Section 5: Descriptions of Data and Approach

In this section, detailed descriptions are provided of our analytical methodology associated with each of the economic time series[1]. Embedded in these descriptions are references to the sources of historical time series data; where relevant, we have provided a hyperlink to these sources.

Inflation model

Inflation (denoted by q) is assumed to follow an Ornstein-Uhlenbeck process of the form (in continuous time):

$$dq_t = \kappa \ (\mu_q - q_t)dt + \sigma dB_q \quad (1)$$

The simulation model samples the discrete form equivalent of this process as:

$$\begin{aligned} \Delta q_t &= q_{t+1} - q_t = \kappa_q(\mu_q - q_t)\Delta t + \varepsilon_q \sigma_q \sqrt{\Delta t} \\ q_{t+1} &= q_t + \kappa_q(\mu_q - q_t)\Delta t + \varepsilon_q \sigma_q \sqrt{\Delta t} \\ &= \kappa_q \Delta t \cdot \mu_q + (1-\kappa_q \Delta t)\cdot q_t + \varepsilon_q \sigma_q \sqrt{\Delta t} \end{aligned}$$

From this last equation, we can see that the expected level of future inflation is a weighted average between the most recent value of inflation (q_t) and a mean reversion level of inflation, μ_q. The speed of reversion is determined by the parameter κ_q. In the continuous model, mean reversion can be seen by considering the first term on the right-hand side of (1) (which is called the drift of the process). If the current level of inflation (q_t) is above the average, the first term is negative. Therefore, equation (1) predicts that the expected change in inflation will be negative; that is, inflation is expected to fall. The second term on the right-hand side of (1) represents the uncertainty in the process. One can think of the Brownian motion term (B_t) as representing a draw from a standardized normal random variable (represented by ε_q in the discrete form of the model). Uncertainty also includes the parameter σ_q, which scales the magnitude of the volatility associated with the inflation process.

We can rearrange the last equation above to show that this process is an autoregressive process.

$$\begin{aligned} q_{t+1} - \mu_q &= \mu_q \kappa_q \Delta t - \mu_q + (1-\kappa_q \Delta t)\cdot q_t + \varepsilon_q \sigma_q \sqrt{\Delta t} \\ &= (1-\kappa_q \Delta t)\cdot q_t - (1-\kappa_q \Delta t)\cdot \mu_t + \varepsilon_q \sigma_q \sqrt{\Delta t} \\ &= \phi_1(q_t - \mu_q) + \varepsilon_q \sigma_q \sqrt{\Delta t} \end{aligned}$$

In order to estimate the parameters of the inflation model, we run the following regression:

$$q_{t+1} = \alpha + \beta q_t + \varepsilon'_{qt} \quad (2)$$

Note that we have not run the regression using the change in inflation as the dependent variable since this would not allow us to simultaneously derive the mean reversion speed (κ_q) and the mean reversion level (μ_q). To derive the parameters of the inflation process, we transform the regression coefficients in (2):

$$\beta = (1 - \kappa_q \Delta t)$$
$$\kappa_q = \frac{1-\beta}{\Delta t} \qquad (3)$$

and

$$\alpha = \kappa_q \mu_q \Delta t = \frac{1-\beta}{\Delta t} \mu_q \Delta t$$
$$\mu_q = \frac{\alpha}{1-\beta}$$

We gathered inflation data from the Bureau of Labor statistics website (www.bls.gov) and ran several regressions of this type to estimate κ_q and μ_q. We considered the frequency of the observations when performing regressions analysis. One concern was that individual monthly CPI levels might contain errors that would bias the regression coefficients. For example, if the CPI level of May 2003 was overstated, then inflation in May would appear "high" while the subsequent inference of inflation would appear "low". If the time series of CPI contained any errors of this type, the mean reversion strength may become overstated.

Given the noisy fluctuations in monthly data, we selected the parameters for the inflation process by looking at annual regressions. By calculating the change in CPI over the course of a year, we have a more stable and accurate depiction of the inflation rate. The popularly reported time series of CPI uses a base period (i.e., resets the index value at 100) between the years 1982 and 1984. Given the fact that the CPI level is only reported to the first decimal place, using the current base does not lend itself to capturing minor changes in inflation in the first half of the 20th-century; a small change in CPI may lead to large swings in inflation when the level of the index is low. The only publicly available series reported on the old base level (1967 = 100) is: Not seasonally adjusted, alternate base (1967), U.S. city averages, all items.

