Estimation of dynamic term structure models

Greg Duffee
Haas School of Business, UC-Berkeley

Joint with Richard Stanton, Haas School

Presentation at IMA Workshop, May 2004
(full paper at http://faculty.haas.berkeley.edu/duffee)

Overview

- Dynamic term structure models

  Specify stochastic evolution of instantaneous interest rate r_t and the compensation investors require to face interest-rate risk.

  Result is a complete dynamic model of the term structure of yields on default-free bonds

- The big question

  How do standard estimation methods behave in finite samples when applied to newer classes of dynamic term structure models?

- The approach

  We use Monte Carlo simulations to answer this question, and uncover some surprising and discouraging results.

Outline

1. Overview of first-generation and second-generation dynamic term structure models
2. Discussion of performance of maximum likelihood estimation
3. Alternatives to ML estimation

First generation of term structure models

One branch: CIR

$$r_t = \delta_0 + x_t$$

equiv m. measure $\quad dx_t = (k\theta - kx_t)dt + \sigma\sqrt{x_t}dz_t$

physical measure $\quad dx_t = (k\theta - (k - \lambda_2)x_t)dt + \sigma\sqrt{x_t}d\tilde{z}_t$

Risk premia are determined by λ_2

- Bond pricing is tractable

$$P_{t,\tau} = E_t^q\left[e^{-\int_t^{t+\tau} r_s ds}\right]$$

- Physical transition density $p(r_{t+s}|r_t)$ is known for $s > 0$.
- Drift under physical, equivalent martingale measures share at least one parameter

First generation of term structure models

The other branch: Vasicek

$$r_t = \delta_0 + x_t$$

$$\text{equiv m. measure} \quad dx_t = (k\theta - kx_t)dt + \sigma dz_t$$

$$\text{physical measure} \quad dx_t = (k\theta + \lambda_1 - kx_t)dt + \sigma d\tilde{z}_t$$

Risk premia determined by λ_1

- Bond pricing is tractable, transition density of r_t is known, drifts share at least one parameter
- For both CIR and Vasicek, generalization to multiple independent $x_{i,t}$'s is simple

Estimation of first-generation models

- Observe yields on bonds with different maturities at dates $t, t+1, \ldots$
- Maximum likelihood is standard technique
- One way to implement
  - Assume as many yields as state variables are observed without error
  - Given parameter vector, can invert to determine states $x_{i,t}$
  - Transition density of yields from t to $t+1$ can then be calculated (Jacobian transformation of transition density of states)
  - Other bond yields observed with normally-distributed error
  - Choose parameter vector to maximize likelihood function
- Existing evidence is that ML estimation works well in finite samples similar in length to real-world data sets

Second-generation models

- Big problem with first-generation models—they do not work

  The dynamics cannot capture real-life variations in expected excess returns to long-maturity bonds

  Forecasts of future bond yields are inferior to random-walk forecasts

- Second-generation models relax key restrictions in first-generation models

  - More flexible specification of bond risk premia; breaks link between physical, equivalent martingale drifts
  - Nonlinear drifts
  - Correlated factors
  - Many of these models do not have known transition densities for discretely-observed bond yields

The first question

- For realistic sample sizes and term structure behavior, how well does ML perform when risk premia specification breaks link between physical, equivalent martingale measures?

  When transition density of discretely-observed data is unknown/intractable, we use simulated ML (simulated transition density)

The second question

- How closely do tractable estimation methods approximate ML?

1. Efficient Method of Moments

   Gallant and Tauchen; auxiliary model is SemiNonParametric (SNP).

2. Linearized extended Kalman filter

Our approach

- We answer these questions in very simple 2nd-generation settings

  Settings are simple enough for ML or simulated ML to be feasible, allows for comparison with alternative techniques

- Discussion today is even simpler – focus almost exclusively on one-factor models with Gaussian dynamics

A key feature of the term structure: persistence

- "True" parameters of physical dynamics of short rate, based on 1970-2000 data

$$dr = 0.065(0.0523 - r_t)dt + 0.0175 dz_t$$

- Half-life of shocks is 11 years
- Monte Carlo simulation of ML estimation of short rate only (ignore info in rest of term structure)
  - 1000 weekly observations (19 years)
  - Mean estimate of k is 0.304, standard dev is 0.239, mean standard error is 0.167

