Incomplete Market Models and The Housing Market

Henrik Petri *

Olof Gustafsson †

2010-06-08

STOCKHOLM SCHOOL OF ECONOMICS

M.Sc. Program in Economics and Business

Thesis in Finance

Abstract

This thesis investigates the characteristics of the market for housing derivatives. We fit two incomplete market models to data using futures contracts on the S&P/Case-Shiller Composite 10 Index. First the index is assumed to follow a Geometric Brownian Motion, second we introduce seasonality through a Geometric Ornstien Uhlenbeck process with drift. Further good deal bounds are studied, and their numerical properties evaluated, for each specification. The GBM model explain 93.4% of the variation within the sample period with an average negative pricing bias of 1.2%. The GOU model explain 80% of the variation within the sample period. Through out the whole thesis we assume a constant market price of risk. For both models this turns out to be a reasonable assumption, even tough we find evidence in favor of a non constant lambda for each model. However due to the poor data quality the results are uncertain.

Tutor: Peter Englund

Discussants: Helena Blom and Marie Lundberg

Date of presentation: 15 June 2010

*21445@student.hhs.se

†21444@student.hhs.se

Acknowledgments

We would foremost like to thank our tutor Peter Englund for patience and valuable guidance. We have also benefited a lot from discussions with Thomas Björk. Lastly we would like to thank Lori Aldinger at CME group who kindly has answered all our questions regarding the data file.

Contents

1 Introduction

1.1 Background

Our interest in housing derivatives originates from a simple observation; In general it is possible to comfortably participate in a market without any particular view on the future development of prices. I.e. it is possible to over some time hold, produce or consume an asset, commodity or good without taking an implicit bet on spot market movements. For example a farmer is able to sell his wheat forward and lock in a certain price when the grain is still in the field. A steel plant may use certificates of delivery to manage the exposure to sudden price spikes in the electricity market. And an asset manager might use options to protect his portfolio from downside risk. Furthermore, in the financial markets the phenomena is taken to the extreme, where almost all risks are tradable. However there is one market where the concept of tradability of risk is still remarkably undeveloped, the real estate market. This fact is in some sense counter intuitive both for the real estate market in general and for the housing market in particular. Although house prices in most western countries have reverted from historical records, most people will still find themselves highly leveraged if deciding to buy instead of rent. This makes the decision the single most important from a portfolio perspective. Without any further analysis at this stage, there seems to be reason to assume that the average house owner would benefit from passing on some of the housing risk to other parts of the economy. At the same time we note that individuals in the rental market certainly would improve their portfolio selection by adding some. With those observations as a starting point we developed an interest in risk management tools for the housing market.

Before dwelling into the mathematical abstractions that are inevitable in the world of derivatives pricing it might be worth explaining some issues in a more heuristic way. As we will see later in this thesis the market for housing derivatives is an incomplete market. In order to understand the implications of this we must first state under which conditions a market is complete. In short a market is complete if every claim can be replicated. This means that the payoff profile of a derivative can be mimiced through taking positions in other assets on the market, the underlying and the bank account. In a complete market derivatives are hence redundant and as a consequence there exist unique derivatives prices. This means that we can derive an exact price for a claim, only by observing information incorporated in the price process for the underlying, and (almost) without any constraints on investor preferences. This is a very important feature of complete markets that distinguish them from the market for the underlying asset itself, which is generally incomplete. For example the price of a stock at a given point in time cannot be determined as easily as the price of a call option on the same. To value the share we must know, among other things, investors attitude towards holding undiversifiable market risk, and even in this setting we will not arrive at a unique price. However in order to price the call option we only need to impose the restriction that the market is arbitrage free. I.e that investors prefer more to less; 1 million sek to 1 sek, or a "free lunch" if such is served. The key to this result is that the price of derivatives must be driven by the same source of risk as the price of the underlying asset and that this risk may be perfectly hedged away through replication.

To make a market incomplete we only need to remove the tradability of the underlying asset. Since the source of risk that drives the price of derivatives no longer can be traded away a simple arbitrage argument is not valid in this setting. Instead we find ourselves in the situation where we need to know investor's risk preferences. Ie. the premium that investors demand for holding on to the non tradable risk that drive the price of the derivative. The way to do this is often to adopt a stochastic model for the underlying asset and derive solutions for the price of derivatives. However as explained these solutions will not be unique but depend on how investors price risk. This is captured in the model by an unknown parameter often referred to as the market price of risk. Using market prices from vanilla contracts this parameter is then inferred. This is commonly referred to as calibrating the model and once completed the model can be used to price more exotic claims.

It might be appropriate to mention that in practice almost all derivative markets are indeed incomplete. For example portfolio rebalancing exhibits transaction costs as a hedger at the very least will suffer from crossing bid offer spreads. In addition buying and selling cannot be done in an instantaneous fashion. As a consequence hedging will suffer from the curse of "buying high and selling low". Volatility is also often time invariant imposing the risk of those costs to rise over the replication period. Further, the underlying asset will most certainly pay some not completely deterministic dividend over the life of the contract. For more exotic derivatives risk exposures are even more complex. From a theoretical perspective the dimensionality of the pricing problem will grow very fast and it will become computationally untraceable trying to incorporate all these, and many more, issues. The way this is handled in practice is to adopt a model that has a reasonable level of sophistication and fit to market prices of vanilla contracts. The price of the derivative is then recalculated tweaking the models parameters and variables from their market values (or calculating partial derivatives if a closed form solution is available) in order to obtain sensitivities to different risk. The individual trader will then review the positions of his book and price those risks that are not easily hedgable, making exotic derivatives trading a mixture of art and science.

With these issues in mind we focus our attention on housing derivatives. We will use the information incorporated in futures prices in order to calibrate two incomplete market models to the housing market. It should be said that previous research in this particular area is limited as exchange traded housing derivatives is a fairly new phenomena. We thus aim to contribute to the literature not only by investigating model calibration but also through evaluating the appropriateness of our suggested models. The latter is equally important as the models are often cited in the literature on derivatives pricing. This is also true for the housing market where for example the benefit of housing derivatives for portfolio optimization has been assessed with the Black Scholes framework as an approximation. It should be stressed that by calibration we mean retrieving the parameter vector necessary to price a given claim in the market. This should be distinguished from the pricing problem where we also need to derive the pricing equation for the claim. We hence limit the scope of the assessment to retrieving the parameter vector and analyzing the model assumptions. As already briefly touched upon the market for housing derivatives is still in development and as we will See not a deep and liquid market. This will of course make the calibration task difficult and render in uncertain parameter estimates. We will therefore use the theory of good deal bounds to add some extra insights to our analysis.

1.2 Purpose

In this thesis we will investigate the pricing of derivatives on the S&P/Case-Shiller Composite 10 index (CSI index). We will do this by specifying two different incomplete market models for the CSI index. The first model is an ordinary Geometrical Brownian Motion, the second a Geometric Ornstein Uhlenbeck process with drift. Our objective is to calibrate the models using price data on futures contracts trading on the index. When completed, the models may in theory price any given claim in the market. As a further step we will also derive good deal bounds for futures contracts in these two different settings.

In short, the purpose of this thesis is hence twofold. First, we specify two models for the index that seem reasonable from a theoretical as well as practical point of view. Second we calibrate these models to market data and try to infer how well they are able to explain observed prices on futures.

1.3 Outline

The rest of this thesis will be structured in six sections. In the first section we present some theory and findings relating to the housing market in general. Next we describe the data set used. Section four and five are devoted to describing model specification and calibration techniques used for each model. We also present results. To keep arguments clear and tractable we choose to integrate relevant derivatives pricing theory in each model section. In part six we derive good deal bounds for each model and investigate their numerical properties. Section seven is devoted to discussion and conclusion.

2 Literature review

This section will present some fundamental findings from the housing market. It is intended for the reader to review in order to get an introduction to the special characteristics of housing derivatives dealt with throughout the thesis.

2.1 The housing market

Representing more than half of the U.S private capital stock and one third of private consumption the housing market is a fundamental driver of U.S wealth- and income distribution (Englund et al, 1999). Home equity typically constitutes the largest part of home owner's net worth leaving owner occupied households sensitive to fluctuations in house prices (ibid). For households below 30 years of age U.S data show a leverage ratio for housing of 3.0 (Vanini et al, 2006). Despite this fact, there are virtually no financial products for managing housing exposure available to the retail investor.

The consequences of these findings are twofold. From an investment perspective it leaves the individual house owner with a suboptimal portfolio allocation and without efficient means of unloading excess housing risk (Case & Shiller, 1993). Evidence from the Swedish market show that for longer holding periods low risk portfolios contain 15%-20% housing whereas efficient short term portfolios contain no housing investment (Englund et al, 2002). Households would therefore benefit significantly from being able to manage their housing exposure (ibid). Likewise individuals in the rental market could substantially improve their portfolio allocation with exposure to real estate investments (Englund et al, 1999). However these findings also have the potential of creating societal imbalances and costs as aggregate housing risk is not distributed to those investors that are best suited for taking it on (Case & Shiller, 1993).

2.1.1 Unloading house price risk

Several financial products have been suggested in order to mitigate the problem of risk allocation in the housing market, see for example Case & Shiller (1993), Schiller & Weiss (1999), Vanini et al (2006). Among these products are exchange traded option and futures contracts, index-linked mortgages as well as OTC contracts designed to offset housing exposure.

Index-linked mortgages can be designed in several ways. Common for them all is that they tie the periodic mortgage payment or the value of the outstanding principal to a predetermined house price index. By introducing positive correlation between interest payments and the local real estate market, the contract function as an effective hedge against house price fluctuations. The index-linked mortgage is thus beneficial to both borrower and lender as it decrease the periodic payment as well as probability of default during house market downturns (Vanini et al, 2006). However the development of a functional and liquid market is conditional upon the existence of trusted house price indices as well as an institutional derivatives market for housing risk (ibid).

Other types of insurance contracts that are suited for the OTC market have also been suggested. The rationale is the same as for index-linked mortgages as they provide a payoff to the policyholder in order to offset a downturn in the housing market. For this benefit the policyholder will pay a periodic or upfront premium. For example Shiller & Weiss (1999) suggest an insurance contract that only have a payoff when a loss is actually realised by the homeowner.

Exchange traded options and futures are well known risk management tools in equity markets, widely used by practitioners and with an extensive coverage in the literature. An investor seeking downside protection may use put options or short futures contracts to manage the risk of his portfolio. In contrast to the OTC market, liquidity, transparency, counter party risk and other transaction costs are reduced through the listing procedure. The development of more complex derivative products for the housing market as those discussed above is to some extent believed to be dependent upon a well functioning options and futures market (Shiller & Weiss, 1999).

2.1.2 Measuring house prices

The accurate measurement of house prices is a fundamental need in the creation of precise and trustworthy indices for derivative contracts to be based upon. This task is however not as straightforward as may be seen at first glance and there are several theoretical and methodological issues to consider (Englund, 2008).

If we contrast the equity and housing markets the problems that arise with respect to the latter become apparent. First the real estate stock consists of a large number of heterogeneous assets. If we claim that for example Ericsson A is trading at 75 SEK this is an unambiguous statement as holding any share of the stock give rise to the same financial claim on Ericsson AB. However claiming that a house in a specified geographical area is trading at a price of 50 000 SEK/m^2 give rise to several questions when inference on the housing stock is intended. The characteristics as well as the precise location of the property will have a great impact on the transaction price for the same. From this point of view it may even be questioned however it is meaningful to speak of a market price for a housing market on which an index level can be based. Further the trading volume in real estate markets is typically of much lower magnitude than in financial markets. For owner-occupied homes Swedish data show that only 3%-5% of the stock is trading each year (Englund, 2008). In connection with the heterogeneity of assets this also gives rise to potential selection biases where units with some specific characteristics are traded more frequently than others. For example, Englund (2008) suggest the potential problem of a lemons market. All these characteristics give rise to noise that influence transaction prices and obstruct meaningful aggregation of house price data.

There are however index methodologies developed for handling these issues. Two major index methodologies are used for owner-occupied homes. Hedonic methods use panel data regression to clear transaction price data from unit specific characteristics (Englund et al, 1999). The procedure however introduces explicit assumptions on the functional form of the price relation as well as assumptions on how the market value of different characteristics changes through time and what characteristics that are priced by the market (ibid). It is further a very data intense technique which requires high quality data.

To overcome these problems repeat sales indices are used. This methodology is today the most widely used in commercial and government applications (Englund et al 1999). By comparing transaction prices of identical housing units, at different points in time, no assumptions need to be made of the functional form of the price relation, or the time dependence of the price of unit specific characteristics.

2.1.3 S&P/Case-Shiller Home Price Indices

The S&P/Case-Shiller Home Price index series are the leading measure of U.S residential real estate prices and consist of 23 headline indices, indices of 20 metropolitan statistical areas and 3 composite indices (S&P, 2010a). The S&P/Case-Shiller National Home Price Index captures approximately 75% of the value of U.S residential housing stock and is published on a quarterly basis (ibid). For the purpose of this thesis we will examine the S&P/Case-Shiller Composite 10 (CSI Index) which consist of repeat sales data from single family homes in the 10 largest U.S metropolitan areas and is published on a monthly basis (S&P, 2010b).

The main variable used in order to calculate index values are price changes between two arms-length sales of the same single-family home. For each transaction in the market a search is conducted to find available information regarding previous sales of the same housing unit. If a search match is found the transaction pair is regarded as a repeat sales pair. Further a weighting scheme of the sales pairs is employed to construct the index. Since the index is intended to capture market trends less weight is given to observations for which price movements are considered not to be market driven or come from idiosyncratic physical changes to a property or neighborhood. The rationale behind the repeat sales procedure is hence to capture the movement of the housing market keeping unit specific characteristics fixed. For a complete technical description of the repeat sales algorithm used for the CSI index we refer the reader to S&P (2010b).

3 Data

Futures and options on the CSI index have been traded on the Chicago Mercantile Exchange (CME) since 2006. The data set consist of daily observations of CSI index futures and and monthly observations of the CSI index. The data set contains primary data from CME where the contracts are traded. Futures prices are observed from 2006-05-22 to 2009-08-03 in a total of 3910 observations. For the CSI index we use monthly observations from 1987 - 2009. On each trading session settlement prices for futures contracts with different maturities are observed. Expiration date for historical futures contracts are not disclosed by CME and may vary according to the occurrence of non scheduled trading days in end of the month. From examining expiration dates of currently outstanding contracts we observe that expiry is typically during the last trading week of each month. As a simplifying assumption we therefore set the expiration date of each contract to the last calendar day of the corresponding month. Figure 1 present the dimensionality of the futures data

Observation date	Contract	Price
5/22/2006	AUG06	231
5/22/2006	FEB07	235
5/22/2006	MAY07	236.2
5/22/2006	NOV06	237.6
5/22/2006	AUG06	228
5/22/2006	FEB07	231
5/22/2006	MAY07	232
5/22/2006	NOV06	227

Figure 1: Dimensionality of futures price data

A common practice in commodity markets is to approximate the spot price with the futures price of the shortest outstanding maturity. This is done since the spot market often is virtually non existing. As we face the similar problem with CSI index updated only monthly we will adopt this practice. Since we continuously use the contract with the shortest maturity as a spot approximation we will lose some observations for each contract. This reduces the set of observed futures prices to 3090 observations. The figure on page 50 in Appendix 9 shows the approximation of the spot price over time using this method.

As true for all large data set we experience some raw data error. The following measures have been taken in order to minimize the effect of these on our estimation. On 2006-08-30 the settlement price for the AUG06 future is 0 and the same is true for the NOV09 future on 2006-11-29. This is handled by omitting these observation dates from the estimation procedure. A troublesome fact is that the market for CSI index derivatives is illiquid. Of course this is far from optimal as it will introduce greater uncertainty into our estimates. Also it might be questionable whether investors is indeed pricing derivatives in a rational way in such a market. One can for example argue that possible arbitrage opportunities cannot be readily exploited and that prices should incorporate a liquidity premium. However for the purpose of this study we will not investigate these issues further. It should also be emphasized that it is the only available source of data and our options to turn elsewhere are hence limited. We also bear in mind that our data set contains the turbulent period of late 2007 and 2008. The great uncertainty prevailing in financial markets during those times will of course be amplified on a market of this nature.

