CONNING®

GEMS® Parameter Estimation Process and Methodology

Prepared for the NAIC

Contents

1 Introduction **3**

1.1 Economic Historical Data Base 4

1.2 Estimator 4

2 Maximum Likelihood Estimation **5**

2.1 Probability Density Function 6

2.2 Likelihood Function 7

2.3 Parameter Estimation 8

2.4 Maximizing the Likelihood Function 8

3 The Kalman Filter **9**

3.1 The Measurement and Filtration Problem 10

4 Kalman Filter and MLE for the Multi-Factor Square Root Process **11**

5 Target Setting and Final Parameter Setting **12**

6 Model Re-Estimation **12**

7 Conclusion **13**

8 References **14**

1 Introduction

This document is intended to give users of GEMS an understanding of the process for estimating the parameters of GEMS real world models. The document contains some high level background information aimed at a moderately technical audience, and some more detailed mathematical content aimed at reasonably expert readers.The information pertains only to the GEMS Expert View Parameterisation and not custom calibration that may have been created using tools supplied by Conning or any other method.

The estimation of the model parameters brings GEMS to life and allows control of the model output to ensure that the models span the maximum range of real world dynamics that are displayed in the market data as well as adequately capturing future events yet to be observed. The estimation of a model is a multi step process, but two considerations are central to the process;

- **Target Setting:** Analysis of the historical data record combined with expert judgment and economic analysis is used to determine what the desired statistical properties of the model should be (e.g. long term or ”steady state” mean, standard deviation etc.). Target setting involves analysing 25-30 years of historical data and applying expert judgment to determine the statistical properties that are most relevant to the current market. Target setting follows a well defined process which removes as much observer bias as possible.

- **Dynamics:** Information from the historical record is used to determine the desirable dynamics that simulated data should exhibit (e.g. yield curve shapes, mean reversion properties, jump frequencies and severity etc.). Model parameters governing dynamics are determined using analytical methods constrained on the targets and using 25-30 years of data where available.

These two considerations are related, and parameter estimation is performed in such a way as to simultaneously match the targets and produce realistic dynamics. In reality there is likely to exist a large number of parameter combinations which will match the model targets reasonably. Which combination of factors one chooses is vital, as this will govern how realistic the model is, and how well it is likely to perform out of sample. Fortunately for the models implemented within GEMS, a range of well understood analytical methods exists to determine the most appropriate and likely combination of parameters given the information contained within time series data.

To effect estimation and validation Conning has created a special purpose environment comprised of both automated and interactive processes including:

- Time Series Database: down loads the relevant economic and financial data on a daily basis. This database includes very long histories of data ($>$ 100 years) for some variables.

- Estimator: automated estimation of model parameters.

