The Choice of Stochastic Process in Real Option Valuation II: Selecting Multiple Factor Models

Luiz de Magalhães Ozorio
Department of Management, Ibmec Business School,
Rio de Janeiro, Brazil
email: lmozorio@ibmecrj.br

Pavel V. Shevchenko
CSIRO Mathematics, Informatics and Statistics,
Sydney, Australia
email: Pavel.Shevchenko@csiro.au

Carlos de Lamare Bastian-Pinto
Department of Management, Ibmec Business School,
Rio de Janeiro, Brazil
email: carbastian@gmail.com

Draft paper, 5 July 2013

The Choice of Stochastic Process in Real Option Valuation II: Selecting Multiple Factor Models

Abstract

The stochastic process choice plays a central role in real option valuation and it can have an impact not only on the project value but also on the investment rule. The first studies on real options used one-factor models such as Geometric Brownian Motion (GBM) and Mean Reversion Models (MRM) to represent uncertainties in the valuation modeling. Selecting the most appropriate model is not always a trivial issue, and besides statistical tools, in general, theoretical considerations are taken for this task. In order to generate more realistic models, in the last decades many authors have presented papers proposing the combination of different kinds of stochastic processes creating multiple factor models. Although these models can be more realistic, the task of selecting among many multiple factor models is more difficult than in case of one-factor models. This paper discusses the choice of multiple factor models in real options valuation, and the main statistical tools and theoretical considerations that can be used for this task.

Key words: stochastic process, real options valuation, multiple factor models, model choice
JEL codes: C15, C53, G13, M21.

1 Introduction

The choice of a stochastic process is an issue of great relevance in the assets valuation modeling aiming to represent uncertainties related to investments. In the case of real options it can have an impact not only on the project value but also on the investment rule (Dixit & Pindyck 1994; Schwartz 1997).

Earlier studies on financial options (Black & Scholes 1973; Cox, Ross & Rubinstein 1979) and real options (Brennan & Schwartz 1985; McDonald & Siegel 1985 and 1986; Paddock, Siegel, & Smith 1988) assumed Geometric Brownian Motion (GBM) for the underlying asset. For valuation of commodities, it is common to use Mean Reversion Models (MRM) (Bhattacharya 1978; Brennan & Schwartz 1985; Dixit & Pindyck 1994), assuming that commodity price might behave randomly in short term but tends to converge to an equilibrium level in the long run reflecting the marginal cost of production.

The task of determining the most appropriate process for the underlying stochastic variables is usually not a trivial question and, in some cases, analysts realize that these uncertainties have elements of more than one type of process. In order to generate more realistic models, several authors proposed models with multiple factors that combine different kinds of processes (Schwartz 1997; Pindyck 1999; Dias & Rocha 1999; Schwartz & Smith 2000). Although the use of these models might increase the accuracy to the asset valuation, it is not straightforward to select the appropriate multi factor model. Moreover, real options typically have a feature that allows exercise at any time before maturity. Thus, valuation of real options requires numerical methods. In the case of one or two-factor models, one can use the finite difference or tree methods to solve corresponding partial differential equations. However, in the case of models with more than two factors, one has to resort to special Monte Carlo approaches such as the least square Monte Carlo method suggested in Longstaff and Schwartz (2001).

This paper discusses the choice of stochastic process in real option valuation and useful tests and considerations to resolve this task: Dickey Fuller Test, Variance Ratio Test, Q-Q plots, autocorrelation, likelihood ratio test, Akaike information criterion, Bayesian information criterion, Bayes factors, direct calculation of model probabilities, deviance information criterion, out-of-sample or cross validation, in-sample errors, hedging errors and sensitivity analysis. We also present a set of empirical examples using real datasets and some simple real option applications in order to discuss the effect of stochastic process choice in the analysis.

The paper is structured as follows. Section 2 presents a bibliographical revision of stochastic processes focusing on multiple factor models applied to real option analysis. In Section 3 we describe some statistical tools that can be used to select the model.

Section 4 presents some empirical applications, and we conclude in Section 5.

2 Multiple Factor Models in Real Options Theory

GBM is the most popular stochastic model used not only in financial derivatives (Black & Scholes 1973; Cox, Ross & Rubinstein 1979) but also in real options analysis (Brennan & Schwartz 1985; McDonald & Siegel 1985, 1986; Paddock, Siegel, & Smith 1988). Its mathematical simplicity that allows to obtain analytical solutions for asset valuation and small number of parameters to calibrate can be considered the main reasons to explain its popularity. In other situations, when the uncertainties in prices depend on an equilibrium level, such as in case of commodities and interest rates, it is debated if the use of GBM would be appropriate (Bhattacharya 1978; Brennan & Schwartz 1985; Dixit & Pindyck 1994). In case of commodities such as oil, copper, sugar and ethanol, it is usual to assume that the price is driven by a mean reversion component, which makes the prices to behave randomly in short term and converge to the equilibrium level associated to the marginal cost of production in the long run.

The task of selecting GBM or MRM in order to represent the main uncertainties involved in a valuation is usually not a trivial issue. Besides statistical diagnostic tools such as Dickey Fuller Test (Dixit & Pindyck 1994) and Variance Ratio Test (Pindyck 1999), some theoretical considerations such as the economic theory and lifetime of assets (Ozorio, Bastian-Pinto & Brandão 2012) can be useful. Nevertheless, in many cases, analysts realize that these uncertainties have elements of more than one process. This has motivated many authors to consider multiple factor models.

One of the pioneer works that presented a multiple factor model was Merton (1976), where GBM and Poisson process are mixed as

$$\frac{dS_t}{S_t} = (\alpha - \lambda k)\, dt + \sigma dz + dq_t. \tag{1}$$

Hereafter, S_t is the price of a financial asset at time t; dz is a Wiener increment; and q_t is a Poisson process with the mean number of events per unit time λ and percentage change in price $\phi - 1$ if the Poisson event occurs. The uncertainty about the size and direction of the jump is represented by random variable $\phi - 1$ with the mean $k = E[\phi - 1]$. That is, dq_t is a Poisson process increment that takes a value $\phi - 1$ with probability λdt and zero with probability $1 - \lambda dt$, i.e. $E[dq_t] = \lambda k dt$. The model parameters α and σ represent the drift and volatility parameters respectively. Here, dq_t, dz and ϕ are assumed to be independent. This model was used for stocks where the effect of common news in the stock prices is represented by GBM while the effect of rare events corresponds to a Poisson jump. Using a compensated Poisson process where

the jumps are non-systematic and the size of jumps is from a lognormal distribution, Merton (1976) derived a closed form formula for European options.

Another work that presents a multiple factor model with jump diffusion process is Dias & Rocha (1999), where the authors proposed a combination of Poisson process and MRM to represent the stochastic behavior of oil prices in real options valuation as

$$\frac{dS_t}{S_t} = [\eta(\bar{S} - S_t)dt - \lambda k]dt + \sigma dz + dq_t, \tag{2}$$

where η is a speed of reversion parameter; $\bar{S}$ is the equilibrium level to which the process reverts in the long run. Here, $E\left[dS_t/S_t\right] = \eta(\bar{S} - S_t)dt$ because $k = E[\phi - 1]$. In this model, similarly to Merton (1976), the common news would cause marginal changes in oil prices, whereas abnormal events (such as crisis, wars and economic booms) would cause discrete jumps. The jumps can be systematic, which does not allow to obtain a risk neutral portfolio, or non-systematic, which allows the use of contingent claims.

There are also many papers proposing the combination of MRM and GBM in order to represent the stochastic behavior of commodity prices; see Gibson & Schwartz (1990), Schwartz (1997), Pindyck (1999), and Schwartz & Smith (2000). These papers claim that besides MRM factor, price processes of some commodities may also have a stochastic upward trend factor. In practical terms, this trend factor would tend to increase the equilibrium level to which the process reverts in the long run as time passes. These increases would have additional motivations to momentary mismatches of supply and demand (captured by MRM) and they would be caused by the progressive exhaustion of natural resources and incremental costs related to new requirements of environmental laws, among other issues. At the same time the improvements in the exploration and production technologies can impose a downward trend of the commodity prices.

Gibson & Schwartz (1990) proposes a two-factor model for pricing financial and real assets contingent on the price of oil, in which the factors are the spot price of oil that follows a GBM and the instantaneous convenience yield that follows a MRM. As explained by the authors, *"the notion of convenience yield, viewed as a net "dividend" accruing to the owner of physical commodity at the margin, has already proven to drive the relationship between future and spot prices of many commodities."* Nevertheless, in order to justify the assumption of stochastic convenience yield the authors postulate that *"the theory of storages posits an inverse relationship between the level of inventories and the net convenience yield which suggests that a constant convenience yield assumption will only hold under very restrictive assumptions."* In Gibson & Schwartz (1990) model, it is assumed that the spot price of oil S_t and the net convenience yield

δ_t follow a joint diffusion process

$$\begin{aligned} dS_t/S_t &= \mu dt + \sigma_1 dz_1, \\ d\delta_t &= \kappa(\alpha - \delta_t)dt + \sigma_2 dz_2, \end{aligned} \tag{3}$$

where μ is the drift parameter of oil price process; σ_1 is the volatility parameter of oil price process; dz_1 and dz_2 are Wiener increments; α is the equilibrium level to which the convenience yield process reverts in the long run; κ is the speed of reversion parameter of convenience yield; σ_2 is the volatility parameter of convenience yield process. Here, dz_1 and dz_2 are correlated and $E[dz_1 dz_2] = \rho dt$, where ρ denotes the correlation coefficient between the two Wiener increments. In order to calibrate the parameters of the MRM (convenience yield process) and the correlation the authors used the seemingly unrelated regression model in conjunction with unrestricted regression model which was used to calibrate the parameters of GBM (oil spot price process). The risk premium of convenience yield was estimated exogenously using computational numerical techniques.