4 GBM Model

To price derivatives on the CSI index, we must adopt a theoretical model that well describes the evolution of the index. As implied by the discussion above, the housing market in some aspects distinguish itself rather radically from financial markets. This makes it hard to justify the standard Black and Scholes Model (BS model) for valuing housing derivatives. Below we will describe some notable features of the housing market and how these affect the choice of process for describing the CSI index.

4.1 Model specification

In order to specify a good model for the CSI index it is essential to understand its characteristics, and hence the underlying sources driving the index process. A common practice, in different contexts, is to assume that the underlying housing index evolves according to a Geometrical Brownian Motion (GBM). De Jong, Driessen and Van Hemmert (2004) adopt a GBM when assessing the economic benefits for investors to have housing futures in their investment portfolio. This despite the fact that the housing market is commonly viewed as an illiquid, heterogeneous market where new information is slowly incorporated into prices. As documented by Capozza, Hendershott and Mack (2004) the housing market ,to some degree, depending on factors such as metropolitan area, constructions costs and population, exhibits serial correlation (positive) and mean reversion. Despite these findings we start out by adopting a GBM for the CSI index. Of course, assuming this simple framework will come at some costs.

The benefit of assuming a GBM (under the $P-measure$), is that if it turns out that the model explains variation in the observed prices well, we will have a simple and computationally friendly model at hand. Also, the index process will be identical to the well know framework of the BS model. As housing derivatives so far has attracted only limited interest from researchers it also seems natural to start out by adopting well known characteristics for the index process. We evaluate the power of this simple model in explaining derivative prices before extending the framework.

4.1.1 The Setup

As discussed above we will assume that the CSI index evolves according to an "ordinary" Geometrical Brownian Motion. If we denote the level of the index by X, the dynamics followed by X will be as follows:

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t)^P \tag{1}$$

where $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$ are constants determined endogenously by the market. The superscript P indicates that $dW(t)^P$ is a wiener process under the objective probability measure P. We will also assume that there exists a risk free asset on the market with the dynamics

$dB(t) = rB(t)dt$, where $r \in \mathbb{R}$denotes the deterministic short rate.

The aim is to price a given claim on the underlying process, i.e. in some sense the claim we are pricing "lives" in the market described above. Let $\mathcal{L} = \phi(X(T))$ be a given claim, where $\phi(-)$is some arbitrary deterministic function. Our aim is to find a reasonable price for this claim $\mathcal{L}$ living on the market (X, B). Thus we will be concerned with studying the price process $\Pi(t, \mathcal{L})$ of this claim.

However before doing that, we recall from earlier discussions, that there are some features of the housing market that distinguishes it from the ordinary Black and Scholes World. The following assumptions, which are made in the Black-Scholes Model, are not valid within the model above:

1. *Trading in the underlying asset.* This assumption is in fact one of the most crucial assumptions made in the BS model. Since the index X cannot be assumed to be an traded asset the usual arbitrage argument made when deriving the BS PDE does not apply. It is nonsensical to talk about buying for example x units of the index, since simply put it is impossible.

2. *Liquid market for the underlying.* This assumption is tightly nested with the assumption of tradability. As already recognized direct trading in the CSI index is not available. However even if it would be theoretically possible to replicate the index using real estate assets the market for real estate is very illiquid.

3. *Portfolios can be carried forward in time.* Fails as a direct consequence of (1). Although it is theoretically possible to carry a portfolio of an index proxy through direct investments in the housing market.

4. *Short selling is possible.* Short selling is not allowed in the housing market.

In short the market described above is an incomplete market. A heuristic argument for this, can be given by applying the Meta Theorem (See Björk 2009) to the market. I.e. the number of random sources $R = 1$ are greater than $N = 0$ the number of assets in the market. Loosely speaking, this means that there are not enough assets on the market to replicate our given claim $\mathcal{L} = \phi(X(T))$. Hence claims living in the market described above are not redundant like in the Black and Scholes world. The standard BS argument fails since we cannot balance a portfolio consisting of the derivative and the underlying index to cancel the randomness in the portfolio. In order to price derivatives on this incomplete market we must instead assume internal consistency between derivatives, i.e. if we take a market consisting of $m > 1$ derivatives as given, we have to be sure that there are no arbitrage possibilities between these derivatives. Differently formulated, it should be impossible to set up a self-financing portfolio[1] consisting of these derivatives, that earns a risk free rate of return that is greater than the return of the bank account. Thus by modifying the usual BS argument slightly we can derive prices for derivatives in incomplete markets. However these prices will not be unique. From the second fundamental theorem of arbitrage it follows that a market is complete if and only if there exists a unique martingale measure. Since the market described above clearly is not complete we will see below that the martingale measure is not unique. This is in fact one of the major problems that we

[1] See Björk (2009) for details

have to handle in this thesis, i.e. calibrating the model derived below as to find a martingale measure Q, such that we in theory can price any given claim living in the housing market in a unique fashion.

Using standard arguments in arbitrage theory one can show that a claim living in the housing market has to satisfy the following PDE:

$$\frac{dF}{dt} + (\mu - \lambda\sigma)x\frac{dF}{dx} + \frac{1}{2}(\sigma x)^2\frac{d^2F}{dx^2} - rF = 0 \tag{2}$$

$$F(T, x) = \Phi(x),\ \forall x \in \mathbb{R}$$

By applying the Feynman Kac representation theorem to the deterministic partial differential equation given above, we see that the price process of every claim in the market is given by: $F(t, X(t)) = e^{-r(T-t)}E^{Q}_{t,x}[\phi(X(T))]$ where the dynamics of X under the risk neutral Measure Q is: $dX(t) = \{\mu - \lambda\sigma\}X(t)dt + \sigma X(t)dW(t)$. Here $dW(t)$ is a wiener increment under Q. For a derivation of the results presented above we refer to the appendix section 7.1.1.

There are a couple of things to note regarding the dynamics of X under Q. The Market price of risk λ is determined by the market and there is a 1:1 correspondence between λ and the $Q - meaure$. Hence determining the market price of risk is the same thing as choosing an equivalent martingale measure. This also means that the price of a particular derivative will depend on the market price of risk. The market price of risk is not determined within the model. It is the market that determines λ. For a further discussion regarding this see for example Björk (2009). Thus in order to determine λ we must conduct an empirical study on the underlying market. Below we will outline the procedure chosen to determine the Q dynamics of X.

By now, we are mainly interested in deriving a pricing formula for futures prices in the market described above. From standard theory (see for example Björk (2009)) we know that the price process for futures contracts is given by $F(t, T) = E^Q[\mathcal{L}] = E^Q[X(T)]$. To compute $E^Q[X(T)]$ we solve an simple ordinary differential equation (See Appendix) and arrive at the following expression for the future price process.

$$F(t, T) = X(t)e^{(\mu - \lambda\sigma)(T-t)}$$

This expression for the price will be used when calibrating our model to the data below. When we computed the price process for a future contract, we assumed that λ is an arbitrary constant. Since the market is incomplete, λ is not known. We will in this thesis be interested in determining λ, or from a mathematical point of view we want to specify the equivalent martingale measure Q, so we in theory are able to price any other derivatives on the market. Below we will present the method adopted in order to do that.

4.2 Model Calibration

The price process of X under the equivalent martingale measure Q is defined as the solution to $dX(t) = \{\mu - \lambda\sigma\} X(t)dt + \sigma X(t)dW(t)$. If we can specify the parameters of the X -dynamics under Q, we will be able to price any other derivative on the market. In order to do that we have to specify the parameter vector $\xi = (\mu, \sigma, \lambda)$. One approach to do this is to calculate theoretical prices of future contracts by guessing the parameter-vector ξ, compare these "guessed" prices to those observed in the market, and then adjust them if it turns out that our guess was wrong. In this special case we do not have to guess the whole parameter list. Recall that the price process of a future contract was given by $F(t,T) = X(t)e^{(\mu-\lambda\sigma)(T-t)}$. If we define $\theta := \mu - \lambda\sigma$, we see that if we can determine a value of θ that can be considered a "good guess", then the drift of the X dynamics under Q will be specified. The upside of just specifying θ is that it relieves us from estimating μ, which usually is estimated with wide confidence intervals. In whatever way, we must still estimate σ to be able to fully specify the Q-dynamics of X. We first show how we estimate θ.

4.2.1 Estimation of theta

To specify θ, we have to calibrate our model to a set of observations. In the literature this is commonly referred to as an inverse problem. In contrast to the pricing problem, where one is interested in pricing derivatives given some model parameters, here the interest is in backing out the risk neutral parameters of the model given observed prices on some class of derivatives. In our case futures prices. The general idea is the following. Given a vector of observed prices $\mathbf{F}^{obs} = (F_1^{obs}, ...F_n^{obs})$, one is aiming to find a parameter value θ, such that the vector of theoretical prices $\mathbf{F}(\theta) = (F_1(\theta), ...F_n(\theta))$ is as close as possible to the observed prices. In other words the goal is to minimize the distance between the two vectors. This distance is usually taken to be the euclidean norm in $\mathbb{R}^n$.

Before we state the problem, which has to be solved, we define all variables of interest. We let Θ denote the parameter space, which in this case is just $\mathbb{R}$. We also index each observation date t by i and each maturity date T by j, where $i, j \in \{1...n\} \times \{1..m\}$ and such that $0 = t_{1,} < .. < t_i < .. < t_n = S$ and $0 < T_1 < .. < T_j < S < .. < T_m$. Hence, $[0, S]$ is the time interval during which our data sample is drawn. 0 is the first observation in our sample and S is the last day on which we can observe any prices. We now state the problem:

$$min_{\theta\in\Theta}||\mathbf{F}^{obs} - \mathbf{F}(\theta)|| = \sum_{i,j}(F^{obs}(t_i, T_j) - F(t_i, T_j, \theta))^2 = \sum_{i,j}(F^{obs}(t_i, T_j) - X(t_i)e^{\theta(T_j - t_i)})^2 \quad (3)$$

where $||(-)||$ is the euclidean norm in $\mathbb{R}^n$and where we in the last two sums, sum over all i and j. This is nothing more than an ordinary nonlinear least squares problem, which in general requires an iterative search algorithm to be solved.

However by taking the logarithm of each observation, and in that way assuming a slightly different error structure, the problem becomes the following:

$$min_{\theta \in \Theta} \sum_{i,j} (ln(F^{obs}(t_i, T_j)) - ln(X(t_i)) - \theta(T_j - t_i))^2 \tag{4}$$

Since this is a *linear* least squares problem (LLP) it has a closed form solution, which is in contrast to the original problem stated above. It can also be shown that the solution to this problem is a global minimum. The linear problem above, is commonly referred to as a regression through the origin. By minimizing the distance between log prices, we will give different weights to the observed values compared to the original minimization problem. In effect, we are making different assumptions about the error structure. In Problem (4) we assume that the error term is multiplicative, while it is assumed to be additive in Problem (3). Thus we will generally not obtain the same solutions when solving these two problems. The minimization problem we choose to solve depends on our assumptions about the residuals.

Solving (4)is easier than solving (3), because we assume an multiplicative error structure in (4). Since they generically yield different estimates of θ, we would like to solve both problems and compare the results. If we try to solve the non linear problem with a suitable search algorithm, like Newtons Method, a problem that arises, is which value we shall use as an initial guess in the search algorithm. One good candidate is the value of θ, obtained when solving problem 2.

To overcome the problem of getting stuck in a local minimum when solving the non linear problem, we choose a range of starting values, and compare the values of the goal function in the different cases.

It may also be worthwhile to comment on the fact that when transforming problem (3) to a linear problem, inference on θ becomes possible if we make some additional assumptions[2]. Interpreting the results may on the other hand be troublesome since we assume a different error structure as opposed to the original problem.

4.2.2 Estimation of sigma

The sample of Index observations is collected once every month. Thus if we order the sample by time, the time difference between two sample points is $\frac{1}{12}$.

Probably the simplest approach to estimate σ is to note that if one define $U_t := ln(\frac{X_t}{X_{t-1}})$ then $U_t \sim N((\alpha - \frac{\sigma^2}{2})(\frac{t}{12} - (\frac{t-1}{12})); \sigma\sqrt{\frac{t}{12} - (\frac{t-1}{12})}$, i.e. $U_t \sim N((\mu - \frac{\sigma^2}{2})(\frac{1}{12}); \sigma\sqrt{\frac{1}{12}})$. This follows directly from $X(t)$ evolving according to a GBM under the physical measure P.

If we let u_t denote an observation of the random variable U_t, which has a standard deviation of $\sigma\sqrt{(\frac{1}{12})}$,then we can estimate the variance of U_t by the sample variance $s(u_t)^2 = \frac{1}{n-1}\sum(u_t - \overline{u})^2$, where $u_t = ln(\frac{x_t}{x_{t-1}})$. Note that u_t $t \in \{0, 1...n\}$ can be regarded as a randomly drawn sample. This follows from the independent increments property of the Brownian motion. From statistical theory we know that $s^2(U_t)$ is an unbiased estimator of σ^2 i.e. $E[s(U_t)^2] = \sigma^2$. However $s(U_t)$ is not an unbiased estimator of σ, but since our sample set is quite large, and the estimator $s(U_t)$ can be shown to be asymptotically unbiased, this gives support to using it as an estimator.

[2]The standard OLS assumptions.

We also see that the variance of U_t is $\frac{\sigma^2}{12}$. Thus to get an estimate of σ^2 we need to multiply s^2 by 12. [3]. It can be shown that this estimator also is the maximum likelihood estimator of σ^2. We also calculate a confidence interval for sigma. We know that $\frac{(n-1)s^2(\mathbf{X})}{\sigma^2} \sim \chi^2(n-1)$, hence a confidence interval for sigma assuming that $\mu$is unknown is:

$$I_\sigma = \left(\frac{\sqrt{(n-1}s}{\sqrt{(\chi^2_{\gamma/2}(n-1)}}, \frac{\sqrt{(n-1}s}{\sqrt{(\chi^2_{1-\gamma/2}(n-1)}} \right) \tag{5}$$

Where γ is the confidence level, n denotes the number of observations in the sample, and χ^2_γ denotes the gamma-quartile of the Chi-Square distribution.

When interpreting the confidence interval of the standard deviation, one should remember, that it is calculated on the premise that our model is correctly specified. Thus even though we would expect the confidence interval above to cover the true σ^2 in 95 out of 100 times (if $\gamma = 0.05$), it is conditional upon the model being the correct specification of the index dynamics.

4.3 Calibration results

4.3.1 Full sample

From the procedure outlined above we calculate an estimate of the unknown parameter θ. We use data from 3090 daily price observations of futures contracts to arrive at an estimate of $\theta = -0.045$ with a standard error equal to 0.0005 when solving the LLP. The estimate obtained through the non linear estimation is almost identical to the estimate from the LLP estimation.

Assuming a multiplicative error structure has the benefit that statistical inference on θ is easy. The estimate of θ, obtained through the LLP, is unbiased and consistent assuming that the independent variable and the residuals are independent, and that the error terms have finite variance. When estimating theta we did not make any explicit assumption about the distribution of the error term. This is necessary in order to do statistical inference. Unfortunately it turns out that the normal assumption of the residuals is a quite strong assumption.[4] In figure 19 (Appendix) it is also seen that the residuals on average are clustered around ≈ -0.02. One can show that the average of the residuals is almost always different from zero when running a regression through the origin. This is because generally a regression through the origin does not provide the best fit to data.