    Implied half-life of shocks 2 1/4 years

- 1st generation models: Estimation of term structure model attenuates finite-sample bias of speed of mean reversion

  - "True" model

    equiv m. measure $dr_t = (0.0085 - 0.065 x_t) dt + 0.0175 dz_t$

    physical measure $dx_t = ((0.0084 - 0.0050) - 0.065 x_t) dt + 0.0175 d\tilde{z}_t$

  - Monte Carlo results (ML estimation, 1000 weeks of data)

    Estimates of all parameters are now unbiased (within Monte Carlo sampling error)

  - Intuition – investors know true model, they price bonds using it

- The 2nd-generation version of the Gaussian one-factor model

$$\text{physical measure} \quad dr = (k\theta - kr_t)dt + \sigma d\tilde{z}_t$$

$$\text{equiv m. measure} \quad dr = (k\theta + \lambda_1 - (k - \lambda_2)r_t)dt + \sigma dz_t.$$

- λ_1 affects average risk premia on bonds
- λ_2 determines how risk premia vary with the level of the term structure
- "True" parameters

  $k\theta = 0.0084, k = 0.065, \sigma = 0.0175, \lambda_1 = 0.005, \lambda_2 = -0.14$
- Physical persistence parameter is $0.065 + 0.14 = 0.205$, half-life of shocks is 3.4 years
- Monte Carlo results

  ML finite-sample estimates of k, $k\theta$ unbiased, but physical speed of mean reversion strongly biased (0.439), bias shows up in price of risk parameter λ_2

Intuition for poor finite-sample performance of ML

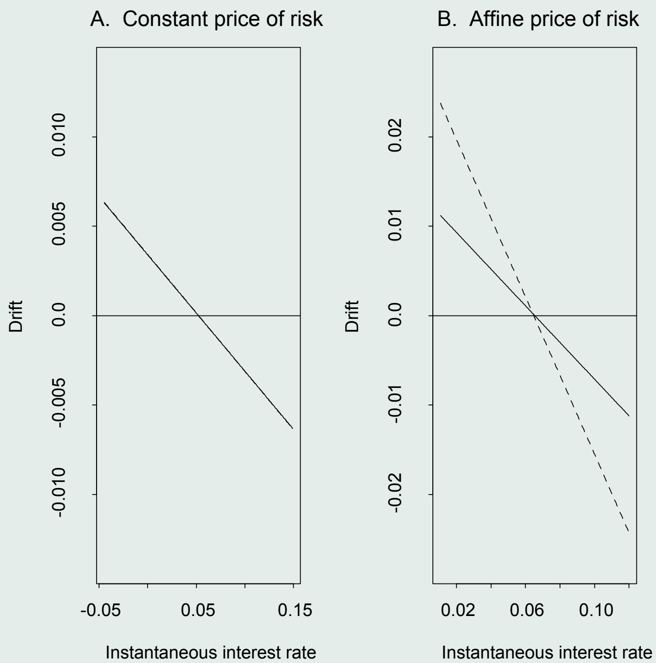
- Drifts of physical, equiv m. measures decoupled with this model

  Bonds are priced as if long-run mean of r_t, speed of mean reversion of r_t are $k\theta/k$, k. Compare to physical values of $(k\theta+\lambda_1)/k$, $k-\lambda_2$.

- Therefore only info about physical drift is from time-series drift of r_t, which is strongly biased

- Here, all the bias shows up in price of risk parameter

- 1st and 2nd generation drifts: true (solid) and mean estimates (dashed)



- Same point carries over to 2nd-generation square root diffusion model
- “True” model

$$r_t = 0.01 + x_t$$

$$\text{equiv m. measure} \quad dx_t = (0.0075 - 0.063x_t)dt + 0.08\sqrt{x_t}dz_t$$