Schwartz (1997) considered three different models for commodity prices with applications to commodity derivatives and commodity production assets valuation. The first model is one-factor model where the log of the price is a MRM. The second model has two stochastic factors and it is similar to the model developed by Gibson & Schwartz (1990) where the convenience yield follows a MRM while the price of commodity follows a GBM. In the third model, in addition to the stochastic commodity price and convenience yield, the author also considered the risk free rate as stochastic (following a MRM). The corresponding model for risk neutral porcesses is

$$\begin{aligned} dS_t &= (r_t - \delta_t)S_t dt + \sigma_1 S_t dz_1^*, \\ d\delta_t &= \kappa(\alpha^* - \delta_t)dt + \sigma_2 dz_2^*, \\ dr_t &= a(m^* - r_t)dt + \sigma_3 dz_3^*, \\ E[dz_1^* dz_2^*] &= \rho_1 dt,\ E[dz_1^* dz_3^*] = \rho_2 dt,\ E[dz_2^* dz_3^*] = \rho_3 dt, \end{aligned} \tag{4}$$

where a is the speed of reversion of the risk free rate; m^* is the equilibrium level to which the risk free rate reverts in the long run. The difficulty with implementation of the commodity price models is that some factors are not directly observable. Often, even the spot price of the commodities is not observable. Other kinds of variables, such as the instantaneous convenience yield, are even harder to estimate. On the other hand, futures prices of commodities are negotiated in many currencies and easy to observe. Schwartz (1997) derived state-space representations of the proposed models and applied Kalman filtering approach with the maximum likelihood method to estimate model parameters for copper, oil and gold. The analysis presented a strong evidence of mean

reversion component in copper and oil prices, but not in gold prices. It also points that the investments tend to be delayed if the mean reversion component is neglected in the real option analysis.

Pindyck (1999) analysed the price behavior of oil, coal and natural gas using 127-year time series and proposed, based on the historical performance, alternative ways for stochastic modeling of these commodity prices. The author comments that it would be ideal to be able to explain the price behavior of these commodities in structural terms by the movement of supply and demand in the market and the variables which determine them. Nevertheless he ponders that the structural models are not appropriately applicable to long term forecasts due to the difficulty related to explanatory variable forecasts of the models. As a result, the long run forecasts of energy commodity prices are made many times assuming that these prices grow on fixed taxes in real terms, in order to reflect the depletion of these natural resources reserves. Typically in such cases, additionally to the drift, stochastic shocks are incorporated in order to reflect the future prices uncertainties, which in practical terms would correspond to assumption a random walk with drift or a GBM for the prices. Alternatively, in many cases it is assumed that in short run the prices may wander randomly due to the momentary pressures of supply and demand, but in the long run they tend to converge to their production marginal cost, which would mean that prices follow a MRM. The identification and choice of the process which best represents the price behavior of such commodities have serious implications in the project valuation, mainly in cases when the real options are being considered in the projects. The Variance Ratio Tests applied to the price series suggested the presence of mean reversion components, despite the difficulty in rejecting the unit root (which would be similar to difficulty in rejecting the GBM). Therefore, at first, the author proposes a mean reversion model where the mean has a quadratic deterministic trend in order to incorporate the increase in time of the production marginal cost of commodities

$$dx_t = [-\gamma(x_t - \alpha_0 - \alpha_1 t - \alpha_2 t^2) + \alpha_1 + 2\alpha_2 t]dt + \sigma dz. \tag{5}$$

Here x_t is the log of commodity price; α_0, α_1, α_2 are the parameters of the quadratic trend of the log prices; and σ is the volatility parameter. Later, the author extended the model so that the level and slope could fluctuate stochastically, and proposed Kalman filter as the adequate approach to the parameter calibration.

Schwartz & Smith (2000) proposed a model with two stochastic factors χ_t and ξ_t (correlated and unobservable) to describe the behavior of commodity prices. The sum of these factors forms the log of commodity prices $\ln S_t = \chi_t + \xi_t$. The first factor, χ_t, is a MRM with null mean reflecting the short run deviations of prices, caused by momentary

mismatches of supply and demand of the commodities. The second factor, ξ_t, represents the long run tendency of prices, influenced by the progressive exhaustion of natural resources and incremental costs related to new requirements of environmental laws, among other issues. Differently from other multiple factor models, this one does not consider the stochastic convenience yield factor, nevertheless the authors commented that it is equivalent to Gibson & Schwartz (1990) with an appropriate calibration. Specifically, the evolution of χ_t and ξ_t is given by the following model

$$\begin{aligned} d\chi_t &= -\kappa\chi_t dt + \sigma_\chi dz_\chi, \\ d\xi_t &= \mu_\xi dt + \sigma_\xi dz_\xi, \\ E[dz_\xi dz_\chi] &= \rho dt, \end{aligned} \tag{6}$$

where κ is the mean reversion parameter of short deviations; σ_χ is the volatility parameter of the short run changes in prices; μ_ξ is the drift parameter of the long run price tendency; σ_ξ is the volatility parameter of the long run price tendency; ρ is the correlation parameter of the two factor increments. The authors estimated the parameters by fitting futures prices of commodities using state-space approach with Kalman filter method.

3 Methods for selecting stochastic processes

Different approaches have been suggested for valuation of real options. The real option analysis is more complex than standard option pricing in financial markets. The difficulties come from the facts such as the asset underlying the option may not be a tradable asset; the investment project can have controllable or uncontrollable cashflows; the project can be or cannot be actively managed; for overview of these issues, see e.g. Sick and Gamba (2005). In general, the evolution of the underlying asset in real time is modelled by some stochastic process (referred to as *real process*) but evaluation of the fair price of financial derivatives driven by this underlying is done under the risk adjusted process (referred to as *risk neutral process*). Roughly speaking, there are two types of models for pricing derivatives: *mark-to-market models* and *spot price models*. In mark-to-market models, the modeller fits the risk neutral process to match exactly a set of market instruments traded today such as today's prices of futures and vanillas. In spot price models, we fit both the real and risk neutral processes to the historical data (e.g. observed futures and vanilla options over some period of time), i.e. we do not fit exactly the market instruments exactly on any specific trading date.

It is important to note that one can assume different mark-to-market risk neutral processes that will match some liquid instruments exactly but will produce different

prices for illiquid derivatives. Also, the model parameters estimated to match say futures curve for a specific trading date will have to change for another trading date due to the change in futures curve. Without entering into further debate, modelling of both the real process and risk neutral process of the underlying are important for real option valuation and we believe that spot models are more appropriate. In this section, after some discussion on spot price and mark-to-market models we present methods that can be used for *model diagnostic*, *model selection* and *model assessment*.

3.1 Mark-to-market models

A simple example of mark-to-market model is geometric Brownian motion with time dependent drift and volatility, i.e. risk neutral process

$$dS_t/S_t = \mu(t)dt + \sigma(t)dz_t^*, \tag{7}$$

where z_t^* is the standard Brownian motion, $\mu(t)$ is calculated to match futures curve, $\sigma(t)$ is calculated to match vanilla options. Specifically, for this model, the futures price at $t = 0$ with maturity T is $F(0,T) = S_0 \exp(\int_0^T \mu(\tau)d\tau)$ and thus

$$\mu(t) = \frac{1}{t}\frac{d\ln(F(0,t)/S_0)}{dt}.$$

Another example, more relevant to modelling commodities is presented in Clewlow and Strickland (1999), where the assumed risk neutral model for futures is

$$dF(t,T) = F(t,T)\sigma e^{-\alpha(T-t)}dz_t^*, \tag{8}$$

which is used to derive the corresponding spot model for $S_t = F(t,t)$

$$\begin{aligned} \frac{dS_t}{S_t} &= \alpha(\mu(t) - \ln S_t)dt + \sigma dz_t^*, \\ \mu(t) &= \frac{\partial \ln F(0,t)}{\partial t} + \ln F(0,t) + \frac{\sigma^2}{4}(1 - e^{-2\alpha t}). \end{aligned} \tag{9}$$

This model will match today's market futures curve exactly because the mean reverting level in spot price is a function of time derived from the futures curve $F(0,t)$. Vanilla prices can be easily calculated using Black-Scholes formula with the variance replaced by $\frac{1}{2}\sigma^2(1 - \exp(-2\alpha(T-t)))/\alpha$ because $\ln S_T$ is normally distributed and thus vanilla options prices can be used easily to estimate σ and α.