In order to assess the validity of the assumption of a time consistent parameter value, we compute market implied theta for each price observation. I.e we solve for theta individually for each observation. We plot the daily estimates of theta against time for all future contracts individually. Figure 3 and 4

[3] In effect we are calculating the sample variance of the random variable defined as $V_t = \sqrt{12}U_t$, which has a variance of σ

[4] See figure 19 in the Appendix. It seen that distribution of the residuals is negatively skewed

show how theta vary with time for the FEB08 futures as well as the distribution of the daily implied theta calculated for all contracts.

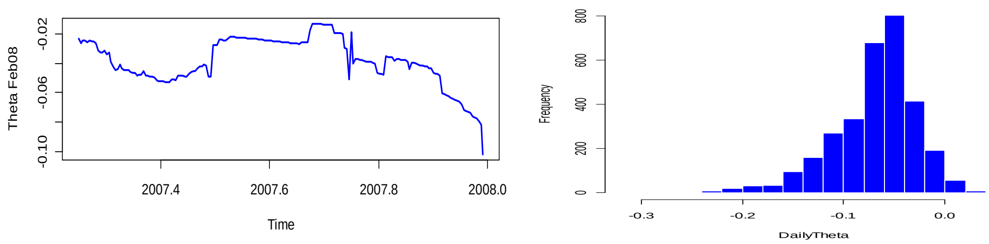

Figure 2: Theta for FEB08 contract over time (left) and histogram of daily theta (right)

As apparent from the figures theta is not constant through time. This is neither to expect as it would imply a perfect fit of the model. What is more interesting is that the distribution of theta seems to be fairly concentrated around a mean of -0.06 (stdev 0.04) suggesting that time consistency may on average be a not too bad assumption when aggregating the information from all contracts. Another reason not expecting a constant θ, is that we use future prices as spot proxies. Thus, we would expect a non constant θ, even tough our model would be correctly specified.

Next we evaluate how well the model is able to describe the futures prices in the data set given the optimized value of $\theta = -0.045$. We estimate how much of the variance in true prices that the model is able to capture on a daily basis. With an R^2 of 93.4% the model seems to have high explanatory power within the sample. This number should be interpreted with caution, since when omitting the intercept term, the usual analysis of variance decomposition $SST = SSR + SSE$ is generally not true.[5] We compute R^2 as:

$$R^2 := 1 - \frac{SSE}{SST} = 1 - \frac{\sum(ln[F^{obs}(t_i,T_j)] - ln[F(t_i,T_j\theta)])^2}{\sum(ln[lnF^{obs}(t_i,T_j)] - ln[F^{avg}])^2}$$

where $ln[F^{avg}]$ is the arithmetic average of all observations. We also compute the daily relative pricing error defined as

$$\frac{\mathbf{F}^{obs} - \mathbf{F}(\theta)}{\mathbf{F}^{obs}}$$

The rationale behind this measure is of practical as well as theoretical nature. So far we have a statistical measure of how well the model is able to capture the variance in the sample when investigating *log prices*. However we would like the pricing error within the estimation period to be of small economical significance when it comes to untransformed prices too. If the pricing error is large in relative terms it would be troublesome to argue why the model even should be considered for pricing other contracts.

[5] SSE, SSR and SST are the common abbreviations for explained sum of squares, residual sum of squares and total sum of squares.

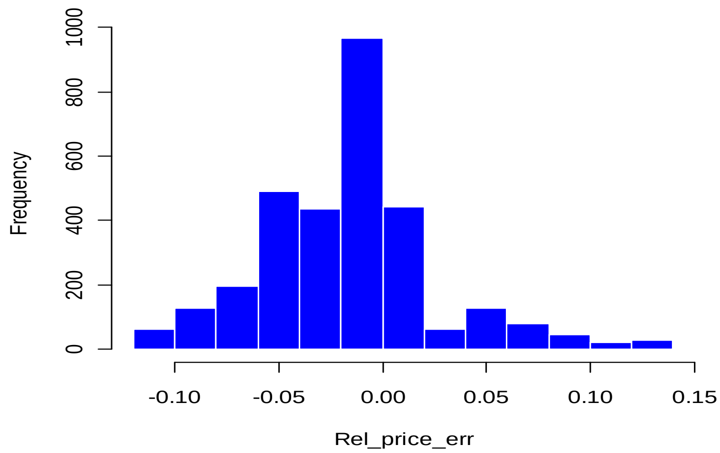

Figure 3: Histogram of relative pricing error.

Figure 5 show the distribution of the relative pricing error with a mean of -1.2%. This is of course a bit troublesome as it suggest that our model on average has a negative pricing bias. As apparent from the histogram the pricing error is not normally distributed. Inference about the significance of the estimate is therefore uncertain given that we do no know the true population distribution. To further investigate this we look at the distribution of $\mathbf{F}^{obs} - \mathbf{F}(\theta)$.[6] Although the observations line up fairly well along the 45 degree line, the normality assumption is rejected at every reasonable level of significance. This is also true even when we omit outliers. On the other hand even if the pricing error is significantly different from zero the bias is economically small considering the nature of the market. We also analyze each futures contract in the sample individually in order to see if the model fit any specific contract in the sample better. The following graph show pricing error for AUG07 and NOV11 futures respectively. As apparent the pricing error differs substantially between contracts within the sample.

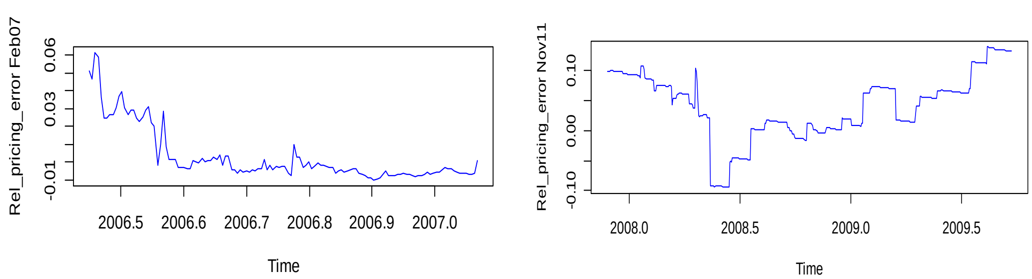

Figure 4: Relative pricing error for AUG07 (left) and NOV11 futures

[6]see appendix for histogram and qq plot

4.3.2 Sub sample

As discussed above we see that theta seems to be fairly consistent over time. We would however like to investigate how the optimal estimate of theta depends on the estimation period. In order to do this we divide the full sample consisting of 3090 observations into a sub sample of 2090 observations. We carry out the same steps as above for the 2090 observations which gives an estimate of $\theta = -0.047$. We now use this to make out of sample predictions for the remaining 1000 observations. We also calculate relative pricing error and get a mean of -1.08%. Figure 21 (Appendix) show the absolute difference between relative pricing error for the NOV11 contract using in sample and out of sample predictions.

From the graphs we can see that the pricing error seems to be consistent for in and out of sample pricing. This is although not completely unexpected recognizing the small difference in the optimal theta for the different estimation periods. Also the futures pricing model depend in a continuous way on theta. This can hence not yield us to draw the same conclusion regarding more complex contracts. However for out of sample prediction the R^2 drops dramatically to only 5%. One should remember that the general market environment prevailing for this period was very volatile. It would hence be naive to expect parameters estimated during preceding time periods to be valid for this period.

Apart from describing the risk neutral drift of the index process theta will determine the shape of the futures curve. Our estimate of a negative theta imply that futures prices in our sample are on average backwardated, meaning that the futures price decline with the maturity of the contract. In a complete market this is often related to market conditions where there is a shortage of the underlying, and/or a convenience yield higher than the cost of carry. However the arbitrage arguments that, together with market characteristics, drive the shape of the futures curve are not valid for our incomplete setting. This follows directly from the non tradability of the CSI index.

4.3.3 Estimation of sigma

We use monthly observations of the CSI index from 1 Jan 1987 to 1 Nov 2009 to arrive at an estimate of $\sigma = 3.13\%$ with a 95% confidence interval of $2.67\% \leq \sigma \leq 3.74\%$. We also perform the estimation on the approximated spot price on the time period 2006-05-22 to 2009-08-03 yielding the following results; daily observations, $\sigma = 8.10\%$, monthly $\sigma = 7.2\%$. The estimate from the CSI index is 3.91% for the same time period.

Further we test the assumption of time invariant volatility. This is done by performing Levene's test of equal variance on the logarithmic returns. One advantage of using Levene's is that it does not require the logarithmic returns to be normally distributed. We divide the return series for CSI index in three sub samples. Sub sample 1 consisting of 1987-01-01 to 1994-09-01 (92 obs), sub sample 2 consisting of 1994-10-01 to 2002-05-01 (92 obs) and sub sample 3 consisting of 2002-06-01 to 2009-11-01 (90 obs). Appendix section 7.2.1 show the output for Levene's test between the three different groups as well as for the first and second half of the sample. As apparent from the table the null hypothesis of equal variance between different time periods is rejected at every reasonable level of significance for all pairs except sub sample 1 and 2. This is however not completely unexpected as we recall the sharp rise and fall in US housing prices during the last decade, implying that this was a period with unusual high volatility. For sub sample 1 and 2 the result is however quite interesting. We cannot reject that the volatility in the housing market was constant from 1987 - 2001, a time period of significant length. This does not mean that the volatility during this period was constant. It is however relieving that our assumption of a constant sigma can not be rejected during this period.

From the above we see that the assumption of time invariant volatility is rejected. This is a well documented phenomena in asset returns and it is also present in the CSI index. On the other hand we also note that for substantial time periods the assumption of a constant σ in the GBM model may be appropriate. There are techniques of handling time varying volatility within the model. Examples are local and stochastic volatility models.

4.4 Model evaluation

4.4.1 Testing for normality in log returns

Below we investigate whether log returns can be assumed to be normally distributed, as predicted by the model. If this is the case, it would yield some support to the Gaussian specification of the noise term in (1). When conducting the test for normality we use data on the CSI index over the period 1987 to 2009. We divide the whole sample into two smaller sub samples. The first sub sample is reaching from the start of the index in 1987 until 2006, and the second sub sample is collected during the period 2006-2009. The reason for subdividing the sample like this is that we want to capture the effect of the rise and fall in house prices during the last decade. As is well known for stock markets, in times of economic downturns, the usual laws in financial markets do not apply. For example it is known that markets are very volatile in these periods and that the normality assumption is rejected.

When testing for normality we conduct the following test for the whole sample and each sub sample:

$$H_0 : ln(S_t/S_{t-1}) \sim Normal$$

$$H_A : ln(S_t/S_{t-1}) \sim non - Normal$$

The empirical distribution for the whole sample of log returns seems to be negatively skewed, i.e. negative returns seems to be larger (in absolute value) than if returns would have been drawn randomly from a normal distribution. When conducting the Shapiro wilk test for normality in R[7], we get a p-value of approximately 0. Thus the null hypothesis of returns coming from a normal distribution is rejected. However by inspecting the graphs on the next page it seems like this is mainly due to the skewness of the Index-returns. The comparatively large negative returns can be seen in all of the three graphs in figure 8 on the following page. By neglecting these observations we would by no means be able to reject the hypothesis of returns coming from a normal distribution.

In figure 9 we show similar graphs as those in figure 8, but for the last subsample (i.e. the period 2006-2009). It is apparent from these graphs that the normal assumption does not hold. The distribution is far more skewed to the left as compared to the whole sample.

An interesting fact is that when evaluating the normal assumption for the period prior to 2006, it is very hard to reject the implicit assumption of log returns coming from a normal assumption. As the graphs display, the normal assumption seems to fit the data well. For example the theoretical and sample-quantiles line up on a 45° line in the qq-plot as predicted by the model. When conducting the wilk test for this sub sample, we get an p-value of 0,31 in R which means that we cannot reject the null hypothesis of returns being lognormaly distributed.

[7] Statistical Package, see references

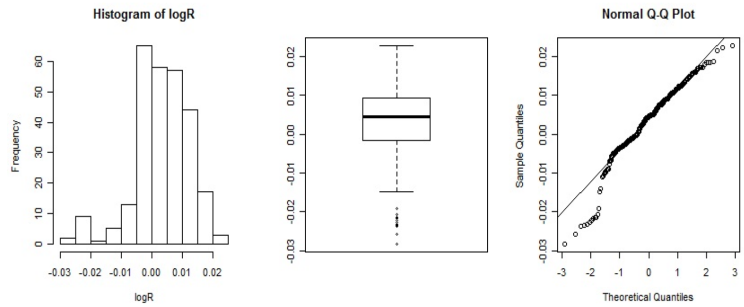

Figure 5: Evaluating the normality U_t during the period 1987-2009

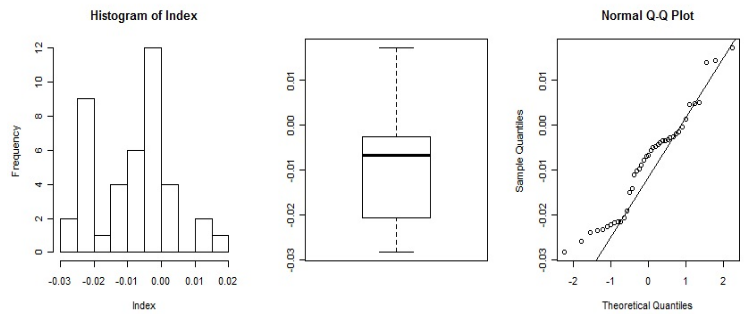

Figure 6: Evaluating the normality of U_t during the period 2006-2009

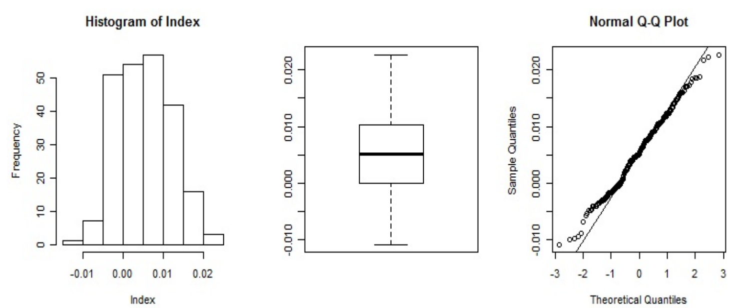

Figure 7: Evaluating the normality of U_t during the period 1987-2006

4.4.2 Testing for mean reversion in log index levels

By applying Ito's lemma to the process defined by $Z(t) := ln[X(t)]$, we get the process followed by log index levels. This process is an arithmetic Brownian motion with drift $\mu - \frac{\sigma^2}{2}$ and diffusion σ. The arithmetic Brownian motion is sometimes referred to as being difference stationary. One then refers to the discrete counterpart of the arithmetic Brownian motion, i.e. "the random walk". By solving the SDE satisfied by $Z(t)$ we get:

$$Z(t) = Z(s) + \left(\mu - \frac{\sigma^2}{2}\right)(t-s) + \sigma(W_t - W_s)$$

Its seen that both the variance and the mean depend on $t-s$, and hence the process is non stationary. A non stationary process is not mean reverting, which differently put means that shocks to the process do not decay. However by plotting log index levels against time[8] we get strong indications that the log index process is stationary with a deterministic trend, i.e. it seems to revert to a linear trend. This could however also be a spurious result, as is common phenomenon for many time series having unit root.

To test the hypothesis that log returns are non stationary, or follows a random walk, we perform a standard dickey fuller test with a constant term. We run the regression

$$\triangle z_t = \alpha + \Phi z_{t-1} + \varepsilon_t \tag{6}$$

where $\triangle z_t = z_t - z_{t-1}$ and test whether Φ is significantly smaller than zero. In other words we test the null hypothesis that the log index series z_t is non stationary, against the alternative hypothesis of a stationary series with a deterministic trend. When performing this test we cannot reject the null hypothesis at any reasonable level of significance. Estimates and standard errors are reported in the Appendix. The power of this test is unfortunately very low, furthermore the test is sensitive to the error terms being white noise. When examining the correllogram of the residuals in figure 17, it is seen that the residuals being white noise seems to be a bad assumption.

