2008

Essays on models for financial volatility

Mihaela Oana Craioveanu
Louisiana State University and Agricultural and Mechanical College

ESSAYS ON MODELS FOR FINANCIAL VOLATILITY

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Economics

by

Mihaela Oana Craioveanu
B.S., University of The West, Timisoara, Romania, 2001
M.S., Louisiana State University, USA, 2005

August 2008

This thesis is dedicated to my parents, Mircea and Alexandra, who have always encouraged and inspired me, and to my son Franco.

Acknowledgments

I want to express my sincere gratitude and respect to my advisor Dr. Eric Hillebrand for everything he has done for me during my graduate studies, for always being involved in my academic pursuit. Without his help and support this dissertation would have been impossible to write. I am also deeply obliged to Dr. Dek Terrell for his constant help and guidance, for making me believe in myself. I am very grateful for having had the opportunity to work under his guidance for these years. I would also like to express my appreciation, respect and admiration to Dr. Carter Hill for his constant support, guidance and advice.

I would like to thank the other committee members Dr. Douglas McMillin and Dr. Ayla Kayhan for their valuable comments. Special thanks also go to Ryan Bilbo for his professional help.

I also want to thank all my friends for their support, and in particular to my dear friend Beatrice who was always there for me and always believed in me.

Finally, this dissertation would not have been possible without the constant inspiration and strength found through my son Franco. Without him I would have not made it this far.

Table of Contents

Abstract

This research is focused on models for volatility. After the introduction of realized volatility as a consistent estimator for daily volatility, time series models without latent variables have been used to model and forecast volatility. The first part of this research provides a critical review of some of the commonly used realized volatility models and addresses the problem of stationarity and lag selection. In the empirical part we apply our methodology to thirty Dow Jones Industrial Average stocks from the NYSE TAQ dataset. We address the lag selection problem for each of the stocks considered. We find that models based on flexible lag structures do not significantly outperform models based on a fixed lag structure.

With respect to latent model specifications for volatility, this study analyzes how the correlation structures in ARCH models relate to those in HARCH models. ARCH models have correlation structures that can be interpreted in the sense of mean reversion. HARCH rely on a specification that includes squared aggregated returns in the conditional variance equation. We find that HARCH is not able to capture correlation scales from ARCH in the mean reverting sense. This finding has implications for persistence. The corresponding persistence measure in HARCH does not capture the persistence of ARCH. In order to address these problems an optimal lag structure is identified. The correspondence between the lag structure and serial correlation is also addressed.

In the last part of this study a Bayesian framework is employed in order to investigate the post storm firm survival after hurricanes Katrina and Rita in the Orleans Parish, Louisiana.

A novelty of this approach is the spatial component in the model specification. Bayesian techniques are employed in order to draw inferences from a spatial probit model on a dataset containing 8,171 firms from the Orleans Parish. We find evidence indicating the presence of spatial components, especially in the quarters immediately following the storms. Other findings are: larger firms are more likely to survive; also, less flooded firms are more likely to survive; finally, sole proprietorships are more likely to reopen than large chain stores.

Chapter 1

Introduction

This dissertation contains three essays on econometric techniques with applications in time series and microeconometrics. The first two parts are directed toward volatility modeling and forecasting, while the last chapter comprises an empirical investigation of firm survival with inferences from a Bayesian framework. The high computational costs associated with obtaining results in this dissertation led to the use of the LSU High Performance Computing facility (HPC) unit.

Volatility modeling has evolved from conditional volatility specifications that treat volatility as unobservable to models based on realized volatility. The first chapter in this dissertation is focused on volatility estimation both from an empirical and a theoretical point of view. In this chapter models for day-to-day realized volatility are discussed, in particular models that capture long memory by aggregation. The models discussed are based on realized volatility, realized variance, and log realized volatility. The problems of stability and lag selection are addressed and the implications of a flexible lag structure on volatility forecasting are investigated. Models typically employed in the literature consider daily, weekly, and monthly time scales. We relax this assumption and propose a computationally intensive method that determines the optimal lag structure. The model specifications incur high computational costs and parallel algorithms are employed. In the empirical section we apply our methodology to twenty-eight Dow Jones Industrial Average stocks from the NYSE TAQ dataset, spanning a period of twelve years. We solve the lag selection problem for each of the stocks considered. We find that allowing for a flexible lag structure does not significantly outperform models based on fixed lag structures.

The second chapter in this thesis analyzes the correlation structure in the context of latent models for volatility. ARCH and GARCH models are discussed. The common persistence measures used in the context of these models can be interpreted in the sense of mean reversion. Another component of the ARCH family, the HARCH model captures the time varying volatility of returns by specifying the conditional variance as a function of aggregated squared returns. This is a very different specification when compared to ARCH models since it builds on the idea of aggregation. Therefore in this chapter the relationship between the correlation structures in HARCH and ARCH models is researched. Simulations find that the HARCH model does not capture correlation structures in the mean reversing sense. The optimal lag structure in the HARCH model is also investigated. Because of the computational cost involved in determining the optimal lags, parallel computing algorithms are employed to consider all these cases. Finally, the correspondence between the lag structure and serial correlation is studied in simulations.

The third chapter is focused on Bayesian inference methods and applications. A spatial probit model is employed with application to the impact of hurricanes Katrina and Rita on Louisiana businesses. The chapter deals with estimating the impact of the storms on firm survival. On a data set containing detailed quarterly data on firms from 2001 to 2007, a Bayesian spatial econometric model is estimated in order to determine firm survival in the wake of the storms. A key feature in this model is that it allows nearby firms' decisions to re-open or not re-open to affect all other firms.

Chapter 2

Models for Daily Realized Stock Volatility Time Series

2.1 Introduction

Volatility is one of the central inputs in a wide range of financial applications, from risk measurement and management to asset and option pricing. One of the most important stylized facts of financial volatility is volatility clustering: large movements tend to be followed by other large movements. This statement about volatility has implications for prediction: current and past values of volatility can be used to predict future volatility. Several types of models have been proposed to account for volatility clustering: models from the ARCH/GARCH family (Engle 1982, Bollerslev 1986), stochastic volatility models (Taylor 1986, Hull and White 1987, 1988, Harvey 1998), and long memory models (Granger and Joyeux 1980, Baillie, Bollerslev, and Mikkelsen 1996). These models rely on daily squared or absolute returns as proxies for true volatility. Volatility is unobservable and modeled as latent.

As high frequency intra-day data became available, an alternative method to estimate volatility was proposed by Andersen and Bollerslev (1998). In their approach, realized volatility is the square root of the sum of intra-day squared returns (i.e. realized variance), which is used as a proxy for daily integrated variance. This measure is less noisy when compared to squared day-to-day returns since it makes use of more data points and has lower variance. The realized volatility estimator is motivated by the common practice in the financial literature to model the log price process of an asset as a continuous semi-martingale. For continuous semi-martingales the sum of squared increments converges to the quadratic variation as the sampling frequency increases. In the case of a continuous time stochastic volatility model, quadratic variation is the same as integrated variance. In practice, market microstructure effects like price discreteness and the bid-ask bounce prevent the use of quadratic variation theory. Realized variance is a consistent estimator of integrated variance in the absence of microstructure noise (Andersen et al. 2003). However, in the presence of market microstructure the estimator becomes inconsistent (Bandi and Russell 2006, Hansen and Lunde 2006, Oomen 2005 and Zhang, Mykland, and Aït-Sahalia 2005). Barndorff-Nielsen and Shephard (2002) have studied the properties of the estimation error for the case of microstructure noise. The realized variance estimator becomes biased, and the bias is increasing with the sampling frequency. Therefore we are confronted with a trade-off: we need a high sampling frequency in order to reduce the measurement error, but because of the market microstructure effect a high sampling frequency means a higher bias.

In order to overcome the problem of microstructure noise, Andersen et al. (2000a, 2001a) propose sparse sampling. The method selects a sampling frequency that delivers an unbiased estimator. The sampling frequency used in the literature varies from 5 minutes to 30 minutes. Bandi and Russell (2005, 2006) and Zhang, Mykland, and Aït-Sahalia (2005) propose an optimal sampling frequency. Zhang, Mykland, and Aït-Sahalia (2005) and Zhang (2006) propose subsampling, while Zhou (1996), Barndorff-Nielsen et al. (2007) propose kernel based estimators of realized volatility and Hansen, Large, and Lunde (2007) propose pre-filtering. Finally, Barucci and Reno (2002) and Malliavin and Mancino (2002) propose Fourier methods. There are now a number of estimators of realized volatility in the presence of microstructure noise: the two-time scales realized volatility estimator proposed by Zhang, Mykland, and Aït-Sahalia (2005), the realized kernel estimator of Barndorff-Nielsen et al. (2007), the modified MA filter of Hansen, Large, and Lunde (2007), and the realized quantile-based estimator of Christensen, Oomen, and Podolskij (2008) that is robust to jumps.

The realized volatility method is very appealing in practice partly because it allows to treat volatility as observable, albeit with measurement error and microstructure bias. Thus, volatility can be modeled and forecast using standard time series models. The distribution of realized volatility is addressed by several studies (Andersen et al. 2001a, 2001b). One of the common findings in these studies is that realized volatility shows strong evidence of high persistence. To capture this long memory property, Andersen et al. (2000b, 2003) use an ARFIMA specification. Martens, van Dijk, and de Pooter (2004) propose a more general fractionally integrated model that allows for nonlinearity, structural breaks, and day of the week effects. An alternative to ARFIMA are models that approximate long memory. Several studies discuss the explanation of long memory by aggregation. This approach dates back to Granger (1980) who proved that aggregating an infinite number of short memory processes induces long memory behavior. LeBaron (2001) shows that the sum of only three autoregressive processes can lead to apparent long memory. Corsi (2004) uses this result to create the HAR realized volatility model (Heterogeneous Autoregressive Model of Realized Volatility), which builds on the HARCH specification proposed by Müller et. al (1997). Volatility is modeled as a sum of different short memory processes at different time horizons: daily, weekly, and monthly. Several studies have followed similar specifications. Andersen, Bollerslev, and Diebold (2007) and Andersen, Bollerslev, and Huang (2007) extend the model to allow for jumps. Corsi et al. (2005) develop a model that accounts for time-varying volatility of realized volatility. Bollerslev et al. (2007) propose a joint model for continuous and jump component using realized volatility and bipower variation (Barndorff-Nielsen and Shephard 2006). McAleer and Medeiros (2007) introduce nonlinearities in realized volatility and develop a multiple regime smooth transition extension. Hillebrand and Medeiros (2007) propose a non-linear estimation framework that can incorporate logistic transition effects and provide asymptotic theory and a linearity test. While some studies model realized volatility (Corsi 2004, Corsi et al. 2005, Andersen, Bollerslev, and Diebold 2007), others specify models based on log realized volatility or realized variance (Andersen, Bollerslev, and Diebold 2007, Andersen, Bollerslev, and Huang 2007, and Bollerslev et al. 2007).

In this chapter we focus on some of the most commonly used models for day-to-day realized volatility. In particular, we consider models that capture long memory by aggregation, motivated by the increasing number of studies that employ similar specifications. We discuss models based on realized volatility, realized variance, and log realized volatility. We start by investigating stability conditions. For linear models based on realized volatility and logarithmic realized volatility, stability conditions of autoregressive processes can be applied. As long as the sum of the autoregressive coefficients is smaller than one, the process is stable. However, for logarithmic and quadratic models it is not clear if this condition holds because of the non-linear specification. We provide simulations for stability for both the quadratic and the logarithmic case. We find that for the quadratic case, the sum of the coefficients rule can be applied, while for the logarithmic case, this does not hold. Furthermore, for the non-linear logarithmic specification, we find that for the same sum of autoregressive coefficients, stability varies depending on the relative coefficient magnitudes. We then investigate the implications of a flexible lag structure on volatility forecasting. Models typically employed in the literature consider daily, weekly, and monthly time scales. We relax this assumption and propose a computationally intensive method that determines the optimal lag structure according to in-sample and out-of-sample fit. The goal is to identify and analyze the different types of time scales found in the data. We apply our methodology to thirty Dow Jones Industrial Average stocks and determine the optimal lag combination for each of the stocks and each of the models considered. The tick-by-tick transaction data are obtained from the NYSE TAQ dataset and cover the period between January 3, 1995 to December 31,

2005. We consider a realized volatility and two log realized volatility model specifications. We look at two different numbers of merit when determining the optimal lag specification: maximum likelihood (in-sample fit) and minimum mean-squared error of the one-day ahead volatility forecast (out-of-sample fit). For all models considered, we find a long time scale and a short time scale. When comparing the forecasting performance between models with daily, weekly, and monthly realized volatility and models based on the optimal lag structure, we find that models based on the fixed lag structure are not significantly outperformed.

The plan for the remainder of the chapter is as follows. Section 2.2 describes the various models considered. The methodology used in identifying the optimal lag structure is presented in section Section 2.3. Section 2.4 describes the dataset used in the empirical application. Section 4.6 presents the identification of the optimal lag structure. This section also includes a forecasting comparison of the optimal lag structure models against the commonly employed model with daily, weekly, and monthly realized volatility. Section 2.6 concludes.

2.2 Models for Day-to-Day Realized Volatility

Let y_t be a consistent and unbiased estimator of the square root of daily integrated variance. We consider different models for daily realized volatility, depending on the measure of volatility chosen: realized volatility, log realized volatility, and realized variance.

Model A (General HAR). Let

$$y_{t,k} = \frac{1}{k}\sum_{i=1}^{k} y_{t-i+1}. \tag{2.1}$$

and consider

$$y_{t+1} = c + \sum_{k_j \in K} \beta_j y_{t,k_j} + w_{t+1}. \tag{2.2}$$

where $K = (k_1, k_2, \ldots, k_N)$ is a set of N indices with $k_1 < k_2 < \cdots < k_N$, $j = 1, \ldots, N$, and $w_{t+1} \sim N(0, \sigma_w^2)$.

By substituting y_{t,k_j} into equation (2.2) we can write

$$y_{t+1} = c + \frac{\beta_1}{k_1}\sum_{i=1}^{k_1} y_{t-i+1} + \frac{\beta_2}{k_2}\sum_{i=1}^{k_2} y_{t-i+1} + \cdots + \frac{\beta_N}{k_N}\sum_{i=1}^{k_N} y_{t-i+1} + w_{t+1}. \tag{2.3}$$

Let $\theta_j = \sum_{i=j}^{N} \beta_i/k_i$ for $j = 1, 2, \ldots, N$. Then equation (2.3) becomes:

$$\begin{aligned} y_{t+1} = c + \theta_1 y_t + \theta_1 y_{t-1} + \cdots + \theta_1 y_{t-k_1+1} + \theta_2 y_{t-k_1} + \cdots + \theta_2 y_{t-k_2+1} + \\ \cdots + \theta_N y_{t-k_{N-1}} + \quad \cdots + \theta_N y_{t-k_N+1} + w_{t+1}. \end{aligned} \tag{2.4}$$

Equation (2.2) can be viewed as a restricted autoregressive model. By analogy with AR models we can state that the model is covariance-stationary if and only if the roots of $1 - \theta_1 z - \cdots - \theta_1 z^{k_1} - \theta_2 z^{k_1+1} - \cdots - \theta_2 z^{k_2} - \cdots - \theta_N z^{k_{N-1}+1} - \cdots - \theta_N z^{k_N} = 0$ lie outside the unit circle. An alternative way to express the stationarity condition is to focus on the sum of the autoregressive coefficients. Let $\phi = \sum_{j=1}^{N} k_j\theta_j = \sum_{j=1}^{N} \beta_j$ be the sum of the autoregressive coefficients. The stationarity condition is $\phi < 1$. The parameter ϕ can be interpreted as a measure of persistence. It is the fraction of the shock that is carried forward in time. The closer ϕ is to one, the more persistent the volatility process will be. High persistence means slow reversion to the mean, while low persistence means fast reversion to the mean. Provided that the time series is stationary, the mean of the process is $c/(1-\phi)$.

Model A is a generalization of the HAR model proposed by Corsi (2004). The HAR model is inspired by the HARCH specification introduced by Müller et al. (1997) in the ARCH framework. HARCH advocates heterogeneity among market participants with respect to their time horizons and models volatility as a function of squared returns aggregated over different time horizons. HAR extends the model in the context of realized volatility. Corsi

(2004) specifies daily realized volatility as a linear function of past daily, weekly, and monthly realized volatility. In our notation this corresponds to $K = (1, 5, 21)$. Andersen, Bollerslev, and Diebold (2007) propose a similar setup but add jump components, while Corsi et al. (2005) employ the same specification in studying the volatility of realized volatility.

Model B (Log linear model). The use of logs of realized volatility is very common in the literature (Andersen et al. 2001a, 2001b). Log realized volatility is approximately normally distributed (Andersen et al. 2003). This finding is confirmed for all stocks considered in our dataset. For this reason we consider a logarithmic model for realized volatility. Consider Model A in equation (2.2). Define $y_{t,k}$ as in (2.1), but let y_t be a consistent and unbiased estimator of the log of the square root of daily integrated variance. This specification is the log version of Model A. Equation (2.2) can be reduced to an autoregressive form and the standard results for autoregressive models can be applied. The sum of the $\beta-$coefficients must be less than one for stability. A similar specification was proposed by Bollerslev et al. (2007).

Model C (Log model). Consider a different definition for log realized volatility. Following Andersen et al. (2007) and Andersen, Bollerslev, and Huang (2007), let

$$\log(y_{t,k}) = \log\left(\frac{1}{k}\sum_{i=1}^{k} y_{t-i+1}\right). \tag{2.5}$$

The proposed logarithmic model is given by

$$\log(y_{t+1}) = c + \sum_{k_j \in K} \beta_j \log(y_{t,k_j}) + w_{t+1}, \tag{2.6}$$

with j, k_j, K as defined in equation (2.2). Because of the log transformation, Model C can no longer be reduced to the autoregressive structure. The stability condition and persistence measure derived from autoregressive models can no longer be applied and a stationarity condition similar to the one for Model A is hard to find.

By Jensen's inequality we have for a $k_l \in K$:

$$\mathbb{E}[\log(y_{t,k_l})] \leq \log \mathbb{E}\left[\frac{1}{k_l}\sum_{i=1}^{k_l} y_{t-i+1}\right] = \log \mathbb{E}\left[\frac{1}{k_l}\sum_{i=1}^{k_l} e^{c+\sum_{k_j \in K}\beta_j \log(y_{t-i,k_j})+w_{t-i+1}}\right]. \qquad (2.7)$$

From the inequality in (2.7) we can conclude that $\log(y_{t,k_l})$ will not be stable unless y_t is stable. However, as can be seen from the equality part in (2.7), y_t is an exponential function of past log volatilities at different time horizons and therefore stability is difficult to analyze.

We investigate the stability of the process in simulations. We simulate 5000 observations based on equation (2.6) with daily, weekly, and monthly lags and corresponding coefficients β_1, β_2, and β_3. We initially set $\beta_1 = 0.01, \beta_2 = 0.09$, and $\beta_3 = 0.70$ such that the coefficients sum up to 0.80. A plot of the sample autocorrelation for logarithmic realized volatility is presented in Figure 2.1, panel (1). The series appears non-stationary indicating that the sum of the coefficients rule may no longer be used as stationarity condition. We analyze stability by estimating the fractional integration parameter of the process. Equation (2.6) specifies the log realized volatility as an aggregate over log realized volatilities at different time horizons. Granger (1980) shows that aggregation of processes induces long memory properties. Therefore $\log(y_t)$ can be approximated by a fractionally integrated process of order d. For $d \in (0, 0.5)$ the process is stationary. If $d \in (0.5, 1)$, the process is non-stationary but mean reverting. If $d \in (-0.5, 0)$ the process is anti-persistent. For $d > 1$ the process is non-stationary and not mean reverting. We treat values of d above 0.5 by differencing the process and estimating the fractional parameter on the differenced series such that we obtain a consistent estimate of $\tilde{d} \in (-0.5, 0)$. Then we estimate d as $\tilde{d}+1$. We generate 100 samples of 5000 observations each based on Model C with daily, weekly, and monthly lags and $\beta-$coefficients set as above. For each generated sample, we estimate an ARFIMA(0,d,0) specification. We employ the Whittle estimator for d. On all 100 samples, the estimates of

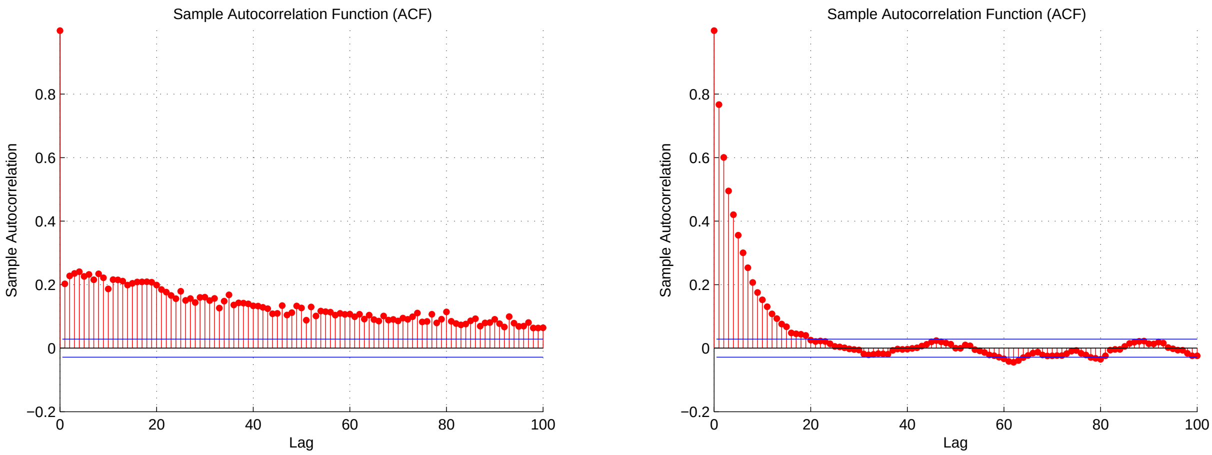
d are larger than 0.5 and below 1, indicating non-stationarity. We repeat the experiment for a sum of coefficients equal to 0.75. For this value, in 78 of the 100 samples the estimates of d are above 0.5 and below 1. The series turns stationary for a sum of coefficients roughly between 0.72 and 0.71. For a value of 0.72, in 65 of the samples the fractional integration parameter was above 0.5, while for a value of 0.71 in 48 of the cases the value was greater than 0.5.

We also examine the effect of different coefficient magnitudes on stability. We find that for the same sum of coefficients, the log series can be either stationary or non-stationary, depending on the magnitude of the coefficient attached to the highest time scale. Suppose we follow the same specification with daily, weekly, and monthly realized volatility, but set $\beta_1 = 0.70, \beta_2 = 0.09$, and $\beta_3 = 0.01$. In this setup the coefficient sum remains the same (0.80), but the realized volatility weights change. The sample autocorrelation for this series is presented in Figure 2.1, panel (2). When compared to panel (1) we see very different dynamics, the decay is slower for the first series. For these values of the coefficients the series looks stationary. We repeat the simulation procedure and find that the estimated fractional coefficient is below 0.5 in all 100 samples. This experiment illustrates that the sum of the coefficients rule cannot be applied when investigating stability for models of type C.

Model D (Quadratic model). Consider the following process

$$(y_{t+1})^2 = c + \sum_{k_j \in K} \beta_j (y_{t,k_j})^2 + w_{t+1}, \tag{2.8}$$

with j, k_j, K as defined in equation (2.2). Similar specifications were employed by Andersen et al. (2007). Equation (2.8) specifies realized variance as a function of past realized variances and can be interpreted as a quadratic specification in terms of sums of realized volatility.



(1) (2)

Figure 2.1: Sample autocorrelation for daily log realized volatility generated from Model C (equation 2.6) with sum of coefficients equal to 0.80. Panel (1): $\beta_1 = 0.01, \beta_2 = 0.09, \beta_3 = 0.70$; Panel (2): $\beta_1 = 0.70, \beta_2 = 0.09, \beta_3 = 0.01$

Stability analysis is complicated by the squared terms. The following inequality holds:

$$(y_{t,k})^2 = \left(\frac{\sum_{i=1}^{k} y_{t-i+1}}{k}\right)^2 \leq \frac{\sum_{i=1}^{k} (y_{t-i+1})^2}{k}. \tag{2.9}$$

From (2.8) and (2.9) we have

$$(y_{t+1})^2 \leq c + \sum_{k_j \in K} \frac{\beta_j}{k_j} \sum_{i=1}^{k_j} (y_{t-i+1})^2 + w_{t+1}. \tag{2.10}$$

The right hand side of (2.10) is an autoregressive model in the realized variance that can be analyzed with the same approach as Model A in equation (2.2). As long as the right hand side is stationary, Model D will also be stationary. Therefore, as long as the sum of the coefficients is less than one, the process will be stable. Figure 2.2 plots 5000 observations generated from equation (2.8) with sum of coefficients equal to 0.99. The series appears stationary. We investigate the stationarity of the series by estimating the fractional integration parameter d. We generate 100 samples of this specification and estimate an ARFIMA(0, d, 0) specification for each sample. On all 100 runs, the parameter d was greater than zero but below 0.5 confirming the stationarity of the process.

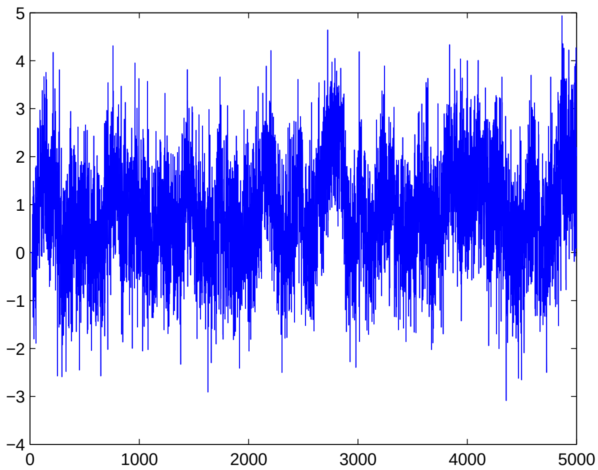


Figure 2.2: Daily realized volatility generated from Model D (equation 2.8) with sum of coefficients equal to 0.99

2.3 Implementation

The goal of this section is to address the problem of the optimal lag structure in the context of Models A, B, and C specified in the previous section. We focus on these models since they have been successfully employed in a number of recent studies. The models employed in practice typically rely on daily, weekly, and monthly realized volatility. The reason for this choice is partly motivated by the simple interpretation of these lags, but also by the high computational cost associated with a flexible lag structure.

We treat the maximum number of lags L allowed as fixed and determine the N relevant lags to be included by examining all combinations of possible lags. The number of models to estimate in this case is $\binom{L}{N}$ for each stock considered. If we always include lag one, the number of models to estimate is reduced to $\binom{L}{N-1}$. This, however, demands significant computing power. Because of the computational intensity we choose a parallel computing framework. Since the estimation processes are independent of each other, the problem is "trivially" parallelizable.

Each model specification is estimated by maximum likelihood. Thus, we estimate $\binom{L}{N-1}$ different models. We search for the best model specification according to two different criteria: in-sample and out-of-sample fit. For the in-sample fit, the decision is based on the maximum of the log-likelihood function. For the out-of-sample fit, we focus on the minimum mean-squared error of the one-day ahead volatility forecast (MSE(1)). For each criterion considered, the estimation algorithm delivers the optimal lag structure and parameter estimates. We apply this procedure to thirty DJIA stocks in the empirical part of the chapter.

The implementation was done in C++ and set up on Louisiana State University's supercomputing framework. On 1 node and 4 processors on each node, which provide a computing power of 42.56 Gflops/second, the determination of the optimal lag structure with $L = 250$ and $N = 3$ took fifteen minutes for each stock and model and criterion of merit.

We start by calibrating the procedure. Therefore, we consider Model A as the data-generating process with daily, weekly, and monthly realized volatility. We simulate 6000 days of realized volatility and use the first 5000 observations for the estimation part and the last 1000 for out-of-sample forecasting. We repeat this process 100 times. On each run we estimate Model A with $L = 250$ and $N = 3$. We always include lag 1, so that we have to estimate $\binom{250}{2} = 31,125$ specifications. For both the in-sample and out-of-sample fit, the average estimated optimal lags over the 100 runs were 5 for the second lag and 21 for the third lag, so we retrieve the data generating lag structure (the first lag is always lag one).

2.4 Data

We use high-frequency tick-by-tick trades on thirty Dow Jones Industrial Average Index stocks as listed in Table 2.1: Alcoa Inc. (AA), American International Group Inc. (AIG), American Express Inc. (AXP), Boeing Co. (BA), Citigroup Inc. (C), Caterpillar Inc. (CAT), Du Pont De Nemours (DD), Walt Disney Co. (DIS), General Electric (GE), General Motors (GM), Home Depot Inc. (HD), Honeywell International (HON), Hewlett Packard (HPQ), International Business Machines Co. (IBM), Intel Co. (INTC), Johnson and Johnson (JNJ), JP Morgan Chase (JPM), Coca Cola (KO), McDonald's (MCD), 3M Company (MMM), Altria Group (MO), Merck Co. (MRK), Microsoft Co. (MSFT), Pfizer Inc. (PFE), Procter and Gamble (PG), AT&T (T), United Tech (UTX), Verizon Communications (VZ), Wal-Mart Stores (WMT), and Exxon Mobil (XOM). The data are obtained from the NYSE TAQ (Trade and Quote) database. The sample period starts in January 3, 1995 and ends in July 31, 2007. In Table 2.1 we report the number of days in the sample and the average number of transactions per day for each of the stocks considered.

In calculating daily realized volatility we employ the realized kernel estimator with modified Tukey-Hanning weights of Barndorff-Nielsen et al. (2007). We start by cleaning the data for outliers. We consider transactions between 9.30 am through 4.00 pm. Following Barndorff-Nielsen et al. (2007) we employ the following 60 second activity fixed tick time sampling scheme: $fq_i = 1 + 60n_i/(t_{0i} - t_{n_i i})$, where fq_i is the sampling frequency, n_i represents the number of transactions for day i, and t_{0i}, $t_{n_i i}$ are the times for the first and last trade for day i. This is tick-time sampling chosen such that the same number of observations is obtained each day.

Figure 2.3 shows plots for daily realized volatility and logarithmic realized volatility series for Walmart Inc. from January 3, 1995 to December 31, 2005. The daily realized volatility is calculated from intraday log returns measured in percentage. From the first two panels we

Table 2.1: **Data description.** The first two columns display the symbols and names of the stocks considered in the empirical investigation. The third column gives the average number of transactions per day. Column 4 shows the number of days in the sample.

Symbol	Stock	Trans per day	No days
aa	Alcoa Inc.	2055	3162
aig	American International Group Inc.	2979	3157
axp	American Express Co.	2599	3164
ba	Boeing Co.	3006	3159
c	Citigroup Inc.	5327	3143
cat	Caterpillar Inc.	3597	2051
dd	Du Pont de Nemours&Co.	2587	3163
dis	Walt Disney Co.	3839	3156
ge	General Electric Co.	8072	3164
gm	General Motors Corp.	2945	3160
hd	Home Depot Inc.	4758	3163
hon	Honeywell International Inc.	1888	3160
hpq	Hewlett-Packard Co.	6480	1314
ibm	International Business Machines Corp.	5117	3160
intc	Intel Co.	43916	3164
jnj	Johnson&Johnson	3551	3156
jpm	JPMorgan Chase&Co.	3400	3155
ko	Coca-Cola Co.	3302	3165
mcd	McDonald's Corp.	2720	3153
mmm	3M Co.	2183	3162
mo	Altria Group Inc.	4031	3153
mrk	Merck&Co. Inc.	4353	3162
msft	Microsoft Co.	40537	3161
pfe	Pfizer Inc.	7029	3159
pg	Procter&Gamble Co.	3062	3163
t	AT&T Inc.	3975	3156
utx	United Technologies Corp.	1834	3162
vz	Verizon Communications Inc.	5388	1775
wmt	Wal-Mart Stores Inc.	4797	3159
xom	Exxon Mobil Corp.	7488	1923

see that each of the two series is characterized by a high degree of volatility clustering with periods of high volatility and low volatility. A common finding in the literature (Andersen et al. 2001a) is that logarithmic realized volatility is close to being normal, while the realized volatility series is not normal. Panels three and four present the QQ- plots for the two series. Although the log realized volatility plot is linear indicating normality, given the long period of time covered in our sample, we observe different periods of volatility, with different means and normality is no longer a good assumption. The last two panels graph the sample autocorrelation function for realized volatility and log realized volatility. Both panels exhibit slow hyperbolic decay indicating the presence of long memory. These findings are consistent across all stocks considered in our study.

2.5 Empirical Application

In this section we identify the optimal lag structure on the dataset described in Section 2.4. We divide the sample period of January 3, 1995 through July 31, 2007 into three parts: an estimation sample, a training sample, and a forecast sample. The estimation sample covers January 3, 1995 through December 31, 2004, the training sample covers January 3, 2005 through December 31, 2005, and the forecast sample covers January 3, 2006 through July 31, 2007. We consider model specifications A, B, and C. We set the number of lags N equal to 3 and the maximum lag L to 250. Searching for all possible combinations of lags means estimating $\binom{250}{3}$ models. We always include lag 1, corresponding to daily realized volatility. The algorithm can choose two free lags, which reduces the estimation cost to $\binom{250}{2} = 31,125$ models. We run the estimation procedure on each dataset and search for the best lag specification.

