Quadratic Variance Swap Models
Theory and Evidence

Damir Filipović
Swiss Finance Institute
Ecole Polytechnique Fédérale de Lausanne
(joint with Elise Gourier and Loriano Mancini)

6th World Congress of the Bachelier Finance Society
Toronto, 26 June 2010
Swissquote Conference on Interest Rate and Credit Risk

Date: Thursday 28th and Friday 29th October 2010

Venue: Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

Scope: Interest rate and credit risk constitute the major risk sources for banks, insurance companies, and other financial institutions. This conference brings together different perspectives and tools for the valuation and managing of these risks. It addresses academics and practitioners, and shall foster the interaction between individuals and across institutions.

Speakers include:

- Tomas Björk (Stockholm School of Economics)
- Giovanni Cesari (UBS)
- Pierre Collin-Dufresne (Columbia University)
- Rama Cont (University Paris 6)
- Jean-Pierre Danthine (Swiss National Bank)
- Mark Davis (Imperial College)
- David Lando (Copenhagen Business School)
- Alex Lipton (Bank of America Merrill Lynch and Imperial College)
- Dilip Madan (University of Maryland)
- Fabio Mercurio (Bloomberg)
- Antoon Pelsser (Maastricht University)
- Kenneth Singleton (Stanford University)

Scientific Committee: Tomas Björk (Stockholm School of Economics), Mark Davis (Imperial College), Damir Filipovic (EPFL), David Lando (Copenhagen Business School), Wolfgang Runggaldier (University of Padova), Kenneth Singleton (Stanford University)
Outline

1. Variance Swaps
2. Quadratic Term Structures
3. Quadratic (Pearson) Diffusions
4. Model Estimation
Outline

1 Variance Swaps

2 Quadratic Term Structures

3 Quadratic (Pearson) Diffusions

4 Model Estimation
Realized Variance

- Filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})\)
- Price process (e.g. S&P 500 index): semimartingale \(S\)
- Annualized realized variance on \(t = t_0 < \cdots < t_k = T\):

\[
RV_{t,T}^2 = \frac{k}{n} \sum_{i=1}^{k} \left( \frac{\log S_{t_i}}{S_{t_{i-1}}} \right)^2
\]

where \(n = \) number of trading days per year
- Approximate by quadratic variation, for \(k \to \infty\):

\[
\sum_{i=1}^{k} \left( \frac{\log S_{t_i}}{S_{t_{i-1}}} \right)^2 \to [\log S]_T - [\log S]_t
\]

- Justified for daily sampling
Realized Variance

- Filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})\)
- Price process (e.g. S&P 500 index): semimartingale \(S\)
- Annualized realized variance on \(t = t_0 < \cdots < t_k = T\):

\[
RV_{t,T}^2 = \frac{k}{n} \sum_{i=1}^{k} \left( \log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2
\]

where \(n = \) number of trading days per year

- Approximate by quadratic variation, for \(k \to \infty\):

\[
\sum_{i=1}^{k} \left( \log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2 \to [\log S]_T - [\log S]_t
\]

- Justified for daily sampling
Realized Variance

- Filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})\)
- Price process (e.g. S&P 500 index): semimartingale \(S\)
- Annualized realized variance on \(t = t_0 < \cdots < t_k = T\):

\[
RV_{t,T}^2 = \frac{k}{n} \sum_{i=1}^{k} \left( \log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2
\]

where \(n = \text{number of trading days per year}\)
- Approximate by quadratic variation, for \(k \to \infty\):

\[
\sum_{i=1}^{k} \left( \log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2 \to [\log S]_T - [\log S]_t
\]
- Justified for daily sampling
Realized Variance

- Filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})\)
- Price process (e.g. S&P 500 index): semimartingale \(S\)
- Annualized realized variance on \(t = t_0 < \cdots < t_k = T\):

\[
RV_{t,T}^2 = \frac{k}{n} \sum_{i=1}^{k} \left( \log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2
\]

where \(n = \) number of trading days per year
- Approximate by quadratic variation, for \(k \to \infty\):

\[
\sum_{i=1}^{k} \left( \log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2 \to [\log S]_T - [\log S]_t
\]

- Justified for daily sampling
Realized Variance

- Filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})\)
- Price process (e.g. S&P 500 index): semimartingale \(S\)
- Annualized realized variance on \(t = t_0 < \cdots < t_k = T\):

\[
RV^2_{t,T} = \frac{k}{n} \sum_{i=1}^{k} \left( \log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2
\]

where \(n = \text{number of trading days per year}\)
- Approximate by quadratic variation, for \(k \to \infty\):

\[
\sum_{i=1}^{k} \left( \log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2 \to [\log S]_T - [\log S]_t
\]

- Justified for daily sampling
Variance Swaps

- A variance swap initiated at $t$ with maturity $T$ pays

$$RV_{t,T} - VS_{t,T}$$

- $VS_{t,T} = \text{variance swap rate fixed at } t$
- Assume deterministic risk-free rate $r$:

$$VS_{t,T}^2 = \frac{1}{T-t} \mathbb{E}_Q \left[ [\log S]_T - [\log S]_t \mid \mathcal{F}_t \right]$$

- Provides hedging instrument against volatility increases, which often coincide with drops of stock prices
A **variance swap** initiated at $t$ with maturity $T$ pays

$$RV_{t,T}^2 - VS_{t,T}^2$$

**$VS_{t,T}^2$** = variance swap rate fixed at $t$

Assume deterministic risk-free rate $r$:

$$VS_{t,T}^2 = \frac{1}{T-t} \mathbb{E}_Q \left( [\log S]_T - [\log S]_t \mid \mathcal{F}_t \right)$$

Provides hedging instrument against volatility increases, which often coincide with drops of stock prices
Variance Swaps

- A variance swap initiated at $t$ with maturity $T$ pays
  \[ RV^2_{t,T} - VS^2_{t,T} \]

- $VS^2_{t,T} = \text{variance swap rate fixed at } t$

- Assume deterministic risk-free rate $r$:
  \[ VS^2_{t,T} = \frac{1}{T-t} \mathbb{E}_Q \left[ [\log S]_T - [\log S]_t \mid \mathcal{F}_t \right] \]

- Provides hedging instrument against volatility increases, which often coincide with drops of stock prices
Variance Swaps

- A variance swap initiated at $t$ with maturity $T$ pays
  \[ RV_{t,T}^2 - VS_{t,T}^2 \]

- $VS_{t,T}^2 = \text{variance swap rate fixed at } t$

- Assume deterministic risk-free rate $r$:
  \[ VS_{t,T}^2 = \frac{1}{T - t} \mathbb{E}_Q \left[ \log S_T - \log S_t \mid \mathcal{F}_t \right] \]

- Provides hedging instrument against volatility increases, which often coincide with drops of stock prices
S-Characteristics

- Assume

\[
\frac{dS_t}{S_t} = r_t \, dt + \sigma_t \, dW_t + \int_{\mathbb{R}} (e^x - 1) (\mu(dt, dx) - \nu_t(dx) \, dt)
\]

- In particular,

\[
\Delta \log S_t = \int_{\mathbb{R}} x \, \mu(dt, dx)
\]

- Hence

\[
[\log S]_T - [\log S]_t = \int_t^T \sigma_s^2 \, ds + \int_t^T \int_{\mathbb{R}} x^2 \, \mu(ds, dx)
\]
S-Characteristics

- Assume
  \[
  \frac{dS_t}{S_{t-}} = r_t \, dt + \sigma_t \, dW_t + \int_{\mathbb{R}} (e^x - 1) \left( \mu(dt, dx) - \nu_t(dx) \right) dt
  \]

- In particular,
  \[
  \Delta \log S_t = \int_{\mathbb{R}} x \mu(dt, dx)
  \]

- Hence
  \[
  [\log S]_T - [\log S]_t = \int_t^T \sigma_s^2 \, ds + \int_t^T \int_{\mathbb{R}} x^2 \mu(ds, dx)
  \]
S-Characteristics

- Assume
  \[
  \frac{dS_t}{S_{t-}} = r_t \, dt + \sigma_t \, dW_t + \int_{\mathbb{R}} (e^x - 1) (\mu(dt, dx) - \nu_t(dx) dt)
  \]

- In particular,
  \[
  \Delta \log S_t = \int_{\mathbb{R}} x \mu(dt, dx)
  \]

- Hence
  \[
  [\log S]_T - [\log S]_t = \int_t^T \sigma_s^2 \, ds + \int_t^T \int_{\mathbb{R}} x^2 \mu(ds, dx)
  \]
Britten–Jones and Neuberger [2], Jiang and Tian [11], Carr and Wu [4], a.o. showed:

\[
\begin{align*}
\left[\log S\right]_T - \left[\log S\right]_t &= \int_0^{F_t} \frac{2}{K^2} (K - S_T)^+ \, dK + \int_{F_t}^{\infty} \frac{2}{K^2} (S_T - K)^+ \, dK \\
&\quad + 2 \int_t^T \left( \frac{1}{F_{S^-}} - \frac{1}{F_t} \right) \, dF_s \\
&\quad - 2 \int_t^T \int_{\mathbb{R}} \left( e^x - 1 - x - \frac{x^2}{2} \right) \mu(ds, dx).
\end{align*}
\]

\[F_t = \frac{S_t}{P(t, T)} = T\text{-futures price of } S\]
Model-Free Replication

- Britten–Jones and Neuberger [2], Jiang and Tian [11], Carr and Wu [4], a.o. showed:

\[
\frac{\log S_T - \log S_t}{F_t} = \int_0^{F_t} \frac{2}{K^2} (K - S_T)^+ \, dK + \int_{F_t}^\infty \frac{2}{K^2} (S_T - K)^+ \, dK \\
+ 2 \int_t^T \left( \frac{1}{F_{s-}} - \frac{1}{F_t} \right) \, dF_s \\
- 2 \int_t^T \int_\mathbb{R} \left( e^x - 1 - x - \frac{x^2}{2} \right) \mu(ds, dx)
\]

