Generalized Transform Analysis of Affine Processes and Asset Pricing Applications

Hui Chen† Scott Joslin‡

April 24, 2009

Abstract

Nonlinearity is an important consideration in many applications of finance and economics, such as pricing securities, computing equilibria, and conducting structural estimations. This paper extends the transform analysis in Duffie, Pan, and Singleton (2000) by providing analytical treatment of a general class of nonlinear transforms for affine jump diffusions. We illustrate the power of the generalized transform through several motivating examples as well as detailed analysis of three problems: (1) pricing defaultable bonds with state-dependent recovery; (2) computing the equilibrium of a Lucas economy with multiple non-i.i.d. trees; and (3) a general class of heterogeneous beliefs models.

†MIT Sloan School of Management, huichen@mit.edu
‡MIT Sloan School of Management, sjoslin@mit.edu

*We thank Ian Martin, Ken Singleton, and the seminar participants at MIT Sloan and Boston University for comments. Tran Ngoc-Khanh provided excellent research assistance and many comments. All the remaining errors are our own.
1 Introduction

In this paper, we provide analytical treatment of a class of transforms for state variables that follow affine jump-diffusions (AJD). These transforms bring analytical and computational tractability to a large class of nonlinear moments, and can be useful in many applications in economics and finance. We illustrate the application of the generalized transform with a variety of examples, including option pricing, term structure modeling, credit risk modeling, method of moment estimations, computing the equilibrium of consumption-based asset pricing models, and a general class of difference-of-opinion models.

For a state variable $X_t$ that follows an affine process, in the sense that the conditional characteristic function is affine$^1$, Duffie et al. (2000), hereafter DPS, derive closed-form expression for the following transform:

$$E_t \left[ \exp \left( - \int_t^T R(X_s, s) \, ds \right) e^{u \cdot X_T} (v_0 + v_1 \cdot X_T) 1_{\{\beta \cdot X_T < y\}} \right],$$  

(1)

where $R(X)$ is an affine function of $X$, which can be interpreted as a stochastic “discount rate”, and $e^{u \cdot X_T} (v_0 + v_1 \cdot X_T) 1_{\{\beta \cdot X_T < y\}}$ is the terminal payoff function at time $T$.

We generalize the DPS result by deriving closed-form expression (up to an integral) for the following generalized transform:

$$E_t \left[ \exp \left( - \int_t^T R(X_s, s) \, ds \right) f(X_T) g(\beta \cdot X_T) \right],$$  

(2)

where $f$ is a polynomial, a log-linear function, or the product of the two; $g$ is a piecewise continuous function with at most polynomial growth (or more generally a tempered distribution) satisfying certain regularity conditions. When $f(X) = e^{u \cdot X} (v_0 + v_1 \cdot X)$ and $g(\beta \cdot X) = 1_{\{\beta \cdot X < y\}}$, we recover the transform of DPS in (1). The flexibility in choosing $g(\cdot)$ and $f(\cdot)$ in (2) makes the generalized transform useful in dealing with generic nonlinearity problems in pricing (nonlinear discount factors or payoffs), estimation (nonlinear moments), and economic modeling.

The primary analytic tool that we use is the Fourier transform. In particular, we utilize knowledge of the conditional characteristic function of the state variable $X_t$ (under certain forward measures) jointly with a Fourier decomposition of the nonlinearity in $g$. This combination brings tractability to our generalized transform by avoiding intermediate Fourier inversions.

$^1$See Duffie et al. (2003) for an elaboration on the characterization via the characteristic function.
We provide several example applications for the generalized transform analysis.

**Option pricing.** When pricing standard European options, the payoff function can be expressed as the product of an exponential function and an indicator function. In this case, we recover the DPS transform as a special case.

**Term structure modeling.** We consider a nonlinear Taylor Rule model that generalizes the model of fed funds target in Piazzesi (2005). With the generalized transform, we have more freedom in modeling the fed policy function. The policy function can be chosen to maintain the requirement that the fed funds target rate, $ff$, is non-negative, has increments in a multiple of 25 basis points, and depend on macro variables such as GDP growth and inflation, but without the restriction that the distribution of $ff$ have the exponential-affine Laplace transform. One application of such a model is to price fed funds futures.

**Credit risk modeling.** We introduce a simple model of default contagion that violates the doubly stochastic assumption in that the default event of one firm affects the default probability of another firm beyond its own default intensity. The flexibility of modeling the default probability for the secondary firm after the default of the primary firm makes it easy to capture the nonlinearity and time-dependence of the contagion effects.