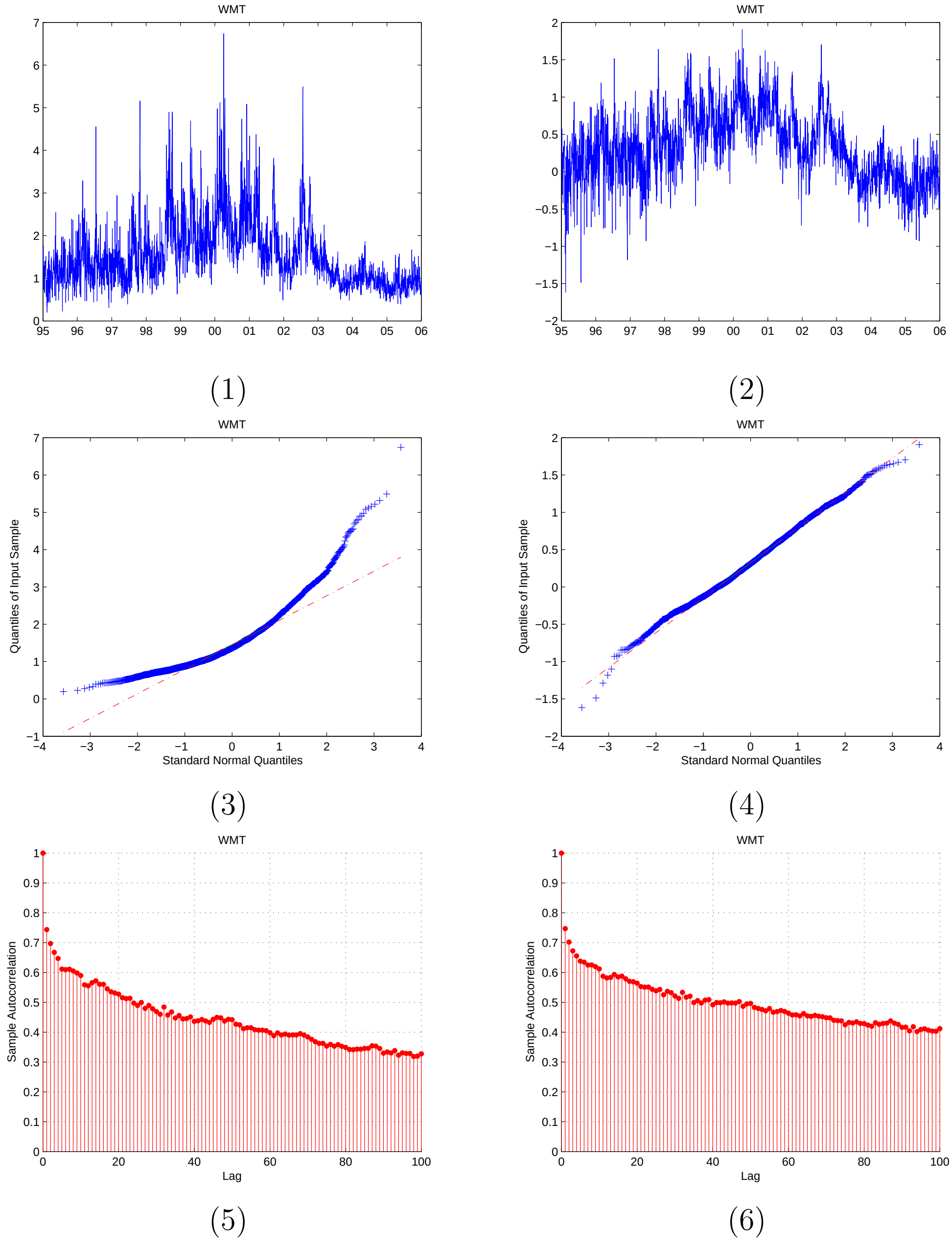


Figure 2.3: WMT series from January 3, 1995 through December 31, 2005. Panel (1): realized volatility; Panel (2): Logarithmic realized volatility; Panel (3): QQ-plot realized volatility; Panel(4): QQ-plot logarithmic realized volatility; Panel (5): sample autocorrelation for realized volatility; Panel (6): sample autocorrelation for logarithmic realized volatility

Model A. We consider the realized volatility specification from equation (2.2)

$$y_{t+1} = c + \sum_{k_j \in K} \beta_j y_{t,k_j} + w_{t+1}.$$

Tables 2.2 and 2.3 present the results for the estimation of the best specification according to maximum likelihood and minimum mean-squared error of the one-day-ahead forecast on the forecast sample. The tables report the optimal lag structure, the corresponding parameter estimates, and standard errors according to Newey and West (1987) for each of the stocks considered. In the last column we list the estimated persistence parameter ϕ calculated as the sum of the estimated coefficients.

With a few exceptions all estimated coefficients are highly significant, for both in-sample and out-of-sample fit. The last column indicates high persistence in realized volatility, as shown by the high values of ϕ. With the exception of hpq, all coefficients are above 0.9 for both criteria. Furthermore, for the majority of stocks the estimated persistence parameter is greater than 0.95.

If we restrict our attention to the in-sample fit there seems to be a substantial influence of the 8-14 lag on daily realized volatility. The long lag is estimated at or close to the boundary of 250, which we interpret as indicating long memory in the time series of realized volatility. For the out-of-sample fit the most frequent lags are the 207-217 lag (6 times) and the 244-250 lag (7 times). For the first lag the influence of 8-14 days is found. This is consistent with the finding for the in-sample fit.

Next we compare the forecasting performance of these models against the benchmark model with lags 1, 5, and 21 usually employed in the literature. We compute Hansen's (2005) Test for Superior Predictive Ability (SPA). The loss function considered is mean-squared error. Table 2.4 presents one-day ahead forecasting error results as well as results from the SPA test. For the lag chosen based on in-sample and out-of-sample, columns 1 and 3 report the mean-squared error (MSE) of the one-day ahead forecast as a percentage of the mean-squared error of the benchmark model. Columns 2 and 4 report the p-values for the SPA test for the in-sample fit (column 2) and out-of-sample fit (column 4) lag specification. The null hypothesis for the SPA test is that the benchmark model of lags (1,5,21) is not inferior to any of the competing models.

If we focus on the in-sample fit criterion we cannot reject the null hypothesis, with a few exceptions. For two out of the thirty stocks considered, the model based on the optimal lag combination outperforms the benchmark model at the ten percent significance level. The specification based on the out-of-sample fit performs better when compared with the benchmark model. At ten percent significance level, for six stocks the model outperforms the benchmark. We can conclude that the benchmark model is not significantly outperformed by neither of the competing models. When comparing the forecast performance between the models based on the in-sample and out-of-sample fit, we observe that, although the model specification based on the out-of-sample fit performs better than the specification based on the in-sample fit, the differences in the MSE are rather small.

Model B. We consider the following realized volatility specification

$$\log(y_{t+1}) = c + \sum_{k_j \in K} \beta_j \log(y_{t,k_j}) + w_{t+1},$$

where $\log(y_{t,k}) = 1/k \sum_{i=1}^{k} \log(y_{t-i+1})$.

Tables 2.5 and 2.6 present the results for the estimation of the best specification for this model according to the two criteria considered. The tables report the optimal lag structure, the corresponding parameter estimates, and standard errors in parenthesis for each of the stocks considered. The last column lists the estimated persistence parameter calculated as the sum of the estimated coefficients.

With a few exceptions all estimated coefficients are highly significant, for both criteria.

The high values of the persistence parameter ϕ found in the last column indicate high persistence in realized volatility. With the exception of gm, all coefficients are above 0.9 for both criteria and above 0.95 for the majority of stocks.

The in-sample fit analysis confirms the influence of the 8-14 lag on the daily realized volatility. The long lag is found close to or at the boundary of 250. For the out-of-sample analysis, the longer lag is no longer found on the boundary. The most frequent lags are the 125-133 lag and the 13-28 lag. The short lag is found at 4-14 days. We can conclude that for both the in-sample and out-of-sample fit, the findings are very similar to the ones found for Model A.

The results for the comparison analysis of the forecasting performance of these models against the benchmark model with lags 1, 5, and 21 are summarized in Table 2.7. The table reports the MSE and p-values from the SPA test. The results are similar with the findings for Model A. With the exception of two stocks, for the in-sample fit criterion we cannot reject the null hypothesis. When comparing the specifications based on the out-of-sample fit against the benchmark model, we see that at ten percent significance level, for two stocks the model outperforms the benchmark. In terms of the MSE performance, the differences between the model specifications based on the in-sample and out-of-sample fit are rather small.

Model C. We repeat the procedure described in the previous sections and apply the same methodology to Model C from equation (2.6):

$$\log(y_{t+1}) = c + \sum_{k_j \in K} \beta_j \log(y_{t,k_j}) + w_{t+1}.$$

In Tables 2.8 and 2.9 we report the results for the estimation of the best specification according to the two criteria considered. The tables report the optimal lag, the corresponding parameter estimates, and the standard errors for each of the stocks considered. The estimated coefficients are significant for the majority of stocks.

The results are in line with the findings for Model A. The second lag is found on the boundary for the in-sample fit. For the first lag we confirm the influence of the 8-14 value. For the out-of-sample fit we find that the long lag is no longer found at the maximum but it converges to some specific value. The influential lags are the 120-127 and the 135-138 lag. For the first lag the influence of 4-14 days is found.

Next we choose as benchmark the model based on lags 1, 5, and 21 and compare this model with the model based on the optimal lag specification in a forecasting competition. Table 2.10 reports the MSE (as percentage of the MSE of the benchmark model) and the results from the SPA test. The lag specification according to the in-sample fit is not significantly outperforming the benchmark model. Again, this is in line with the previous findings for Models A and B. Only for one stock the model considered outperforms the benchmark model. For the specification obtained by the out-of-sample fit we find the same result, the benchmark is outperformed for one stock. If we compare the model specification based on the in-sample fit with the one based on the out-of-sample fit in terms of their MSE, we conclude that the differences are small.

2.6 Conclusion

This chapter provides a critical review of models for daily realized volatility, motivated by the extensive use of these models in practice. We address the problem of stationarity and lag selection. We focus on three specifications: a linear one based on realized volatility, a log linear realized volatility specification, and one based on log realized volatility. We propose a computationally intense procedure that allows the selection of lags based on maximum log-likelihood and minimum mean-squared error. We find under both optimality criteria considered a short lag (8-14 days for the linear models and 4-14 days for the log linear model) and a long lag (at 120-127 days and 135-138 days, or at the maximum lag allowed). When comparing the models based on the optimal lag structure against the benchmark model with daily, weekly, and monthly realized volatility we find that the benchmark model is not significantly outperformed by these models.

Table 2.2: **Model A: In-Sample Fit Estimation.** The first column displays the stock symbol. Column 2 presents the optimal lags according to the in-sample fit. Columns 3 to 6 give the parameter estimates. The last column is the persistence measure, calculated as the sum of the β coefficients. The figures in parentheses are standard errors.

	lags	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{c}$	$\hat{\phi}$
aa	(1,5,250)	0.252(.029)	0.578(.036)	0.134(.027)	0.061(.034)	0.964
aig	(1,6,246)	0.330(.056)	0.514(.053)	0.103(.033)	0.078(.031)	0.948
axp	(1,5,249)	0.368(.041)	0.468(.045)	0.137(.035)	0.045(.032)	0.973
ba	(1,8,250)	0.388(.035)	0.420(.037)	0.153(.029)	0.065(.042)	0.961
c	(1,10,248)	0.391(.043)	0.465(.042)	0.115(.031)	0.046(.033)	0.972
cat	(1,13,250)	0.285(.042)	0.562(.072)	0.100(.065)	0.098(.075)	0.947
dd	(1,6,250)	0.338(.051)	0.498(.057)	0.132(.026)	0.051(.025)	0.968
dis	(1,6,248)	0.272(.029)	0.555(.029)	0.135(.029)	0.064(.031)	0.963
ge	(1,8,248)	0.413(.031)	0.435(.040)	0.116(.028)	0.054(.031)	0.965
gm	(1,10,248)	0.311(.032)	0.546(.041)	0.084(.035)	0.086(.039)	0.941
hd	(1,7,249)	0.375(.035)	0.480(.035)	0.105(.030)	0.069(.036)	0.96
hon	(1,8,250)	0.344(.052)	0.470(.060)	0.148(.034)	0.067(.037)	0.962
hpq	(1,4,246)	0.242(.064)	0.300(.088)	0.219(.041)	0.248(.070)	0.761
ibm	(1,7,247)	0.398(.030)	0.442(.036)	0.133(.033)	0.039(.038)	0.973
intc	(1,9,247)	0.468(.025)	0.408(.033)	0.102(.023)	0.046(.030)	0.978
jnj	(1,8,249)	0.343(.032)	0.504(.039)	0.079(.035)	0.097(.038)	0.926
jpm	(1,9,250)	0.322(.029)	0.540(.034)	0.104(.026)	0.057(.027)	0.967
ko	(1,4,250)	0.314(.036)	0.500(.041)	0.142(.034)	0.062(.026)	0.956
mcd	(1,14,248)	0.338(.029)	0.502(.041)	0.115(.042)	0.068(.041)	0.955
mmm	(1,14,250)	0.406(.031)	0.473(.042)	0.086(.037)	0.047(.036)	0.965
mo	(1,9,250)	0.317(.030)	0.456(.037)	0.130(.043)	0.145(.053)	0.903
mrk	(1,8,245)	0.382(.038)	0.449(.041)	0.096(.032)	0.107(.039)	0.927
msft	(1,7,249)	0.469(.029)	0.382(.038)	0.133(.024)	0.024(.027)	0.984
pfe	(1,10,248)	0.374(.028)	0.439(.036)	0.123(.040)	0.102(.046)	0.935
pg	(1,9,248)	0.413(.035)	0.450(.040)	0.109(.027)	0.039(.027)	0.972
t	(1,9,249)	0.362(.035)	0.436(.039)	0.163(.031)	0.070(.035)	0.961
utx	(1,8,248)	0.358(.048)	0.487(.057)	0.119(.025)	0.057(.028)	0.964
vz	(1,7,250)	0.300(.070)	0.638(.057)	0.020(.038)	0.062(.038)	0.958
wmt	(1,9,249)	0.359(.036)	0.497(.041)	0.118(.033)	0.042(.032)	0.973
xom	(1,4,248)	0.265(.071)	0.605(.070)	0.110(.036)	0.016(.030)	0.98

Table 2.3: **Model A: Out-Of-Sample Fit Estimation.** The first column displays the stock symbol. Column 2 presents the optimal lags according to the out-of-sample fit. Columns 3 to 6 give the parameter estimates. The last column is the persistence measure, calculated as the sum of the β coefficients. The figures in parentheses are standard errors.

	lags	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{c}$	$\hat{\phi}$
aa	(1,11,131)	$0.344_{(.035)}$	$0.519_{(.037)}$	$0.089_{(.036)}$	$0.08_{(.036)}$	0.952
aig	(1,16,117)	$0.431_{(.056)}$	$0.446_{(.056)}$	$0.065_{(.040)}$	$0.086_{(.039)}$	0.942
axp	(1,7,209)	$0.409_{(.041)}$	$0.439_{(.042)}$	$0.126_{(.035)}$	$0.042_{(.031)}$	0.974
ba	(1,8,246)	$0.389_{(.035)}$	$0.418_{(.037)}$	$0.155_{(.029)}$	$0.062_{(.042)}$	0.962
c	(1,10,217)	$0.391_{(.043)}$	$0.46_{(.041)}$	$0.123_{(.030)}$	$0.042_{(.031)}$	0.974
cat	(1,10,247)	$0.284_{(.040)}$	$0.529_{(.066)}$	$0.136_{(.062)}$	$0.097_{(.070)}$	0.949
dd	(1,11,250)	$0.401_{(.047)}$	$0.467_{(.052)}$	$0.099_{(.029)}$	$0.053_{(.028)}$	0.967
dis	(1,5,154)	$0.256_{(.028)}$	$0.537_{(.030)}$	$0.172_{(.030)}$	$0.056_{(.029)}$	0.966
ge	(1,10,209)	$0.431_{(.031)}$	$0.425_{(.040)}$	$0.111_{(.029)}$	$0.051_{(.031)}$	0.966
gm	(1,2,27)	$0.21_{(.036)}$	$0.301_{(.067)}$	$0.397_{(.045)}$	$0.131_{(.041)}$	0.908
hd	(1,11,207)	$0.419_{(.037)}$	$0.454_{(.037)}$	$0.088_{(.032)}$	$0.067_{(.037)}$	0.961
hon	(1,5,127)	$0.309_{(.045)}$	$0.426_{(.053)}$	$0.225_{(.037)}$	$0.07_{(.035)}$	0.959
hpq	(1,13,102)	$0.317_{(.061)}$	$0.369_{(.102)}$	$0.203_{(.082)}$	$0.138_{(.060)}$	0.89
ibm	(1,6,244)	$0.39_{(.031)}$	$0.435_{(.037)}$	$0.149_{(.034)}$	$0.036_{(.037)}$	0.975
intc	(1,12,138)	$0.485_{(.026)}$	$0.384_{(.040)}$	$0.103_{(.030)}$	$0.057_{(.030)}$	0.972
jnj	(1,10,208)	$0.364_{(.031)}$	$0.492_{(.040)}$	$0.072_{(.036)}$	$0.094_{(.036)}$	0.928
jpm	(1,3,129)	$0.229_{(.033)}$	$0.494_{(.044)}$	$0.24_{(.032)}$	$0.06_{(.028)}$	0.964
ko	(1,6,247)	$0.373_{(.040)}$	$0.463_{(.041)}$	$0.124_{(.033)}$	$0.055_{(.027)}$	0.96
mcd	(1,5,28)	$0.275_{(.037)}$	$0.296_{(.050)}$	$0.354_{(.045)}$	$0.113_{(.026)}$	0.924
mmm	(1,8,248)	$0.371_{(.028)}$	$0.465_{(.037)}$	$0.129_{(.031)}$	$0.046_{(.030)}$	0.966
mo	(1,2,64)	$0.187_{(.029)}$	$0.345_{(.043)}$	$0.379_{(.047)}$	$0.129_{(.040)}$	0.911
mrk	(1,16,56)	$0.44_{(.037)}$	$0.399_{(.068)}$	$0.074_{(.063)}$	$0.123_{(.039)}$	0.913
msft	(1,8,246)	$0.485_{(.028)}$	$0.367_{(.037)}$	$0.133_{(.025)}$	$0.022_{(.027)}$	0.985
pfe	(1,20,199)	$0.42_{(.027)}$	$0.425_{(.046)}$	$0.089_{(.049)}$	$0.103_{(.051)}$	0.934
pg	(1,10,244)	$0.421_{(.035)}$	$0.447_{(.040)}$	$0.102_{(.027)}$	$0.04_{(.027)}$	0.971
t	(1,6,63)	$0.349_{(.044)}$	$0.338_{(.058)}$	$0.269_{(.038)}$	$0.073_{(.029)}$	0.956
utx	(1,3,70)	$0.234_{(.031)}$	$0.468_{(.044)}$	$0.256_{(.045)}$	$0.064_{(.026)}$	0.958
vz	(1,2,68)	$0.213_{(.066)}$	$0.487_{(.071)}$	$0.262_{(.054)}$	$0.054_{(.043)}$	0.962
wmt	(1,10,212)	$0.365_{(.036)}$	$0.493_{(.043)}$	$0.115_{(.034)}$	$0.042_{(.031)}$	0.973
xom	(1,12,13)	$0.321_{(.062)}$	$0.458_{(.348)}$	$0.165_{(.357)}$	$0.072_{(.031)}$	0.944

Table 2.4: **Model A: Forecasting results.** The first column displays the stock symbol. Column 2 reports the mean-squared error (MSE) for the one-day ahead forecast of realized volatility using the model chosen by the in-sample fit criterion. Column 3 reports the p-values of the Superior Predictive Ability (SPA) test of Hansen (2005) for the in-sample fit criterion. The null hypothesis for the SPA test is that the benchmark model is the best forecasting model. The benchmark considered is the (1,5,21) model specification. The last two columns present the MSE and p-values for the model specification based on the out-of-sample fit. The MSE are calculated as a percentage of the MSE of the benchmark model.

	$MSE(1)^{LL}$	$p-val^{LL}$	$MSE(1)^{FCH}$	$p-val^{FCH}$
aa	1.000	0.889	0.975	0.671
aig	1.011	0.879	0.981	0.630
axp	0.977	0.426	0.969	0.314
ba	0.961	0.511	0.959	0.515
c	0.878	0.602	0.872	0.584
cat	0.981	0.766	0.970	0.503
dd	1.008	0.692	0.981	0.696
dis	1.010	0.939	0.976	0.698
ge	0.948	0.254	0.933	0.133
gm	1.025	0.589	0.970	0.032
hd	1.009	0.575	0.980	0.474
hon	0.992	0.410	0.957	0.006
hpq	0.988	0.477	0.959	0.618
ibm	0.935	0.285	0.927	0.171
intc	0.988	0.713	0.967	0.736
jnj	0.963	0.042	0.941	0.002
jpm	0.997	0.670	0.959	0.590
ko	1.000	0.131	0.958	0.169
mcd	1.001	0.614	0.991	0.035
mmm	0.994	0.136	0.963	0.528
mo	3.557	0.219	0.982	0.506
mrk	1.003	0.888	0.984	0.897
msft	0.921	0.307	0.920	0.325
pfe	0.996	0.007	0.981	0.024
pg	0.978	0.378	0.977	0.349
t	1.033	0.207	0.943	0.112
utx	0.990	0.516	0.955	0.024
vz	1.041	0.826	0.966	0.470
wmt	0.963	0.296	0.944	0.286
xom	1.052	0.371	0.964	0.880

Table 2.5: **Model B: In-Sample Fit Estimation.** The first column displays the stock symbol. Column 2 presents the optimal lags according to the in-sample fit. Columns 3 to 6 give the parameter estimates. The last column is the persistence measure, calculated as the sum of the β coefficients. The figures in parentheses are standard errors.

	lags	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{c}$	$\hat{\phi}$
aa	(1,6,250)	0.279 (.027)	0.535 (.034)	0.152 (.026)	0.014 (.009)	0.966
aig	(1,5,247)	0.281 (.033)	0.555 (.038)	0.114 (.027)	0.016 (.007)	0.950
axp	(1,12,248)	0.422 (.028)	0.472 (.032)	0.097 (.025)	0.000 (.008)	0.991
ba	(1,13,250)	0.375 (.027)	0.469 (.034)	0.120 (.030)	0.015 (.010)	0.964
c	(1,10,250)	0.373 (.029)	0.489 (.033)	0.122 (.025)	0.004 (.007)	0.985
cat	(1,14,250)	0.278 (.030)	0.566 (.050)	0.119 (.047)	0.026 (.017)	0.963
dd	(1,10,229)	0.327 (.028)	0.537 (.035)	0.113 (.025)	0.008 (.007)	0.977
dis	(1,7,249)	0.286 (.028)	0.541 (.034)	0.139 (.024)	0.016 (.006)	0.966
ge	(1,10,250)	0.375 (.028)	0.500 (.034)	0.102 (.024)	0.007 (.006)	0.977
gm	(1,10,249)	0.281 (.025)	0.568 (.031)	0.100 (.033)	0.015 (.009)	0.949
hd	(1,11,249)	0.366 (.027)	0.509 (.030)	0.096 (.030)	0.013 (.009)	0.971
hon	(1,9,234)	0.300 (.031)	0.510 (.040)	0.156 (.033)	0.016 (.009)	0.966
hpq	(1,4,27)	0.282 (.042)	0.430 (.065)	0.266 (.046)	0.005 (.013)	0.977
intc	(1,9,247)	0.455 (.024)	0.407 (.030)	0.119 (.022)	0.012 (.009)	0.981
ibm	(1,7,182)	0.374 (.026)	0.465 (.034)	0.149 (.025)	0.001 (.006)	0.988
jnj	(1,8,250)	0.311 (.026)	0.527 (.034)	0.112 (.029)	0.010 (.006)	0.950
jpm	(1,10,250)	0.328 (.029)	0.549 (.031)	0.099 (.024)	0.010 (.007)	0.976
ko	(1,6,249)	0.306 (.030)	0.514 (.037)	0.146 (.030)	0.008 (.006)	0.966
mcd	(1,14,250)	0.312 (.027)	0.526 (.038)	0.128 (.037)	0.011 (.010)	0.966
mmm	(1,14,250)	0.379 (.025)	0.500 (.034)	0.094 (.035)	0.005 (.007)	0.973
mo	(1,19,250)	0.380 (.024)	0.491 (.037)	0.052 (.042)	0.024 (.012)	0.923
mrk	(1,8,242)	0.345 (.024)	0.470 (.034)	0.116 (.030)	0.022 (.009)	0.931
msft	(1,7,176)	0.399 (.025)	0.438 (.031)	0.148 (.023)	0.002 (.008)	0.985
pfe	(1,9,249)	0.355 (.024)	0.462 (.034)	0.126 (.034)	0.022 (.010)	0.943
pg	(1,10,248)	0.411 (.025)	0.462 (.029)	0.107 (.025)	0.004 (.006)	0.980
t	(1,9,240)	0.323 (.025)	0.456 (.032)	0.178 (.028)	0.023 (.008)	0.957
utx	(1,6,249)	0.261 (.026)	0.568 (.031)	0.131 (.021)	0.016 (.007)	0.960
vz	(1,10,231)	0.317 (.037)	0.615 (.046)	0.047 (.036)	0.000 (.011)	0.979
wmt	(1,8,249)	0.318 (.029)	0.534 (.036)	0.130 (.027)	0.006 (.007)	0.982
xom	(1,4,27)	0.210 (.038)	0.554 (.053)	0.192 (.032)	0.006 (.006)	0.956

Table 2.6: **Model B: Out-Of-Sample Fit Estimation.** The first column displays the stock symbol. Column 2 presents the optimal lags according to the out-of-sample fit. Columns 3 to 6 give the parameter estimates. The last column is the persistence measure, calculated as the sum of the β coefficients. The figures in parentheses are standard errors.

	lags	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{c}$	$\hat{\phi}$
aa	(1,11,132)	0.325 (.027)	0.511 (.036)	0.122 (.033)	0.017 (.009)	0.958
aig	(1,9,15)	0.343 (.030)	0.339 (.064)	0.237 (.069)	0.023 (.007)	0.919
axp	(1,7,133)	0.378 (.026)	0.438 (.034)	0.170 (.027)	0.003 (.007)	0.986
ba	(1,4,131)	0.318 (.027)	0.366 (.040)	0.268 (.032)	0.020 (.010)	0.951
c	(1,8,131)	0.354 (.027)	0.454 (.032)	0.175 (.027)	0.005 (.007)	0.982
cat	(1,11,195)	0.283 (.028)	0.523 (.049)	0.149 (.045)	0.029 (.015)	0.954
dd	(1,13,250)	0.345 (.028)	0.541 (.034)	0.090 (.027)	0.009 (.008)	0.975
dis	(1,5,154)	0.247 (.027)	0.519 (.033)	0.209 (.030)	0.010 (.007)	0.974
ge	(1,5,140)	0.328 (.028)	0.457 (.036)	0.190 (.026)	0.007 (.006)	0.975
gm	(1,4,14)	0.234 (.025)	0.208 (.047)	0.456 (.042)	0.028 (.006)	0.898
hd	(1,2,15)	0.230 (.034)	0.189 (.049)	0.509 (.035)	0.031 (.007)	0.928
hon	(1,5,127)	0.256 (.029)	0.452 (.041)	0.255 (.034)	0.016 (.008)	0.963
hpq	(1,2,16)	0.231 (.048)	0.318 (.059)	0.420 (.039)	0.011 (.013)	0.968
ibm	(1,6,126)	0.372 (.027)	0.432 (.037)	0.176 (.027)	0.005 (.006)	0.979
intc	(1,16,90)	0.483 (.024)	0.361 (.041)	0.127 (.032)	0.019 (.009)	0.971
jnj	(1,10,247)	0.332 (.025)	0.520 (.033)	0.105 (.030)	0.007 (.006)	0.956
jpm	(1,3,126)	0.245 (.027)	0.452 (.038)	0.279 (.033)	0.008 (.007)	0.976
ko	(1,6,125)	0.311 (.028)	0.471 (.038)	0.181 (.034)	0.009 (.005)	0.962
mcd	(1,5,28)	0.244 (.024)	0.309 (.041)	0.377 (.039)	0.022 (.008)	0.930
mmm	(1,6,125)	0.309 (.028)	0.465 (.039)	0.183 (.031)	0.009 (.005)	0.956
mo	(1,2,40)	0.220 (.031)	0.300 (.042)	0.403 (.030)	0.021 (.006)	0.923
mrk	(1,16,56)	0.401 (.022)	0.402 (.054)	0.123 (.051)	0.020 (.008)	0.925
msft	(1,8,137)	0.415 (.024)	0.417 (.031)	0.148 (.022)	0.006 (.008)	0.982
pfe	(1,20,209)	0.402 (.022)	0.453 (.037)	0.094 (.041)	0.019 (.011)	0.948
pg	(1,8,68)	0.376 (.027)	0.436 (.034)	0.149 (.029)	0.009 (.005)	0.961
t	(1,6,63)	0.286 (.026)	0.390 (.037)	0.289 (.031)	0.016 (.007)	0.964
utx	(1,4,70)	0.222 (.025)	0.458 (.034)	0.282 (.031)	0.013 (.007)	0.961
vz	(1,2,21)	0.208 (.044)	0.296 (.058)	0.456 (.038)	0.013 (.010)	0.959
wmt	(1,24,212)	0.416 (.032)	0.479 (.047)	0.084 (.040)	0.006 (.009)	0.979
xom	(1,12,13)	0.376 (.035)	0.878 (.294)	-0.312 (.296)	0.009 (.007)	0.942

Table 2.7: **Model B: Forecasting results.** The first column displays the stock symbol. Column 2 reports the mean-squared error (MSE) for the one-day ahead forecast of realized volatility using the model chosen by the in-sample fit criterion. Column 3 reports the p-values of the Superior Predictive Ability (SPA) test of Hansen (2005) for the in-sample fit criterion. The null hypothesis for the SPA test is that the benchmark model is the best forecasting model. The benchmark considered is the (1,5,21) model specification. The last two columns present the MSE and p-values for the model specification based on the out-of-sample fit. The MSE are calculated as a percentage of the MSE of the benchmark model.

