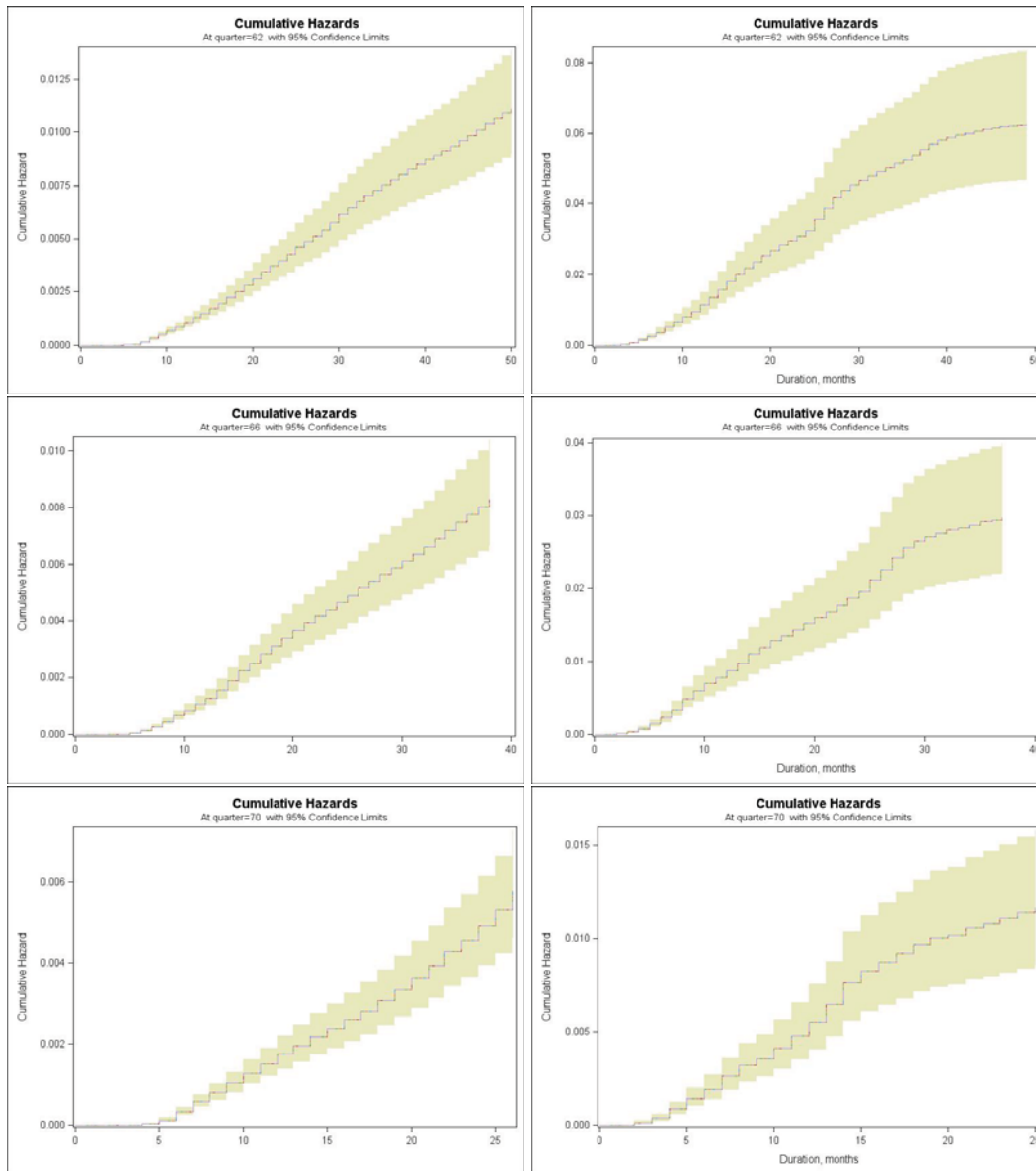


Figure 3.5

Cumulative baseline hazards of default (left) and prepayment (right) estimated under the 'new' regime

Nelson-Aalen estimates of cumulative baseline hazard (covariate values are set to 0) are depicted for the following strata: first row – Q2 2005, second row – Q2 2006, third row – Q2 2007.



Figure 3.6

Ratio of the number of foreclosed mortgages to the total number of observed mortgages for an 'old' regime cohort versus a 'new' regime cohort
 Average monthly number of observed mortgages for Q2 2002 cohort in the analyzed period - 581,
 for Q2 2006 cohort - 14,381.

Table 3.15
Estimates of the (empirical) cumulative hazard of default

	'New' regime	'Old' regime
Panel A. Trended values of dynamic LTV during the period of the house price drop.		
Contract A 2-year ahead	9.38%	5.88%
Contract A 4.7-year ahead	15.47%	13.30%
Panel B. Actual covariate values.		
Contract A 2-year ahead	10.32%	6.30%
Contract A 4.7-year ahead	23.50%	17.50%

Values of the cumulative hazard of default are obtained using Nelson-Aalen estimates of the baseline hazard for non-stratified models of default and prepayment for the respective regime. Contract characteristics for contract 'A': origination – San Diego, October 2005, loan amount – \$ 543,268 (in constant year 2000 dollars), LTV at origination – 85%, contract rate at origination – 6.75 %, margin – 3.03%, lifetime cap – 13.75%, periodic cap – 1 perc. pt, lifetime floor - 6.75%; purchase loan, FICO at origination – 657. After 9 months the actual values of house prices have been replaced by the 12-month moving average.

Table 3.16
Summary statistics for the yields

Maturity	Minimum	Median	Maximum	Mean	St. dev.	Kurtosis
3 M	0.0027	0.0595	0.1603	0.0294	0.8641	4.240
6 M	0.0031	0.0613	0.1652	0.0297	0.8092	4.035
12 M	0.0039	0.0633	0.1581	0.0292	0.6816	3.654
5 Y	0.0184	0.0702	0.1500	0.0260	0.6987	3.308
10 Y	0.0271	0.0742	0.1532	0.0254	0.8892	3.461

Number of observations - 468.

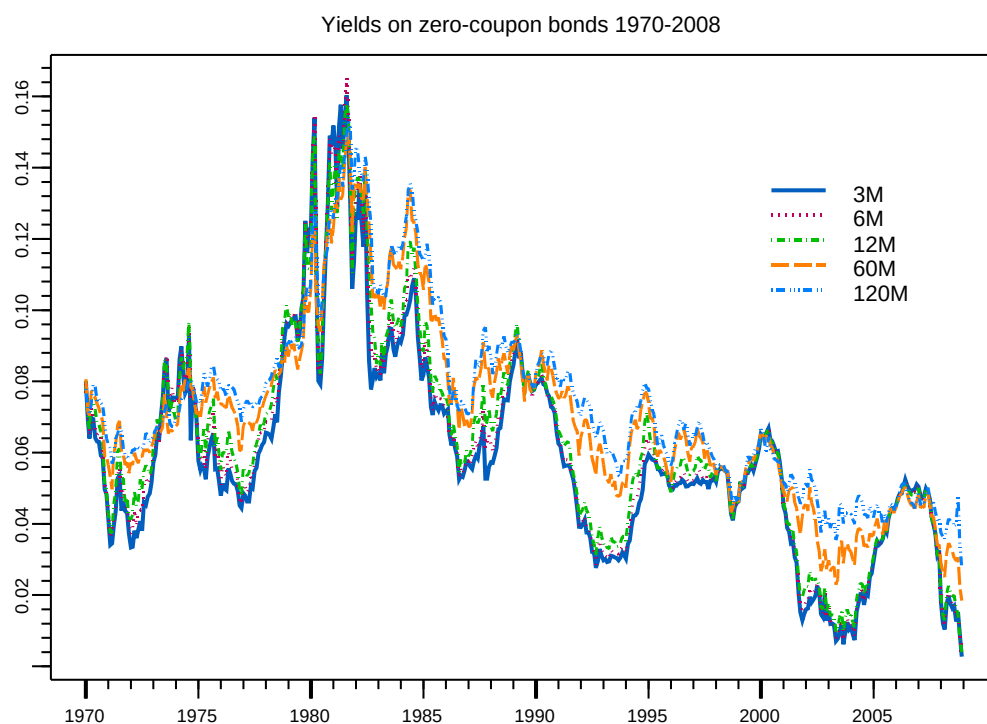


Figure 3.7

Monthly series of yields for 5 maturities: January 1970 - December 2008
 Sources of data : CRSP, University of Chicago and J. Huston McCulloch's website
<http://www.econ.ohio-state.edu/jhm/ts/>

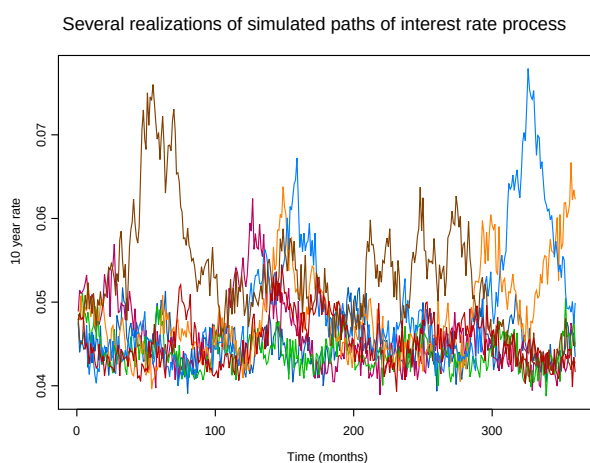


Figure 3.8

Simulated paths of yield on 10-year bond under estimated parameters of CIR interest rate process

Table 3.17

Estimates of the parameters of the affine term structure model

Parameter	Estimate	Std error	<i>t</i> -statistic
κ_1	0.8514	0.0234	6.0114
θ_1	0.0275	5.27e-4	6.0733
σ_1	0.1295	0.0001	28.1143
α_1	-0.3004	0.0136	-2.1204
κ_2	0.0144	0.0018	0.5610
θ_2	0.0006	0.0013	0.0715
σ_2	0.0539	7.40e-6	22.0558
α_2	-0.0427	0.0003	-1.6573
$\sigma_{\epsilon_{3m}}$	2.5e-3	6.0e-9	—
$\sigma_{\epsilon_{6m}}$	1.0e-8	2.7e-8	—
$\sigma_{\epsilon_{12m}}$	2.2e-3	5.3e-9	—
$\sigma_{\epsilon_{5y}}$	2.1e-3	1.0e-8	—
$\sigma_{\epsilon_{10y}}$	1.1e-3	1.2e-8	—
Log-likelihood	-10,355		

State-space routines from *SsfPack 3.0* library by S.J. Koopman, N. Shephard and J. Doornik were used in the estimation. *t*-stats are based on the OPG standard errors. N obs. = 468.

Table 3.18

Pearson correlation coefficients between HPA indices

Region	LA	SDiego	Tampa	Detroit	NYC	Comp. 10
<i>LA</i>	1	0.9839	0.9892	0.7327	0.9729	0.9909
<i>SDiego</i>		1	0.9669	0.8156	0.9671	0.9884
<i>Tampa</i>			1	0.7442	0.9734	0.9874
<i>Detroit</i>				1	0.7485	0.7836
<i>NYC</i>					1	0.9917
<i>Comp. 10</i>						1

Pearson correlation coefficients between selected individual Case-Shiller HPA indices the composite index for 10 largest MSAs. Sample size: 249 obs.

Yields on 3M and 10Y zero-coupon bonds and Kalman filter estimate of the smoothed values of the state

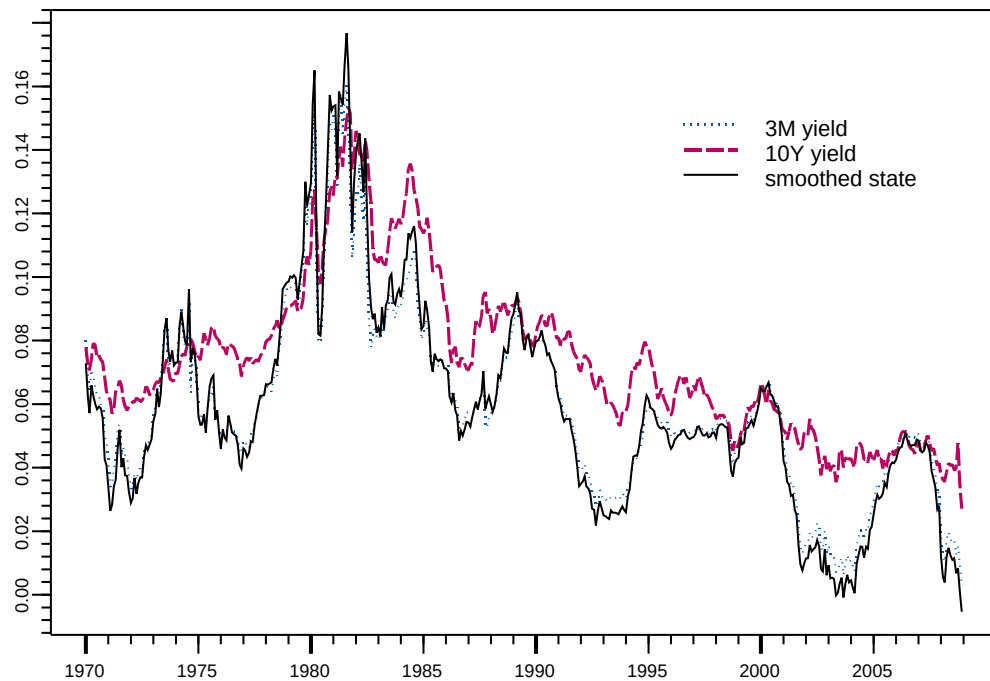


Figure 3.9

Smoothed state estimates from Kalman filter, yield on 3-month zero-coupon bond and yield on 10-year zero-coupon bond

Sum of the smoothed state estimates($\hat{y}_{j,t}$, $j = 1, 2$) for the two factors of the term structure is depicted by the solid black line.