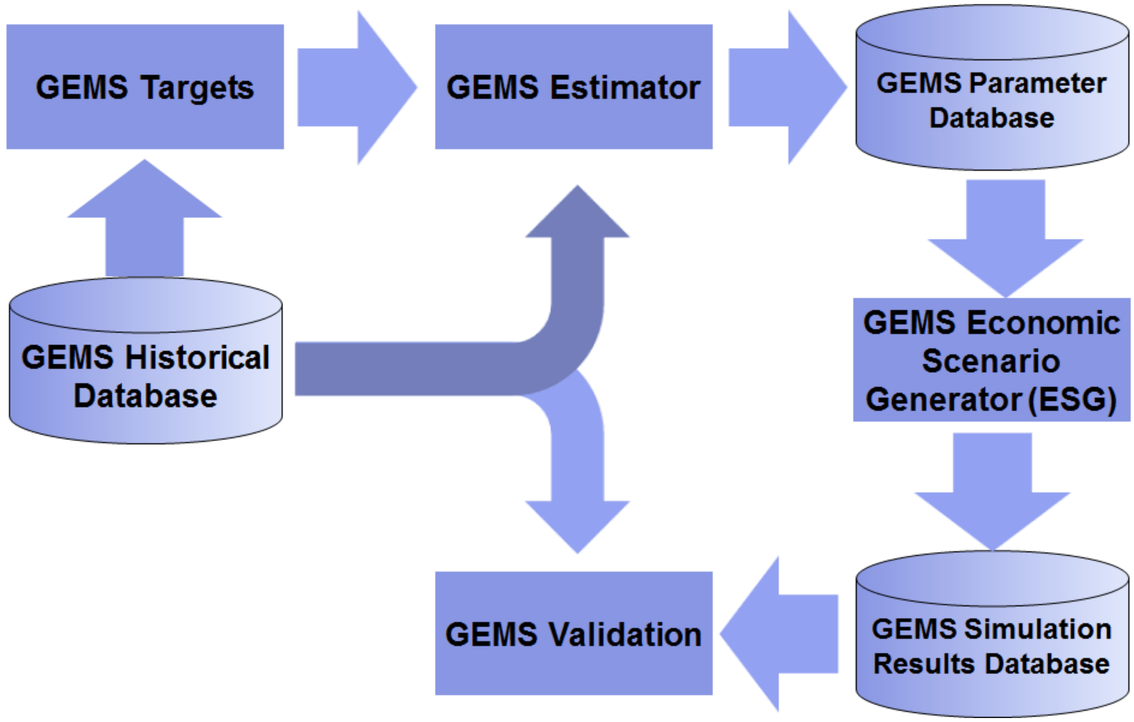

Figure 1. GEMS parameter estimation and validation process.

- Validator: a tool to check GEMS output against the historical record.

A schematic diagram of the process is shown in figure 1. In the following sections the estimation aspects are explained in more detail.

1.1 Economic Historical Data Base

Conning maintains a time series database (TSDB) that contains all the relevant market data that is used for estimating the GEMS parameters. The market data in the database is updated on a daily, a monthly or a quarterly basis (depending on the kind of data) with automated feeds from market data vendors (e.g. Bloomberg). The database is supplemented with deep histories.

If automated update process fails, for example some time series data is missing, then the automated updating application will try to retrieve the data over the next few days by generating further data request messages. If all these fail, the GEMS team gets an automated alert and the problem is managed manually.

The database is stored on a server in Hartford, Connecticut. It is mirrored on a daily basis on a server in Cologne.

1.2 Estimator

The estimator takes a combination of market data and target data and uses this information to determine the model parameters that represent the best fit to both. The following summarizes the methodology used to estimate the models discussed in this report. Further information on the relevant theory can be found in the reference section at the end of the report.

- **Government Yield Curve:** estimated using maximum likelihood in conjunction with a Kalman filter.
- **Common stock:** estimated using approximate maximum likelihood in conjunction with a Kalman filter.
- **Corporate Yields and Spreads:** estimated using a combination of maximum likelihood, Kalman filter and moment fitting.

Once estimation is complete, the parameters have to be made available to GEMS. This is accomplished with an application called the 'Populator' that loads the estimated parameters and historical data into a definition file that can be used within GEMS and ADVISE. In the following sections an introduction to the main principles of maximum likelihood estimation and the Kalman filter is given.

2 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is a standard approach to estimating model parameters in statistics. MLE is used in a wide variety of fields and is a powerful analytical technique for inferring the parameters of a model from a set of sample data. MLE is used because it has many optimal properties for parameter estimation namely:

- **Sufficiency -** the complete information about a given parameter of interest is contained in its MLE estimator.
- **Consistency -** the true parameter value that generated the modeled data is recovered asymptotically, (i.e. for sufficiently large data samples).
- **Efficiency -** the method delivers the lowest possible variance of parameter estimates.
- **Parameterization invariance -** the solution to MLE is independent of the parameterisation used.

The purpose of this section is to give users enough information to understand and adequately describe the method both internally and to external bodies such as regulators. More in depth information can be found in a number of reference materials given at the end of this paper [1,2,3]. We now proceed to introduce the two key concepts, the probability density function and the likelihood function, which must be understood before MLE can be fully described.