Defaultable bonds are a good example of a contingent claim with nonlinear payoffs, since the recovery rate at default has to be between 0 and 1, and has empirically been shown to nonlinearly depend on macro and firm-specific variables. We introduce a class of state-dependent recovery models into the reduced-form models of default, which substantially relaxes the “recovery of market value” assumption standard in current literature. We derive closed form solutions for the pricing of defaultable zero-coupon bond. The model can also be used to price other credit derivatives such as credit default swaps or recovery locks\(^2\). Our example of Cauchy recovery model demonstrates that ignoring the correlation between recovery rates and default intensity can lead to substantial deviations in credit spreads, especially for bonds with high or low credit quality.

**Method of moments estimation.** The need to compute unconditional and conditional moments of nonlinear functions often arises in the method of moments or GMM estimations. We provide a GMM estimator to a model with latent state variables.

---

\(^2\)This is a forward contract that requires no upfront or running payments, and allows purchase or sale of underlying bonds at a predetermined price if a credit event occurs.
Computing the equilibrium of asset pricing models. Besides nonlinear payoffs, nonlinear discount factors also arise in asset pricing models. We use the example of the Lucas model with two trees. We show that it is straightforward to extend the two tree models of Cochrane et al. (2008) to allow for mean reversion and conditional heteroscedasticity in dividend growth, as well as jumps with time-varying intensities. We use the model to analyze the equilibrium effects of growth rates on prices and risk premia.

Difference-of-opinion models. Even when fundamentals follow simple dynamics, nonlinearity in discount factors may arise due to heterogeneous beliefs among agents. For example, agents may disagree regarding expected changes and volatility of fundamentals and also the likelihood and magnitude of jumps in the state of the economy. In equilibrium, beliefs then become important determinants of asset prices. We provide a general class of difference-of-opinion models that are tractable using the generalized transform, and examine the effects of disagreement regarding jump intensity on asset prices.

Our method is closely related to the transform analysis of Duffie et al. (2000), which is widely used in term structure modeling, option pricing, and structural estimations. Examples include Singleton (2001), Pan (2002), Piazzesi (2005), and Joslin (2009), among many others. The generalized transform in this paper extends the DPS method to a large class of nonlinear functions.

Computing the expectations of nonlinear functions such as in (2) is a very common task in economics and finance. When the nonlinear moments cannot be computed analytically, commonly used alternative methods include simulations or solving numerically the partial differential equations arising from the expectations via the Feynman-Kac methodology. Both methods can be time-consuming and lacking accuracy, especially in high dimensional cases, making them less suitable for estimations or doing comparative statics.

Fourier transform methods are also commonly used to compute the moments of affine processes. A direct approach is to compute the conditional density of the affine state variables through Fourier inversion of the conditional characteristic function, which in turn can be computed using the transform analysis of DPS and is available in closed form in some special cases. One can then evaluate the nonlinear moments directly using the density function. Papers that take this approach include Heston (1993), Chen and Scott (1995), Bates (1996), Bakshi et al. (1997), Dumas et al. (2009), among others. A limitation of this approach is that, for general moments, it involves multiple numerical integrals and is subject to the curse of dimensionality.
A second approach takes the Fourier transform of the nonlinear functions and replaces the nonlinear moments with an integral of the conditional characteristic functions. For example, Bakshi and Madan (2000) price a class of (at most linear growth) derivative securities by spanning the payoffs via characteristic functions. Carr and Madan (1999) addresses the nonlinearity in a European option payoff by taking the Fourier transform of the payoff function with respect to the strike price. Martin (2008) takes the Fourier transform of a nonlinear pricing kernel that arises in the two tree model of Cochrane et al. (2008). In the last two cases, the conditional characteristic functions from the Fourier transform are known in closed form, which helps one obtain option prices or the value of consumption claims.

To the best of our knowledge, this paper is the first to generalize this approach and provide analytical treatment of generic nonlinear moments of affine processes by combining the Fourier method with the transform analysis of DPS. While applications of the Fourier method have so far mostly concentrated in the option pricing literature, we show that the generalized transform can be applied to many more areas of economics and finance.

The paper proceeds as follows. Section 2 gives the main theoretical results on the transform and its two extensions. Section 3 outlines several motivating examples for the transform analysis. Section 4 investigates three concrete example applications of the generalized transform analysis. Section 5 concludes.