Note that time dependent drift $\mu(t)$ derived from the current futures curve $F(0,t)$ cannot be used for another trading date due to the change in futures curve between trading dates in real time. Generally speaking, traditional statistical approach cannot be used to validate this type of model because there is no observation/measurement

errors between observed and model predicted prices. However, one can validate hedging strategy, i.e. calculate difference between replication portfolio and instrument (hedging error) using historically observed prices and compare the models using, for example, the mean squared error for the hedging error.

3.2 Spot price models

In general, in spot price models, we assume that spot price S_t is some function of underlying state variables $\boldsymbol{Y}_t$ that can be observable or not observable. Then we have to assume a stochastic processes for $\boldsymbol{Y}_t$; these are typically modelled as continuous time Ito processes in real time

$$dY_t^{(i)} = \mu^{(i)}(\boldsymbol{Y}_t, t)dt + \sigma^{(i)}(\boldsymbol{Y}_t, t)dz_i, \tag{10}$$

where $\mu^{(i)}(\boldsymbol{Y}_t, t)$ and $\sigma^{(i)}(\boldsymbol{Y}_t, t)$ are the drift and volatility of $Y_t^{(i)}$ that can be functions of the underlyings $\boldsymbol{Y}_t$ and time t, and $E[dz_i dz_j] = \rho_{ij}dt$. The risk neutral process which is used to value derivatives (i.e. options, futures, etc) driven by $\boldsymbol{Y}_t$ is obtained from no-arbitrage considerations. In general, it can be written as

$$dY_t^{(i)} = (\mu^{(i)}(\boldsymbol{Y}_t, t) - \lambda^{(i)}(\boldsymbol{Y}_t, t)\sigma(\boldsymbol{Y}_t, t))dt + \sigma(\boldsymbol{Y}_t, t)dz_i^*, \tag{11}$$

where $\lambda^{(i)}(\boldsymbol{Y}_t, t)$ is the risk premium that can be function of $\boldsymbol{Y}_t$ and time t, and $E[dz_i^* dz_j^*] = \rho_{ij}dt$. One can consider adding Poisson jumps to the above processes; also stochastic volatility can be one of the unobserved factors.

Two-factor model. The well known Schwartz and Smith (2000) two-factor model for commodity futures assumes that the log spot price of a commodity is $\ln S_t = \xi_t + \chi_t$, where χ_t is unobservable short-term deviation in prices and ξ_t is an unobservable long-term equilibrium price level with the following real processes

$$\begin{aligned} d\chi_t &= -\kappa\chi_t dt + \sigma_\chi dz_\chi, \\ d\xi_t &= \mu_\xi dt + \sigma_\xi dz_\xi, \\ E[dW_\chi dW_\xi] &= \rho dt. \end{aligned} \tag{12}$$

One can add a seasonality component $f(t)$ so that $\ln S_t = \xi_t + \chi_t + f(t)$. Then, the corresponding risk neutral process used to value futures and options is

$$\begin{aligned} d\chi_t &= (-\kappa\chi_t - \lambda_\chi)dt + \sigma_\chi dz_\chi^*, \\ d\xi_t &= (\mu_\xi - \lambda_\xi)dt + \sigma_\xi dz_\xi^*, \\ E[dz_\chi^* dz_\xi^*] &= \rho dt, \end{aligned} \tag{13}$$

where λ_χ and λ_ξ are the risk premia that typically assumed to be constant but in general can be functions of χ_t and ξ_t. If the risk premia are linear functions of state variables, then the price of the future contracts with maturity T is

$$F_{0,T} = E^*[S_T] = \exp(\xi_0 + B(T)\chi_0 + A(T)), \tag{14}$$

where expectation is calculated under the risk neutral process. $A(T)$ and $B(T)$ are simple functions of time. Given that log spot is normally distributed, $A(T)$ and $B(T)$ functions are easily calculated; for details see Schwartz and Smith (2000).

Remark 3.1 *It is important to note that the real process can be mean reverting while risk neutral is not mean reverting and vice versa.*

Three factor model. A popular extension of the above two-factor model is adding extra factor. Namely allowing drift of the long term factor to be stochastic itself

$$\begin{aligned}
&d\chi_t = -\kappa\chi_t dt + \sigma_\chi dz_\chi, \\
&d\xi_t = \mu_t dt + \sigma_\xi dz_\xi, \\
&d\mu_t = \gamma(\mu_\xi - \mu_t)dt + \sigma_\mu dz_\mu, \\
&E[dz_\chi dz_\xi] = \rho_{\chi\xi}dt, \;\; E[dz_\chi dz_\mu] = \rho_{\chi\mu}dt, \;\; E[dz_\mu dz_\xi] = \rho_{\mu\xi}dt.
\end{aligned} \tag{15}$$

Risk neutral processes are obtained by including risk premia into the drift terms. If risk premia are linear functions of the state variables then the price of the future contracts with maturity T becomes

$$F_{0,T} = E^*[S_T] = \exp(B_\xi(T)\xi_0 + B_\chi(T)\chi_0 + B_\mu(T)\mu_0 + A(T)), \tag{16}$$

where expectation is calculated under the risk neutral process and all functions of time $A(T)$, $B_\mu(T)$, $B_\xi(T)$ and $B_\chi(T)$ are easily calculated in closed form.

Multi factor affine models. In general, if the stochastic risk neutral model for the underlying variables $\boldsymbol{Y}_t = (Y_t^{(1)}, \ldots, Y_t^{(M)})$ is exponentially affine model, i.e. drifts and covariances in (11) are linear functions of $\boldsymbol{Y}_t$, and log spot price is a linear function with respect to $\boldsymbol{Y}_t$, then the futures price can always be calculated as

$$F_{T,t} = E^*[S_T] = \exp(B_1(T-t)Y_t^{(1)} + \ldots + B_M(T-t)Y_t^{(M)} + A(T-t)),$$

where functions of time $A(T-t), B_1(T-t), \ldots, B_M(T-t)$ are calculated from the system of ODEs.

State-space representation. In general, spot price multi factor models can be formulated as a state-space model

$$\boldsymbol{Y}_t = g(\boldsymbol{Y}_{t-1}, \boldsymbol{\varepsilon}_t); \quad \text{state equation}, \tag{17}$$

$$\boldsymbol{X_t} = h(\boldsymbol{Y}_t, \boldsymbol{\epsilon}_t); \quad \text{space/measurement equation}, \tag{18}$$

where $g(\cdot)$ and $h(\cdot)$ are some functions, $\boldsymbol{X_t} = (X_{t,1}, \ldots, X_{t,n})$ are observations on the trading date t (e.g. futures, vanillas, etc), $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\epsilon}_t$ are the vectors of serially independent normally distributed errors with zero mean and some covariances. In the case of the above two/three factor models, state and space equations are linear in $\boldsymbol{Y}_t$ and in error terms and can be written in the form

$$\boldsymbol{Y}_t = \boldsymbol{a} + \boldsymbol{G}\boldsymbol{Y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{19}$$

$$\ln \boldsymbol{F_t} = \boldsymbol{d}_t + \boldsymbol{c}\boldsymbol{Y}_t + \boldsymbol{\epsilon}_t, \tag{20}$$

where $\boldsymbol{F_t} = (F_{t,T_1}, \ldots, F_{t,T_n})$ are observed futures prices at trading date t; $\boldsymbol{a}$ and $\boldsymbol{\epsilon}_t$ are M dimensional vectors and $\boldsymbol{G}$ is $M \times M$ matrix; $\boldsymbol{d}_t$ and $\boldsymbol{\epsilon}_t$ are n-dimensional vectors and $\boldsymbol{c}_t$ is $n \times M$ matrix. Using Kalman filter procedure, one can calculate the density of the observed data (so-called likelihood) and fit the model using frequentist of Bayesian inference methods as described in the following sections; for application examples, see e.g. Schwartz & Smith (2000), Schwartz (1997). For a detailed discussion of state-space models and Kalman filter, see Harvey (1989). In the case of nonlinear relationships and non-Gaussian errors, one can try nonlinear Kalman filter or particle filter Monte Carlo methods; see Peters et al (2012)

The modeller should choose the model (i.e. the number of factors and model parameters for the risk neutral and real processes) and fit the model parameters to the observed data. In general, fitting can be done using the frequentist or Bayesian approaches. Using calibration results for different models, the user can make the model choice based on standard statistical criteria. Note that here, we aim to validate both the real process and risk neutral process of the underlying variables.