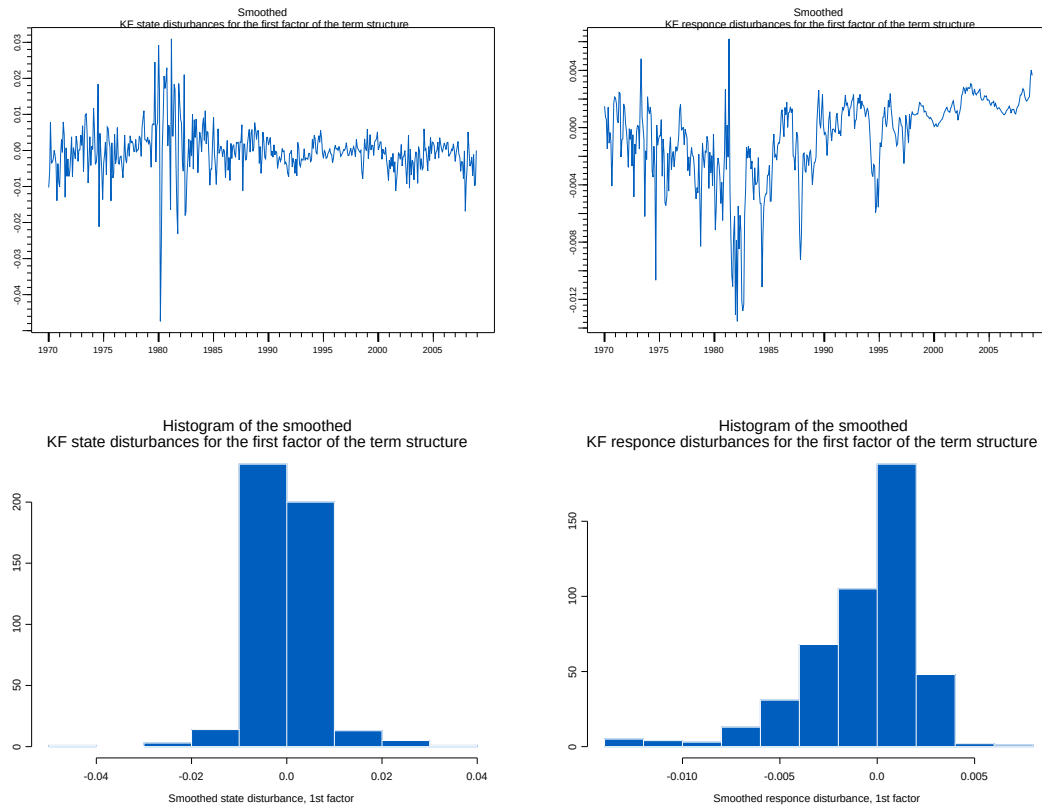


Figure 3.10

Series of smoothed state and response disturbances for the first factor of the CIR term structure

Smoothed state disturbance estimates $\hat{\eta}_t$ are in the left column; smoothed response disturbance estimates $\hat{\epsilon}_t$ are in the right column.

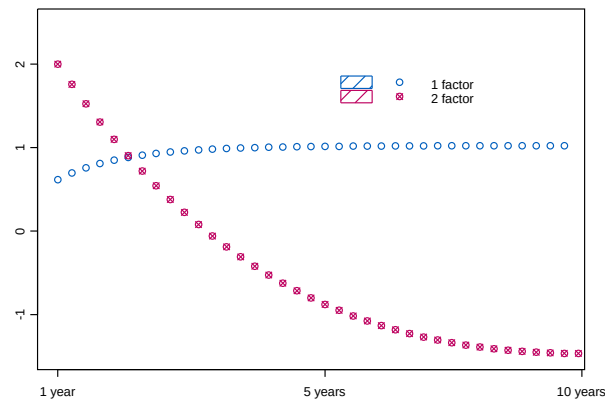


Figure 3.11

Factor loadings for the estimated parameters of two latent factors
Loadings on the first and the second factor are drawn as the functions of the bond's maturity

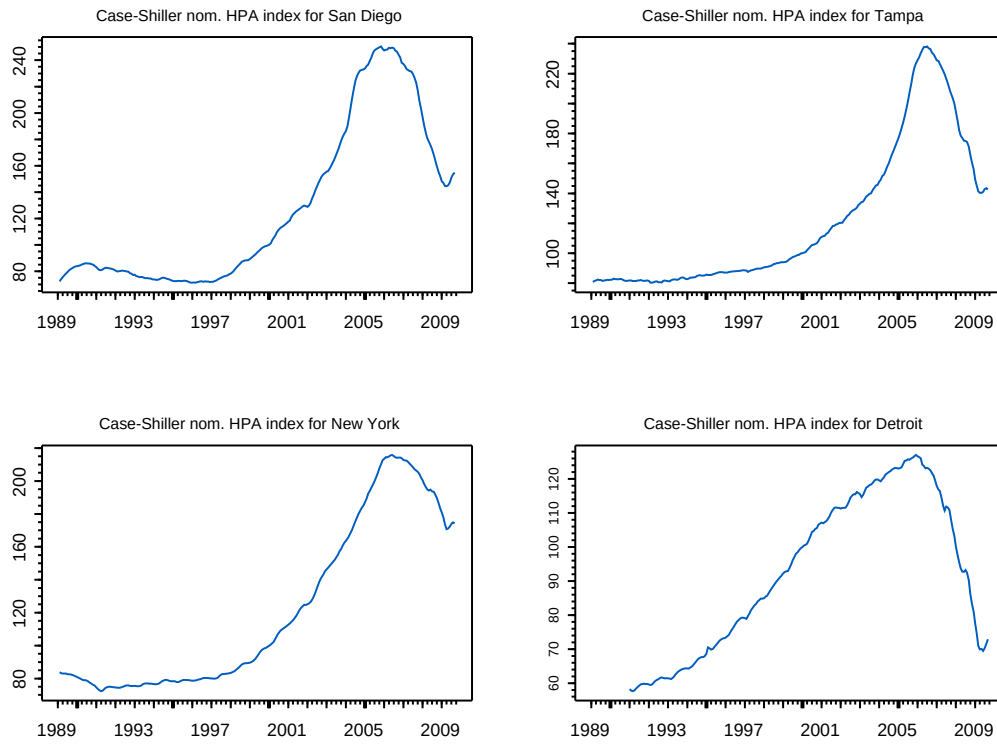


Figure 3.12
Case-Shiller HPA indices for selected MSAs
Nominal indices for 1989-2009, January 2000 value equal to 100

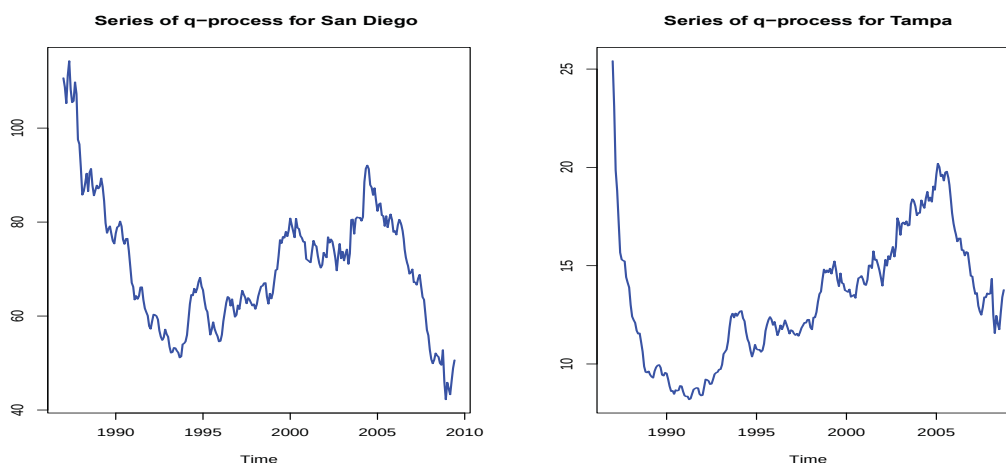


Figure 3.13
Series of relative returns (q) for selected MSAs

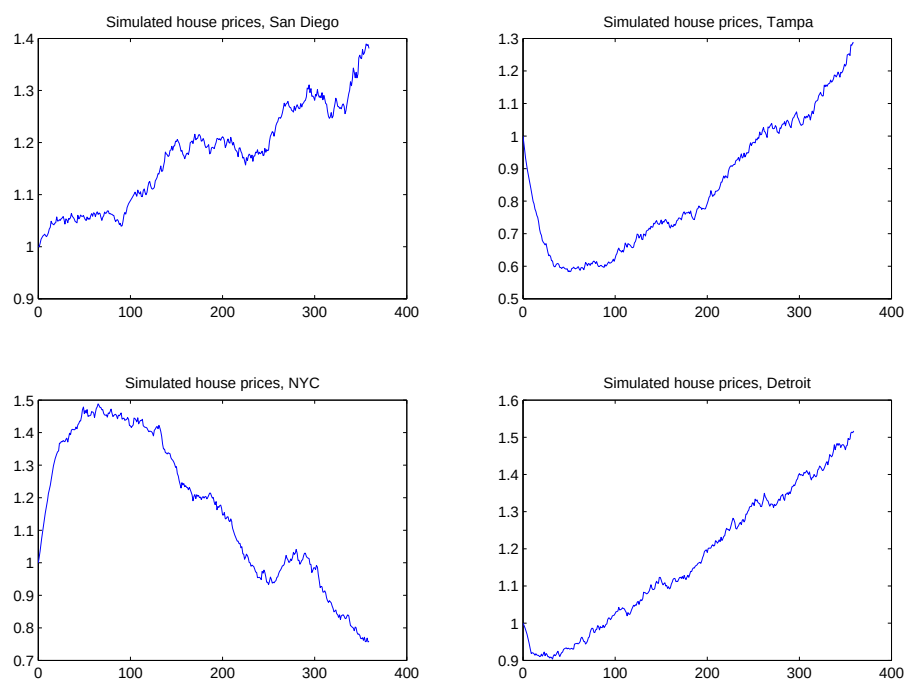


Figure 3.14
Simulated house price appreciation indices for selected MSAs
Means of 100 simulated house price paths are depicted. Level of house prices in January 2006 is taken to be 1. Simulation is carried 360 months forward.

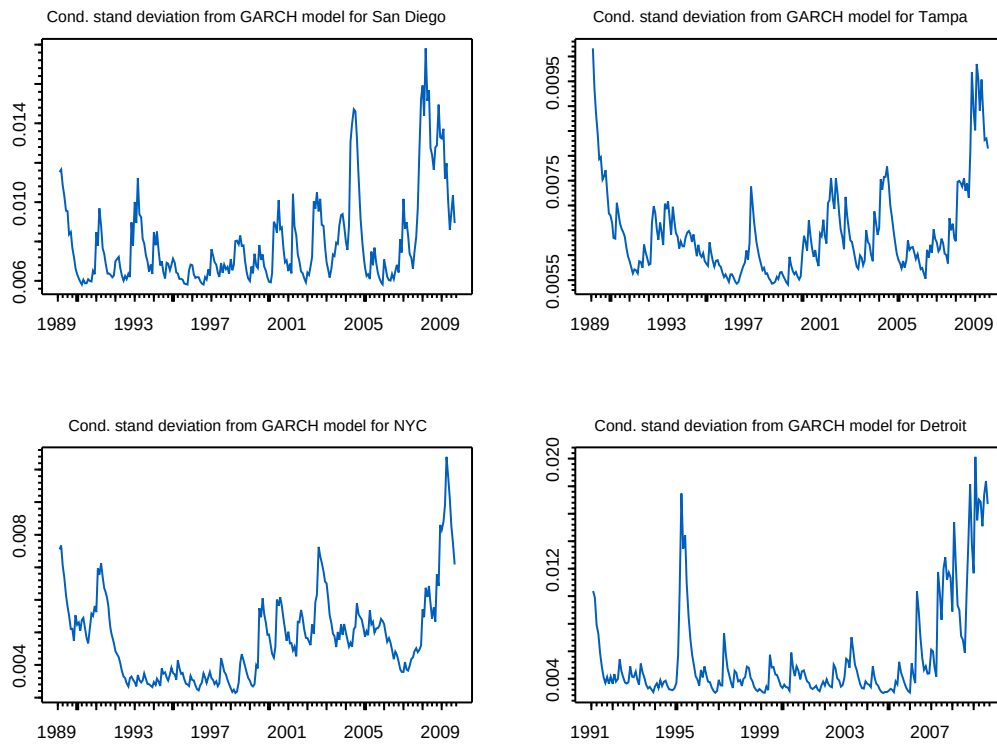


Figure 3.15
Estimates of conditional standard deviation from the GARCH models for selected MSAs

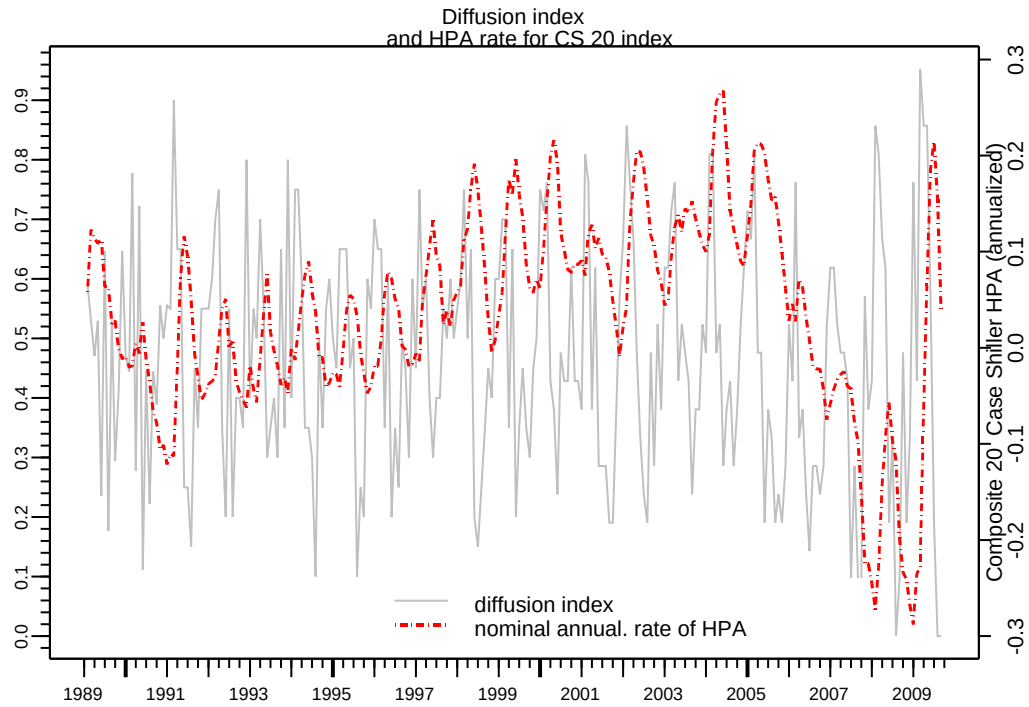


Figure 3.16

The diffusion index for annualized nominal returns across 20 MSAs and the nationwide rate of nominal house price appreciation

Diffusion index (solid gray) is the average of indicators which take the value of 1 if prices in a given region in a given period increased compared to their level in the previous month and 0 if prices decreased. The nationwide nominal rate of house price appreciation for 1989-2009 is depicted by the dotted red line.