	$MSE(1)^{LL}$	$p-val^{LL}$	$MSE(1)^{FCH}$	$p-val^{FCH}$
aa	0.994	0.960	0.983	0.717
aig	1.03	0.906	0.992	0.496
axp	1.002	0.268	0.988	0.735
ba	0.998	0.703	0.98	0.767
c	0.946	0.762	0.943	0.892
cat	0.991	0.959	0.985	0.803
dd	0.994	0.725	0.983	0.891
dis	1.007	0.942	0.978	0.736
ge	0.987	0.374	0.978	0.509
gm	1.025	0.760	0.985	0.888
hd	0.991	0.634	0.98	0.825
hon	0.989	0.758	0.941	0.061
hpq	1.001	0.516	0.988	0.760
ibm	0.992	0.563	0.976	0.563
intc	0.995	0.750	0.964	0.678
jnj	0.982	0.494	0.959	0.128
jpm	1.007	0.424	0.961	0.447
ko	0.967	0.506	0.962	0.277
mcd	0.999	0.625	0.994	0.108
mmm	1.015	0.300	0.962	0.527
mo	1.018	0.465	0.987	0.415
mrk	0.991	0.900	0.967	0.907
msft	0.976	0.623	0.97	0.641
pfe	0.998	0.077	0.979	0.241
pg	0.996	0.625	0.984	0.896
t	1.043	0.138	0.965	0.138
utx	0.972	0.095	0.952	0.070
vz	1.01	0.277	0.983	0.601
wmt	0.998	0.473	0.964	0.898
xom	1.016	0.248	0.943	0.895

Table 2.8: **Model C: In-Sample Fit Estimation** The first column displays the stock symbol. Column 2 presents the optimal lags according to the in-sample fit. Columns 3 to 6 give the parameter estimates. The last column is the persistence measure, calculated as the sum of the β coefficients. The figures in parentheses are standard errors.

	lags	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{c}$
aa	(1,6,229)	0.267(.027)	0.542(.036)	0.168(.027)	-0.012(.011)
aig	(1,5,227)	0.254(.033)	0.585(.040)	0.127(.027)	-0.009(.009)
axp	(1,11,248)	0.439(.030)	0.449(.034)	0.075(.029)	-0.005(.012)
ba	(1,13,250)	0.371(.026)	0.459(.034)	0.150(.032)	-0.016(.014)
c	(1,10,250)	0.365(.029)	0.506(.034)	0.144(.028)	-0.035(.011)
cat	(1,13,250)	0.262(.035)	0.558(.060)	0.145(.059)	-0.006(.021)
dd	(1,10,236)	0.322(.031)	0.543(.040)	0.126(.029)	-0.019(.010)
dis	(1,8,249)	0.289(.024)	0.544(.030)	0.139(.028)	-0.007(.009)
ge	(1,10,250)	0.367(.027)	0.487(.034)	0.110(.025)	-0.008(.010)
gm	(1,10,145)	0.316(.025)	0.537(.036)	0.118(.033)	-0.015(.010)
hd	(1,11,250)	0.378(.028)	0.512(.032)	0.100(.032)	-0.023(.012)
hon	(1,8,247)	0.310(.031)	0.514(.041)	0.178(.033)	-0.030(.011)
hpq	(1,3,227)	0.221(.060)	0.361(.083)	0.295(.049)	-0.042(.028)
ibm	(1,5,249)	0.339(.028)	0.467(.035)	0.170(.028)	-0.006(.010)
intc	(1,5,247)	0.393(.029)	0.429(.034)	0.160(.021)	-0.003(.009)
jnj	(1,8,250)	0.308(.025)	0.540(.033)	0.123(.035)	-0.015(.011)
jpm	(1,11,250)	0.335(.029)	0.540(.034)	0.117(.025)	-0.022(.010)
ko	(1,6,250)	0.291(.032)	0.537(.039)	0.153(.030)	-0.014(.008)
mcd	(1,9,250)	0.285(.026)	0.507(.039)	0.184(.037)	-0.018(.011)
mmm	(1,14,250)	0.371(.025)	0.505(.034)	0.098(.034)	-0.014(.008)
mo	(1,16,250)	0.370(.025)	0.477(.034)	0.082(.043)	-0.005(.016)
mrk	(1,8,242)	0.337(.027)	0.477(.035)	0.123(.029)	0.000(.010)
msft	(1,6,247)	0.375(.036)	0.463(.037)	0.164(.022)	-0.024(.008)
pfe	(1,9,249)	0.352(.024)	0.456(.030)	0.134(.032)	0.003(.013)
pg	(1,10,249)	0.385(.023)	0.492(.028)	0.096(.024)	-0.010(.007)
t	(1,9,240)	0.330(.026)	0.445(.033)	0.187(.028)	-0.004(.009)
utx	(1,6,249)	0.236(.027)	0.593(.032)	0.140(.021)	-0.016(.009)
vz	(1,6,234)	0.262(.040)	0.661(.049)	0.083(.039)	-0.031(.016)
wmt	(1,8,249)	0.309(.028)	0.535(.034)	0.139(.028)	-0.017(.009)
xom	(1,3,28)	0.191(.041)	0.565(.059)	0.209(.033)	-0.015(.007)

Table 2.9: **Model C: Out-Of-Sample Fit Estimation** The first column displays the stock symbol. Column 2 presents the optimal lags according to the in-sample fit. Columns 3 to 6 give the parameter estimates. The figures in parentheses are standard errors.

	lags	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{c}$
aa	(1,11,116)	0.316(.026)	0.516(.037)	0.137(.035)	-0.008(.011)
aig	(1,15,136)	0.389(.032)	0.488(.044)	0.075(.034)	-0.009(.010)
axp	(1,7,218)	0.334(.029)	0.468(.034)	0.142(.028)	-0.006(.011)
ba	(1,4,135)	0.326(.028)	0.365(.042)	0.291(.032)	-0.015(.013)
c	(1,8,245)	0.356(.028)	0.477(.033)	0.163(.027)	-0.021(.011)
cat	(1,10,245)	0.258(.036)	0.526(.062)	0.190(.058)	-0.009(.020)
dd	(1,13,249)	0.340(.030)	0.540(.039)	0.105(.031)	-0.015(.011)
dis	(1,5,120)	0.257(.028)	0.469(.044)	0.248(.032)	-0.012(.009)
ge	(1,5,122)	0.316(.028)	0.470(.037)	0.198(.026)	-0.021(.008)
gm	(1,14,193)	0.498(.026)	0.385(.036)	0.150(.036)	0.008(.011)
hd	(1,2,20)	0.216(.037)	0.234(.049)	0.492(.036)	0.004(.009)
hon	(1,5,127)	0.247(.030)	0.454(.041)	0.270(.035)	-0.017(.011)
hpq	(1,20,21)	0.458(.045)	1.176(.584)	-0.660(.590)	-0.007(.022)
ibm	(1,6,126)	0.367(.028)	0.411(.038)	0.174(.029)	-0.004(.009)
intc	(1,16,106)	0.486(.024)	0.355(.038)	0.133(.030)	0.002(.009)
jnj	(1,10,137)	0.316(.024)	0.535(.036)	0.110(.036)	-0.019(.009)
jpm	(1,4,126)	0.279(.025)	0.444(.034)	0.262(.030)	-0.019(.008)
ko	(1,6,138)	0.300(.031)	0.490(.041)	0.180(.034)	-0.018(.007)
mcd	(1,10,28)	0.283(.026)	0.367(.053)	0.282(.052)	-0.002(.009)
mmm	(1,4,63)	0.260(.028)	0.419(.041)	0.270(.033)	-0.012(.006)
mo	(1,5,59)	0.300(.023)	0.330(.036)	0.322(.037)	-0.014(.010)
mrk	(1,57,58)	0.279(.025)	0.444(.034)	0.262(.030)	-0.019(.008)
msft	(1,7,137)	0.386(.033)	0.440(.036)	0.158(.023)	-0.014(.007)
pfe	(1,20,213)	0.401(.023)	0.442(.037)	0.104(.042)	-0.001(.015)
pg	(1,5,66)	0.292(.027)	0.458(.036)	0.240(.027)	-0.017(.005)
t	(1,6,70)	0.304(.028)	0.397(.038)	0.285(.031)	-0.016(.008)
utx	(1,4,69)	0.202(.026)	0.450(.037)	0.298(.032)	-0.016(.009)
vz	(1,2,19)	0.215(.051)	0.285(.068)	0.462(.045)	-0.009(.013)
wmt	(1,10,121)	0.328(.027)	0.505(.039)	0.136(.032)	-0.010(.009)
xom	(1,12,20)	0.36(.037)	0.661(.083)	-0.092(.076)	0.003(.008)

Table 2.10: **Model C: Forecasting results.** The first column displays the stock symbol. Column 2 reports the mean-squared error (MSE) for the one-day ahead forecast of realized volatility using the model chosen by the in-sample fit criterion. Column 3 reports the p-values of the Superior Predictive Ability (SPA) test of Hansen (2005) for the in-sample fit criterion. The null hypothesis for the SPA test is that the benchmark model is the best forecasting model. The benchmark considered is the (1,5,21) model specification. The last two columns present the MSE and p-values for the model specification based on the out-of-sample fit. The MSE are calculated as a percentage of the MSE of the benchmark model.

	$MSE(1)^{LL}$	$p-val^{LL}$	$MSE(1)^{FCH}$	$p-val^{FCH}$
aa	0.986	0.961	0.986	0.450
aig	0.907	0.847	0.860	0.749
axp	0.728	0.297	0.703	0.860
ba	1.010	0.782	0.962	0.798
c	0.953	0.837	0.953	0.964
cat	0.972	0.765	0.972	0.568
dd	1.005	0.666	0.985	0.847
dis	1.000	0.926	0.980	0.218
ge	0.980	0.413	0.959	0.492
gm	1.019	0.706	0.959	0.680
hd	1.002	0.484	0.985	0.293
hon	1.003	0.725	0.949	0.010
hpq	1.028	0.555	0.982	0.504
ibm	0.982	0.205	0.982	0.484
intc	1.013	0.808	0.958	0.602
jnj	0.966	0.671	0.933	0.145
jpm	1.026	0.441	0.950	0.319
ko	0.955	0.362	0.955	0.204
mcd	1.007	0.065	0.991	0.177
mmm	1.012	0.307	0.950	0.440
mo	1.009	0.338	0.338	0.412
mrk	0.986	0.819	0.934	0.960
msft	0.979	0.601	0.973	0.639
pfe	0.997	0.082	0.975	0.228
pg	0.482	0.550	0.469	0.420
t	1.037	0.140	0.954	0.146
utx	0.970	0.502	0.949	0.530
vz	1.027	0.764	0.989	0.751
wmt	0.928	0.503	0.893	0.508
xom	1.014	0.533	0.921	0.928

Chapter 3

The Serial Correlation Structure of Heterogeneous ARCH and Standard ARCH Models

3.1 Introduction

Volatility is a very important determinant in many financial applications, with a range extending from option and asset pricing to portfolio and risk management. It is very well known that volatility is changing over time. Therefore, understanding the temporal dependence present in the second moment of asset returns and finding a suitable model that captures this feature of the data becomes crucial.

Many models have been proposed to capture the change in volatility over time. Probably the most widely used model is ARCH (Auto Regressive Conditional Heteroscedasticity) developed by Engle in 1982. Since the development of the ARCH model, a lot of research has been employed in the direction of extending this model and developing other volatility models. Among the ARCH family, the most important and frequently used models are GARCH, EGARCH (Nelson 1991), IGARCH (Bollerslev and Engle 1986) and FIGARCH (Baillie, Bollerslev and Mikkelsen 1996). In practice, the ARCH model requires a large number of parameters in order to accurately describe the volatility process. The GARCH model (Generalized ARCH) was proposed by Bollerslev (1986) in order to overcome this problem. The EGARCH model proposed by Nelson (1991) accounts for the negative correlation between volatility and stock return changes. Other extensions of the ARCH model are: GARCH-t (Bollerslev 1987), ARCH-M (Engle, Lilien and Robins 1987), AGARCH (Engle 1990), NGARCH (Higgins and Bera 1992), QARCH (Sentana 1992), PGARCH (Bollerslev and Ghysels 1996).

The most important task of any volatility model is to be able to forecast volatility. Therefore any volatility model should be able to replicate the stylized facts about volatility: high persistence, mean reversion, asymmetric impact of positive and negative innovations and possible exogenous variable influences on volatility (Engle and Patton 2001). Persistence and mean reversion are connected notions since different persistence levels of shocks will imply different mean reversion times for the process to revert to the mean level of volatility and eliminate the effect of shocks. Recent studies reveal a new stylized fact of financial volatility data, namely that in addition to a long correlation structure it also features a short correlation structure that reverts to the mean within a few days (Hillebrand 2006). Furthermore, there is evidence that fluctuations with long mean reversion and fluctuations with short mean reversion are connected (Müller et al. 1993).

Different models were proposed to account for the different patterns observed in the data. The most reported stylized fact about volatility is long memory or high persistence (Ding, Engle, and Granger 1993). This property is reflected in the slow decay observed in the autocorrelations of absolute or squared returns or in sums of autoregressive coefficients close to unity. Many researchers report estimations based on GARCH specifications with an estimated sum of the autoregressive coefficients very close to one. The IGARCH (Integrated GARCH) class of models was proposed by Engle and Bollerslev (1986) to capture this empirical fact of very high volatility persistence. Other authors resort to fractionally integrated processes like ARFIMA (Granger an Joyeux 1980) to capture long memory. In the context of GARCH, the FIGARCH model (Fractionally Integrated GARCH) model proposed by

Baillie et al. (1996) incorporates the long memory property found in the autocorrelation of squared or absolute returns into the ARCH framework.

ARFIMA and (G)ARCH models with high values for the persistence parameter are one way to model long memory. Other sources of long memory are structural breaks and aggregation (Hyung, Poon, and Granger 2005). Several studies discuss the consequences of structural breaks in data generating parameters. In the context of GARCH models this problem was addressed, for example by Diebold (1986), Lamoureux and Lastrapes (1990), Mikosch and Starica (2004), and Hillebrand (2005). The common finding in these studies is that neglecting parameter changes can induce high persistence. Granger (1980) shows that long memory can be approximated by aggregating processes with short memory. Examples of such models are the HARCH model (Müller et al. 1997) in the context of latent model specifications and the HAR-RV model (Corsi 2004) in the context of realized volatility models.

In this chapter I focus on the long memory property, therefore I consider the HARCH model (Heterogeneous ARCH). The HARCH model considers volatility as a sum of different components, each component representing a different time interval. In this setup, long term fluctuations are used to forecast short term volatility. Therefore HARCH models capture conditional volatility in a different manner than ARCH-GARCH models in terms of how past returns are entered into the conditional volatility equation. While HARCH uses the idea of aggregation, (G)ARCH relies on high levels for the persistence parameter. Therefore, before abandoning ARCH specifications in favor of HARCH, we need a clarification in terms of the correspondence between the two models. We start by investigating how the correlation structure from the ARCH-GARCH models relates to the correlation structure from the HARCH model. Since HARCH nests ARCH, we employ simulations to determine if the HARCH model is able to capture the correlation structure present in ARCH models and vice versa. Our simulation studies show that the HARCH model does not seem to be able to pick up the correlation scales in the mean reverting sense present in ARCH models.

The above finding has important implications for persistence. There are different ways to define persistence in the context of (G)ARCH models (Engle and Patton 2001). The sum of the autoregressive coefficients is the most common volatility measure employed in the literature. The closer this sum is to one, the more persistent the influence of a shock on volatility will be. Following this idea a similar persistence measure can be constructed for HARCH. Our scientific interest is if the corresponding HARCH persistence measure can capture the persistence of the ARCH model. Our simulations show that this is not the case. Understanding how different models capture long memory is very important. Therefore we also investigate the relationship between HARCH models and other models that capture long memory. We consider ARFIMA, IGARCH and GARCH models with break points. Since these types of models were developed to account for the high persistence phenomenon found in the data it is interesting to understand in which way the HARCH model captures the characteristics of such data.

In order to address these problems an optimal lag structure must be identified. A parsimonious model of the lag structure is often determined on an ad hoc basis. Because there is no theoretical reason to choose one set of structures over others, we treat the maximum number of lags allowed as fixed and determine the lags by examining all combinations of possible lags. This, however, demands substantial computing power and we employ parallel computing algorithms to consider all these cases. This computational approach will allow us to determine the optimal lag structure and therefore identify the corresponding different time scales found in the data. In doing this we must address the problem of the correspondence between the lag structure and serial correlation. This is a very important problem since it is not clear how the two measures relate and if the optimal lag structure captures the serial correlation measure. We find that long memory in the process translates into large lag values in HARCH rather than in large values of the sum of the coefficients.

The HARCH lags do not capture the serial correlation of the ARCH model, but capture the influence of the lags used in the data generating process. Finally, HARCH is not able to capture GARCH, neither by the persistence measure, nor by the lags.

The remainder of this chapter is organized as follows. In Section 3.2 we discuss ARCH, GARCH, and HARCH models. Section 3.3 discusses the design used in our simulations and the results from running these simulations. Section 3.3.3 presents the optimal lag methodology and results.

3.2 ARCH Models

The ARCH(p) model proposed by Engle (1982) is given by:

$$\log S_t - \log S_{t-1} = r_t = \mu + \varepsilon_t, \tag{3.1}$$

$$\varepsilon_t|_{\mathcal{F}_{t-1}} \sim \mathcal{N}(0,\, h_t), \tag{3.2}$$

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2. \tag{3.3}$$

where r_t are the log returns, μ is the conditional mean and h_t is the time-dependent conditional variance. The ARCH model was proposed to capture the volatility clustering phenomenon. This is achieved by modeling the conditional variance as a function of past squared returns.

The serial correlation in the ARCH model is captured by the sum of the autoregressive coefficients α_j. The stationarity condition is given by $\sum_{j=1}^{p} \alpha_j < 1$. The closer the sum of the autoregressive coefficients gets to one, the more persistent a shock will be on the conditional variance. Let $\phi = \sum_{j=1}^{p} \alpha_j$. Therefore ϕ can be interpreted as a measure of persistence. It is the fraction of the shock that is carried forward in time. Alternatively, $1 - \phi$ is the fraction of the shock that is "washed out" in each period and $1/(1 - \phi)$ is the amount of time necessary for the process to return to the mean. The more persistent a process will be, the longer it will take for the process to revert to the unconditional mean given by $\mathbb{E}h_t = w/(1 - \sum_{j=1}^{p} \alpha_j)$.

In practice, ARCH models require a large number of parameters in the conditional variance equation in order to capture the features found in the data. The GARCH(p, q) model (Bollerslev 1986) was proposed to address this problem. The difference between ARCH and GARCH models is that the latter allows past lagged values of variance in the conditional variance equation. The GARCH model is given by:

$$\log S_t - \log S_{t-1} = r_t = \mu + \varepsilon_t, \tag{3.4}$$

$$\varepsilon_t|_{\mathcal{F}_{t-1}} \sim \mathcal{N}(0,\, h_t), \tag{3.5}$$

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i}. \tag{3.6}$$

The stationarity condition for the GARCH model is analogous to the ARCH condition. As long as the sum $\phi = \sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i < 1$, the process is stationary. There are several different ways to define persistence in GARCH models. The coefficient ϕ calculated as the sum of the autoregressive parameters is the most common measure of persistence used. The unconditional mean is then given by $\mathbb{E}h_t = w/(1 - \sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i)$.

Both ARCH and GARCH models use a sum of past squared returns in the conditional variance equation. The HARCH model differs in this respect, since it uses sums of squared returns aggregated over different intervals of time. The HARCH model as proposed by Müller et al. (1997) is given by:

$$\log S_t - \log S_{t-1} = \mu + \varepsilon_t, \tag{3.7}$$

$$\varepsilon_t|_{\mathcal{F}_{t-1}} \sim \mathcal{N}(0,\, h_t), \tag{3.8}$$

$$h_t = \omega + \sum_{k=1}^{n} \alpha_k \left(\sum_{i=1}^{k} \varepsilon_{t-i} \right)^2 . \tag{3.9}$$

The model as described specifies the inclusion of all lags from 1 to n. This can be too restrictive. A modified version of the specification described above allows the inclusion of a specified number of lags. Let $K = (k_1, k_2, \ldots, k_m)$ be a set of m indices with $k_1 < k_2 < \cdots < k_m$. Then equation (3.9) becomes:

$$h_t = \omega + \sum_{j \in K} \alpha_j \left(\sum_{i=1}^{j} \varepsilon_{t-i} \right)^2 . \tag{3.10}$$

The persistence measure for HARCH models is derived by Müller et al.(1997). Equation (3.10) can be written as:

$$h_t = \omega + \sum_{j \in K} \alpha_j (\varepsilon_{t-1} + \varepsilon_{t-2} \ldots + \varepsilon_{t-j})^2 = \omega + \sum_{j \in K} \alpha_j \left(\sum_{i=1}^{j} \varepsilon_{t-i}^2 + 2 \sum_{\substack{i,k=1 \\ i \neq k}}^{j} \varepsilon_{t-i} \varepsilon_{t-k} \right) . \tag{3.11}$$

Taking expectations in the previous equation we obtain:

$$\mathbb{E}h_t = \omega + \sum_{j \in K} \alpha_j \left(\sum_{i=1}^{j} \mathbb{E}\varepsilon_{t-i}^2 + 2 \sum_{\substack{i,k=1 \\ i \neq k}}^{j} \mathbb{E}(\varepsilon_{t-i} \varepsilon_{t-k}) \right) . \tag{3.12}$$

Let $\varepsilon_t = \sqrt{h_t}\eta_t$, with $\eta_t \sim iid(0, 1)$. Because $\mathbb{E}\eta_t = 0$ and $\mathbb{E}\eta_t^2 = 1$ the expectation of the cross product is zero and (3.12) becomes:

$$\mathbb{E}h_t = \omega + \sum_{j \in K} \alpha_j \left(\sum_{i=1}^{j} \mathbb{E}h_{t-i} \right) . \tag{3.13}$$

If h_t is stationary and $\mathbb{E}h_t = \mathbb{E}h_{t-i}$ for $i = 1 \ldots j$ then

$$\mathbb{E}h_t = \frac{\omega}{\sum_{j \in K} j\alpha_j}. \tag{3.14}$$

Let

$$\phi = \sum_{j \in K} j\, \alpha_j. \tag{3.15}$$

By analogy with ARCH models, a necessary condition for stationarity is given by $\phi < 1$. Müller et al.(1997) and Embrechts et al.(1998) provide proofs for the sufficiency of this condition. The persistence measure is thus given by ϕ and the unconditional mean of the process is $\mathbb{E}h_t = 1/(1-\phi)$.

Looking at equation (3.10) we observe that HARCH nests the ARCH specification. To see this write (3.10) as:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \sum_{\substack{j \in K \\ j \geq 2}} \alpha_j \left(\sum_{i=1}^{j} \varepsilon_{t-i} \right)^2. \tag{3.16}$$

By setting $\alpha_j = 0, j \geq 2$ in the previous equation we obtain

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2. \tag{3.17}$$

Equation (3.17) is an ARCH(1) specification. For $\alpha_j \neq 0, j \geq 2$, to see the relationship between HARCH and ARCH models it is useful to write (3.16) in the form of (3.12):

$$h_t = \omega + \sum_{j \in K} \alpha_j \left(\sum_{i=1}^{j} \varepsilon_{t-i}^2 \right) + 2 \sum_{j \in K} \alpha_j \left(\sum_{\substack{i,k=1 \\ i \neq k}}^{j} \varepsilon_{t-i} \varepsilon_{t-k} \right). \tag{3.18}$$

The first term $\sum_{j \in K} \alpha_j \left(\sum_{i=1}^{j} \varepsilon_{t-i}^2 \right)$ in the previos equation corresponds to an ARCH specification. The difference between (3.18) and an ARCH(2) models is given by the cross-terms $\varepsilon_{t-i}\varepsilon_{t-k}$.

3.3 Research Design and Methods

3.3.1 Design

The objective in this section is to understand in which way HARCH captures long memory. For this purpose we compare the performance of the HARCH model in terms of the correlation structure with other data generating processes. Long memory is defined as hyperbolic decay of the autocorrelation function (Baillie 1996) and is measured in a sum of autocorrelation coefficients very close to one. Therefore we use data generating processes that have this feature and estimate HARCH on it. The interest is to understand in what way the HARCH model can capture the correlation structure used in the data generating process. There are two candidates for measuring high persistence in the context of the HARCH model: the corresponding persistence measure ϕ and the lags. Since we know the characteristics of the data generating process, by estimating a HARCH specification on these data we can examine how the HARCH model captures these features of the data. In particular we are interested in understanding whether the HARCH model can capture the correlation structure in either the persistence measure or the lags, and what the relation between these two measures of persistence is.

In section 3.3.2 we look at data generating processes from different models and estimate a HARCH (2) specification on these data. We set different values for the data generating coefficients. In this way we can explore the way in which a HARCH (2) specification captures different levels of persistence. We always include lag 1 such that the specification nests an ARCH(1) specification. For each data generating process we generate 5000 samples of length $T = 10000$ points each. On each run we estimate a HARCH model with lags 1 and 2. Each model specification is estimated by maximum likelihood. The maximum likelihood estimation is coded in C++ using a quasi-Newton routine with analytical first derivatives and numerical second derivatives. The simulation results in 5000 parameter estimates for α_1 and α_2. Our interest is whether the data-generating persistence measure is reflected in the corresponding HARCH persistence measure $\phi = \alpha_1 + k\alpha_2$ with $k = 2$ (we always consider lags 1 and 2 in this section) or in the lag value $k = 2$.

We start by calibrating our procedure. Next, we focus on understanding how the correlation structure of the HARCH model relates to the correlation structure of ARCH(1) and ARCH(2) models. We expect HARCH to capture the correlation structure of these models in the corresponding persistence measure, since the two specifications are nested. We find that HARCH captures the persistence better in the lag values than in the corresponding measure ϕ. Next we employ models that approximate long memory by aggregation. For this purpose we use as data generating process an ARCH(2)+HARCH(2) specification. The corresponding HARCH persistence measure ϕ works well in this case, especially for high persistence. We also look at fractionally integrated models such as ARFIMA. Since these types of models are able to generate hyperbolic decay in the autocorrelations, we expect that HARCH should capture long memory by high values of the lags since HARCH is based on the idea of aggregation. We choose to estimate a HARCH(4) specification for this experiment. The results section indicate that the measure ϕ does not capture the data generating process. The coefficients attached to higher lag values are very small therefore HARCH cannot capture ARFIMA structures through lags. Finally, we explore the type of correlation structures captured by ARCH when applied on synthetic data obtained from a HARCH (2) process. Our results show that ARCH performs well on HARCH data. This result is not surprising since HARCH nests ARCH.

In Section 3.3.3 we take the analysis one step further and allow for the number of lags to be variable. In particular, we treat the maximum number of lags allowed as fixed and search for the best HARCH specification by choosing the specification resulting in the highest value for the log-likelihood. Again, we research whether the data generating correlation structure is reflected by ϕ or the optimal lag value. In addition to ARCH, we also consider other data generating specifications. We start by looking at GARCH models. We expect HARCH to capture the correlation structure of such models by high values of ϕ. Our findings show that HARCH is not able to capture GARCH dynamics. We also look at models with change points. Such models induce spurious high persistence when estimated without accounting for the break point. Therefore we expect large lag values or high ϕ values. We confirm large values for the HARCH optimal lags, but not for ϕ. This finding also holds when HARCH is applied on IGARCH data. For ARFIMA data generating processes we find that HARCH is not able to capture such dynamics.

3.3.2 Results

Calibration

We consider the following HARCH(2) specification as data generating process:

$$h_t = 6\text{e-}5 + 0.2\varepsilon_{t-1}^2 + 0.15\left(\sum_{j=1}^{2}\varepsilon_{t-j}\right)^2. \tag{3.19}$$

We generate 5000 samples of 10000 observations each. On each run, we estimate a HARCH(2) specification with lags $k_1 = 1$ and $k_2 = 2$. The average estimates for the two parameters are $\alpha_1 = 0.200$ and $\alpha_2 = 0.150$ implying an average persistence measure of 0.95. We can conclude from this experiment that we are able to retrieve the data generating parameters. Histograms for the two parameter estimates are presented below in Figure 3.1.

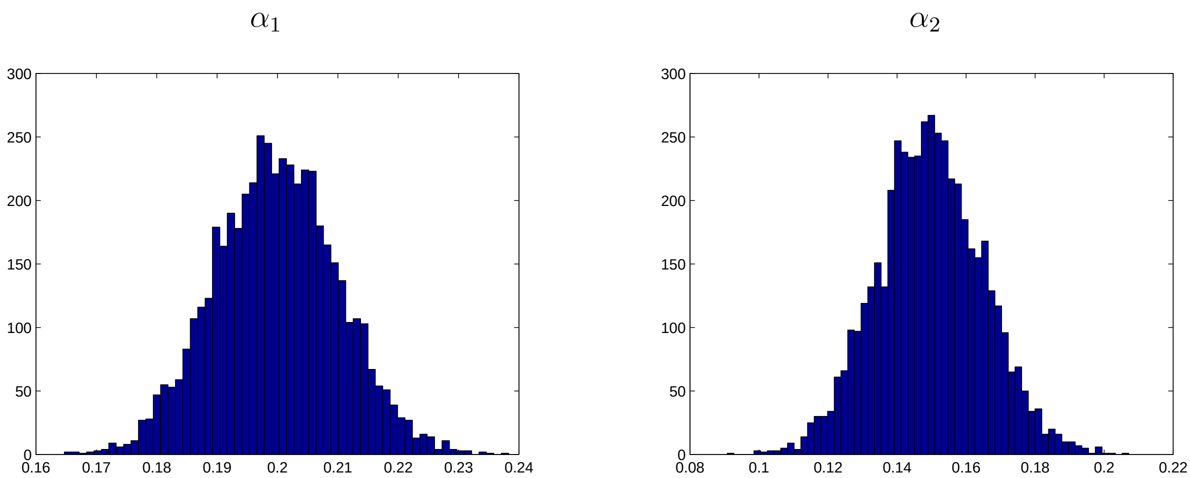


Figure 3.1: Histograms for calibration parameters.

ARCH(1)

As a data-generating process, we start by considering the following ARCH(1) model with a conditional variance process:

$$h_t = w + \alpha \varepsilon_{t-1}^2.$$

As previously stated, α is a measure of the persistence of the process. The closer α is to one, the more persistent the process will be. This means that volatility reverts very slowly to its unconditional mean after a shock. We vary the values used in the data generating process for α such that the synthetic data generated exhibits different correlation scales. The values used in the DGP vary according to Table 3.1. For the first scenario we use a value of $\alpha = 0.99$ which corresponds to high persistence. The average time for the process to revert to its mean of $8e - 2$ is 100 days, which corresponds to annualized volatility of 123%. For the next two scenarios we consider processes with faster mean reversion times of 4 units and 2 units of time corresponding to an annualized volatility of 24% and 16%.

We generate 5000 samples of length $T = 10000$ points each for each scenario. For each sample and for each scenario, we estimate the following HARCH(2) specification:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \left(\sum_{j=1}^{2} \varepsilon_{t-j} \right)^2 . \tag{3.20}$$

Table 3.1 reports the results from the HARCH estimation for all three scenarios considered. For the first scenario we see that the average estimate of α_1 is 0.91 and the average estimate of α_2 is 0.0006, corresponding to a persistence measure of 0.91. This value is very low when compared to the data generating persistence of 0.99 and does not correspond to a mean reversion time of 100 units. For the second scenario with the sum of the autoregressive parameters of 0.75, we obtain an average of 4 units of time for the process to revert to the unconditional mean 2.4e-4. The average estimates for α_1 is 0.73, the average estimate of α_2 is 0.001, and the estimated persistence measure is 0.73. This value is close to the mean reversion of 4 units of time used in the simulation. The third scenario uses a persistence measure of 0.35 and an average time for the process to revert to the unconditional mean of 1.5. The average estimates of α_1 and α_2 are 0.344 and 0.005 and the estimated persistence measure is 0.354 which corresponds to a mean reversion time of 1.54 units. A HARCH(2) specification with $\alpha_2 = 0$ is identical to an ARCH(1) specification. This explains the small coefficient values obtained for the second parameter. The estimated HARCH persistence level captures the influence of the first lag since the influence of the second one is negligible given the small values for the α_2 coefficients.

These simulations reveal that the HARCH(2) models captures the influence of the first lag used in the data generating process. In terms of the estimated persistence measure ϕ, for high levels of persistence, the estimated persistence measure in HARCH(2) does not capture the data generating correlation structures from ARCH(1) processes.

The corresponding histograms for the parameter estimates are presented in Figure 3.2. From the histograms we can see that the second parameter is very small, while the first parameter is capturing the persistence measure of the process. This is what we would expect since the ARCH(1) model is nested in the HARCH(2) specification when the second parameter is zero. Therefore we can conclude that HARCH captures the correlation structure of the data generating process through the influence of the first lag.

Table 3.1: Descriptive statistics of the estimates from HARCH on ARCH(1) data.

w	ϕ	$\frac{1}{1-\phi}$		$\hat{w}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\phi}$
	Scenario1						
6.00e-05	0.99	100	mean	0.0001	0.9161	0.0007	0.9175
	Scenario2		st.dev.	0	0.0846	0.0014	0.0855
6.00e-05	0.75	4	mean	0.0001	0.7306	0.0014	0.7335
	Scenario3		st.dev.	0	0.0432	0.0024	0.044
6.00e-05	0.35	1.5	mean	0.0001	0.3449	0.0046	0.3542
			st.dev.	0	0.0344	0.0143	0.0185

ARCH(2)

As a data-generating process, we consider an ARCH(2) model with a mean return of w and a conditional variance processes with different levels of persistence.

$$h_t = w + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2.$$

The corresponding persistence measure for this model is $\phi = \alpha_1 + \alpha_2$. We simulate data from this model according to Table 3.2 with correlation structures of 100, 4, and 2 units of time. On these data we estimate the HARCH(2) model specification from equation (3.1). The second panel of the table reports the estimation results.

The average estimate of α_1 and α_2 for the first scenario correspond to a persistence measure of 0.99, which corresponds to the data generating value of 0.99. For the second and third scenarios we obtain persistence measures of 0.72 and 0.36, which do not correspond to the data generating values of 0.75 and 0.45.