3.3 Frequentist approach

Fitting model parameters using data via the frequentist approach is a classical problem described in many textbooks. Under the frequentist approach a modeller says that parameters are fixed while their estimators have associated uncertainties that typically converge to zero when a sample size increases. The most popular approach to fit the parameters of the assumed model is the maximum likelihood method. Given the

model parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)$, assume that the joint density of data $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is $f(\boldsymbol{x}|\boldsymbol{\theta})$. Then the *likelihood function* is the joint density $f(\boldsymbol{x}|\boldsymbol{\theta})$ considered as a function of parameters $\boldsymbol{\theta}$, formally defined as

$$\ell_{\boldsymbol{x}}(\boldsymbol{\theta}) = f(\boldsymbol{x}|\boldsymbol{\theta}). \tag{21}$$

The maximum likelihood estimators $\widehat{\boldsymbol{\theta}}^{\mathrm{MLE}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{X})$ of the parameters $\boldsymbol{\theta}$ are the values of these parameters that maximize the log-likelihood function $\ln \ell_{\boldsymbol{x}}(\boldsymbol{\theta})$. Under the suitable regularity conditions, as the sample size increases, MLEs converge to the true value and are distributed from the K-variate normal distribution $\mathcal{N}_K(\cdot)$ as

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}^{\mathrm{MLE}} - \boldsymbol{\theta}) \to \mathcal{N}_K\left(0, [\boldsymbol{I}(\boldsymbol{\theta})]^{-1}\right). \tag{22}$$

Here, $[\boldsymbol{I}(\boldsymbol{\theta})]^{-1}$ is the inverse matrix of the expected Fisher information matrix whose matrix elements are typically approximated by the *observed information matrix*

$$\widehat{\boldsymbol{I}}(\widehat{\boldsymbol{\theta}})_{km} = -\frac{1}{n} \left. \frac{\partial^2 \ln \ell_{\boldsymbol{x}}(\boldsymbol{\theta})}{\partial\theta_k \partial\theta_m} \right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}. \tag{23}$$

That is, standard errors (and covariances between errors) of $\widehat{\boldsymbol{\theta}}^{\mathrm{MLE}}$ are estimated by covariance matrix $n^{-1}\boldsymbol{I}(\boldsymbol{\theta})^{-1}$. For precise details on regularity conditions and proofs, see Lehmann (1983, Theorem 6.2.1 and 6.2.3); these can also be found in many other books. Though very useful and widely used, these asymptotic approximations are usually not accurate enough for small samples, that is the distribution of parameter errors can be materially different from normal and MLEs may have significant bias.

Typically, maximisation of the likelihood (or minimisation of some distances in other methods) must be done numerically. Popular numerical optimisation algorithms include simplex method, Newton methods, expectation maximisation (EM) algorithm, and simulated annealing.

Example 3.1 *Consider geometric Brownian motion real process*

$$dS_t/S_t = \mu dt + \sigma dW_t, \tag{24}$$

observed at discrete equally spaced times $t_0, t_1, \ldots, t_n$ i.e.

$$\ln S_i = \ln S_{i-1} + (\mu - \sigma^2/2)\delta t + \sigma\sqrt{\delta t}\epsilon_i,$$

where $\delta t - t_i - t_{i-1}$ and ϵ_i are iid from the standard normal distribution. Then the likelihood of return data $R_i = \ln(S_i/S_{i-1})$, $i = 1, \ldots, N$ is

$$\ell(\mu, \sigma) \propto \prod_{i=1}^{N} \frac{1}{\sigma} \exp\left(-\frac{1}{2\sigma^2\delta t}(R_i - (\mu - \frac{1}{2}\sigma^2)\delta t)^2\right).$$

Maximizing $\ln \ell(\mu, \theta)$ *with respect to* μ *and* σ *gives the following MLEs*

$$\widehat{\sigma^2} = \frac{1}{N\delta t}\sum_{i=1}^{N}(R_i - \overline{R})^2, \quad \widehat{\mu} = \frac{1}{\delta t}\overline{R} + \frac{1}{2}\widehat{\sigma^2}, \quad \overline{R} = \frac{1}{N}\sum_{i=1}^{N}\ln R_i.$$

Example 3.2 *As an another example, consider the mean reverting real time process*

$$dS_t = (\omega - \theta S_t)dt + \sigma dW_t,$$

observed at discrete equally spaced times $t_0, t_1, \ldots, t_n$, *i.e.*

$$S_i = S_{i-1}\rho + \frac{\omega}{\theta}(1-\rho) + v\epsilon_i, \quad v^2 = \frac{1}{2\theta}(1-\rho^2).$$

Then the likelihood of S_i, $i = 1, \ldots, N$ *is*

$$\ell(\omega, \theta, \sigma) \propto \prod_{i=1}^{N}\frac{1}{v}\exp\left(-\frac{\epsilon_i^2}{2v^2}\right).$$

Maximizing $\ln \ell(\omega, \theta, \sigma)$ *with respect to* (ω, θ, σ) *gives the following MLEs*

$$\widehat{\rho} = \frac{1}{det}\left(\sum_i S_i \sum_i S_{i-1} - N\sum_i S_i S_{i-1}\right),$$

$$\widehat{\mu} = \frac{1}{det}\left(\sum_i S_{i-1}\sum_i S_i S_{i-1} - \sum_i S_{i-1}^2 \sum_i S_i\right),$$

$$det = \sum_i S_i \sum_i S_i - N\sum_i S_i^2.$$

Using the above estimators, calculate $\hat{v}^2 = \frac{1}{N}\sum_i \hat{\epsilon}_i^2$ *and finally* $\hat{\theta} = -\ln\hat{\rho}/dt$, $\hat{\omega} = \hat{\mu}\hat{\theta}/(1-\hat{\rho})$, *and* $\sigma^2 = 2\hat{\theta}\hat{v}^2/(1-\hat{\rho}^2)$.

For a general multifactor models the likelihood is not so easy to calculate. However, in the case of state variables from Gaussian distribution and linear measurement equation, i.e. linear state-space model (19), the likelihood can be calculated using Kalman filter recursion. In particular, using Kalman filter procedure, one can calculate the probability density function of $\boldsymbol{F_t}$ for given $\boldsymbol{F_{t-1}}$, i.e. $f(\boldsymbol{F_t}|\boldsymbol{F_{t-1}})$. Then the likelihood of all data is

$$\ell_{\boldsymbol{F_{1:T}}}(\boldsymbol{\theta}) = \prod_{t=1}^{T} f(\boldsymbol{F_t}|\boldsymbol{F_{t-1}}), \tag{25}$$

where $\boldsymbol{\theta}$ are models parameters (drift, volatility, correlations and risk premia). Once the likelihood function is calculated, the parameters can be estimated using maximum likelihood method with numerical optimization. One can also use Bayesian approach with Markov chain Monte Carlo (MCMC) methods described in the next section. In the case of nonlinear relationships and non-Gaussian errors, one can try nonlinear Kalman filter or particle filter Monte Carlo methods, see Peters et al (2012), but this goes beyond the purpose of this paper.

3.4 Bayesian approach

There is a broad literature covering Bayesian inference and its applications; for a good introduction, see e.g. Berger (1985). In the Bayesian approach, both data and parameters are considered to be random. A convenient interpretation is to think that parameter is a random variable with some distribution and the true value (which is deterministic but unknown) of the parameter is a realisation of this random variable. Consider a random vector of data $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ whose density, for a given vector of parameters $\boldsymbol{\theta}$, is $f(\boldsymbol{x}|\boldsymbol{\theta})$. Then the joint density of the data and parameters is

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|\boldsymbol{x})f(\boldsymbol{x}), \tag{26}$$

where $\pi(\boldsymbol{\theta})$ is the density of parameters (a so-called *prior density*); $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ is the density of parameters given data $\boldsymbol{X} = \boldsymbol{x}$ (a so-called *posterior density*); $f(\boldsymbol{x}, \boldsymbol{\theta})$ is the joint density of the data and parameters; $f(\boldsymbol{x}|\boldsymbol{\theta})$ is the density of the data given parameters $\boldsymbol{\theta}$, i.e. it is a likelihood function $\ell_{\boldsymbol{x}}(\boldsymbol{\theta}) = f(\boldsymbol{x}|\boldsymbol{\theta})$; $f(\boldsymbol{x})$ is the marginal density of $\boldsymbol{X}$. If $\pi(\boldsymbol{\theta})$ is continuous, then $f(\boldsymbol{x}) = \int f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ and if $\pi(\boldsymbol{\theta})$ is a discrete, then the integration should be replaced by a corresponding summation.

Using (26), the posterior density can be calculated as

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})/f(\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{27}$$

Here, $f(\boldsymbol{x})$ plays the role of a normalisation constant and the posterior can be viewed as a combination of prior knowledge (contained in $\pi(\boldsymbol{\theta})$) with information from the data (contained in the likelihood $f(\boldsymbol{x}|\boldsymbol{\theta})$). Using the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{x})$, one can easily construct a probability interval for $\boldsymbol{\theta}$ to contain the true value with the required probability, which is the analogue for confidence intervals under the frequentist approach. Sometimes the posterior density can be calculated in closed form, but in general, one should use Gaussian approximation or MCMC methods.