The annual rate of inflation was measured as:

$$q_t = \ln \frac{CPI_t}{CPI_{t-1}}$$

Where CPI_t is the reported index value for year t and CPI_{t-1} is the prior year's reported index value of the same month. We ran two annual regressions: (1) all available data and (2) the years after World War II.

Time Period	κ_q	μ_q	σ_q
1913-2001	0.37	3.3%	4.0%
1946-2001	0.47	4.8%	3.0%

For use in our projections, we selected κ_q to be 0.4 and the mean reversion level to be 4.8% to capture the post war economic period. Although it might appear that the speed of mean reversion over the second half of the 20th-century has increased, it should be noted that the standard error of the estimate of κ_q is higher than over the larger time period.

Instead of being concerned with the annualized, instantaneous level of inflation, bond investors are more concerned with the expected level of inflation over the life of their investment. Given the existing level of inflation (q_t) and the parameters of the assumed process in (1), we can derive expectations of future inflation over various horizons. Vasicek (1977) uses an Ornstein-Uhlenbeck process assumes to model interest rates and provides a closed-form solution for long-term bond yields as a function of the current interest-rate and the model parameters. According to Vasicek (1977), the time *t* price of a bond, $P(t,T)$ that matures at time *T* is:

$$P(t,T) = A(t,T)e^{-rB(t,T)} \tag{4}$$

where $A(t,T)$ and $B(t,T)$ are functions of the parameters of the assumed process for interest rate movements.

Our process for inflation follows the same Ornstein-Uhlenbeck process, so we can develop a “term structure” of inflation analogous to Vasicek (1977). This term structure posits an expected, inflation rate over various horizons. A term structure of inflation is needed to generate nominal interest rates, since investors are concerned not only about the time value of money, but also about the expected level of inflation over the life of bonds.

Real Interest rates

A significant amount of work has been done in the area of interest rate modeling. The role of the financial scenario generator is not to explain past movements in interest rates, but rather to develop a model that posits plausible projections of future interest-rate levels. (It might also be noted that trying to develop a model that mimics past movements may be a futile exercise since, despite the volume of research in the area, no tractable model has yet been shown to be satisfactory in accurately explaining historical interest-rate movements.)

In selecting a term structure model, we were concerned about two specific issues. First, we considered the number of stochastic factors to incorporate that generates term structure movements. Choosing the number of stochastic factors for a term structure model represents a balance between (1) having a large number of factors to adequately emulate empirical rate movements and (2) limiting the number of factors so the resulting model is simple enough to be tractable. With one-factor term structure models, the dynamics of the entire yield curve are completely driven by the single source of uncertainty. Resulting yield curve movements are subsequently constrained: yields of all maturities are perfectly correlated to the one stochastic factor and the range of potential yield curves is limited. Introducing additional sources of uncertainty (such as allowing the long end of the curve to fluctuate and/or introducing stochastic volatility) provide for a fuller variety of yield curve shapes. The downside is that introducing multiple dimensions of yield curve movements increases the complexity of the model quickly.

Our second concern was in choosing a term structure model that has closed form solutions for bond prices so that the entire term structure can be quickly and easily retrieved from the existing the values of the stochastic factors. When closed form solutions for bond yields are available, this allows users of the term structure model to track various points on the yield curve during a simulation. For example, users of a term structure model who are interested mortgage prepayment rates will be interested in the refinancing rate, which may be closely related to bond yields of specific maturities (such as 10 years). Without some explicit closed form solution, the modeler has no foundation to imply yields of different maturities from a limited set of stochastic factors.