To accommodate the issues with the dickey fuller test, we also apply the augmented dickey fuller test. We report the results of these tests in the Appendix. As can be seen we are only able to reject the null hypothesis of non stationary returns, when the number of lags are equal to 9.

[8] Figure attached in the Appendix.

5 Geometric Ornstein Uhlenbeck Model

In this section the GBM framework is extended to account for mean reversion and linear trends in house price indices. We will at the outset specify the model. We then take this model to data and calibrate its parameters. The calibration procedure will differ from that of the GBM model, which also implies that the results of the two model calibrations will not be directly comparable. In the end we will discuss the empirical results of the calibration of the GOU model, and comment on how they compare to the GBM model.

5.1 Specification of the GOU model.

Although the results in section 1, regarding the GBM specification of the CSI index showed some explanatory power of prices, we cannot deny the fact that there seems to be a strong linear trend that the logarithm of the index revert to. Fabozzi et. al. (2010) suggest that an Geometric Ornstein Uhlenbeck (GOU) process with drift should be adopted to account for mean reversion in house price indices. We chose to apply a GOU process with drift that has the following representation under the objective measure:

$$dZ(t) = \{\{b(\varphi(t) - Z(t)) + \frac{d\varphi(t)}{dt}\}dt + \sigma d\widetilde{W}(t) \tag{7}$$

Using the dynamics above we choose to model log prices directly. The reason for doing this is that it is computationally easier to calibrate the model if we directly model the logarithm of the index levels. However the process followed by the index $X(t)$ under this representation, could be found by applying Ito's lemma to the process defined by the function $f(Z(t),t) = e^{Z(t)}$. In the model specified above, we model the long run mean reversion level of log prices, as a deterministic function of time $\varphi : \mathbb{R}^+ \to \mathbb{R}.^+$The parameter, $b > 0$ can be interpreted as the mean-reversion speed. As the notation suggests, and for reasons that will become apparent later, it is chosen to be a constant. The long run mean reversion is dependent on t. We have also added the term $\frac{d\varphi(t)}{dt}$ in the SDE above. This is because, if not, long run index levels may not revert to long run mean reversion levels (Dornier 2004).

We have not yet chosen an explicit specification for $\varphi(t)$. As mentioned above there is evidence (Fabozzi et.al 2010) that house price indices exhibit log linear trends. These findings are also supported for the CSI index in section 4.4.2. and is further visualized in figure 23 in the Appendix. Thus we deem a specification of the form $\varphi(t) := pt + q$ as appropriate.

There are however dangers of adopting this specification. By assuming a log linear trend it is possible that one over/under estimates the actual change in the index. For example, during the economic turbulence post 2006, it is not implausible that a GOU model with linear trend would underestimate increases in the index, by forcing Index levels to stay close to their long run mean. To mitigate this issue, one solution is to choose another specification for $\varphi(t)$. We will elaborate more on this later. From now on we let $\varphi(t) := pt + q$, as suggested by Fabozzi et. al. (2010)

In general, the advantages of the GOU model are similar to those of the Geometrical Brownian Motion. The GBM model is nested in the GOU model. This is seen by setting $b = 0$ and letting $\varphi(t) :=$

$pt + q$. Precisely as the GBM model it is tractable from a computational perspective. For example futures prices are easily computed. Another feature of the GOU model is that it, rests on the normal assumption of log returns. Hence, the model evaluation of the GBM model, concerning the normality of log returns, also applies to the GOU model. If we assume that the index evolves according to a GOU process, the risk neutral dynamics of Z, assuming a constant λ "market price of risk", will be as follows:

$$dZ(t) = \{b(pt + q - Z(t)) - \lambda\sigma + p\}dt + \sigma dW(t) \tag{8}$$

where $dW(t)$ is a wiener increment under the risk neutral measure (compare to $d\widetilde{W}(t)$ above which is under P). It can be shown that the solution to (8) is given by: $Z(T) = pT + q_1 + (z(t) - pt - q_1)e^{b(t-T)} + \sigma e^{-bT}\int_t^T e^{bu}dW(u)$. One can also show that $Z(T)$ is normally distributed, and that the first and second moments of $Z(T)$ conditioned at the information at time t are given by[9]:

$$E^Q[Z(T)] = pT + q_1 + (z(t) - pt - q_1)e^{b(t-T)} \tag{9}$$

$$Var^Q[Z(T)] = \frac{\sigma^2}{2b}(1 - e^{2b(t-T)}) \tag{10}$$

We would now like to calculate future prices in the market, where the underlying object is governed by a GOU process. The price of a future contract is given by: $F(t,T) = E^Q[X(T)]$. If we use that $Z := ln[X(t)]$, we can calculate the price of a future contract. Since X is log normally distributed, its first moment is given by $E[X(T)] = exp\left(E[Z(T)] + \frac{Var[Z(T)]}{2}\right)$. Thus the future price, when the underlying process is a GOU-process, is:

$$F(t,T) = exp(pT + q_1 + (ln[x(t)] - pt - q_1)e^{b(t-T)} + \frac{\sigma^2}{4b}(1 - e^{2b(t-T)})) \tag{11}$$

where $q_1 := q - \lambda\frac{\sigma}{b}$. Note that q_1 is not known, since lambda is unknown, and therefore we can not calculate prices in the incomplete market described by the dynamics above. In order to do that we must estimate λ. That is however not enough, to fully specify the Q-dynamics of $Z(t)$ we must specify the whole parameter list $\nu = (\sigma, b, p, q, \lambda)$. In the following section, we will in detail explain, how the calibration of the GOU model will be conducted.

[9] Derivation of the solution to (8), and the moments of $Z(T)$, is shown in the Appendix

5.2 Model calibration

The calibration of the GBM model turned out to be relatively easy, since by defining $\theta := \mu - \lambda\sigma$, it relieved us from estimating μ. To calibrate the GOU model, we have to estimate the whole parameter list in order to specify the Q dynamics of our model. We will carry out our "calibration scheme" in two steps:

1. We estimate the parameter list $\nu_1 = (\sigma, b, p, q)$, describing the model under P, by the Maximum Likelihood Method.
2. We then use this parameter list as an input in a minimization problem corresponding to the one that we solved for the GBM model, and in this way obtaining an estimate of λ.

5.2.1 P-measure parameters

The process that we are observing is according to our model assumptions defined as the solution to the SDE given by (7). Note that we now our trying to estimate the parameter list v_1 directly under P. The SDE in (8) is solved in a similar manner to the one solved in (7) (see Appendix for details). As mentioned above, the conditional distribution of $Z(T)$, given the information generated by the internal filtration up to time t is normal, with moments as in (9)and (10). Thus, in the case of the Geometric Ornstein Uhlenbeck process we have explicit expressions for the transition densities , i.e. the conditional density functions from one given period s to another period t s.t. $t > s$. The observations on the CS Index Z are collected at equidistant discrete points in time $0 < t_1 < t_2 < ... < t_n$,where $t_i - t_{i-1} = 1/12$. Since the underlying process satisfies the so called Markov property and the diffusion term is Gaussian, we can arrive at an explicit expression for the Likelihood function by an iterative application of Bayes Theorem (See Appendix):

$$L(\nu_1) = \prod_{i \in I} p(Z_{t_i} | Z_{t_{i-1}}; \nu_1) \tag{12}$$

where $I = \{1, 2, ..., n\}$. The transition density is normal, with mean and variance as shown in (9) and (10), i.e:

$$p(Z_{t_i} | Z_{t_{i-1}}; \nu_1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(z_{t_i} - \mu_i)^2}{2\sigma_1^2}} \tag{13}$$

where $u_i = pt_i + q + (z_{t_{i-1}} - pt_{i-1} - q)e^{b(t_{i-1} - t_i)}$ and $\sigma_1 = \frac{\sigma^2}{2b}(1 - e^{2b(t_{i-1} - t_i)})$. Observe that $\sigma_1$has the sub-script 1, not i. This is to indicate that σ_1 is independent of i, and to distinguish it from σ, which is one of the parameters in (7). (note that σ_1 is independent of time since we have observations sampled at equidistant points in time). We set $\Delta := t_i - t_{i-1}$. The explicit expression for the log likelihood function now becomes:

$$l(\nu_1) := ln[L(\nu_1)] = -n\frac{ln(2\pi)}{2} - nln(\sigma_1) - \frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(z_{t_i} - u_i)^2 =$$

$$= -n\frac{ln(2\pi)}{2} - nln(\sigma_1) - \frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(z_{t_i} - (pt_i + q + (z_{t_{i-1}} - pt_{i-1} - q)e^{-b\Delta}))^2 \tag{14}$$

The goal is to find the parameter vector ν_1 that maximizes the log-likelihood function (14). This vector can be found by numerical methods.

Estimates of the standard errors are found by taking the square root of the diagonal elements, of the Inverse of the Fisher Information Matrix. More formally the standard errors are obtained as:

$$\hat{\sigma}_{\nu_{1i}} = \left(-E\left[\left(\frac{d^2L(\nu)}{d\nu^2}\right)_{ii}\right]\right)^{-1}$$

where $\hat{\sigma}_{\nu_{1i}}$ denotes the standard error for the $i:th$ element of the parameter vector ν_1, and $\left(\frac{d^2L(\nu)}{d\nu^2}\right)_{ii}$ is the $i:th$ diagonal element of the hessian matrix.

5.2.2 Q-measure parameter

As a first step towards the calibration of the GOU model, we extracted the parameters under P, i.e. we estimated the parameter list $\nu_1 = (\sigma, b, p, q)$. However knowing this parameter list is just a necessary condition in order price derivatives living on the market (X, B). To fully specify the $Q-dynamics$ of the model, and to be able to attach a unique price to every claim on the market, we also need to know λ, the market price of risk. We can determine λ by almost exactly mimicking the procedure, by which, we calibrated the GBM-model to data.

Thus, what we want to do is to find the scalar λ, which minimizes the distance between observed future prices and theoretical future prices, i.e in symbols we want to solve the following problem:

$$min_{\lambda\in\Theta}||\mathbf{F}^{obs} - \mathbf{F}(\lambda)|| = \sum_{i,j}(F^{obs}(t_i,T_j) - F(t_i,T_j,\lambda))^2$$

$$= \sum_{ij}(F^{obs}(t_i,T_j) - exp(pT_j + q_1 + (ln[x(t_i)] - pt_i - q_1)e^{b(t_i-T_j)} + \frac{\sigma^2}{4b}(1 - e^{2b(t_i-T_j)}))^2 \tag{15}$$

Note, that the parameter-space Θ in this case is $\mathbb{R}$. It is also important to emphasize that the sample obtained and used as input, differs from that used when extracting the parameter vector under P. In the problem above $F^{obs}(t_i, T_j)$, denotes the observed futures price at time t_i, with a maturity at time T_j. At every date t_i we have several observations of futures contracts with different maturity. This is why we have indexed the observations above with a double index. What also distinguishes this sample from the one when estimating the P measure parameters, is that it is not sampled at equally spaced points in time. (For a more detailed description of the dimensionality of the data set, and the indexing of the observations, see the data section and the section on the calibration of the GBM model.)

Now, remember that we have estimated all parameters of the model except λ. As with the GBM model, we could try to minimize the distance of log prices instead of solving the problem in(15). In that case there will be a closed form solution for the estimate of λ. This estimate, can then be used as a starting value when trying to solve the problem above with for example Newtons Method (see the calibration of the GBM model).

5.3 Calibration results

5.3.1 P-measure parameters

When performing the MLE optimization, we obtained the following estimates of the parameters:

Parameters	Estimates/(St.err)
p	0.05/(0.017)
q	4.05/(0.0007)
b	4.11/(1.524)
σ_1^2	0.003/(0.368)

Figure 8: *Estimates* of the GOU- model with $\varphi(t) = pt + q$

A heuristic interpretation of the parameters is that, q is the “base-level” of mean-reversion, i.e.starting at $t = 0$, we would expect the process to oscillate around 4. The parameter p can be interpreted as the increase in mean index level per time, i.e. after one year, we would expect the process to oscillate around $4 + 0.053 = 4.053$, and continuing in the same fashion. Differently put this means that the mean level of the Index, increases approximately 5% per year. As explained earlier b is heuristically thought of as the speed of mean reversion. By comparing with figure 23 these estimates feels fairly reasonable. For example it is seen that in 1987, log prices was at a level of 4, and they did in fact steadily increase the whole sample period (except for the last few years) to a level around 5. A simple calculation shows that the mean increase per year in log prices was approximately $1/20 = 0.05$. This calculation may seem arbitrary. It can be justified by noting how smooth the evolution of the Index have been ever since the start of the index in 1987.

In the above table, the standard errors of the estimates are reported in parenthesis. A small standard error, roughly indicates a high statistical significance.

5.3.2 Q-measure parameter

When minimizing the distance of log prices we get an estimate of lambda equal to -0.03. Since, we as in the GBM case, can not theoretically justify any explicit assumptions about the distribution of the residuals in the above least squares minimization, it is not fair to talk about the statistical significance of the parameter λ. On the other hand we can get a feeling of how well the model prices futures by evaluating how much of the variation in the observed prices that the model can explain. We compute the following statistic to capture this:

$$R^2 := 1 - \frac{\sum(ln[F^{obs}(t_i, T_j)] - ln[lnF(T_j, t_i, \lambda)])^2}{\sum(ln[F^{obs}(t_i, T_j)] - ln[F^{avg}])^2}$$

where $ln[F^{avg}] = \frac{\sum ln[F^{obs}]}{nr.obs.}$. We compute a R squared of 0,8 for the OU model. This can be compared to the R^2 computed for the GBM model, which was 0,9. Since the GBM model is nested in the GOU model we expected the R^2 to be higher for the GOU model than for the GBM model, but this is not the case. One explanation for this contradictory result is that the calibration methods used for the models differs. First of all, when it comes to the GBM model we backed out θ by comparing observed and theoretical prices for the period 2006-2009. Knowing θ was sufficient in order to price futures on the GBM market. On the other hand when we calibrated the GOU model we estimated the $P - measure$ parameters of the model for the period 1987-2009, and then compared observed and theoretical future prices, for the period 2006-2009, to back out the parameter λ. Thus when calibrating the GOU model we used a longer history of observations from the underlying Index (p measure parameters).

In the end, the fact that we use a longer history of observations from the underlying Index when calibrating the GOU model, makes the results of the calibrations incomparable. It would have been possible to calibrate the GBM model in precisely the same way as we calibrated the GOU model. However it is empirically well known that the estimates obtained for μ usually have very wide confidence intervals. Thus we would not gain anything by adopting a similar procedure to the GBM model.

To further asses the validness of the estimate of lambda, we compute daily implied lambda for each contract over time. It is evident from the figure below that daily implied lambda for the whole sample is uniformly distributed around its estimate of -0.03. This is encouraging since it suggests that investors on *average* are pricing derivatives with a constant lambda. What is less encouraging is that there is a wide dispersion in lambda. Thus the assumption of a constant lambda may be fallacious. However there are reasons not to expect a constant lambda:

1. Investors are irrational, i.e they cannot use the information that they posses efficiently. For example if investors had very short memories, a stochastic lambda would be realistic, since they would constantly revalue their estimates of lambda, and use much less information from the past.

2. The model assumptions are bad. However identifying a good model would not automatically imply that we should find a constant lambda. For example we approximate the spot prices with future prices with the shortest maturity.

3. Lambda is time varying representing changes in investor sentiment.

Another notable feature in the figure below is that there are many observations with a large negative lambda. Thus investors are willing to pay to take on the risks associated with some of the future contracts. Why would any investor be willing to do that? One possible explanation is that they do not understand that they are paying to engage in these bets (See Wilmott and Ahmad). Another reason is that investors seek exposure to this risk factor in order to diversify their portfolio.