Gaussian Approximation for Posterior. For a given data realisation $\boldsymbol{X} = \boldsymbol{x}$, denote the mode of the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ by $\widehat{\boldsymbol{\theta}}$. If the prior is continuous at $\widehat{\boldsymbol{\theta}}$, then a Gaussian approximation for the posterior is obtained by a second-order Taylor series expansion around $\widehat{\boldsymbol{\theta}}$:

$$\ln \pi(\boldsymbol{\theta}|\boldsymbol{x}) \approx \ln \pi(\widehat{\boldsymbol{\theta}}|\boldsymbol{x}) + \frac{1}{2}\sum_{i,j} \left.\frac{\partial^2 \ln \pi(\boldsymbol{\theta}|\boldsymbol{x})}{\partial\theta_i \partial\theta_j}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} (\theta_i - \widehat{\theta}_i)(\theta_j - \widehat{\theta}_j). \tag{28}$$

Under this approximation, $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ is a multivariate normal distribution with the mean $\widehat{\boldsymbol{\theta}}$ and covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{I}^{-1},\ (\boldsymbol{I})_{ij} = -\frac{\partial^2 \ln \pi(\boldsymbol{\theta}|\boldsymbol{x})}{\partial\theta_i \partial\theta_j}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}. \tag{29}$$

In the case of improper constant priors, this approximation is comparable to the Gaussian approximation for the MLEs (22). Also, note that in the case of constant priors, the mode of the posterior and the MLE are the same. This is also true if the prior is uniform within a bounded region, provided that the MLE is within this region.

Once the posterior density $\pi(\theta|\boldsymbol{x})$ is found, for given data $\boldsymbol{X}$, one can define point estimators of θ. The mode and mean of the posterior are the most popular point estimators. The median of the posterior is also often used as a point estimator for θ.

Sometimes there is no prior knowledge about the model parameter $\boldsymbol{\theta}$, or we would like to rely on data only and avoid an impact from any subjective information. In this case we need a *noninformative prior* (sometimes called *vague prior*) that attempts to represent a near-total absence of prior knowledge. A natural noninformative prior is the uniform density

$$\pi(\theta) \propto \text{const} \quad \text{for all } \theta. \tag{30}$$

If parameter θ is restricted to a finite set, then this $\pi(\theta)$ corresponds to a proper uniform distribution. However, if the parameter θ is not restricted, then a constant prior is not a proper density (since $\int f(\theta)d\theta) = \infty$). Such a prior is called an *improper prior*. It is not a problem to use improper priors as long as the posterior is a proper distribution.

With respect to multi-factor models, once the likelihood is derived (e.g. the Kalman filter likelihood for linear models (25)), then one can use the Bayesian approach. Posterior cannot be found in closed form but various MCMC methods can be used to get samples from the posterior. The easiest to implement is Metropolis-Hastings algorithm which is a universal algorithm used to generate a Markov chain $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots\}$ with a stationary distribution $\pi(\boldsymbol{\theta}|\boldsymbol{x})$. It has been developed by Metropolis *et al* (1953) in mechanical physics and generalised by Hastings (1970) in a statistical setting. Given a density $\pi(\boldsymbol{\theta}|\boldsymbol{x})$, known up to a normalisation constant, and a conditional proposal density $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$, the method generates $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots\}$ using the following algorithm.

1. Initialise $\boldsymbol{\theta}^{(l=0)}$ with any value within a support of $\pi(\boldsymbol{\theta}|\boldsymbol{x})$;
2. For $l = 1, \dots, L$
   (a) Set $\boldsymbol{\theta}^{(l)} = \boldsymbol{\theta}^{(l-1)}$;
   (b) Generate a proposal $\boldsymbol{\theta}^*$ from $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(l)})$;

(c) Accept proposal with the acceptance probability

$$p(\boldsymbol{\theta}^{(l)}, \boldsymbol{\theta}^*) = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*|\boldsymbol{x})q(\boldsymbol{\theta}^{(l)}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{(l)}|\boldsymbol{x})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(l)})}\right\}, \tag{31}$$

i.e. simulate U from the uniform distribution function $\mathcal{U}(0,1)$ and set $\boldsymbol{\theta}^{(l)} = \boldsymbol{\theta}^*$ if $U < p(\boldsymbol{\theta}^{(l)}, \boldsymbol{\theta}^*)$. Note that the normalisation constant of the posterior does not contribute here;

3. Next l (i.e. do an increment, $l = l + 1$, and return to step 2).

3.5 Model diagnostic checking

Once the model parameters are calibrated, the model assumptions should be checked. Typical assumptions include independence and normality of the model residuals. For example, for linear state space model (19), one should check that error terms ε_t and ϵ_t are serially independent and are from the standard normal distribution. For this task, the following statistical methods are often used.

- **Unit root testing**. In general, unit root test is defined for autoregressive process of the order k. Here, for simplicity we consider $k = 1$, i.e. the model

$$x_t = \alpha x_{t-1} + \delta_t,$$

where δ_t are serially independent normal variables. Unit root testing is testing of null hypothesis that $\alpha = 1$. If $|\alpha| < 1$, then time series x_t is stationary. If null hypothesis is rejected, then we can estimate α by some fitting procedure such as maximum likelihood. There are many tests for unit root such as Dickey-Fuller test and augmented Dickey Fuller tests. One can also perform Bayesian inference approach and estimate the posterior for α.

- **Q-Q plots**. It is also common to check quantiles of the sample (y-coordinate) against model assumed quantiles (x-coordinate), this (x, y) plot is referred to as Q-Q plot. If model assumption is correct then the points of the plot should be close to $x = y$ line. For example, if we check that $x_1, \ldots, x_N$ are from the standard normal distribution, then we plot the sample order statistics $x_{(1)}, \ldots, x_{(N)}$ against the quantiles of the standard normal distribution $y_i = F_N^{-1}((i - 0.5)/N)$, $i = 1, \ldots, N$. If the assumption of normality is valid then correlation R between y_i and x_i should be close to one. Formal percentage points for the R^2 for samples from normal distribution is given in Shapiro and Francia (1972). For example for $N = 200$, $\Pr[R^2 < 0.987] = 0.05$. One can also apply popular goodness of fit tests including Kolmogorov-Smirnov, Anderson-Darling and chi-square tests.

- **Autocorrelation**. In addition to visual inspection of the model residuals, it is useful to monitor the serial correlation. For a given sample $x_1, \ldots, x_N$, the autocorrelation at lag k is estimated as

$$\widehat{\mathrm{ACF}}(k) = \frac{1}{N\widehat{\sigma}^2} \sum_{i=1}^{N-k} (\theta_i - \widehat{\mu})(\theta_{i+k} - \widehat{\mu}), \tag{32}$$

  where $\widehat{\mu}$ and $\widehat{s}^2$ are the mean and variance of a sample $x_1, \ldots, x_N$. Of course, it is biased estimate because we use sample estimators $\widehat{\mu}$ and $\widehat{s}^2$, and model parameter point estimators to calculate the residuals, but for large N and consistent point estimators it will converge to the true autocorrelation. It is possible to estimate the variance of the autocorrelations due to the finite sample. For example, if x_i are iid, then for large N, $\widehat{\mathrm{ACF}}(1), \ldots, \widehat{\mathrm{ACF}}(m)$ are iid normal variables with zero mean and variance $1/N$. That is, autocorrelations should be within bounds $\pm 1.96/\sqrt{N}$ with 0.95 confidence. This is usually used to check if residuals are iid. It is also a good idea to check absolute residuals. Often, returns in financial time series exhibit small autocorrelations while absolute returns have significant autocorrelations. This is typically an indication of time dependent volatility and can be removed by GARCH models for volatility if required depending on time horizon of the model use.

3.6 Model Selection

Given several competing models that passed diagnostic check, the modeller should decide which model to be used. Here, one can use the following procedures depending on the calibration approach selected to fit the model. Typically, under the frequentist approach, the modeller takes likelihood ratio tests and Akaike Information Criterion; under the Bayesian approach, the modeller often calculates the Bayes factors and Deviance Information Criterion. These are briefly described below.

- **In-sample errors** Once the statistical model parameterized by $\boldsymbol{\theta}$ is fitted to a data sample, we can calculate the difference between model predicted and observed values in the data sample. Typically one calculates the following quantities between observations X_t and predicted values $X_t^{pred} = \mathrm{E}[X_t | X_1, \ldots, X_{t-1}; \widehat{\boldsymbol{\theta}}]$:
  - squared correlation coefficient/coefficient of determination/R-squared (R^2), $(Corr[X_t, X_t^{pred}])^2$; larger value indicates better model;
  - mean squared error (MSE), $E[(X_t - X_t^{pred})^2]$; lower value indicates better model;

- Root mean square error (RMSE), $\sqrt{E[(X_t - X_t^{pred})^2]}$; lower value indicates better model;
- mean absolute percentage error, $E[|X_t - X_t^{pred}|/X_t]$; lower value indicates better model.

Here, the model parameters estimated $\widehat{\boldsymbol{\theta}}$ are obtained using the full dataset $X_{1:T}$.