2 Generalized Transforms

In this section, we outline our theoretical results. Full details are deferred to the Appendix. As in DPS, we begin by fixing a probability space \((\Omega, \mathcal{F}, P)\) and an information filtration \(\{\mathcal{F}_t\}\), satisfying the usual conditions (see e.g., Protter (2003)), and suppose that \(X\) is a Markov process in some state space \(D \subset \mathbb{R}^n\) satisfying the stochastic differential equation

\[
\begin{aligned}
dx_t &= (K_0 + K_1 X_t) dt + \sqrt{H_0 + H_1 \cdot X_t} dW_t + dZ_t, 
\end{aligned}
\]

where \(W\) is an \(\mathcal{F}_t\)-standard \(n\)-dimensional Brownian motion and \(Z\) is pure jump process with arrival intensity \(\lambda_t = \lambda_0 + \lambda_1 \cdot X_t\) with fixed \(D\)-invariant distribution \(\nu\). Whenever needed, we also assume that there is an affine discount rate function \(R(X_t) = \rho_0 + \rho_1 \cdot X_t\). For brevity, let \(\Theta\) denote the

\footnote{In the \(N\)-tree case, \(N > 2\), Martin (2008) also provides an \((N-2)\)-dimensional integral to compute the associated \((N-1)\)-dimensional transform.}
parameters of the process \((K_0, K_1, H_0, H_1, \lambda_0, \lambda_1, \nu, \rho_0, \rho_1)\). Alternatively, we can define the process in terms of the infinitesimal generator or, as Duffie et al. (2003) and Singleton (2001) stress, in terms of the conditional characteristic function.

2.1 Transform Analysis

In order to establish our main result, let us first review some basic concepts from distribution theory. A function \(f : \mathbb{R}^N \rightarrow \mathbb{R}\) which is smooth and rapidly decreasing in the sense that for any multi-index \(\alpha\) and any \(N \in \mathbb{N}\), \(\|f\|_{N,\alpha} \equiv \sup_x |\partial^\alpha f(x)|(1 + |x|)^N < \infty\) is referred to as a Schwartz function. The collection of all Schwartz functions is denoted \(S\). \(S\) is endowed with the topology generated by the family of semi-norms \(\|f\|_{N,\alpha}\). The dual of \(S\), denoted \(S^*\) and also called the set of tempered distributions, is the set of continuous linear functionals on \(S\). Any continuous function which has at most polynomial growth in the sense that \(|g(x)| < |x|^p\) for some \(p\) and \(x\) large enough is seen to be a tempered distribution through the map

\[
g : S \rightarrow \mathbb{R} \quad g : f \mapsto \int_{x \in \mathbb{R}^N} g(x)f(x)dx
\]

Many tempered distributions do not arrive from functions. An important example is the \(\delta\)-function, \(\delta : h \mapsto h(0)\). For our considerations, the key property is that the set of tempered distributions is suitable for Fourier analysis. For example, a function which is bounded may not have a Fourier transform in the sense of a function, but will possess a Fourier transform that is a tempered distribution. An example is the Heaviside function:

\[
f(x) = 1_{\{x \leq 0\}} \Rightarrow \hat{f}(s) = \frac{1}{2} \delta(s) - \frac{1}{2\pi s}
\]

where integrating against \(1/s\) is to be interpreted as the principle value of the integral. Considering distributions allows us to consider functions which are not integrable and thus in particular may not decay at infinity and may not even be bounded.

We now state our main result:

**Theorem 1.** Suppose that \(f(s) = \exp(s), g \in S^*\) and \((\Theta, \alpha, \beta)\) satisfies Assumption 1 and As-
**sumption 2 in Appendix A. Then**

\[ H(f,g,\alpha,\beta) = E_0 \left[ \exp \left( - \int_0^T R(X_u)du \right) f(\alpha \cdot X_T) g(\beta \cdot X_T) \right] \]

\[ = \frac{1}{2\pi} \langle \hat{g}, G(\alpha + \beta i) \rangle \]  

\hspace{1cm} (6)

where \( \hat{g} \in S^* \) and \( G(\alpha + \beta i) \) denotes the function

\[ s \mapsto G(\alpha + s\beta i) = E_0 \left[ e^{-\int_0^T R(X_u)du} e^{(\alpha + is\beta) \cdot X_T} \right]. \]  

\hspace{1cm} (7)

The function \( G \) is the transform given in DPS. Recalling their result,

\[ G(\alpha + is\beta) = e^{A(T;\alpha + is\beta,\Theta) + B(T;\alpha + is\beta,\Theta) \cdot X_0} \]

where \( A, B \) satisfy the ODE/BVP:

\[ \dot{B} = K_{1^T} B + \frac{1}{2} B^T H_1 B - \rho_1 + \lambda_1 (\phi(B) - 1) \quad B(0) = \alpha + is\beta \]  

\hspace{1cm} (8)

\[ \dot{A} = K_{0^T} B + \frac{1}{2} B^T H_0 B - \rho_0 + \lambda_0 (\phi(B) - 1) \quad A(0) = 0 \]  

\hspace{1cm} (9)

where \( \phi(c) = E_\nu[e^{cZ}] \), the moment-generating function of the jump distribution. Solving the ODE system of (8-9) adds little complication to the transform. The solution is available in closed form in some cases, and can generally be quickly and accurately computed using standard numerical methods.