Table 3.19
Estimates of the relative housing returns (q) process

MSA	κ_q estimate	$\kappa_q \theta_q$ estimate	σ_q estimate	φ_{01}	φ_{11}
Phoenix	0.0787 (0.005)	0.6080 (0.049)	0.1381 (0.38)	-4.35E-06	2.07E-03
Los Angeles	0.0402 (0.008)	0.6477 (2.29)	0.1359 (0.080)	-0.0894	5.31E-03
San Diego	0.0303 (0.004)	2.0325 (0.771)	0.2387 (0.007)	-0.0541	8.11E-04
San Francisco	0.0453 (0.004)	0.4296 (0.004)	0.1559 (0.007)	-0.0089	1.27E-03
Denver	0.1229 (0.004)	0.5615 (0.019)	0.1560 (0.005)	-0.0073	2.42E-03
Washington	0.0312 (0.004)	0.3948 (0.067)	0.1804 (0.008)	-0.0062	7.93E-04
Miami	0.0559 (0.004)	0.4996 (0.052)	0.1711 (0.007)	-0.0029	8.64E-04
Tampa	0.0370 (0.008)	0.4216 (0.034)	0.1212 (0.008)	-0.0632	5.42E-03
Atlanta	0.0552 (0.009)	0.2513 (0.054)	0.1294 (0.003)	-0.0083	1.10E-03
Chicago	0.0545 (0.004)	0.4193 (0.089)	0.1141 (0.008)	-0.0013	1.21E-03
Boston	0.0880 (0.005)	0.1041 (0.087)	0.0778 (0.003)	-0.0009	6.37E-03
Detroit	0.0902 (0.009)	1.1777 (0.145)	0.1318 (0.006)	-0.0218	1.60E-03
Minneapolis	0.0703 (0.005)	0.1291 (0.010)	0.0921 (0.002)	-0.0177	4.89E-03
Charlotte	0.0667 (0.005)	0.1119 (0.011)	0.0654 (0.003)	-0.0085	7.92E-03
Las Vegas	0.0600 (0.006)	0.2151 (0.018)	0.0986 (0.004)	-0.0098	4.74E-03

Ratio of returns from holding housing to returns from investing in discount bond (continuously 'rolling over' the latter) is assumed to obey the SDE:

$dq(t) = q(t) \left[\kappa_q(\theta_q - q(t))dt - \sigma_q \sqrt{q(t)} dW_q(t) \right]$. Parameters φ_{01} and φ_{11} are calibrated so that empirical series of q appear relatively trendless. Standard errors of diffusion parameters or combination thereof are reported in parenthesis below the estimates.

Table 3.20
Estimates of the relative housing returns (q) process

MSA	κ_q estimate	$\kappa_q \theta_q$ estimate	σ_q estimate	φ_{01}	φ_{11}
New York	0.0681 (0.005)	0.1137 (0.011)	0.0809 (0.004)	-0.0054	6.2E-03
Cleveland	0.0338 (0.008)	0.0710 (0.004)	0.0562 (0.006)	-0.0010	6.95E-03
Portland	0.1056 (0.008)	0.0949 (0.005)	0.0614 (0.005)	-0.0052	1.28E-02
Dallas	0.1487 (0.042)	1.1422 (0.147)	0.0944 (0.003)	-0.0195	3.24E-03
Seattle	0.1302 (0.007)	0.2789 (0.064)	0.0778 (0.004)	-0.0119	7.31E-03

Ratio of returns from holding housing to returns from investing in discount bond (continuously 'rolling over' the latter) is assumed to obey the SDE:

$dq(t) = q(t) \left[\kappa_q(\theta_q - q(t))dt - \sigma_q \sqrt{q(t)} dW_q(t) \right]$. Parameters φ_{01} and φ_{11} are calibrated so that empirical series of q appear relatively trendless. Standard errors of diffusion parameters or combination thereof are reported in parenthesis below the estimates.

Table 3.21
Estimated parameter values of the GARCH models of housing returns

Parameter	Estimate	Std err.	Estimate	Std err.	Estimate	Std err.
Panel A.	Phoenix		San Francisco		Denver	
Const. (cond. mean)	3.02E-04	1.85E-04	6.39E-05	3.36E-04	1.07E-03	2.87E-04
AR(1) (cond. mean)	9.19E-01	2.27E-02	8.89E-01	2.82E-02	7.65E-01	4.56E-02
Const. (cond. var.)	7.71E-07	5.42E-07	3.21E-06	2.34E-06	4.62E-07	5.86E-07
ARCH(1)	3.09E-01	1.36E-01	1.49E-01	7.8E-03	1.06E-01	5.33E-02
GARCH(1)	6.69E-01	1.21E-01	7.77E-01	1.1E-01	8.77E-01	7.35E-02
AIC	-2126		-1858		-2040	
N. obs.	248		248		248	
Panel B.	Washington		Miami		Tampa	
Const. (cond. mean)	-2.31E-06	2.38E-04	1.48E-04	2.51E-04	4.13E-04	4.58E-04
AR(1) (cond. mean)	8.82E-01	2.52E-02	9.47E-01	2.32E-02	8.45E-01	4.55E-02
Const. (cond. var.)	7.28E-06	2.99E-06	4.25E-06	2.22E-06	5.84E-06	7.58E-06
ARCH(1)	3.57E-01	1.50E-01	2.86E-01	1.34E-01	9.71E-02	9.03E-02
GARCH(1)	2.97E-01	1.77E-01	5.85E-01	1.32E-01	7.82E-01	2.33E-01
AIC	-2018		-1965		-1850	
N. obs.	248		248		248	
Panel C.	Atlanta		Chicago		Minneapolis	
Const. (cond. mean)	6.90E-04	2.02E-04	1.16E-03	2.87E-04	1.09E-03	3.11E-04
AR(1) (cond. mean)	7.73E-01	3.91E-02	7.07E-01	4.25E-02	7.19E-01	4.03E-02
Const. (cond. var.)	5.02E-07	3.47E-07	1.26E-06	8.36E-07	3.23E-06	2.10E-06
ARCH(1)	1.56E-01	5.32E-02	1.51E-01	5.56E-02	2.27E-01	1.05E-01
GARCH(1)	7.78E-01	7.66E-02	7.96E-01	7.31E-02	6.80E-01	1.43E-01
AIC	-1923		-1970		-1898	
N. obs.	232		248		248	
Panel D.	Charlotte		Las Vegas		Cleveland	
Const. (cond. mean)	1.22E-03	2.93E-04	4.01E-04	3.0E-04	1.27E-03	3.02E-04
AR(1) (cond. mean)	5.32E-01	6.44E-02	8.47E-01	3.6E-02	5.68E-01	6.29E-02
Const. (cond. var.)	2.17E-06	1.71E-06	2.34E-06	1.3E-06	8.51E-07	9.19E-07
ARCH(1)	1.90E-01	9.89E-02	2.02E-01	8.1E-02	1.46E-01	5.36E-02
GARCH(1)	7.06E-01	1.60E-01	7.47E-01	8.4E-02	8.37E-01	7.22E-02
AIC	-2014		-1855		-1914	
N. obs.	248		248		248	

Table 3.22

Estimated parameter values of the GARCH models of housing returns (continued)

Parameter	Estimate	Std err.	Estimate	Std err.	Estimate	Std err.
Panel A.	Los Angeles		San Diego		Boston	
Const. (cond. mean)	2.7E-03	7.97E-04	3.1E-03	8.50E-04	3.88E-03	7.09E-04
MA(1) (cond. mean)	0.5.9E-01	7.48E-02	5.4E-01	7.45E-02	5.66E-01	6.15E-02
Const. (cond. var.)	1.2E-05	7.27E-06	1.3E-05	7.90E-06	2.24E-06	2.56E-06
ARCH(1)	1.8E-01	7.71E-02	1.9E-01	8.64E-02	9.30E-02	5.53E-02
GARCH(1)	5.5E-01	2.09E-01	5.9E-01	1.98E-01	8.62E-01	9.27E-02
AIC	-1719		-1692		-1765	
N. obs.	248		248		248	
Panel B.	New York		Portland		Detroit	
Const. (cond. mean)	2.92E-03	5.3E-04	5.06E-03	5.06E-03	9.11E-04	2.78E-04
MA(1) (cond. mean)	6.54E-01	6.3E-02	7.23E-01	7.23E-01	7.91E-01 [†]	7.26E-02
MA(2) (cond. mean)			6.91E-01	6.91E-01	1.43E-01 [‡]	8.72E-02
Const. (cond. var.)	1.34E-06	1.1E-06	7.81E-07	7.81E-07	3.43E-06	1.99E-06
ARCH(1)	1.26E-01	5.9E-02	8.63E-02	8.63E-02	4.15E-01	2.00E-01
GARCH(1)	8.09E-01	9.0E-02	8.89E-01	8.89E-01	5.57E-01	1.34E-01
AR(3) (cond. mean)					-4.14E-01	7.90E-02
AR(4) (cond. mean)					2.08E-01	6.18E-02
AIC	-1931		-1948		-1781	
N. obs.	248		248		224	
Panel C.	Dallas		Seattle			
Const. (cond. mean)	5.9E-04	4.03E-04	1.12E-03	3.35E-04		
AR(1) (cond. mean)	6.9E-01	4.85E-02	7.83E-01	4.03E-02		
Const. (cond. var.)	3.3E-06	3.40E-06	2.07E-06	1.62E-06		
ARCH(1)	1.1E-01	9.40E-02	1.21E-01	6.38E-02		
GARCH(1)	6.9E-01	2.76E-01	7.91E-01	1.11E-01		
AIC	-875		-1822			
N. obs.	116		232			

[†] – AR(1), [‡] – AR(2)

Table 3.23
ML estimation of the latent termination processes.

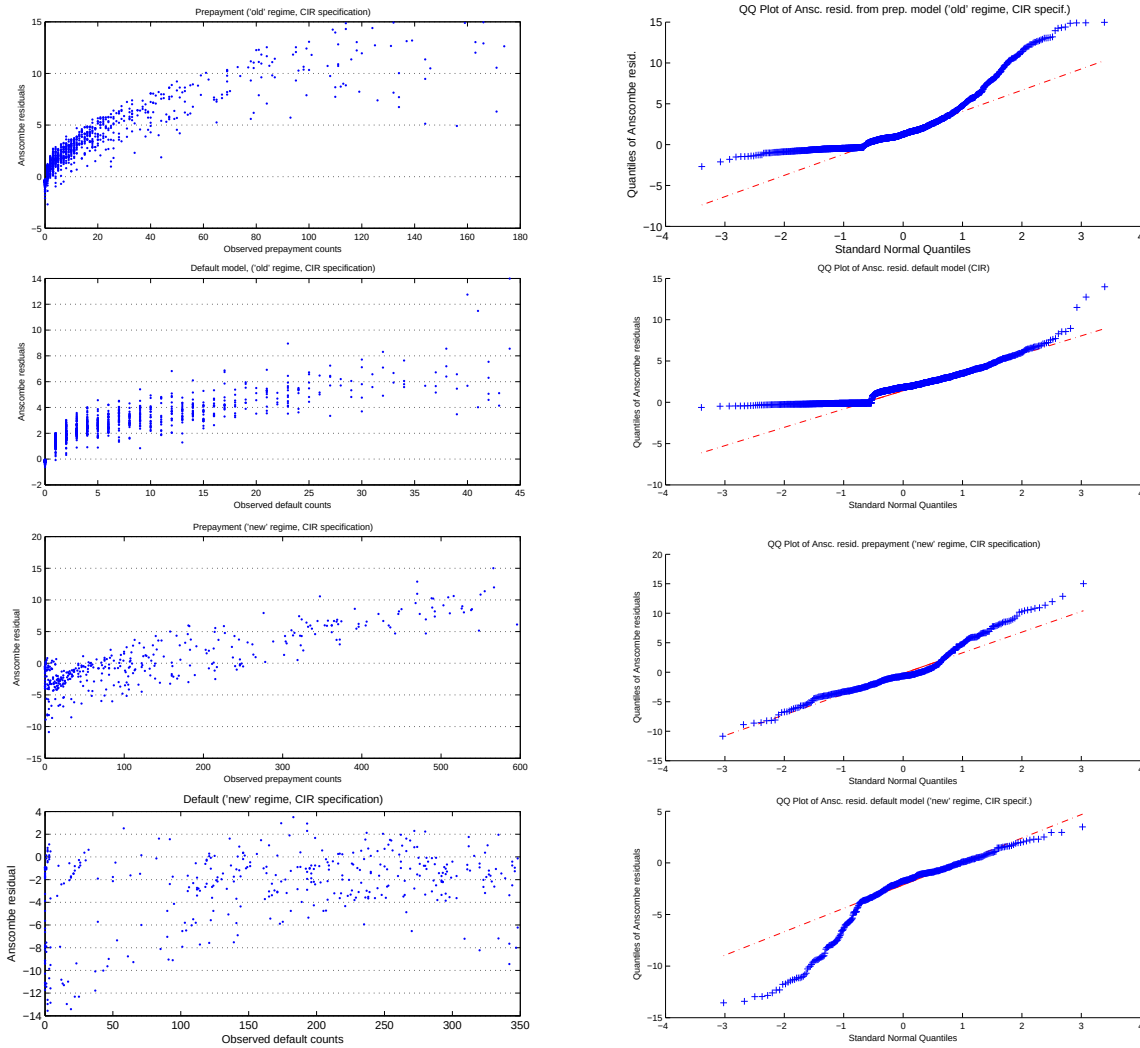
Parameter	Estim.	Std err.	Parameter	Estim.	Std err.
Panel A. Log-OU (Black-Karasinski) specification. 'Old' regime (up to 2004)					
κ^p	0.1047	0.012	κ^d	0.1746	0.008
μ^p	-6.5887	0.322	μ^d	-5.3822	0.349
σ^p	0.2620	0.010	σ^d	0.4067	0.008
Number of observations	1,429			1,429	
Log-likelihood	-15,142			-15,057	
Panel B. Log-OU (Black-Karasinski) specification. 'New' regime (2005 and later)					
κ^p	0.1017	0.018	κ^d	0.0869	0.009
μ^p	-7.0758	0.277	μ^d	-8.2614	0.439
σ^p	0.2826	0.011	σ^d	0.3376	0.009
Number of observations	439			439	
Log-likelihood	-5,548			-5,247	

Simulated ML estimation with likelihood approximated by APF with 5,000 support points (particles).

Table 3.24
ML estimation of the latent termination processes.

Parameter	Estim.	Std err.	Parameter	Estim.	Std err.
Panel A. CIR specification. 'Old' regime (up to 2004).					
κ^p	0.2248	0.030	κ^d	0.6242	0.062
μ^p	1.6381	0.121	μ^d	2.3075	0.021
σ^p	0.8778	0.008	σ^d	1.5920	0.001
Number of observations	1,429			1,429	
Log-likelihood	-20,127			-26,329	
Panel B. CIR specification. 'New' regime (2005 and later)					
κ^p	0.4933	0.010	κ^d	1.2017	0.009
μ^p	2.3975	0.045	μ^d	4.1618	0.023
σ^p	1.0430	0.002	σ^d	0.9957	0.008
Number of observations	439			439	
Log-likelihood	-5,658			-5,425	

Simulated ML estimation with likelihood approximated by APF with 5,000 support points (particles). Parameters are scaled by the factor 1E-4 in the default estimation and by the factor 1E-3 in the prepayment estimation.

**Figure 3.17**

Anscombe residuals from default and prepayment models (CIR specification)

Anscombe transformation is defined as $A(y) = \int_{-\infty}^y \text{Var}^{-\frac{1}{3}}(t) dt$. It is aimed at transforming data to approximately normally distributed with standard deviation 1.