- **Likelihood Ratio test**. It is a statistical test comparing two models. The test statistic for the null model with parameters $\widehat{\boldsymbol{\theta}}^{(0)}$ and alterative model with $\widehat{\boldsymbol{\theta}}^{(1)}$ is

$$LR = -2\ln\left(\ell(\widehat{\boldsymbol{\theta}}^{(0)})/\ell(\widehat{\boldsymbol{\theta}}^{(1)})\right), \tag{33}$$

where $\ell(\widehat{\boldsymbol{\theta}}^{(0)})$ and $\ell(\widehat{\boldsymbol{\theta}}^{(1)})$ are the likelihoods of the models. The distribution of statistic LR is chi-squared distribution with degrees of freedom $m_1 - m$, where m_0 and m_1 are the number of parameters in the null model and alternative model respectively. The models should be nested, i.e. more complex model can be reduced to a simpler model via constraints on the parameters.

- **Akaike Information Criterion (AIC)**. It is a measure of the relative goodness of fit of a statistical model introduced by Akaike (1983):

$$AIC = 2m - 2\ell(\widehat{\boldsymbol{\theta}}), \tag{34}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ are model parameters, m is the number of parameters, and $\ell(\boldsymbol{\theta})$ is likelihood function for the data maximized at $\widehat{\boldsymbol{\theta}}$. Point estimators $\widehat{\boldsymbol{\theta}}$ are the maximum likelihood estimators. The best model within a set of K candidate models for the data corresponds to the smallest $AIC^* = \min(AIC_1, \ldots, AIC_K)$. Note that AIC penalises for the increase in the number of parameters while rewards for goodness of fit. The quantity $\exp((AIC^* - AIC_k)/2)$ can be interpreted as the relative likelihood of the ith model. It looks similar to the likelihood ratio test if the number of parameters in the candidate models is the same but note that likelihood ratio test is used for nested models. To account for the number of observations N used to fit the model, the criteria is adjusted as

$$\widetilde{AIC} = 2m - 2\ell_D(\widehat{\boldsymbol{\theta}}) + \frac{2m(m+1)}{N-m-1}. \tag{35}$$

- **Bayesian Information Criteria**. Bayesian information criterion (BIC) or Schwarz criterion is a criterion for model selection among a finite set of models. It is closely related to AIC and is an asymptotic result for the data from the exponential family distribution. Formally it is given by

$$BIC = -2 \ln \ell_D(\widehat{\boldsymbol{\theta}}) + m \ln N. \tag{36}$$

The model with the lower value of BIC is the one to be preferred. Note that there is no requirement for compared models to be nested.

- **Bayes factors**. Consider a model M with parameter vector $\boldsymbol{\theta}$. The model likelihood with data $\boldsymbol{x}$ can be found by integrating out the parameter $\boldsymbol{\theta}$

$$\pi(\boldsymbol{x}|M) = \int \pi(\boldsymbol{x}|\boldsymbol{\theta}, M)\pi(\boldsymbol{\theta}|M)d\boldsymbol{\theta}, \tag{37}$$

where $\pi(\boldsymbol{\theta}|M)$ is the prior density of $\boldsymbol{\theta}$ in M. Given a set of K competing models $(M_1, \ldots, M_K)$ with parameters $\boldsymbol{\theta}_{[1]}, \ldots, \boldsymbol{\theta}_{[K]}$ respectively, the Bayesian alternative to traditional hypothesis testing is to evaluate and compare the posterior probability ratio between the models. Assuming we have some prior knowledge about the model probability $\pi(M_i)$ (if no knowledge is available one can assign equal probabilities to the models), we can compute the posterior probabilities for all models using the model likelihoods

$$\pi(M_i|\boldsymbol{x}) = \frac{\pi(\boldsymbol{x}|M_i)\ \pi(M_i)}{\sum_{k=1}^{K} \pi(\boldsymbol{x}|M_k)\ \pi(M_k)}. \tag{38}$$

Consider two competing models M_1 and M_2, parameterised by $\boldsymbol{\theta}_{[1]}$ and $\boldsymbol{\theta}_{[2]}$ respectively. The choice between the two models can be based on the posterior model probability ratio, given by

$$\frac{\pi(M_1|\boldsymbol{x})}{\pi(M_2|\boldsymbol{x})} = \frac{\pi(\boldsymbol{x}|M_1)\ \pi(M_1)}{\pi(\boldsymbol{y}|M_2)\ \pi(M_2)} = \frac{\pi(M_1)}{\pi(M_2)} B_{12}, \tag{39}$$

where $B_{12} = \pi(\boldsymbol{x}|M_1)/\pi(\boldsymbol{x}|M_2)$ is the Bayes factor, the ratio of the posterior odds of model M_1 to that of model M_2. As shown by Lavine and Scherrish (1999), an accurate interpretation of the Bayes factor is that the ratio B_{12} captures the change of the odds in favour of model M_1 as we move from the prior to the posterior. A Bayes factor $B_{12} > 10$ is considered strong evidence in favour of M_1. Kass and Raftery (1995) give a detailed review of the Bayes factors. Typically, the integral (37) required by the Bayes factor is not analytically tractable, and sampling based methods must be used to obtain estimates of the model likelihoods. There are quite a few methods in the literature for direct computation of the Bayes factor or indirect construction of the Bayesian model selection criterion, both based on MCMC outputs. The popular methods are direct estimation of

the model likelihood thus the Bayes factor; indirect calculation of an asymptotic approximation as the model selection criterion; and direct computation of the posterior model probabilities, as discussed below. Also, given MCMC samples from the posterior distribution obtained through MCMC, there is a *reciprocal importance sampling estimator* (RISE) proposed in Gelfand and Dey (1994) to approximate the model likelihood that can be regarded as a generalization of the *harmonic mean estimator* suggested by Newton and Raftery (1994).

- **Direct calculation of model probabilities**. Accurate estimation of the required posterior distributions usually involves development of a Reversible Jump MCMC framework. This type of Markov chain sampler is complicated to develop and analyse. It goes beyond the scope of this paper but interested reader can find details in Green (1995). In the case of small number of models, Congdon (2006) suggests to run a standard MCMC for each model separately and use the obtained MCMC samples to estimate $\pi(M_k|\boldsymbol{x})$. It was adopted in Peters, et al (2009) for modelling claims reserving problem in the insurance. Using the Markov chain results for each model, in the case of equiprobable nested models, this procedure calculates the posterior model probabilities $\pi(M_i|\boldsymbol{x})$ as

$$\pi(M_i|\boldsymbol{x}) = \frac{1}{L}\sum_{l=1}^{L} \frac{f\left(\boldsymbol{x}|M_i, \boldsymbol{\theta}_{[i]}^{(l)}\right)}{\sum_{j=1}^{K} f\left(\boldsymbol{x}|M_j, \boldsymbol{\theta}_{[j]}^{(l)}\right)}, \tag{40}$$

where $\boldsymbol{\theta}_{[i]}^{(l)}$ is the MCMC posterior sample at Markov chain step l for model M_i, $f(\boldsymbol{x}|M_i, \boldsymbol{\theta}_{[i]}^{(l)})$ is the joint density of the data $\boldsymbol{x}$ given the parameter vector $\boldsymbol{\theta}_{[i]}^{(l)}$ for model M_i, and L is the total number of MCMC steps after burn-in period.

- **Deviance Information Criterion**. For a dataset $\boldsymbol{X} = \boldsymbol{x}$ generated by the model with the posterior density $\pi(\boldsymbol{\theta}|\boldsymbol{x})$, define the deviance

$$D(\boldsymbol{\theta}) = -2\ln\pi(\boldsymbol{x}|\boldsymbol{\theta}) + C, \tag{41}$$

where the constant C is common to all candidate models. Then the *deviance information criterion* (DIC) is calculated as

$$\begin{aligned} DIC &= 2\mathrm{E}[D(\boldsymbol{\theta})|\boldsymbol{X}=\boldsymbol{x}] - D(\mathrm{E}[\boldsymbol{\theta}|\boldsymbol{X}=\boldsymbol{x}]) \\ &= \mathrm{E}[D(\boldsymbol{\theta})|\boldsymbol{X}=\boldsymbol{x}] + (\mathrm{E}[D(\boldsymbol{\theta})|\boldsymbol{X}=\boldsymbol{x}] - D(\mathrm{E}[\boldsymbol{\theta}|\boldsymbol{X}=\boldsymbol{x}])), \end{aligned} \tag{42}$$

where $\mathrm{E}[\cdot|\boldsymbol{X}=\boldsymbol{x}]$ is the expectation with respect to the posterior density of $\boldsymbol{\theta}$. The expectation $\mathrm{E}[D(\boldsymbol{\theta})|\boldsymbol{X}=\boldsymbol{x}]$ is a measure of how well the model fits

the data; the smaller this is, the better the fit. The difference $\mathrm{E}[D(\boldsymbol{\theta})|\boldsymbol{X} = \boldsymbol{x}] - D(\mathrm{E}[\boldsymbol{\theta}|\boldsymbol{X} = \boldsymbol{x}])$ can be regarded as the effective number of parameters. The larger this difference, the easier it is for the model to fit the data. The DIC criterion favours the model with a better fit but at the same time penalises the model with more parameters. Under this setting the model with the smallest DIC value is the preferred model. DIC is a Bayesian alternative to *Akaike's information criterion*. For more details on the above-mentioned criteria, see e.g. Robert (2001, Chapter 7).