- \( F_t = S_t / P(t, T) = T \)-futures price of \( S \)
Model-Free Valuation

• Taking $\mathbb{Q}$-expectation:

$$V S^2_{t, T} = \frac{2}{T - t} \int_0^\infty \frac{\Theta_t(K, T)}{P(t, T) K^2} dK + \epsilon$$

with error term

$$\epsilon = - \frac{2}{T - t} \mathbb{E}_Q \left[ \int_t^T \int_{\mathbb{R}} \left( e^x - 1 - x - \frac{x^2}{2} \right) \nu_s(dx)ds \mid \mathcal{F}_t \right]$$

• $\Theta_t(K, T) = \text{out-of-the-money option price}$
Model-Free Valuation

• Taking $Q$-expectation:

$$VS_{t,T}^2 = \frac{2}{T-t} \int_0^\infty \frac{\Theta_t(K, T)}{P(t, T)K^2} dK + \epsilon$$

with error term

$$\epsilon = -\frac{2}{T-t} \mathbb{E}_Q \left[ \int_t^T \int_\mathbb{R} \left( e^x - 1 - x - \frac{x^2}{2} \right) \nu_s(dx)ds \mid \mathcal{F}_t \right]$$

• $\Theta_t(K, T)$ = out-of-the-money option price
• Chicago Board Options Exchange Volatility Index (VIX)

\[ \text{VIX}_t = \sqrt{V S^2_{t, t+30 \text{ days}}} \times 100 \% \]

calculated as weighted blend of options on S&P 500 index

• Introduced in 1993, revised in 2003
• Industry benchmark for market volatility
VIX

- Chicago Board Options Exchange Volatility Index (VIX)

\[ VIX_t = \sqrt{\text{VS}^2_{t, t+30 \text{ days}}} \times 100 \% \]

calculated as weighted blend of options on S&P 500 index

- Introduced in 1993, revised in 2003

- Industry benchmark for market volatility
• Chicago Board Options Exchange Volatility Index (VIX)

\[ VIX_t = \sqrt{VS_{t,t+30 \text{ days}}^2} \times 100 \% \]

calculated as weighted blend of options on S&P 500 index

• Introduced in 1993, revised in 2003

• Industry benchmark for market volatility
Term-Structure Models

- OTC variance swaps at many different maturities available ⇒ design and estimate term-structure models of variance swap rates and risk premiums!

Figure: Variance swap rates \( \sqrt{V_{S_t}^2} \) on the S&P 500 index from Jan 4, 1996 to Apr 2, 2007. Source: Bloomberg
Forward Variance

• Define the forward variance

\[
f(t, T) = \mathbb{E}_Q \left[ \sigma^2_T + \int_{\mathbb{R}} x^2 \nu_T(dx) \mid \mathcal{F}_t \right]
\]

• Then the variance swap rates equal

\[
VS_{t,T}^2 = \frac{1}{T-t} \mathbb{E}_Q \left[ \int_t^T \sigma^2_s ds + \int_t^T \int_{\mathbb{R}} x^2 \mu(ds, dx) \mid \mathcal{F}_t \right] = \frac{1}{T-t} \int_t^T f(t, s) \, ds
\]

• The spot variance is

\[
\nu_t = \lim_{T \downarrow t} VS_{t,T}^2 = f(t, t)
\]

• Note the analogy to yields vs. forward rates
Forward Variance

- Define the forward variance

\[ f(t, T) = \mathbb{E}_Q \left[ \sigma_T^2 + \int_{\mathbb{R}} x^2 \nu_T(dx) \mid \mathcal{F}_t \right] \]

- Then the variance swap rates equal

\[ \text{VS}_{t, T}^2 = \frac{1}{T - t} \mathbb{E}_Q \left[ \int_t^T \sigma_s^2 ds + \int_t^T \int_{\mathbb{R}} x^2 \mu(ds, dx) \mid \mathcal{F}_t \right] \]

\[ = \frac{1}{T - t} \int_t^T f(t, s) ds \]

- The spot variance is

\[ \nu_t = \lim_{T \downarrow t} \text{VS}_{t, T}^2 = f(t, t) \]

- Note the analogy to yields vs. forward rates
Forward Variance

- Define the forward variance

\[ f(t, T) = \mathbb{E}_Q \left[ \sigma_T^2 + \int_\mathbb{R} x^2 \nu_T(dx) \mid \mathcal{F}_t \right] \]

- Then the variance swap rates equal

\[
\text{VS}^2_{t,T} = \frac{1}{T-t} \mathbb{E}_Q \left[ \int_t^T \sigma_s^2 ds + \int_t^T \int_\mathbb{R} x^2 \mu(ds, dx) \mid \mathcal{F}_t \right] \\
= \frac{1}{T-t} \int_t^T f(t, s) ds
\]

- The spot variance is

\[ \nu_t = \lim_{T \downarrow t} \text{VS}^2_{t,T} = f(t, t) \]

- Note the analogy to yields vs. forward rates
Forward Variance

- Define the forward variance

\[ f(t, T) = \mathbb{E}_Q \left[ \sigma_T^2 + \int_{\mathbb{R}} x^2 \nu_T(dx) \mid \mathcal{F}_t \right] \]

- Then the variance swap rates equal

\[
VS_{t,T}^2 = \frac{1}{T-t} \mathbb{E}_Q \left[ \int_{t}^{T} \sigma_s^2 \, ds + \int_{t}^{T} \int_{\mathbb{R}} x^2 \mu(ds, dx) \mid \mathcal{F}_t \right] \\
= \frac{1}{T-t} \int_{t}^{T} f(t, s) \, ds
\]

- The spot variance is

\[ \nu_t = \lim_{T \downarrow t} VS_{t,T}^2 = f(t, t) \]

- Note the analogy to yields vs. forward rates
Term-Structure Models: Program

- Exogenous (factor) model of $f(t, T)$, or $\nu_t$, under $\mathbb{P}$ and $\mathbb{Q}$

- Define index $\mathbb{Q}$-dynamics ...

$$\frac{dS_t}{S_t} = r_t \, dt + \sigma_t \, dW_t + \int \left( e^{\delta_t(\xi)} - 1 \right) (\mu(dt, d\xi) - \nu_t(d\xi)dt)$$

- ... such that spot variance satisfies

$$\nu_t = \sigma_t^2 + \int \delta_t(\xi)^2 \nu_t(d\xi)$$

and

$$d[\nu, \log S]_t \leq 0 \quad ("leverage effect")$$

- E.g. Buehler [3], Egloff et al. [9], Cont and Kokholm [7], a.o.
Term-Structure Models: Program

- Exogenous (factor) model of \( f(t, T) \), or \( \nu_t \), under \( \mathbb{P} \) and \( \mathbb{Q} \)
- Define index \( \mathbb{Q} \)-dynamics . . .

\[
\frac{dS_t}{S_t} = r_t \, dt + \sigma_t \, dW_t + \int \left( e^{\delta_t(\xi)} - 1 \right) \left( \mu(dt, d\xi) - \nu_t(d\xi) \right) dt
\]

. . . such that spot variance satisfies

\[
\nu_t = \sigma_t^2 + \int \delta_t(\xi)^2 \nu_t(d\xi)
\]

and

\[
d[\nu, \log S]_t \leq 0 \quad ("leverage effect")
\]

- E.g. Buehler [3], Egloff et al. [9], Cont and Kokholm [7], a.o.
Term-Structure Models: Program

- Exogenous (factor) model of $f(t, T)$, or $\nu_t$, under $\mathbb{P}$ and $\mathbb{Q}$
- Define index $\mathbb{Q}$-dynamics...

\[
\frac{dS_t}{S_{t^-}} = r_t \, dt + \sigma_t \, dW_t + \int \left( e^{\delta_t(\xi)} - 1 \right) \left( \mu(dt, d\xi) - \nu_t(d\xi) \, dt \right)
\]

- ...such that spot variance satisfies

\[
\nu_t = \sigma_t^2 + \int \delta_t(\xi)^2 \, \nu_t(d\xi)
\]

and

\[
d[\nu, \log S]_t \leq 0 \quad ("\text{leverage effect"})
\]

- E.g. Buehler [3], Egloff et al. [9], Cont and Kokholm [7], a.o.
Term-Structure Models: Program

- Exogenous (factor) model of $f(t, T)$, or $\nu_t$, under $\mathbb{P}$ and $\mathbb{Q}$
- Define index $\mathbb{Q}$-dynamics...

$$\frac{dS_t}{S_{t-}} = r_t \, dt + \sigma_t \, dW_t + \int \left( e^{\delta_t(\xi)} - 1 \right) (\mu(dt, d\xi) - \nu_t(d\xi)dt)$$

- ...such that spot variance satisfies

$$\nu_t = \sigma_t^2 + \int \delta_t(\xi)^2 \nu_t(d\xi)$$

and

$$d[\nu, \log S]_t \leq 0 \quad \text{("leverage effect")}$$

- E.g. Buehler [3], Egloff et al. [9], Cont and Kokholm [7], a.o.
Outline

1 Variance Swaps

2 Quadratic Term Structures

3 Quadratic (Pearson) Diffusions

4 Model Estimation
Factor Model

- Forward variance factor model
  \[ f(t, T) = g(T - t, X_t) \]

- State space \( X \subseteq \mathbb{R} \) open

- Jump-diffusion state process
  \[ dX_t = b(X_t)dt + \Sigma(X_t)dW_t + \int_{\mathbb{R}} \xi(\mu(dt, d\xi) - \nu(X_{t-}, d\xi)dt) \]

- \( b, \Sigma, \gamma, g \) nice enough \(...\) in particular linear growth
  \[ b(x)^2 + \Sigma(x)^2 + \int_{\mathbb{R}} \xi^2 \nu(x, d\xi) \leq K(1 + x^2) \]
Factor Model

- Forward variance factor model

\[ f(t, T) = g(T - t, X_t) \]