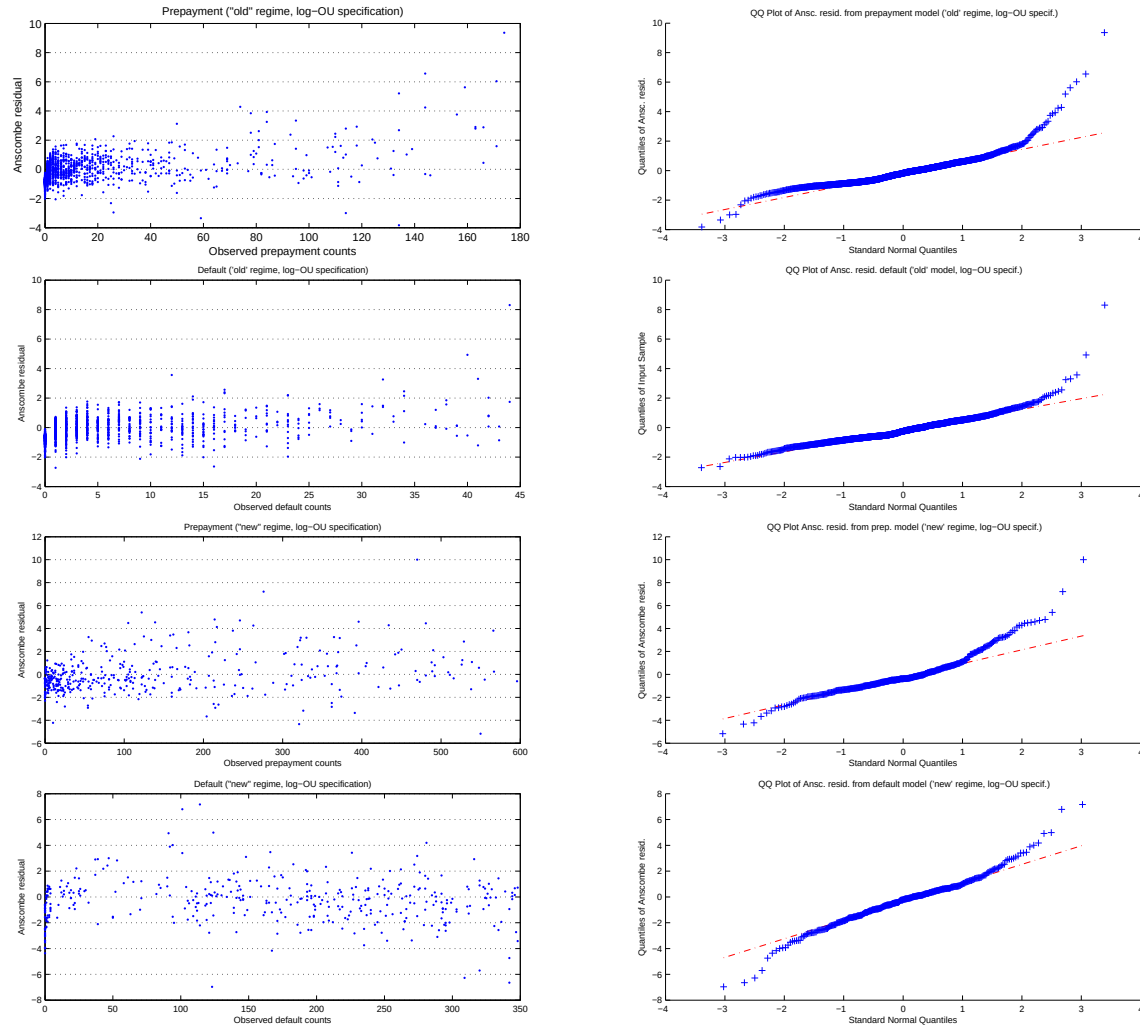


Figure 3.18

Anscombe residuals from default and prepayment models (log-OU specification)
 Anscombe transformation is defined as $A(y) = \int_{-\infty}^y \text{Var}^{-\frac{1}{3}}(t)dt$. It is aimed at transforming data to approximately normally distributed with standard deviation 1.

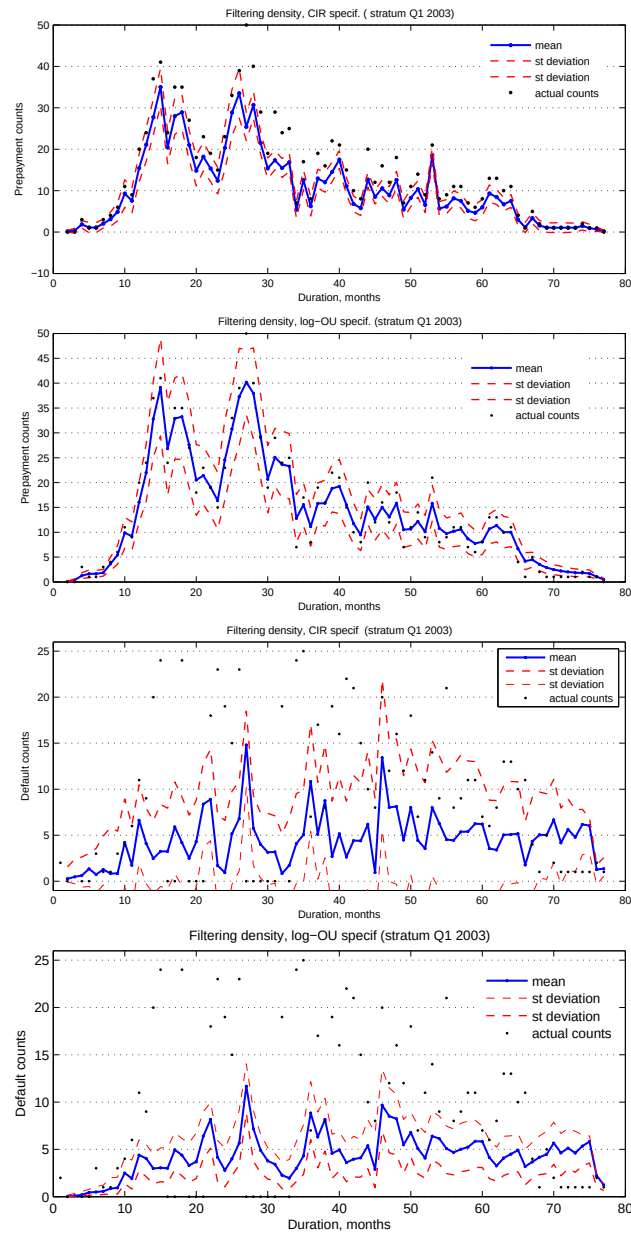


Figure 3.19

One-step-ahead predictive density and actual counts of default and prepayment. Solid blue line depicts sample mean of predictive distribution (2.59), as approximated by 5,000 discrete support points, for a contracts from a given stratum (originated in Q1 2003). Dashed red line is sample standard deviation. Actual number of counts is represented by black dots.

Table 3.25
Pearson GOF statistic for different model specifications

Pearson statistic	Default model ($\hat{y}_{jn}$)	Default model ($\hat{\hat{y}}_{jn}$)	Prepaym. model ($\hat{y}_{jn}$)	Prepaym. model ($\hat{\hat{y}}_{jn}$)
Panel A.	Log-OU (Black-Karasinski) specification. 'Old' regime (up to 2004)			
G_g	2,192 (1.00)	773.6 (<0.05)	2,741 (1.00)	746.7 (<0.05)
Panel B.	Log-OU (Black-Karasinski) specification. 'New' regime (2005 and later)			
G_g	2,656 (1.00)	156.6 (<0.05)	3,592 (1.00)	122.1 (<0.05)
Panel C.	CIR specification. 'Old' regime (up to 2004).			
G_g	5,303 (1.00)	2,091 (0.997)	5,824 (1.00)	2,747 (1.00)
Panel D.	CIR specification. 'New' regime (2005 and later).			
G_g	3,945 (1.00)	878.0 (1.00)	3,185 (1.00)	107.6 (<0.05)

P -values calculated under the assumption that test statistic is asymptotically $\chi^2(jn - p)$ distributed are reported in parenthesis. Number of degrees of freedom (with number of estimated parameters p taken to be 19) for Panels A & C is 1,410, for Panels B & D – 420. Statistic G_g in the columns 2 and 4 is calculated based on simple arithmetic mean, while the one in the columns 3 and 5 is based on the filter output.

CHAPTER 4

CALIBRATION OF THE PRICING MODEL TO THE OBSERVED MARKET PRICES OF RESIDENTIAL MORTGAGES

4.1 PRICE OF A MORTGAGE CONTRACT AS A CONDITIONAL EXPECTATION

Mortgages in our sample have been originated in different states with varying legal provisions with regard to foreclosure and sale of property. There is also certain variation with regard to mortgage insurance within our data set. Nevertheless, common feature of all the contracts is that the amount of lender's claim is known with certainty at any time during the term of a contract. However, foreclosure process and subsequent sale imply significant costs to the lender. Therefore, it is reasonable to assume that only a fraction of the amount of the legal claim is recovered, which amounts to the assumption of recovery of face value. In practice lender does not necessarily receive the payoff at the time of default as it is implied by RFV, however, this simplification should not pose a major problem. Valuation of the mortgage amounts to evaluation the conditional expectation under the risk-neutral (spot martingale) measure that makes state processes martingales. Default and prepayment in the model are driven by two processes based on independent¹ Brownian motions; in addition, there are two latent factors of the term structure of interest rates and latent factor governing house price dynamics (house prices enter the model via concurrent loan-to-value ratio parameter). Monte Carlo valuation procedure is based on the application of the basic no-arbitrage principle: the amount of funds loaned should equal

¹An assumption of independence of the state processes under risk-neutral measure is not uncommon for the reduced-form models with multidimensional state space, e.g. Jarrow, Li, Liu, and Wu (2006).

the expected present value of the discounted cash flows over the 'life span' of the mortgage contract (given by pricing equation (2.17)) where discount rate takes into account likelihood of premature termination by default or prepayment. The formula that takes into consideration amortizing nature of mortgage can be represented as:

$$\begin{aligned}
 V(0, T) &= \mathbb{E} \left(\sum_{n=1}^N \mathbb{Q}(\tau > t_{n-1}) e^{-\int_0^{t_n} \tilde{r}_s ds} CF(t_n) \right) \\
 \mathbb{Q}(\tau > t_{n-1}) &= \prod_{m=1}^{n-1} \left(1 - \sum_{\ell} \lambda_m^{(\ell)} \right), \quad \ell = d, p \\
 \tilde{r}_s &= (1 - \text{tax})r_s + \text{lq} \\
 CF(t_n) &= \lambda_n^{(d)} W(n) + \lambda_n^{(p)} A(n) + (1 - \sum_{\ell} \lambda_n^{(\ell)}) M \\
 \lambda_n^{(\ell)} &= l_n^{(\ell)} \exp(\mathbf{z}(t_n) \hat{\boldsymbol{\beta}}^{(\ell)})
 \end{aligned} \tag{4.1}$$

where the expectation is taken with respect to a martingale measure $\mathbb{Q}$ and $N = 360$ for a standard 30-year mortgage. r_s is an instantaneous spot rate, $\text{tax} = \text{tax}_F + \text{tax}_{ST}$. M stands for monthly mortgage payment adjusted for taxes:

$$M = (1 - \text{tax}_F) (1 - \text{tax}_{ST}) (1 + mr) \text{Prin}(n-1) + (L \cdot mc - (1 + mr) \text{Prin}(n-1)) \tag{4.2}$$

where tax_F is federal income tax rate, tax_{ST} – state income tax rate², mr is annual contract rate, $\text{Prin}(n)$ stands for unpaid principal after the n^{th} payment. mc is mortgage constant which is equivalent to:

$$mc = \frac{\frac{mr}{12}}{1 - \left(\frac{1}{1 + \frac{mr}{12}} \right)^{360}}$$

In case of prepayment (which may occur at any point in time but is recorded at monthly frequency) lender receives mortgage balance left unpaid after the n^{th} payment and regular monthly payment: $A(n) = \text{Prin}(n) + M$. $W(n)$ is the recovery value in case of default. I assume time-independent loss rate and parameterize random recovery rate as $\Delta = 1 - Z \frac{\text{Prin}(n)}{H(0)}$, where Z is a random variable distributed as *Uniform* (0,1), $H(0)$ is the

²While in principle these parameters could also be calibrated to the market data, for considerations of tractability they were chosen to be fixed at 28% (federal rate) and 4.32% (state rate).

value of the collateral (house) at the inception of the contract. Therefore, loss is assumed to be proportional to the ratio of unpaid balance to the initial value of the collateral. Schönbucher (2003) shows that the local expected value of recovery rate can be substituted for the value of random variable as long as stochastic recovery rate process is independent of other stochastic processes in the model and pertinent payoff is linearly related to recovery rates (*cf* Proposition 6.6). The resulting formula for the recovery rate obtains as:

$$\delta^e = 1 - \zeta^e \frac{Prin(n)}{H(0)} \quad (4.3)$$

where $\zeta^e = \mathbb{E}^{\mathbb{Q}}(Z)$. Let a be the percentage of the mortgage contract that is covered by insurance provided by the government agencies or private insurers (in this case $a = 1 - LTV_{>0.8}$). In case of default lender receives $W(n) = \min\{1, a + \delta^e\}A(n)$. This formulation for the recovery value implies that for the lender the cash flow in case of default on 100%-insured mortgage is the same as the cash flow in case of prepayment (ignoring any possibility of differential tax treatments).

The mapping between martingale intensity processes³ $l_{\mathbb{P}}^{(\ell)}$ and $l_{\mathbb{Q}}^{(\ell)}$ is given by the version of Cameron-Martin-Girsanov theorem (*cf* Duffie (2001), ch.11.I). If a nonexplosive counting process N_t on a filtered probability space $(\Omega, (\mathcal{G}_t), \mathbb{P})$, where filtration $\mathbb{G}$ is driven by d -dimensional Brownian motion $\mathbf{W}$, admits martingale intensity process $l_{\mathbb{P}}$ and ψ is a strictly positive predictable process such that $\int_0^T \psi_s l_{\mathbb{P},s} ds < \infty$ *a.s.*, the local martingale ξ is defined as:

$$\xi_t = \exp \left(\int_0^t (1 - \psi_s) l_{\mathbb{P},s} ds \right) \prod_{\{i: T(i) \leq t\}} \psi_{T(i)}, \quad t \leq T \quad (4.4)$$

where $T(i)$ is the i^{th} jump time of N . If ψ and $l_{\mathbb{P}}$ are bounded then ξ is a martingale. An equivalent probability measure $\mathbb{Q}$ is defined by $\frac{d\mathbb{Q}}{d\mathbb{P}} = \xi_T$. Counting process N admits intensity $l_{\mathbb{Q}} = \psi l_{\mathbb{P}}$ on $[0, T]$ and is doubly stochastic under risk-neutral measure $\mathbb{Q}$ with respect to the filtration $\mathbb{F} = \mathbb{G} \vee \sigma(\{N_s : 0 \leq s \leq t\})$. Furthermore, Theorem 4.8 in (Schönbucher 2003) shows states that there exists a predictable process φ such that

³This is the term used, for example, by McNeil, Frey, and Embrechts (2005).