2.1 Probability Density Function

If we have a set of data represented by the data vector $y = (y_1....y_m)$, we can consider this from a statistical standpoint a random sample from an unknown population. The goal is to produce a model that represents the population that is most likely to have generated the data. The population can be conveniently thought of in terms of its corresponding probability distribution. Each probability distribution is associated with unique values of the model parameters. As the parameters change in value, different probability distributions will be generated. Formally, a model can be defined as the group of probability distributions spanned by the model parameters.

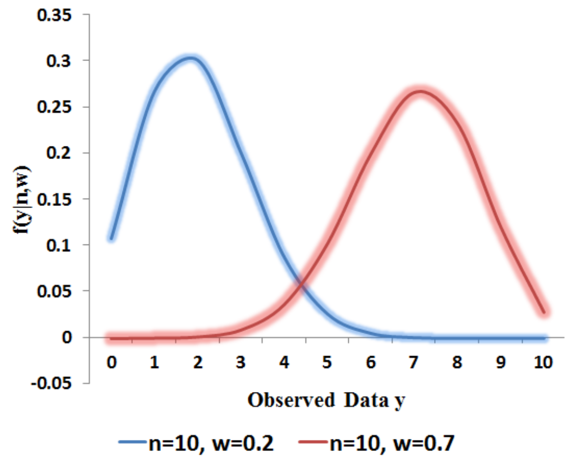

Figure 2. Binomial probability density functions for two parameterisations with w=0.2 and w=0.7 and a sample size n=10

Let $f(y|w)$ denote the probability density function (PDF) that defines the probability of observing data y given the parameter vector w. The parameter vector $w = (w_1....w_k)$ is defined on a multi-dimensional parameter space and contains all of the model parameters. If individual observations, y_i, are independent of one another, then according to probability theory, the PDF for the data $y = (y_1....y_m)$ given the parameter vector w can be expressed as the product of PDFs for individual observations;

$$f(y = (y_1, y_2, ..., y_m)|w) = f_1(y_1|w)f_2(y_2|w)...f_m(y_m|w) \tag{1}$$

As an illustrative example, consider a simple case with one observation and one model parameter, that is, $m = k = 1$. Suppose that the data y represents the number of successes in a sequence of 10 Bernoulli trials (e.g. tossing a biased coin 10 times) with the probability of a success on any one trial, w, of 0.2. The PDF for this process is given by;

$$f(y|n = 10, w = 0.2) = \frac{10!}{y!(10-y)!}0.2^y 0.8^{10-y} \tag{2}$$

where $y = (0, 1, ..., 10)$ is the number of successes after the 10 trials. This is of course the well known binomial distribution with parameters $n = 10$ and $w = 0.2$. The form of this PDF is shown in figure 2 along with the distribution if the parameter value w is changed to 0.7 in which case the PDF is given by;

$$f(y|n = 10, w = 0.7) = \frac{10!}{y!(10-y)!}0.7^y 0.3^{10-y} \tag{3}$$

The PDF for this particular process can then be generalized for all admissible values of w and n as;

$$f(y|n,w) = \frac{n!}{y!(n-y)!} w^y (1-w)^{n-y} \tag{4}$$

with $(0 \leq w \leq 1; y = 0, 1, ..., n)$, which as a function of y gives the probability of observing data y for a given value of n and w. The collection of all such PDFs generated by varying the parameters across its admissible range (in this case 0 to 1 for w, $n \geq 1$) defines the model of such a process.