In the special case that \( \hat{g} \) defines a function, we can write the result as

\[ H = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{g}(s)G(\alpha + is\beta)ds \]

\hspace{1cm} (10)

Fourier transforms of many functions are known in closed form. See, for example, Folland (1984). Additionally, standard rules allow for differentiation, integrations, products, convolutions and other operations to be conducted while maintaining closed form expressions. However, even in the case that the function \( \hat{g} \) is not known in closed form, it can be computed readily.

In some cases of interests, Assumption 2 may be violated. \( \beta \cdot X_T \) may have heavy tails so that, for example, \( E[(\beta \cdot X_T)^4] = \infty \). Another example would be in a pure-jump process where the density may not be continuous. Depending on the case, our result can often be extended by
limiting arguments or by considering different function spaces (such as Sobolev spaces for non-smooth densities.)

It is worth noting that there is some flexibility in the choice of $\alpha$ and $g$. This is because $e^{\alpha \cdot X_T} g(\beta \cdot X_T) = e^{(\alpha - c \beta) \cdot X_T} \tilde{g}(\beta \cdot X_T)$ where $\tilde{g}(s) = e^{cs} g(s)$. This flexibility may be useful in the case where $g$ is not integrable but decreases rapidly as $s$ approaches either $+\infty$ or $-\infty$ (e.g., the logit function). In this case, it may be computationally beneficial to use such a transformation to apply (10) and avoid the use of distributions.

### 2.2 Extensions of Generalized Transforms

The result above can be extended in a number of ways. First, we introduce a class of \textit{pl-linear} (polynomial-log-linear) functions:

$$f(\alpha, v, X) = \sum_i p_i(v_i \cdot X) e^{\alpha_i \cdot X},$$

where $\{p_i\}$ are arbitrary polynomials and $\{\alpha_i\}$ are complex vectors.\footnote{Allowing complex eigenvalues allows one to have oscillatory \textit{sine} and \textit{cosine} terms.} We will refer to any function which cannot be expressed as the product of a polynomial and a log-linear function in (11) as \textit{non-pl-linear}. The following proposition extends the function $f$ in Theorem 1 to any \textit{pl-linear} functions.

**Proposition 1.** Suppose that $v \in \mathbb{R}^N$, $n \in \mathbb{N}$, $g \in \mathcal{S}^*$ and $(\Theta, \alpha, \beta)$ satisfies Assumption 1’ and Assumption 2’ in Appendix A. Then

$$H(f, g, \alpha, \beta, v, n) = E_0 \left[ \exp \left( - \int_0^T R(X_u) du \right) (v \cdot X_T)^n e^{\alpha \cdot X_T} g(\beta \cdot X_T) \right] = \frac{1}{2\pi} \langle \hat{g}, G(\alpha + \cdot \beta i; v, n) \rangle$$

where $\hat{g} \in \mathcal{S}^*$ and $G(\alpha + \cdot \beta i; v, n)$ denotes the function

$$s \mapsto G(\alpha + s\beta i; v, n) = E_0 \left[ e^{-\int_0^T R(X_u) du} (v \cdot X_T)^n e^{(\alpha + is\beta) \cdot X_T} \right].$$

The function $G$ is computed by solving the associated ODE/BVP in Appendix A.

The assumption that the function $g$ in the generalized transform be a tempered distribution might appear restrictive at first sight, as $g$ is required to be rapidly decreasing (see our earlier
discussions of Schwartz functions). However, as Proposition 1 demonstrates, by specifying \( f \) and \( g \) appropriately, we can let \( f \) “absorb” any polynomial or exponential growth in a moment function, leaving \( g \) to be admissible to the transform. We will demonstrate this feature with examples in Section 4.

Next, at the expense of a multi-dimensional Fourier inversion, we can allow \( g \) to depend on \( X \) in more general ways. That is, instead of \( g(\beta \cdot X) \), we consider \( g(\beta_1 \cdot X, \cdots, \beta_M \cdot X) \) for \( M \in \mathbb{N} \).

**Proposition 2.** Suppose that \( f(s) = \exp(s) \), \( g \in \mathcal{S}_M^\ast \) (an \( M \)-dimensional tempered distribution), \( \alpha \in \mathbb{R}^N \), \( \beta \in \mathbb{R}^{M \times N} \) and \((\Theta, \alpha, \beta)\) satisfies Assumption 1 and Assumption 2 in Appendix A. Then

\[
H(f,g,\alpha,\beta) = E_0 \left[ \exp \left( -\int_0^T R(X_s) ds \right) f(\alpha \cdot X_T) g(\beta X_T) \right] = \left( \frac{1}{2\pi} \right)^M \langle \hat{g}, G_M(\alpha + \cdot \beta i) \rangle
\]

where \( \hat{g} \in \mathcal{S}_M \) and \( G_M(\alpha + \cdot \beta i) \) denotes the function

\[
G_M : \mathbb{C}^M \rightarrow \mathbb{C} \ , \ s \mapsto G_M(\alpha + s^\top \beta i) = E_0 \left[ e^{-\int_0^T R(X_u) du} e^{(\alpha + is^\top \beta) \cdot X_T} \right].
\]