- State space \( \mathcal{X} \subseteq \mathbb{R} \) open

- Jump-diffusion state process

\[ dX_t = b(X_t)dt + \Sigma(X_t)dW_t + \int_{\mathbb{R}} \xi(\mu(dt, d\xi) - \nu(X_{t-}, d\xi)dt) \]

- \( b, \Sigma, \gamma, g \) nice enough . . . in particular linear growth

\[ b(x)^2 + \Sigma(x)^2 + \int_{\mathbb{R}} \xi^2 \nu(x, d\xi) \leq K(1 + x^2) \]
Factor Model

- Forward variance factor model
  \[ f(t, T) = g(T - t, X_t) \]

- State space \( \mathcal{X} \subseteq \mathbb{R} \) open

- Jump-diffusion state process
  \[ dX_t = b(X_t)dt + \Sigma(X_t)dW_t + \int_{\mathbb{R}} \xi (\mu(dt, d\xi) - \nu(X_t, d\xi))dt \]

- \( b, \Sigma, \gamma, g \) nice enough \ldots in particular linear growth
  \[ b(x)^2 + \Sigma(x)^2 + \int_{\mathbb{R}} \xi^2 \nu(x, d\xi) \leq K(1 + x^2) \]
Factor Model

- Forward variance factor model
  \[ f(t, T) = g(T - t, X_t) \]

- State space \( \mathcal{X} \subseteq \mathbb{R} \) open

- Jump-diffusion state process
  \[
  dX_t = b(X_t)dt + \Sigma(X_t)dW_t + \int_{\mathbb{R}} \xi (\mu(dt, d\xi) - \nu(X_{t-}, d\xi)dt)
  \]

- \( b, \Sigma, \gamma, g \) nice enough . . . in particular linear growth
  \[
  b(x)^2 + \Sigma(x)^2 + \int_{\mathbb{R}} \xi^2 \nu(x, d\xi) \leq K(1 + x^2)
  \]
Quadratic Term Structure

**Theorem 2.1.**
The forward variance model admits a **quadratic term structure**:

\[ g(t, x) = \phi(t) + \psi(t)x + \pi(t)x^2 \]

if (and essentially only if) the state process \( X \) is **quadratic**:

\[ b(x) = b + \beta x \]

\[ \Sigma^2(x) = a + \alpha x + Ax^2 \]

\[ \nu(x, d\xi) = n(d\xi) + \nu(d\xi)x + N(d\xi)x^2. \]

Moreover, …
Theorem 2.1 (cont’d).

...the functions $\phi$, $\psi$, $\pi$ satisfy the linear ODE

$$\frac{d}{dt} \begin{pmatrix} \phi \\ \psi \\ \pi \end{pmatrix} = \begin{pmatrix} 0 & b & a + \int_{\mathbb{R}} \xi^2 n(d\xi) \\ 0 & \beta & 2b + \alpha + \int_{\mathbb{R}} \xi^2 \nu(d\xi) \\ 0 & 0 & 2\beta + A + \int_{\mathbb{R}} \xi^2 N(d\xi) \end{pmatrix} \begin{pmatrix} \phi \\ \psi \\ \pi \end{pmatrix}$$

with initials $\phi_0$, $\psi_0$, $\pi_0$ determined by the spot variance function

$$g_0(x) \equiv g(0, x) = \phi_0 + \psi_0 x + \pi_0 x^2.$$
Proof

- **Kolmogorov backward equation:**

\[
\frac{\partial g(t, x)}{\partial t} = b(x) \frac{\partial g(t, x)}{\partial x} + \frac{1}{2} \Sigma^2(x) \frac{\partial^2 g(t, x)}{\partial x^2} \\
+ \int_{\mathbb{R}} \left(g(t, x + \xi) - g(t, x) - \frac{\partial g(t, x)}{\partial x} \xi\right) \nu(x, d\xi)
\]

- Reads here:

\[
\phi'(t) + \psi'(t)x + \pi'(t)x^2 = b(x) (\psi(t) + 2\pi(t)x) + \Sigma^2(x)\pi(t) \\
+ \int_{\mathbb{R}} \pi(t)\xi^2 \nu(x, d\xi) \\
= \psi(t)P_1(x) + \pi(t)P_2(x)
\]
Proof

- Kolmogorov backward equation:

\[
\frac{\partial g(t, x)}{\partial t} = b(x) \frac{\partial g(t, x)}{\partial x} + \frac{1}{2} \Sigma^2(x) \frac{\partial^2 g(t, x)}{\partial x^2}
\]

\[
+ \int_{\mathbb{R}} \left( g(t, x + \xi) - g(t, x) - \frac{\partial g(t, x)}{\partial x} \xi \right) \nu(x, d\xi)
\]

- Reads here:

\[
\phi'(t) + \psi'(t)x + \pi'(t)x^2 = b(x) (\psi(t) + 2\pi(t)x) + \Sigma^2(x)\pi(t)
\]

\[
+ \int_{\mathbb{R}} \pi(t)\xi^2 \nu(x, d\xi)
\]

\[
= \psi(t)P_1(x) + \pi(t)P_2(x)
\]
• Hence

\[ P_1(x) = b(x) \]

\[ P_2(x) = 2b(x)x + \Sigma^2(x) + \int_{\mathbb{R}} \xi^2 \nu(x, d\xi) \]

are quadratic polynomials in \( x \)

• Necessity follows for diffusion case (\( \nu(x, d\xi) = 0 \))

• Separate terms in 1, \( x \), \( x^2 \) yields the result
• Hence

\[ P_1(x) = b(x) \]

\[ P_2(x) = 2b(x)x + \Sigma^2(x) + \int_{\mathbb{R}} \xi^2 \nu(x, d\xi) \]

are quadratic polynomials in \( x \)

• Necessity follows for diffusion case \( (\nu(x, d\xi) = 0) \)

• Separate terms in 1, \( x \), \( x^2 \) yields the result
Proof

• Hence

\[ P_1(x) = b(x) \]
\[ P_2(x) = 2b(x)x + \Sigma^2(x) + \int_{\mathbb{R}} \xi^2 \nu(x, d\xi) \]

are quadratic polynomials in \( x \)

• Necessity follows for diffusion case (\( \nu(x, d\xi) = 0 \))

• Separate terms in 1, \( x \), \( x^2 \) yields the result
Theorem 2.2.
The process $X$ is quadratic if and only if

$$
\mathbb{E} [X_t^n \mid X_0 = x] = \sum_{k=0}^{n} M_{kn}(t)x^k
$$

for all $n \geq 0$. Moreover, the $(n + 1) \times (n + 1)$-matrix $M$ solves the ODE

$$
\frac{d}{dt} M(t) = BM(t)
$$

$$
M(0) = Id
$$

That is, $M(t) = e^{Bt}$, where . . .
Theorem 2.2 (cont’d).

…the matrix $B$ is upper triangular, and reads for the diffusion case (for simplicity):

$$
B = \begin{pmatrix}
0 & b & 2\frac{a}{2} & 0 & \cdots & 0 \\
0 & \beta & 2 \left( b + \frac{\alpha}{2} \right) & 3 \cdot 2\frac{a}{2} & 0 & \vdots \\
0 & 0 & 2 \left( \beta + \frac{A}{2} \right) & 3 \left( b + 2\frac{\alpha}{2} \right) & \ddots & 0 \\
0 & 0 & 0 & 3 \left( \beta + 2\frac{A}{2} \right) & \ddots & n(n-1)\frac{a}{2} \\
\vdots & \ddots & 0 & \ddots & n \left( b + (n-1)\frac{\alpha}{2} \right) \\
0 & \cdots & \cdots & 0 & n \left( \beta + (n-1)\frac{A}{2} \right)
\end{pmatrix}
$$

- Zhou [13], Forman and Sørensen [10], Cuchiero et al. [8]
**Theorem 2.2 (cont’d).**

...the matrix $B$ is upper triangular, and reads for the diffusion case (for simplicity):

$$B = \begin{pmatrix}
0 & b & 2\frac{a}{2} & 0 & \cdots & 0 \\
0 & \beta & 2(b + \frac{\alpha}{2}) & 3 \cdot 2\frac{a}{2} & 0 & \vdots \\
0 & 0 & 2(\beta + \frac{A}{2}) & 3(b + 2\frac{\alpha}{2}) & \cdots & 0 \\
0 & 0 & 0 & 3(\beta + 2\frac{A}{2}) & \cdots & n(n - 1)\frac{a}{2} \\
\vdots & \ddots & 0 & \cdots & n(b + (n - 1)\frac{\alpha}{2}) \\
0 & \cdots & 0 & \cdots & 0 & n(\beta + (n - 1)\frac{A}{2})
\end{pmatrix}$$

- Zhou [13], Forman and Sørensen [10], Cuchiero et al. [8]
Eigenpolynomials and Moments

- Spectral decomposition

\[ B = S \text{ diag } (\lambda_0, \ldots, \lambda_n) S^{-1} \]

with eigenvalues \( \lambda_k = B_{kk}, \, k = 0, \ldots, n \)

- Gives eigenpolynomials \( p_k(x) = \sum_{j=0}^{k} S_{jk} x^j \) s.t.

\[ \mathbb{E} [p_k(X_t) | X_0 = x] = e^{\lambda_k t} p_k(x), \quad k = 0, \ldots, n \]

- Stationary moments given by first row of \( S^{-1} \):

\[ \mathbb{E} \left[ X_t^k \right] = S_{0k}^{-1}, \quad k = 0, \ldots, n \]

- Very efficiently computable! (e.g. with Mathematica)
Eigenpolynomials and Moments

- Spectral decomposition

\[ B = S \text{ diag } (\lambda_0, \ldots, \lambda_n) S^{-1} \]

with eigenvalues \( \lambda_k = B_{kk}, \; k = 0, \ldots, n \)

- Gives eigenpolynomials \( p_k(x) = \sum_{j=0}^{k} S_{jk} x^j \) s.t.

\[ \mathbb{E} [p_k(X_t) \mid X_0 = x] = e^{\lambda_k t} p_k(x), \quad k = 0, \ldots, n \]