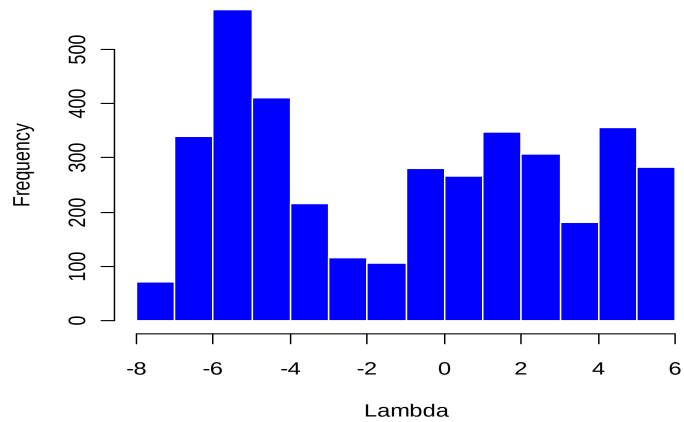

Figure 9: Graph showing distribution of daily implied lambda.

As a last step in examining the validness of the estimated λ, we investigate the size and sign of implied lambda for the different contracts, by plotting lambda against time for each contract (See figure Appendix). It is readily seen that a majority of the contracts with a maturity date before 2007 are valued with a negative lambda, and contrary a majority of the contracts with a maturity ex post 2007 are valued with a positive implied lambda. It is also seen, that lambda seems to be fairly constant or smooth over time for each contract. This result suggests a reversal in risk appetite and is also reflected in the rapid decline of house prices. As suggested by Wilmott and Ahmad, a negative market price of risk is often associated with economic boom periods of excessive risk taking and bubble building. Thus the wide dispersion in lambda seen in the histogram above, may be due to a time varying (increasing over time) lambda.

The R^2-value is calculated on the premise that lambda is constant, what other implications in terms of pricing does this assumption have? The graphs below show future prices implied by the model, versus observed future prices, assuming that $\lambda = -0.03$ and assuming that all other parameters are set to their estimates in the table on page 31. These graphs confirms one of the potential problems with the GOU model that we saw earlier. Assuming that $\varphi(t)$ is linear in t, could have the implication that future prices implied by the model are under/over priced relative to those observed in the market. This is clearly seen below. The model systematically under prices the August 06 contract and over prices the November 09 contract. This can be explained by the rapid increase of the Case Shiller Index, starting around 2006. The model "forces" prices to stay close to their long run average. This has the effect that when prices increase faster than predicted by the model, future prices and index levels are set too low, and vice versa.

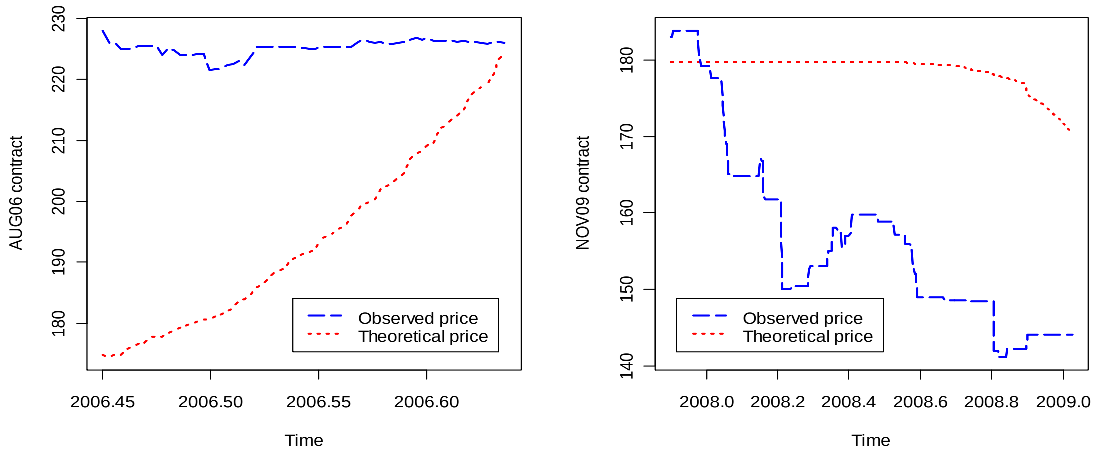

Figure 10: Graphs showing future prices implied by the model vs. observed future prices for a AUG06 contract and a NOV09 contract.

6 Good Deal Bounds

We have so far calibrated both the GBM model and the GOU model to data with mixed results. If we believe that these models can be properly calibrated to observed prices,why do we feel the need for calculating good deal bounds?

These two methods, in one sense, represent different views on part of the implementer, regarding the prices observed in the market. When calibrating a given model, we believe that those prices observed in the market are the "true" ones, and we back out the market price of risk by comparing them to the theoretical prices. However when calculating good deal bounds we make an assumption about the market price of risk, and then go to the market to see if there are any prices that represent too good deals. If this is the case, we regard these derivatives as miss priced. In short when calibrating a model we believe in the derivative market, when calculating good deal bounds we have more faith in the theoretical model of the underlying process.

Knowing good deal bounds also has value implications for many different parties in the financial markets. For example a financial institution can benefit from good deal bounds analysis, since they by knowing them can limit their risks in derivative transactions.

Recently there has been a rapid development in the theory of Good Deal Bounds. The basic idea of this theory is, as the name suggests, to exclude prices on incomplete markets that yield too good deals. The need for this theory, stems from the fact that by just precluding arbitrage, the pricing bounds obtained are too wide to be useful from a practical perspective.

In essence, the theory aims at finding the highest and lowest arbitrage free processes subject to a constraint on the Sharpe ratios of the traded derivative.

$$|SR(t)| < C^2$$

However this problem is intractable from a mathematical point of view, and it also permits portfolios, composed of the derivative, the underlying, and the money account that has very high sharpe ratios. This problem can be resolved by putting a bound on the right hand side of the Hansen Jagannathan inequality, i.e. we bound $||\lambda(t)|| < C^2$, utilizing the HJB inequility $|SR(t)| \leq ||\lambda(t)||$ (see Björk 2006), we then have the following chain of inequalities:

$$|SR(t)| \leq ||\lambda(t)|| < C^2$$

Thus, the problem translates to that of finding the highest and lowest arbitrage free price process subject to a bound on the market price of risk. The upper good deal bound for the case when the underlying follows a GBM, is then defined as the solution, to the following maximization problem:

Problem 5.1

$max_{\lambda}\ E^{Q}[X(T)]$ (5.1)

$dX(t) = \{\mu - \lambda\sigma\}X(t)dt + \sigma X(t)dW(t)$

$||\lambda||^2 < C^2$ where $C \in \mathbb{R}$

For general specifications of λ this is a standard stochastic optimal control problem. However in the scalar case presented above, where the market price of risk is assumed to be constant (that is why we omit the time -argument of λ), we note that the optimization problem has a trivial solution. We see that value of $\lambda \in \mathbb{R}$ that solves the problem lies in the interval $[-C, C]$.

Since we can solve for the expectation in 5.1 explicitly, i.e. $E_t^Q[X(T)] = X(t)e^{(\mu-\lambda\sigma)(T-t)}$ and we know that the exponential function is strictly increasing in its argument, we know that the solution to the problem must be a boundary solution, more precisely the lower boundary (again because e^x is increasing in x) i.e. $\lambda = -C$. In a similar way one sees that $\lambda = C$ solves the corresponding minimization problem.

Thus in the Geometrical Brownian Motion setting, where $\lambda$is assumed to be constant, we have the following pricing bounds for a future contract:

$$X(t)e^{(\mu-C\sigma)(T-t)} \leq F(t,T) \leq X(t)e^{(\mu+C\sigma)(T-t)} \tag{16}$$

In a similar manner we obtain the corresponding pricing bounds in the Ornstein Uhlenbeck case, also assuming a constant market price of risk. The problem is formulated exactly the same way, apart from the dynamics defining $X(t)$.

Problem 5.2

$$max_\lambda \ E^Q[X(T)] \ (5.1)$$

$$dX(t) = (b(pt + q_1 - ln[X(t)]) + p + \frac{\sigma^2}{2})X(t)dt + \sigma X(t)dW(t)$$

$$||\lambda||^2 < C^2 \text{ where } C \in \mathbb{R}$$

In this case $E^Q[X(T)] = exp(pT + q_1 + (ln[x(t)] - pt - q_1)e^{b(t-T)} + \frac{\sigma^2}{4b}(1 - e^{2b(t-T)}))$ as derived above. Reshuffling terms $E^Q[X(T)] = exp\{q_1(1 - e^{b(t-T)}) + pT + (ln[x(t)] - pt)e^{b(t-T)} + \frac{\sigma^2}{4b}(1 - e^{2b(t-T)})\}$. Recall that $q_1 := q - \frac{\lambda\sigma}{b}$. Now it is obvious that the second, third and fourth term in the expectation do not depend on λ, and the first term is a linear function of λ. Thus, since the exponential-function is a strictly increasing function, we can equivalently maximize the exponent. whether $\lambda = \pm C$ depends on whether $1 - e^{b(t-T)}$ is greater or less than 0, or equivalently whether b, the mean-reversion speed, is greater or less than 0. However b is by definition always taken to be positive. Thus $1 - e^{b(t-T)}$ is positive, and therefore the solution to the problem above is $\lambda = -C$.

Thus the pricing bounds in the OU-model are:

$$exp\{q_1^L(1 - e^{b(t-T)}) + \omega\} \leq F(t,T) \leq exp\{q_1^H(1 - e^{b(t-T)}) + \omega\} \tag{17}$$

where $q_1^L := q - C\frac{\sigma}{b}$ and $q_1^H := q + C\frac{\sigma}{b}$ and $\omega := pT + (ln[x(t)] - pt)e^{b(t-T)} + \frac{\sigma^2}{4b}(1 - e^{2b(t-T)})$.

6.1 Numerical properties of the pricing bounds

This section investigates some numerical properties of the pricing bounds derived for the GOU model and the GBM model. Most of these properties are fairly obvious when looking at the expressions for the pricing bounds in 16 and 17. However we will mention all of them for the sake of completeness. By investigating the properties of the good deal bounds we can also get a better understanding of the models, especially the GOU model.

Properties of the pricing bounds for the GBM model

In the case of the GBM model it is easy to see how changes in the different parameters will translate into changes in the pricing bounds. Just look at the expression for the pricing bounds in 16. It is obvious that an increase in σ (keeping all other variables fixed) will widen the pricing bounds. The same is true for λ. It is on the other hand not that obvious how changes in μ will affect the pricing bounds. An increase in μ will *both* shift the pricing bounds upwards and widen the bounds. One way to see this is the following:

The pricing bounds for the GBM model are:

$$X(t)e^{(\mu-C\sigma)(T-t)} \leq F(t,T) \leq X(t)e^{(\mu+C\sigma)(T-t)} \tag{18}$$

Say that μ increases by $x > 0$ units to $u + x$. The "new" pricing bounds are now :

$$X(t)e^{(\mu+x-C\sigma)(T-t)} \leq F(t,T) \leq X(t)e^{(\mu+x+C\sigma)(T-t)} \tag{19}$$

It is true that

$X(t)e^{(\mu-C\sigma)(T-t)} \leq X(t)e^{(\mu+x-C\sigma)(T-t)}$ and $X(t)e^{(\mu+C\sigma)(T-t)} \leq X(t)e^{(\mu+x+C\sigma)(T-t)}$ so the pricing bounds shifts upwards.

We also have:

$$X(t)e^{(\mu+x+C\sigma)(T-t)} - X(t)e^{(\mu+x-C\sigma)(T-t)} = X(t)e^{(\mu+x)}(e^{C\sigma(T-t)} - e^{-C\sigma(T-t)}) >$$

$$> X(t)e^{\mu}(e^{C\sigma(T-t)} - e^{-C\sigma(T-t)}) = X(t)e^{(\mu+C\sigma)(T-t)} - X(t)e^{(\mu-C\sigma)(T-t)}$$

i.e. the new pricing bounds are also wider.[10]

Properties of the pricing bounds for the GOU model

We will now investigate the properties of the pricing bounds for the GOU model.The parameters of the GOU model are p, q, b, σ and λ. By just looking at the analytical expressions, it is not easy to see, as for the GBM model, how changes in the different parameters will cause changes in the pricing bounds. Therefore we will with the aid of some graphs illustrate this. However to be able to do that we must chose a benchmark case, to which all the pricing bounds can be compared. In section 5.3.1-5.3.2

[10]The inequalities are reversed if $x < 0$.

we estimated the parameter list for the GOU model. It seems natural to chose, as the benchmark case, the pricing bounds obtained when setting the parameter vector equal to the estimated vector in section 5.3.1. If we let $v_2 = (b, p, q, \sigma, C)$, then v_2 is set equal to $(4.11, 0.05, 4.05, 0.05, 1)$ in our base case. Furthermore we let the level of the log Index levels $ln[X(t)]$ vary between 0 and 6, and the risk free rate r is set equal to $0,04$[11] We also assume that we are standing at $t = 0$ and that all future prices that we observe are prices on one year contracts, i.e. $T = 1$.

We have set $C = 1$ in our base case. This is for two reasons. First of all lambda was estimated to be -0.03 for the GOU model, secondly empirical evidence suggests that values of C greater than 2 are rare on mature markets (See Murgoci (2009)). The value of C that the implementer of the model chose should in general depend on the severity of the market incompleteness. Since the market for housing derivatives is highly illiquid, this motivates a higher value of C. Implications of changes in C on the pricing bounds will be investigated below.

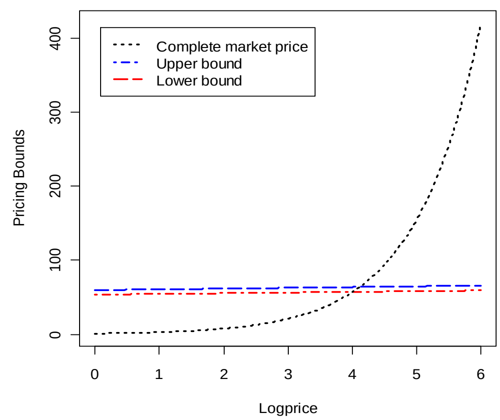

Figure 11: Showing good deal bounds for our base case. As apparent from the figure there is huge difference between the complete and the incomplete market prices.

We start off by investigating how the pricing bounds changes to changes in the parameter b. All the graphs in this section are found in the Appendix. Figure (24) show good deal bounds for $b = 4.11$, $b = 2$ and $b = 0.05$ The graphs should be interpreted in the following way: "Given that we stand at $t = 0$ and that we have a parameter vector $(b, 0.05, 4.05, 0.003, 1)$, we will, depending on the log prices at this time, expect to see future contracts trading at prices in the interval obtained by intersecting the graphs with a vertical line at the observed spot prices at $t = 0$." For example if $b = 4.11$ and the log price at $t = 0$ was 4, then we would expect future prices to trade somewhere in the interval $[56, 62]$. It is hard to tell from the graphs in the Appendix whether the pricing bounds widens or not when b increases, but a close examination confirms that they do widen. It is also seen that changes in b causes the pricing bounds to shift. In general the lower the value of b the more the GOU model resembles the GBM model. This partially explains why we have Upper Bound $>$ Complete price $>$ Lower bound,

[11] We need to specify r in order to be able to compute future prices in a complete market setting where $F(t,T) = X(t)e^{r(T-t)}$.

when b is low. In the limit when $b = 0$, we have the GBM model with a local mean of return equal to $p = 0.05$, and hence the pricing bounds in this case can also be given by :

$$X(t)e^{(0.05-\sigma)(T-t)} \leq F(t,T) \leq X(t)e^{(0.05+\sigma)(T-t)}$$

In the complete setting future prices can be calculated as $F(t,T) = X(t)e^{r(T-t)}$. Since $r = 0.04$ and $\sigma > 0.01$[12] we now see why Upper bound>Complete price>Lower bound is true when b is low.