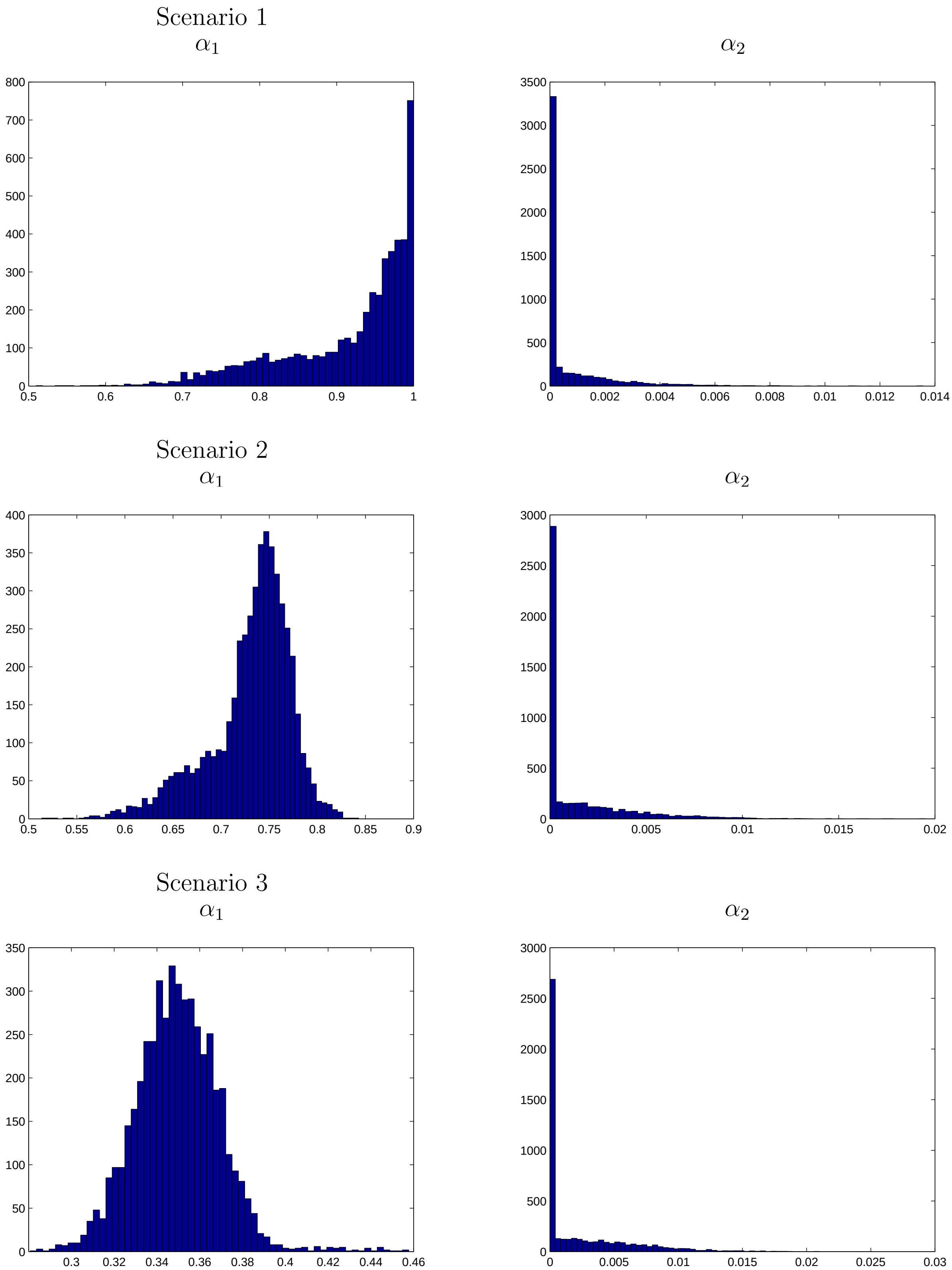


Figure 3.2: Histograms of the HARCH parameter estimates constructed according to Table 3.1.

We can conclude that the HARCH(2) model seems to be able to capture the data generating correlation structure better for high levels of persistence, than for low values of persistence when applied on ARCH(2) data. The model captures the influence of the second lag used in the ARCH specification. Compared to the results from the ARCH(1) scenarios, the estimated coefficient corresponding to the second lag is larger in magnitude. For the same level of persistence (0.99) the HARCH(2) model captures the influence of the lags used in the data generating process. Histograms for the two parameters are presented in Figure 3.3.

Table 3.2: Descriptive statistics of the HARCH(2) estimates on ARCH(2) data.

w	α_1	α_2	ϕ	$\frac{1}{1-\phi}$		$\hat{w}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\phi}$
	Scenario1								
6.00e-05	0.16	0.83	0.99	100	mean	0.0001	0.3816	0.3086	0.9988
					st.dev.	0	0.0322	0.0153	0.011
	Scenario2								
6.00e-05	0.1	0.65	0.75	4	mean	0.0001	0.1995	0.2592	0.7179
					st.dev.	0	0.0268	0.0169	0.0489
	Scenario3								
6.00e-05	0.1	0.35	0.45	2	mean	0.0001	0.1467	0.1054	0.3575
					st.dev.	0	0.0194	0.0098	0.0263

HARCH

In the previous section we investigated the way in which the HARCH model captures different correlation structures from the ARCH model. We now investigate how the ARCH model behaves when estimated on HARCH data. We consider as data generating process a HARCH(2) specification:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \left(\sum_{j=1}^{2} \varepsilon_{t-j} \right)^2 .$$

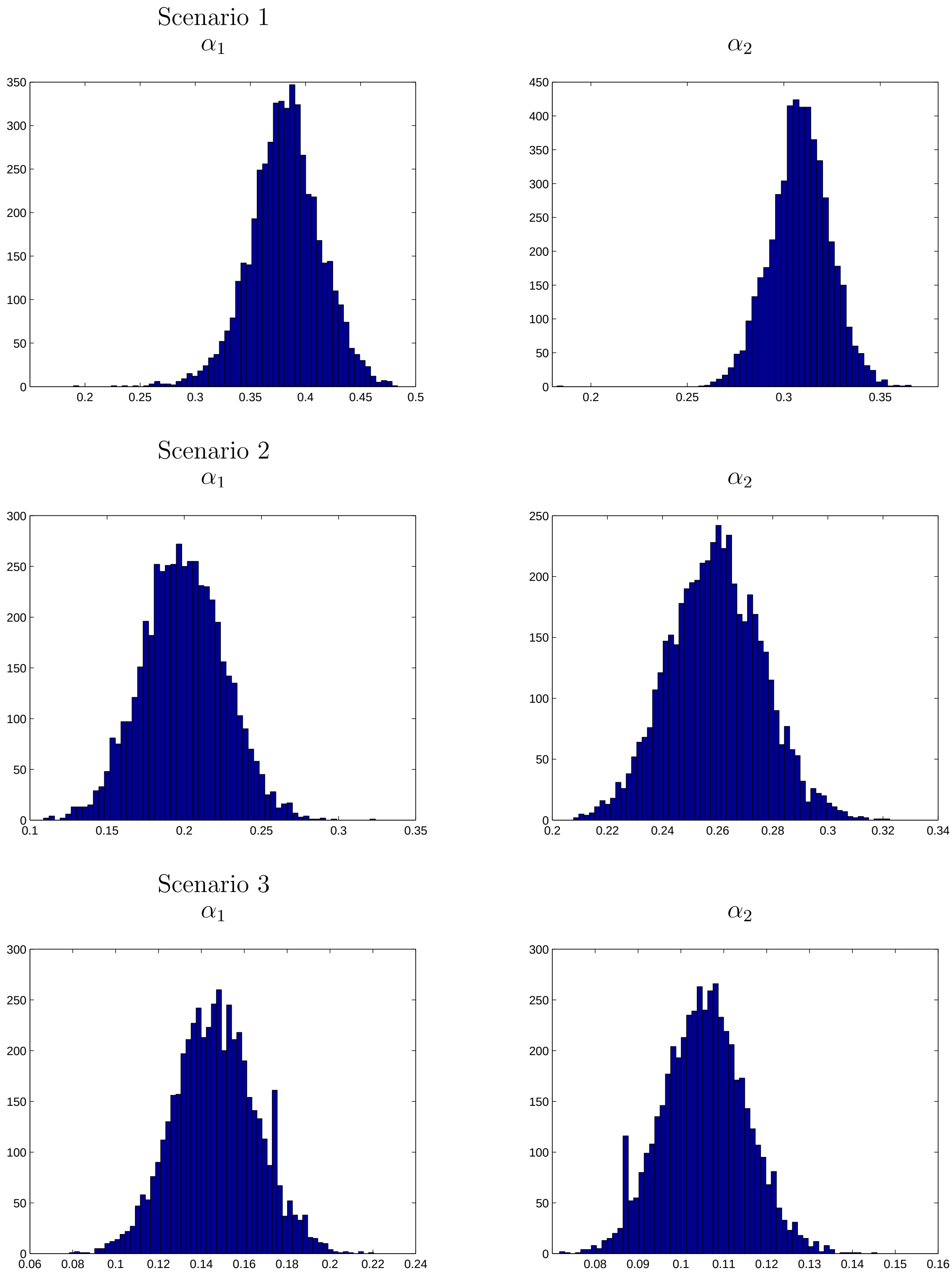


Figure 3.3: Histograms of the HARCH parameter estimates when the DGP is an ARCH(2) process constructed according to Table 3.2.

We look at different values for the data generating parameters according to Table 3.3. On each of these data generating processes we estimate an ARCH(2) specification. The second part of the table reports the results from the estimation. In all three scenarios considered, we obtain estimates of ϕ that are close to the true values used in the data generating process. In terms of the correlation structures, scenarios 2 and 3 come close to the data generation correlation structures. We conclude from this simulation that the ARCH(2) model performs well when applied to HARCH data. The histograms for the two parameter estimates for the scenarios considered are presented in Figure 3.4.

Table 3.3: Descriptive statistics of the estimates from ARCH on HARCH data.

w	α_1	α_2	ϕ	$\frac{1}{1-\phi}$		$\hat{w}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\phi}$
	Scenario1								
6.00e-05	0.29	0.35	0.99	100	mean	0.0001	0.6278	0.3440	0.9718
					st.dev.	0.0000	0.0218	0.0198	0.0199
	Scenario2								
6.00e-05	0.05	0.35	0.75	4	mean	0.0001	0.3995	0.3492	0.7487
					st.dev.	0.0000	0.0214	0.0200	0.0277
	Scenario3								
6.00e-05	0.15	0.15	0.45	2	mean	0.0001	0.2996	0.1494	0.4491
					st.dev.	0.0000	0.0179	0.0148	0.0218

ARCH(2)+HARCH(2)

We explore next the correlations captured by the HARCH model estimated on long memory data. The interest is in how an HARCH(2) model can capture the correlation structure from additive processes. We consider the following data generating process:

$$h_t = 6\text{e-}5 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \alpha_3 \left(\sum_{j=1}^{2} \varepsilon_{t-j} \right)^2 .$$

The stationarity condition for this model is $\alpha_1 + \alpha_2 + 2\alpha_3 < 1$. The persistence measure

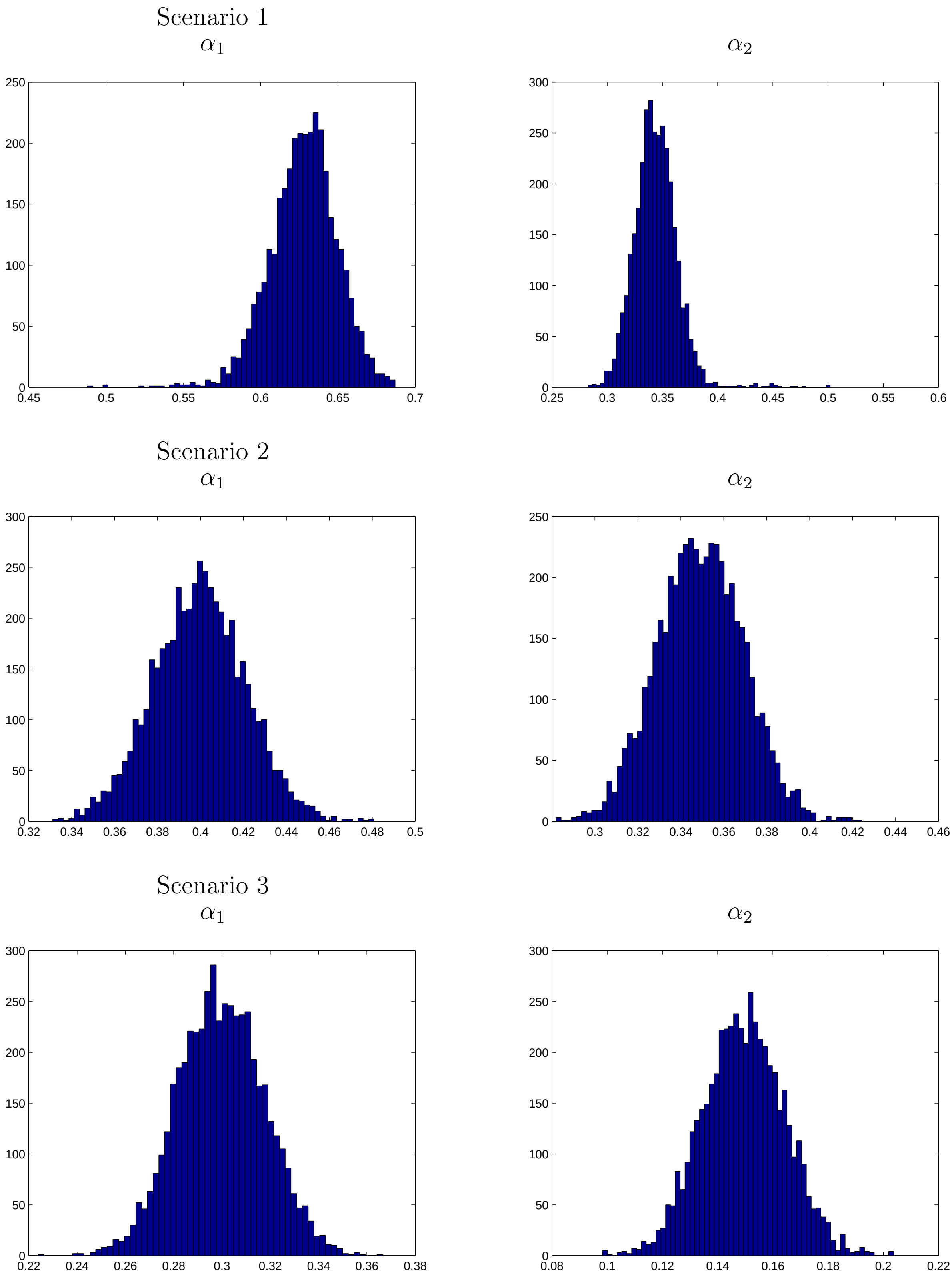


Figure 3.4: Histograms for ARCH(2) parameter estimates when the DGP is an HARCH(2) process constructed according to Table 3.3.

becomes $\phi = \alpha_1 + \alpha_2 + 2\alpha_3$, and the average time required for the process to revert to the unconditional mean is given by $1/(1-\phi)$.

The different scenarios considered are presented in Table 3.4. For the first scenario we use a data generating process that results in a persistence measure of 0.99. When we estimate a HARCH(2) model specification on the data we obtain an average HARCH persistence measure of 0.9945 which corresponds to the value used in the data generating process. For scenarios 2 and 3 with persistence levels of 0.65 and 0.35 the average estimated persistence measures from the HARCH(2) model are 0.63 and 0.29 respectively. We can conclude that the HARCH(2) model performs well on data with high persistence. The histograms for the two parameter estimates are presented in Figure 3.5.

Table 3.4: Descriptive statistics of the estimates from a HARCH(2) estimation when the data generating process is an HARCH(2)+ARCH(2).

w	α_1	α_2	α_3	ϕ	$\frac{1}{1-\phi}$		$\hat{w}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\phi}$
	Scenario1									
6.00e-05	0.34	0.15	0.25	0.99	100	mean	0.0001	0.3990	0.2978	0.9946
						st.dev.	0.0000	0.0224	0.0100	0.0133
	Scenario2									
6.00e-05	0.3	0.15	0.1	0.65	3	mean	0.0001	0.3179	0.1605	0.6390
						st.dev.	0.0000	0.0198	0.0100	0.0250
	Scenario3									
6.00e-05	0.12	0.15	0.04	0.35	2	mean	0.0001	0.1161	0.0892	0.2945
						st.dev.	0.0000	0.0163	0.0082	0.0204

ARFIMA

ARFIMA is another class of models that are able to generate long memory. Both ARFIMA and ARCH models are mean reversion models, but one difference between the models lies in the autocorrelation function. For an ARFIMA process, autocorrelations exhibit a slow hyperbolic decay, while ARCH models exhibit geometrical decay. The hyperbolic

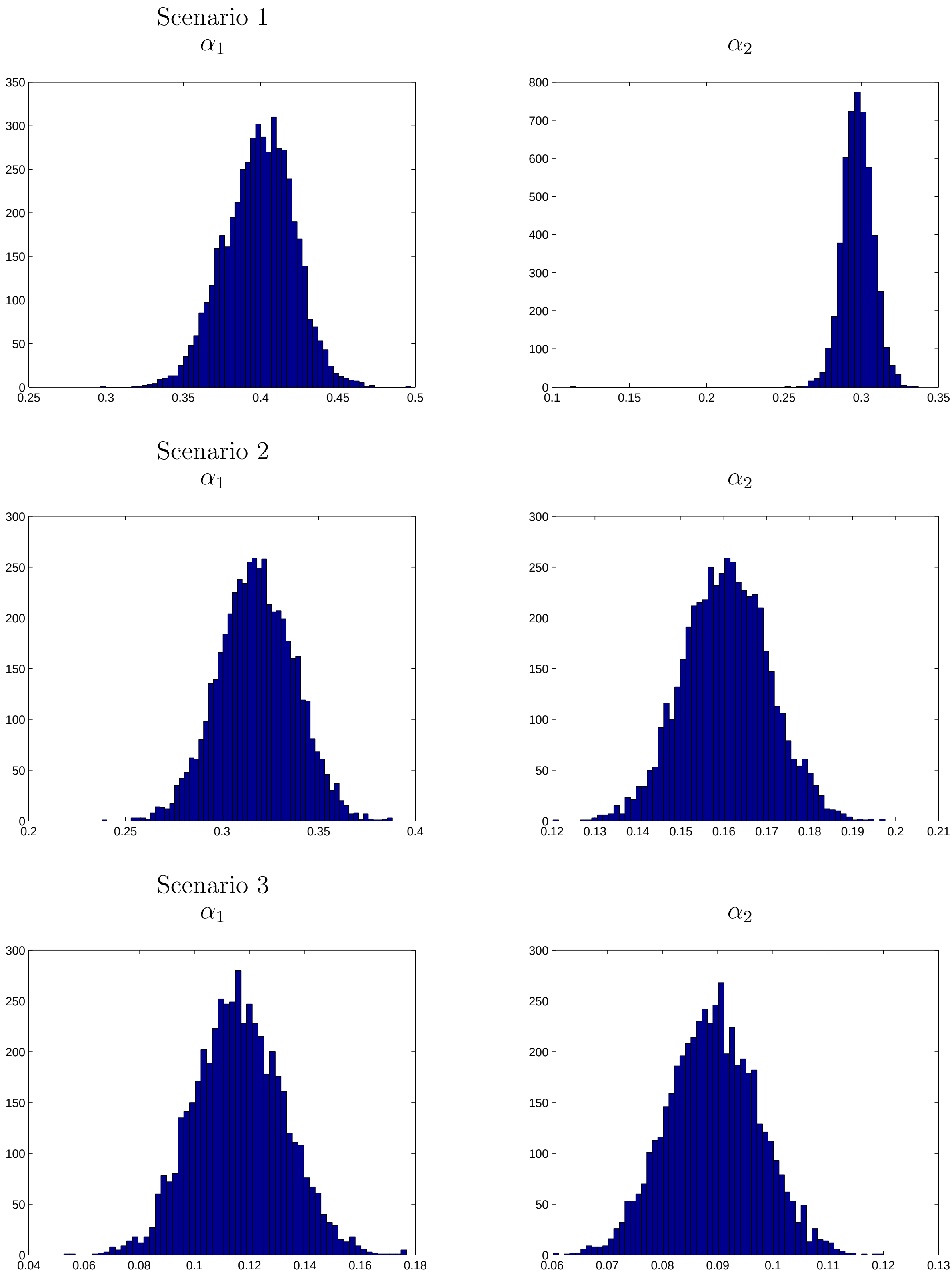


Figure 3.5: Histograms for HARCH parameter estimates when the DGP is an ARCH+HARCH process constructed according to Table 3.4.

decay of the autocorrelations is found often in financial data, therefore we wish to research how HARCH models behave when estimated on such data.

Granger and Joyeux (1980) started the literature on integrated processes in econometrics. A process is said to be integrated of order d if it has an ARMA representation after differencing d times. Fractionally integrated processes are obtained for values of d which are not integers. For such processes we obtain a slower decay in the autocorrelation function.

Consider the ARFIMA(0,d,0) process $(1-L)^d y_t = \varepsilon_t$, where L is the lag operator, ε_t is white noise with zero mean and constant variance, and d is the fractional integration parameter. A value of $d \in (0, 0.5)$ indicates stationary long memory. We explore different values of d in the data generating process according to Table 3.5.

The HARCH model often requires high values of n in practice in order to capture the high correlation scales found in the data. Therefore, we estimate the following HARCH(4) specification on each data generating process:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \left(\sum_{j=1}^{5} \varepsilon_{t-j} \right)^2 + \alpha_3 \left(\sum_{j=1}^{20} \varepsilon_{t-j} \right)^2 + \alpha_4 \left(\sum_{j=1}^{60} \varepsilon_{t-j} \right)^2 .$$

This specification allows for influences on the conditional volatility at time t of 1-day returns, one week returns, one-month returns, and three-month returns. The second column in Table 3.5 presents the estimation results. For a value of d=0.1 the average HARCH persistence measure is 0.1437, which is very low. This is consistent with the data generating process since a low value for the fractional integration parameter was used. However, for the other two cases considered we reach a different conclusion: the persistence levels are again very low (0.1458 for the first scenario and 0.1526 for the second scenario considered) and are not able to capture the long memory property found in the data. We would have expected the HARCH model to capture a high persistence value in the case of $d = 0.499$. The low persistence estimates for all three cases considered are determined by $\hat{\alpha}_1$, since the other parameters are estimated at values close to zero. Given the aggregation of returns in the conditional volatility equation we would have expected higher persistence values. We conclude that the HARCH(4) model is not able to capture ARFIMA dynamics.

Table 3.5: Descriptive statistics of HARCH estimates on ARFIMA data.

d		$\hat{w}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\alpha_3}$	$\hat{\alpha_4}$	$\hat{\phi}$
Scenario1							
0.1	mean	0.0280	0.1235	0.0029	0.0001	0.0001	0.1437
	st.dev.	0.0009	0.0254	0.0013	0.0001	0.0001	0.0206
Scenario2							
0.3	mean	0.0223	0.1263	0.0028	0.0001	0.0001	0.1458
	st.dev.	0.0009	0.0259	0.0013	0.0001	0.0001	0.0219
Scenario3							
0.499	mean	0.0002	0.1330	0.0028	0.0001	0.0001	0.1526
	st.dev.	0.0000	0.0200	0.0011	0.0001	0.0001	0.0150

3.3.3 Optimal Lag

The goal of this section is to address the problem of the optimal lag structure in the context of the HARCH model described in Section 3.2. The simulations in the previous section were very useful in understanding how the correlation structure from the HARCH model relates to the one from an ARCH data generating process, but in order to fully understand the relationships between the models we need to identify the optimal lag structure in the context of these models. In the simulations presented in the previous section we specified a lag structure each time we estimated the HARCH model. Now we relax this assumption and treat the maximum number of lags L allowed as fixed and determine the N relevant lags to be included by examining all combinations of possible lags. The number of models to estimate in this case is $\binom{L}{N}$. If we always include lag one, the number of models to estimate is reduced to $\binom{L}{N-1}$. This, however, demands significant computing power. Because of the

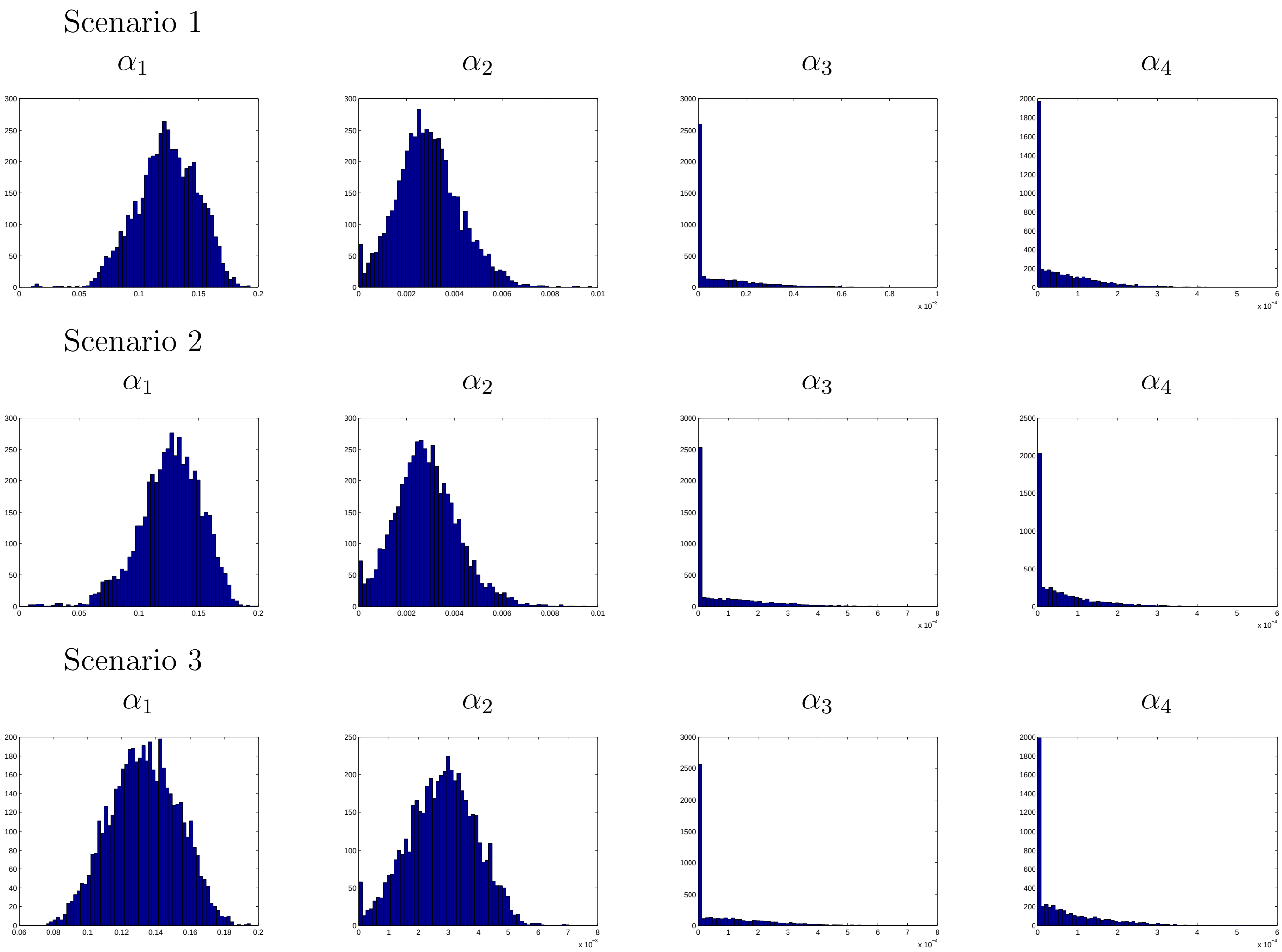


Figure 3.6: Histograms of the HARCH parameter estimates when the DGP is an ARFIMA(0,d,0) process constructed according to Table 3.5.

computational intensity we choose a parallel computing framework. Since the estimation processes are independent of each other, the problem is "trivially" parallelizable.

Each model specification is estimated by maximum likelihood. Thus, we estimate $\binom{L}{N-1}$ different models. We search for the best model specification according to the maximum of the log-likelihood function. The estimation algorithm delivers the optimal lag structure, parameter estimates, and the associated persistence measure. Once we identify the optimal lag structure for each model considered we can make inferences about the correlation structure. In particular, we are interested in how the HARCH model captures the data generating correlation structure by either the associated persistence measure or by the optimal lag.

The implementation was done in C++ and set up on Louisiana State University's supercomputing framework. On 20 nodes and 4 processors on each node, which provide a computing power of 851.4 Gflops/second, the determination of the optimal lag structure for a 1,000 samples of 5,000 observations each and with $L = 60$ and $N = 2$ took twenty minutes for each data generating process.

Calibration

We start by calibrating the procedure. Therefore, we consider as the data-generating process the following HARCH(2) model specification:

$$h_t = 6\text{e-}5 + 0.2\varepsilon_{t-1}^2 + 0.15\left(\sum_{j=1}^{5}\varepsilon_{t-j}\right)^2. \tag{3.21}$$

We generate samples of 5000 observations and repeat this process 1000 times. On each run, we estimate a HARCH(2) model with $L = 60$ and $N = 2$. We always include lag 1, so that we have to estimate $\binom{60}{1} = 3,540$ model specifications. The average estimated optimal lag over the 1000 runs was 5 for the second lag, which is very close to 5. The average estimates for the two parameters of interest are 0.23 for α_1 and 0.148 for α_2. The average persistence measure for the HARCH model is 0.97. Therefore we conclude that we are able to retrieve both the data generating lag structure (the first lag is always lag one) and the parameter estimates. Histograms for the two parameters are presented in Figure 3.9.

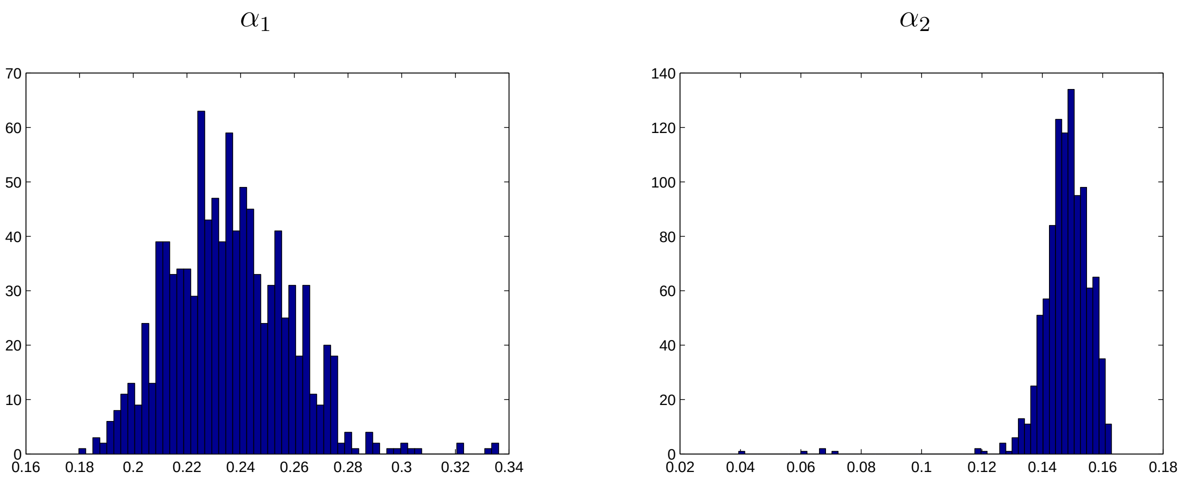


Figure 3.7: Histograms for calibration parameters.

Next we explore how the HARCH model captures different correlation structures by generating samples from ARCH, GARCH, IGARCH, and ARFIMA models. For each data-generating process considered, we simulate 1000 samples of length $T = 5000$ points each. For each sample, we estimate a HARCH(2) specification. The estimation procedure is searching for the lags that result in the highest log-likelihood value.

ARCH

We employ an ARCH(2) model with different values for the parameters in the conditional variance equation. We start by employing a mean return of zero and conditional variance process $h_t = 6\text{e-}5 + 0.15\varepsilon_{t-1}^2 + 0.83\varepsilon_{t-2}^2$ as the data-generating process. The sum of the autoregressive parameters is 0.98, implying a reversion to the mean of $1/(1-0.98) = 50$ units of time. The value of 0.98 indicates relatively high persistence of shocks. Next we explore other scenarios with different levels of persistence and different levels of mean time reversion. The different data generating processes considered are summarized in Table 3.6.

For each model specification presented in Table 3.6, we estimate a HARCH model with $L = 60$ and $N = 2$ and lag 1 fixed on the simulated data. The average values for the lags and parameter estimates can be found in Table 3.6.

The estimated persistence values confirm our findings from the previous section where we estimated a HARCH(2) model on an ARCH(2) data generating process. We confirm that in terms of capturing the persistence from the data generating process, the HARCH model performs better for high persistence values. However, what we can show here is that the model captures the influence of the second lag used in the data generating process as well. On all three scenarios the optimal lag was estimated as being 2 which corresponds to the second lag used in the data generating process. We conclude that the HARCH model captures the direct influence of the lags used in the data generating process. The histograms for the two parameter estimates and the corresponding persistence measure are presented in Figure 3.8.