To derive real interest rates, we selected the two-factor Vasicek term structure model, which is a simple case of the two-factor Hull-White model. In the two-factor Vasicek model, the short-term rate (denoted by *r*) reverts to a long-term rate (denoted by *l*) that is itself stochastic.

$$\begin{aligned} dr_t &= \kappa_1 (l_t - r_t) dt + \sigma_1 dB_1 \\ dl_t &= \kappa_2 (\mu - l_t) dt + \sigma_2 dB_2 \end{aligned}$$

In order to estimate the parameters of the model, we look at the discrete analog of the model:

$$\begin{aligned} \Delta r_t &= \kappa_1 (l_t - r_t) \Delta t + \sigma_1 \varepsilon_{1t} \\ \Delta l_t &= \kappa_2 (\mu - l_t) \Delta t + \sigma_2 \varepsilon_{2t} \end{aligned}$$

$$\begin{aligned} r_{t+1} - r_t &= \kappa_1 (l_t - r_t) \Delta t + \sigma_1 \varepsilon_{1t} = (\kappa_1 l_t + \kappa_1 r_t) \Delta t + \sigma_1 \varepsilon_{1t} \\ l_{t+1} - l_t &= \kappa_2 (\mu - l_t) \Delta t + \sigma_2 \varepsilon_{2t} = (\kappa_2 \mu - \kappa_2 l_t) \Delta t + \sigma_2 \varepsilon_{2t} \end{aligned}$$

We can also see how each of the stochastic factors is updated over time, analogous to the situation for inflation presented above. First, we rearrange the discrete form of the two-factor Vasicek term structure model:

$$\begin{aligned} r_{t+1} &= r_t + (\kappa_1 l_t + \kappa_1 r_t) \Delta t + \sigma_1 \varepsilon_{1t} \\ &= \kappa_1 \Delta t \cdot l_t + (1 - \kappa_1 \Delta t) \cdot r_t + \sigma_1 \varepsilon_{1t} \\ l_{t+1} &= l_t + (\kappa_2 \mu - \kappa_2 l_t) \Delta t + \sigma_2 \varepsilon_{2t} \\ &= \kappa_2 \Delta t \cdot \mu_l + (1 - \kappa_2 \Delta t) \cdot l_t + \sigma_2 \varepsilon_{2t} \end{aligned} \qquad (5)$$

From these equations, we can see that the short rate is again a weighted average between the current level r_t and the mean reversion factor l_t. The mean reversion factor is itself a weighted average of some long-term mean and its current value.

Hibbert, Mowbray, and Turnbull (2001) (HMT) also use this process for real interest rates. They derive closed form solutions for bond prices (and therefore yields), which are considerable more complicated than the one-factor Ornstein-Uhlenbeck process for inflation:

$$P^r(t,T) = A^r(t,T)e^{-r_t B_1(t,T) - l_t B_2(t,T)} \tag{6}$$

where r_t and l_t are the values for the short and long real interest rate and A^r, B_1, and B_2 are functions of underlying parameters in the two-factor Vasicek specification.

Estimating the equations in (5) is a difficult procedure since real interest rates are not directly observable in the market. We compute *ex post* real interest rates based on the difference between nominal rates observed in the market less the monthly (annualized) inflation rate. We use the three-month Constant Maturity Treasury (CMT) as a proxy for the instantaneous short rate and the 10-year CMT yield as a proxy for the long rate. (We also looked at longer Treasury yields as a proxy for the long rate. Results were not sensitive to the choice of maturity.) Nominal interest rates are from the Federal Reserve's historical database (http://www.federalreserve.gov/releases/).

There are several issues with the Federal Reserve's interest rate data. First, at the long end of the yield curve, there are significant gaps in many of the series. For example, the 20-year CMT was discontinued in 1987; yields in 20-year securities after this date would have to be interpolated from other yields. Also, given the decision of the Treasury to stop issuing 30-year bonds, the future of 30-year rate data is uncertain (in fact, the Fed's data stops reporting 30-year CMT data in the early 2002). At the short end of the yield curve, it was difficult to determine a proxy for the short rate. Ideally, one would want an interest-rate that most closely resembles an instantaneous rate. While the one-month CMT is reported back only to 2001, the 3-month rate is available beginning in 1982. While we could have reverted to a private, proprietary source of data to create a longer time series, we restricted ourselves to only publicly available data sources that would be available to any user of the model.

We use the following regressions on monthly data from 1982 to 2001:

$$\begin{aligned} r_{t+1} &= \alpha_1 + \alpha_2 l_t + \alpha_3 r_t + \varepsilon'_{rt} \\ l_{t+1} &= \beta_1 + \beta_2 l_t + \varepsilon'_{lt} \end{aligned}$$

Traditional OLS regressions are not possible since the short rate process is dependent upon the long rate; these are simultaneous equations. Instead, we use two-stage least squares estimation. In order to estimate the short-rate equation, we first obtain estimates for the long-rate $\hat{l}_t$.

$$\begin{aligned} &\text{Stage 1: } l_{t+1} = \beta_1 + \beta_2 l_t + \varepsilon'_{2t} \\ &\text{Stage 2: } \Delta r_{t+1} = \alpha_1(\hat{l}_t - r_t) + \varepsilon'_{1t} \end{aligned}$$

The resulting parameters were selected from the regression results.