- Stationary moments given by first row of \( S^{-1} \):

\[ \mathbb{E} \left[ X_t^k \right] = S_{0k}^{-1}, \quad k = 0, \ldots, n \]

- Very efficiently computable! (e.g. with Mathematica)
Eigenpolynomials and Moments

- Spectral decomposition

\[ B = S \text{ diag } (\lambda_0, \ldots, \lambda_n) S^{-1} \]

with eigenvalues \( \lambda_k = B_{kk}, \ k = 0, \ldots, n \)

- Gives eigenpolynomials \( p_k(x) = \sum_{j=0}^{k} S_{jk}x^j \) s.t.

\[ \mathbb{E} [p_k(X_t) | X_0 = x] = e^{\lambda_k t} p_k(x), \quad k = 0, \ldots, n \]

- Stationary moments given by first row of \( S^{-1} \):

\[ \mathbb{E} \left[ X_t^k \right] = S_{0k}^{-1}, \quad k = 0, \ldots, n \]

- Very efficiently computable! (e.g. with Mathematica)
Eigenpolynomials and Moments

- Spectral decomposition

\[ B = S \text{ diag } (\lambda_0, \ldots, \lambda_n) S^{-1} \]

with eigenvalues \( \lambda_k = B_{kk}, \ k = 0, \ldots, n \)

- Gives eigenpolynomials \( p_k(x) = \sum_{j=0}^{k} S_{jk}x^j \) s.t.

\[ \mathbb{E}[p_k(X_t) \mid X_0 = x] = e^{\lambda_k t} p_k(x), \ k = 0, \ldots, n \]

- Stationary moments given by first row of \( S^{-1} \):

\[ \mathbb{E}[X_t^k] = S_{0k}^{-1}, \ k = 0, \ldots, n \]

- Very efficiently computable! (e.g. with Mathematica)
Quadratic Interest Rate vs. Variance Swap Models

• An interest rate factor model \( r = r(X) \) admits a quadratic term structure if

\[
\mathbb{E} \left[ e^{-\int_0^t r(X_s) ds} \mid X_0 = x \right] = e^{\Phi(t) + \Psi(t)x + \Pi(t)x^2}
\]

• Ahn, Dittmar, and Gallant [1], Leippold and Wu [12], a.o.

• Chen, Filipović, and Poor [5]: the only (!) consistent jump-diffusion state process \( X \) is Gaussian

\[
dX_t = (b + \beta X_t) dt + \Sigma dW_t
\]

• Quadratic variance swap term-structure models are much more flexible!
Quadratic Interest Rate vs. Variance Swap Models

• An interest rate factor model \( r = r(X) \) admits a quadratic term structure if

\[
\mathbb{E} \left[ e^{-\int_0^t r(X_s) ds} \mid X_0 = x \right] = e^{\Phi(t) + \Psi(t) x + \Pi(t) x^2}
\]

• Ahn, Dittmar, and Gallant [1], Leippold and Wu [12], a.o.

• Chen, Filipović, and Poor [5]: the only (!) consistent jump-diffusion state process \( X \) is Gaussian

\[
dX_t = (b + \beta X_t) \, dt + \Sigma \, dW_t
\]

• Quadratic variance swap term-structure models are much more flexible!
Quadratic Interest Rate vs. Variance Swap Models

- An interest rate factor model \( r = r(X) \) admits a quadratic term structure if
  \[
  \mathbb{E} \left[ e^{-\int_0^t r(X_s)ds} \mid X_0 = x \right] = e^{\Phi(t)+\Psi(t)x+\Pi(t)x^2}
  \]

- Ahn, Dittmar, and Gallant [1], Leippold and Wu [12], a.o.
- Chen, Filipović, and Poor [5]: the only (!) consistent jump-diffusion state process \( X \) is Gaussian
  \[
dX_t = (b + \beta X_t) dt + \Sigma \, dW_t
  \]

- Quadratic variance swap term-structure models are much more flexible!
An interest rate factor model \( r = r(X) \) admits a quadratic term structure if

\[
\mathbb{E} \left[ e^{-\int_0^t r(X_s)ds} \mid X_0 = x \right] = e^{\Phi(t)+\Psi(t)x+\Pi(t)x^2}
\]

Ahn, Dittmar, and Gallant [1], Leippold and Wu [12], a.o.

Chen, Filipović, and Poor [5]: the only (!) consistent jump-diffusion state process \( X \) is Gaussian

\[
dX_t = (b + \beta X_t) dt + \Sigma dW_t
\]

Quadratic variance swap term-structure models are much more flexible!
Model Identification

• The quadratic property of $X$ is invariant w.r.t. affine transformations
  \[ X \mapsto c + \gamma X \]

• Can be offset with affine transformation of the quadratic forward variance function:
  \[ \phi + \psi x + \pi x^2 \mapsto \left( \phi + \psi c + \pi c^2 \right) + \left( \psi \gamma + 2\pi \gamma \right) x + \pi \gamma^2 x^2 \]

• Need canonical representation of $X$ for econometric model identification!
The quadratic property of $X$ is invariant w.r.t. affine transformations

$$X \mapsto c + \gamma X$$

Can be offset with affine transformation of the quadratic forward variance function:

$$\phi + \psi x + \pi x^2 \sim (\phi + \psi c + \pi c^2) + (\psi \gamma + 2\pi \gamma) x + \pi \gamma^2 x^2$$

Need canonical representation of $X$ for econometric model identification!
Model Identification

- The quadratic property of $X$ is invariant w.r.t. affine transformations
  \[ X \mapsto c + \gamma X \]

- Can be offset with affine transformation of the quadratic forward variance function:
  \[ \phi + \psi x + \pi x^2 \mapsto (\phi + \psi c + \pi c^2) + (\psi \gamma + 2\pi \gamma) x + \pi \gamma^2 x^2 \]

- Need canonical representation of $X$ for econometric model identification!
Outline

1. Variance Swaps
2. Quadratic Term Structures
3. Quadratic (Pearson) Diffusions
4. Model Estimation
Canonical Representation

• Forman and Sørensen [10]

**Theorem 3.1.**

Denote by $D = \alpha^2 - 4aA$ the discriminant of the diffusion function of the quadratic diffusion process

$$dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t^2} \, dW_t.$$

Suppose $A > 0$. Then $X$ falls in one of the following three equivalence classes . . .
Canonical Representation

- Forman and Sørensen [10]

**Theorem 3.1.**
Denote by $D = \alpha^2 - 4aA$ the discriminant of the diffusion function of the quadratic diffusion process

$$dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t^2} \, dW_t.$$ 

Suppose $A > 0$. Then $X$ falls in one of the following three equivalence classes . . .
Class 1: $D < 0$

- **State space** $\mathcal{X} = \mathbb{R}$
- Canonical representative:

  $$dX_t = (b + \beta X_t)dt + \sqrt{1 + AX_t^2} \, dW_t$$

  for $b \geq 0$ and $\beta \in \mathbb{R}$

- If $\beta < 0$: $\exists$ stationary density

  $$\mu(x) \propto \left(1 + Ax^2\right)^{\beta/A-1} \exp \left[ \frac{2b}{\sqrt{A}} \arctan[\sqrt{A}x] \right]$$

(Pearson’s type IV, or skew $t$-distribution)

- For $A \to 0$: **Gaussian** limit with stationary density

  $$\mu(x) \propto \exp \left[ 2bx + \beta x^2 \right]$$
Class 1: $D < 0$

- **State space** $\mathcal{X} = \mathbb{R}$
- **Canonical representative:**

$$dX_t = (b + \beta X_t)dt + \sqrt{1 + AX_t^2} dW_t$$

for $b \geq 0$ and $\beta \in \mathbb{R}$

- If $\beta < 0$: $\exists$ stationary density

$$\mu(x) \propto (1 + Ax^2)^{\beta/A-1} \exp \left[ \frac{2b}{\sqrt{A}} \arctan[\sqrt{A}x] \right]$$

(Pearson’s type IV, or skew $t$-distribution)

- For $A \to 0$: **Gaussian** limit with stationary density

$$\mu(x) \propto \exp [2bx + \beta x^2]$$
Class 1: \( D < 0 \)

- State space \( \mathcal{X} = \mathbb{R} \)
- Canonical representative:

\[
dX_t = (b + \beta X_t)dt + \sqrt{1 + AX_t^2} \, dW_t
\]

for \( b \geq 0 \) and \( \beta \in \mathbb{R} \)
- If \( \beta < 0 \): \( \exists \) stationary density

\[
\mu(x) \propto (1 + Ax^2)^{\beta/A-1} \exp \left[ \frac{2b}{\sqrt{A}} \arctan[\sqrt{A}x] \right]
\]

(Pearson’s type IV, or skew \( t \)-distribution)
- For \( A \to 0 \): Gaussian limit with stationary density

\[
\mu(x) \propto \exp[2bx + \beta x^2]
\]
Class 1: $D < 0$

- **State space** $\mathcal{X} = \mathbb{R}$
- **Canonical representative:**

$$dX_t = (b + \beta X_t) dt + \sqrt{1 + AX_t^2} dW_t$$

for $b \geq 0$ and $\beta \in \mathbb{R}$

- If $\beta < 0$: $\exists$ stationary density

$$\mu(x) \propto (1 + Ax^2)^{\beta/A - 1} \exp \left[ \frac{2b}{\sqrt{A}} \arctan[\sqrt{A}x] \right]$$

(Pearson’s type IV, or skew $t$-distribution)

- For $A \to 0$: **Gaussian** limit with stationary density

$$\mu(x) \propto \exp [2bx + \beta x^2]$$
Class 2: \( D = 0 \)

- **State space** \( \mathcal{X} = (0, \infty) \)
- Canonical representative:

\[
dX_t = (b + \beta X_t)dt + \sqrt{AX_t^2} \, dW_t
\]

for \( b \geq 0 \) and \( \beta \in \mathbb{R} \)

- If \( b > 0 \) and \( \beta < 0 \): \( \exists \) stationary density

\[
\mu(x) \propto x^{2\beta/A - 2} \exp \left[ -\frac{2b}{Ax} \right]
\]