Table 3.6: Descriptive statistics of the estimates from HARCH estimation on ARCH(2) data.

w	α_1	α_2	ϕ	$\frac{1}{1-\phi}$		$\hat{Lag}$	$\hat{w}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\phi}$
	Scenario1									
6e-5	0.15	0.83	0.98	50	mean	2	0.0001	0.4436	0.2771	0.9977
					st.dev.	0	0	0.0338	0.0158	0.0119
	Scenario2									
6e-5	0.10	0.65	0.75	4	mean	2	0.0001	0.1982	0.2572	0.7129
					st.dev.	0.0316	0	0.034	0.0226	0.0634
	Scenario3									
6e-5	0.10	0.35	0.45	2	mean	2	0.0001	0.1474	0.1042	0.3571
					st.dev.	0.182	0	0.0243	0.016	0.0338

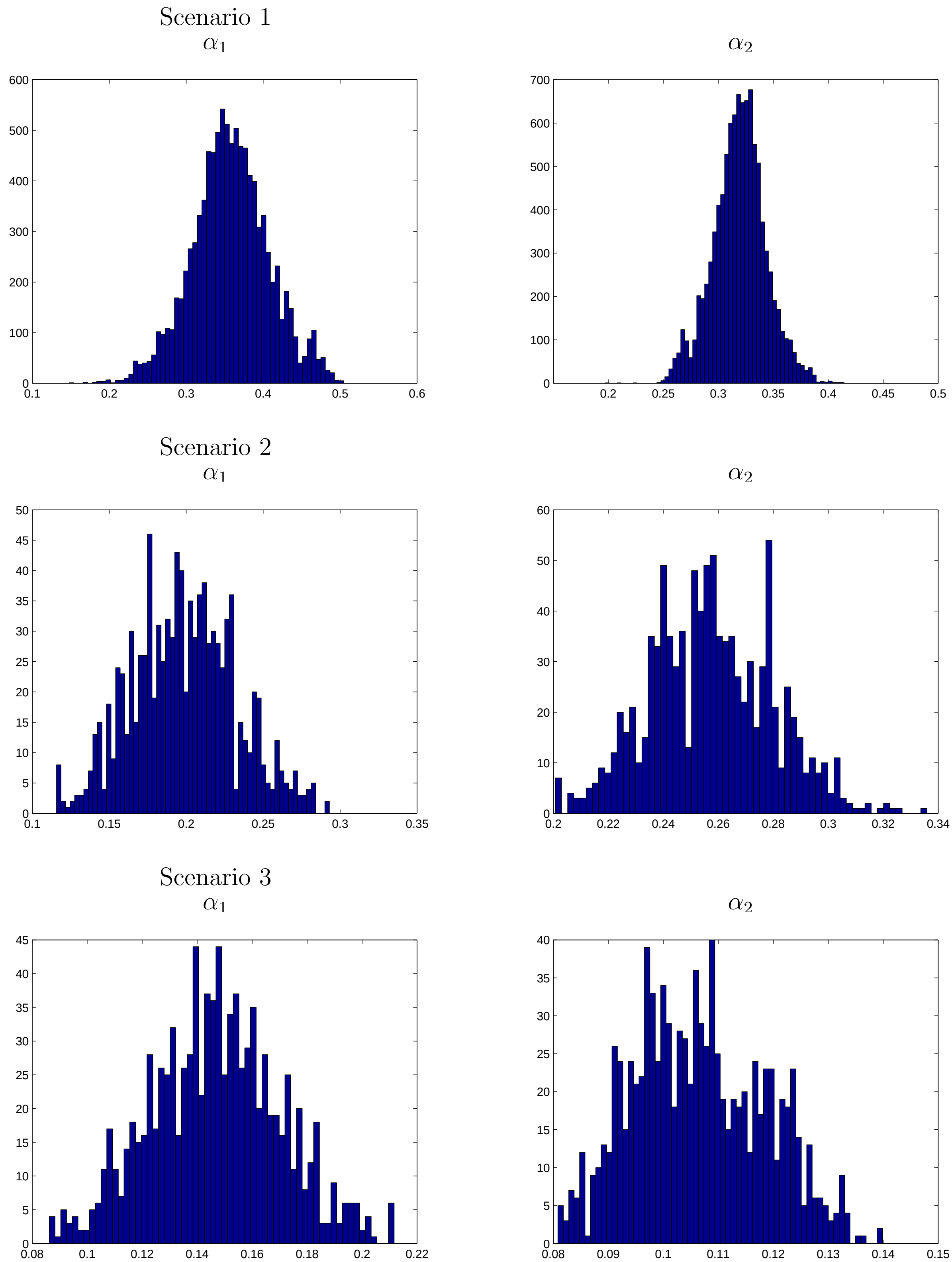


Figure 3.8: Histograms of the HARCH parameter estimates and optimal lag when the DGP is an ARCH(2) process constructed according to Table 3.6.

GARCH(1,1)

Let the data-generating process be a GARCH(1,1) model with $h_t = 6\text{e-}5 + 0.18\varepsilon_{t-1}^2 + 0.80h_{t-1}$. The sum of the autoregressive coefficients is 0.98, indicating high persistence. The average time of the process to return to the unconditional mean is $1/(1-0.98) = 50$ units of time. We consider two alternative scenarios with a level of persistence of 0.75 and one with a level of persistence of 0.35.

When estimating HARCH on these data, for the first scenario the average optimal lag is 7. The persistence measure is 0.63 which is very low when compared to the 0.99 value used in the DGP. For scenarios 2 and 3 we get similar results, the optimal lags are estimated to be 4 and 16, and the corresponding persistence measures are 0.22 and 0.16. In neither of the two cases the HARCH specification was able to capture the data generating correlation structure. Interestingly, the optimal lag is found to be 16 for the scenario that assumes a low level of persistence of 0.35. We conclude that the HARCH model is not able to capture GARCH dynamics. One explanation for this phenomenon could be that since the two models are not nested, the HARCH model is not able to capture the correlation structure from the data generating process.

Table 3.7: Descriptive statistics of HARCH estimates on a GARCH(1,1) data generating process.

w	α_1	α_2	ϕ	$\frac{1}{1-\phi}$		$\hat{Lag}$	$\hat{w}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\phi}$
	Scenario1									
6e-5	0.15	0.83	0.98	50	mean	7	0.0011	0.3504	0.0574	0.6357
					st.dev.	5.0574	0.0001	0.055	0.0363	0.0985
	Scenario2									
6e-5	0.10	0.65	0.75	4	mean	4	0.0002	0.1472	0.0286	0.225
					st.dev.	5.4288	0	0.0223	0.014	0.0255
	Scenario3									
6e-5	0.10	0.25	0.45	2	mean	16	0.0001	0.1457	0.0074	0.1641
					st.dev.	17.1515	0	0.021	0.0093	0.0291

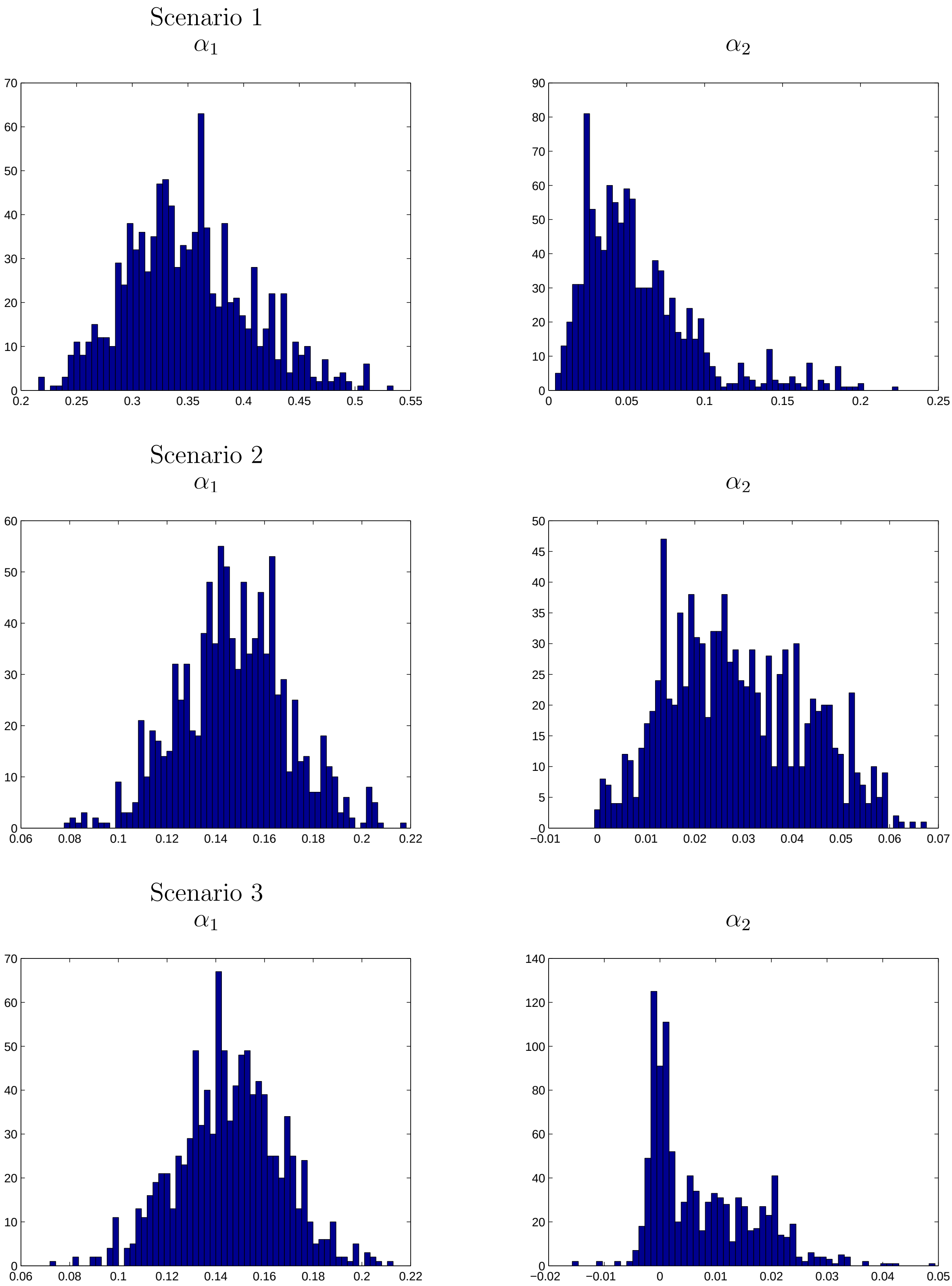


Figure 3.9: Histograms of the HARCH parameter estimates when the DGP is a GARCH(1,1) process constructed according to Table 3.7.

Data generating processes with change points

We wish to explore how the HARCH model captures correlation structures generated from data with structural breaks. For this purpose we use as data generating process a GARCH(1,1) model with a switch in the constant. Let:

$$h_t = \begin{cases} \omega_1 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} & \text{if } t = 1, \cdots, T_1 \\ \omega_2 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} & \text{if } t = T_1 + 1, \cdots, T \end{cases}$$

Let $\alpha_1 = 0.10$ and $\beta_1 = 0.50$. For the first 3000 observations we set $\omega_1 = 5e - 3$. We assume a parameter change in the constant term that occurs at observation 3001 and change ω to $\omega_2 = 0.0125$. Estimating a HARCH specification and ignoring the break, the average optimal lag is 19. The average values for the two parameter estimates are .169 for α_1 and 0.0171 for α_2. The persistence measure associated with the two parameter estimates is 0.277, which is very low. However, the optimal lag is obtained at a value of 19 indicating that the HARCH model requires a large number of lags in order to capture long memory. From this experiment we can conclude that the serial correlation found in the data generating process is not captured by the persistence measure, but by the high value of the lag.

IGARCH(1,1)

In the IGARCH model of Engle and Bollerslev (1986), the coefficients are estimated by maximizing the likelihood function subject to the constraint that the sum of the autoregressive coefficients is one. For an IGARCH(1,1) model this translates in imposing the restriction that $\alpha + \beta = 1$. In this model, shocks to the volatility have an indefinite memory and do not die out over time.

As data generating process we consider the following IGARCH(1,1) model for the conditional variance process:

$$h_t = 6\text{e-}5 + 0.05\varepsilon_{t-1}^2 + 0.95h_{t-1}.$$

Table 3.8: Results from estimating HARCH on IGARCH data.

	lag	$\hat{w}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\phi}$
mean	26	0.0278	0.3288	0.0240	0.6819
std.dev	17.7211	0.0223	0.1108	0.0226	0.2190

Table 3.8 reports the estimates from this simulation. The estimated optimal lag is found at a value of 26. The parameter estimates translate into an average persistence measure of 0.68 which corresponds to a correlation structure of approximately 4 days. This value is very low and does not capture the features of the data generating process. It is interesting to mention that again the optimal lag was found at a high value of 26. We conclude that the serial correlation structure of the data generating process is not captured by the HARCH persistence measure, but by the high value of the estimated lag.

ARFIMA

We also want to explore the type of lags and persistence captured by the HARCH model when applied to long memory data generated by ARFIMA. We consider an ARFIMA(0,0.45,0) process: $(1-L)^{0.45}y_t = \varepsilon_t$, where L is the lag operator and ε_t is white noise with zero mean and constant variance. We set $d = 0.45$ which indicates stationary long memory. We generate 1000 samples of 6000 observations each based on this specification and estimate a HARCH specification with $L = 60$ and $N = 2$ on each sample. We always include lag 1.

The results are summarized in Table 3.3.3. Both the average optimal lag and the estimated persistence measure are very low. The average optimal lag over the 1000 sample paths is 6 and the average persistence is 0.27, which indicates a low level of persistence and corresponds to a correlation structure of approximately 2 days. We can conclude from these results that the HARCH (2) model with a variable second lag is not able to capture the features of ARFIMA data by neither the persistence measure, nor the lag value.

Table 3.9: Results from estimating HARCH on ARFIMA data.

	lag	$\hat{w}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\phi}$
mean	6	0.9538	0.2906	0.0172	0.3250
std.dev	2.2445	0.0344	0.0400	0.0112	0.0276

3.4 Conclusion

This chapter provides an analysis of the HARCH model correlation structure. We study how the HARCH correlation structure relates to the correlation structure from different models proposed in the literature, with emphasis on models that are able to capture the high persistence phenomenon found in financial data. Therefore we estimate HARCH models on data generated by these models. We find that HARCH is only able to capture correlation structures obtained from ARCH data. However, our findings indicate that HARCH does not seem to be able to pick up correlation scales from ARCH data in the mean reverting sense, but it captures the influence of the lags used in the data generating process. When compared to GARCH models, our simulations show that the HARCH model is not able to capture the data correlation structure for such models. We also investigate the type of correlations found when in the presence of long memory models and models that approximate long memory by aggregation. We find that although the optimal lags are estimated at higher values for these data, the estimated persistence measure for the HARCH model does not correspond to the data-generating persistence.

Chapter 4

The Impact of Storms on Firm Survival: A Bayesian Spatial Econometric Model for Firm Survival

4.1 Introduction

In the form of Hurricanes Katrina and Rita, New Orleans and the gulf coast faced perhaps the most devastating natural disasters in the history of the United States. The disasters left policy makers with difficult questions not addressed in the academic literature. Fortunately, the disaster also left researchers with empirical data from a natural experiment of epic proportions.

This study addresses one key policy question, the determinant of business survival and recovery in the aftermath of a large scale natural disaster. According to the White House, [1] the Federal Government has provided over $ 114 billion in resources ($ 127 billion including tax relief) to the Gulf States to assist in rebuilding. State and Federal government officials faced the challenge of quickly implementing programs to minimize business failures and aide in the recovery process. Much of the academic literature focuses on business survival under

[1] www.whitehouse.gov/infocus/katrina

normal operating conditions.

One body of literature is based on the theoretical model developed by Jovanovic (1982) which predicts a positive relationship between firm survival and firm age. The implications predicted by this model have been tested empirically by several authors. Dunne, Roberts, and Samuleson (1989) use The Census of Manufacturers dataset to study survival rates for 219,754 plants from the manufacturing industry and find that survival increases with age and size. Audretsch (1991) finds the same relationship between firm survival, firm size and age by analyzing survival rates for 11,000 firms across different manufacturing industries using the U.S. Small Business Data Base. The study also finds that differences in survival rates are due to differences in technological regimes and industry specific characteristics such as scale economies and capital intensity. The aggregation to the industry level is motivated by data limitations. Audretsch and Mahmood (1995) address this problem and extend the analysis by allowing firm specific characteristics to influence survival rates. Using a dataset compiled by the U.S. Small Business Administration, the authors estimate a hazard duration model for 12,251 firms in the manufacturing sector and find that survival rates depend not only on industry specific characteristics such as technological conditions and scale economies, but also on establishment specific characteristics. The establishment specific characteristics identified are ownership structure and size. The study also confirms the positive relationship between firm survival and firm size and age. Caves (1998), Sutton (1997), and Geroski (1995) present ample surveys of the relevant literature and offer a summary of the main stylized facts. For other countries similar findings are found for: Canada (Baldwin and Gorecki 1991, Baldwin 1995, Baldwin and Rafiquzzaman 1995), Portugal (Mata, Portugal, and Guimaraes 1995, Mata and Portugal 1994), and Germany (Wagner 1994).

The second body of literature has evolved in the direction of analyzing firm survival at the product market level. A novelty of these studies is that firm survival is analyzed in the context of an evolutionary product market. The idea was first introduced by Gort and Kleppe (1982) who identify five stages of product life cycle based on net entry in the market. The authors conclude that firm survival is determined by technological changes as the market evolves over the life cycle of the products. Argawal (1996) and Agarwal and Gort (1996) use the Thomas Register of American Manufacturers database and analyze firm survival in the product life cycle framework. Argawal (1997) follows the same framework and considers the influence on firm survival of both firm specific characteristics and market product characteristics. The common finding across these studies is that the probability of survival changes across different stages of product life cycle development. Agarwal and Audretsch (2001) use the Thomas Register of American Manufacturers and analyze the relationship between firm survival and size in the context of the product life cycle framework. The study finds that while there is a positive relationship between size and survival in the early stages of development of the market, this relationship is no longer true for later stages of development. Agarwal and Gort (2002) conduct an analysis on firm survival by grouping the data according to the different stages of the product life cycle. The authors separate the different impacts on firm survival in industry specific life cycle factors and firm specific life cycle ones and take into account the effect of the two on each other. Their findings confirm the importance of both product and firm life cycle in determining firm survival.

Both of these strains of literature provide general guidance for our study, but do not specifically address the issue of business survival in a large scale disaster. One exception is Dahlhamer and Tierney (1997), who investigate the impact of the Northridge earthquake on 1,110 Los Angeles firms. Dahlhamer and Tierney find that the key factors predicting business performance were business size, disruption of operations, earthquake shaking intensity, and utilization of post-disaster aid. Much of the other literature on economic consequences of disasters focuses on community level effects (Friesma, et. al. 1979, Rossi, et. al. 1983, Wright et. al. 1979). Another approach is case study or qualitative analysis. For example, Runyan's (2006) qualitative analysis of Katrina is based on face-to-face interviews of seventeen small business owners affected by the storm. Another related study is the street survey of businesses after hurricane Katrina conducted by Campanella (2007). On a dataset containing 651 businesses established before the storm hit and 56 new businesses over a period of 15 months, the author conducts weekly street surveys to assess the status of New Orleans businesses recovery. Although the study is mainly based on summary statistics, since the geographical area under investigation is identical to ours, this study is of particular interest for our research. The author finds that locally owned businesses opened faster than large chain stores and businesses offering luxury items opened faster than businesses offering necessity goods. Finally, businesses located in less flooded areas opened faster when compared to ones located in more heavily flooded areas.

This chapter is organized as follows. In the next section we describe the dataset used. Sections 4.3 and 4.4 describe the spatial probit model specification and the methodology used. Section 4.5 describes the non-spatial probit model. The results are presented in Section 4.6. Section 4.7 concludes.

4.2 Data

This chapter examines the impact of Hurricanes Katrina and Rita on firm survival in Orleans Parish, Louisiana. In particular, we focus on explaining firm survival for the whole parish and by industry. Hurricane Katrina was characterized as one of the deadliest hurricanes to make landfall in the United States. The most affected area was New Orleans, Louisiana, both in terms of loss of life and property destruction. The cause was the failure of the levee system resulting in flooding for most of the city and surrounding areas.

The data set used for this study spans the period of 2004Q3 to 2007Q3. The most basic unit of observation in our dataset is an establishment. An establishment is a particular firm situated at a single geographical location. Some establishments are independent, while other ones are linked to a parent firm in which case they are called reporting units. Because Hurricane Katrina hit on August 29, 2005, we estimate our model for each quarter following the 2005Q2 quarter when the hurricanes hit: 2005Q3, 2005Q4, 2006Q1, 2006Q2, 2006Q3, 2006Q4, 2007Q1, 2007Q2 and 2007Q3. For presentation purposes we present detailed results for 2005Q4, 2006Q2, and 2007Q2.

Just prior to the storm, a total of 9,592 firms reported employment or wages to the Louisiana Department of Labor in Orleans Parish in 2005Q2.[2] Following Terrell and Bilbo, this study considers firms as open if they reported either employment or wages in any month of a quarter to the Louisiana Department of Labor for unemployment insurance purposes. Out of the 9,592 employers open in 2005Q2, 8,171 had valid latitudes and longitudes that could be used to determine location and append elevation data. This results in a sample of 8,171 employers with detailed quarterly data from 2004 to 2007, including employment, wages, and location. Using GIS maps, we are able to append flood depths to this data. Table 4.1 reports the total number of firms open in our sample in each quarter by industry type in Orleans Parish. The loss in terms of employers for the 3 chosen quarters are: 3,208 employers (39.26%) for 2005Q4, 3,055 (37.39%) for 2006Q2, and 3,146 (38.50%) for 2007Q2.

The primary dependent variable for our study is a a binary variable assuming values of 1 or 0, depending on whether the firm is open or not in a particular quarter. In assigning this value for each firm we follow the methodology proposed in Terrell and Bilbo.[3] Louisiana firms are required by law to report employment and wage data to the Louisiana Department of Labor (LDOL). This data is reported on a quarterly basis and is the basis for the Quarterly Census of Employment and Wages (QCEW). Several issues must be addressed to assess whether businesses are open or closed. First, the LDOL removes a firm from the data only after that particular firm fails to file a report for seven consecutive quarters or requests

[2]Terrell and Bilbo, A Report on the Impact of Hurricanes Katrina and Rita on Louisiana Businesses: 2005Q2-2006Q4, found at www.bus.lsu.edu/ded

[3]www.bus.lsu.edu/ded

removal from the database. The standard BLS measure of number of employers is based on a count of the number of employers in the QCEW database. Typically, this provides a reasonable measure of the number of firms, and offers a potential advantage by not removing seasonal firms or those simply failing to report in a given quarter.

A second important issue is that some businesses report zero employment and wages, but are still considered open by the LDOL. For the purpose of our study these firms should be considered as not operating. Third, in some cases LDOL estimates the employment and wages for some firms that fail to report. These three issues might be unimportant for most purposes. However, in the wake of an event such as Hurricane Katrina, particularly when the goal is to determine the patterns of entry and exit, these issues are crucial. This study follows Terrell and Bilbo's method of using a very conservative measure to determine whether an employer is open. The methodology uses the fact that the QCEW data includes a variable describing the way in which the data was obtained (whether it was estimated or reported by the employer). Based on this variable, we define employers as open only if they report positive values for employment or wages in at least one month in a particular quarter.

The next task consists of defining explanatory variables. One obvious factor that may affect the probability of being open is the flood depth. To focus on areas where flooding is relatively easy to measure, this study is limited to the city limits of New Orleans or equivalently Orleans Parish, Louisiana. Within the city, there are two distinct geographic areas, the East Bank and West Bank. The West Bank levees held and thus the area experienced minimal flooding. The levees failed in the East Bank where the majority of businesses existed. As a result this area filled with water much like a bowl. Elevation of these employers is thus a reasonable predictor of flood damage. Based on this logic, a flood value of zero is assigned to all West Bank employers, while latitude and longitude is used to assign flood elevations of East Bank employers.

Table 4.1: **Firms by Industry and Quarter.** The first column displays the industry. Columns 1-13 report the number of of firms open for each quarter.

	04Q3	04Q4	05Q1	05Q2	05Q3	05Q4	06Q1	06Q2	06Q3	06Q4	07Q1	07Q2	07Q3
1. Agriculture, Forestry, Fishing & Hunting	9	10	10	10	10	9	9	8	9	9	8	9	9
2. Mining	36	35	39	40	33	33	31	32	28	29	29	27	26
3. Utilities	5	5	8	19	3	4	2	6	10	8	7	7	2
4. Construction	269	275	277	298	199	194	207	216	212	226	222	220	214
5. Manufacturing	172	173	188	196	146	126	131	132	134	140	135	134	135
6. Wholesale Trade	328	336	342	356	292	275	261	257	243	253	240	239	238
7. Retail Trade	1,186	1,201	1,275	1,336	932	609	630	680	682	694	711	716	687
8. Transportation & Warehousing	195	207	212	222	178	167	164	168	171	162	153	154	146
9. Information	107	111	115	128	97	82	80	74	77	71	68	68	58
10. Finance & Insurance	396	414	425	466	356	305	304	317	284	312	298	291	282
11. Real Estate, Rental & Leasing	360	362	379	402	295	249	237	231	232	232	229	226	211
12. Professional, Scientific & Technical Services	1,125	1,161	1,195	1,264	970	943	932	959	940	968	921	919	905
13. Management of Companies & Enterprises	29	30	32	37	23	25	21	21	19	20	19	21	16
14. Administrative, Support, Waste Management & Remediation	344	352	371	389	307	260	262	274	267	272	259	251	245
15. Educational Services	78	81	84	91	72	59	66	60	61	64	65	65	65
16. Health Care & Social Assistance	753	786	800	838	576	487	445	462	454	461	456	465	444
17. Arts, Entertainment & Recreation	134	139	138	149	113	100	99	96	95	96	93	97	92
18. Accommodation & Food Services	848	880	926	981	726	559	566	583	571	586	595	585	557
19. Other Services	708	732	748	784	526	392	426	450	437	452	443	453	440
20. Public Administration	139	141	155	165	108	85	86	90	83	83	81	78	79
Total	7,221	7,431	7,719	8,171	5,962	4,963	4,959	5,116	5,009	5,138	5,032	5,025	4,851

More specifically, a second data set of Orleans Parish elevations was obtained from the Louisiana CADGIS Laboratory. This data set consists of something called LIDAR Edited Points – a massive data set of three dimensional points: latitude, longitude, and elevation. These points are considered to be "edited" points which means that ground obstructions such as vegetation foliage, man-made structures, etc. have been removed. The data set is intended only to contain land elevations. The LIDAR and QCEW data sets were combined using a GIS software package (ESRI's ArcView 9.2) and each employer was assigned the elevation of the point nearest to it from the LIDAR edited points data. This provides elevation to $8,171$ firms in Orleans Parish. The elevation is measured in feet relative to the sea level. The elevation variable was then used to calculate flood depth for all firms in our sample. As previously stated, West Bank employers were assigned a flooding variable of zero, while East Bank employer's flooding can be measured based on the elevation of the firm. The average flooding in New Orleans was roughly two feet above sea level. Therefore the flood depth was calculated as two minus the elevation variable. Terrell, Bilbo, and Lam (2007) conduct a study in order to determine how accurate the measure of flood depth based on elevation is in determining whether businesses were flooded or not. The results are based on a phone survey of 1,833 Orleans Parish businesses. Each business was asked if they were flooded or not. Then the authors compare the results based on the phone survey to the results based on the elevation measure. Their findings confirm that we have a good measure for flood depth.

We expect heavily flooded establishments to reopen more slowly than the less flooded ones. In order to test this hypothesis we construct a categorical variable capturing the feet of water as following: no flood, between 0 and 2 feet of water, between 2 and 4 feet of water, between 4 and 6 feet of water, between 6 and 8 feet of water, and finally above 8 feet of water.

One of the main contributions of this chapter is the analysis of the spatial interactions between firms. We allow for a firm's decision to reopen to be influenced by the decision to reopen of nearby firms. Therefore we need information on neighboring firms. The latitude and longitude data was used to identify the nearest neighbors for each firm in our sample. Based on this we construct a 8,171× 8,171 spatial weight matrix (W) for every combination of firms in our dataset. We rely on a spatial contiguity relationship between firms in constructing the matrix W. Therefore the weight matrix reflects the spatial relationship between firms and is constructed such that each element w_{ij} of the matrix is assigned a value of 1 if firm j and firm i have a contiguity relationship and 0 in the absence of such a relationship. When we use the term contiguity relationship we follow the spatial literature and refer to the fact that firms i and j have a common border and therefore are considered neighbors. The diagonal elements were all set to zero. Next we row standardize the matrix by dividing each element w_{ij} in the matrix by the row sum such that all rows sum to one. The row standardization does not change the relative spatial dependency among observations. By dividing each element of the matrix by the row sum we implicitly assume that the decision of reopening for each firm is a weighted average of the same decision of nearby firms and that all nearby firms are assigned the same weight. Other more complicated weighting schemes are possible, depending on how one wishes to quantify the degree of contiguity between firms. For the purpose of this chapter we simply want to account for spatial effects in the reopening decision, therefore any type of spatial dependency is acceptable.

Before proceeding any further we want to provide the reader with some intuition regarding the importance of the spatial weight matrix. A related concept in spatial econometrics is the spatial lag concept. While the first order contiguity matrix W provides information about each firm's neighbors, the spatial lag matrix provides information about the neighbors of neighbors. For the purpose of this study, this concept is very important since by using spatial lags the initial impact of neighbors on the decision to reopen propagates through space and has an impact on the decision of reopening of neighbors of neighbors.

The size of each establishment is another factor affecting the probability of reopening. We construct 4 categories based on the average employment across the three months of that quarter: size1 includes firms with average employment between 1 and 4 employees; size2 between 5 and 49 employees; size3 between 50 and 249 employees, and size4 includes firms with more than 250 employees.

The relative size of the establishment is also a factor that could affect the reopening decision. The variable relative size is calculated to make a distinction between locally owned businesses and chain stores. This hypothesis was also tested by Campanella (2007) who finds that locally owned businesses are reopening sooner than large chain stores. We calculate the variable relative size for each quarter by dividing the average employment across the three months of that quarter for each establishment by the the sum of average employment across all Louisiana establishments with the same reporting unit. Therefore a value close to one implies that we are looking at a locally owned business, while a value close to zero indicates a chain store. We also construct interactions between this variable and flooding variables (rel size&flood).

The type of industry is also expected to affect the firm reopening decision. We expect establishments in certain industries to open faster than in other ones. For this purpose we construct dummy variables for each of the 20 business categories presented in Table 4.1.

Summary statistics for all these variables are presented in Table 4.2.

4.3 Spatial Probit Model Specification

This section focuses on the statistical model for whether an establishment is open conditional on that establishment's characteristics. As previously stated, an "establishment" denotes a single location for an employer. We use a modified version of the spatial probit model introduced by Smith and LeSage (2002). We model the establishment's decision to

Table 4.2: **Descriptive statistics.** The first column displays the variable symbol. Column 2 reports the number of observations. Column 3 and 4 report the mean and standard deviation. The last two columns present the min and max values.

	Obs	Mean	Std. Dev.	Min	Max
open 2005 Q4	8171	0.607	0.488	0	1
open 2006 Q2	8171	0.626	0.484	0	1
open 2007 Q2	8171	0.615	0.487	0	1
rel size	8171	0.890	0.298	0.0001	1
rel size&flood	8171	0.409	0.487	0.0000	1
size1	8171	0.486	0.500	0	1
size2	8171	0.435	0.496	0	1
size3	8171	0.065	0.247	0	1
size4	8171	0.011	0.106	0	1
Ind1	8171	0.001	0.035	0	1
Ind2	8171	0.005	0.070	0	1
Ind3	8171	0.002	0.048	0	1
Ind4	8171	0.036	0.187	0	1
Ind5	8171	0.024	0.153	0	1
Ind6	8171	0.044	0.204	0	1
Ind7	8171	0.164	0.370	0	1
Ind8	8171	0.027	0.163	0	1
Ind9	8171	0.016	0.124	0	1
Ind10	8171	0.057	0.232	0	1
Ind11	8171	0.049	0.216	0	1
Ind12	8171	0.155	0.362	0	1
Ind13	8171	0.005	0.067	0	1
Ind14	8171	0.048	0.213	0	1
Ind15	8171	0.011	0.105	0	1
Ind16	8171	0.103	0.303	0	1
Ind17	8171	0.018	0.134	0	1
Ind18	8171	0.120	0.325	0	1
Ind19	8171	0.096	0.295	0	1
Ind20	8171	0.020	0.140	0	1
flood 0-2	8171	0.152	0.359	0	1
flood 2-4	8171	0.109	0.311	0	1
flood 4-6	8171	0.101	0.302	0	1
flood 6-8	8171	0.044	0.204	0	1
flood 8	8171	0.067	0.251	0	1

stay in business or not as a function of temporally and spatially varying observable and unobservable factors. The goal is to characterize the probability that an establishment is open in a given time period.