From figure 11 it also evident that, if we discard the pricing bounds, a high level of b implies that future prices are valued almost at a constant level, independent of what spot price we actually observe at $t = 0$. In contrast, a low b implies that future contracts are valued very differently depending on the spot price observed. An explanation for this is the following. If the market believes that b is high, then they anticipate that the market will revert back very fast to the mean reversion level at this time (i.e. approximately 4 at $t = 1$). Thus if we at $t = 0$ observe a spot price above 4, say 5, futures prices will be valued almost as if the spot prices were 4, since it is expected that the spot prices revert back to the long run mean. In contrast if b is low, the market do not believe that the market will revert back to its long run mean *immediately*. This explains why the *pricing ranges* (not pricing bounds) are much wider in the graphs where b is high than in the graphs where b is low.

As expected an increase in q, letting all other variables being fixed to the base case, causes the pricing bounds to shift upwards. This is because if q increases the base level of mean reversion increases, and hence index levels are expected to oscillate around a higher level. This also causes future prices to increase. The pricing bounds are less sensitive to changes in q when b is low. This is because when b is low a change in q will not cause Index levels to oscillate around the new level of q immediately.

Changes in p has the same effect on the pricing bounds as changes in q, this could easily be seen investigating the analytical expressions for the pricing bounds. Reshuffling terms in (17) we see that the pricing bounds shifts upwards if $T - te^{b(t-T)}$is positive, which it is.

It is also apparent in the graphs that the higher the spot prices are, the wider are the good deal pricing bounds. This is merely due to the exponential nature of the future price process. From a economical point of view this is not very easy to explain. One idea is the following: The higher the price of the underlying the greater is the potential downside, and hence the uncertainty is greater.

From the graphs in the Appendix, we see that the pricing bounds widens, when increasing the volatility parameter (and vice versa). This is to expect since the uncertainty of the payoffs are greater in a high volatility environment.

By lastly looking at figure 25 and 26 it is also readily seen that the higher we set C, the bound on the market price of risk, the wider are the pricing bounds. We can also see this by inspecting the expressions for the pricing bounds in equation 16 and 17.

[12] Note that this is not σ_1.

7 Discussion

Although we know from previous studies that the housing market exhibits characteristics that would imply that the GBM framework should not be considered, or at least modified, we start out by examining it. The reason is that calibrating the GBM model may be viewed as a point of reference. Although this thesis is first of all concerned with the calibration problem (as opposed to the pricing problem) the two are clearly intimately connected. Even though in theory any model specification may be calibrated to market prices we do not only try to parametrize the market's view on housing risk, but also test how well our different models describes the index evolution. Thus we are interested both in the parameters estimated through the calibration and how well the pricing dynamics implied by the models fits available market data. This may give a hint of the validness of the model assumptions and is also a first step towards introducing more realistic models describing housing index dynamics.

The results from the GBM calibration are mixed. We reject the assumption of normality of log returns a of the index process. The assumption of log index levels being non stationary is not rejected when performing the dickey fuller test and the augmented dickey fuller test. It should be recognized that the evaluation of each of these model assumptions, by them selves, could constitute a whole thesis. Thus any further investigations are welcomed.

Also the level and movement of futures prices seem to be captured by the model. This is encouraging as it suggests that this simple framework is able to capture some of the dynamics of futures prices in the housing market. The parameter theta, which under the GBM specification determines the risk neutral drift of the index process, cannot be regarded as constant. However the information that we get from calculating theta on a daily basis suggests that the distribution is concentrated around the estimate. Optimally we would be able be to isolate the impact of measurement error in input parameters from the variation in theta. I.e the error that stem from inaccurate measurement of the index levels through introducing the spot approximations as well as the liquidity and market imperfections hidden in futures prices. Although we do not perform such analysis we qualitatively argue that theta is fairly constant considering these sources of disturbance. Of course this may however not necessarily mean that the model is good. It could simply be an implication of low variation in the input variables. Considering the high level of explained variation of price predictions within the sample we do not see this as likely. However the model needs to be further assessed with other contracts to complete the picture.

As a further step, we calibrate the GOU model to data, where also the mean reversion properties of the housing market is incorporated. We have assumed that the mean reversion is linearly dependent on t. Previous research (Shiller 2010) and the results in section 4.4.2 lends support to this specification. What is evident from the calibration of the GOU model is that estimating the parameter list and introducing mean reversion does not necessarily give a better fit to observed prices. We saw that the model was unable to give reasonable prices, if at some time the index deviated from its long run mean. This means that one has to be specially careful of the data set used when estimating the parameters. This also raises the question whether there is a more realistic specification of the mean reversion level $\varphi(t)$. Assuming a regime switching mean reversion level may be something to consider in future research, since the housing market seems sensitive to market shocks.

In line with the GBM setup, we also explicitly estimate lambda, also known as “the market price of risk”, for the GOU model. In both models we impose the restriction that lambda is constant through time although it is possible to allow for a time dependent, or even stochastic, market price of risk. In this thesis we find some evidence in favor of a non constant lambda.

In both models we have also assumed constant volatility, an assumption that was later proven to fail for the GBM model. There are ways of handling time dependent volatility within the model for example through modeling the volatility structure as a function of the index level (local volatility models) or as an independent stochastic process (stochastic volatility models). We however leave these questions open for further research. Of course we cannot ignore the fact that for both models our results probably are biased by the recent boom and bust in the US housing market. Sadly no market for housing derivatives was active prior to this period limiting our data set. It would have been very interesting to perform the same analysis on the period of 1987-2000 when housing prices were more stable.

Although theoretically appealing, one cannot neglect the fact the well functioning markets for housing derivatives are yet to be developed. One, by many possible reasons, that also have modeling implications, may be that the idiosyncratic variations of housing prices is diversified away in the index construction. For the individual homeowner this means that hedging may be difficult using index derivatives. From a qualitative perspective it is reasonable to believe that engaging in derivatives transactions is too costly if the “beta” of the individual housing unit is too low. As explained the CSI index is a composite of 10 of the largest U.S metropolitan areas and may thus not constitute a good hedging benchmark for individual homeowners. On the other hand we recognize the difficulties in getting trading volume in smaller and more local indices. From a modeling perspective it would be possible to introduce several risk factors to capture the certainly more interesting behaviours of such local indices. The risk when adopting such a model is that it becomes complicated and intractable. In the end the calibration depend not only on a correct specification of the underlying process but equally upon the availability of long time series of high quality data. The later of which we were reminded of during the credit crisis.

8 Appendix

8.1 Arbitrage Theory Review

We assume the following model:

$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t)$

$dB(t) = rB(t)dt$

We will from now on write μ and σ as shorthand for $\mu(t, X(t))$ and $\sigma(t, X(t))$. If we assume that there already exists a liquid market for one particular derivative written on the Case shiller Index, then we can take

two claims $\mathcal{L}$ and $\mathcal{Z}$ as given, where:

$\mathcal{L} = \phi(X(T))$

$\mathcal{Z} = \psi(X(T))$

We then make the following assumptions (following Björk.T 2009):

There is an liquid and frictionless market for both of the claims given above.

Assume that $F : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ and $G : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ are smooth real valued functions of class $C^{1,2}$, where $\prod(t, \mathcal{L}) = F(t, X(t))$ and $\prod(t, \mathcal{Z}) = G(t, X(t))$. Applying the Ito Formula to the transformed processes F and G, we get that:

$dF(t) = \alpha_F F dt + \beta_F F d\tilde{W}(t)$

$dG(t) = \alpha_G G dt + \beta_G G d\tilde{W}(t)$

where $\alpha_F \cdot F = F_t + \mu F_x + \frac{1}{2}\sigma^2 F_{xx}$ and $\beta_F \cdot F = \sigma F_x$and similar conditions holds for G. We have here used the same notation as in Björk (2009).

We now form a self financing portfolio consisting solely of F and G, i.e. we set up a portfolio $dV(t) = V(t)(u^G(t)\frac{dG(t)}{G(t)} + u^F(t)\frac{dF(t)}{F(t)})$,where $u^R(t)$ $R \in \{G, F\}$ are the relative portfolio weights defined as $u^R(t) := \frac{h^R(t)R(t)}{V(t)}$ for $R \in \{G, F\}$. $h(t) = (h^G(t), h^F(t))$is an portfolio strategy, intuitively describing the dynamics of the number of derivative contracts hold in the portfolio.

Inserting the expressions for dF and dG given by Ito's lemma above, we get the following dynamics for the self-financing portfolio. $dV = V\{u^F(\alpha_F dt + \beta_F d\tilde{W}(t)) + u^G(\alpha_G dt + \beta_G d\tilde{W}(t))\} = V\{(u^F\alpha_F + u^G\alpha_G)dt + (u^F\beta_F + u^G\beta_G)d\tilde{W}(t)\}$

Then in order to cancel all the randomness in the portfolio, we have to impose the following conditions on the portfolio weights:

$u^F\beta_F + u^G\beta_G = 0$

$u^F + u^G = 1$

which has a solution if and only if its determinant is nonzero i.e. $\beta_F \neq \beta_G$. Solving this system for u^F and u^G we get:

$$u^F = \frac{-\beta_G}{\beta_F - \beta_G} \tag{20}$$

$$u^G = \frac{\beta_F}{\beta_F - \beta_G} \tag{21}$$

In order to avoid arbitrage we need to have $k = r$ for this portfolio, i.e. the evolution of the deterministic portfolio $dV_t = kV_t dt$ has to have the same drift as the Bank account $dB(t) = rB(t)dt$ in order to preclude arbitrage opportunities.

By setting the drift term of the self financing portfolio equal to the deterministic short rate, and inserting the expressions for the weights in (20) and (21) we get:

$u^F \alpha_F + u^G \alpha_G = r \Longrightarrow \frac{-\beta_G}{\beta_F - \beta_G}\alpha_F + \frac{\beta_F}{\beta_F - \beta_G}\alpha_G = r \Longrightarrow -\beta_G \alpha_F + \beta_F \alpha_G = r(\beta_F - \beta_G) \Longrightarrow -\beta_G \alpha_F + r\beta_G = r\beta_F - \beta_F \alpha_G$ multiplying both sides by $\frac{-1}{\beta_F \beta_G}$ we have : $\frac{\alpha_F - r}{\beta_F} = \frac{\alpha_G - r}{\beta_G}$.

Thus in the end we arrive at the conclusion that : Assuming that the market (X, B) is free of arbitrage, there exists a process $\lambda(t, X(t))$ such that for all $t \in [0, T]$we need to have $\frac{\alpha_F(t, X(t)) - r}{\beta_F(t, X(t))} = \lambda(t, X(t))$ for any derivative F in the market. From this relation by inserting the expressions of $\alpha_F(t)$ and $\beta_F(t)$ we can derive the PDE stated in the methodology section. For simplicity and without better knowledge, we have maken the following assumption of the functional form of lambda $\lambda(t, X(t)) = \lambda \vee t \in [0, T]$ i.e lambda is time invariant and deterministic. This is motivated in the main text.

8.2 Heuristic derivation of future price when the underlying is a GBM

As explained the dynamics of X under Q is :

$$dX(t) = \{\mu - \lambda\sigma\} X(t)dt + \sigma X(t)dW(t) \tag{22}$$

For the purpose of pricing a futures contract we are interested in determining $E^Q[X(t)]$. If we define $z_t := E^Q[X(t)]$. Then we would expect that $dz_t = dE^Q[X(t)] = E^Q[dX(t)]$, because of the linearity of the expectation operator. Furthermore $E^Q[dX(t)] = E^Q[\{\mu - \lambda\sigma\} X(t)dt + \sigma X(t)dW(t)] = \{\mu - \lambda\sigma\} E^Q[X(t)]dt + E^Q[\sigma X(t)dW(t)]$. And because $E^Q[\sigma X(t)dW(t)] = 0$ for any sufficiently integrable process $X(t)$,we have:

$$dz_t = \{\mu - \lambda\sigma\} E^Q[X(t)]dt + E^Q[\sigma X(t)dW(t)] = \{\mu - \lambda\sigma\} E^Q[X(t)]dt = \{\mu - \lambda\sigma\} z_t dt$$

From basic calculus we know that $dz_t = \{\mu - \lambda\sigma\} z_t dt$ has the solution $z_t = Ce^{(u - \lambda\sigma)t}$, where $C = X(0)$, since $z_0 = E^Q[X(0)|\mathcal{F}_0] = X(0)$. Thus we have that $F(t, T) = E^Q[X(T)|\mathcal{F}_t] = X(t)e^{(\mu - \lambda\sigma)(T - t)}$. To derive the future price process in a more formal way, it is possible to apply the same steps as when deriving the GOU model below.

8.3 Dynamics of $X(t)$

Assuming that $dZ(t) = \{b(pt + q - Z(t)) + p\}dt + \sigma dW(t)$, and defining $X(t) := f(t, Z(t)) = e^{Z(t)}$, then X has a stochastic differential given by:

$df = \frac{df}{dt}dt + \frac{df}{dz}dZ + \frac{1}{2}\frac{d^2f}{dz^2}(dZ)^2 = 0 + \frac{df}{dz}(\{b(pt + q - Z(t)) + p\}dt + \sigma dW(t)) + \frac{1}{2}\frac{d^2f}{dx^2}(\{b(pt + q - Z(t)) + p\}dt + \sigma dW(t))^2 =$

$= \{b(pt + q - Z(t)) + \frac{\sigma^2}{2} + p\}e^{Z(t)}dt + \sigma e^{Z(t)}dW(t)$. Thus

$$dX(t) = (b(pt + q - ln[X(t)]) + \frac{\sigma^2}{2} + p)X(t)dt + \sigma X(t)dW(t) \tag{23}$$

where we have used the formal multiplication table $(dW)^2 = dt$,$dW \cdot dt = 0$ and $(dt)^2 = 0$ presented in Björk (2009). Note that the above argument by no means constitute a full proof, we have just given a heuristic explanation why the process followed by Z is as in (23).

8.3.1 Solution to the SDE in (7), where $\varphi : \mathbb{R}^+ -> \mathbb{R}^+$ is an arbitrary sufficiently "smooth" function.

Recall that the dynamics followed by Z under P was:

$$dZ(t) = \{b(\varphi(t) - Z(t)) + \frac{d\varphi(t)}{dt}\}dt + \sigma dW(t)$$

If we let $X(t) = Z(t) - \varphi(t)$ then $dX(t) = dZ(t) - \frac{d\varphi(t)}{dt}dt = -bX(t)dt + \sigma dW(t)$ (apply Ito's theorem) and thus:

$$e^{bt}dX(t) + be^{bt}X(t)dt = \sigma e^{bt}dW(t) \tag{24}$$

It follows that $d(e^{bt}X(t)) = \sigma e^{bt}dW(t)$ (note that $d()$ is the Ito-differential and the latest implication is seen by for example applying Ito's lemma). If we set $M(t) := e^{bt}X(t)$, then $M(t) = M(s) + \int_s^t \sigma e^{bu}dW(u)$ for $s < t$. Remembering that $Z(t) = X(t) + \varphi(t) = e^{-bt}M(t) + \varphi(t) = e^{-bt}(M(s) + \int_s^t \sigma e^{bu}dW(u)) + \varphi(t)$. Since by definition $M(s) = e^{bs}X(s) = e^{bs}(z(s) - \varphi(s))$, we have $Z(t) = e^{-bt}(e^{bs}(z(s) - \varphi(s)) + \int_s^t \sigma e^{bu}dW(u)) + \varphi(t) = \varphi(t) + e^{b(s-t)}(z(s) - \varphi(s)) + e^{-bt}\sigma\int_s^t e^{bu}dW(u))$ (Note the upper and lower integration- limits are s and t in all integrals above)

Remark: If we would have assumed that the drift was dependent on time, i.e. that it was a mapping $b : \mathbb{R}^+ \to \mathbb{R}^+$, then to solve this SDE, we could, with a minor modification, apply exactly the same argument as above. We would just have to multiply with an integrating factor $e^{B(t)}$ in (24) above, where $B(t)$ is the anti derivative of $b(t)$.