Real Interest Rate Process
Estimated from 1982 - 2001

κ_r	μ_r	σ_r	κ_l	σ_l
6.1	2.8%	10.0%	5.1	10.0%

These parameters indicate a high level of volatility that is tempered by strong levels of mean reversion.

Nominal interest rates

Fisher (1930) provides a thorough presentation of the interaction of real interest rates and inflation and their effects on nominal interest rates. He argues that nominal interest rates compensate investors not only for the time value of money, but also for the erosion of purchasing power that results from inflation. In the model presented here, the underlying movements in inflation and real interest rates generate the process for nominal interest rates. If bonds are priced using expectations of inflation and real interest rates until the bond's maturity, then nominal interest rates are implied by combining the term structure of inflation (eq. (4)) and the term structure of real interest rates (eq. 6). Therefore:

$$P^i(t,T) = P^r(t,T) \times P^q(t,T)$$

where *i* refers to nominal interest rates and the superscripts on the bond prices correspond to the underlying stochastic variables.

Unfortunately, the parameters for the real interest rate process shown above generate a distribution that severely restricts the range of potential future nominal interest rates. For example, the 1st percentile of the distribution for the 20-year nominal rate is 5.9% and the 99th percentile is 8.2%. There are several candidates for problems with real interest rates that may lead to this seemingly unrealistic distribution of future nominal rates: (1) the use of *ex post* real interest rate measures is unsuitable, (2) monthly measurement of real interest rates contains mean reverting errors which exaggerate mean reversion speed, or (3) the time period used to measure real interest rates is too short.

As a result, the parameters for real interest rates were altered to allow nominal interest rates to more accurately reflect historical volatility. Specifically, mean reversion speed was dramatically reduced. Given that mean reversion speed and volatility work together to affect the range of interest rate projections, volatility was also reduced. The following parameters are used as the "base case" in the model. These parameters are in line with what was used in Hull (2001).

κ_r	μ_r	σ_r	κ_l	σ_l
1.0	2.8%	1.00%	0.1	1.65%

Equity Model – Regime Switching

In order to capture the fat tails of the equity return distribution that have been observed historically, we use a regime-switching model to capture equity returns. To motivate the rationale for the model, consider October 1987. This single observation may appear to be too "extreme" and very unlikely given a single variance assumption. Instead, suppose that equity returns at any point in time are generated from two distinct distributions, a "high volatility" regime or a "low volatility" regime. The chance of switching from one regime to the other over the next time step is dictated by transition probabilities. During times of economic instability, the returns on equities may be more uncertain, representing a transition to the high volatility regime. Thus, the observation from October 1987 may simply be a draw from the high volatility regime.

In the financial scenario model, equity returns are based on the risk-free nominal interest rate ($q + r$) and a risk premium or the excess equity return attributable to capital appreciation (x):

$$s_t = q_t + r_t + x_t$$

To estimate the parameters of the regime switching equity return model, we used the approach of Hardy (2001). In her model, Hardy assumes that stock prices are lognormally distributed under each regime. But while Hardy looks at total equity returns, including dividends and the underlying compensation from the risk free rate, we use the excess equity returns x. Following the procedure outlined in Hardy (2001), we then maximize the likelihood function implied from the regime switching process.

We allow the returns of small stocks and large stocks to be generated independently and estimate each equity process separately. Numerous web sites are available to capture the time series of capital appreciation of these indices (see for example, finance.yahoo.com). We used the longest time series available for large stocks (1871-2002), available at Robert Shiller's web site (http://www.econ.yale.edu/~shiller/data/ie_data.htm). To look at small stocks, we used Ibbotson data captured from 1926-1999. As expected, the risk and return of small stocks appears higher than large stocks under both regimes. Using the procedure of Hardy (2001), we developed the following parameter estimates:

Excess Monthly Returns

	Large Stocks (1871-2002)		***Small Stocks (1926-1999)***	
	Low Volatility Regime	***High Volatility Regime***	***Low Volatility Regime***	***High Volatility Regime***
Mean	0.8%	-1.1%	1.0%	0.3%
Variance	3.9%	11.3%	5.2%	16.6%
Probability of Switching	1.1%	5.9%	2.4%	10.0%

Note that while the expected return in the high volatility regime is lower, it is more likely that if the high volatility regime is ever reached, the equity market will revert back to the low volatility regime since the probability of switching is higher.