2.2 Likelihood Function

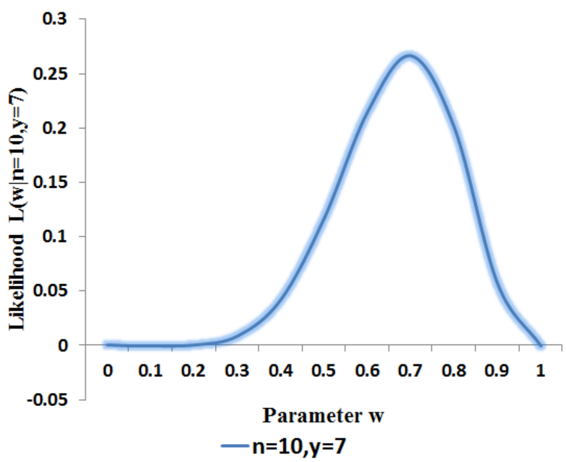

Figure 3. Binomial probability density functions for two parameterisations with w=0.2 and w=0.7 and a sample size n=10

For a model defined by a set of parameter values, the PDF will show that some outcomes are more probable than others. In the previous section, the PDF with $w = 0.2$ implies the outcome $y = 2$ (i.e. 2 successes) is more likely to occur than $y = 5$ (i.e. 5 successes) (0.302 vs. 0.026). In most practical modeling applications however (e.g. financial modeling), we have already observed the data. Accordingly, we must solve the inverse problem. That is, given the observed data and a model, find the PDF (described by the model parameters), among all the probability densities that the model prescribes, that is most likely to have produced the observation. To solve this problem, we must define a so called likelihood function. This is done by effectively reversing the roles of the data vector y and the parameter vector w in $f(y|w)$;

$$\mathcal{L}(w|y) = f(y|w) \tag{5}$$

The likelihood function $\mathcal{L}(w|y)$ represents the likelihood of the parameter w given the observed data y. For the example of a process defined by a binomial distribution given above, the likelihood function for $y = 7$ and $n = 10$ is given by;

$$\mathcal{L}(w|n=10, y=7) = \frac{10!}{7!3!} w^7 (1-w)^3 \tag{6}$$

with $(0 \leq w \leq 1)$. The form of this likelihood function is shown in figure 3.

It is interesting to consider the differences between the PDF $f(y|w)$ and the likelihood function $\mathcal{L}(w|y)$, illustrated in figures 2 and 3. The main difference is that the functions are defined on different axes, and therefore are not directly comparable to each other. The PDFs in figure 2 are a function of the data y given a particular set of parameter values w. The likelihood function however, is a function of the parameter given a particular set of observed data. In other words, the PDF tells us the probability of a particular observation for a given parameter, whereas the likelihood function tells us the likelihood of a particular parameter value for a given observed data set. In the example above the likelihood function is a curve because there is only one unknown parameter. For a model with two parameters, the likelihood function will be a surface on the parameter space. In general, for a model with k parameters, the likelihood function $\mathcal{L}(w|y)$ takes the form of a k dimensional surface sitting above a k dimensional hyper plane spanned by the parameter vector $w = (w_1....w_k)$

2.3 Parameter Estimation

Having introduced the concepts of the probability density function and likelihood function, it is now possible to extend the discussion to the analytical method for estimating the parameters, w, of a model, given a set of sample data y, known as maximum likelihood estimation. Once data have been collected and the likelihood function of the model is computed, statistical inferences can be made about the population, and the probability distribution that underlies the data. Given that different parameter values lead to different PDFs, we are interested in finding the parameter values that corresponds to the population probability distribution.

The principle of maximum likelihood estimation (MLE) states that the desired probability distribution is the one that makes the observed data most likely. To put this within the context of the above example, this means that we must find the parameter vector that maximizes the likelihood function $\mathcal{L}(w|y)$. The resulting parameter vector, referred to as the MLE estimate, is denoted by $w_{MLE} = (w_{1,MLE}...w_{k,MLE})$. In figure 3 for example, the MLE estimate is $w_{MLE} = 0.7$, for which the value of the maximum likelihood function is $\mathcal{L}(w_{MLE} = 0.7|n = 10, y = 7) = 0.267$. The probability distribution corresponding to this MLE estimate is shown in figure 2. Applying the MLE principle implies that this is the population that is most likely to have generated the observed data, $y = 7$. To summarize, maximum likelihood estimation is an analytical method to find the model parameters which lead to the model producing a probability distribution that makes the observed data most likely.