If we let $\varphi(t) := pt + q$, then it is seen, that in this special case, the solution to the SDE is:

$$Z(t) = pt + q + e^{b(s-t)}(z(s) - ps - q) + e^{-bt}\sigma \int_s^t e^{bu} dW(u)) \tag{25}$$

8.3.2 The solution to the corresponding SDEunder Q

Assuming that $\varphi(t) := pt + q$, the risk-neutral dynamics followed by Z can be seen to be $dZ(t) = \{b(pt + q - Z(t)) - \lambda\sigma + p\}dt + \sigma dW(t)$

Reshuffling terms we can rewrite this as:

$$dZ(t) = \{b(pt + q_1 - Z(t)) + p\}dt + \sigma dW(t)$$

where $q_1 := q - \lambda\frac{\sigma}{b}$

Now we know from above that the solution to this SDE simply is (replace q with $q_1$everywhere in (25)):

$$Z(t) = pt + q_1 + e^{b(s-t)}(z(s) - ps - q_1) + e^{-bt}\sigma \int e^{bu} dW(u)) \tag{26}$$

8.3.3 Finding the moments of the GOU process under Q

In order to find theoretical future prices of futures written on the Case shiller Index, we need to know the moments of the GOU-process. By a well known lemma (See Björk Chapter 4), we have that if the process $X(t)$ is defined by the following SDE :

$$X(t) = \int \sigma(s) dW(s)$$

where $\sigma(t)$ is a given deterministic function, then $X(t)$ is normally distributed with zero mean and $Var[X(t)] = \int \sigma^2(s)ds$.

By applying the above lemma to $Z(t)$ in (26), it is not hard to see that the moments of $Z(t)$, when $\varphi(t) := pt + q$, are the following:

$$E^Q[Z(T)] = pT + q_1 + (z(t) - pt - q_1)e^{b(t-T)}$$

$$Var^Q[Z(T)] = \frac{\sigma^2}{2b}(1 - e^{2b(t-T)})$$

Note: $t < T$ and that t, T has the same meaning as s, t in all the derivations in this Appendix.

8.4 The likelihood function

The likelihood function is given by $L(\nu_1) = p(Z_{t_i}, Z_{t_{i-1}}, ..., Z_{t_1}, \nu_1)$ By Bayes Theorem we can rewrite this as:$p(Z_{t_i}, Z_{t_{i-1}}, ..., Z_{t_1}, \nu_1) = p(Z_{t_i}|Z_{t_{i-1}}, ..., Z_{t_1}, \nu_1) \cdot p(Z_{t_{i-1}}, ..., Z_{t_1}, \nu_1)$. By iteratively using Bayes theorem in the same fashion we arrive at: $p(Z_{t_i}, Z_{t_{i-1}}, ..., Z_{t_1}, \nu_1) = p(Z_{t_i}|Z_{t_{i-1}}, ..., Z_{t_1}, \nu_1) \cdot p(Z_{t_{i-1}}|Z_{t_{i-2}} ...., Z_{t_1}, \nu_1) \cdots p(Z_{t_2}|Z_{t_1}; \nu_1) \cdot p(Z_{t_1}; \nu_1)$. Since the Ornstein Uhlenbeck process satisfies the Markov property, we also have that $p(Z_{t_i}|Z_{t_{i-1}}, ..., Z_{t_1}, \nu_1) = p(Z_{t_i}|Z_{t_{i-1}}, \nu_1)$ and finally we see that $p(Z_{t_i}, Z_{t_{i-1}}, ..., Z_{t_1}, \nu_1) = \prod p(Z_{t_n}|Z_{t_{n-1}}, \nu_1)$ where we multiply over all $n \in \{1, 2...i\}$.

9 Appendix

9.1 Tables and Graphs

9.1.1 Levene's test for equal variances

Sub sample 1 1987-01-01 to 1994-09-01 (92 obs),

Sub sample 2 1994-10-01 to 2002-05-01 (92 obs)

Sub sample 3 2002-06-01 to 2009-11-01 (90 obs)

Test	P-Value
Sub sample 1 vs Sub sample 2	0.278
Sub sample 2 vs Sub sample 3	$<$0.001
Sub sample 1 vs Sub sample 3	$<$0.001
First half vs Second half of Sample	$<$0.001

Figure 12: Table showing output from Levene´s test

9.1.2 Shapiro Wilk tests for normality conducted in R

Sub Sample	Test-Statistic	P-Value
1987-2006	0.993	0.306
2006-2009	0.952	0.083
Whole sample	0.951	$<$0.001

Figure 13: Showing output of Shapiro Wilk Test

9.1.3 Histogram and qqplot for the pricing errors

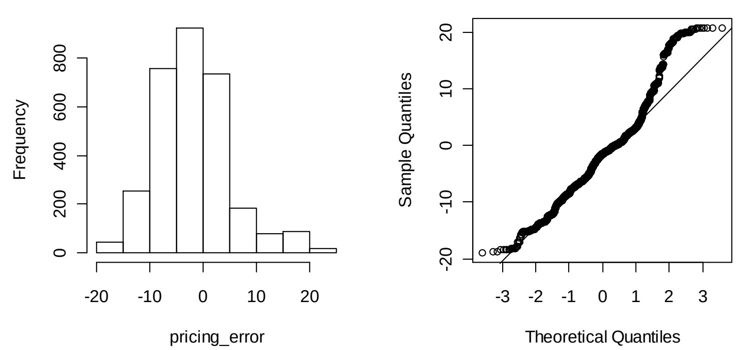

Figure 14: Histogram and qq plot showing pricing error.

9.1.4 Results of the Dickey fuller tests

Coefficients	Estimate	Std. Error
Intercept	0.0124	0.00643
Φ	-0.00193	0.00137

Figure 15: Results of the "standard" dickey fuller test with a constant term.

Lag order	P-value
1	0.442
2	0.907
3	0.943
4	0.782
5	0.546
6	0.441
7	0.261
8	0.096
9	0.029

Figure 16: The augumented dickey fuller test for 1-9 lags.

9.1.5 Correlogram of the residuals in the dickey fuller regression

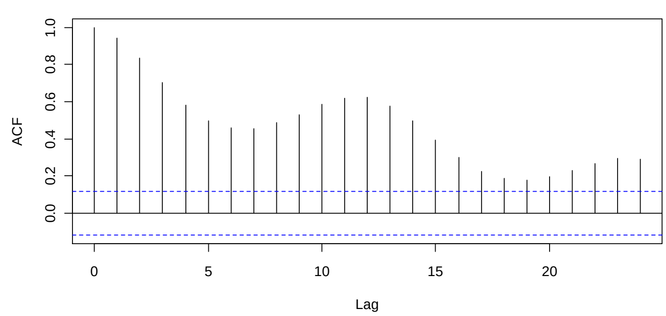

Figure 17: Showing correlogram of the residuals in regression6.

9.1.6 Graph showing spot approximation over time

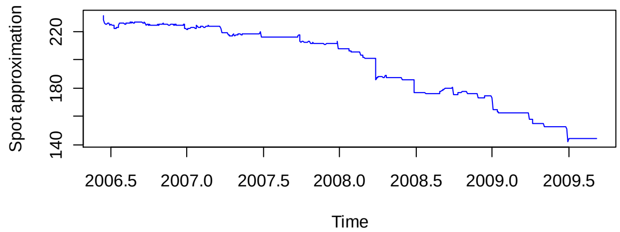

Figure 18: approximation of spot price

9.1.7 Plot and histogram of residuals

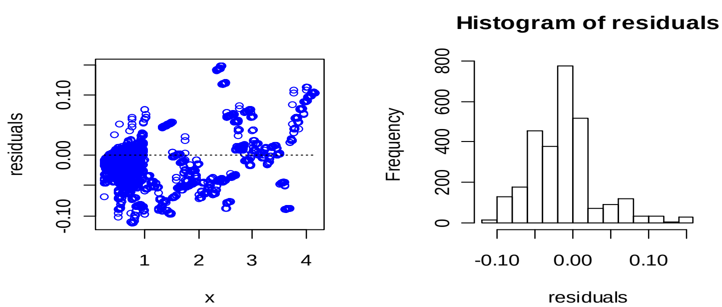

Figure 19: Left graph showing plot of residuals where $x = T - t$, right graph is a histogram of the residuals. It is readily seen that the distributtion of the residuals is negatively skewed. Furthermore assuming a homoscedastic variance of the residuals seems troublesome.

9.1.8 Relative pricing error for the GBM model.

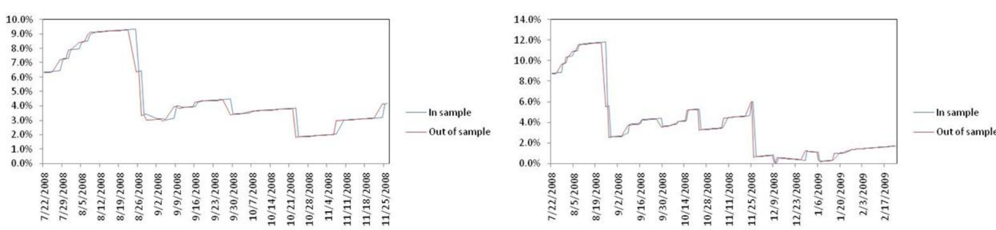

Figure 20: Graphs showing in/out of sample pricing error for Feb 09 contract (left) and May 09 contract (right)

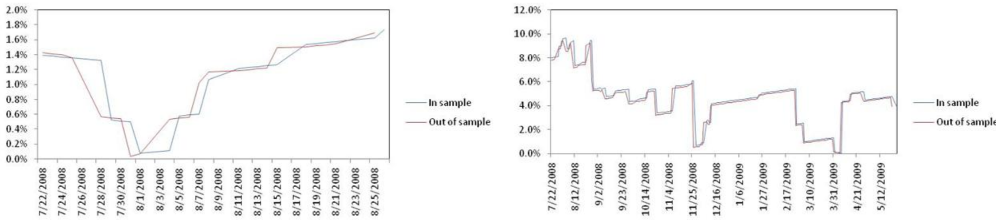

Figure 21: Graphs showing in/out of sample pricing errors for Nov 08 contract (left) and Nov 09 contract (right)

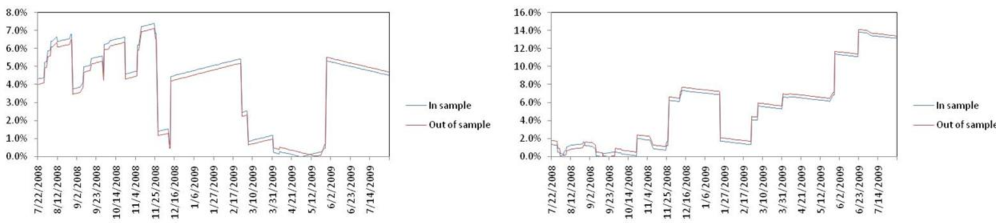

Figure 22: Graphs showing in/out of sample pricing error for Nov 10 contract (left) and Nov 11 contract (right)

9.1.9 Plot of log index levels

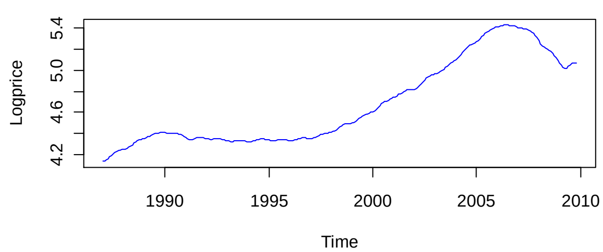

Figure 23: Graph displaying log prices over time for the Case Shiller Index

9.1.10 Good deal bounds for different values of b

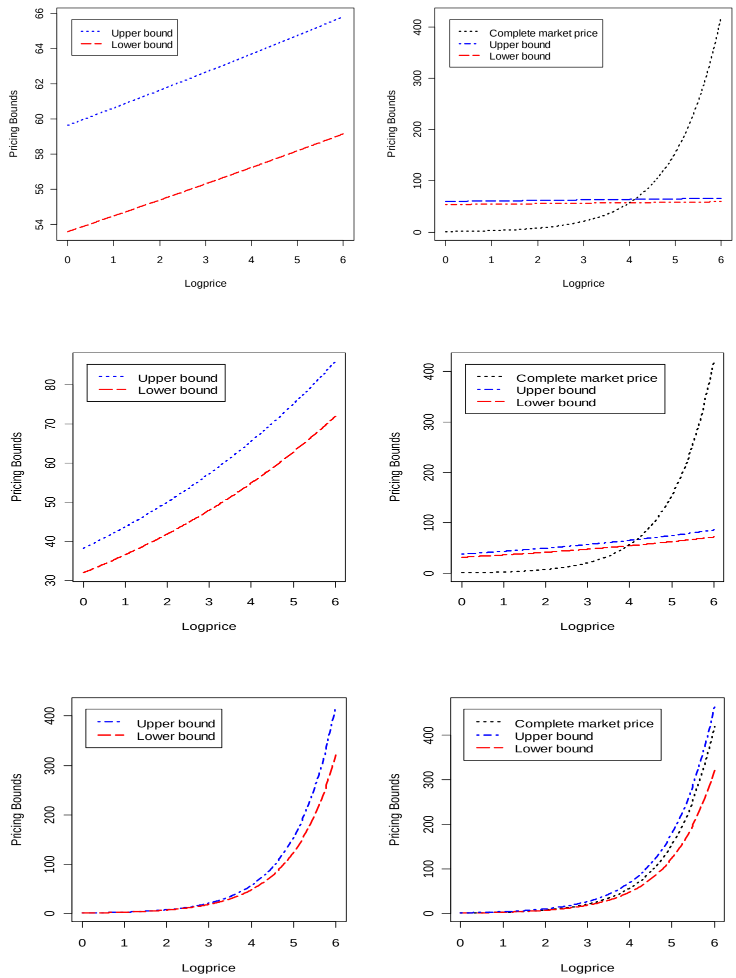

Figure 24: Showing good deal bounds for $b = 4,11$ (top), $b = 2$ (middle) and $b = 0,05$ (bottom). All other parameters are set to their values in the base case.

9.1.11 Good deal bounds for different values of C

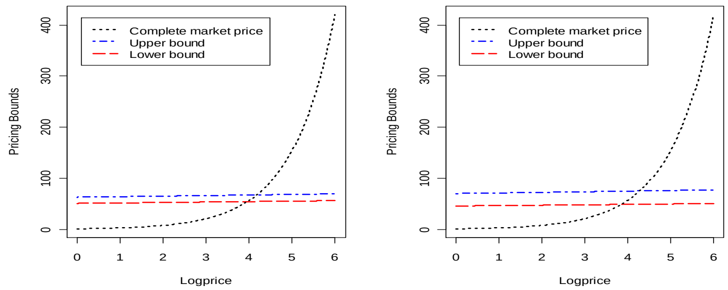

Figure 25: Pricing bounds when $C = 2$ left and $C = 4$ right. All other values are fixed to their values in the base case.

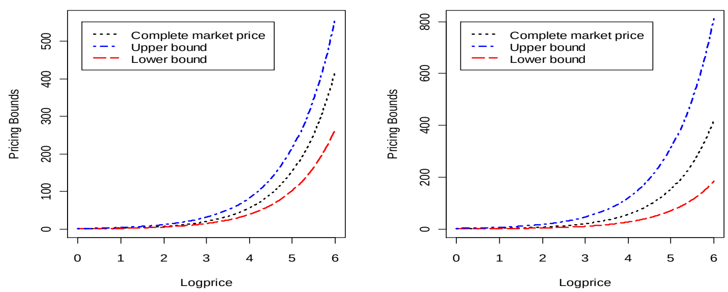

Figure 26: Note that the scale on the y axis differ. Pricing bounds when $C = 2$ left and $C = 4$ right and b is set to 0,05. All other values are fixed to their values in the base case.