We start by introducing the main assumptions in the model and the notation that will be used for the rest of the chapter. Let m be the number of individual establishments. Each establishment is confronted in each period with choosing among two alternatives, labeled as 0 for closed and 1 for open. For each establishment we observe whether the firm is open or closed and model it as the realization of a random variable y_i. The decision to open after the storm ranges from consideration of profits of one store in a large chain by a manager of a fortune 500 company to a sole proprietorship's decision to reopen. Economic theory suggests that the decision to reopen is primarily made to maximize the discounted value of future profits.[4] However, the decision to open may be the same as the decision to return to the city for some proprietors who rely on business income as their primary source of funds. For ease of exposition, assume that the choice of whether to be open or closed is the result of an entrepreneur's decision to maximize their utility. An event will occur with a certain probability p if the utility derived from choosing that alternative is greater than the utility from the other alternative. Let z_i be the difference in utility from alternatives 1 and 0. The difference in utility is modeled as:

$$z_i = x_i\beta + \theta_i + \epsilon_i. \tag{4.1}$$

where $i = 1 \ldots m$, x_i is a vector of observed establishment specific attributes, β is a vector of unobserved parameters to be estimated, θ_i is an unobserved random effect component, and ϵ_i is the stochastic error term with $\epsilon_i \sim N(0,1)$. We do not observe z_i, but only observe

[4]We address differences in behavior across the ownership class variable relative size (see discussion in the data section) measuring employment at this establishment as a ratio of total employment at this location to that of all establishments under the same ownership in Louisiana.

the sign of z_i. We observe the establishment choice y_i being equal to 1 or 0, depending on whether z_i has a positive sign indicating the higher utility from this alternative or a negative sign associated with the lower utility associated with this alternative. Therefore we observe:

$$y_i = \begin{cases} 1 & \text{if } z_i > 0; \\ 0 & \text{if } z_i \leq 0. \end{cases} \tag{4.2}$$

The probability of choosing alternative 1 is given by:

$$P_i = P(y_i = 1) = P(z_i > 0). \tag{4.3}$$

The distinction between this model and the standard probit model is the term θ_i. The unobserved component θ_i is constructed such that it allows for spatial correlation across establishments. In other words we assume that differences in utilities are similar for neighboring establishments. This is obtained by specifying θ_i according to a spatial autoregressive structure:

$$\theta_i = \rho \sum_{j=1}^{m} w_{ij}\theta_j + u_i. \tag{4.4}$$

with $u_i \sim N(0, \sigma^2)$, $W = (w_{ij} : i, j = 1 \dots m)$ is a row standardized spatial weight matrix such that $\sum_{j=1}^{m} w_{ij} = 1$. ρ can be interpreted as the degree of spacial dependence across establishments. The spatial autocorrelation is thus determined by both ρ and W. We can write equation (4.4) in matrix notation:

$$\theta = \rho W \theta + u. \tag{4.5}$$

where $u \sim N(0, \sigma^2 I_m)$ and I_m is the identity matrix.

Let $B_\rho = I_m - \rho W$. We can obtain a solution for θ using (4.5):

$$\theta = B_\rho^{-1} u. \tag{4.6}$$

Note that the matrix B_ρ^{-1} plays a role similar to a lag polynomial in time series econometrics. This matrix captures the fact that spatial shocks (u) affect neighbors in space in much the same way that time series shocks affect observations close in time. Given our weight matrix, a shock to one firm has a first order impact of ρ on contiguous establishments, ρ^2 on establishments contiguous to those establishments, and so forth.

From (4.6) we see that the distribution for θ is given by:

$$\theta|(\rho, \sigma^2) \sim N(0, \sigma^2(B_\rho^{'} B_\rho)^{-1}). \tag{4.7}$$

The error term ϵ is assumed to be conditionally independent of the spatial unobserved component such that $\epsilon|\theta \sim N(0, \sigma_\epsilon^2)$ and we assume $\sigma_\epsilon^2 = 1$.

The full model in matrix notation is given by:

$$Z = X\beta + \theta + \epsilon. \tag{4.8}$$

4.4 Bayesian Inference in the Spatial Probit Model Specification

Our statistical approach is a simplification of the LeSage and Smith (2002) model assuming a homoscedastic ϵ_i. Bayesian inference is preferred in this setting primarily because it is easier to implement than the EM algorithm suggested by McMillen (1992) for the analogous frequentist model. In addition, the Bayesian approach provides exact small sample inferences.

Prior distributions for the unknown parameters complete the statistical model. Following LeSage and Smith, we assume

$$\beta \sim N(c, T) \tag{4.9}$$

$$H_p = 1/\sigma^2 \sim \Gamma(\alpha, v) \tag{4.10}$$

$$\rho \sim U[(\lambda_{min}^{-1}, \lambda_{max}^{-1})] \tag{4.11}$$

Given the statistical model summarized in section 4.3, LeSage and Smith (2002) provide the full conditionals required to the model by Monte Carlo Markov Chain (MCMC) methods. The MCMC method arrives at the target distribution of the unknown parameters by sequentially sampling from a set of conditional distributions of the parameters. This is very useful since usually it is difficult to find an analytical result for the posterior densities. The MCMC method provides a sample from the posterior density and we can use this sample to draw inferences about the parameters of interest. Under mild regularity conditions satisfied in this application, these samples converge to sample from the posterior distribution.

The Bayesian framework uses the idea of a loss function. The loss function is a measure of the loss incurred when comparing the true value of the parameter with the estimated value. The Bayesian estimator is obtained by minimizing the loss function. Suppose that we are interested in estimating $g(\mu)$, where g is the function of interest. In order to obtain the estimate of g we minimize the expected value of the loss function. In the case of a quadratic loss function this is reduced to minimizing:

$$\int (g(\hat{\mu}) - g(\mu))^2 p(\mu|y) d\mu. \tag{4.12}$$

By differentiating (4.12) with respect to $g(\mu)$ and equating to zero we obtain:

$$\hat{g(\mu)} = \int g(\mu)p(\mu|y)d\mu. \tag{4.13}$$

Therefore the point estimator for $g(\mu)$ is the posterior mean $\hat{g(\mu)} = E[g(\mu)|y]$. Then for a sample of size N from the posterior distribution we can approximate the posterior mean by:

$$E(g(\mu)) = \frac{1}{N}\sum_{i=1}^{N} g(\mu_i) \overset{N\to\infty}{\to} \int g(\mu)p(\mu|y)d\mu. \tag{4.14}$$

Following the same approach we can approximate the posterior variance by:

$$Var(g(\mu)) = \frac{1}{N}\sum_{i=1}^{N}[g(\mu_i) - E(g(\mu))]^2. \tag{4.15}$$

The MCMC algorithm follows that of Smith and LeSage (2002) and primarily a Gibbs sampling approach. For clarity, the notation used in this chapter is identical to that introduced by Smith and LeSage (2002). The problem consists of constructing a sampling algorithm for the set of unknown parameters given by (β, ρ, σ^2). Implementing the MCMC method also requires data augmentation to sample θ and z.

Intuitively, one can see that conditional on θ and the latent variable z, the equation

$$z_i - \theta_i = x_i\beta + \epsilon_i \tag{4.16}$$

is simply a linear regression model.

Thus, the conditional posterior distribution of β is proportional to the multinormal density:

$$\beta \mid (\theta, \rho, \sigma^2, z, y) \sim N(A^{-1}b, A^{-1}) \tag{4.17}$$

where $A = X'X + T^{-1}$ and $b = X'(z - \theta) + T^{-1}c$.

The conditional distribution of θ also follows a normal distribution:

$$\theta \mid (\beta, \rho, \sigma^2, z, y) \sim N(A_0^{-1} b_0, A_0^{-1}) \tag{4.18}$$

where $A_0 = \sigma^{-2} B_\rho' B_\rho$ and $b_0 = z - X\beta$.

The conditional posterior distribution of σ^2 (or the related precision H_p) is related to a chi-squared distribution in the following way:

$$H_p = \frac{1}{\sigma^2} \mid (\beta, \theta, \rho, z, y) \sim \frac{\chi^2(m + 2\alpha)}{\theta' B_\rho' B_\rho \theta + 2v} \tag{4.19}$$

The conditional posterior distribution of ρ is given by:

$$\rho \mid (\beta, \theta, \sigma^2, z, y) \propto \mid B_\rho \mid \exp(-\frac{1}{2\sigma^2} \theta' B_\rho' B_\rho \theta) \tag{4.20}$$

where $\rho \in [\lambda_{min}^{-1}, \lambda_{max}^{-1}]$ and λ_{min} and λ_{max} are the minimum and maximum eigenvalues of W.

The distribution in (4.20) is non standard and therefore we cannot sample from it directly. One solution to this problem is to use a Metropolis-Hastings algorithm. Smith and LeSage (2002) suggest using univariate numerical integration rather than a Metropolis-Hastings algorithm in this setting. In particular, we use the properties of the inverted gamma distribution to integrate out the nuisance parameter σ^2. Then equation (4.20) can be written as:

$$\rho \mid (\beta, \theta, z, y) \propto \mid B_\rho \mid [m^{-1} \theta' B_\rho' B_\rho \theta]^{-m/2} \pi(\rho) \tag{4.21}$$

Before sampling from this posterior distribution for ρ we need to calculate the normalizing constant that transforms (4.21) in a proper density function that integrates to one. The normalizing constant can be found by integrating (4.21) over a grid of ρ values chosen from the interval $[\lambda_{min}^{-1}, \lambda_{max}^{-1}]$. The conditional posterior distribution for the grid of ρ values can be obtained by integrating the normalized density. The updated value for the unknown parameter ρ can be obtained by drawing from this distribution using the inversion method. In the estimation part of the chapter we will use this method for updating the values of ρ. For a comparison between this method and the M-H method see Smith and LeSage (2002).

Finally, we need a conditional posterior distribution for the latent variable z. This distribution is a truncated normal distribution where the truncation depends on the observed choice for each firm:

$$z_i \mid (\beta, \theta, \rho, \sigma^2, V, -z_i, y) \sim \begin{cases} TN_{(0,\infty)}(x_i'\beta + \theta_i, 1) & \text{if } y_i = 1; \\ TN_{(-\infty,0)}(x_i'\beta + \theta_i, 1) & \text{if } y_i = 0. \end{cases} \tag{4.22}$$

The Gibbs sampler is given by the following iterative process:

1. Set starting values for the parameters $\beta_0, \theta_0, \rho_0, \sigma_0^2$ and the latent variable z_0.
2. Sample $\beta_1 \mid (\theta_0, \rho_0, \sigma_0^2, z_0)$ from the multinormal distribution given by equation (4.17).
3. Sample $\theta_1 \mid (\beta_1, \rho_0, \sigma_0^2, z_0)$ from the multinormal distribution given by equation (4.18).
4. Sample $\sigma_1^2 \mid (\beta_1, \theta_1, \rho_0, z_0)$ using equation (4.19).
5. Sample $\rho_1 \mid (\beta_1, \theta_1, \sigma_1^2, z_0)$ using numerical integration to obtain the conditional distribution for ρ using equation (4.21).
6. Sample $z_1 \mid (\beta_1, \theta_1, \sigma_1^2)$ from the truncated normal distribution given by equation (4.22).
7. Return to the first step and iterate to generate the posterior sample. Discard the burn-in period of the sampler to avoid dependence on the starting values.

Before proceeding, it is useful to note that the full conditionals may differ substantially from the marginal densities for each of these parameters. For example, the fact that the conditional density of θ_i is mean zero does not imply that the posterior mean of θ_i is zero. In fact, we expect the posterior mean for parameter θ_i to differ substantially across firms to capture the impact of other open or closed businesses on the probability that firm i is open.

4.5 Non-Spatial Probit Model

4.5.1 Model Specification

The only difference between the probit spatial model specification presented in Section 4.3 and a standard frequentist probit model is the unobserved spatial component θ_i. Abstracting away from the spatial interactions between neighboring firms simplifies the model so that each firm's decision is a just a function of firm specific attributes. The random utility model described in Section 4.3 continues to be of interest in explaining each firm's decision to reopen. Specifically, a particular firm will decide to reopen if the utility from reopening is higher than the utility from staying out of business. Again, we only observe which one of the two alternatives was chosen, that is we observe the sign of the random variable and not the actual value it takes. As before, let y_i^* be the difference in utility from the two alternatives 1 and 0. The difference in utility is modeled as:

$$y_i^* = x_i\beta + \epsilon_i. \tag{4.23}$$

where $i = 1 \dots m$, x_i is a vector of observed establishment specific attributes, β is a vector of unknown parameters to be estimated, and $\epsilon_i \sim N(0, \sigma^2)$.

We do not observe y_i^*, but only observe the sign of y_i being equal to 1 or 0, depending on the sign of y_i^*:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0; \\ 0 & \text{if } y_i^* \leq 0. \end{cases} \tag{4.24}$$

The probability of choosing alternative 1 is given by:

$$P_i = P(y_i = 1) = P(y_i^* > 0). \tag{4.25}$$

From equation(4.23) we can re-write the probability of choosing alternative 1 as:

$$P_i = P(x_i\beta + \epsilon_i > 0) = P(\epsilon_i > -x_i\beta) = P(\epsilon_i \leq x_i\beta). \tag{4.26}$$

We can further write:

$$P_i = P(\frac{\epsilon_i}{\sigma} \leq \frac{x_i\beta}{\sigma}) = \Phi(\frac{x_i\beta}{\sigma}). \tag{4.27}$$

where Φ is the cumulative distribution function of the normal distribution.

Let $\sigma^2 = 1$. Then,

$$P_i = \Phi_i = \Phi(x_i\beta) = \int_{-\infty}^{x_i\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt. \tag{4.28}$$

4.5.2 Maximum Likelihood Estimation

The model presented in the previous section is usually estimated by maximum likelihood methods. The maximum likelihood estimator (MLE) of β is the vector that maximizes the likelihood function. By construction, the random variable y follows a Bernoulli distribution. Since $P_i = \Phi(x_i\beta)$ we can write the probability density function for y_i as:

$$f(y_i) = \Phi_i^{y_i}(1 - \Phi_i)^{1-y_i}. \tag{4.29}$$

where Φ_i is given by (4.28). For our sample of m independent observations, the likelihood function is given by:

$$L(\beta) = \prod_{i=1}^{m} \Phi_i^{y_i} (1 - \Phi_i)^{1-y_i}. \tag{4.30}$$

It is more convenient to work with the log likelihood function:

$$\ln L(\beta) = \sum_{i=1}^{m} y_i \ln \Phi_i + (1 - y_i) \ln(1 - \Phi_i). \tag{4.31}$$

The MLE estimator of β is obtained by maximizing equation (4.31) with respect to β. Because of nonlinearity in the first order conditions the estimates cannot be obtained directly. Therefore numerical optimization methods are used to obtain the MLE estimates. The results presented in this chapter were obtained in STATA using a Newton-Raphson algorithm. The maximum likelihood estimator is consistent and asymptotically efficient, with the following asymptotic distribution (Judge, Hill, Griffith, Lütkepohl, Lee 1985):

$$\hat{\beta}_{MLE} \sim N\left(\beta, -\left[\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} |_{\beta = \tilde{\beta}}\right]^{-1}\right). \tag{4.32}$$

where $\tilde{\beta}$ represents the final set of parameter estimates.

4.5.3 Bayesian Inference

The steps used in order to estimate the probit model from a Bayesian perspective are very similar to the ones described in Section 4.4 in the context of the spatial probit model. We start by specifying a prior distribution for our parameters of interest. We only need a prior for β because σ^2 is set to one. The prior distribution reflects the knowledge about the parameter of interest before looking at our sample. We set the prior for β as $\beta \sim N(0, \Sigma_p)$. We choose the following prior precision: $H_p = \Sigma_p^{-1} = 0.001 I_k$, where I_k is the identity matrix and $k = 30$ is the number of unknown parameters to estimate. The prior parameter precision is chosen to imply an uninformative prior.

After choosing the prior for β, the conditional posterior distribution of β is given by: Thus, the conditional posterior distribution of β is proportional to the multinormal density:

$$\beta \mid y^* \sim N(\bar{\beta}, \bar{H}) \tag{4.33}$$

where $\bar{\beta} = \bar{H}^{-1}[H_p\beta_p + x'x\hat{\beta}]$, $\bar{H} = H_p + x'x$, and $\hat{\beta} = (x'x)^{-1}x'y^*$.

The conditional posterior distribution for the latent variable y^* is given by a truncated normal where the truncation depends on the observed choice such that:

$$y_i^* \mid (\beta, y) \sim \begin{cases} TN_{(0,\infty)}(x_i'\beta, 1) & \text{if } y_i = 1; \\ TN_{(-\infty,0)}(x_i'\beta, 1) & \text{if } y_i = 0. \end{cases} \tag{4.34}$$

The Gibbs algorithm is then given by:

1. Set a starting value for y_0^*.

2. Sample $\beta_1 \mid y_0^*$ using equation (4.33).

3. Sample $y_i^* \mid \beta_1$ using equation (4.34).

4. Return to the first step and iterate to generate the posterior sample. Discard the burn-in period of the sampler to avoid dependence on the starting values.

4.6 Results

In this section we present results from the estimation of the models discussed in the previous sections. We start by presenting results from a frequentist non-spatial probit model obtained by maximum likelihood estimation. We compare these results to the ones obtained using a Bayesian framework. Therefore, we estimate both a non-spatial probit specification and a spatial probit specification using Bayesian techniques. For comparison purposes we present results for 2005Q6. Next we proceed to estimate the spatial probit specification presented in Section 4.3. Detailed results are presented for all three quarters of interest: 2005Q4, 2006Q2, and 2007Q2.

4.6.1 Comparison between Non-Spatial and Spatial Model

Our objective is to compare the maximum likelihood estimator of the non-spatial probit model to the non-spatial probit and spatial probit Bayesian results. The results are presented in Table 4.6.1 for 2005Q4. The table reports maximum likelihood estimates and corresponding standard errors in the first portion of the table. The table also reports the posterior means and standard deviations for the same coefficients but based on Bayesian inference for the non-spatial probit model. The MLE estimates and posterior means are very similar as one would expect given our priors and sample size. This comparison is included primarily to validate our results from both algorithms. The same result holds when comparing the standard errors with the posterior standard deviations and the confidence intervals with the highest posterior density region.[5] This result is not surprising since the MLE estimates and the posterior means should be asymptotically equivalent for the probit model and our choice of priors . Because our sample consists of 8,171 establishments, even with a somewhat informative prior the resulting posterior means would be very similar with the MLE estimates.

[5]Similar results for the case of probit models were found by Griffith, Hill, and O'Donnell(2006) and Ogunc (2002).

Table 4.3: **Estimation results for 2005Q4 using different models.** The first column displays the variable symbol. For each model, columns 1, 2, 3, and 4 report estimates, standard errors and 95% confidence Intervals for the MLE, and posterior means, posterior standard deviations and highest posterior density intervals for the other 2 models.

	Maximum Likelihood Estimation				Bayesian Probit				Spatial Bayesian Probit			
	estimate	Std. Err.	95%	CI	p.mean	p Std.Dev	2.5%	97.5%	p.mean	p Std.Dev	2.5%	97.5%
Intercept	0.596	0.172	0.260	0.933	0.603	0.174	0.259	0.940	0.826	0.238	0.353	1.286
rel size	0.139	0.077	-0.012	0.291	0.138	0.078	-0.016	0.290	0.238	0.115	0.007	0.463
rel size&flood	-0.147	0.099	-0.340	0.046	-0.145	0.099	-0.340	0.052	-0.247	0.146	-0.526	0.048
Size1	-0.733	0.131	-0.990	-0.475	-0.739	0.132	-0.999	-0.486	-1.040	0.187	-1.411	-0.693
Size2	-0.150	0.131	-0.408	0.107	-0.156	0.132	-0.415	0.099	-0.186	0.187	-0.553	0.166
Size3	0.366	0.145	0.082	0.651	0.366	0.147	0.075	0.655	0.551	0.203	0.145	0.938
Ind1	1.245	0.527	0.213	2.277	1.318	0.544	0.330	2.454	2.073	0.846	0.561	3.866
Ind2	0.851	0.265	0.331	1.371	0.865	0.268	0.345	1.404	1.212	0.366	0.505	1.942
Ind3	-0.882	0.346	-1.561	-0.204	-0.903	0.349	-1.597	-0.222	-1.289	0.515	-2.337	-0.312
Ind4	0.466	0.128	0.216	0.717	0.466	0.127	0.216	0.715	0.709	0.187	0.338	1.071
Ind5	0.289	0.140	0.015	0.563	0.288	0.138	0.019	0.557	0.467	0.202	0.058	0.864
Ind6	0.825	0.127	0.576	1.074	0.827	0.125	0.580	1.070	1.210	0.186	0.840	1.570
Ind7	-0.165	0.108	-0.376	0.047	-0.164	0.106	-0.375	0.045	-0.185	0.156	-0.488	0.121
Ind8	0.604	0.140	0.331	0.878	0.608	0.138	0.335	0.879	0.884	0.201	0.492	1.282
Ind9	0.357	0.154	0.056	0.659	0.358	0.152	0.065	0.662	0.523	0.219	0.101	0.955
Ind10	0.462	0.119	0.229	0.695	0.463	0.119	0.228	0.693	0.680	0.170	0.349	1.007
Ind11	0.352	0.121	0.116	0.589	0.354	0.120	0.116	0.590	0.554	0.177	0.207	0.910
Ind12	0.700	0.109	0.486	0.915	0.703	0.108	0.492	0.915	0.997	0.155	0.700	1.296
Ind13	0.274	0.236	-0.189	0.738	0.280	0.235	-0.184	0.751	0.395	0.343	-0.266	1.078
Ind14	0.350	0.122	0.110	0.589	0.352	0.122	0.111	0.597	0.525	0.175	0.186	0.866
Ind15	0.321	0.172	-0.016	0.657	0.329	0.170	-0.008	0.664	0.482	0.246	-0.002	0.961
Ind16	0.273	0.112	0.054	0.492	0.274	0.111	0.056	0.494	0.434	0.163	0.118	0.754
Ind17	0.330	0.149	0.038	0.622	0.334	0.149	0.042	0.619	0.500	0.214	0.080	0.922
Ind18	-0.123	0.110	-0.340	0.093	-0.124	0.110	-0.340	0.091	-0.153	0.157	-0.459	0.154
Ind19	0.024	0.112	-0.195	0.243	0.025	0.110	-0.194	0.241	0.047	0.159	-0.270	0.357
flood 0-2	-0.279	0.098	-0.470	-0.087	-0.280	0.098	-0.474	-0.085	-0.325	0.146	-0.622	-0.038
flood 2-4	-0.509	0.101	-0.706	-0.312	-0.513	0.101	-0.710	-0.314	-0.675	0.152	-0.968	-0.379
flood 4-6	-0.448	0.102	-0.647	-0.248	-0.450	0.101	-0.649	-0.249	-0.580	0.153	-0.876	-0.280
flood 6-8	-0.442	0.114	-0.665	-0.219	-0.446	0.113	-0.666	-0.219	-0.646	0.173	-0.997	-0.320
flood 8	-0.654	0.102	-0.855	-0.454	-0.659	0.102	-0.857	-0.456	-0.965	0.158	-1.286	-0.658
σ^2									1.031	0.069	0.896	1.169
ρ									0.454	0.109	0.202	0.633

When comparing the MLE estimates and the posterior means from the spatial model with the MLE estimates and posterior means from the non-spatial specification we see that all posterior means from the spatial specification are larger in absolute value. This is rather unexpected since we would expect to see larger magnitudes in the non-spatial specification since this specification ignores any potential spatial affects (see LeSage and Smith 2002). The posterior mean for the autocorrelation parameter ρ is 0.454 indicating the existence of spatial correlation between establishments.

One important property of this model is that it generates heteroskedasticity in the errors if the spatial component is omitted. In the context of a linear regression, ordinary least squares would still be consistent. In the context of the probit model, heteroscedasticity would translate into inconsistency of maximum likelihood estimates (Greene 2002). This fact may explain the discrepancy.

Sampling distributions of model parameters for the MLE and marginal densities from the posterior densities are presented in Figure 4.1 through 4.5.

4.6.2 Spatial Bayesian Results

This section discusses results from the spatial probit model specification developed in Section 4.3. Table 4.6.2 contains results for all three quarters. While the coefficients are informative in indicating the direction of change in probabilities, their magnitudes are not very informative. Therefore we also report marginal effects in Table 4.6.2. Additional reports for particular firms can be found in the Appendix.

Perhaps the most surprising finding is the relationship between the relative size variable and the probability of reopening. Campanella (2007) reports results from data gathered during bicycle tours over a 15 month period. One interesting result from Campanella's study was that locally owned businesses were more likely to reopen than large chain businesses.

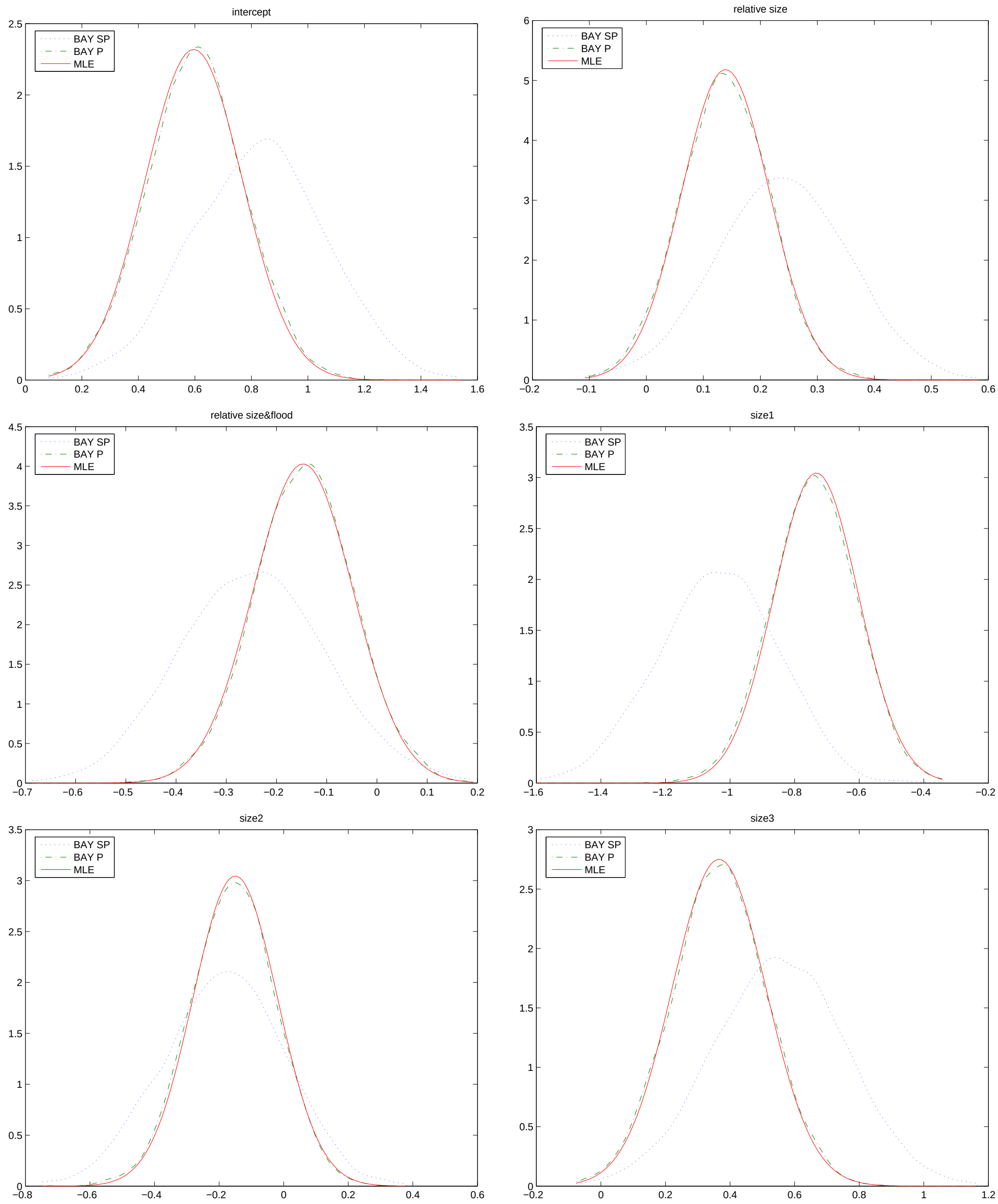


Figure 4.1: Sampling distributions for parameters for 2005Q4.

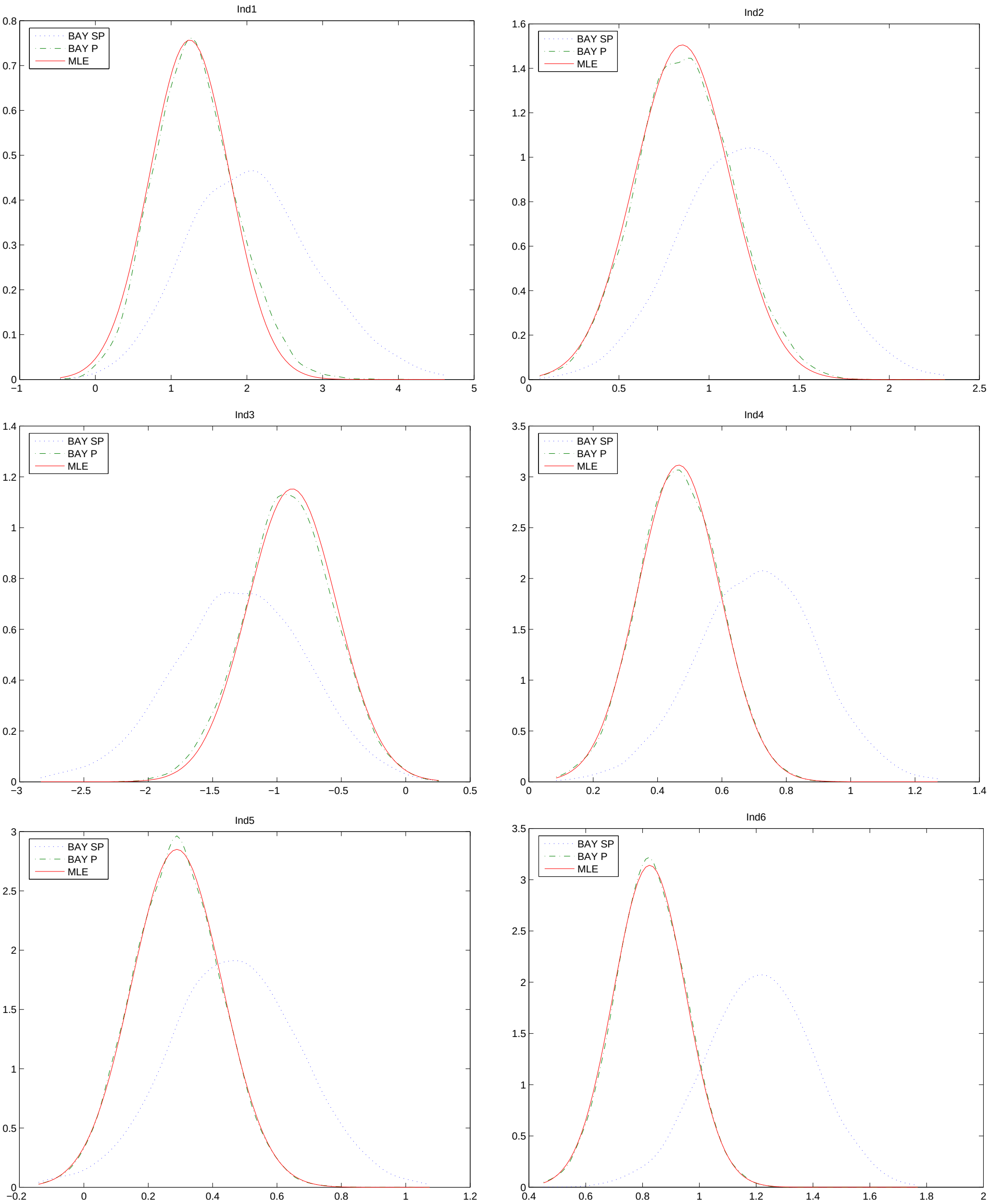


Figure 4.2: Sampling distributions for parameters for 2005Q4.

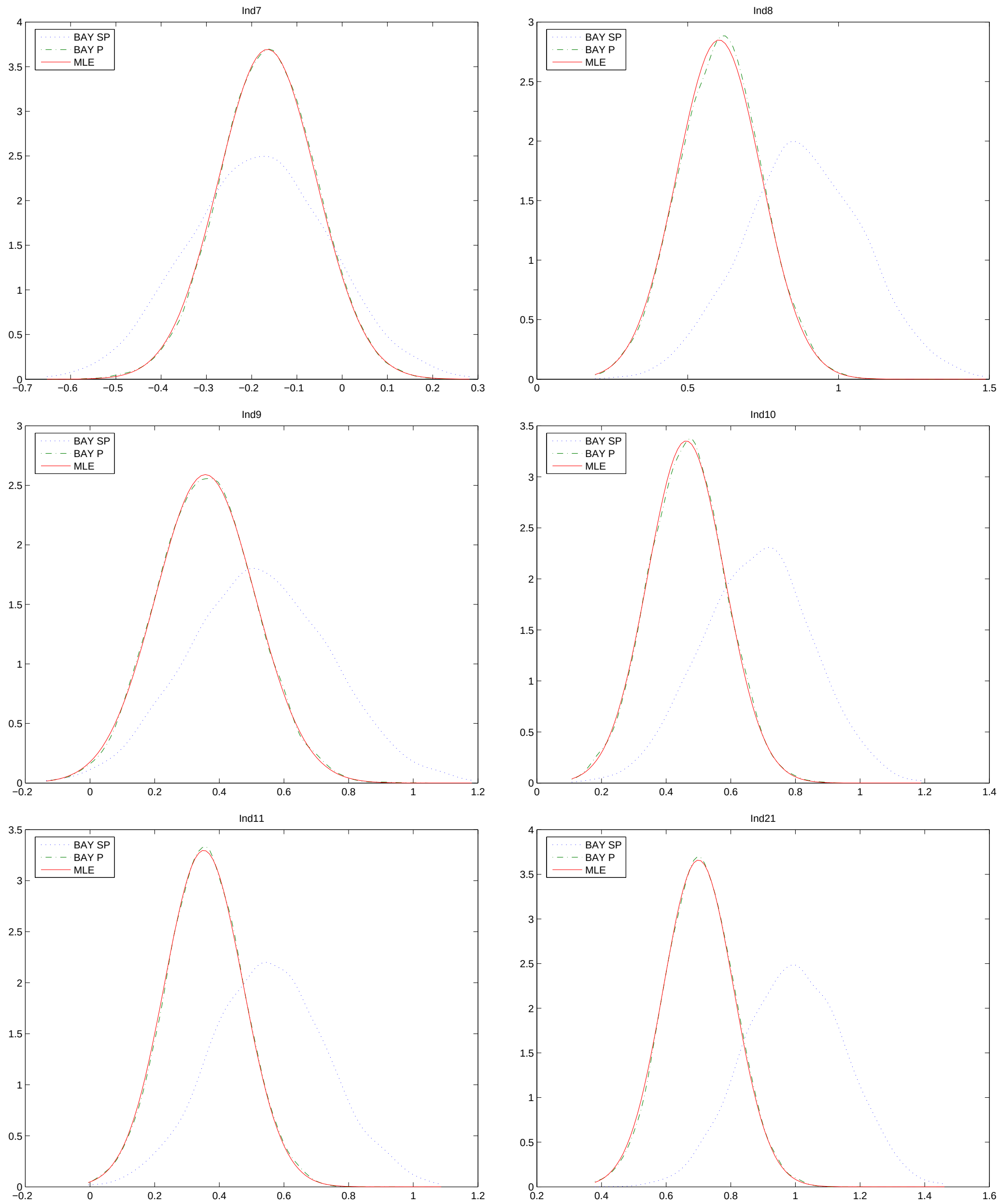


Figure 4.3: Sampling distributions for parameters for 2005Q4.