2.4 Maximizing the Likelihood Function

We now show an example of how in practice the MLE estimate is computed. It is often convenient for the MLE estimate to be obtained by maximizing the log of the likelihood function, ln $\mathcal{L}(w|y)$. Because the two functions, ln $\mathcal{L}(w|y)$ and $\mathcal{L}(w|y)$ are monotonically related to each other, one would however obtain the same result by maximizing either one. In order to maximize the log likelihood function we must find the value of w which satisfies the following partial differential equation which is referred to as the likelihood equation;

$$\frac{\partial ln\mathcal{L}(w|n)}{\partial w_i} = 0 \tag{7}$$

at $w_i = w_{i,MLE}$ for all $i = 1, ..., k$. The existence of and MLE estimate requires that the above likelihood equation has a real solution. To ensure that $\mathcal{L}(w|y)$ is a maximum and not a minimum, an additional condition must be satisfied. To be a maximum, the log-likelihood function should be a peak and not a trough in the parameter space local to w_{MLE}. This condition can be ascertained from the second derivatives of $\mathcal{L}(w|y)$ which will be negative at $w_i = w_{i,MLE}$ for $i = 1, ..., k$;

$$\frac{\partial^2 ln\mathcal{L}(w|n)}{\partial w_i^2} < 0 \tag{8}$$

To illustrate the MLE procedure, consider the previous one-parameter binomial example given a fixed value of n. First, taking the logarithm of the likelihood function $\mathcal{L}(w|n = 10, y = 7)$, we obtain the log likelihood;

$$ln\mathcal{L}(w|n = 10, y = 7) = ln\frac{10!}{7!3!} + 7lnw + 3ln(1 - w) \tag{9}$$

We then calculate the first derivative of the log likelihood function;

$$\frac{dln\mathcal{L}(w|n = 10, y = 7)}{dw} = \frac{7}{w} - \frac{3}{1 - w} \tag{10}$$

The MLE estimate is obtained as $w_{MLE} = 0.7$ by requiring this equation to be zero. To make sure that the solution represents a maximum, not a minimum, the second derivative of the log likelihood function is evaluated at $w = w_{MLE}$;

$$\frac{d^2 ln\mathcal{L}(w|n)}{dw^2} = \frac{7}{w^2} - \frac{3}{(1 - w)^2} \tag{11}$$

which for $w = w_{MLE} = 0.7$ takes a negative value as desired.

For more sophisticated models involving multiple parameters such as many of those used in GEMS, it is usually not possible to obtain a precise analytic solution for the MLE estimate. For these models the MLE estimate is found using numerical techniques such as nonlinear optimization. These techniques allow for the optimal parameters to be found, that maximize the log-likelihood.

3 The Kalman Filter

The Kalman filter is a mathematical method that provides an efficient means of recursively estimating the parameters of a model based on a given data set. The method comprises a set of well defined mathematical equations that together have some desirable properties applicable to a wide range of modeling problems including the modeling of financial processes. In particular the model has the following useful properties:

- Computationally efficient method of describing the state of a system containing both information and noise