Equity Dividend Yields

Similar to the process used by HMT and Wilkie (1986), we assume that (the log of) the dividend yield follows an autoregressive process.

$$d(\ln y_t) = \kappa_y(\mu_y - \ln y_t)dt + \sigma_y dB_{yt}$$

Estimation of this process is analogous to the inflation process described above. One source of difficulty for estimating the dividend yield process is in obtaining data. There is no long time series of dividend yields and that is publicly available for equity indices. To obtain this information, we used a proprietary source of financial data (Telerate). However, one may be able to estimate the dividend yield of indices that contained a limited number of stocks (such as the Dow Jones industrial average).

Real Estate (Property)

Given the that the real estate portfolios of insurers are dominated by commercial properties, we use the National Council of Real Estate Investment Fiduciaries (NCREIF) pricing index to capture the quarterly returns on commercial properties (see www.ncreif.com). The NCREIF data is generated from market appraisals of various property types including apartment, industrial, office, and retail. While the use of appraisal data may only approximate sharp fluctuations in market valuation, publicly obtainable transaction-based real estate data was not available.

Using *quarterly* return data from NCREIF from 1978 to 2001, we estimated two separate Ornstein-Uhlenbeck models for real estate: the first model included inflation while the second model did not. While we expected inflation to be a driver of real estate returns, the results were not significant.

Unemployment

There are many plausible ways to link unemployment rates to other economic variables. One approach to estimating unemployment is based on the well-known Phillips curve. The Phillips curve illustrates a common inverse relationship between unemployment and inflation. (It should be noted that since the original publication of the Phillips curve in 1958, recent data has not fit the inverse relationship as well.) The approach taken by Phillips seems plausible: As the economy picks up, inflation increases to help temper the demand driven economy. At the same time, unemployment falls as firms hire to meet the increasing demand. When the economy slows down, unemployment rises, and inflationary pressures subside.

We include a first-order autoregressive process in the Phillips curve:

$$du_t = \kappa_u(\mu_u - u_t)dt + \alpha_u dq_t + \sigma_u \varepsilon_{ut}$$

It is expected that when inflation increases (dq_t >0), unemployment decreases (the coefficient on inflation changes (α_u) is negative). One may argue that there is a lag between inflation and unemployment. To keep the model simple, we did not pursue any distributed lag approach.

The discrete form of the unemployment model:

$$\begin{aligned} u_{t+1} &= u_t + \kappa_u \mu_u - \kappa_u \Delta t \cdot u_t + \alpha_u (q_{t+1} - q_t) + \sigma_u \varepsilon_{ut} \\ &= \kappa_u \mu_u + (1 - \kappa_u \Delta t) \cdot u_t + \alpha_u (q_{t+1} - q_t) + \sigma_u \varepsilon_{ut} \end{aligned}$$

Which suggests the following regression:

$$u_{t+1} = \beta_1 + \beta_2 u_t + \beta_3 (q_{t+1} - q_t) + \sigma_u \varepsilon_{ut}$$

We use inflation data as described above and retrieve monthly unemployment data from the Bureau of Labor Statistics (www.bls.gov). Using data from 1948 to 2001 and transforming the regression coefficients as in (3), we get:

$$du_t = 0.13 \times (6.1\% - u_t)dt - 0.72 dq_t + 0.76\% \times dB_{ut}$$

---

[1] In the Financial Scenario Model software, there exist options for the user to override the stochastically simulated variables in certain circumstances. Specifically:

- The user may limit the following variables to non-negative values only: nominal interest rates, real interest rates, and inflation.
- The user may input selected specific scenarios for the following variables: nominal interest rates, inflation, and equity returns.

Additional discussion of these issues is provided in *Section 6 – Discussion of Issues* and in *Appendix A – User's Guide* of this report.