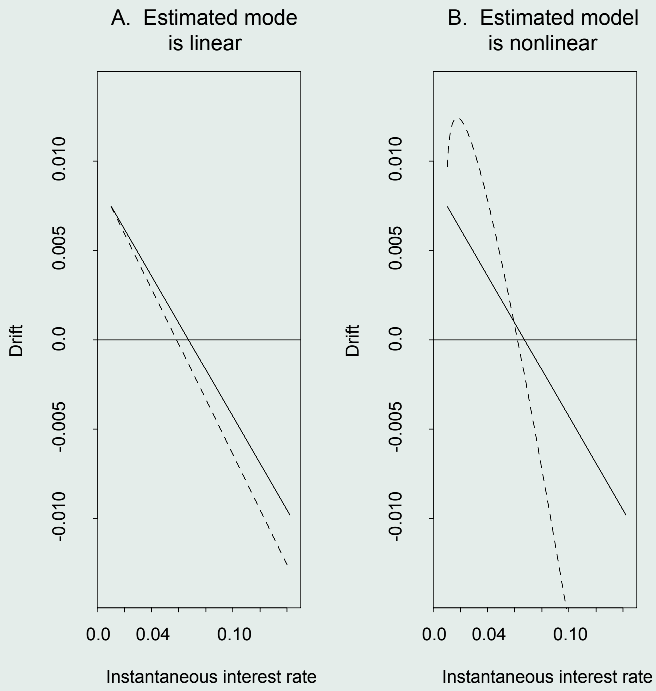
$$\text{physical measure} \quad dx_t = (0.0075 - (0.063 - (-0.068))x_t)dt + 0.08\sqrt{x_t}d\tilde{z}_t$$

- Estimated model allows for nonlinear physical dynamics with more general risk premium specification

$$\text{physical measure} \quad dx_t = (k\theta + \lambda_1\sqrt{x_t} - (k - \lambda_2)x_t)dt + \sigma\sqrt{x_t}d\tilde{z}_t$$

- Drifts implied by parameters estimated with ML from Monte Carlo simulation (next slide)

- 1st and 2nd generation drifts: true (solid) and mean estimates (dashed)



- Conclusion: With 2nd-generation models (allowing for general specification of dynamics of risk premia), ML produces strongly biased estimates of risk premia

  Therefore models produce bad estimates of expected excess returns to bonds

- Bias is qualitatively equivalent to bias in speed of mean reversion of near-unit-root processes

Question 2: Tractable alternatives to ML

- Commonly-used technique in term structure modeling is Efficient Method of Moments
- Our conclusion is that it performs very poorly

  With highly persistent processes, EMM breaks down
- Overview of EMM is next, followed by some results

Efficient Method of Moments

- Path simulation estimation technique

  Useful in settings where continuous dynamics of data are known, but not discrete dynamics

- Denote history of observed yields through t as vector Y_t.
- True density function is denoted $g_{Y_t}(Y_t;\rho_0)$; may be unknown or intractable
- $f(y_t|Y_{t-1};\gamma_0)$ is auxiliary function that approximately expresses log density of y_t as a function of Y_{t-1} and auxiliary parameter vector γ_0
- First step in EMM: maximize auxiliary log-likelihood function

$$\frac{1}{T}\sum_{t=1}^{T}\left[\left.\frac{\partial f(y_t|Y_{t-1};\gamma)}{\partial\gamma}\right|_{\gamma=\tilde{\gamma}_T}\right]=0.$$

- Central Limit Theorem

$$\sqrt{T}(\tilde{\gamma}_T - \gamma_0) \xrightarrow{d} N\left(0, d^{-1}Sd^{-1}\right),$$

$$S = E\left[\left(\frac{\partial f}{\partial \gamma}\right)\left(\frac{\partial f}{\partial \gamma'}\right)\bigg|_{\gamma=\gamma_0}\right],$$

$$d = E\left(\frac{\partial f}{\partial \gamma \partial \gamma'}\bigg|_{\gamma=\gamma_0}\right).$$

- Second step in EMM: Simulate long time series $\hat{Y}_N(\rho) = (\hat{y}_1(\rho)', \ldots, \hat{y}_N(\rho)')'$ using true dynamic term structure model with params ρ

- Calculate expectation of score vector of auxiliary model, evaluated at ρ

$$m_T(\rho, \tilde{\gamma}_T) = \frac{1}{N}\sum_{\tau=1}^{N}\frac{\partial}{\partial \gamma} f[\hat{y}_\tau(\rho) \mid \hat{Y}_{\tau-1}(\rho); \tilde{\gamma}_T].$$

$$\lim_{N\to\infty} m_T(\rho, \tilde{\gamma}_T) = E\left(\frac{\partial f(y_t(\rho)|Y_{t-1}(\rho);\gamma)}{\partial \gamma}\bigg|_{\gamma=\tilde{\gamma}_T}\right)$$