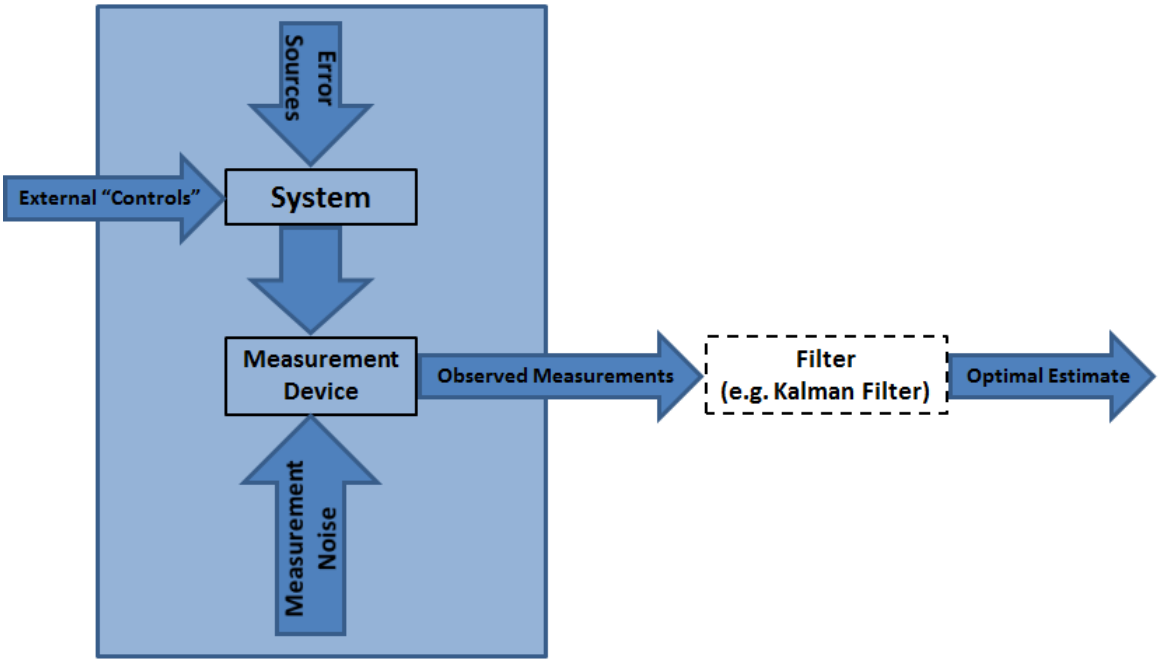

Figure 4. Schematic diagram of the data measurement and filtration problem.

- Produces an estimate that minimizes the mean square error of the estimator given all information available
- Enables the estimation of models based on non observable state variables (e.g. GEMS n-factor Affine model of the non defaultable term structure)

Kalman filter dynamics results from the consecutive cycles of predicting the state of an observed variable based on a model, comparing that prediction with the realized outcome in the historical or observed data and updating the parameters to achieve optimal predictive power. The update step is referred to as filtering, hence the method is referred to as a Kalman filter. The change to the filter at each iteration step, represents the novel information conveyed to it by the last observation. A complete description of the Kalman filter is beyond the scope of this document, but the following sections should give the reader enough information to adequately explain the functional use of the method as it pertains to the estimation of GEMS models. For further information interested readers are directed to the resources listed in the bibliography section of this document [4-12].

3.1 The Measurement and Filtration Problem

Figure 4, illustrates the problem that the Kalman Filter is designed to solve. A physical system, (e.g. a given financial market) is driven by a set of external drivers or controls (e.g. economic data, trading activity, market shocks etc.) and its outputs are measured by an observer (e.g. a data vendor), such that knowledge of the systems behavior is governed solely by the system controls and the observed or measured outputs. The