(inverse Gamma distribution)

- Also called GARCH diffusion
Class 2: $D = 0$

- State space $\mathcal{X} = (0, \infty)$
- Canonical representative:

$$dX_t = (b + \beta X_t)dt + \sqrt{AX_t^2} \, dW_t$$

for $b \geq 0$ and $\beta \in \mathbb{R}$

- If $b > 0$ and $\beta < 0$: $\exists$ stationary density

$$\mu(x) \propto x^{2\beta/A-2} \exp \left[-\frac{2b}{Ax}\right]$$

(inverse Gamma distribution)

- Also called GARCH diffusion
Class 2: $D = 0$

- State space $\mathcal{X} = (0, \infty)$
- Canonical representative:

$$dX_t = (b + \beta X_t)dt + \sqrt{AX_t^2} \, dW_t$$

for $b \geq 0$ and $\beta \in \mathbb{R}$

- If $b > 0$ and $\beta < 0$: $\exists$ stationary density

$$\mu(x) \propto x^{2\beta/A-2} \exp \left[ -\frac{2b}{Ax} \right]$$

(inverse Gamma distribution)

- Also called GARCH diffusion
Class 2: $D = 0$

- State space $\mathcal{X} = (0, \infty)$
- Canonical representative:

$$dX_t = (b + \beta X_t)dt + \sqrt{AX_t^2} \, dW_t$$

for $b \geq 0$ and $\beta \in \mathbb{R}$
- If $b > 0$ and $\beta < 0$: \exists stationary density

$$\mu(x) \propto x^{2\beta/A-2} \exp \left[ -\frac{2b}{Ax} \right]$$

(inverse Gamma distribution)
- Also called GARCH diffusion
Class 3: $D > 0$

- **State space** $\mathcal{X} = (0, \infty)$
- Canonical representative:

$$dX_t = (b + \beta X_t)dt + \sqrt{X_t + AX_t^2} \, dW_t$$

for $b \geq \frac{1}{2}$ and $\beta \in \mathbb{R}$

- If $\beta < 0$: $\exists$ stationary density

$$\mu(x) \propto x^{2b-1}(1 + Ax)^{2\beta/A-2b-1}$$

(scaled $F$-distribution)

- For $A \to 0$: affine limit case with stationary density

$$\mu(x) \propto x^{2b-1} \exp[2\beta x]$$

(Gamma distribution)
Class 3: $D > 0$

- State space $\mathcal{X} = (0, \infty)$
- Canonical representative:

$$dX_t = (b + \beta X_t)dt + \sqrt{X_t + AX_t^2} \, dW_t$$

for $b \geq \frac{1}{2}$ and $\beta \in \mathbb{R}$

- If $\beta < 0$: $\exists$ stationary density

$$\mu(x) \propto x^{2b-1}(1 + Ax)^{2\beta/A - 2b - 1}$$

(scaled $F$-distribution)

- For $A \to 0$: affine limit case with stationary density

$$\mu(x) \propto x^{2b-1} \exp[2\beta x]$$

(Gamma distribution)
Class 3: $D > 0$

- State space $\mathcal{X} = (0, \infty)$
- Canonical representative:

$$dX_t = (b + \beta X_t)dt + \sqrt{X_t + AX_t^2} \, dW_t$$

for $b \geq \frac{1}{2}$ and $\beta \in \mathbb{R}$

- If $\beta < 0$: $\exists$ stationary density

$$\mu(x) \propto x^{2b-1}(1 + Ax)^{2\beta/A - 2b - 1}$$

(scaled $F$-distribution)

- For $A \to 0$: affine limit case with stationary density

$$\mu(x) \propto x^{2b-1} \exp[2\beta x]$$

(Gamma distribution)
Class 3: $D > 0$

- State space $\mathcal{X} = (0, \infty)$
- Canonical representative:

$$dX_t = (b + \beta X_t)dt + \sqrt{X_t + AX_t^2} \, dW_t$$

for $b \geq \frac{1}{2}$ and $\beta \in \mathbb{R}$

- If $\beta < 0$: exists stationary density

$$\mu(x) \propto x^{2b-1}(1 + Ax)^{2\beta/A - 2b - 1}$$

(scaled $F$-distribution)

- For $A \to 0$: affine limit case with stationary density

$$\mu(x) \propto x^{2b-1} \exp[2\beta x]$$

(Gamma distribution)
Measure Change

- **Aim:** equivalent change of measure $\mathbb{Q} \sim \mathbb{P}$ preserving quadratic property of $X$:

$$dX_t = \left( b^Q + \beta^Q X_t \right) dt$$

$$+ \sqrt{a + \alpha X_t + AX_t^2} \left( dW_t + \frac{\ell + \lambda X_t}{\sqrt{a + \alpha X_t + AX_t^2}} dt \right)$$

$$= dW_t^Q$$

- With return variance risk premium parameters

$$\ell = b - b^Q, \quad \lambda = \beta - \beta^Q$$

- Problem: Novikov’s condition fails in general

- But equivalent measure change works here, due to …
Measure Change

- Aim: equivalent change of measure $\mathbb{Q} \sim \mathbb{P}$ preserving quadratic property of $X$:

$$dX_t = \left( b^Q + \beta^Q X_t \right) dt$$

$$+ \sqrt{a + \alpha X_t + AX_t^2} \left( dW_t + \frac{\ell + \lambda X_t}{\sqrt{a + \alpha X_t + AX_t^2}} dt \right)$$

$$= dW_t^Q$$

- With return variance risk premium parameters

$$\ell = b - b^Q, \quad \lambda = \beta - \beta^Q$$

- Problem: Novikov's condition fails in general
- But equivalent measure change works here, due to ...
Measure Change

- Aim: equivalent change of measure $\mathbb{Q} \sim \mathbb{P}$ preserving quadratic property of $X$:

$$dX_t = \left( b^Q + \beta^Q X_t \right) dt$$

$$+ \sqrt{a + \alpha X_t + AX_t^2} \left( dW_t + \frac{\ell + \lambda X_t}{\sqrt{a + \alpha X_t + AX_t^2}} dt \right)$$

$$= dW_t^Q$$

- With return variance risk premium parameters

$$\ell = b - b^Q, \quad \lambda = \beta - \beta^Q$$

- Problem: Novikov’s condition fails in general
  - But equivalent measure change works here, due to ...
Measure Change

- Aim: equivalent change of measure $\mathbb{Q} \sim \mathbb{P}$ preserving quadratic property of $X$:

$$dX_t = \left( b^Q + \beta^Q X_t \right) dt$$

$$+ \sqrt{a + \alpha X_t + AX_t^2} \left( dW_t + \frac{\ell + \lambda X_t}{\sqrt{a + \alpha X_t + AX_t^2}} dt \right)$$

$$= dW_t^Q$$

- With return variance risk premium parameters

$$\ell = b - b^Q, \quad \lambda = \beta - \beta^Q$$

- Problem: Novikov’s condition fails in general

- But equivalent measure change works here, due to . . .
Theorem 3.2 (Cheridito, Filipović, and Yor [6]).

Let $b$, $\sigma$, $\lambda$ be locally bounded functions on $\mathcal{X}$, and

$$X_t = X_0 + \int_0^t b(X_s) \, ds + \int_0^t \Sigma(X_s) \, dW_s.$$  

Assume that the martingale problem for

$$\tilde{A}f(x) = (b(x) + \Sigma(x)\lambda(x)) f'(x) + \frac{1}{2} \Sigma(x)^2 f''(x)$$

is well posed in $\mathcal{X}$. Then stochastic exponential

$$\mathcal{E}_t(\lambda(X)^\top \cdot W) = \exp \left( \int_0^t \lambda(X_s)^\top dW_s - \frac{1}{2} \int_0^t \|\lambda(X_s)\|^2 \, ds \right)$$

is a martingale.
Application: Martingality of $S$

- Recall quadratic state process

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t^2} \, dW_t^1 \]

- Let $\rho : \mathcal{X} \rightarrow [-1, 1]$ be Lipschitz and s.t. for some $x^* > 0$:

\[ \rho(x) \begin{cases} 
\leq 0, & x \geq x^*, \\
\geq 0, & x \leq -x^*
\end{cases} \]

- Model the discounted S&P 500 index process as

\[ \frac{dS_t}{S_t} = \sqrt{g_0(X_t)} \left( \rho(X_t) \, dW_t^1 + \sqrt{1 - \rho^2(X_t)} \, dW_t^2 \right) \]

- It satisfies (with high probability) the “leverage effect”

\[ d[g_0(X), \log S]_t = g_0'(X_t) \sqrt{g_0(X_t)} \rho(X_t) \leq 0 \]
Application: Martingality of $S$

- Recall quadratic state process

$$dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t^2} \, dW_t^1$$

- Let $\rho : \mathcal{X} \rightarrow [-1, 1]$ be Lipschitz and s.t. for some $x^* > 0$:

$$\rho(x) \begin{cases} 
  \leq 0, & x \geq x^*, \\
  \geq 0, & x \leq -x^* 
\end{cases}$$

- Model the discounted S&P 500 index process as

$$\frac{dS_t}{S_t} = \sqrt{g_0(X_t)} \left( \rho(X_t) \, dW_t^1 + \sqrt{1 - \rho^2(X_t)} \, dW_t^2 \right)$$

- It satisfies (with high probability) the “leverage effect”

$$d[g_0(X), \log S]_t = g_0'(X_t) \sqrt{g_0(X_t)} \rho(X_t) \leq 0$$
Application: Martingality of $S$

- Recall quadratic state process
  \[
  dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t^2} \, dW_t^1
  \]

- Let $\rho : \mathcal{X} \to [-1, 1]$ be Lipschitz and s.t. for some $x^* > 0$:
  \[
  \rho(x) \begin{cases} 
  \leq 0, & x \geq x^*, \\
  \geq 0, & x \leq -x^*
  \end{cases}
  \]

- Model the discounted S&P 500 index process as
  \[
  \frac{dS_t}{S_t} = \sqrt{g_0(X_t)} \left( \rho(X_t) \, dW_t^1 + \sqrt{1 - \rho^2(X_t)} \, dW_t^2 \right)
  \]

- It satisfies (with high probability) the "leverage effect"
  \[
  d[g_0(X), \log S]_t = g_0'(X_t) \sqrt{g_0(X_t)\rho(X_t)} \leq 0
  \]
Application: Martingality of $S$

- Recall quadratic state process

$$dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t^2} \, dW_t^1$$

- Let $\rho : \mathcal{X} \to [-1, 1]$ be Lipschitz and s.t. for some $x^* > 0$:

$$\rho(x) \begin{cases} 
\leq 0, & x \geq x^*, \\
\geq 0, & x \leq -x^*
\end{cases}$$

- Model the discounted S&P 500 index process as

$$\frac{dS_t}{S_t} = \sqrt{g_0(X_t)} \left( \rho(X_t) \, dW_t^1 + \sqrt{1 - \rho^2(X_t)} \, dW_t^2 \right)$$

- It satisfies (with high probability) the "leverage effect"

$$d[g_0(X), \log S]_t = g_0'(X_t) \sqrt{g_0(X_t)} \rho(X_t) \leq 0$$
Application: Martingality of $S$

- **Question:** is $S$ a true $\mathbb{Q}$-martingale? (vital for pricing!)