$\int_0^t \|\varphi(s)\| ds < \infty$ which satisfies:

$$dW_{\mathbb{P}} = dW_{\mathbb{Q}} - \varphi(t)dt \quad (4.5)$$

One functional form choice for the martingale intensity process that satisfies the requirement of non-negativity is that of log-OU (Black-Karasinski⁴) diffusion process:

$$\begin{aligned} dx_t^{(\ell)} &= [a^{(\ell)}(t) - b^{(\ell)}(t)x_t^{(\ell)}]dt + \sigma^{(\ell)}(t)dW_t^{(\ell)} \\ \lambda_{0,t}^{(\ell)} &= \exp(x_t^{(\ell)}) \end{aligned} \quad (4.6)$$

With the choice $\sigma(t) = \sigma$, $b(t) = \kappa$, and $a(t) = \kappa\theta$ this becomes:

$$dx_{\mathbb{Q},t}^{(\ell)} = \psi^{(\ell)} \left(\kappa_{\mathbb{Q}}^{(\ell)} \left[\theta_{\mathbb{Q}}^{(\ell)} - x_{\mathbb{Q},t} \right] dt + \sigma^{(\ell)} dW_{\mathbb{Q},t}^{(\ell)} \right) \quad (4.7)$$

With "essentially affine" price of risk $\varphi(t) = \varphi_{11} + \varphi_{21}x_t$; $\varphi_{21} = \frac{\kappa_{\mathbb{Q}} - \kappa_{\mathbb{P}}}{\sigma}$; $\varphi_{11} = \frac{\kappa_{\mathbb{P}}\theta_{\mathbb{P}} - \kappa_{\mathbb{Q}}\theta_{\mathbb{Q}}}{\sigma}$ the mean-reverting character of the process under the physical probability measure is preserved under the risk-neutral measure⁵. I further assume simple specification for the price of the jump risk process⁶ $\psi(t)$: $\psi = \psi_{11}$. Jarrow, Lando, and Yu (2005) argue in favor of conditionally diversifiable risk premium for jump in value ($\psi = 1$), however, the empirical evidence so far is at best mixed.

4.2 MONTE CARLO IMPLEMENTATION

Absence of arbitrage implies that expected value of payoff of the mortgage contract under martingale measure $\mathbb{Q}$ at the time of origination equates market value of the mortgage. Since data on contractual amounts of loans, contract rates and upfront fees is at our disposal it is possible to evaluate conditional expectation in (4.2) using simulation methods.

⁴Berndt (2007) reports that in the empirical reduced-form study on large sample of corporate bonds this specification for default intensity exhibited best fit.

⁵For the discussion of the behavior of the default intensity under different measures with the specification for the price of risk process similar to the one used above see, e.g., Pan and Singleton (2008).

⁶Similar assumption is made, e.g., in Driessen (2005).

Simulation approach is appealing for several reasons, most importantly because of dependency of the payoff on the path of interest rates, house prices and latent termination factors and considerable number of state variables in the model. Conditional independence of the termination processes across the borrowers given observable characteristics which were used in the estimation of the empirical hazard models further facilitates Monte Carlo pricing procedure. This procedure is based on the simulation of risk-neutral default and prepayment intensity processes as specified in (4.7), interest rates (spot rate (2.42) and benchmark (index) rate) and future house prices (3.17). Large number (N) of independent sample paths of the state processes are simulated using either one of the appropriate discretization methods for SDE or, as in this case, using random variates from known transitional density. Using simulated values of state variables samples $v_1 = V(\omega_1), \dots, v_N = V(\omega_N)$ of the model-predicted value of the i^{th} mortgage contract are evaluated according to (4.2). Unbiased estimator for the conditional expectation $\mathbb{E}^Q$ of a random variable $V^{(i)}(\omega)$ obtains as $\frac{1}{N} \sum_{n=1}^N v_n^{(i)}$. Monte Carlo error $V - \hat{V}$ tends in the limit as $N \rightarrow \infty$ to mean zero normal random variable with standard deviation equal to the standard deviation of V divided by $\sqrt{N}$ according to the central limit theorem. CLT applies since samples x_n generated using pseudo-random number generator are independent.

Calibration of the remaining model parameters $\Theta = (\psi^{(d)}, \psi^{(p)}, \varphi^{(d)}, \varphi^{(p)}, \zeta^e, lq)$ to the observed market prices of mortgages is achieved via stochastic multivariate optimization. Key identity that should be satisfied in the no-arbitrage environment is represented by:

$$V(0, T) = Loan \cdot (1 - pts) \quad (4.8)$$

where *Loan* is the face value of the loan at origination, *pts* are upfront fees on the loan. In our data the information on points is not available; we have opted to use zero value for points, the resulting discrepancy being absorbed in the total pricing error. Hence, the

optimization problem can be formulated as :

$$\arg \min_{\Theta} \sum_{i=1}^I \left(\frac{V_i(0, T)}{Loan_i} - (1 - pts_i) \right)^2 \quad (4.9)$$

I have chosen random subsets of mortgages originated within the same calendar months. Since the MSAs are represented in the sample unequally in different time periods and I deemed it desirable that contracts from all MSAs be represented in the calibration subsample at any time period, the size of subsample was chosen to be in the range of 250-320. The calibration procedure is quite expensive computationally, therefore the upper limit on the size of a subsample was set by available computational resources⁷. Since the dimensionality of the problem is quite high, I have opted to choose an evolutionary algorithm instead of more traditional methods based on stochastic gradient approximation. The particular method was differential evolution (DE) which involves fitness-based selection from a population of S vectors each of which represents a trial solution to the problem (Brabazon and O'Neill 2006), (Price, Storn, and Lampinen 2006). Fitness function is given by sum of the Euclidian distances of the calibrated value (mean over N Monte Carlo samples; in my case $N = 1000$) to the actual value of the loan for the subsample of K loans:

$$Fitness = \sum_{i=1}^K \left(\frac{1}{N} \sum_{n=1}^N v_n^{(i)} - Loan^{(i)} \right)^2 \quad (4.10)$$

The algorithm proceeds as follows:

Algorithm 3 Differential evolution optimization.

Step 1 Choose S 'individuals' (vectors Θ) at random from some distribution (I chose uniform) that covers the range of the possible solutions

Step 2 Evaluate fitness of each 'individual'

⁷Calibration was performed on Intel Xeon quad-core 3.16 GHz workstation with 16 Gb of RAM.

Step 3 Mutation: combine characteristics of randomly (or purposefully) selected 'individuals' (vectors) into a *variant* vector

Step 4 Crossover: combine the variant vector with each 'individual' forming S *trial* vectors

Step 5 Evaluate fitness of each trial 'individual' and perform selection: keep old 'individual' if his fitness score is not surpassed by that of the new vector, otherwise, keep the new vector

The Matlab implementation of the DE optimization procedure uses free software developed by R. Storn, K. Price and others which accompanies the book by Price, Storn, and Lampinen (2006).

As a practical matter, I have restricted the range of the possible values of $\psi^{(\ell)}$ close to unity⁸ in accordance with conditional diversification argument of Jarrow, Lando, and Yu (2005). Values of the liquidity parameter are in the range (1E-6,1E-2), the upper bound for the expected loss ζ^e is set at 0.75 – approaching this boundary is considered to be the sign of the model's inability to explain observed market prices without resorting to implausibly high assumptions about losses incurred in the case of default.

4.3 APPLICATION OF THE MODEL TO NON-PRIME MORTGAGE MARKET

The valuation model described above is calibrated using subsamples of mortgages originated throughout 2002-2007. If there was indeed a shift in the unobserved characteristics of non-prime borrowers which revealed itself in the altered default and prepayment behavior, then there are at least three possible scenarios in regard to the market reaction. Since all the mortgages in our sample come from the MBS pools the 'market' is understood in the sense of the secondary market. First, market might have been unaware of the changes in the borrowers' traits. If that was the case then testable implications will be as

⁸The range of possible values is limited to (0.99, 1.25).

follows: the "old" pricing model is going to apply to the contracts originated under the "new" regime. Alternatively, secondary market might have suspected that the regime had changed, but there was considerable uncertainty as to what the true characteristics of that "new" regime were. In this case one may expect that the performance of the model (in terms of parameter 'sanity' and / or accuracy of pricing) estimated under the "old" regime on the subsamples originated under "new" regime would be worse, but the "new" model may not be sufficiently adequate as well. Third possibility which I am considering is that the secondary market might have learned of a regime shift and of the features of the "new" regime – this hypothesis is consistent with inadequacy of the "old" pricing model for the subsamples originated under "new" regime. On the contrary, the "new" model is expected to perform adequately. Results presented in tables 4.1 and 4.2 are largely consistent with the latter hypothesis. Observationally, there is variation in the additive risk adjustments φ_{11} , φ_{21} , both in default and prepayment prices of risk. However, the cumulative value of price of risk remains negative⁹ in all cases. Loss parameter ζ^e seems to be exhibiting the greatest degree of variation: calibration early in the "old" regime yields values typically in the range of (0.25, 0.5) , however, as we're approaching the end of 2004, the expected loss¹⁰ rises above 0.5 and exceeds 0.7 in early 2005 subsample.

For the "new" model, which we begin to calibrate on the subsample from December 2004, initially we observe high values of expected loss and liquidity parameters. However, as we move on in calendar time these values settle in the range 0.3-0.5 for expected loss and 40-60 basis points for liquidity premium. The calibrated loss value begin to increase in 2007 subsamples and is already very high for mortgages originated in Spring (at about the same time problems of many subprime mortgage originators became public knowledge). Indeed, market perception of riskiness of subprime mortgage pools and, consequently, spreads of subprime-backed MBS was changing rapidly about that time (see figure 4.1).

⁹I report the value of total additive risk adjustment as $\varphi_{11}^{(\ell)} + \varphi_{21}^{(\ell)} \hat{\theta}_{\mathbb{P}}^{(\ell)}$.

¹⁰Recall that this is expected value of loss for the lender in case of default under the martingale measure $\mathbb{Q}$ rather than under the real-world probability measure.

For the subsamples of loans originated in June-July 2007 our calibration procedure does not allow to obtain satisfactory estimates as even after dozens of iterations the pricing error still remains much greater in magnitude than that reported for the earlier periods¹¹. It seems that secondary mortgage market processed information in an efficient manner so that "new" model started to dominate the old one soon after the time period in which, as the results of statistical tests suggest, the switch from the "old" regime to the "new" one occurred.

4.3.1 PROBABILITIES OF DEFAULT: MONTE CARLO ESTIMATES

In Chapter 3 I calculated default hazards for a randomly chosen contract under the old regime and under the new one using empirical estimates of the baseline hazards of the two regimes and graphed these hazards on figure 3.3. This supposedly should have given an idea of how the "new" hazard differs from the "old" one. Results presented in table 4.3 extend this experiment building upon the estimates obtained in Chapters 5 and 6. For randomly chosen mortgages from two different regions I calculate the forecast default probabilities (cumulative hazards of default) for different time horizons under the objective probability measure $\mathbb{P}$ using estimates of the parameters of latent baseline processes (table 3.23), term structure of interest rates (table 3.17), and housing processes for these regions (backed from the estimates of relative housing returns process reported in table 3.19) as means of 2,000 Monte Carlo scenarios. Comparison of the cumulative hazards under the two regime supports the results of the analysis in Chapter 3. Simulated hazard of default is uniformly greater under the second regime (column 3 of table 4.3). Depending on the region and the time horizon the ratio of the two varies in the range from 1.1 to 1.8. However, this increase, though noticeable, still is insufficient to fully account for the serious increase in default rates of the latest vintages of ARMs (*cf* figure 3.6). However, when house price process is perturbed by the 30% shock (which roughly corresponds to the

¹¹One iteration of evolutionary algorithm with population of 50-60 'individuals' currently executes about 70-80 minutes.

magnitude of the decrease in the house prices in the MSAs under consideration) the cumulative hazard of default increases 2-3 times compared to the base case ("old" regime – column 2) and by 70%-100% compared to the "new" regime (column 3). Thus simulations based on our estimated model parameters tend to confirm the point of view that the surge in subprime mortgage default can be for the most part attributed to the precipitous fall in house prices (primarily in the regions that experienced highest house price appreciation in the previous years). It should be noted, though, that probability of house price swings of such magnitude is quite low in our model¹², even though these swings are actually part of the data used for the estimation of the house process parameters. It is not improbable that models that were used by mortgage originators and other market participants attached even smaller *ex ante* probabilities to such house price scenarios.

¹²For San Diego the probability of decline in prices similar to one that occurred in 2006-2009 is below 8%.

Table 4.1

Calibrated risk adjustments, liquidity and expected loss parameters for the 'old' regime model (estimated on Q1 1997- Q2 2004 data)

Param.	June 2002	May 2003	March 2004	Sept. 2004	Oct. 2004	Dec. 2004	Feb. 2005
ψ^p	1.1348	1.1932	1.1901	1.0640	1.1551	1.1676	1.0073
φ_{11}^p	-11.240	-9.9847	-8.5336	-9.4549	-9.7261	-9.1911	-11.565
φ_{21}^p	-0.0524	0.1765	0.0888	0.1369	-0.0767	-0.0120	0.0746
$\varphi_{11}^p + \varphi_{21}^p \theta^p$	-10.894	-11.148	-9.1187	-10.357	-9.2210	-9.1123	-12.057
ψ^d	1.1860	1.0713	1.1213	1.0535	1.0415	1.1517	0.9909
φ_{11}^d	0.2315	-0.3496	0.2522	-0.2482	-0.1783	0.7704	0.4526
φ_{11}^d	0.1214	0.0012	0.0914	-0.0270	-0.0281	0.1566	0.0933
$\varphi_{11}^d + \varphi_{21}^d \theta^d$	-0.4219	-0.3558	-0.2399	-0.1029	-0.0269	-0.0724	-0.0494
Lq	0.0045	0.0066	0.0064	0.0059	0.0054	0.0056	0.0064
Loss	0.3144	0.2627	0.3381	0.4807	0.4992	0.5937	0.7114
RMSE, %	0.59	0.54	0.65	0.45	0.62	0.50	0.54
No. mtg	283	304	284	275	260	261	303

Normalized RMSE is calculated as $\frac{RMSE}{Loan_{max} - Loan_{min}}$. ψ^ℓ denotes multiplicative price of risk parameter, φ_{11}^ℓ and φ_{21}^ℓ – additive price of risk parameters, while $\varphi_{11}^\ell + \varphi_{21}^\ell \theta^\ell$ represents approximate cumulative value of additive drift adjustment. Lq stands for liquidity adjustment in 4.2; Loss – for the expected loss (ζ^e in 4.3).

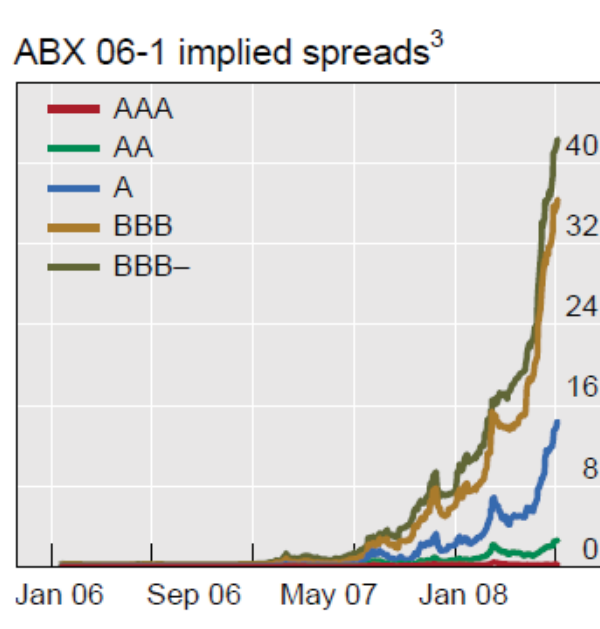


Figure 4.1

Spreads between subprime credit default swaps index (ABX) and LIBOR, thousands basis points

Reproduced from page 69 in Fender, Ingo and Martin Scheicher "The ABX: how do the markets price subprime mortgage risk?" *BIS Quarterly Review*, September 2008, pp 67–81.