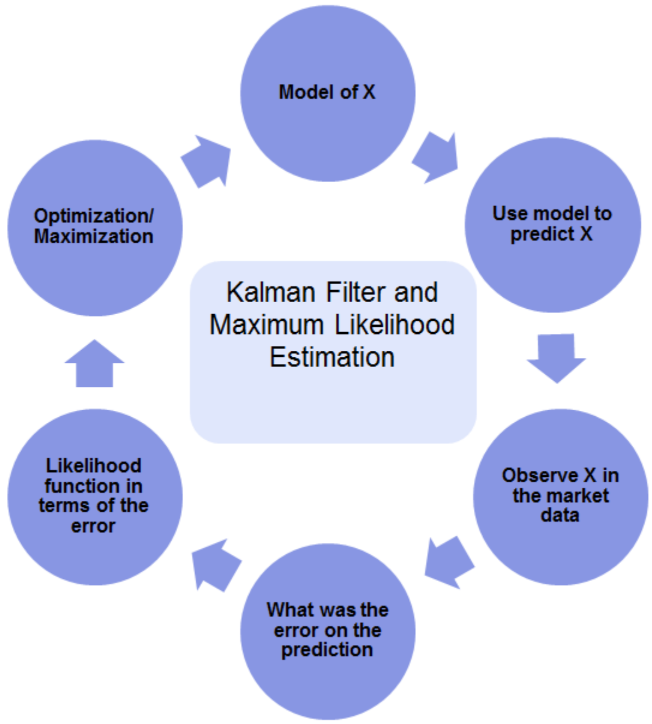

Figure 6. Schematic diagram of the parameter estimation methodology using Kalman Filtering and Maximum Likelihood Method.

observations and measurements contain both "true" information about the system, but also the errors and uncertainties in the process, namely measurement noise (e.g. from taking an average asset price across a number of brokers) and the system errors (e.g. in financial markets from liquidity effects).

The filtration problem can be summarized as finding the estimate of a system's state, based on the available information (controls and observations) which optimizes some defined criteria. Put another way, the problem is to find a filter which achieves the aforementioned. It is beyond the scope of this document but in many situations regarding the efficient estimation of financial market models the optimal filtration can be shown to be a Kalman Filter.

4 Kalman Filter and MLE for the Multi-Factor Square Root Process

The Kalman filter and MLE can be used in practice for the estimation of a multi-factor square root diffusion process such as that used in the non defaultable term structure model of Cox, Ingersoll, Ross (CIR) upon which the GEMS model is based. The CIR model takes the general form that the factors of the model evolve according to the following dynamics:

$$dy_i(t) = \kappa_i(\theta_i - y_i(t))dt + \sigma_i\sqrt{y_i(t)}dW_1(t) \tag{12}$$

A complete mathematical description of the process for estimating te parameters of this model are described in [12] and the interested reader is directed to this source for more details.

Briefly it is important to understand that since the factors themselves and the parameters (i.e. κ, θ and σ) are not directly observable in the market we must find a way of estimating the most appropriate parameters by effectively mapping the unobserved state variables onto the observed data, which in this case are the prices or equivalently the yields of zero coupon bonds. The Kalman filter, enables us to perform this mapping, and maximum likelihood allows us to minimize the estimation error of the parameters at each step in the Kalman filter cycle. Figure 6 shows the steps in the estimation cycle which are detailed more precisely in [12].

5 Target Setting and Final Parameter Setting

In the estimation of GEMS models the final parameter set, β, is further constrained to be close to a set of defined target values for the mean and standard deviation of the simulated variables. These targets are set from analysis of the available data and ensure that the models produce simulated values which are consistent across asset classes and not biased, for instance by short data histories. These targets are met by constraining the MLE estimation to combinations of parameters that have expectation values close to the target values. In practice this is implemented by including a penalty function in the final MLE functions such that we maximize the function;

$$\mathcal{L}(\beta)' = \mathcal{L}(\beta) - (\sum^{k} P^k_{mean} + P^k_{variance}) \tag{13}$$

Where P^k_{mean} and $P^k_{variance}$ are the values of the penalty function for the k^{th} target and the penalty functions are given by the square of the difference between the model value for the parameter set β and the target value;

$P^k_{mean} = (ModelMean(\beta) - TargetMean)^2$

$P^k_{variance} = (ModelVariance(\beta) - TargetVariance)^2$

For most models used in GEMS the expectation values of mean and variance given a particular set of parameter values, β, have explicit formula which aids efficient estimation.