This extension relaxes an important restriction for the nonlinear function \( g(\cdot) \) in Theorem 1, where \( g(\cdot) \) can only depend on \( X \) through the linear combination \( \beta \cdot X \). In that case, the marginal impact of \( X_i \) on \( g \) will be proportional to \( \beta_i \), which might be too restrictive in some cases. Proposition 2 removes this restriction, with the caveat that the problem is subject to the curse of dimensionality as \( M \) gets large.

Finally, it is also immediate to further extend the transform in Proposition 2 by replacing \( f(\alpha \cdot X) \) with a \textit{pl-linear} function as in Proposition 1.

### 3 Applications

#### 3.1 Option Pricing

We first show that the generalized transform analysis can recover the DPS result as a special case. We use option pricing as an example.

As shown in DPS, for pricing European options, we want to evaluate the transform:

\[
E_t^Q \left[ e^{-\int_t^T r_s ds + \alpha \cdot X_T} g_y(\beta \cdot X_T) \right],
\]
where \( g_y(x) = 1_{\{x \leq y\}} \) is nonlinear and non-integrable. For example, for an European put option with strike \( K \), \( X_t \) will be the log stock price, \( y = \log K \), and the option price is

\[
P_t = E_t^Q \left[ e^{-\int_t^T r_s ds + y g_y(X_T)} \right] - E_t^Q \left[ e^{-\int_t^T r_s ds + X_T g_y(X_T)} \right].
\]

(16)

However, the Fourier transform is defined as a distribution:

\[
\hat{g}_y(s) = \frac{1}{2} \delta(s) + \frac{e^{i\pi y s}}{2\pi i s}
\]

(17)

where \( \delta \), the dirac-\( \delta \) function, is the distribution defined by the relation

\[
\int \delta(x) h(x) dx = h(0)
\]

(18)

After applying symmetry, this replicates the formula given in DPS obtained by Levy-inversion.\(^5\) Our method applies more generally to any exotic whose payoff is an arbitrary function of the terminal stock value.

3.2 Nonlinear Taylor Rule

Through open market operations, the federal reserve targets the federal funds rate, the interest rate at which depository institutions lend balances at the Federal Reserve to other depository institutions overnight. The target is set by the Federal Open Market Committee which holds eight regularly scheduled meeting throughout the year as well as additional meetings whenever needed. The federal funds target rate can be viewed as a primitive in determining the yield curve. A common approach to modeling fed policy is in the form of a simple Taylor rule (Taylor (1993)),

\[
f_t = \beta_0 + \beta_1 \pi_t + \beta_2 g_t + \epsilon_t
\]

(19)

where \( f_t \) denotes the fed funds target rate at time \( t \), \( \pi_t \) denotes inflation, \( g_t \) represents a measure of the output gap, and \( \epsilon_t \) is the monetary policy shock. This gives a simple representation of the Fed’s goal of price stability and sustainable economic growth.

Several potential deficiencies with such a simple rule are apparent. The fed target rate nec-

\(^5\)In DPS, they arrive at this equation by effectively computing the forward density by Fourier transform (a 1-dimensional integral) and then integrating over the payoff region (now a 2-dimensional integral. In this case, Fubini and limiting arguments allow this 2-dimensional integral to reduced to a 1-dimensional integral as in the standard Lévy inversion formula.
nessarily must be non-negative. This directly induces a nonlinearity. Additionally, the fed must implicitly consider this lower bound for future policy in setting current policy. Also, the target is typical a multiple of 25 basis points. Finally, the policy rule may incorporate other variables such as credit conditions or forward-looking variables.

Piazzesi (2005) uses pure jump processes with deterministic jump times to model fed funds target $f_t$. Let $X_t$ be a vector summarizing economic conditions, which may contain inflation and growth measures, as in a Taylor rule, as well as possibly other macroeconomic variables such as unemployment. Piazzesi (2005) computes the prices of fixed income securities in this setting when the jumps in $f$ occur with stochastic intensity during FOMC meetings. In this specification, the distribution of jump size does not depend on the state variable. This implies that the moment-generating function, conditional on pre-meeting information, maintains an affine form

$$E_{t_0}[e^{a_f t}] = e^{A a + B a \cdot X_{t_0}}. \quad (20)$$

This assumption implies, among other things, that the expected policy response is linear in the state. In other words, in expectation a linear Taylor rule holds in the case that the state variables are the Taylor rule inputs, but there may be non-normal policy shock deviations from the linear rule. However, as elaborated below, our specification allows for a nonlinear policy response and thus inherently nonlinear expectations. We elaborate below on the empirical relevance of such nonlinearities.