Table 4.2

Calibrated risk adjustments, liquidity and expected loss parameters for the 'new' regime model (estimated on post-2004 data)

Parameter	Dec. 2004	Febr. 2005	May 2005	July 2005	Nov. 2005
Panel A. Subsamples of mortgages originated in Dec. 2004 - Nov. 2005					
ψ^p	1.0000	1.0896	1.0742	1.1883	1.0648
φ_{11}^p	-5.9075	-7.5450	-7.7872	-3.3601	-15.174
φ_{21}^p	0.1974	-0.0798	-0.0796	0.1634	0.0231
$\varphi_{11}^p + \varphi_{21}^p \theta^p$	-7.5259	-6.8904	-7.1348	-4.7001	-15.363
ψ^d	0.9921	1.0498	1.1915	1.1762	1.1181
φ_{11}^d	0.4471	0.3751	0.0266	0.7776	0.4161
φ_{21}^d	0.0702	0.0490	0.0145	0.1185	0.0643
$\varphi_{11}^d + \varphi_{21}^d \theta^d$	-0.1282	-0.0267	-0.0927	-0.1943	-0.1113
Lq	0.0072	0.0067	0.0052	0.0061	0.0047
Loss	0.6754	0.5881	0.3021	0.3156	0.3813
RMSE, %	0.49	0.54	0.72	0.53	0.84
No. mtg	261	303	293	299	301
Panel B. Subsamples of mortgages originated in Feb. 2006 - March 2007					
	Feb. 2006	June 2006	Oct. 2006	Jan. 2007	March 2007
ψ^p	1.0584	1.1988	1.1748	1.1905	1.1866
φ_{11}^p	-6.0607	-12.612	-10.890	-14.885	-11.415
φ_{21}^p	-0.0245	-0.0466	0.1495	0.1763	0.1978
$\varphi_{11}^p + \varphi_{21}^p \theta^p$	-5.8877	-12.283	-11.946	-16.131	-12.813
ψ^d	1.0844	0.9897	1.0253	1.0977	1.0761
φ_{11}^d	0.1259	0.4540	1.1793	0.1446	0.6608
φ_{21}^d	0.0201	0.0670	0.1573	0.0377	0.0822
$\varphi_{11}^d + \varphi_{21}^d \theta^d$	-0.0391	-0.0950	-0.1107	-0.1646	-0.0132
Lq	0.0041	0.0063	0.0063	0.0070	0.0062
Loss	0.4779	0.4226	0.4819	0.6953	0.7156
RMSE, %	0.71	0.70	1.12	0.83	1.14
No. mtg	301	301	291	299	311

Normalized RMSE is calculated as $\frac{RMSE}{Loan_{max} - Loan_{min}}$. ψ^ℓ denotes multiplicative price of risk parameter, φ_{11}^ℓ and φ_{21}^ℓ – additive price of risk parameters, while $\varphi_{11}^\ell + \varphi_{21}^\ell \theta^\ell$ represents approximate cumulative value of additive drift adjustment. Lq stands for liquidity adjustment in 4.2; Loss – for the expected loss (ζ^e in 4.3).

Table 4.3
Predicted hazard of default under the 'old' and the 'new' regime

	'New' regime	'Old' regime
Panel A. In the absence of house price drop		
Contract 'A'	5.05%	4.86%
2-year horizon		
Contract 'A'	15.84%	9.91%
4.7-year horizon		
Contract 'B'	4.74%	4.28%
2-year horizon		
Contract 'B'	13.22%	9.50%
4.7-year horizon		
Panel B. With 30% house price drop		
Contract 'A'	5.05%	4.86%
2-year horizon		
Contract 'A'	30.08%	17.33%
4.7-year horizon		
Contract 'B'	4.74%	4.28%
2-year horizon		
Contract 'B'	22.46%	16.28%
4.7-year horizon		

Contract characteristics for contract 'A': origination – San Diego, October 2005, loan amount – \$ 543,268 (in constant year 2000 dollars), LTV at origination – 85%, contract rate at origination – 6.75 %, margin – 3.03%, lifetime cap – 13.75%, periodic cap – 1 perc. pt, lifetime floor - 6.75%; purchase loan, FICO at origination – 657.

Contract characteristics for contract 'B': Tampa, November 2005, loan amount – \$ 92,742.86, LTV at origination – 90%, contract rate at origination – 6.625 %, margin – 4%, lifetime cap – 12.625%, periodic cap – 1 perc. pt, lifetime floor - 4 %; loan is taken out for refinancing, subject to prepayment penalty, FICO at origination – 629.

Reported cumulative hazards of default are means of the 2,000 simulations (under the real-world probability measure). House price drop is assumed to occur 35 months past the origination of the contract.

CHAPTER 5

CONCLUSION

Reduced-form approach to the valuation of contingent claims subject to credit risk is considered a flexible alternative to the traditional structural models. In the mortgage pricing context, it unites duration analysis (competing risk variant thereof) with unobserved components models. In the present study, I have applied a reduced-form model in the spirit of Duffie and Singleton (1999) to the loan-level data set which contains payment histories of non-prime adjustable rate residential mortgages. Conditional on observed loan and borrower's characteristics and changing interest rate and house prices, probability of borrower's default (prepayment) is dependent on the unobserved factors evolving over time. These latent factors can be interpreted quite broadly: first, they can reflect changing economic conditions to the extent these are not captured by dynamic covariates in the duration model, second, they may incorporate dynamic observed and intangible costs of refinancing and default. The latter dynamics may be due to technological innovation or to changing attitudes and mores. The model can also be interpreted as a dynamic frailty one (see, e.g. Aalen, Borgan, and Gjessing (2008, ch. 11) for an extended discussion). The probability of mortgage termination is considered to be a composition of a deterministic part (estimated covariates' effects) and a stochastic part (baseline default (prepayment)). An assumption that the actual occurrence of default (prepayment) given the probability of this event is uncertain adds an additional 'layer' of stochasticity to the model¹.

Present work, while building on model presented in Kau, Keenan, and Smurov (2004), adds an important component to default and prepayment models – namely, house prices.

¹This latter 'layer' reflects idiosyncratic risk of prepayment (default).

Numerous empirical studies² have corroborated the prediction of structural models that at a certain point decline in the value of home equity is the dominant determinant of the decision to default on a home loan. The stochastic house price process (which is linked to the term structure of interest rates by design) enters into the model in a straightforward way - via the estimate of borrower's equity, that is represented by a covariate (covariates).

A number of studies documented mean reversion in the time series of house prices (see, e.g., Glaeser, Gyourko, and Saiz (2008)), therefore, the mean-reverting characterization of the house price process seems a natural choice. Interest rates follow mean-reverting process as well; to preserve the link between house prices and interest rates and to ensure that the changes in interest rates induce opposite changes in house prices, the mean-reverting form is posited for the ratio of returns to holding house to returns from investment in long-term bonds³. This parsimonious specification at the same time allows for the estimation of a separate set of parameters for each MSA.

The pricing model developed in this study was applied to the subsamples of mortgages originated earlier in the decade and later on (in the last 2-3 years of subprime boom). Our analysis suggests that there is limited evidence in favor of the shift in the hidden characteristics (behavior) of the subprime borrowers towards higher propensity to default. Results of experiment presented in table 3.15 suggest that although the cumulative hazard of default for a representative contract increased under the second regime to 24% from 11% under the first regime the increase was largely due to the worsening of the environment. The hazard would have increased only to 15%, had house prices been constant. This experiment was based on the empirical (Nelson-Aalen) estimates of the baseline hazards for both regimes. The estimation of the complete set of the model parameters allowed me to repeat this experiment, this time using the house process model. The

²Deng, Quigley, and Van Order (2000), Guiso, Sapienza, and Zingales (2009), among others.

³Several contributions to the 'structural' credit risk literature used continuous-time mean-reverting specifications for the ratios, for instance, Collin-Dufresne and Goldstein (2001). In the asset-pricing literature, Piazzesi, Schneider, and Tuzel (2007) is an example of a model which parameters authors estimate using long-run price-dividend ratios.

findings reported in table 4.3 are largely consistent with the aforementioned results. Simulated hazard of default under the 'new' regime holding other factors constant increased by 13%-59% relative to the 'old' regime. This is to be compared with 75%-100% increase in the hazard of default under the new regime as a result of the 30% shock to simulated house prices. Thus, the application of our model to the non-prime adjustable rate mortgage data indicates that the greater share of the surge in defaults in the recent years can be explained by the effect of observable variables, first and foremost by the steep drop in house prices.

Speaking of the potential avenues of the future research, it might be attractive to augment the set of the state variables by other factors that have an impact on the decision to default (prepay), such as, for example, unemployment⁴ or by macroeconomic variables representing the business cycle such as gross national⁵ or regional product. Another possible extension is to preserve the independency of default and prepayment latent factors from other stochastic state variables, but to allow for more sophisticated correlation structure of the latent variables: either for instantaneous correlation between the two and / or for dependency of the 'random draws' of the termination processes⁶. Finally, while there are certain benefits of consistency in using the same data set for the estimation of physical conditional probabilities (hazards) of default and prepayment and subsequent risk-adjustments calibration, certain data limitations (such as the lack of information on upfront fees (points)) complicate this calibration procedure. In principle, one might use market data such as series of prices of credit-default swaps for this purpose, were such data available.

⁴Unemployment represented by a stochastic process driven by independent Brownian motion is relatively straightforward extension of the model. Alternatively, the form of each of the state processes can be chosen to be simple as, e.g., in de Jong, Driessen, and Van Hemert (2007), however, explicit instantaneous correlation between the various sources of uncertainty can be introduced.

⁵As, for example, in Kolbe and Zagst (2008).

⁶In terms of estimation, this would introduce dependence not only on the previous observation in the same stratum, but on observations from earlier strata.

BIBLIOGRAPHY

- Aalen, Odd, Ørnulf Borgan, and Håkon Gjessing, 2008, *Survival and Event History Analysis: A Process Point of View* (Springer: Berlin , Heidelberg, New York).
- Abbring, Jaap, 2001, Econometric duration and event-history analysis, Lecture Notes, University of Chicago.
- , and Gerard J. Van Den Berg, 2007, The unobserved heterogeneity distribution in duration analysis, *Biometrika* 94, 87–99.
- Abraham, Jesse, and William Schauman, 1991, New evidence on house prices from Freddie Mac repeat sales, *AREUEA Journal* 19, 333–352.
- Aït-Sahalia, Yacine, 2002, Maximum likelihood estimation of discretely sampled diffusions: A closed form approximation approach, *Econometrica* 70, 223–262.
- Akritis, M., 1988, Pearson-type goodness-of-fit tests: the univariate case, *Journal of the American Statistical Association* 83, 222–230.
- Ambrose, Brent, and Anthony Sanders, 2003, Commercial mortgage-backed securities: Prepayment and default, *Journal of Real Estate Finance and Economics* 26, 179–196.
- Andersen, Per, Ørnulf Borgan, Richard Gill, and Nils Keiding, 1993, *Statistical Models Based on Counting Processes* (Springer: Berlin , Heidelberg, New York).
- Andrieu, Christoph, Arnaud Doucet, and Roman Holenstein, 2007, Particle Markov Chain Monte Carlo, Working paper, Department of Mathematics, University of Bristol.

- Artzner, Philippe, and Freddy Delbaen, 1995, Default risk insurance and incomplete markets, *Mathematical Finance* 5, 187–195.
- Azizpour, Shahriar, and Kay Giesecke, 2008, Premia for correlated default risk, Working Paper, Stanford University.
- Babbs, Simon H., and K. Ben Nowman, 1999, Kalman filtering of generalized Vasicek term structure models, *Journal of Financial and Quantitative Analysis* 34, 115–130.
- Bajari, Patrick, Chenghuan Sean Chu, and Minjung Park, 2008, An empirical model of subprime mortgage default from 2000 to 2007, NBER Working Paper 14625.
- Bakshi, Gurdip, Dilip Madan, and Frank Zhang, 2006, Investigating the role of systematic and firm-specific factors in default risk: Lessons from empirically evaluating credit risk models, *Journal of Business* 79, 1955–1987.
- Bauwens, Luc, and Nikolaus Hautsch, 2006, Stochastic conditional intensity processes, *Journal of Financial Econometrics* 4, 450–493.
- , 2007, Modelling financial high frequency data using point processes, CRC Discussion Paper No.2007-066.
- Bauwens, Luc, and David Veredas, 2004, The stochastic conditional duration model: A latent factor model for the analysis of financial durations, *Journal of Econometrics* 119, 381–412.
- Belanger, Alain, Steven E. Shreve, and Dennis Wong, 2004, A general framework for pricing credit risk, *Mathematical Finance* 14, 317–350.
- Bennett, Paul, Richard Peach, and Stavros Peristiani, 2001, Structural change in the mortgage market and the propensity to refinance, *Journal of Money, Credit, and Banking* 33, 955–975.

- Berndt, Antje, 2007, Specification analysis of reduced-form credit risk models, Working paper.
- , Rohan Douglas, Darrell Duffie, Mark Ferguson, and David Schranz, 2008, Measuring default risk premia from default swap rates and EDFs, Working paper.
- Bielecki, Tomasz R., and Marek Rutkowski, 2002, *Credit Risk: Modeling, Valuation and Hedging* (Springer: Berlin , Heidelberg, New York).
- Bingham, Nicholas H., and Rudiger Kiesel, 2004, *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives* (Springer: Berlin , Heidelberg, New York).
- Björk, Tomas, 2004, *Arbitrage Theory in Continuous time* (Oxford University Press: Oxford, UK).
- Black, Fisher, and John Cox, 1976, Valuing corporate securities: Some effects of bond indenture provisions, *Journal of Finance* 31, 351–367.
- Black, Fisher, and Piotr Karasinski, 1991, Bond and option pricing when short rates are log-normal, *Financial Analysts Journal* 47, 52–59.
- Black, Fisher, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 81–98.
- Brabazon, Anthony, and Michael O’Neill, 2006, *Biologically Inspired Algorithms for Financial Modelling* (Springer: Berlin, Heidelberg, New York).
- Brandt, Michael, and Pedro Santa-Clara, 2002, Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets, *Journal of Financial Economics* 63, 161–210.
- Brigo, Damiano, and Fabio Mercurio, 2006, *Interest Rate Models – Theory and Practice* (Springer: Berlin , Heidelberg, New York).

Cameron, Adrian C., and Pravin K. Trivedi, 1998, *Regression Analysis of Count Data* (Cambridge University Press: New York, NY) Econometric society monograph No. 30.

———, 2005, *Microeconometrics: Methods and Applications* (Cambridge University Press: New York, NY).