6 Model Re-Estimation

As a general principal we do not wish to re-estimate the parameters of models every week, month, quarter or even year. Because the targets used in GEMS are at a long term horizon we should expect that they are reasonably stable over short periods of time. In the model building and maintenance process a balance must be struck between parameter stability and keeping the models current. When using 25-30 years of history even several years of new data may lead to only small changes in the model parameters and these may have little material impact on the final simulated distributions. Nevertheless the model distributions in GEMS are reassessed if there is a major market event, and are assessed anyway every two years. However the principal that is generally applied is that a models parameters are only changed if it is apparent that significantly better results could be obtained by reestimating the model with more current data. In order to ensure parameters are stable, model parameters are generally only updated when one of the following occurs;

- The model no longer matches the target values within tolerance.
- A specific problem is identified by which the model is no longer producing results consistent with what is observed in the market.
- A model higher up in the cascade structure has been re-estimated requiring the downstream models to be re-estimated in order to keep simulated values close to targets values.
- Significant new data in the market suggests a structural change or change in dynamics of the market.

When a parameter change is made it is communicated in the quarterly release note for the economy effected. Tools are also available to aid users in reparameterising the model to match a particular view of the market not already incorporated within GEMS.

It is however important that the initial conditions of models (e.g. the initial yield curves for government bonds) are updated prior to model use, and this task is performed on a quarterly basis by Conning and tools are available such that users can update the initial conditions of the models at higher frequency.

7 Conclusion

The maximum likelihood method and Kalman filtering are powerful techniques when applied the estimation of a system's dynamics, which can be combined to estimate the parameters of a model of the system. In this document we have introduced both techniques and some general principles of the filtering problem. The Kalman filter, under certain conditions, represents the optimal filter for extracting novel information from a sequence of observations (e.g. a time series of financial data). The final solution for the model parameters is determined using optimization, however in practice one must exercise great care to ensure any theoretical boundary conditions are not violated by the optimizer for a given model. The quality of the GEMS parameterisation over other parameterisations originates primarily from the precise details of the implementation of the methods described here, and from extensions such as the ability to simultaneously fit yield and bond return dynamics. The result of the parameterisation process in GEMS is a model that captures a wide range of the features and dynamics of the real data, and which performs well out of sample.

8 References

[1] Bickel, P. J. and Doksum, K. A. (1997) *Mathematical Statistics*, Oakland, CA: Holden-day Inc.
[2] DeGroot, M. H. and Schervish, M. J. (2002), *Probability and Statistics (3rd edition).* Boston, M A: Adison-Wesley
[3] Spanos, A (1999), *Probability theory and statistical inference.* Cambridge, UK: Cambridge University Press
[4] Gelb, A. (1974). *Applied Optimal Estimation.* MIT Press.
[5] Kalman, R.E. (1960). *A new approach to linear filtering and prediction problems.* Journal of Basic Engineering 82 (1): 3545.
[6] Kalman, R.E.; Bucy, R.S. (1961). *New Results in Linear Filtering and Prediction Theory.*
[7] Harvey, A.C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.
[8] Simon, D. (2006). *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches.* Wiley-Interscience.
[9] Liu, W.; Principe, J.C. and Haykin, S. (2010). *Kernel Adaptive Filtering: A Comprehensive Introduction.* John Wiley.
[10] Jazwinski, Andrew H. (1970). *Stochastic Processes and Filtering.* Mathematics in Science and Engineering. New York: Academic Press. pp. 376. ISBN 0123815509.
[11] Maybeck, Peter S. (1979). *Stochastic Models, Estimation, and Control.* Mathematics in Science and Engineering. 141-1. New York: Academic Press. pp. 423. ISBN 0124807011.
[12] Bolder, David Jamieson, (2001). *Affine Term-Structure Models: Theory and Implementation*

Disclosures/Confidentiality Notice

Conning (www.conning.com) is a leading investment management firm with a long history of serving the insurance industry. Conning supports institutional investors, including pension plans, with investment solutions and asset management offerings, award-winning risk modeling software, and industry research. Founded in 1912, Conning has investment centers in Asia, Europe and North America.