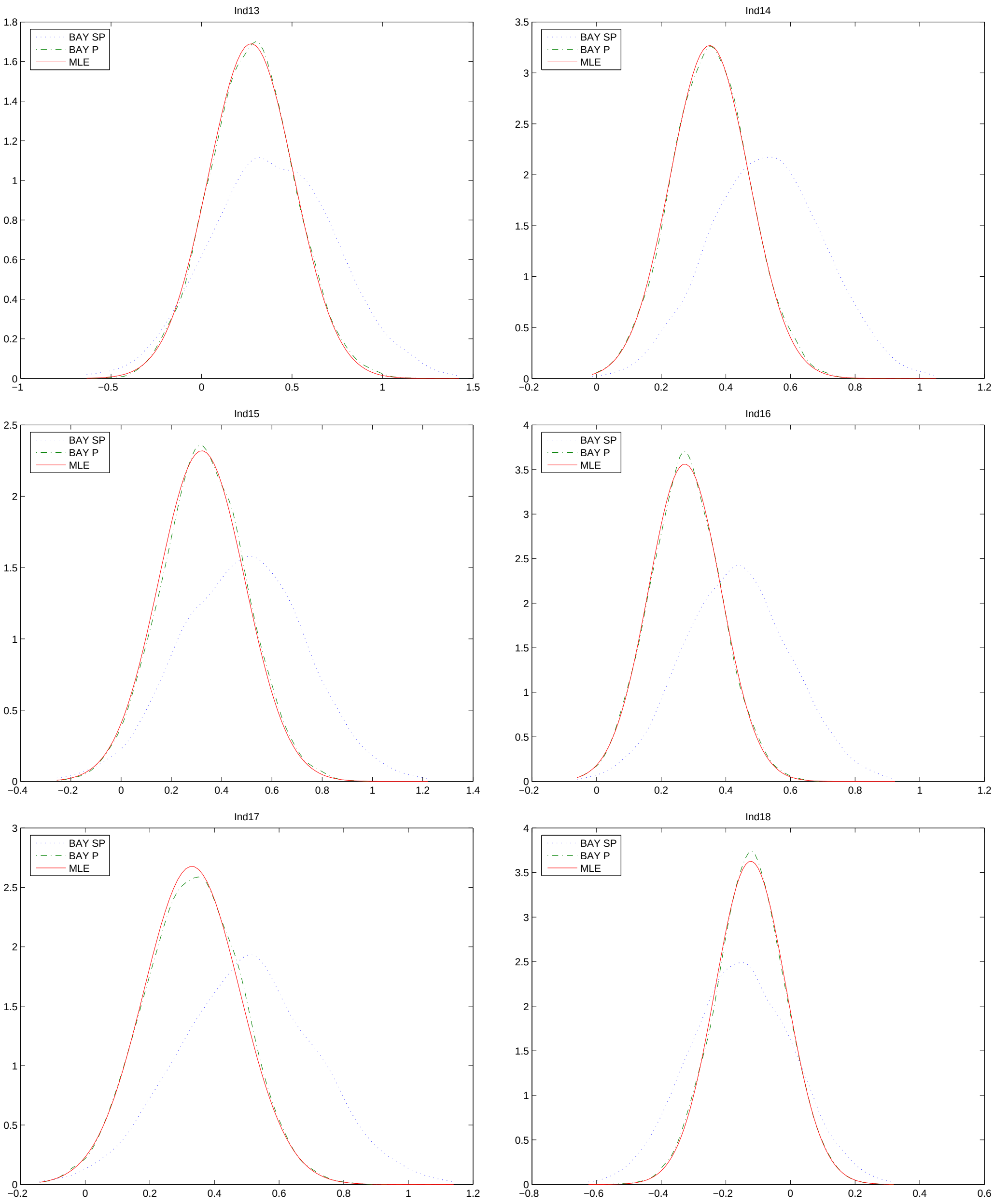


Figure 4.4: Sampling distributions for parameters for 2005Q4.

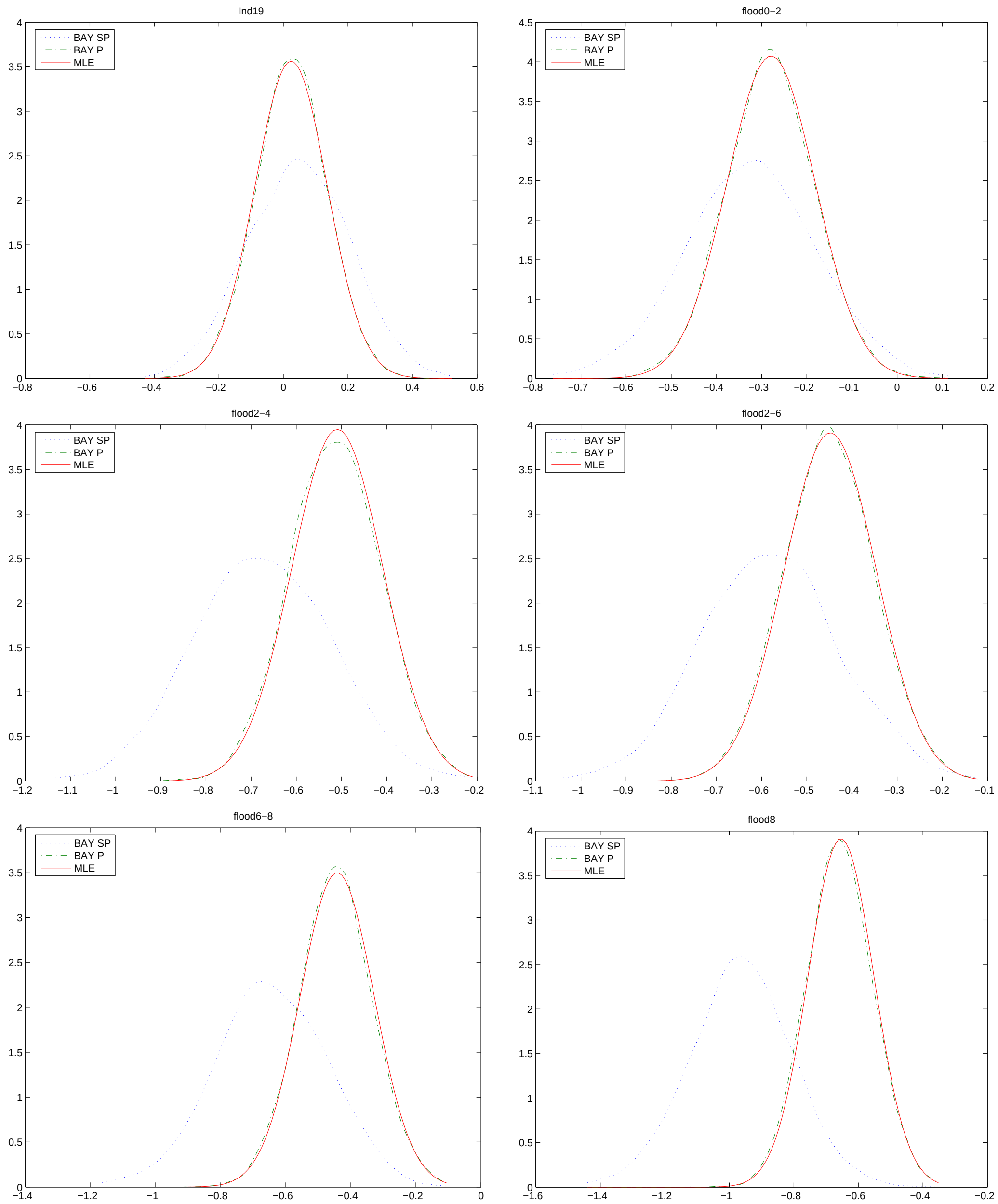


Figure 4.5: Sampling distributions for parameters for 2005Q4.

The relative size variable has a positive sign in all 3 quarters. To interpret the relative size variable, it is useful to think of a simple example where a firm may have multiple locations with an identical number of employees. In this case, the relative size variable is simply 1 divided by the number of locations. For a sole proprietorship relative size is one, with two locations it takes a value one-half, with twenty locations 0.05 and so forth. Thus, the change of 1 is roughly moving from a very large chain to sole proprietorship and implies an increase in probability by 8.6% for 2004Q4, by a factor of 26.5% in 2006Q2, and by a factor of 22.4% in 2007Q2 holding other things constant. Going from 2 locations to 1 location would imply an increase in probability by a factor of 4.3% for 2005Q4, while going from 20 locations to 1 implies an increase in probability of 8.2% for that same quarter. Campanella's (2007) study provides a good point of reference for our results. He finds that 75% of local businesses had reopened compared to 59% of national chains over a 15 month period ending November 2006. Though he is not using statistical analysis to hold other factors constant, the fact that the similarity between our 26.5% and the 26% difference in his study is reassuring.

The interaction term between the relative size of the firm and the flood variable has a negative sign in the first quarter. The sign flips for the following 2 quarters considered. Two years later after Katrina hit, locally owned businesses that were flooded are more likely to reopen.

With respect to the size of the firm, Tables 4.3 and 4.6.2 contain three dummy variables with over 250 employees as the omitted group. Recall that the literature predicts higher survival rates for larger firms. With regard to very small employers, our results conform to this prediction. Table 4.6.2 predicts that firms with less than five employees (Size1=1) were 38% less likely to be open in 2005Q4 or 2006Q2 and 15% less likely in 2007Q2. The pattern varies across time periods for firms with five to forty-nine employees (Size2=1). Firms with fifty to 249 were more likely to be open in all three quarters than the largest firms (18% more likely in 2005Q4 and 2007Q2 and 13% more likely in 2006Q2).

When we examine the relationship between the industry category and the reopening decision we find the following. All industry types except utilities and accommodation and food services were more likely to reopen immediately after the storm when compared to public administration businesses. Firms in construction had a higher probability of reopening by a factor of 0.2 in all three quarters considered when compared to public administration businesses.

Not surprisingly the coefficients attached to all flood variables have a negative sign and are generally quite large for all 3 quarters considered. For example, having been flooded with less than 2 feet of water compared with no flooding decreases the reopening probability by 9.9% for 2005Q4. The magnitude increases for the next 2 quarters considered, in 2007Q2 the probability of reopening decreases by a factor of 3.9%. All the flood variables increase in magnitude over time. The largest magnitudes occur for employers with eight or more feet of flooding. Holding other things constant, this level of flooding reduces the probability of opening by 34% in 2005Q4, 51% in 2006Q4, and 66% in 2007Q2 relative to firms with no flooding. The growing impact of flooding on firm survival is somewhat surprising and may indicate that some firms tried to reopen in areas with heavy damage, only to fail after a short period.

The error term attached to the spatial component θ is just over one in all 3 quarters. The spatial autocorrelation term ρ diminishes in magnitude as time passes. This finding suggests that spatial interactions between establishments were very important in the quarters immediately after the storm, but that the spatial component loses importance as time passes.

Table 4.4: **Estimation results.** The first column displays the variable symbol. For each quarter, columns 1, 2, 3, and 4 report posterior means, posterior standard deviations, 2.5% and 97.5% percentile

	Quarter 6				Quarter 8				Quarter 12			
	p.mean	p.Std dev	2.5%	97.5%	p.mean	p.Std dev	2.5% p	97.5%	p.mean	p.Std dev	2.5%	97.5%
Intercept	0.826	0.238	0.353	1.286	0.592	0.254	0.090	1.095	0.023	0.248	-0.440	0.507
rel size	0.238	0.115	0.007	0.463	0.762	0.115	0.538	0.990	0.626	0.114	0.409	0.853
rel size&flood	-0.247	0.146	-0.526	0.048	0.363	0.147	0.076	0.644	0.778	0.150	0.489	1.078
Size1	-1.040	0.187	-1.411	-0.693	-1.037	0.197	-1.438	-0.664	-0.464	0.185	-0.847	-0.110
Size2	-0.186	0.187	-0.553	0.166	-0.142	0.191	-0.529	0.227	0.383	0.186	0.006	0.732
Size3	0.551	0.203	0.145	0.938	0.348	0.211	-0.064	0.752	0.470	0.203	0.075	0.880
Ind1	2.073	0.846	0.561	3.866	0.886	0.684	-0.376	2.257	1.579	0.794	0.143	3.227
Ind2	1.212	0.366	0.505	1.942	0.928	0.365	0.221	1.656	0.515	0.347	-0.151	1.181
Ind3	-1.289	0.515	-2.337	-0.312	-0.498	0.494	-1.478	0.428	0.138	0.455	-0.735	1.026
Ind4	0.709	0.187	0.338	1.071	0.632	0.194	0.260	1.030	0.756	0.193	0.377	1.131
Ind5	0.467	0.202	0.058	0.864	0.174	0.200	-0.212	0.566	0.321	0.199	-0.062	0.720
Ind6	1.210	0.186	0.840	1.570	0.596	0.186	0.237	0.957	0.415	0.182	0.059	0.775
Ind7	-0.185	0.156	-0.488	0.121	-0.220	0.158	-0.533	0.094	-0.032	0.159	-0.341	0.276
Ind8	0.884	0.201	0.492	1.282	0.700	0.206	0.304	1.103	0.503	0.200	0.109	0.890
Ind9	0.523	0.219	0.101	0.955	0.114	0.221	-0.304	0.553	0.062	0.219	-0.374	0.478
Ind10	0.680	0.170	0.349	1.007	0.697	0.179	0.351	1.051	0.590	0.174	0.255	0.937
Ind11	0.554	0.177	0.207	0.910	0.175	0.181	-0.176	0.541	0.213	0.181	-0.132	0.580
Ind12	0.997	0.155	0.700	1.296	0.704	0.166	0.385	1.037	0.651	0.163	0.340	0.965
Ind13	0.395	0.343	-0.266	1.078	-0.092	0.338	-0.786	0.571	0.084	0.340	-0.619	0.717
Ind14	0.525	0.175	0.186	0.866	0.361	0.183	0.009	0.732	0.203	0.180	-0.156	0.559
Ind15	0.482	0.246	-0.002	0.961	0.242	0.263	-0.264	0.765	0.444	0.261	-0.054	0.961
Ind16	0.434	0.163	0.118	0.754	0.038	0.166	-0.285	0.361	0.100	0.164	-0.223	0.420
Ind17	0.500	0.214	0.080	0.922	0.088	0.212	-0.327	0.493	0.268	0.214	-0.158	0.677
Ind18	-0.153	0.157	-0.459	0.154	-0.339	0.163	-0.650	-0.011	-0.183	0.162	-0.499	0.129
Ind19	0.047	0.159	-0.270	0.357	-0.024	0.165	-0.350	0.297	0.077	0.163	-0.243	0.398
flood 0-2	-0.325	0.146	-0.622	-0.038	-0.747	0.144	-1.028	-0.458	-1.109	0.149	-1.407	-0.817
flood 2-4	-0.675	0.152	-0.968	-0.379	-1.079	0.151	-1.374	-0.785	-1.210	0.150	-1.513	-0.925
flood 4-6	-0.580	0.153	-0.876	-0.280	-1.174	0.156	-1.482	-0.866	-1.325	0.160	-1.630	-1.011
flood 6-8	-0.646	0.173	-0.997	-0.320	-1.359	0.180	-1.725	-1.009	-1.589	0.191	-1.970	-1.224
flood 8	-0.965	0.158	-1.286	-0.658	-1.645	0.167	-1.962	-1.316	-1.956	0.178	-2.310	-1.619
σ^2	1.031	0.069	0.896	1.169	1.033	0.133	0.750	1.245	1.090	0.140	0.757	1.302
ρ	0.454	0.109	0.202	0.633	0.318	0.185	0.001	0.616	0.288	0.176	-0.025	0.566

Table 4.5: **Marginal effects.** The first column displays the variable symbol. For each quarter, columns 1, 2, 3, and 4 report the posterior means, posterior standard deviations, 2.5% and 97.5% percentile.

	Quarter 6				Quarter 8				Quarter 12			
	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%
rel size	0.086	0.041	0.003	0.166	0.265	0.039	0.189	0.342	0.224	0.040	0.147	0.303
rel size&flood	-0.089	0.053	-0.190	0.017	0.126	0.051	0.027	0.223	0.279	0.053	0.176	0.384
Size1	-0.382	0.070	-0.514	-0.246	-0.375	0.074	-0.517	-0.233	-0.149	0.067	-0.292	-0.032
Size2	-0.070	0.070	-0.204	0.065	-0.054	0.073	-0.197	0.089	0.143	0.067	0.003	0.267
Size3	0.183	0.073	0.045	0.330	0.126	0.078	-0.021	0.279	0.178	0.074	0.029	0.323
Ind1	0.351	0.099	0.159	0.546	0.232	0.163	-0.141	0.499	0.476	0.180	0.053	0.743
Ind2	0.305	0.085	0.144	0.481	0.270	0.096	0.074	0.456	0.194	0.129	-0.053	0.435
Ind3	-0.425	0.137	-0.651	-0.119	-0.183	0.174	-0.500	0.156	0.057	0.161	-0.233	0.386
Ind4	0.221	0.065	0.099	0.353	0.210	0.068	0.083	0.353	0.286	0.070	0.146	0.421
Ind5	0.156	0.069	0.020	0.295	0.064	0.074	-0.079	0.210	0.120	0.074	-0.023	0.267
Ind6	0.313	0.072	0.181	0.460	0.201	0.067	0.076	0.337	0.156	0.067	0.023	0.288
Ind7	-0.070	0.059	-0.186	0.046	-0.084	0.061	-0.202	0.037	-0.013	0.057	-0.125	0.094
Ind8	0.259	0.068	0.134	0.404	0.228	0.071	0.095	0.374	0.190	0.074	0.043	0.333
Ind9	0.171	0.072	0.036	0.316	0.042	0.082	-0.116	0.204	0.023	0.079	-0.132	0.175
Ind10	0.215	0.061	0.103	0.339	0.228	0.065	0.105	0.363	0.224	0.064	0.099	0.349
Ind11	0.182	0.062	0.066	0.309	0.065	0.067	-0.065	0.199	0.078	0.067	-0.051	0.212
Ind12	0.282	0.065	0.162	0.413	0.231	0.063	0.113	0.361	0.247	0.059	0.133	0.358
Ind13	0.127	0.110	-0.100	0.328	-0.037	0.126	-0.297	0.204	0.035	0.121	-0.197	0.269
Ind14	0.174	0.062	0.060	0.299	0.129	0.067	0.003	0.265	0.074	0.065	-0.058	0.199
Ind15	0.159	0.081	-0.001	0.317	0.087	0.094	-0.098	0.275	0.167	0.098	-0.020	0.358
Ind16	0.148	0.059	0.038	0.271	0.015	0.063	-0.106	0.139	0.035	0.059	-0.085	0.149
Ind17	0.165	0.071	0.027	0.305	0.033	0.079	-0.124	0.183	0.099	0.080	-0.060	0.252
Ind18	-0.058	0.060	-0.174	0.058	-0.131	0.062	-0.249	-0.004	-0.064	0.057	-0.180	0.042
Ind19	0.018	0.059	-0.099	0.134	-0.009	0.063	-0.132	0.114	0.027	0.059	-0.092	0.138
flood 0-2	-0.099	0.048	-0.198	-0.012	-0.175	0.051	-0.281	-0.086	-0.390	0.056	-0.496	-0.281
flood 2-4	-0.225	0.057	-0.340	-0.118	-0.289	0.066	-0.417	-0.168	-0.428	0.055	-0.533	-0.319
flood 4-6	-0.189	0.056	-0.301	-0.084	-0.325	0.068	-0.456	-0.198	-0.470	0.056	-0.572	-0.358
flood 6-8	-0.214	0.065	-0.350	-0.098	-0.396	0.077	-0.545	-0.244	-0.557	0.059	-0.665	-0.435
flood 8	-0.337	0.062	-0.458	-0.216	-0.506	0.071	-0.633	-0.359	-0.656	0.046	-0.740	-0.558

4.7 Conclusion

In this chapter a Bayesian framework is used in order to investigate the post storm survival of firms in the Orleans Parish. A novelty of our approach is the spatial component in the model specification. In particular, we model each firm's decision of reopening as a function of firm characteristic variables and as a function of neighboring firms' decision to reopen. We estimate a spatial probit model on a dataset containing quarterly data on 8,171 firms from the Orleans Parish and find evidence indicating the presence of spatial components, especially in the quarters immediately following the storms. Other findings are: larger firms are more likely to survive; also, less flooded firms are more likely to survive; finally, sole proprietorships are more likely to reopen than large chain stores.

Chapter 5

Summary and Conclusions

This dissertation is focused on econometric techniques and statistical methods for analyzing data. It contains two essays discussing models for financial volatility and one essay on Bayesian inferences of business survival in New Orleans after the hit of the hurricanes Katrina and Rita. Highly computational methods were necessary for all three essays in order to obtain the results presented in this dissertation.

In the first essay models for daily realized volatility are discussed, with special focus on models that approximate long memory by aggregation. Although these models are widely used in practice and are able to replicate the most important features of financial data, little research has been done in the direction of understanding their main properties. Therefore in the first essay we discuss stationarity conditions and find that depending on the model, the standard stationarity condition that compares the sum of the autoregressive coefficients to one is not always a good indicator for stability. In this essay the lag selection problem is also discussed. Typically realized volatility models are employed with a daily, weekly, and monthly specification. We relax this specification and an optimal lag structure is researched. We allow for a maximum of three time scales and set the maximum scale to a value of 250. A computationally intense method based on the in-sample fit and out-of-sample fit is employed in the context of thirty DJIA stocks. We start by constructing daily volatility estimates from the intraday data. Next we find an optimal lag structure for each stock considered. We find under both optimality criteria considered a short lag and a long lag. The results reveal gains in forecasting when comparing the models based on the optimal lag structure against the benchmark model with daily, weekly, and monthly realized volatility. In terms of future research, specifications with 2 lag components and 4 lag components must be researched and compared to the specification with 3 lags employed in this essay. The models described in this chapter are based on an autoregressive specification. The long memory property found in the data is approximated by aggregation. Therefore, the way in which these models approximate long memory is by aggregating AR models with short memory. A logical step would be to analyze MA specification in realized volatility.

The second essay is also written in the context of financial volatility. The essay presents results regarding the correlation structure in the context of ARCH models. The difference between ARCH models and the models described in the first essay is that ARCH models volatility as unobservable. This chapter is focused on understanding the correlation structure in the context of the HARCH model. This model is chosen because of the way in which it captures the high persistence. Unlike other ARCH models that rely on past squared returns in the volatility equation, the HARCH model specifies volatility as a function of past aggregated squared returns. Therefore, in the ARCH world the HARCH specification is of particular interest. We make use of extensive simulations to generate synthetic data in order to analyze the relationship between the HARCH correlation structure and the correlation structure from other models proposed in the literature. We find that HARCH is only able to capture correlation structures obtained from ARCH data. However, our findings indicate that HARCH does not seem to be able to capture correlation scales from ARCH data in the mean reverting sense, but it captures the influence of the lags used in the data generating process. Further studies need to extend this analysis to higher lag structures where specifications with 3 lags are employed.

The last essay is dealing with estimating the impact of storms Katrina and Rita on firm survival in New Orleans. Previous studies on the problem of business survival in New Orleans are based on summary statistics. Besides the main practical interest generated by the results for the affected area of New Orleans, this essay is of particular interest because of its econometric methodology. We use a Spatial Bayesian Probit specification. The novelty is the spatial component. In particular, we allow for a firm's decision to reopen to be influenced by the same decision of nearby firms. The results indicate the presence of spatial components, especially in the quarters immediately following the storms. Other findings in this essay are: larger firms are more likely to survive, less flooded firms are more likely to survive and sole proprietorships are more likely to reopen than large chain stores. In terms of future work, we plan to extend the model in order to be able to use it to assess the impact of various government programs on firm survival.

Bibliography

Agarwal, R. 1996. Technological activity and survival of firms. Economic Letters 52 (1): 101–108.

Agarwal, R. 1997. Survival of Firms over the Product Life Cycle. Southern Economic Journal 63 (3): 571–584.

Agarwal, R., Audretsch D.B. 2001. Does Entry Size Matter? The Impact of the Life Cycle and Technology on Firm Survival. The Journal of Industrial Economics 49 (1): 21–43.

Agarwal, R., Gort M. 1996. The Evolution of Markets and Entry, Exit and Survival of Firms. The Review of Economics and Statistics 78 (3): 489–498.

Agarwal, R., Gort, M. 2002. Firm and Product Life Cycles and Firm Survival. The American Economic Review, Papers and Proceedings of the One Hundred Fourteenth Annual Meeting of the American Economic Association 92 (2): 184–190.

Andersen, T., Bollerslev, T. 1998. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. International Economic Review 39: 885-906.

Andersen, T., Bollerslev, T., Diebold, F.X. 2007. Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility. The Review of Economics and Statistics. 89(4): 701-720.

Andersen, T., Bollerslev, T., Diebold, F.X. and Ebens, H. 2001a. The Distribution of Realized Stock Return Volatility. Journal of Financial Economics 61: 43-76.

Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P. 2003. Modeling and Forecasting Realized Volatility. Econometrica 71: 529-626.

Andersen, T. Bollerslev, T., Diebold, F.X. and Labys, P. 2001b. The Distribution of Realized Exchange Rate Volatility. Journal of the American Statistical Association 96: 42-55.

Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P. 2000a. Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian. Multinational Finance Journal 4: 159-179.

Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P. 2000b. Great Realizations. Risk 18: 105-108.

Andersen, T., T.Bollerslev, T., Huang, X. 2007. A Reduced Form Framework for Modeling Volatility of Speculative Prices based on Realized Variation Measures. Manuscript. CREATES 2007-14.

Andersen, T., T.Bollerslev, T., Huang, X. 2006. A Semiparametric Framework for Modeling and Forecasting Jumps and Volatility in Speculative Prices. Manuscript. Duke University.

Audretsch, D. B. 1991. New-Firm Survival and the Technological Regime. The Review of Economics and Statistics 73 (3): 441–450.

Audretsch, D. B., Mahmood, T. 1995. New Firm Survival: New Results Using a Hazard Function. The Review of Economics and Statistics 77 (1): 97–103.

Baillie, R., Bollerslev, T., Mikkelsen, H. 1996. Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 74: 3-30.

Baldwin, J. R. 1995. The Dynamics of Industrial Competition. Cambridge University Press, Cambridge.

Baldwin, J. R., Gorecki, P. K. 1991. Firm Entry and Exit in the Canadian Manufacturing Sector, 1970-1982. The Canadian Journal of Economics / Revue canadienne d'Economique 24 (2): 300–323.

Baldwin, J. R., Rafiquzzaman, M. 1995. Selection versus Evolutionary Adaptation: Learning and Post-Entry Performance. International Journal of Industrial Organization 13: 501–522.

Bandi, F., Russell, J. 2005. Market Microstructure noise, integrated variance estimators, and the limitations of asymptotic approximations: a solution. Manuscript. Graduate School of Business. University of Chicago.

Bandi, F., Russell, J. 2005. Microstructure noise, realized volatility, and optimal sampling. Journal of Financial Economics 79 (3): 655-92.

Bandi, F., Russell, J. 2006. Separating Microstructure Noise from Volatility. Journal of Financial Economics 79 (3): 655-92.

Barndorff-Nielsen, O., Shephard, N. 2002. Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models. Journal of the Royal Statistical Society 64: 253-280.

Barndorff-Nielsen, O., Shephard, N. 2006. Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. Journal of Financial Econometrics 4 (1): 1-30.

Barndorff-Nielsen, O., Hansen P., Lunde A., Shephard, N. 2007. Designing realised kernels to measure ex-post variation of equity prices in the presence of noise. Manuscript. SSRN 620203.

Barndorff-Nielsen, O., Hansen P., Lunde A., Shephard, N. 2007. Designing realised kernels to measure ex-post variation of equity prices in the presence of noise. Empirical appendix: Description of data and estimations. Manuscript. SSRN 620203.

Barndorff-Nielsen, O., Hansen P., Lunde A., Shephard, N. 2007. Subsampling realised kernels. Manuscript. SSRN 927483.

Barucci, E., Reno, R. 2002. On measuring volatility of diffusion processes with high frequency data. Economics Letters 74: 371-378.

Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 21: 307-328.

Bollerslev, T. 1987. A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. The Review of Economics and Statistics. 69 (3): 542-547.

Bollerslev, T., Ghysels, E. 1996. Periodic Autoregressive Conditional Heteroskedasticity. Journal of Business and Economic Statistics. 14: 139–151.

Bollerslev, T., Kretschmer, U., Pigorsch, C., Tauchen, G. 2007. A Discrete-Time Model for Daily S&P 500 Returns and Realized Variations: Jumps and Leverage Effects. Forthcoming. Journal of Econometrics.

Campanella, R. 2007. Street Survey of Business Reopenings in Post-Katrina New Orleans. CBR Whitepaper funded by National Science Foundation Award 0554937. www.kerrn.org/pdf/campanella2.pdf.

Caves R. E. 1998. Industrial Organization and New Findings on the Turnover and Mobility of Firms. Journal of Economic Literature 36 (4): 1947–1982.

Christensen, K., Oomen, R., Podolskij, M. 2008. Realised Quantile-Based Estimation of the Integrated Variance. Manuscript. SSRN 1085553.

Corsi, F. 2004. A Simple Long Memory Model of Realized Volatility. Manuscript. University of Southern Switzerland.

Corsi, F., Kretschmer,U., Mittnik, S., Pigorsch, C. 2005. The Volatility of Realized Volatility. Econometric Reviews, forthcoming.

Dahlhamer, J.M., Tierney, K. J. 1996. Winners and losers: predicting business disaster recovery outcomes following the Northridge Earthquake. University of Delaware Disaster Research Center.

Diebold, F. X. 1986. Modeling the persistence of conditional variances: A comment. Econometric Reviews 5: 51-56.

Ding, Z., Granger C. W. J., Engle R. F. 1993. A Long Memory Property of Stock Market Returns and a New Model. Journal of Empirical Finance 1: 83-106.

Ding, Z., Granger C. W. J. 1996. Modeling Volatility Persistence of Speculative returns. A New Approach. Journal of Econometrics 73 (1): 185-215.

Embrechts, P., Samorodnitsky, G., Dacorogna, M. M., Müller, U. A. 1998. How heavy are the tails of a stationary HARCH(k) process? A study of the moments. Stochastic Processes and Related Topics: In memory of Stamatis Cambanis. Birkhäuser.

Engle, R. F. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica 50 (4): 987-1007.

Engle, R. F. 1990. Stock Volatility and the Crash of 87. The Review of Financial Studies 50 (4): 987–1007.

Engle, R.F., Lilien, D.M., and Robins, R.P. 1987. Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model. Econometrica 55 (2): 391–408.

Engle, R. F., Patton, A. J. 2001. What good is a volatility model? Quantitative Finance 1(2): 237–245.

Forsberg, L., Bollerslev, T. 2002. Bridging the Gap Between the Distribution of Realized (ECU) Volatility and ARCH Modeling (of the Euro): The GARCH-NIG Model. Journal of Applied Econometrics 17: 535-548.

Geroski, P. A. 1995. What Do We Know About Entry. International Journal of Industrial Organization 13: 450–456.

Gort, M., Klepper, S. 1982. Time Paths in the Diffusion of Product Innovations. The Economic Journal 92 (367): 630–653.

Granger, C. 1980. Long memory relationships and the aggregation of dynamic models. Journal of Econometrics. 14: 227-38.

Granger, C., Joyeux, R. 1980. An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis. 1: 15-29.

Greene, W. H. 2003. Econometric Analysis. 5th edition, Prentice Hall, Englewood Cliffs, NJ

Greenberg, E. 2008. Introduction to Bayesian Econometrics. Cambridge University Press.

Griffiths, W. E., Hill, R. C., O'Donnell, C. J. 2006. A Comparison of Bayesian and Sampling Theory Inferences in a Probit Model. Berkeley Electronic Press.