- **Yes!** Write $S$ as stochastic exponential

$$S_t = S_0 \mathcal{E}_t \left( \lambda(X)^T \cdot W \right)$$

with

$$\lambda(x) = \sqrt{g_0(x)} \left( \frac{\rho(x)}{\sqrt{1 - \rho^2(x)}} \right)$$

- ...and note that the martingale problem for

$$\tilde{A}f(x) = \left( b + \beta x + \sqrt{a + \alpha x + Ax^2} \sqrt{g_0(x)\rho(x)} \right) f'(x)$$

$$+ \frac{1}{2} \left( a + \alpha x + Ax^2 \right) f''(x)$$

is well posed in $\mathcal{X}$ (Yamada–Watanabe)
Application: Martingality of $S$

- Question: is $S$ a true $\mathbb{Q}$-martingale? (vital for pricing!)
- Yes! Write $S$ as stochastic exponential

$$S_t = S_0 \mathcal{E}_t \left( \lambda(X)^T \cdot W \right)$$

with

$$\lambda(x) = \sqrt{g_0(x)} \left( \frac{\rho(x)}{\sqrt{1 - \rho^2(x)}} \right)$$

... and note that the martingale problem for

$$\tilde{A}f(x) = \left( b + \beta x + \sqrt{a + \alpha x + Ax^2} \sqrt{g_0(x) \rho(x)} \right) f'(x)$$

$$+ \frac{1}{2} \left( a + \alpha x + Ax^2 \right) f''(x)$$

is well posed in $\mathcal{X}$ (Yamada–Watanabe)
Application: Martingality of $S$

- Question: is $S$ a true $\mathbb{Q}$-martingale? (vital for pricing!)
- Yes! Write $S$ as stochastic exponential

$$S_t = S_0 \mathcal{E}_t \left( \lambda(X)^\top \cdot W \right)$$

with

$$\lambda(x) = \sqrt{g_0(x)} \left( \frac{\rho(x)}{\sqrt{1 - \rho^2(x)}} \right)$$

- ... and note that the martingale problem for

$$\tilde{A}f(x) = \left( b + \beta x + \sqrt{a + \alpha x + Ax^2} \sqrt{g_0(x)\rho(x)} \right) f'(x) + \frac{1}{2} \left( a + \alpha x + Ax^2 \right) f''(x)$$

is well posed in $\mathcal{X}$ (Yamada–Watanabe)
Outline

1 Variance Swaps

2 Quadratic Term Structures

3 Quadratic (Pearson) Diffusions

4 Model Estimation
Generalized Method of Moments

- Model parameters

\[ \theta = (a(= 0, 1), \alpha(= 0, 1), A, b, \beta, \ell, \lambda, \phi_0, \psi_0, \pi_0) \]

- Observations: 5-vector of VS rates

\[ Y_t = (VS_{t,t+\tau_1}^2, \ldots, VS_{t,t+\tau_5}^2)^\top, \quad t = 1, \ldots, T \]

- Define (martingale increments) vector-valued function

\[ h_t = h(Y_t, Y_{t-1}, \theta) \]

such that

\[ \frac{1}{T} \sum_{t=1}^{T} h_t \approx \mathbb{E}_{\theta_0} [h(G(X_t), G(X_{t-1}), \theta_0)] = 0 \]

where \( G(x) = (g(\tau_1, x), \ldots, g(\tau_5, x))^\top \)

- Notice: eigenpolynomials of \( X_t \) in closed form: Forman and Sørensen [10] provide explicit optimal martingale estimating functions for \( X_t \). Problem: \( X_t \) is not observed
Generalized Method of Moments

- Model parameters
  \[ \theta = (a(= 0, 1), \alpha(= 0, 1), A, b, \beta, \ell, \lambda, \phi_0, \psi_0, \pi_0) \]

- Observations: 5-vector of VS rates
  \[ Y_t = (VS_{t,t+\tau_1}^2, \ldots, VS_{t,t+\tau_5}^2)^T, \quad t = 1, \ldots, T \]

- Define (martingale increments) vector-valued function
  \[ h_t = h(Y_t, Y_{t-1}, \theta) \]
  such that
  \[ \frac{1}{T} \sum_{t=1}^T h_t \approx \mathbb{E}_{\theta_0} [h(G(X_t), G(X_{t-1}), \theta_0)] = 0 \]

where \( G(x) = (g(\tau_1, x), \ldots, g(\tau_5, x))^T \)

- Notice: eigenpolynomials of \( X_t \) in closed form: Forman and Sørensen [10] provide explicit optimal martingale estimating functions for \( X_t \). Problem: \( X_t \) is not observed.
Generalized Method of Moments

- Model parameters
  \[ \theta = (a(=0, 1), \alpha(=0, 1), A, b, \beta, \ell, \lambda, \phi_0, \psi_0, \pi_0) \]

- Observations: 5-vector of VS rates
  \[ Y_t = (V S_{t,t+\tau_1}^2, \ldots, V S_{t,t+\tau_5}^2)^\top, \quad t = 1, \ldots, T \]

- Define (martingale increments) vector-valued function
  \[ h_t = h(Y_t, Y_{t-1}, \theta) \]
  such that
  \[ \frac{1}{T} \sum_{t=1}^{T} h_t \approx \mathbb{E}_{\theta_0}[h(G(X_t), G(X_{t-1}), \theta_0)] = 0 \]
  where \( G(x) = (g(\tau_1, x), \ldots, g(\tau_5, x))^\top \)

- Notice: eigenpolynomials of \( X_t \) in closed form: Forman and Sørensen [10] provide explicit optimal martingale estimating functions for \( X_t \). Problem: \( X_t \) is not observed
Generalized Method of Moments

- Model parameters
  \[ \theta = (a(= 0, 1), \alpha(= 0, 1), A, b, \beta, \ell, \lambda, \phi_0, \psi_0, \pi_0) \]

- Observations: 5-vector of VS rates
  \[ Y_t = (V S_{t,t+\tau_1}^2, \ldots, V S_{t,t+\tau_5}^2)\top, \quad t = 1, \ldots, T \]

- Define (martingale increments) vector-valued function
  \[ h_t = h(Y_t, Y_{t-1}, \theta) \]

  such that
  \[ \frac{1}{T} \sum_{t=1}^{T} h_t \approx \mathbb{E}_{\theta_0} [h(G(X_t), G(X_{t-1}), \theta_0)] = 0 \]

  where \( G(x) = (g(\tau_1, x), \ldots, g(\tau_5, x))\top \)

- Notice: eigenpolynomials of \( X_t \) in closed form: Forman and Sørensen [10] provide explicit optimal martingale estimating functions for \( X_t \). Problem: \( X_t \) is not observed
Extended Kalman Filter

- **De facto standard estimation method**
  - **Pros:**
    - Faster and more stable than GMM (simulation study)
    - Uses all data
    - Filters latent factor $X_t$ (useful for pricing)
  - **Cons:** asymptotically inconsistent (but here finite sample!)
  - **Approximate state transition equation (QML):**
    $X_{t+1} \sim N\left(\mathbb{E}_\theta[X_{t+1} | X_t], \text{var}_\theta[X_{t+1} | X_t]\right)$
  - **Linearize observation equation:**
    
    $$Y_{t+1} = G(\hat{X}_{t+1|t}, \theta) + G'(\hat{X}_{t+1|t}, \theta)(X_{t+1} - \hat{X}_{t+1|t}) + \epsilon_{t+1}$$

    where $\hat{X}_{t+1|t} = \text{predicted state}$
Extended Kalman Filter

- De facto standard estimation method
- Pros:
  - Faster and more stable than GMM (simulation study)
  - Uses all data
  - Filters latent factor $X_t$ (useful for pricing)
- Cons: asymptotically inconsistent (but here finite sample!)
- Approximate state transition equation (QML):
  \[ X_{t+1} \sim \mathcal{N}(\mathbb{E}_\theta[X_{t+1} \mid X_t], \text{var}_\theta[X_{t+1} \mid X_t]) \]
- Linearize observation equation:
  \[ Y_{t+1} = G(\hat{X}_{t+1|t}, \theta) + G'(\hat{X}_{t+1|t}, \theta)(X_{t+1} - \hat{X}_{t+1|t}) + \epsilon_{t+1} \]
  where $\hat{X}_{t+1|t} = \text{predicted state}$
Extended Kalman Filter