- **Model averaging**. If competing models have significant probabilities, then the modeller may choose to average the results across the models (weighted by model probabilities) instead of selecting the best model. Here, one can use the model probabilities implied by the above discussed AIC or Bayes factors.

3.7 Model assessment

Once the best model is selected using the above described statistical criteria, the modeller should perform the assessment of the selected model. Typically it involves estimating the prediction error on new data.

- **Out-of-sample or cross validation**. Cross-validation is a technique for assessing the accuracy of the prediction of the fitted statistical model. It involves partitioning a sample of data into complementary subsets, performing the analysis on one subset and validating the analysis on the other subset. To reduce variability, multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over the rounds. Often, one step prediction are calculated. Typically one calculates the following quantities between observations X_t and predicted values $X_t^{pred} = \mathrm{E}[X_t|X_1, \ldots, X_{t-1}]$:
  - squared correlation or pseudo R^2: $(Corr[X_t, X_t^{pred}])^2$;
  - mean quadratic error: $E[(X_t - X_t^{pred})^2]$;
  - mean absolute percentage error: $E[|X_t - X_t^{pred}|/X_t]$.

  Here, we assume that the model is parameterized by $\boldsymbol{\theta}$ estimated using one data subset and the above prediction errors are calculated for another data subset.

- **Hedging errors**. Given that risk neutral process used for the pricing of real option is based on assumption of replication portfolio, it is worthwhile to check the hedging errors (difference between the option/future and replicated portfolio) for observed data realization. This is especially critical for mark-to-market models.

- **Sensitivity**. The modeller should check the sensitivity of the model output (i.e. price of real option) to the inputs. For example, it is typical to observe that for a long investment project, short-term factor is not important and pricing can be done using a model for a long term factor only.

4 Empirical Applications

The data we use to test the models are weekly observations of futures prices of crude oil, contract CL traded on the New York Mercantile Exchange (NYMEX), from 23 November 1990 till 10 May 2013. For illustration example we use the first 12 contracts available for each observation date although for most of the dates there are about 20 contracts and for some dates there are more than 70 contracts. We fit the following two models (one-factor and two-factor models) for the log spot price of a commodity $\ln S_t$.

- **Model 1 (Two-factor model)**. Assume that $\ln S_t = \xi_t + \chi_t$, where χ_t is unobservable short-term deviation in prices and ξ_t is an unobservable long-term equilibrium price level with the following real processes

$$\begin{aligned} d\chi_t &= -\beta\chi_t dt + \sigma_\chi dz_\chi, \\ d\xi_t &= (\mu_\xi - \gamma\xi_t)dt + \sigma_\xi dz_\xi, \\ E[dW_\chi dW_\xi] &= \rho dt. \end{aligned} \tag{43}$$

  Corresponding risk neutral processes (used to value futures and options) are

$$\begin{aligned} d\chi_t &= (-\widetilde{\beta}\chi_t - \lambda_\chi)dt + \sigma_\chi dz^*_\chi, \\ d\xi_t &= (\widetilde{\mu}_\xi - \widetilde{\gamma}\xi_t)dt + \sigma_\xi dz^*_\xi, \\ E[dz^*_\chi dz^*_\xi] &= \rho dt. \end{aligned} \tag{44}$$

  This is just a well known Schwartz and Smith (2000) two-factor model extended to have mean reversion in a long term factor and more general (linear in state variables) risk premia. Under this risk neutral process, the price of the future contract at time t with maturity at time T can be easily calculated as

$$\begin{aligned} F_{t,T} &= E^*[S_T] = \exp(B_1(T-t)\xi_t + B_2(T-t)\chi_t + A(T-t)); \qquad (45) \\ B_1(\tau) &= \exp(-\widetilde{\gamma}\tau); \quad B_2(\tau) = \exp(-\widetilde{\beta}\tau); \\ B_0(\tau) &= \frac{\widetilde{\mu}_\xi}{\widetilde{\gamma}}(1-e^{\widetilde{\gamma}\tau}) - \frac{\widetilde{\lambda}_\chi}{\widetilde{\beta}}(1-e^{-\widetilde{\beta}\tau}) + \frac{\sigma_\xi^2}{4\widetilde{\gamma}}(1-e^{-2\widetilde{\gamma}\tau}) + \frac{\sigma_\chi^2}{4\widetilde{\beta}}(1-e^{-2\widetilde{\beta}\tau}) \\ &\quad + \frac{\sigma_\xi\sigma_\chi\rho}{\widetilde{\gamma}+\widetilde{\beta}}(1-e^{-(\widetilde{\gamma}+\widetilde{\beta})\tau}). \end{aligned}$$

- **Model 2 (One-factor model)**. Here we assume that there is only one-factor, i.e. $\ln S_t = \xi_t$, that follows real and risk neutral processes

$$\begin{array}{l} d\xi_t = (\mu_\xi - \gamma\xi_t)dt + \sigma_\xi dz_\xi, \\ d\xi_t = (\widetilde{\mu}_\xi - \widetilde{\gamma}\xi_t)dt + \sigma_\xi dz^*_\xi, \end{array} \tag{46}$$

correspondingly. Then, the price of the futures contract is

$$\begin{aligned} F_{t,T} &= E^*[S_T] = \exp(B_1(T-t)\xi_t + A(T-t)); \\ B_1(\tau) &= \exp(-\widetilde{\gamma}\tau); \quad B_0(\tau) = \frac{\widetilde{\mu}_\xi}{\widetilde{\gamma}}(1 - e^{\widetilde{\gamma}\tau}) + \frac{\sigma^2_\xi}{4\widetilde{\gamma}}(1 - e^{-2\widetilde{\gamma}\tau}), \end{aligned} \tag{47}$$

that can be obtained from (45) by setting $\chi_t = \lambda_\chi = \widetilde{\beta} = \sigma^2_\chi = \rho = 0$.

A statistically sound method to fit the above models is the Kalman filter procedure typically used in the academic literature. In this paper, we choose more simple, fast and easy to implement procedure suggested in Cortazar and Schwartz (2003). In the case of crude oil futures, this procedure produces results that are not materially different from the Kalman filter results. Of course the advantage of the Kalman procedure is that it calculates the model likelihood which is used to get the point estimates of the parameters, confidence intervals for the estimates and can be used to apply formal model selection criteria such as Akaike criteria. Under the simplified procedure, we have to resort to in-sample and out-of-sample tests. The fitting procedure estimates risk neutral mean reversion parameters $(\widetilde{\gamma}, \widetilde{\beta})$, volatilities $(\sigma_\xi, \sigma_\chi)$ and correlation ρ using historical covariances between futures contracts via nonlinear least square method. Then, risk neutral drift parameters $(\lambda_\chi, \widetilde{\mu}_\xi)$ are estimated by nonlinear least square method minimizing

$$\sum_i \sum_j [\ln F(t_i, T_j) - \ln F_{obs}(t_i, T_j)]^2$$

where unobservable factors ξ_{t_i} and χ_{t_i} for each trading date t_i are calculated in closed form by least square method minimizing,

$$\sum_j [\ln F(t_i, T_j) - \ln F_{obs}(t_i, T_j)]^2.$$

Finally, obtained time series for ξ_t and χ_t are used to estimate real process parameters $(\mu_\xi, \gamma, \beta, \sigma_\xi, \sigma_\chi, \rho)$ by maximum likelihood method for mean reversion process; see Example 3.2. The calibration for the one-factor model (Model 2) is easily obtained from this procedure by setting $\chi_t = \lambda_\chi = \widetilde{\beta} = \beta = \sigma^2_\chi = \rho = 0$. Calibration results are summarized in Table 1 and Table 2. The percentage root mean square error (RMSE) between model and market log prices (across different contracts and total across all

contracts and trading dates) is significantly smaller for Model 1 indicating that two-factor model is much superior than one-factor model. Other model selection statistics (R^2 and AIC) are also in favour of Model 1; see Table 1. Figure 1 shows prediction error for some contracts, estimated long and short factors, and predicted value for the 12th contract (F12) versus time.

Model 1 (Two-Factor Model)
$\beta = 6.06,\ \mu = 7.62,\ \gamma = 1.69,\ \sigma_\xi = 0.31,\ \sigma_\chi = 0.20,\ \rho = -0.26,\ \widetilde{\mu} = 1.04,\ \widetilde{\beta} = 6.73,\ \widetilde{\gamma} = 0.23$
RMSE=0.45%, R^2=0.999, AIC=-25175
Model 2 (One-Factor Model)
$\mu = 7.27,\ \gamma = 1.62,\ \sigma_\xi = 0.29,\ \widetilde{\mu} = 0.69,\ \widetilde{\gamma} = 0.15$
RMSE=1.65%, R^2=0.99, AIC=-18509

Table 1: Parameter estimates for two-factor model (Model 1) and one-factor model (Model 2). RMSE is the percentage root mean square error between model and market log prices across all contracts and all trading dates in the dataset.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Model 1	0.6%	0.7%	0.5%	0.4%	0.4%	0.3%	0.2%	0.2%	0.2%	0.3%	0.5%	0.6%
Model 2	4.2%	1.8%	0.8%	0.6%	0.7%	0.8%	0.8%	0.9%	1.0%	1.2%	1.4%	1.7%

Table 2: The percentage root mean square error (RMSE) between model and market log prices across different contracts. F1 corresponds to the 1st available contract, F2 to the 2nd available contract, etc.