Campbell, John Y., and Joao Cocco, 2003, Household risk management and optimal mortgage choice, *Quarterly Journal of Economics* 118, 1449–1494.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton University Press: Princeton, NJ).

Capozza, Dennis R., Patric H. Hendershott, and Charlotte Mack, 2004, An anatomy of price dynamics in illiquid markets: Analysis and evidence from local housing markets, *Real Estate Economics* 31, 1–32.

Cappé, Olivier, Eric Moulines, and Tobias Rydén, 2005, *Inference in Hidden Markov Models* (Springer: NY).

Carlin, Brad, Nicholas Polson, and D. Stoffer, 1992, A Monte Carlo approach to nonnormal and nonlinear state-space modeling, *Journal of the American Statistical Association* 87, 493–500.

Case, Karl E., and Robert J. Shiller, 1989, The efficiency of the market for single-family homes, *American Economic Review* 79, 125–137.

———, 1996, Mortgage default risk and real estate prices : the use of index-based futures and options in real estate, *Journal of Housing Research* 7, 243–258.

Chan, K.C., G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209–1227.

- Chen, Ren-RAW, and Louis Scott, 1992, Pricing interest rate options in a two-factor Cox-Ingersoll-Ross model of the termstructure, *Review of Financial Studies* 5, 613–636.
- , 2003, Multi-factor Cox-Ingersoll-Ross model of the termstructure: Estimates and tests from a Kalman filter model, *Journal of Real Estate Finance and Economics* 27, 143–172.
- Chib, Siddhartha, Michael K. Pitt, and Neil Shephard, 2006, Likelihood based inference for diffusion driven state space models, Working Paper.
- Chopin, Nicolas, 2004, Central limit theorem for Sequential Monte Carlo methods and its application to Bayesian inference, *The Annals of Statistics* 32, 2385–2411.
- Ciochetti, Brian A., Yongheng Deng, Bin Gao, and Rui Yao, 2002, The termination of mortgage contracts through prepayment and default in the commercial mortgage markets: A proportional hazard approach with competing risks, *Real Estate Economics* 30, 595–633.
- Clapp, John M., Yongheng Deng, and Xudong An, 2006, Unobserved heterogeneity in models of competing mortgage termination risks, *Real Estate Economics* 34, 243–273.
- Clapp, John M., Gerson M. Goldberg, John P. Harding, and Michael LaCour-Little, 2001, Movers and shuckers: Interdependent prepayment decisions, *Real Estate Economics* 29, 411–450.
- Clebaner, Fima C., 1998, *Introduction to Stochastic Calculus with Applications* (Imperial College Press: London).
- Cocco, Joao, 2005, Portfolio choice in the presence of housing, *Review of Financial Studies* 18, 535–567.
- Collin-Dufresne, Pierre, and Rita Goldstein, 2001, Do credit spreads reflect stationary leverage ratios?, *Journal of Finance* 56, 1929–1958.

- Collin-Dufresne, Pierre, R. Goldstein, and J. Hugonnier, 2004, A general formula for valuing defaultable securities, *Econometrica* 72, 1377–1407.
- Cox, David R., 1972, Regression models and life tables (with discussion), *Journal of the Royal Statistical Society, series B* 34, 187–220.
- , and Valerie Isham, 1980, *Point Processes* (Chapman and Hall: London and New York).
- Cox, David R., and David Oakes, 1984, *Analysis of Survival Data* (Chapman and Hall: London and New York).
- Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–407.
- Crawford, Gordon W., and Michael C. Fratantoni, 2003, Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices, *Real Estate Economics* 31, 223–243.
- Cunningham, Donald F., and Charles A. Capone, 1990, The relative termination experience of adjustable to fixed-rate mortgages, *Journal of Finance* 45, 1687–1703.
- Cunningham, Donald F., and Patrick H. Hendershott, 1984, Pricing FHA mortgage default insurance, *Housing Finance Review* 3, 383–392.
- Dai, Qiang, and Kenneth Singleton, 2003, Term structure dynamics in theory and reality, *Review of Financial Studies* 16, 631–678.
- Daley, Daryl J., and David Vere-Jones, 2003, *An Introduction to the Theory of Point Processes* (Springer: New York) 2 ed.
- Das, Sanjiv R., Darrell Duffie, Nikunj Kapadia, Leandro Saita, and Ke Wang, 2007, Common failings: How corporate defaults are correlated, *Journal of Finance* 62, 93–118.

- de Jong, Frank, Joost Driessen, and Otto Van Hemert, 2007, Hedging house price risk: Portfolio choice with housing futures, Discussion paper, New York University.
- DeGiorgi, Enrico, and Jurg Burkhard, 2006, An intensity based non-parametric default model for residential mortgage portfolios, *Journal of Risk* 8, 57–95.
- Dell’Ariccia, Giovanni, Deniz Igan, and Luc Laeven, 2008, Credit booms and lending standards: Evidence from the subprime mortgage market, Discussion paper, IMF.
- Demyanyk, Yuliya, and Otto Van Hemert, 2009, Understanding the subprime mortgage crisis, *Review of Financial Studies* forthcoming.
- Deng, Yongheng, 1997, Mortgage termination: An empirical hazard model with a stochastic term structure, *Journal of Real Estate Finance and Economics* 14, 309–331.
- , John M. Quigley, and Robert Van Order, 2000, Mortgage termination, heterogeneity and the exercise of mortgage options, *Econometrica* 68, 275–307.
- Dolde, Walter, and Dogan Tirtiroglu, 2002, Housing price volatility changes and their effects, *Real Estate Economics* 30, 41–66.
- Doucet, Arnaud, M. Briers, and S. Senecal, 2006, Efficient block sampling strategies for sequential Monte Carlo, *Journal of Computational and Graphical Statistics* 15, 693–711.
- DOUCET, ARNAUD, NANDO DE FREITAS, AND NEIL GORDON (eds.), 2001, *Sequential Monte Carlo Methods in Practice* (Springer).
- Doucet, Arnaud, and Adam M. Johansen, 2008, A tutorial on particle filtering and smoothing: fifteen years later, Working Paper.
- Downing, Chris, Dwight Jaffee, and Nancy Wallace, 2010, Is the market for mortgage-backed securities a market for lemons?, *Review of Financial Studies* forthcoming.

- Downing, Christopher, Richard Stanton, and Nancy E. Wallace, 2007, Volatility, mortgage default, and CMBS subordination, Discussion paper, University of California, Berkeley.
- Driessen, Joost, 2005, Is default event risk priced in corporate bonds?, *Review of Financial Studies* 18, 165–195.
- Duan, Jin-Chuan, 1997, Augmented GARCH (p,q) process and its diffusion limit, *Journal of Econometrics* 79, 97–127.
- , and Jean-Guy Simonato, 1999, Estimating and testing exponential-affine term structure models by Kalman filter, *Review of Quantitative Finance and Accounting* 13, 111–135.
- Duffee, Gregory, 1999, Estimating the price of default risk, *Review of Financial Studies* 12, 197–226.
- Duffie, Darrell, 2001, *Dynamic Asset Pricing Theory* (Princeton University Press: Princeton, NJ).
- , Andreas Eckner, Guillaume Horel, and Leandro Saita, 2009, Frailty correlated default, *Journal of Finance* 64, 2089–2123.
- Duffie, Darrell, and Nicolae Gârleanu, 2001, Risk and valuation of collateralized debt obligations, *Financial Analysts Journal* 57, 41–59.
- Duffie, Darrell, and Rui Kan, 1996, A yield factor model of interest rates, *Mathematical Finance* 6, 379–406.
- Duffie, Darrell, and David Lando, 2001, Term structures of credit spreads with incomplete accounting information, *Econometrica* 69, 633–664.
- Duffie, Darrell, L. Pedersen, and Kenneth J. Singleton, 2003, Modeling credit spreads in sovereign debt: A case study of Russian bonds, *Journal of Finance* 55, 119–159.

- Duffie, Darrell, Leandro Saita, and Ke Wang, 2006, Multi-period corporate default prediction with stochastic covariates, *Journal of Financial Economics* 83, 635–665.
- Duffie, Darrell, and Kenneth J. Singleton, 1997, An econometric model of the term structure of interest rate swap spreads, *Journal of Finance* 52, 1287—1321.
- , 1999, Modeling term structures of defaultable bonds, *Review of Financial Studies* 12, 687–720.
- , 2003, *Credit Risk: Pricing, Measurement, and Management* (Princeton University Press: Princeton, NJ).
- Dunn, Kenneth B., and John J. McConnell, 1981, Valuation of GNMA mortgage-backed securities, *Journal of Finance* 36, 599–616.
- Dunn, Kenneth B., and Chester S. Spatt, 2005, The effect of refinancing costs and market imperfections on the optimal call strategy and the pricing of debt contracts, *Real Estate Economics* 33, 595–617.
- Durbin, James, and Siem Jan Koopman, 2001, *Time Series Analysis by State Space Methods* (Oxford University Press: Oxford).
- Durham, Garland, and A. Ronald Gallant, 2002, Numerical techniques for maximum likelihood estimation of continuous time diffusion processes, *Journal of Business and Economic Statistics* 20, 297–338.
- Elerian, Ola, Siddhartha Chib, and Neil Shephard, 2001, Likelihood inference for discretely observed nonlinear diffusions, *Econometrica* 69, 959–993.
- Elliot, G., T. G. Rothenberg, and J.H. Stock, 1996, Efficient tests for an autoregressive unit root, *Econometrica* 64, 813–836.
- Eom, Young H., Jean Helwege, and Jing Huang, 2004, Structural models of corporate bond pricing: An empirical analysis, *Review of Financial Studies* 17, 499–544.

- Ericsson, Jan, Kris Jacobs, and Rodolfo Oviedo-Helfenberger, 2004, The determinants of credit default swap premia, *CIRANO*, 2004s-55.
- Firestone, Simon, Robert Van Order, and Peter Zorn, 2007, The performance of low-income and minority mortgages, *Real Estate Economics* 35, 479–504.
- Flavin, Marjorie, and Takashi Yamashita, 2002, Owner-occupied housing and the composition of the household portfolio, *American Economic Review* 92, 345–362.
- Fleming, Thomas R., and David P. Harrington, 1991, *Counting Processes and Survival Analysis* (Wiley: New York).
- Florens, Jean-Pierre, Denis Fougere, and Michel Mouchart, 2008, Duration models and point processes, in L. Matyas, and P. Sevestre, ed.: *The Econometrics of Panel Data* . vol. 46 (Springer).
- Florens-Zmirou, Danielle, 1993, On estimating the diffusion coefficient from discrete observations, *Journal of Applied Probability* 30, 790–804.
- Foote, Christopher L., Kristopher Gerardi, and Paul S. Willen, 2008, Negative equity and foreclosure: Theory and evidence, *Journal of Urban Economics* 64, 234–245.
- Gagliardini, Patrick, and Christian Gourieroux, 2005, Stochastic migration models with application to corporate risk, *Journal of Financial Econometrics* 3, 188–226.
- Gallant, Ronald, and George Tauchen, 2005, Simulated score methods and indirect inference for continuous-time models, in Y. Ait-Sahalia, and L. Hansen, ed.: *Handbook of Financial Econometrics* (Elsevier Science) Amsterdam.
- Gamerman, Dani, and Hedibert F. Lopes, 2006, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (Chapman & Hall / CRC: Boca Raton, London, New York).
- Gaure, Simen, Knut Røed, and Tao Zhang, 2007, Time and causality: A Monte Carlo assessment of the timing-of-events approach, *Journal of Econometrics* 141, 1159–1195.

- Gerardi, Kristopher, Adam H. Shapiro, and Paul S. Willen, 2008, Subprime outcomes: Risky mortgages, homeownership experiences, and foreclosures, Federal Reserve Bank of Boston, Working Paper 07-15.
- Gerardi, Kristopher S., Andreas Lehnert, Shane M. Sherland, and Paul S. Willen, 2009, Making sense of the subprime crisis, Federal Reserve Bank of Atlanta, Working Paper 2009-2.
- Geske, Robert, 1977, The valuation of corporate liabilities as compound options, *Journal of Financial and Quantitative Analysis* 12, 541–552.
- Geyer, Alois L., and Stefan Pichler, 1999, A state-space approach to estimate and test multifactor Cox-Ingersoll-Ross models of the term structure, *Journal of Financial Research* 22, 107–130.
- Gibbons, Michael R., and Krishna Ramaswamy, 1993, A test of the Cox, Ingersoll and Ross model of the term structure, *Review of Financial Studies* 6, 620–658.
- Giesecke, Kay, 2004, Credit risk modeling and valuation: An introduction, in D. Shimko, ed.: *Credit Risk: Models and Management*, Vol. 2 (Risk Books).
- , and Lisa Goldberg, 2008, The market price of credit risk: The impact of asymmetric information, Working paper, Stanford University.
- Gilks, Walter, and Carlo Berzuini, 2001, Following a moving target – Monte Carlo inference for dynamic Bayesian systems, *Journal of the Royal Statistical Society, Series B* 63, 127–146.
- Glaeser, Edward L., Joseph Gyourko, and Albert Saiz, 2008, Housing supply and housing bubbles, Working paper, University of Pennsylvania and Harvard University.
- Glasserman, Paul, 2004, *Monte Carlo Methods in Financial Engineering* (Springer: Berlin, New York, Heidelberg).

- Goncharov, Yevgeny, 2006, An intensity-based approach to the valuation of mortgage contracts and computation of the endogenous mortgage rate, *International Journal of Theoretical and Applied Finance* 9, 889–914.
- Goodhart, Charles, and Boris Hofmann, 2007, *House Prices and the Macroeconomy: Implications for Banking and Price Stability* (Oxford University Press: Oxford, New York).
- Gordon, Neil, D. Salmond, and Adrian Smith, 1993, Novel approach to nonlinear / non-Gaussian Bayesian state estimation, *IEE Proceedings* 18, 1434–1443.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2009, Moral and social constraints to strategic default on mortgages, NBER, Working Paper 15145.
- Heckman, James J., and Burton Singer, 1984, A method for minimising the impact of distributional assumptions in econometric models for duration data, *Econometrica* 52, 271–320.
- Heston, Steven L., 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* 6, 327–343.
- Hjort, N. L., 1990, Goodness-of-fit tests in models for life history data based on cumulative hazard rates, *AS* 18, 1221–1258.
- Holly, Sean, M. Hashem Pesaran, and Takashi Yamagata, 2006, Spatio-temporal model of house prices in the US, Discussion paper, University of Cambridge.
- Houweling, Patrick, and Ton Vorst, 2005, Pricing default swaps: Empirical evidence, *Journal of International Money and Finance* 24, 1200–1225.
- Huang, Wenyi, and Jan Ondrich, 2002, Stay, pay, or walk away: A hazard rate analysis of FHA-insured multifamily mortgage terminations, *Journal of Housing Research* 13, 85–117.