Figure 1 plots the historical quarterly time series of the fed funds rate. Also plotted are the fed funds rate prescribed by two Taylor Rules:

$$r_t = 1\% + 1.5 \pi_t + .5 g_t + \epsilon_t \quad (21)$$

$$r_t = 1.19\% + 1.49 \pi_t + .09 g_t + \epsilon_t \quad (22)$$

The parameters in (21) correspond to the parameters in Taylor (1993) while the parameters in (22) are computed by an OLS regression. The federal funds rate is from Federal Reserve release H.15. The output gap is the real GDP output gap as computed by the Congressional Budget Office.

For both cases, the rule underestimates the response when actual rates are lowest (see Figure 2.) This suggests a nonlinearity in the form of more aggressive action in extreme situations. However, the natural lower limit of 0% also limits the policy action in extreme situations, introducing another nonlinearity. Thus it may be very natural to generate a simple nonlinear policy rule to model the
Fed’s policy action.  

The generalized transform technique allows us to consider more flexible nonlinear policy rules. Define a policy rule $G(\beta \cdot X_T)$ so that at a meeting at date $T$, the fed sets a target of $G(\beta \cdot X_T)$. The function $G$ can be chosen such that it generates movements in $f_t$ in a multiple of 25 basis points. Suppose the short rate is $r_t$, and the spread between the short rate and the target is $s_t = r_t - f_t$. Consider a federal funds futures contract which, for simplicity, we assume pays off $f_T$ at some future FOMC meeting time $T$. The price of such a contract is given by:

$$P_t = E_t^Q \left[ \exp \left( - \int_t^T r_s ds \right) f_T \right] = E_t^Q \left[ \exp \left( - \int_t^T r_s ds \right) G(\beta \cdot X_T) \right]. \quad (23)$$

This expectation is easily mapped to the generalized transform (6), with

$$f(\alpha \cdot X) = 1 \quad (24)$$

$$g(\beta \cdot X) = G(\beta \cdot X_T) \quad (25)$$

6Another approach would be to have a full economic model of optimal policy rather than a simple rule. As we see empirically, such a rule is likely to be nonlinear as well.
3.3 Default Contagion

Next, we consider a model of correlated default that violates the doubly stochastic assumption in that the default event of one firm affects the default probability of another firm beyond its own default intensity. This model is an extension of the primary-secondary framework for counterparty risk in Jarrow and Yu (2001).

Suppose two firms $A$, $B$ are from the same industry. Firm $A$ is a primary player in the industry, while $B$ is a secondary small firm. Let the default times of $A$ and $B$ be $\tau_A$ and $\tau_B$. Under the doubly stochastic assumption, conditional on the information in the driving filtration that determines the intensities $\lambda_A$ and $\lambda_B$, $\tau_A$ and $\tau_B$ are independent.

\[
\Pr (\tau_A > s | \mathcal{F}_t) = E_t \left[ \exp \left( - \int_t^s \lambda_A (X_u) \, du \right) \right]
\]

(26)

\[
\Pr (\tau_B > s | \mathcal{F}_t) = E_t \left[ \exp \left( - \int_t^s \lambda_B (X_u) \, du \right) \right]
\]

(27)

However, imagine that the event of the primary firm’s default brings major impacts on the rest of the firms in the industry, while the failure of the secondary firm has little effect on others. The impact of $A$ on $B$ can be due to direct business ties that are more important to $B$ than to $A$, a form of counterparty risk in Jarrow and Yu (2001), or it could be due to $A$’s default changes the
perception of risk in other firms, as in Collin-Dufresne et al. (2002). In this case, \( \tau_A \) remains solely dependent on \( \lambda_A \), while \( \tau_B \) depends on \( \lambda_B \) and the status of \( A \).

\[
\Pr(\tau_B < s|\mathcal{F}_t) = \Pr(\tau_B < s, \tau_A > \tau_B) + \Pr(\tau_B < s, \tau_A < \tau_B) \\
= E_t \left[ \int_t^s e^{-\int_u^s \mu(X_v) dv} \lambda_B(X_u) du \right] \\
+ E_t \left[ \int_t^s 1_{\{\tau_A > u, \tau_B > u\}} \lambda_A(X_u) \cdot \Pr(\tau_B < s|\mathcal{F}_{u^-}, \tau_A = u, \tau_B > u) du \right] \\
(28)
\]

where

\[
\mu(X) = \lambda_A(X) + \lambda_B(X).
\]