- De facto standard estimation method
- Pros:
  - Faster and more stable than GMM (simulation study)
  - Uses all data
  - Filters latent factor $X_t$ (useful for pricing)
- Cons: asymptotically inconsistent (but here finite sample!)
- Approximate state transition equation (QML):
  $$X_{t+1} \sim \mathcal{N}(\mathbb{E}_\theta[X_{t+1} | X_t], \text{var}_\theta[X_{t+1} | X_t])$$
- Linearize observation equation:
  $$Y_{t+1} = G(\hat{X}_{t+1|t}, \theta) + G'(\hat{X}_{t+1|t}, \theta)(X_{t+1} - \hat{X}_{t+1|t}) + \epsilon_{t+1}$$

where $\hat{X}_{t+1|t} =$ predicted state
Extended Kalman Filter

- De facto standard estimation method
- Pros:
  - Faster and more stable than GMM (simulation study)
  - Uses all data
  - Filters latent factor $X_t$ (useful for pricing)
- Cons: asymptotically inconsistent (but here finite sample!)
- Approximate state transition equation (QML):
  \[
  X_{t+1} \sim \mathcal{N}\left(\mathbb{E}_\theta[X_{t+1} \mid X_t], \text{var}_\theta[X_{t+1} \mid X_t]\right)
  \]
- Linearize observation equation:
  \[
  Y_{t+1} = G(\hat{X}_{t+1|t}, \theta) + G'(\hat{X}_{t+1|t}, \theta)(X_{t+1} - \hat{X}_{t+1|t}) + \epsilon_{t+1}
  \]
  where $\hat{X}_{t+1|t}$ = predicted state
Extended Kalman Filter

- **De facto standard estimation method**
- **Pros:**
  - Faster and more stable than GMM (simulation study)
  - Uses all data
  - Filters latent factor $X_t$ (useful for pricing)
- **Cons:** asymptotically inconsistent (but here finite sample!)
- **Approximate state transition equation (QML):**

  $$X_{t+1} \sim \mathcal{N}(\mathbb{E}_\theta[X_{t+1} \mid X_t], \text{var}_\theta[X_{t+1} \mid X_t])$$

- **Linearize observation equation:**

  $$Y_{t+1} = G(\hat{X}_{t+1|t}, \theta) + G'(\hat{X}_{t+1|t}, \theta)(X_{t+1} - \hat{X}_{t+1|t}) + \epsilon_{t+1}$$

  where $\hat{X}_{t+1|t} =$ predicted state
Extended Kalman Filter

- De facto standard estimation method
- Pros:
  - Faster and more stable than GMM (simulation study)
  - Uses all data
  - Filters latent factor $X_t$ (useful for pricing)
- Cons: asymptotically inconsistent (but here finite sample!)
- Approximate state transition equation (QML):

  $$X_{t+1} \sim \mathcal{N}(\mathbb{E}_\theta[X_{t+1} \mid X_t], \text{var}_\theta[X_{t+1} \mid X_t])$$

- Linearize observation equation:

  $$Y_{t+1} = G(\hat{X}_{t+1|t}, \theta) + G'(\hat{X}_{t+1|t}, \theta)(X_{t+1} - \hat{X}_{t+1|t}) + \epsilon_{t+1}$$

where $\hat{X}_{t+1|t} = \text{predicted state}$
Extended Kalman Filter

- De facto standard estimation method
- Pros:
  - Faster and more stable than GMM (simulation study)
  - Uses all data
  - Filters latent factor $X_t$ (useful for pricing)
- Cons: asymptotically inconsistent (but here finite sample!)
- Approximate state transition equation (QML):
  \[ X_{t+1} \sim \mathcal{N} (\mathbb{E}_\theta[X_{t+1} \mid X_t], \text{var}_\theta[X_{t+1} \mid X_t]) \]
- Linearize observation equation:
  \[ Y_{t+1} = G(\hat{X}_{t+1 \mid t}, \theta) + G'(\hat{X}_{t+1 \mid t}, \theta)(X_{t+1} - \hat{X}_{t+1 \mid t}) + \epsilon_{t+1} \]

where $\hat{X}_{t+1 \mid t} = \text{predicted state}$
Extended Kalman Filter

- De facto standard estimation method
- Pros:
  - Faster and more stable than GMM (simulation study)
  - Uses all data
  - Filters latent factor $X_t$ (useful for pricing)
- Cons: asymptotically inconsistent (but here finite sample!)
- Approximate state transition equation (QML):

  $$X_{t+1} \sim \mathcal{N}\left(\mathbb{E}_{\theta}[X_{t+1} \mid X_t], \text{var}_{\theta}[X_{t+1} \mid X_t]\right)$$

- Linearize observation equation:

  $$Y_{t+1} = G(\hat{X}_{t+1|t}, \theta) + G'(\hat{X}_{t+1|t}, \theta)(X_{t+1} - \hat{X}_{t+1|t}) + \epsilon_{t+1}$$

where $\hat{X}_{t+1|t} = \text{predicted state}$
Figure: Variance swap rates $\sqrt{V_{S_t}^2, t+\tau}$ on the S&P 500 index from Jan 4, 1996 to Apr 2, 2007. Source: Bloomberg
Summary Statistics

Panel A: Variance swap rates

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
<th>$Q_{22}$</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20.76</td>
<td>6.80</td>
<td>0.87</td>
<td>4.09</td>
<td>49,806.04</td>
<td>−3.84</td>
</tr>
<tr>
<td>3</td>
<td>20.90</td>
<td>6.54</td>
<td>0.78</td>
<td>3.87</td>
<td>52,308.54</td>
<td>−3.65</td>
</tr>
<tr>
<td>6</td>
<td>21.48</td>
<td>6.32</td>
<td>0.78</td>
<td>3.93</td>
<td>54,570.82</td>
<td>−3.45</td>
</tr>
<tr>
<td>12</td>
<td>22.25</td>
<td>6.06</td>
<td>0.62</td>
<td>3.19</td>
<td>56,549.61</td>
<td>−3.07</td>
</tr>
<tr>
<td>24</td>
<td>22.86</td>
<td>5.90</td>
<td>0.55</td>
<td>2.75</td>
<td>57,460.86</td>
<td>−2.86</td>
</tr>
</tbody>
</table>

Panel B: Calculated VIX

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
<th>$Q_{22}$</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20.68</td>
<td>6.14</td>
<td>0.71</td>
<td>3.38</td>
<td>51,405.71</td>
<td>−3.63</td>
</tr>
<tr>
<td>3</td>
<td>20.71</td>
<td>5.80</td>
<td>0.63</td>
<td>3.20</td>
<td>51,697.76</td>
<td>−3.40</td>
</tr>
<tr>
<td>6</td>
<td>20.78</td>
<td>5.23</td>
<td>0.47</td>
<td>2.79</td>
<td>55,149.03</td>
<td>−3.22</td>
</tr>
</tbody>
</table>

**Table:** Panel A: Summary statistics of the variance swap rates on the S&P 500 index at different maturities (in months) from January 4, 1996 to April 2, 2007, for a total of 2832 observations.
Summary Statistics cont’d

• Panel A cont’d: The table reports mean, standard deviation (Std), skewness (Skew), kurtosis (Kurt); the Ljung–Box portmanteau test for up to 22nd order autocorrelation, $Q_{22}$, 10% critical value is 30.81; the augmented Dickey–Fuller test for unit root involving 22 augmentation lags, a constant term and time trend, ADF, 10% critical value is $-3.16$.

• Panel B: summary statistics of the two-, three- and six-month VIX calculated using SPX options and applying the revised CBOE VIX methodology.
Summary Statistics cont’d

Panel A cont’d: The table reports mean, standard deviation (Std), skewness (Skew), kurtosis (Kurt); the Ljung–Box portmanteau test for up to 22nd order autocorrelation, $Q_{22}$, 10% critical value is 30.81; the augmented Dickey–Fuller test for unit root involving 22 augmentation lags, a constant term and time trend, ADF, 10% critical value is $-3.16$.

Panel B: summary statistics of the two-, three- and six-month VIX calculated using SPX options and applying the revised CBOE VIX methodology.
Principal Component Analysis

- PCA of variance swap curve $\tau \mapsto \sqrt{\text{VS}_{t,t+\tau}^2}$
- One major factor (level), explains **96% of variance**
- Second factor (slope), explains **3% of variance**

**Figure:** First two variance swap curve loadings
Principal Component Analysis

- PCA of variance swap curve $\tau \mapsto \sqrt{\text{VS}_t^2}$
- One major factor (level), explains **96% of variance**
- Second factor (slope), explains 3% of variance

**Figure**: First two variance swap curve loadings
Principal Component Analysis

- PCA of variance swap curve $\tau \mapsto \sqrt{\text{VS}_{t,t+\tau}^2}$
- One major factor (level), explains 96% of variance
- Second factor (slope), explains 3% of variance

**Figure:** First two variance swap curve loadings
Model Specifications

- Full specification: no restrictions

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t} \, dW_t \]

\[ f(t, t) = \phi_0 + \psi_0 X_t + \pi_0 X_t^2 \]

- Nested restricted specifications:
  - \( A = 0 \): affine \( X \)-dynamics
  - \( \pi = 0 \): linear spot variance function
  - \( \psi_0^2 = 4\phi_0\pi_0 \): spot variance function has exactly one zero
  - \( \phi_0 = \psi_0 = 0 \): sv function has exactly one zero at \( x = 0 \)

- Report likelihood ratio

\[ LR = 2 (\log L_{\text{Full}} - \log L_{\text{REST}}) \sim \chi^2_{\# \text{ rest}} \]
Model Specifications

- Full specification: no restrictions

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t} \, dW_t \]

\[ f(t, t) = \phi_0 + \psi_0 X_t + \pi_0 X_t^2 \]

- Nested restricted specifications:
  - \( A = 0 \): affine \( X \)-dynamics
  - \( \pi = 0 \): linear spot variance function
  - \( \psi_0 = 4\phi_0\pi_0 \): spot variance function has exactly one zero
  - \( \phi_0 = \psi_0 = 0 \): sv function has exactly one zero at \( x = 0 \)

- Report likelihood ratio

\[ LR = 2 (\log L_{\text{Full}} - \log L_{\text{REST}}) \sim \chi^2_{\# \text{ rest}} \]
Model Specifications