Griffiths, W., Hill, R. C., Judge, G. 1993. The Theory and Practice of Econometrics. 2nd edition, John Wiley and Sons, NY.

Hamilton, J.D. 1994. Time Series Analysis. Princeton University Press.

Hansen P. 2005. A Test for Superior Predictive Ability. Journal of Business and Economic Statistics 23: 365-380.

Hansen, P., Large, J., Lunde, A. 2007. Moving average-based estimators of integrated variance. Econometric Reviews, forthcoming.

Hansen, P., Large, J., Lunde A. 2006. Realized variance and market microstructure noise (with discussion). Journal of Business and Economic Statistics 24: 127-128.

Higgins, M. L., Bera, A.K. 1992. A class of non-linear ARCH models. International Economic Review. 33: 137–158.

Hillebrand, E., Medeiros, M.C. 2007. Forecasting Realized Volatility Models: The Benefits of Bagging and Nonlinear Specifications. Manuscript. Department of Economics PUC-Rio (Brazil).

Hillebrand, E. 2006. Overlaying Time Scales in Financial Volatility Data. Fomby, T.B. and D. Terrell: Advances in Econometrics, Vol. 20: Econometric Analysis of Financial and Economic Time Series/Part B. Elsevier 153–178.

Hillebrand, E. 2005. Neglecting Parameter Changes in GARCH Models. Journal of Econometrics 12: 121-138.

Hillebrand, E. 2003. Mean Reversion Models of Financial Markets. Dissertation.

Harvey, A.C. 1998. Forecasting Volatility in the Financial Markets. 2nd ed. Butterworth-Heinemann. 307-320.

Hull, J., White, A. 1987. The pricing of options on assets with stochastic volatilities. Journal of Finance 42 (2): 281-300.

Hull, J., White, A. 1988. An analysis of the bias in option pricing caused by a stochastic volatility. Journal of Finance. 42 (2): 281-300.

Hyung, N., Poon, S.-H., Granger, C. 2005. A Source of Long Memory in Volatility. Manuscript. .

Jeffreys, A. 1967. Theory of Probability. London: Oxford University Press (3rd edn.).

Judge, G., Griffiths, W., Hill, R. C., Lütkepohl, H., Lee, T. 1985. Learning and Practicing Econometrics. New York:Wiley.

Lamoureux, C.G., Lastrapes, W. D. 1990. Persistence in variance, structural change, and the GARCH model. Journal of Business and Economic Statistics 8: 225-234.

LeBaron, B. 2001. Stochastic volatility as a simple generator of financial power laws and long memory. Quantitative Finance 1: 621-631.

LeSage, J. P. 1998. Spatial Econometrics. Manuscript.

LeSage, J. P. An Introduction to Spatial Econometrics. forthcoming, Revue d'Economie Industrielle, a special issue devoted to spatial econometrics, edited by Julie LeGallo, Anne Plunket, and Nadine Massard.

Malliavin, P., Mancino, M.E. 2002. Fourier series method for measurement of multivariate volatilities. Finance and Stochastics 6: 49-61.

Mandelbrot, B. 1963. The Variation of Certain Speculative Prices. Journal of Business 36(4): 394-419.

Martens, M., van Dijk, D., De Pooter, M. 2004. Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity. Manuscript. Tinbergen Institute.

Mata, J. and Portugal, P. 1994. Life Duration of New Firms. Journal of Industrial Economics 27: 227–246.

Mata, J., Portugal, P. and Guimaraes, P. 1995. The Survival of New Plants: Start-up Conditions and Post-Entry Evolution. International Journal of Industrial Organization 13: 459–481.

McAleer, M., Medeiros, M.C. 2007. A Multiple Regime Smooth Transition Heterogeneous Autoregressive Model for Long Memory and Asymmetries. Journal of Econometrics, forthcoming.

McAleer, M., Medeiros, M.C. 2006. Realized Volatility: A Review. Econometric Reviews forthcoming.

Merton, R.C. 1980. On Estimating the Expected Return on the Market: An Ex-ploratory Investigation. Journal of Financial Economics 8: 323-361.

Mikosch, U. A., Starica, C. 2004. Nonstationarities in financial time series, the long range dependence, and the IGARCH effects. The Review of Economics and Statistics 86: 378-390.

Müller, U. A., Dacorogna, M. M., Dave, R. D., Olsen, R. B., Pictet, O. V., and J. E. von Weizsaecker. 1997. Volatilities of different time resolutions - Analyzing the dynamics of market components. Journal of Empirical Finance 4: 213-239.

Nelson, D.B. 1990. Stationarity and Persistence in the GARCH(1,1) Model. Econometric Theory 6: 318–334.

Newey, W.K., West, K.D. 1987. A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55: 703-708.

Ogunc, A. K. 2002. Essays on the Bayesian Inequality Restricted Estimation. Dissertation, Louisiana State University.

Oomen, R. 2006. Properties of realized variance under alternative sampling schemes. Journal of Business and Economic Statistics. 24: 219-237.

Phaneuf, D., Palmquist, R. B. 2003. Estimating Spatially and Temporally Explicit Land Conversion Models using discrete duration. manuscript .

Runyan, R. C. 2006. Small Business in the Face of Crisis: Identifying Barriers to Recovery from a Natural Disaster. Journal of Contingencies and Crisis Management 14 (1): 12 –26.

Smith, T. E., LeSage J.P. 2004. A Bayesian Probit Model with Spatial Dependencies. LeSage, J.P. and Pace, R. K.: Advances in Econometrics, Vol. 18: Advances in Econometrics Spatial And Spatiotemporal Econometrics. Elsevier 127–160.

Sentana, E. 1995. Quadratic ARCH Models. Review of Economic Studies 62: 639–661.

Sutton J. 1997. Gibrat's Legacy. Journal of Economic Literature 35 (1): 40–59.

Taylor, S.J. 1986. Modeling Financial Time Series. 1st ed. John Wiley & Sons Ltd.

Terrell, M.D., Bilbo, R. 2007. A Report on the Impact of Hurricanes Katrina and Rita on Louisiana Businesses: 2005Q2-2006Q4. www.bus.lsu.edu/ded.

Terrell, M.D., Bilbo, R., Lam, N. 2007. A Comparison of Self-Reported Flooding from Hurricane Katrina to GIS Elevation Based Measured Flooding. Manuscript.

Wagner, J. 1994. The Post–Entry Performance of New Small Firms in German Manufacturing Industries. Journal of Industrial Economics 62: 141-154.

Zhang, L. 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Bernoulli 12 (6), 1019-1043.

Zhang, L, Mykland P., Aït-Sahalia Y. 2005. A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. Journal of the American Statistical Association 100: 1394-13411.

Zhou, B. 1996. High frequency data and volatility in foreign-exchange rates. Journal of Business and Economic Statistics 14: 45-52.

Appendix: Supplementary Tables for Chapter 4

Table 5.1: **Marginal effects.** Marginal effects for firm with probability of opening in the 25% percentile and no open neighbors. The first column displays the variable symbol. For each quarter, columns 1, 2, 3, and 4 report the posterior means, posterior standard deviations, 2.5% and 97.5% percentile.

	Quarter 6				Quarter 8				Quarter 12			
	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%
rel size	0.070	0.045	0.001	0.162	0.233	0.082	0.046	0.366	0.191	0.071	0.039	0.312
rel size&flood	-0.073	0.053	-0.188	0.011	0.111	0.058	0.012	0.230	0.237	0.089	0.049	0.391
Size1	-0.291	0.052	-0.399	-0.197	-0.049	0.018	-0.092	-0.021	-0.116	0.044	-0.206	-0.033
Size2	-0.072	0.072	-0.215	0.061	-0.021	0.029	-0.092	0.021	0.098	0.046	0.002	0.185
Size3	0.180	0.060	0.054	0.289	0.026	0.017	-0.007	0.061	0.117	0.047	0.023	0.210
Ind1	0.358	0.085	0.173	0.506	0.036	0.037	-0.049	0.093	0.211	0.075	0.038	0.344
Ind2	0.311	0.076	0.157	0.459	0.046	0.020	0.012	0.093	0.121	0.079	-0.046	0.265
Ind3	-0.425	0.132	-0.637	-0.122	-0.099	0.110	-0.374	0.040	0.024	0.128	-0.269	0.232
Ind4	0.227	0.063	0.105	0.353	0.041	0.019	0.013	0.086	0.170	0.052	0.076	0.279
Ind5	0.160	0.070	0.020	0.296	0.017	0.020	-0.019	0.060	0.088	0.056	-0.017	0.203
Ind6	0.320	0.062	0.203	0.446	0.040	0.019	0.011	0.084	0.110	0.052	0.014	0.217
Ind7	-0.071	0.059	-0.184	0.047	-0.026	0.019	-0.062	0.013	-0.008	0.050	-0.101	0.093
Ind8	0.265	0.064	0.145	0.393	0.043	0.019	0.015	0.088	0.128	0.054	0.027	0.236
Ind9	0.176	0.073	0.035	0.318	0.011	0.023	-0.031	0.058	0.018	0.067	-0.116	0.145
Ind10	0.220	0.060	0.106	0.340	0.043	0.019	0.014	0.088	0.145	0.051	0.055	0.250
Ind11	0.187	0.063	0.066	0.316	0.017	0.019	-0.014	0.062	0.063	0.054	-0.037	0.176
Ind12	0.288	0.059	0.178	0.406	0.044	0.019	0.015	0.089	0.156	0.050	0.069	0.258
Ind13	0.131	0.112	-0.102	0.336	-0.017	0.045	-0.126	0.053	0.018	0.102	-0.208	0.190
Ind14	0.178	0.062	0.059	0.301	0.030	0.019	0.001	0.073	0.060	0.054	-0.042	0.169
Ind15	0.162	0.081	-0.001	0.318	0.020	0.023	-0.026	0.069	0.113	0.065	-0.015	0.242
Ind16	0.152	0.060	0.040	0.271	0.005	0.018	-0.025	0.046	0.032	0.050	-0.063	0.134
Ind17	0.169	0.072	0.027	0.313	0.009	0.022	-0.035	0.053	0.075	0.060	-0.046	0.194
Ind18	-0.058	0.060	-0.173	0.061	-0.045	0.022	-0.089	-0.002	-0.059	0.052	-0.159	0.044
Ind19	0.019	0.060	-0.098	0.137	-0.001	0.018	-0.034	0.039	0.025	0.050	-0.071	0.129
flood 0-2	-0.126	0.056	-0.241	-0.015	-0.136	0.038	-0.218	-0.072	-0.411	0.052	-0.510	-0.307
flood 2-4	-0.260	0.057	-0.368	-0.148	-0.238	0.054	-0.353	-0.142	-0.446	0.051	-0.544	-0.347
flood 4-6	-0.225	0.058	-0.335	-0.111	-0.271	0.059	-0.392	-0.165	-0.485	0.052	-0.580	-0.381
flood 6-8	-0.249	0.064	-0.377	-0.125	-0.339	0.073	-0.486	-0.206	-0.561	0.054	-0.666	-0.451
flood 8	-0.358	0.055	-0.467	-0.250	-0.449	0.073	-0.589	-0.313	-0.644	0.045	-0.727	-0.552

Table 5.2: **Marginal effects.** Marginal effects for firm with probability of opening in the 25% percentile and all neighbors open. The first column displays the variable symbol. For each quarter, columns 1, 2, 3, and 4 report the posterior means, posterior standard deviations, 2.5% and 97.5% percentile.

	Quarter 6				Quarter 8				Quarter 12			
	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%
rel size	0.062	0.043	0.000	0.155	0.204	0.096	0.016	0.362	0.180	0.076	0.021	0.312
rel size&flood	-0.065	0.051	-0.176	0.010	0.097	0.060	0.004	0.223	0.223	0.096	0.025	0.391
Size1	-0.168	0.022	-0.210	-0.126	-0.117	0.017	-0.152	-0.087	-0.032	0.010	-0.052	-0.010
Size2	-0.060	0.060	-0.188	0.042	-0.038	0.049	-0.150	0.043	0.028	0.012	0.001	0.048
Size3	0.114	0.034	0.038	0.172	0.057	0.032	-0.014	0.110	0.032	0.011	0.007	0.053
Ind1	0.189	0.032	0.118	0.239	0.086	0.068	-0.098	0.154	0.047	0.014	0.013	0.067
Ind2	0.174	0.029	0.110	0.224	0.107	0.027	0.041	0.150	0.031	0.019	-0.018	0.056
Ind3	-0.453	0.169	-0.732	-0.096	-0.158	0.156	-0.507	0.074	-0.002	0.050	-0.136	0.052
Ind4	0.136	0.026	0.081	0.183	0.091	0.020	0.047	0.129	0.043	0.008	0.027	0.060
Ind5	0.100	0.036	0.016	0.162	0.030	0.036	-0.053	0.089	0.024	0.014	-0.007	0.046
Ind6	0.179	0.020	0.139	0.219	0.088	0.021	0.043	0.125	0.030	0.011	0.006	0.050
Ind7	-0.058	0.050	-0.163	0.032	-0.056	0.043	-0.150	0.019	-0.006	0.019	-0.048	0.024
Ind8	0.155	0.024	0.105	0.200	0.096	0.020	0.054	0.134	0.034	0.011	0.010	0.053
Ind9	0.109	0.037	0.027	0.172	0.018	0.043	-0.076	0.089	0.002	0.023	-0.053	0.036
Ind10	0.134	0.025	0.082	0.179	0.097	0.018	0.059	0.133	0.038	0.009	0.021	0.055
Ind11	0.115	0.029	0.052	0.168	0.031	0.033	-0.042	0.087	0.017	0.014	-0.016	0.040
Ind12	0.166	0.020	0.127	0.206	0.098	0.018	0.063	0.133	0.040	0.008	0.025	0.056
Ind13	0.079	0.068	-0.083	0.179	-0.034	0.083	-0.240	0.090	-0.001	0.039	-0.104	0.046
Ind14	0.111	0.030	0.047	0.164	0.060	0.026	0.002	0.106	0.016	0.015	-0.019	0.040
Ind15	0.101	0.044	-0.001	0.170	0.038	0.044	-0.065	0.108	0.030	0.015	-0.006	0.054
Ind16	0.096	0.030	0.030	0.151	0.005	0.035	-0.071	0.065	0.008	0.016	-0.027	0.033
Ind17	0.105	0.037	0.022	0.167	0.013	0.043	-0.083	0.080	0.020	0.016	-0.019	0.045
Ind18	-0.038	0.040	-0.111	0.044	-0.059	0.026	-0.106	-0.002	-0.016	0.014	-0.041	0.014
Ind19	0.010	0.043	-0.082	0.085	-0.008	0.037	-0.091	0.055	0.006	0.016	-0.032	0.032
flood 0-2	-0.104	0.051	-0.214	-0.011	-0.227	0.055	-0.339	-0.123	-0.253	0.056	-0.367	-0.151
flood 2-4	-0.234	0.060	-0.355	-0.120	-0.354	0.062	-0.475	-0.233	-0.288	0.059	-0.411	-0.181
flood 4-6	-0.198	0.059	-0.318	-0.086	-0.392	0.063	-0.514	-0.264	-0.331	0.065	-0.459	-0.209
flood 6-8	-0.223	0.067	-0.361	-0.101	-0.463	0.070	-0.599	-0.324	-0.433	0.079	-0.591	-0.282
flood 8	-0.348	0.062	-0.471	-0.226	-0.567	0.059	-0.670	-0.444	-0.574	0.071	-0.704	-0.428

Table 5.3: **Marginal effects.** Marginal effects for firm with probability of opening in the 50% percentile and no open neighbors. The first column displays the variable symbol. For each quarter, columns 1, 2, 3, and 4 report the posterior means, posterior standard deviations, 2.5% and 97.5% percentile.

	Quarter 6				Quarter 8				Quarter 12			
	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%
rel size	0.057	0.043	0.000	0.153	0.171	0.102	0.005	0.350	0.141	0.086	0.003	0.301
rel size&flood	-0.060	0.051	-0.175	0.008	0.083	0.061	0.001	0.218	0.176	0.108	0.004	0.376
Size1	-0.015	0.008	-0.034	-0.004	-0.111	0.037	-0.190	-0.049	-0.067	0.029	-0.132	-0.018
Size2	-0.011	0.013	-0.046	0.005	-0.036	0.048	-0.151	0.041	0.057	0.029	0.001	0.116
Size3	0.012	0.006	0.003	0.027	0.053	0.032	-0.014	0.117	0.067	0.030	0.012	0.130
Ind1	0.015	0.008	0.004	0.036	0.081	0.070	-0.089	0.188	0.108	0.047	0.023	0.206
Ind2	0.015	0.008	0.004	0.035	0.101	0.040	0.032	0.183	0.066	0.046	-0.033	0.152
Ind3	-0.198	0.146	-0.569	-0.014	-0.152	0.152	-0.496	0.070	0.004	0.086	-0.215	0.126
Ind4	0.013	0.007	0.004	0.030	0.085	0.030	0.034	0.151	0.092	0.032	0.040	0.164
Ind5	0.011	0.006	0.002	0.026	0.028	0.036	-0.050	0.096	0.049	0.032	-0.012	0.116
Ind6	0.015	0.008	0.005	0.035	0.082	0.030	0.032	0.148	0.061	0.029	0.010	0.122
Ind7	-0.010	0.011	-0.036	0.004	-0.054	0.043	-0.152	0.017	-0.010	0.034	-0.089	0.049
Ind8	0.014	0.007	0.004	0.033	0.090	0.032	0.038	0.159	0.070	0.030	0.018	0.138
Ind9	0.011	0.007	0.002	0.027	0.016	0.042	-0.078	0.094	0.007	0.044	-0.094	0.084
Ind10	0.013	0.007	0.004	0.030	0.091	0.030	0.039	0.157	0.079	0.029	0.032	0.144
Ind11	0.012	0.006	0.003	0.028	0.028	0.032	-0.042	0.090	0.034	0.031	-0.030	0.094
Ind12	0.015	0.008	0.004	0.034	0.092	0.030	0.042	0.159	0.085	0.029	0.038	0.150
Ind13	0.008	0.010	-0.014	0.026	-0.033	0.080	-0.235	0.086	0.004	0.069	-0.166	0.107
Ind14	0.011	0.006	0.003	0.027	0.056	0.029	0.002	0.120	0.032	0.031	-0.034	0.092
Ind15	0.029	0.016	0.000	0.063	0.054	0.059	-0.065	0.168	0.109	0.062	-0.015	0.229
Ind16	0.010	0.006	0.002	0.025	0.004	0.034	-0.073	0.065	0.015	0.031	-0.053	0.072
Ind17	0.011	0.007	0.002	0.026	0.012	0.042	-0.081	0.087	0.041	0.035	-0.036	0.107
Ind18	-0.008	0.010	-0.034	0.005	-0.086	0.050	-0.197	-0.002	-0.043	0.041	-0.136	0.025
Ind19	0.001	0.007	-0.015	0.013	-0.008	0.036	-0.091	0.054	0.012	0.031	-0.058	0.070
flood 0-2	-0.019	0.014	-0.056	-0.001	-0.216	0.061	-0.340	-0.103	-0.347	0.069	-0.478	-0.214
flood 2-4	-0.055	0.029	-0.124	-0.015	-0.340	0.071	-0.473	-0.201	-0.386	0.071	-0.519	-0.244
flood 4-6	-0.043	0.024	-0.101	-0.010	-0.376	0.071	-0.512	-0.234	-0.430	0.071	-0.561	-0.284
flood 6-8	-0.052	0.030	-0.128	-0.012	-0.447	0.076	-0.586	-0.293	-0.528	0.074	-0.661	-0.375
flood 8	-0.101	0.044	-0.203	-0.036	-0.552	0.065	-0.661	-0.410	-0.648	0.056	-0.743	-0.524

Table 5.4: **Marginal effects.** Marginal effects for firm with probability of opening in the 50% percentile and all neighbors open. The first column displays the variable symbol. For each quarter, columns 1, 2, 3, and 4 report the posterior means, posterior standard deviations, 2.5% and 97.5% percentile.

	Quarter 6				Quarter 8				Quarter 12			
	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%
rel size	0.072	0.043	0.002	0.163	0.222	0.088	0.028	0.363	0.187	0.072	0.027	0.311
rel size&flood	-0.075	0.052	-0.187	0.012	0.106	0.059	0.006	0.229	0.233	0.091	0.033	0.393
Size1	-0.163	0.045	-0.262	-0.086	-0.383	0.068	-0.513	-0.246	-0.161	0.071	-0.312	-0.034
Size2	-0.057	0.059	-0.182	0.044	-0.043	0.061	-0.163	0.081	0.131	0.068	0.002	0.266
Size3	0.182	0.078	0.039	0.343	0.100	0.058	-0.022	0.210	0.112	0.047	0.021	0.207
Ind1	0.620	0.177	0.173	0.845	0.176	0.128	-0.132	0.375	0.202	0.079	0.038	0.349
Ind2	0.426	0.129	0.162	0.655	0.210	0.073	0.066	0.348	0.113	0.075	-0.049	0.249
Ind3	-0.167	0.058	-0.292	-0.064	-0.186	0.174	-0.516	0.126	0.018	0.124	-0.265	0.213
Ind4	0.241	0.075	0.097	0.389	0.167	0.051	0.074	0.271	0.160	0.046	0.080	0.255
Ind5	0.152	0.075	0.015	0.306	0.052	0.061	-0.077	0.169	0.081	0.050	-0.018	0.181
Ind6	0.430	0.072	0.279	0.558	0.160	0.050	0.067	0.263	0.102	0.045	0.017	0.192
Ind7	-0.044	0.038	-0.117	0.034	-0.080	0.059	-0.200	0.030	-0.012	0.049	-0.118	0.076
Ind8	0.307	0.081	0.152	0.464	0.180	0.053	0.085	0.288	0.119	0.047	0.029	0.216
Ind9	0.172	0.082	0.028	0.344	0.032	0.070	-0.114	0.165	0.013	0.065	-0.128	0.131
Ind10	0.230	0.069	0.100	0.364	0.181	0.048	0.091	0.280	0.135	0.042	0.058	0.223
Ind11	0.183	0.069	0.057	0.326	0.052	0.056	-0.064	0.157	0.055	0.048	-0.043	0.148
Ind12	0.351	0.065	0.223	0.468	0.182	0.047	0.097	0.278	0.146	0.042	0.073	0.233
Ind13	0.132	0.117	-0.063	0.387	-0.040	0.115	-0.300	0.158	0.012	0.098	-0.215	0.171
Ind14	0.172	0.067	0.050	0.308	0.105	0.051	0.003	0.209	0.053	0.048	-0.050	0.144
Ind15	0.106	0.060	0.000	0.237	0.085	0.092	-0.099	0.263	0.150	0.086	-0.019	0.316
Ind16	0.139	0.060	0.032	0.267	0.010	0.055	-0.105	0.112	0.026	0.047	-0.074	0.113
Ind17	0.164	0.079	0.021	0.330	0.025	0.068	-0.121	0.148	0.067	0.055	-0.050	0.169
Ind18	-0.036	0.039	-0.109	0.046	-0.124	0.062	-0.248	-0.004	-0.061	0.055	-0.178	0.038
Ind19	0.016	0.045	-0.064	0.111	-0.010	0.057	-0.130	0.092	0.020	0.047	-0.080	0.108
flood 0-2	-0.073	0.034	-0.145	-0.009	-0.279	0.056	-0.384	-0.164	-0.405	0.056	-0.509	-0.292
flood 2-4	-0.129	0.037	-0.211	-0.068	-0.399	0.053	-0.497	-0.290	-0.440	0.054	-0.541	-0.330
flood 4-6	-0.191	0.058	-0.309	-0.081	-0.249	0.059	-0.361	-0.135	-0.213	0.057	-0.328	-0.109
flood 6-8	-0.125	0.039	-0.213	-0.062	-0.486	0.055	-0.594	-0.374	-0.558	0.055	-0.663	-0.445
flood 8	-0.159	0.044	-0.258	-0.085	-0.558	0.050	-0.653	-0.458	-0.644	0.047	-0.730	-0.547

Table 5.5: **Marginal effects.** Marginal effects for firm with probability of opening in the 75% percentile and no open neighbors. The first column displays the variable symbol. For each quarter, columns 1, 2, 3, and 4 report the posterior means, posterior standard deviations, 2.5% and 97.5% percentile.

	Quarter 6				Quarter 8				Quarter 12			
	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%
rel size	0.044	0.041	0.000	0.142	0.144	0.103	0.003	0.339	0.194	0.070	0.040	0.316
rel size&flood	-0.046	0.046	-0.162	0.006	0.069	0.059	0.001	0.205	0.240	0.087	0.051	0.396
Size1	-0.378	0.054	-0.476	-0.269	-0.382	0.059	-0.489	-0.260	-0.111	0.054	-0.235	-0.020
Size2	-0.067	0.067	-0.192	0.066	-0.050	0.068	-0.179	0.089	0.088	0.050	0.001	0.195
Size3	0.214	0.076	0.056	0.353	0.136	0.082	-0.025	0.290	0.113	0.059	0.015	0.247
Ind1	0.516	0.105	0.220	0.625	0.301	0.202	-0.133	0.595	0.487	0.245	0.026	0.871
Ind2	0.414	0.093	0.198	0.558	0.339	0.116	0.087	0.527	0.134	0.102	-0.024	0.362
Ind3	-0.317	0.079	-0.422	-0.112	-0.145	0.140	-0.352	0.169	0.049	0.101	-0.079	0.305
Ind4	0.272	0.066	0.134	0.393	0.245	0.071	0.101	0.383	0.203	0.068	0.079	0.348
Ind5	0.182	0.077	0.023	0.326	0.068	0.077	-0.078	0.223	0.073	0.051	-0.010	0.188
Ind6	0.424	0.048	0.321	0.508	0.232	0.070	0.091	0.361	0.097	0.051	0.011	0.207
Ind7	-0.068	0.056	-0.172	0.047	-0.078	0.055	-0.179	0.036	-0.003	0.027	-0.049	0.057
Ind8	0.330	0.065	0.194	0.446	0.269	0.074	0.119	0.406	0.123	0.060	0.020	0.253
Ind9	0.203	0.082	0.040	0.357	0.046	0.085	-0.109	0.218	0.017	0.043	-0.051	0.113
Ind10	0.262	0.061	0.138	0.375	0.269	0.065	0.137	0.388	0.148	0.056	0.050	0.268
Ind11	0.216	0.066	0.081	0.343	0.069	0.070	-0.064	0.213	0.047	0.042	-0.021	0.141
Ind12	0.294	0.036	0.222	0.362	0.222	0.045	0.130	0.307	0.072	0.015	0.043	0.101
Ind13	0.153	0.128	-0.100	0.392	-0.027	0.120	-0.246	0.224	0.028	0.068	-0.073	0.188
Ind14	0.205	0.066	0.073	0.329	0.142	0.071	0.003	0.285	0.044	0.041	-0.024	0.138
Ind15	0.187	0.093	-0.001	0.359	0.095	0.101	-0.096	0.295	0.109	0.075	-0.008	0.276
Ind16	0.170	0.063	0.045	0.291	0.016	0.063	-0.102	0.142	0.022	0.033	-0.034	0.095
Ind17	0.195	0.081	0.031	0.346	0.036	0.081	-0.116	0.195	0.061	0.052	-0.025	0.176
Ind18	-0.056	0.058	-0.164	0.061	-0.118	0.053	-0.212	-0.004	-0.026	0.023	-0.064	0.024
Ind19	0.019	0.061	-0.100	0.141	-0.007	0.062	-0.125	0.116	0.017	0.032	-0.036	0.089
flood 0-2	-0.116	0.049	-0.208	-0.015	-0.232	0.035	-0.297	-0.157	-0.091	0.011	-0.114	-0.071
flood 2-4	-0.222	0.041	-0.294	-0.137	-0.296	0.029	-0.350	-0.238	-0.093	0.011	-0.116	-0.073
flood 4-6	-0.196	0.044	-0.275	-0.103	-0.310	0.028	-0.364	-0.253	-0.095	0.012	-0.119	-0.074
flood 6-8	-0.213	0.047	-0.299	-0.117	-0.332	0.029	-0.388	-0.274	-0.098	0.012	-0.122	-0.076
flood 8	-0.287	0.034	-0.351	-0.219	-0.356	0.026	-0.404	-0.305	-0.100	0.012	-0.124	-0.078

Table 5.6: **Marginal effects.** Marginal effects for firm with probability of opening in the 75% percentile and all neighbors open. The first column displays the variable symbol. For each quarter, columns 1, 2, 3, and 4 report the posterior means, posterior standard deviations, 2.5% and 97.5% percentile.

	Quarter 6				Quarter 8				Quarter 12			
	Marg. eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%	Marg.eff	p.Std dev	2.5%	97.5%
rel size	0.035	0.037	0.000	0.133	0.121	0.100	0.001	0.329	0.119	0.087	0.001	0.292
rel size&flood	-0.036	0.042	-0.148	0.004	0.058	0.056	0.000	0.195	0.148	0.108	0.002	0.359
Size1	-0.109	0.014	-0.138	-0.083	-0.124	0.015	-0.155	-0.096	-0.113	0.037	-0.178	-0.032
Size2	-0.047	0.048	-0.152	0.031	-0.039	0.050	-0.151	0.045	0.096	0.042	0.002	0.165
Size3	0.077	0.022	0.028	0.115	0.060	0.033	-0.014	0.113	0.113	0.041	0.022	0.183
Ind1	0.119	0.020	0.078	0.151	0.091	0.071	-0.100	0.162	0.201	0.057	0.040	0.266
Ind2	0.112	0.018	0.073	0.143	0.113	0.028	0.043	0.155	0.113	0.069	-0.048	0.210
Ind3	-0.422	0.182	-0.756	-0.077	-0.162	0.158	-0.512	0.077	0.018	0.124	-0.271	0.197
Ind4	0.091	0.017	0.055	0.122	0.095	0.020	0.050	0.133	0.161	0.028	0.099	0.211
Ind5	0.068	0.024	0.011	0.108	0.031	0.038	-0.054	0.092	0.082	0.046	-0.019	0.162
Ind6	0.115	0.013	0.090	0.142	0.092	0.021	0.047	0.129	0.103	0.039	0.018	0.171
Ind7	-0.045	0.039	-0.127	0.023	-0.059	0.044	-0.154	0.020	-0.013	0.050	-0.117	0.076
Ind8	0.102	0.015	0.070	0.131	0.101	0.020	0.057	0.138	0.120	0.039	0.032	0.185
Ind9	0.074	0.024	0.020	0.114	0.018	0.045	-0.080	0.091	0.013	0.064	-0.129	0.121
Ind10	0.089	0.016	0.056	0.120	0.102	0.018	0.064	0.134	0.136	0.030	0.070	0.190
Ind11	0.078	0.019	0.037	0.112	0.032	0.034	-0.044	0.090	0.056	0.047	-0.042	0.139
Ind12	0.314	0.058	0.205	0.426	0.217	0.059	0.110	0.335	0.236	0.062	0.118	0.356
Ind13	0.053	0.047	-0.063	0.116	-0.035	0.086	-0.244	0.092	0.013	0.099	-0.222	0.162
Ind14	0.076	0.019	0.033	0.110	0.063	0.027	0.002	0.110	0.053	0.047	-0.050	0.135
Ind15	0.069	0.029	0.000	0.114	0.040	0.046	-0.068	0.110	0.105	0.054	-0.017	0.190
Ind16	0.066	0.020	0.022	0.102	0.005	0.036	-0.075	0.066	0.026	0.047	-0.073	0.108
Ind17	0.072	0.025	0.015	0.113	0.014	0.045	-0.089	0.084	0.068	0.053	-0.052	0.153
Ind18	-0.037	0.039	-0.119	0.029	-0.093	0.050	-0.200	-0.003	-0.062	0.055	-0.176	0.037
Ind19	0.007	0.031	-0.064	0.059	-0.009	0.039	-0.096	0.057	0.020	0.047	-0.081	0.103
flood 0-2	-0.082	0.042	-0.175	-0.008	-0.232	0.055	-0.342	-0.127	-0.413	0.054	-0.517	-0.301
flood 2-4	-0.194	0.055	-0.306	-0.093	-0.361	0.060	-0.478	-0.245	-0.449	0.053	-0.550	-0.345
flood 4-6	-0.161	0.052	-0.269	-0.066	-0.399	0.061	-0.519	-0.276	-0.488	0.053	-0.582	-0.380
flood 6-8	-0.184	0.061	-0.316	-0.076	-0.469	0.067	-0.599	-0.336	-0.567	0.052	-0.662	-0.456
flood 8	-0.302	0.062	-0.431	-0.183	-0.572	0.055	-0.671	-0.457	-0.653	0.036	-0.715	-0.577

Vita

Mihaela Craioveanu was born in Timisoara, Romania. She received her bachelor degree in finance and insurance from University of the West, Timisoara, Romania, in 2001. She received a Master of Science in economics from Louisiana State University in 2005. While in the Economics Department at Louisiana State University she was awarded an Economic Development Assistantship from 2004 to 2007. During this period she collaborated on various projects for the Division of Economic Development under the supervision of Dr. Dek Terrell. The last part of her dissertation was the result of one of these projects. Mihaela will complete the degree of Doctor of Philosophy in August 2008.