Let \( \lambda_{B,t}^+ \) be the default intensity of \( B \) after \( A \) defaults. Then, the probability that \( B \) defaults before \( s \) conditional on \( A \) defaulting at time \( u < s \) is

\[
\Pr(\tau_B < s|\mathcal{F}_{u^-}, \tau_A = u, \tau_B > u) = 1 - E_u \left[ \exp \left( - \int_u^s \lambda_{B,t}^+(X_v) dv \right) \right].
\]

Jarrow and Yu (2001) consider the following specification:

\[
\lambda_{B,t}^+ = \lambda_{B,t} + c
\]

(29)

The restriction is that \( A \)'s default permanently increases the default intensity of \( B \) by a constant amount, which can be relaxed in several ways. For example, we can assume that

\[
\lambda_{B,t}^+ = \lambda_{B,t} + c(X_t)
\]

(30)

for some \( c \) which is affine in \( X \). In this case, the second expectation in (28) can be computed in closed form (up to a single integral) using the DPS transform.

Alternatively, we can directly assume that

\[
\Pr(\tau_B < s|\mathcal{F}_{u^-}, \tau_A = u, \tau_B > u) = g(\beta(s - u) \cdot X_u, s - u).
\]

(31)

where \( g \) satisfies the regularity conditions in Theorem 1 plus \( \frac{\partial g}{\partial s} > 0 \). This specification is attractive when one wants to capture particular empirical patterns of default rates over different horizons (e.g., at the different phases of firm \( A \)'s bankruptcy process), or the dissipation of the initial impact from \( A \) to \( B \) over time, which could be infeasible under (30). It could also provide a reduced-form solution
to the challenges in modeling the nonlinear dependence of default intensity on state variables found in Duffie et al. (2007), which can become more prominent in a period of market distress following a large firm’s default.

One possible choice for \( g \) is a logistic function, as used by Campbell et al. (2008):

\[
g(X, t) = \frac{1}{1 + \exp(-\alpha(t) - \beta(t) \cdot X)},
\]

where \( \alpha(t) \) and \( \beta(t) \) satisfy

\[
\alpha'(t) + \beta'(t) \cdot X > 0.
\]

The evaluation of (28) given (31) depends on whether \( X \) jumps at the default event of firm A, which corresponds to the “no-jump” condition as discussed by Duffie et al. (1996). If \( X \) does not jump, the second term in (28) is equal to:

\[
\Pr(\tau_B < s, \tau_A < \tau_B) = V_t = E_t \left[ \int_t^s e^{-\int_u^s \mu(X_v) dv} \lambda_A(X_u) g(\alpha(s - u), \beta(s - u) \cdot X_u) du \right],
\]

for which the generalized transform readily applies. If \( X \) jumps at \( \tau_A \), then a correction term is needed for (33) (see Proposition 1 of Duffie et al. (1996)):

\[
\Pr(\tau_B < s, \tau_A < \tau_B) = V_t - E_t [\Delta V_{\tau_A}],
\]

which again can computed with the generalized transform.

This simple model of contagion can also be used to capture jumps in default intensities and default correlation across firms following major events, with \( \lambda_A \) being the arrival intensity of such events.

### 3.4 GMM Estimation

The need to compute unconditional and conditional moments of nonlinear functions often arises in the method of moments and GMM estimations. Consider an econometric model given by

\[
u_t = y_t - h(X_t, w_t; \theta),
\]

where \( \theta \) is a vector of unknown parameters, \( X_t \) is a vector of latent state variables that follows a stationary affine process, \( y_t \) and \( w_t \) are variables observed at time \( t \). For example, \( h(X_t, w_t; \theta) \)
can be the price of a security at time $t$ given latent state variable $X_t$, observable state variable $w_t$, and parameters $\theta$, and $y_t$ can be the observed price of this security in the market. The moment condition is

$$E[u_t z_t] = 0$$  \hspace{1cm} (36)$$

for any valid instrument $z_t$.

A method of moments estimator $\hat{\theta}_{MM}$ is the solution to

$$\frac{1}{T} \sum_{t=1}^{T} z_t \left[ y_t - E \left[ h(X_t, w_t; \hat{\theta}_{MM}) \right] \right] = 0.$$  \hspace{1cm} (37)$$

where we have taken the unconditional expectation with respect to the latent state variable $X_t$. Assuming that $h(\cdot)$ can be decomposed into the product of $f$ and $g$ satisfying the regularity conditions in Theorem 1, then we can compute the expectation using the generalized transform. Similarly, a generalized method of moments estimator $\hat{\theta}_{GM}$ minimizes the objective function

$$Q(\theta; Y_T) = \left[ \frac{1}{T} \sum_{t=1}^{T} z_t \left[ y_t - E \left[ h(X_t, w_t; \theta) \right] \right] \right]' W_T \left[ \frac{1}{T} \sum_{t=1}^{T} z_t \left[ y_t - E \left[ h(X_t, w_t; \theta) \right] \right] \right].$$  \hspace{1cm} (38)$$

4 Three Concrete Examples

In this section, we provide in-depth analysis of three examples to illustrate the application of the generalized transform analysis.