- Full specification: no restrictions
  \[ dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t} \, dW_t \]
  \[ f(t, t) = \phi_0 + \psi_0 X_t + \pi_0 X_t^2 \]

- Nested restricted specifications:
  - \( A = 0 \): affine \( X \)-dynamics
  - \( \pi = 0 \): linear spot variance function
  - \( \psi_0^2 = 4\phi_0\pi_0 \): spot variance function has exactly one zero
  - \( \phi_0 = \psi_0 = 0 \): sv function has exactly one zero at \( x = 0 \)

- Report likelihood ratio
  \[ LR = 2 (\log L_{\text{Full}} - \log L_{\text{REST}}) \sim \chi^2_{\text{rest}} \]
Model Specifications

- Full specification: no restrictions

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t} \, dW_t \]

\[ f(t, t) = \phi_0 + \psi_0 X_t + \pi_0 X_t^2 \]

- Nested restricted specifications:
  - \( A = 0 \): affine \( X \)-dynamics
  - \( \pi = 0 \): linear spot variance function
    - \( \psi_0^2 = 4\phi_0\pi_0 \): spot variance function has exactly one zero
    - \( \phi_0 = \psi_0 = 0 \): sv function has exactly one zero at \( x = 0 \)

- Report likelihood ratio

\[ LR = 2 (\log L_{\text{Full}} - \log L_{\text{REST}}) \sim \chi^2_{\# \text{ rest}} \]
Model Specifications

• Full specification: no restrictions

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t} \, dW_t \]
\[ f(t, t) = \phi_0 + \psi_0 X_t + \pi_0 X_t^2 \]

• Nested restricted specifications:
  • \( A = 0 \): affine \( X \)-dynamics
  • \( \pi = 0 \): linear spot variance function
  • \( \psi_0^2 = 4\phi_0\pi_0 \): spot variance function has exactly one zero
  • \( \phi_0 = \psi_0 = 0 \): sv function has exactly one zero at \( x = 0 \)

• Report likelihood ratio

\[ LR = 2 (\log L_{\text{Full}} - \log L_{\text{REST}}) \sim \chi^2_{\# \text{ rest}} \]
Model Specifications

- Full specification: no restrictions

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t} \, dW_t \]
\[ f(t, t) = \phi_0 + \psi_0 X_t + \pi_0 X_t^2 \]

- Nested restricted specifications:
  - \( A = 0 \): affine \( X \)-dynamics
  - \( \pi = 0 \): linear spot variance function
  - \( \psi_0^2 = 4\phi_0\pi_0 \): spot variance function has exactly one zero
  - \( \phi_0 = \psi_0 = 0 \): sv function has exactly one zero at \( x = 0 \)

- Report likelihood ratio

\[ LR = 2 (\log L_{\text{Full}} - \log L_{\text{REST}}) \sim \chi^2_{\# \text{ rest}} \]
Model Specifications

- Full specification: no restrictions

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t} \, dW_t \]

\[ f(t, t) = \phi_0 + \psi_0 X_t + \pi_0 X_t^2 \]

- Nested restricted specifications:
  - \( A = 0 \): affine \( X \)-dynamics
  - \( \pi = 0 \): linear spot variance function
  - \( \psi_0^2 = 4\phi_0\pi_0 \): spot variance function has exactly one zero
  - \( \phi_0 = \psi_0 = 0 \): sv function has exactly one zero at \( x = 0 \)

- Report likelihood ratio

\[ LR = 2 (\log L_{\text{Full}} - \log L_{\text{REST}}) \sim \chi^2_{\# \text{ rest}} \]
## Estimation Results: Overview

<table>
<thead>
<tr>
<th>Type</th>
<th>Full</th>
<th>$A = 0$</th>
<th>$\pi = 0$</th>
<th>$\psi_0^2 = 4\phi_0\pi_0$</th>
<th>$\phi_0 = \psi_0 = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$A$</td>
<td>0.402</td>
<td>0</td>
<td>1.620</td>
<td>0.462</td>
<td>0.320</td>
</tr>
<tr>
<td>$b$</td>
<td>2.005</td>
<td>0.529</td>
<td>4.083</td>
<td>0.979</td>
<td>0.098</td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.742</td>
<td>-0.264</td>
<td>-0.839</td>
<td>-1.034</td>
<td>-0.849</td>
</tr>
<tr>
<td>$\ell$</td>
<td>-0.023</td>
<td>-0.003</td>
<td>-1.273</td>
<td>0.300</td>
<td>0.032</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-0.243</td>
<td>0.020</td>
<td>-1.503</td>
<td>-0.590</td>
<td>-0.458</td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>0.016</td>
<td>0.000</td>
<td>0.693</td>
<td>0.002</td>
<td>0</td>
</tr>
<tr>
<td>$\psi_0$</td>
<td>-0.002</td>
<td>0.007</td>
<td>0.008</td>
<td>0.010</td>
<td>0</td>
</tr>
<tr>
<td>$\pi_0$</td>
<td>0.002</td>
<td>0.009</td>
<td>0.003</td>
<td>0.013</td>
<td>1.640</td>
</tr>
<tr>
<td>LR</td>
<td>0</td>
<td><strong>386</strong></td>
<td><strong>620</strong></td>
<td>264</td>
<td>286</td>
</tr>
</tbody>
</table>
Estimation Results

• Full model of class 3 gives best fit

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{X_t + AX_t^2} \, dW_t \]

• All nested restricted specifications strongly rejected, in particular the affine ones (“\(A = 0\)”, “\(\pi = 0\)"

• Class 3 combines affine behavior for small \(X_t\) and quadratic behavior for large \(X_t\)

• Quadratic terms allow for extreme movements and hump shaped VS term structure

• Drawback: spot variance function bounded away from zero (\(\geq 0.1^2\))
Estimation Results

- Full model of class 3 gives best fit

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{X_t + AX_t^2} \, dW_t \]

- All nested restricted specifications strongly rejected, in particular the affine ones ("A = 0", "\pi = 0")

- Class 3 combines affine behavior for small \( X_t \) and quadratic behavior for large \( X_t \)

- Quadratic terms allow for extreme movements and hump shaped VS term structure

- Drawback: spot variance function bounded away from zero (\( \geq 0.1^2 \))
Estimation Results

- Full model of class 3 gives best fit

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{X_t + AX_t^2} \, dW_t \]

- All nested restricted specifications strongly rejected, in particular the affine ones ("A = 0", "\pi = 0")

- Class 3 combines affine behavior for small \( X_t \) and quadratic behavior for large \( X_t \)

- Quadratic terms allow for extreme movements and hump shaped VS term structure

- Drawback: spot variance function bounded away from zero (\( \geq 0.1^2 \))
Estimation Results

- Full model of class 3 gives best fit

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{X_t + AX_t^2} \, dW_t \]

- All nested restricted specifications strongly rejected, in particular the affine ones ("A = 0", "\pi = 0")

- Class 3 combines affine behavior for small \( X_t \) and quadratic behavior for large \( X_t \)

- Quadratic terms allow for extreme movements and hump shaped VS term structure

- Drawback: spot variance function bounded away from zero (\( \geq 0.1^2 \))
Estimation Results

- Full model of class 3 gives best fit

\[ dX_t = (b + \beta X_t) \, dt + \sqrt{X_t + AX_t^2} \, dW_t \]

- All nested restricted specifications strongly rejected, in particular the affine ones ("A = 0", "\pi = 0")
- Class 3 combines affine behavior for small \( X_t \) and quadratic behavior for large \( X_t \)
- Quadratic terms allow for extreme movements and hump shaped VS term structure
- Drawback: spot variance function bounded away from zero (\( \geq 0.1^2 \))
In-Sample Analysis: Predicted VS

Figure: Very good fit of predicted (filtered) variance swap rates vs. data for 6 months maturity
In-Sample Analysis: Humps

Figure: Hump shaped VS term structure on 15-Dec-1998. Quadratic model (left), linear model ($\pi = 0$) right.
Out-of-Sample Analysis: VIX Futures Pricing

Figure: Challenging exercise for the model: VIX futures on Feb 15, 2007. Maturities 30, 60, 90, 120, 150, 180, 270, 420 days. Systematic pricing error for VIX due to jumps. Acceptable result.
**Figure:** Simulation of a spot volatility trajectory over 100 years. In black the linearized model. We see volatility spikes and clustering. Do we need jumps after all?
Conclusion

- **Need for variance swap term-structure models**
- Quadratic term structure (closed form) led to quadratic factor process
- Quadratic models are much more flexible than linear-affine models
- Data imply strong statistical evidence for quadratic terms
- Quadratic models capture nonlinear phenomena: rare events, volatility clustering
Conclusion

- Need for variance swap term-structure models
- Quadratic term structure (closed form) led to quadratic factor process
- Quadratic models are much more flexible than linear-affine models
- Data imply strong statistical evidence for quadratic terms
- Quadratic models capture nonlinear phenomena: rare events, volatility clustering
Conclusion

- Need for variance swap term-structure models
- Quadratic term structure (closed form) led to quadratic factor process
- Quadratic models are much more flexible than linear-affine models
  - Data imply strong statistical evidence for quadratic terms
  - Quadratic models capture nonlinear phenomena: rare events, volatility clustering
Conclusion

- Need for variance swap term-structure models
- Quadratic term structure (closed form) led to quadratic factor process
- Quadratic models are much more flexible than linear-affine models
- Data imply strong statistical evidence for quadratic terms
- Quadratic models capture nonlinear phenomena: rare events, volatility clustering
Conclusion

- Need for variance swap term-structure models
- Quadratic term structure (closed form) led to quadratic factor process
- Quadratic models are much more flexible than linear-affine models
- Data imply strong statistical evidence for quadratic terms
- Quadratic models capture nonlinear phenomena: rare events, volatility clustering
References I


References III


References IV

G. J. Jiang and Y. S. Tian.
The model-free implied volatility and its information content.

M. Leippold and L. Wu.
Asset pricing under the quadratic class.

H. Zhou.
Itô conditional moment generator and the estimation of short-rate processes.