- Hull, John C., and Alan White, 1987, The pricing of options on assets with stochastic volatilities, *Journal of Finance* 42, 281–300.
- Iacovello, Matteo, 2000, House prices and the macroeconomy in Europe: results from a structural VAR analysis, Discussion Paper, 18 European Central Bank.
- Iacus, Stefano M., 2008, *Simulation and Inference for Stochastic Differential Equations* (Springer: New York).
- Jacobsen, Martin, 1982, *Statistical Analysis of Counting Processes* (Springer: New York, Heidelberg, Berlin).
- James, Jessica, and Nick Webber, 2000, *Interest Rate Modelling* (Wiley: Chichester, West Sussex).
- Janosi, Tibor, Robert A. Jarrow, and Yildiray Yildirim, 2002, Estimating expected losses and liquidity discounts implicit in debt prices, *Journal of Risk* 5, 1–38.
- Jarrow, Robert, Haitao Li, Sheen Liu, and Chunchi Wu, 2006, Reduced-form valuation of callable corporate bonds: Theory and evidence, Discussion paper, Cornell University Working paper.
- Jarrow, Robert A., David Lando, and Fan Yu, 2005, Default risk and diversification: Theory and empirical applications, *Mathematical Finance* 15, 1–26.
- Jarrow, Robert A., and Stuart M. Turnbull, 1995, Pricing derivatives on financial securities subject to credit risk, *Journal of Finance* 50, 53–86.
- Jegadeesh, Narasimhan, and Xiongwei Ju, 2000, A non-parametric prepayment model and valuation of mortgage-backed securities, *Journal of Fixed Income* 10, 50–67.
- Jenkins, Stephen P., 2005, Survival analysis, Lecture Notes, Institute for Social and Economic Research, University of Essex.

- Jensen, Anders T., 2005, Statistical inference for doubly stochastic poisson processes, Ph.D. thesis Department of Applied Mathematics and Statistics, University of Copenhagen.
- Johannes, Michael, and Nicholas Polson, 2005, MCMC methods for financial econometrics, in Yacine Aït-Sahalia, and Lars Hansen, ed.: *Handbook of Financial Econometrics* (Elsevier).
- Jones, E. Philip, Scott Mason, and Eric Rosenfeld, 1984, Contingent claims analysis of corporate capital structures: An empirical investigation, *Journal of Finance* 39, 611–625.
- Kalbfleisch, John D., and Ross L. Prentice, 2002, *The Statistical Analysis of Failure Time Data* (Wiley).
- Kau, James B., Donald C. Keenan, and Taewon Kim, 1994, Default probabilities for mortgages, *Journal of Urban Economics* 35, 278–296.
- Kau, James B., Donald C. Keenan, Walter J. Muller, and James F. Epperson, 1992, A generalized valuation model for fixed-rate residential mortgages, *Journal of Money, Credit, and Banking* 24, 279–299.
- , 1995, The valuation at origination of fixed-rate residential mortgages with default and prepayment, *Journal of Real Estate Finance and Economics* 11, 5–36.
- Kau, James B., Donald C. Keenan, and Alexey Smurov, 2004, Reduced-form mortgage valuation, Working paper, University of Georgia.
- , 2006, Reduced form mortgage pricing as an alternative to option-pricing models, *Journal of Real Estate Finance and Economics* 33, 183–196.
- Kessler, Bo Martin, and Michael Sørensen, 1999, Estimating equations based on eigenfunctions for a discretely observed diffusion process, *Bernoulli* 5, 299–314.

- Keys, Benjamin, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2010, Securitization and screening: Evidence from subprime mortgage backed securities, *Quarterly Journal of Economics* 125.
- Kijima, Masaaki, 2003, *Stochastic Processes with Applications to Finance* (Chapman and Hall / CRC: Boca Raton , FL).
- Kolbe, Andreas, and Rudi Zagst, 2008, A hybrid-form model for the prepayment-risk-neutral valuation of mortgage-backed securities, *International Journal of Theoretical and Applied Finance* 11, 635–656.
- Koopman, Siem Jan, and Andre Lucas, 2008, A non-Gaussian panel time series model for estimating and decomposing default risk, *Journal of Business and Economic Statistics* 26, 510–525.
- , and Andre Monteiro, 2008, The multi-state latent factor intensity model for credit rating transitions, *Journal of Econometrics* 142, 399–424.
- Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics* 54, 159–178.
- LaCour-Little, Michael, and Stephen Malpezzi, 2003, Appraisal quality and residential mortgage default: Evidence from Alaska, *Journal of Real Estate Finance and Economics* 27, 211–233.
- Lancaster, Tony, 1990, *The Econometric Analysis of Transition Data* (Cambridge University Press: Cambridge) Econometric Society Monographs No. 17.
- Lando, David, 1998, On Cox processes and credit risky securities, *Review of Derivatives Research* 2, 99–120.

- , 2004, *Credit Risk Modeling: Theory and Applications* (Princeton University Press: Princeton, NJ).
- Leland, Hayne E., and Klaus Toft, 1996, Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads, *Journal of Finance* 51, 987–1019.
- Leung, Wai K., and C. F. Sirmans, 1990, A lattice approach to fixed-rate mortgage pricing with default and prepayment options, *AREUEA Journal* 14, 91–104.
- Liao, Szu-Lang, Ming-Shann Tsai, and Shu-Ling Chiang, 2008, Closed-form mortgage valuation using reduced-form model, *Real Estate Economics* 36, 313–347.
- Liesenfeld, Roman, and Jean-Francois Richard, 2003, Univariate and multivariate stochastic volatility models: Estimation and diagnostics, *Journal of Empirical Finance* 10, 505–531.
- Lo, Andrew W., 1988, Maximum likelihood estimation of generalized Itô processes with discretely sampled data, *Econometric Theory* 4, 231–247.
- Longstaff, Francis, 2003, Optimal recursive refinancing and the valuation of mortgage-backed securities, in *Derivatives 2003: Reports from the Frontiers* vol. 2.
- , and Eduardo Schwartz, 1995, Valuing risky debt: A new approach, *Journal of Finance* 50, 789–820.
- Madan, Dilip, and Haluk Unal, 1998, Pricing the risks of default, *Review of Derivatives Research* 2, 121–160.
- Mayer, Christopher J., Karen M. Pence, and Shane M. Sherlund, 2010, The rise in mortgage default: Facts and myths, *Journal of Economic Perspectives* forthcoming.
- McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts, 2005, *Quantitative Risk Management: Concepts, Techniques and Tools* (Princeton University Press: Princeton).

- McNeil, Alexander J., and Jonathan P. Wendin, 2007, Bayesian inference for generalized linear mixed models of portfolio credit risk, *Journal of Empirical Finance* 14, 131–149.
- Meen, Geoffrey P., 2001, *Modelling Spatial Housing Markets: Theory, Analysis and Policy* (Kluwer: Boston, Dordrecht, London).
- Merton, Robert, 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* 29, 449–470.
- Miller, Norman, and Liang Peng, 2006, Exploring metropolitan housing price volatility, *Journal of Real Estate Finance and Economics* 33, 5–18.
- Modigliani, Franco, and Merton Miller, 1958, The cost of capital, corporation finance, and the theory of investment, *American Economic Review* 48, 261–297.
- Møller, Jesper, and Rasmus P. Waagepetersen, 2003, *Statistical Inference and Simulation for Spatial Point Processes* (Chapman & Hall / CRC: Boca Raton, FL).
- , 2007, Modern statistics for spatial point processes, *Scandinavian Journal of Statistics* 34, 643–684.
- Musiela, Marek, and Marek Rutkowski, 2005, *Martingale Methods in Financial Modelling* (Springer: Berlin, Heidelberg, New York).
- Nelson, Daniel B., 1990, ARCH models as diffusion approximations, *Journal of Econometrics* 45, 7–38.
- Ortalo-Magne, François, and Sven Rady, 2006, Housing market dynamics: on the contribution of income shocks and credit constraints, *Review of Economic Studies* 73, 459–485.
- Pain, Nigel, and Peter Westaway, 1997, Modelling structural change in the UK housing-market: A comparison of alternative house price models, *Econometric Modelling* 14, 587–610.

- Pan, Jun, and Kenneth J. Singleton, 2008, Default and recovery implicit in the term structure of sovereign CDS spreads, *Journal of Finance* 63, 2345–2384.
- Patruno, Gregg N., 1994, Mortgage prepayment: A new model for the new era, *Journal of Fixed Income* 4, 42–50.
- Pavlov, Andrey D., 2001, Competing risks of mortgage termination: Who refinances, who moves, and who defaults?, *Journal of Real Estate Finance and Economics* 23, 185–211.
- Pearson, Neil D., and Tong-Sheng Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model, *Journal of Finance* 49, 1279–1304.
- Pedersen, Asger R., 1995, A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Scandinavian Journal of Statistics* 22, 55–71.
- Pennington-Cross, Anthony, 2003, Credit history and the performance of prime and non-prime mortgages, *Journal of Real Estate Finance and Economics* 27, 279–301.
- Piazzesi, Monika, Martin Schneider, and Selale Tuzel, 2007, Housing, consumption and asset pricing, *Journal of Financial Economics* 83, 531–569.
- Pitt, Michael, 2002, Smooth particle filters for likelihood evaluation and maximization, Discussion paper, University of Warwick Working paper.
- , and Neil Shephard, 1999, Filtering via simulation: Auxiliary particle filter, *Journal of the American Statistical Association* 94, 590–599.
- , 2001, Auxiliary variable based particle filter, in Arnaud Doucet, Nando de Freitas, and Neil Gordon, ed.: *Sequential Monte Carlo Methods in Practice* (Springer).
- Pliska, Stanley R., 2006, Mortgage valuation and optimal refinancing, in A. N. Shiryaev, M. R. Grossinho, P. E. Oliveira, and M. L. Esquível, ed.: *Stochastic Finance* (Springer).

- Price, Kenneth, Rainer Storn, and Jouni Lampinen, 2006, *Differential Evolution – A Practical Approach to Global Optimization* (Springer: Berlin, Heidelberg, New York).
- R Development Core Team, 2007, *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing Vienna, Austria ISBN 3-900051-07-0.
- Rebonato, Riccardo, 1996, *Interest-Rate Option Models* (Wiley: Chichester, West Sussex) 2 edn.
- Richard, Scott F., and Richard Roll, 1989, Prepayments on fixed-rate mortgage-backed securities, *Journal of Portfolio Management* 15, 73–82.
- Ridder, Geert, and Insan Tunali, 1999, Stratified partial likelihood estimation, *Journal of Econometrics* 92, 193–232.
- Robert, Christian P., and George Casella, 2004, *Monte Carlo Statistical Methods* (Springer: New York).
- Russell, J. R., 1999, Econometric modeling of multivariate irregularly-spaced high-frequency data, Working paper.
- Schönbucher, Philipp J., 1998, The term structure of defaultable bond prices, *Review of Derivatives Research* 2, 161–192.
- , 2003, *Credit Derivatives Pricing Models: Models, Pricing and Implementation* (Wiley: Chichester, West Sussex).
- Schwartz, Eduardo S., and Walter N. Torous, 1989, Prepayment and the valuation of mortgage-backed securities, *Journal of Finance* 44, 375–392.
- , 1992, Prepayment, default, and the valuation of mortgage pass-through securities, *Journal of Business* 65, 221–239.

- , 1993, Mortgage prepayment and default decisions: A Poisson regression approach, *AREUEA Journal* 21, 431–449.
- Sharp, Nicholas J., David P. Newton, and Peter W. Duck, 2008, An improved fixed-rate mortgage valuation methodology with interacting prepayment and default options, *Journal of Real Estate Finance and Economics* 36, 307–342.
- Singleton, Kenneth J., 2006, *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment* (Princeton University Press: Princeton, NJ).
- Smith, Adrian F., and Alan E. Gelfand, 1992, Bayesian statistics without tears: A sampling-resampling perspective, *The American Statistician* 46, 84–88.
- Smurov, Alexey, 2004, A reduced-form approach to mortgage valuation, Ph.D. thesis University of Georgia.
- Spall, James C., 2003, *Introduction To Stochastic Search And Optimization* (Wiley: Chichester, West Sussex).
- Stanton, Richard, 1995, Rational prepayment and value of mortgage-backed securities, *Review of Financial Studies* 8, 677–708.
- , and Nancy Wallace, 1998, Mortgage choice: What's the point?, *Real Estate Economics* 26, 173–205.
- Strickland, Chris M., Catherine S. Forbes, and Gael M. Martin, 2006, Bayesian analysis of the stochastic conditional duration model, *Computational Statistics & Data Analysis* 50, 2247–2267.
- Titman, Sheridan, and Walter N. Torous, 1989, Valuing commercial mortgages: An empirical investigation of the contingent-claims approach to pricing risky debt, *Journal of Finance* 44, 345–373.

- Trussell, James, and Toni Richards, 1985, Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure, *Sociological Methodology* 15, 242–276.
- van Dijk, Frank, Philip H. Frances, Richard Paap, and Dick van Dijk, 2007, Modeling regional house prices, Discussion paper, Econometric Institute.
- Winkelmann, Rainer, 2008, *Econometric Analysis of Count Data* (Springer: Berlin, Heidelberg, New York).
- Wu, Liuren, 2008, Modeling financial security returns using lévy processes, in John Birge, and Vadim Linetsky, ed.: *Financial Engineering* . vol. 15 (Elsevier, North-Holland) Handbooks in Operations Research and Management Science.
- Yang, Tyler T., Henry Buist, and Isaac F. Megbolugbe, 1998, An analysis of ex-ante probabilities of mortgage prepayment and default, *Real Estate Economics* 26, 651–676.
- Zhou, Chunsheng, 2004, The term structure of credit spreads with jump risk, *Journal of Banking and Finance* 25, 2015–2040.
- Zivot, Eric, and Jiahui Wang, 2006, *Modeling Financial Time Series with S-Plus* (Springer: New York).

APPENDIX

DRIFT AND DIFFUSION OF THE BOND PROCESS

SDE for the *zero-coupon* bond under the physical measure is:

$$dp(t, T) = p(t, T) \left[r(t) dt - D_1(t, T) \sigma_1 \sqrt{y_1(t)} dW_1^{\mathbb{P}}(t) - D_2(t, T) \sigma_2 \sqrt{y_2(t)} dW_2^{\mathbb{P}}(t) \right] \quad (1)$$

where short rate $r(t)$ is the sum of the 2 latent factors:

$$r(t) = y_1(t) + y_2(t) \quad (2)$$

Hence $\mu_b(\cdot)$ is given by:

$$\mu_b(t, T, y_1(t), y_2(t)) = p(t, T) r(t) dt \quad (3)$$

and $\sigma_b(\cdot)$ is given by:

$$\sigma_b(t, T, y_1(t), y_2(t)) = (D_1^2(t, T) \sigma_1^2 y_1(t) + D_2^2(t, T) \sigma_2^2 y_2(t))^{(-\frac{1}{2})} \quad (4)$$

where σ_1 and σ_2 are volatility parameters of the latent factors. Using independence between the latent factors of the term structure, we can write the Brownian motion for the bond process as:

$$W_b^{\mathbb{P}}(t) = (D_1^2(t, T) \sigma_1^2 y_1(t) + D_2^2(t, T) \sigma_2^2 y_2(t))^{(-\frac{1}{2})} (-D_1(t, T) \sigma_1 \sqrt{y_1(t)} dW_1^{\mathbb{P}}(t) - D_2(t, T) \sigma_2 \sqrt{y_2(t)} dW_2^{\mathbb{P}}(t)) \quad (5)$$