9.1.12 Good deal bounds for different values of q

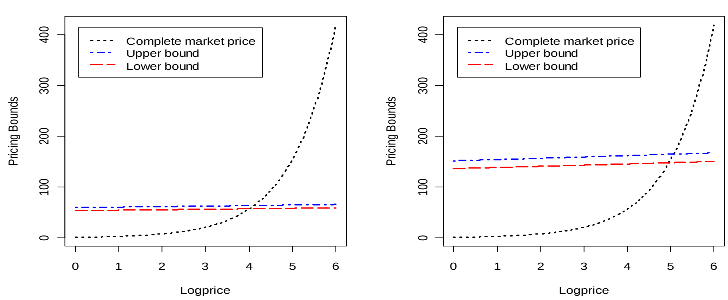

Figure 27: Pricing bounds when $q = 4$ left and $q = 5$ right. All other values are fixed to their values in the base case.

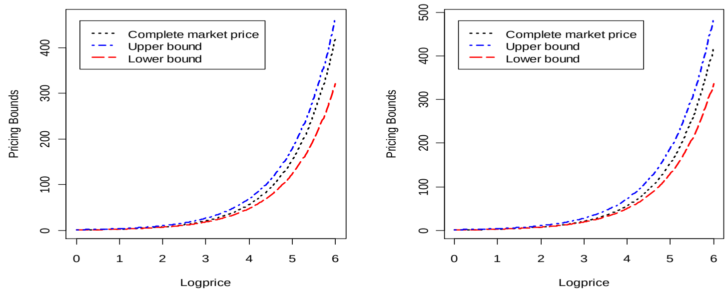

Figure 28: Note the scale for the y axis differ for the graphs! Pricing bounds when $q = 4$ left and $q = 5$ right and b is set to 0,05. All other values are fixed to their values in the base case.

9.1.13 Good deal bounds for different values of σ

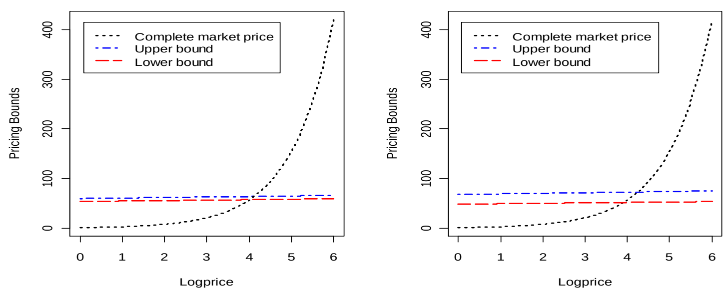

Figure 29: Note the scale for the y axis differ for the graphs! Pricing bounds when $\sigma_1^2 = 0,003$ left and $\sigma_1^2 = 0,03$ right. All other values are fixed to their values in the base case.

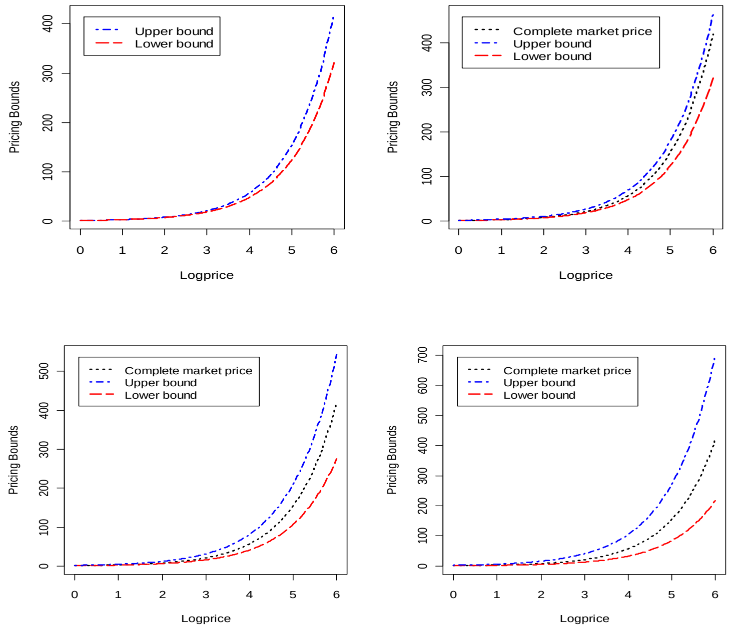

Figure 30: Showing pricing bounds when $\sigma_1^2 = 0.003$ (top),$\sigma_1^2 = 0.01$ (bottom left) and $\sigma_1^2 = 0.03$ (bottom right). In all graphs $b = 0.05$ and the rest of the parameters are set to their base case values.

9.1.14 Graphs showing implied lambda for the GOU model over time

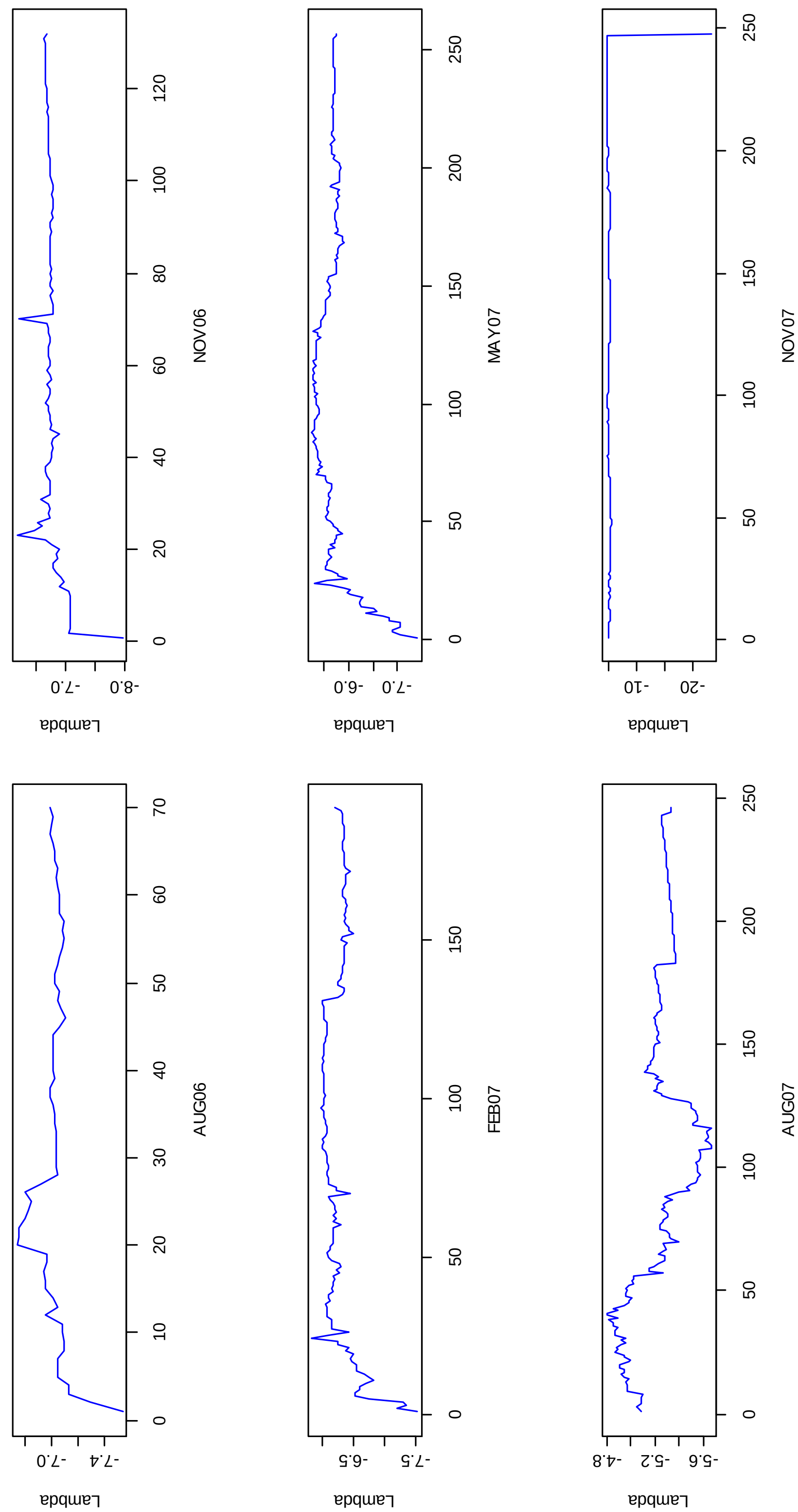

Figure 31: Graphs showing implied lambda over time for som different contracts ordered by maturity date.

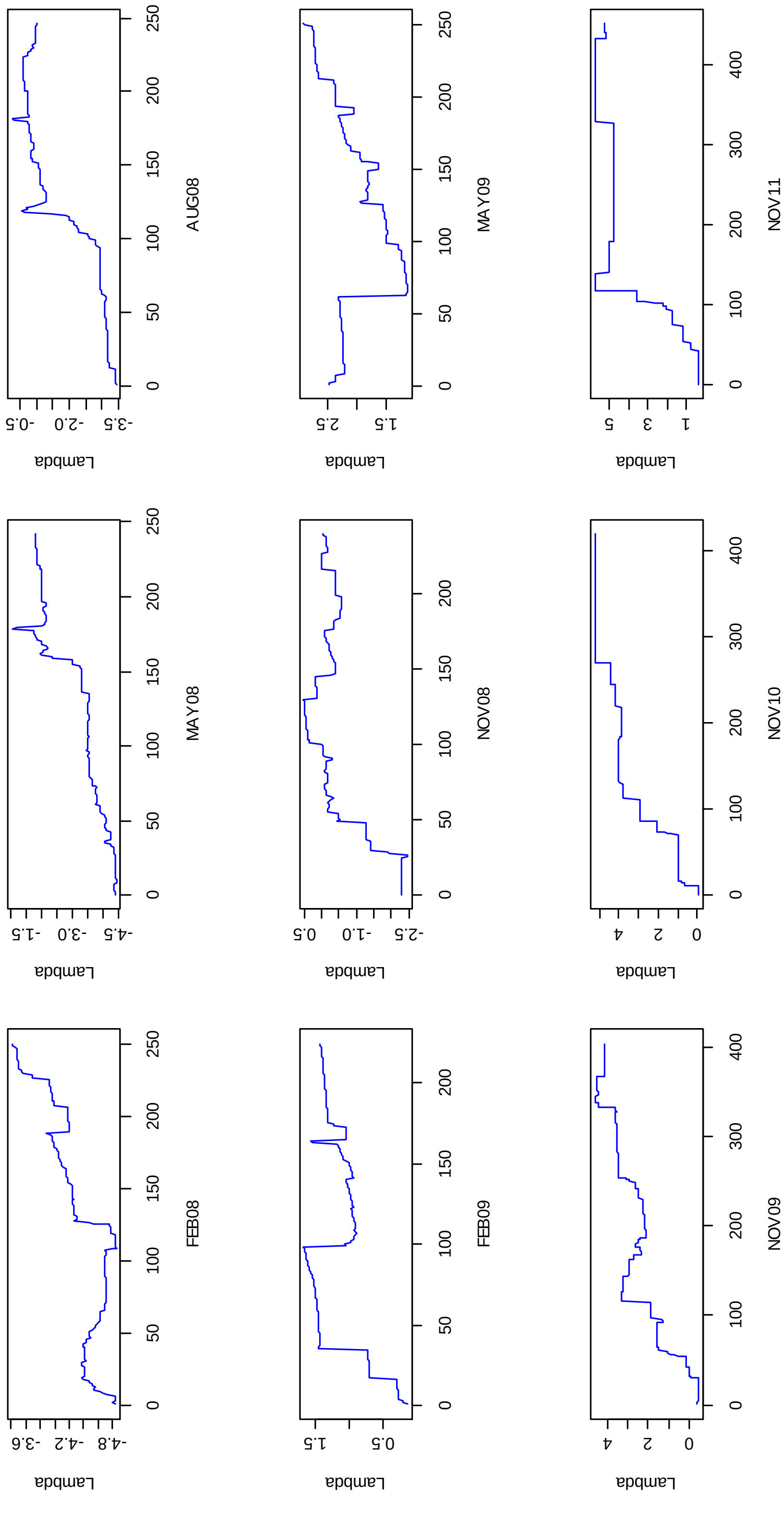

Figure 32: Graphs showing implied lambda over time for som different contracts ordered by maturity date

References

[1] Ahmad, R. Wilmott, P. The Market Price of Interest-rate Risk: Measuring and Modelling Fear and Greed in the Fixed-income Markets, WILMOTT magazine, 64-70

[2] Björk, T. (2009) *Arbitrage Theory in Continuous Time*, third edition, Oxford University Press, Oxford.

[3] Björk, T. Slinko, I. (2006) Towards a General Theory of Good Deal Bounds, Review of Finance, Springer.

[4] Capozza, D.R Hendershott, P.H. Mack, C. (2004) An Anatomy of Price Dynamics in Illiquid Markets: Analysis and Evidence from Local Housing Markets, Real Estate Economics, 32:1, 1-32

[5] Case, K. E. Shiller, R.J. Weiss, A.N. (1993) Index-Based Futures and Options in Real Estate, The Journal of Portfolio Management, 83-92

[6] Chong, E K P. Zak, S H. (2001) An Introduction to Optimization, second edition, Wiley & Sons.

[7] Cochrane, J. and Saá Requejo , J. (2000) Beyond arbitrage: Good-deal asset price bounds in incomplete markets, Journal of political Economy, 79-119.

[8] Djehiche, B. Alaton, P. (2002) On modelling and pricing weather Derivatives, Applied Mathematical Finance, Volume 9, 1 - 20.

[9] Englund, P. Hwang, M. Quigley, J.M. (2002) Hedging Housing Risk, Journal of Real Estate Finance and Economics, 24:1/2, 167-200

[10] Englund, P. Gordon, T. Quigley, J.M. (1999) The Valuation of Real Capital: A Random Walk down Kungsgatan, Journal of Housing Economics, 8, 205-216

[11] Fabozzi, F J. Shiller, R J. Tunaru, R S. (2010) Property Derivatives for Managing European Real-Estate Risk, European Financial Management, 8-26.

[12] Murgoci, A. (2009) *Essays in Mathematical Finance*, Dissertation Stockholm School of Economics.

[13] Phillips, P. C. B. and Yu, J. *Maximum Likelihood and Gaussian Estimation of Continous Time models in Finance*, Cowles Foundation Discussion Papers.

[14] Resnick, S I. (2009) *A Probability Path,* Birkhäuser.

[15] Schwartz, E S. The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging, The Journal of Finance, Vol. 52, No. 3, 923-973

[16] Shiller, R.J Weiss, A.N. (1999) Home Equity Insurance, Journal of Real Estate Finance and Economics, 19:1, 21-47

[17] Syz, J. Vanini, P. Salvi, M. (2006) Property Derivatives and Index-Linked Mortgages, Working paper 373, National Centre of Competence in Research Financial Valuation and Risk Management

[18] Önalan, Ö. (2009) Financial Modelling with Ornstein–Uhlenbeck Processes Driven by Lévy Process, Proceedings of the World Congress on Engineering.

[19] Englund, P. (2008) Lecture Notes in Real Estate Finance, Stockholm School of Economics, fall 2008

[20] Standard & Poors (2010a) S&P/Case-Shiller Home Price Indices fact Sheet, websource, http://www.standardandpoors.com, accessed 2010-06-07

[21] Standard & Poors (2010b) S&P/Case-Shiller Home Price Indices Index Methodology, websource, http://www.standardandpoors.com, accessed 2010-06-07