EMM Asymptotics

- Central Limit Theorem

$$\sqrt{T} m_T(\rho_0, \tilde{\gamma}_T) \xrightarrow{d} N\left(0, C(\rho_0) d^{-1} S d^{-1} C(\rho_0)\right)$$

$$C(\rho) = \lim_{T \to \infty} \left(\left. \frac{\partial m_T(\rho, \tilde{\gamma}_T)}{\partial \gamma'} \right|_{\gamma = \tilde{\gamma}_T} \right) = \left. \frac{\partial m_T(\rho, \gamma)}{\partial \gamma} \right|_{\gamma = \gamma_0}$$

- Key to simplification: $C(\rho_0) = d$

$$\sqrt{T} m_T(\rho_0, \tilde{\gamma}_T) \xrightarrow{d} N(0, S).$$

- Logic leads to EMM estimator

$$\tilde{\rho}_T = \operatorname*{argmin}_{\rho} m_T(\rho, \tilde{\gamma}_T)' \tilde{S}_T^{-1} m_T(\rho, \tilde{\gamma}_T).$$

- $\tilde{S}_T$ is sample counterpart to S

More about EMM

- Variance-covariance matrix of parameter estimates is

$$\tilde{\Sigma}_T = \frac{1}{T}[(\tilde{M}_T)'\tilde{S}_T^{-1}(\tilde{M}_T)]^{-1},$$

$$\tilde{M}_T = \left.\frac{\partial m_T(\rho, \tilde{\gamma}_T)}{\partial \rho'}\right|_{\rho=\tilde{\rho}_T}.$$

- EMM is a GMM estimator; standard GMM test uses overidentifying restrictions to evaluate adequacy of model
- Auxiliary function is unspecified
  - Common choice is semi-nonparametric (SNP); vector-autoregression used to describe conditional mean, non-normal innovations with GARCH effects
  - If true likelihood function is used as auxiliary function, parameter estimates and asymptotic SDs are same as in ML case

Summary of Monte Carlo results for EMM/SNP

- Overidentifying restrictions reject 1st generation Gaussian model at the 5% level in 40% of the simulations
- As models get more complicated, biases in EMM parameter estimates and standard errors grow unacceptably large

- Reason for failure of EMM: A bad weighting matrix for the moments
  - Recall asymptotic variance-covariance matrix of EMM moment vector:

$$\sqrt{T}m_T(\rho_0, \tilde{\gamma}_T) \xrightarrow{d} N\left(0, C(\rho_0)d^{-1}Sd^{-1}C(\rho_0)\right)$$

    d is 2nd deriv of auxiliary function evaluated at sample data + true auxiliary params

    C is 2nd deriv of auxiliary function evaluated at infinite amount of "true" data + true auxiliary params

    S is variance-covariance matrix of auxiliary function score vector
  - Asymptotically, C and d^{-1} cancel

- But when data are highly persistent, curvature of auxiliary function at sample data typically differs substantially from expected curvature

  Result is inefficient parameter estimates, bad test statistics

- This can be fixed by constructing sample estimates of d, C, but in practice this is possible only when original likelihood function is tractable

- Our conclusion: EMM should not be used to estimate parameters of a highly persistent process

- We recommend as an alternative a varient of the Kalman filter

Kalman filter

- Usual Kalman filter setting

  1. Linear relation between observables (yields), unobservables (state vector)
  2. Linear conditional mean of unobservables
  3. Gaussian innovations of unobservables and noise in observables; constant variances

- 2nd generation term structure models retain (1), not necessarily (2) or (3)

- If not,

  1. Linearize instantaneous drift of unobservables; use as proxy for conditional mean
  2. Use instantaneous variance of unobservables, scaled by time, as proxy for discrete-time variances; treat as Gaussian

- Inconsistent

- Our Monte Carlo results show . . .

1. In presence of stochastic volatility and/or nonlinear drifts, estimation with the Kalman filter is less efficient than ML estimation

   Less precision, somewhat greater bias in parameters

2. But in settings where simulations are necessary to implement ML, run time is 25–60 times faster than ML

3. Since examination of finite-sample properties is important before interpreting estimation results, run-time considerations are paramount

Conclusions

1. 2nd generation term structure models present estimation difficulties not present in 1st generation models
   - With ML, strong biases in risk premia
   - ML may require simulation
2. The linearized Kalman filter is a reasonable alternative to ML, but EMM is not

Latest version of paper is at http://faculty.haas.berkeley.edu/duffee