5 Conclusions

We considered the use of multiple factor models for valuation of real options. The choice of underlying stochastic model is certainly important for valuation of real options especially for projects with long lifetime. There are many statistical tools that can help to resolve this issue. In this paper, we reviewed and discussed methods of model selection, model assessment and model diagnostics.

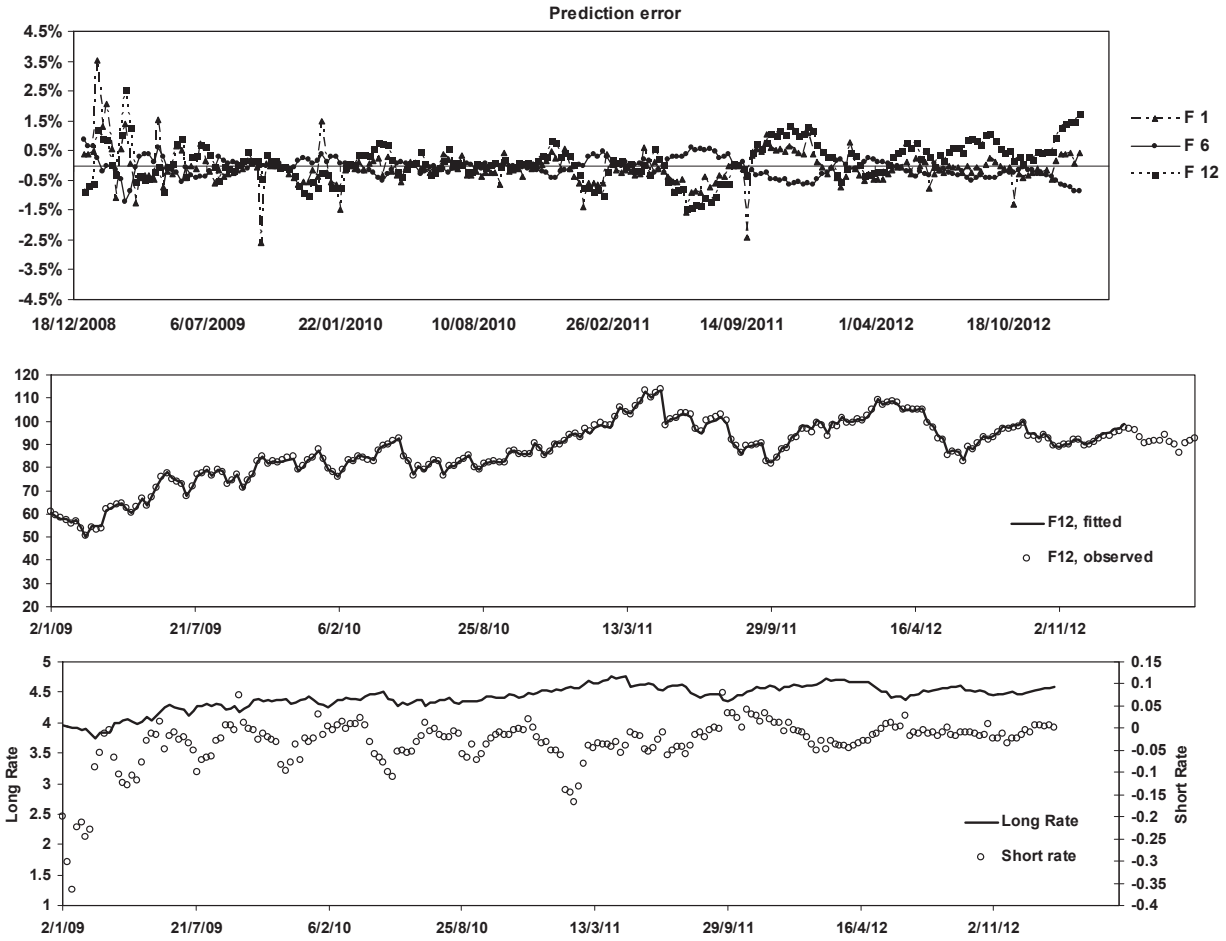

Figure 1: Fitting results for two-factor model (Model 1).

References

[1] Akaike, H. Information measure and model selection. *Bull. Int. Statist. Inst. 50* (1983), 277–290.

[2] Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York, 1985.

[3] Bhattacharya, S. Project valuation with mean-reverting cash flow streams. *Journal of Finance 33*, 5 (1978), 1317–1331.

[4] Black, F., and Scholes, M. The pricing of options and corporate liabilities. *Journal of Political Economy 81*, May-June (1973), 637–659.

[5] Brennan, M. J., and Schwartz, E. S. Evaluating natural resource investments. *The Journal of Business 58*, 2 (1985), 135–157.

[6] Clewlow, L., and Strickland, C. Valuing energy options in a one factor model fitted to forward prices. *Preprint* (1999).

[7] Congdon, P. *Bayesian Statistical Modelling*, 2nd ed. John Wiley & Sons, Ltd, 2006.

[8] Cortazar, G., and Schwartz, E. S. Implementing a stochastic model for oil futures prices. *Energy Economics 25* (2003), 215238.

[9] Cox, J., Ross, S., and Rubinstein, M. Option pricing: A simplified approach. *Journal of Financial Economics 7* (1979), 229–264.

[10] Dias, M. A. J., and Rocha, K. Petroleum concessions with extendible options using mean reversion with jumps to model oil prices. Paper presented at the 3rd Real Option Conference.

[11] Dixit, A., and Pindyck, R. *Investment under uncertainty.* Princeton University Press, Princeton, NJ, 1994.

[12] Gelfand, A., and Dey, D. Bayesian model choice: Asymptotic and exact calculations. *Journal of the Royal Statistical Society, series B 56* (1994), 501–514.

[13] Gibson, R., and Schwartz, E. S. Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance 45*, 3 (1990), 959–976.

[14] Green, P. Reversible jump MCMC computation and Bayesian model determination. *Biometrika 82* (1995), 711–732.

[15] Harvey, A. C. *Forecasting, structural time series models, and the Kalman filter.* Cambridge University Press, 1989.

[16] Hastings, W. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57* (1970), 97–109.

[17] Lavin, M., and Scherrish, M. Bayes factors: what they are and what they are not. *The American Statistician 53* (1999), 119–122.

[18] Lehmann, E. L. *Theory of Point Estimation.* Wiley & Sons, New York, 1983.

[19] Longstaff, F. A., and Schwartz, E. S. Valuing American options by simulation: a simple least-square approach. *The Review of Financial Studies 14*, 1 (2001), 113–147.

[20] McDonald, R., and Siegel, D. Investment and the valuation of firms when there is an option to shut down. *International Economic Review June* (1985), 331–49.

[21] McDonald, R., and Siegel, D. The value of waiting to invest. *The Quartely Journal of Economics 101* (1986), 707–728.

[22] Merton, R. C. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics 3* (1976), 125–144.

[23] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equations of state calculations by fast computing machines. *Journal of Chemical Physics 21* (1953), 1087–1091.

[24] Newton, M., and Raftery, A. Approximate Bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, series B 56* (1994), 1–48.

[25] Ozorio, L. M., Bastian-Pinto, C. L., Baydia, T. N., and Brandão, L. E. T. Reversão à média com tendência e opções reais na siderurgia. *Revista Brasileira de Finanças 10* (2012), 1–27.

[26] Paddock, J. L., Siegel, D. R., and Smith, J. L. Option valuation of claims on real assets: The case of offshore petroleum leases. *Quarterly Journal of Economics 103* (1988), 479–508.

[27] Peters, G. W., Briers, M., Shevchenko, P., and Doucet, A. Calibration and filtering for multi factor commodity models with seasonality: Incorporating

panel data from futures contracts. *Methodology and Computing in Applied Probability* (2012). DOI 10.1007/s11009-012-9286-7.

[28] Peters, G. W., Shevchenko, P. V., and Wüthrich, M. V. Model uncertainty in claims reserving within Tweedie's compound Poisson models. *ASTIN Bulletin 39*, 1 (2009), 1–33.

[29] Pindyck, R. S. The long-run evolution of energy prices. *Energy Journal 20*, 2 (1999), 1–27.

[30] Robert, C. P. *The Bayesian Choice*. Springer, New York, 2001.

[31] Schwartz, E., and Smith, J. E. Short-term variations and long-term dynamics in commodity prices. *Management Science 46,*, 7 (2000), 893–911.

[32] Schwartz, E. S. The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance 52*, 3 (1997), 923–973.

[33] Sick, G., and Gamba, A. Some important issues involving real options: An overview. *Preprint* (2005).