4.1 Recovery Risk

The value of a credit-risky security (e.g., defaultable bonds or credit default swaps) depends on the discount rate, default probability, and the recovery value of the security in the event of default. Recovery risk refers to the uncertainty about the recovery rate. Due to the great amount of difficulty in forecasting the recovery rate far ahead of the default event, academics and practitioners have often treated recovery risk as a secondary consideration. We introduce a new class of stochastic recovery models and study the impact of state-dependent recovery risks on pricing.

We fix a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\{\mathcal{F}_t : t \geq 0\}$. Following the reduced-form models, the default time is assumed to be a totally inaccessible $\mathcal{F}$-stopping time $\tau : \Omega \rightarrow (0, +\infty]$. For simplicity, we assume that under the risk neutral measure $Q$, $\tau$ is doubly-stochastic driven by
the filtration \( \{ \mathcal{F}_t : t \geq 0 \} \), with intensity \( \lambda_t \). The instantaneous riskfree rate is \( r_t \).

Consider a \( T \) year defaultable zero-coupon bond with face value of 1, and the recovery value at default \( \varphi_t \) is a bounded predictable process that is adapted to the filtration \( \{ \mathcal{F}_t : t \geq 0 \} \). The price of the bond is:

\[
V_0 = E^Q \left[ e^{-\int_0^T R_u du} 1_{\{ \tau \leq T \}} \varphi_\tau \right] + E^Q \left[ e^{-\int_0^T R_u du} 1_{\{ \tau > T \}} \right] \\
= E^Q \left[ \int_0^T e^{-\int_0^u (R_v + \lambda_v) dv} \lambda_t \varphi_t dt + e^{-\int_0^T (R_u + \lambda_u) du} \right].
\]

(39)

The second equality follows from the doubly-stochastic assumption and certain regularity conditions (see Duffie (2005) for details).

Duffie and Singleton (1999) discuss three types of recovery models:

1. “recovery of treasury” (RT):
   \[
   \varphi_t = (1 - L_t) P_t,
   \]
   where \( P_t \) is the price at time \( t \) of an otherwise equivalent default-free bond; \( L_t \) is a value between 0 and 1.

2. “recovery of face value” (RFV):
   \[
   \varphi_t = (1 - L_t) F,
   \]
   where \( F \) is the face value of the bond.

3. “recovery of market value” (RMV):
   \[
   \varphi_t = (1 - L_t) V_{t-},
   \]
   where \( V_{t-} \) is the market value of the security immediately before default.

Duffie and Singleton (1999) show that under the RMV specification and a suitable no-jump condition,\(^8\) one can price defaultable claims with the default-adjusted discount rate, \( r_t + \lambda_t L_t \). Moreover, if one directly specifies the mean-loss rate \( \lambda_t L_t \) as affine, the standard results for affine term structure models apply. In contrast, the RT and RFV models are generally less tractable.

\(^7\)See Duffie (2005) for a survey on the reduced form approach for modeling credit risk and the doubly-stochastic property.

\(^8\)See also Duffie et al. (1996) and Collin-Dufresne et al. (2004) for discussions on the no-jump condition.
While analytically appealing, the RMV assumption has some limitations. First, since \( \lambda_t \) and \( L_t \) enter the default-adjusted discount rate symmetrically, we cannot separately identify the effect of default intensity and recovery using information on prices alone. Second, when pricing bonds of different seniorities from the same issuer, it is more natural to separately model default intensity \( \lambda \) (same across different bonds) and recovery rates \( L \) (depending on seniority). Finally, data on recovery rates are usually quoted as fraction of face value instead of market value. For example, Moody’s database of corporate defaults estimates defaulted debt recovery rates using the ratio of market bid prices observed roughly 30 days after the date of default to par value.

Bakshi et al. (2006) study a class of RT and RFV models for which \( \varphi \) is exponential affine in the default intensity as well as the class of completely monotone functions. They solve for bond prices using the DPS transform. We show that a much wider range of RFV models becomes tractable using the generalized transform analysis.\(^9\) The added flexibility in modeling recovery rates allows us to introduce explicit dependence of recovery rates on macro and firm-specific variables.

### 4.1.1 Model Setup

Next, we consider a concrete example. We directly specify the dynamics of state variables \( X_t = [\lambda_t, Y_t]' \) under the risk neutral probability measure \( Q \):

\[
\begin{align*}
d\lambda_t &= \kappa_{\lambda}(\theta_{\lambda} - \lambda_t)dt + \sigma_{\lambda}\sqrt{\lambda_t}dW^\lambda_t & (43) \\
dY_t &= \kappa_Y(\theta_Y - Y_t)dt + \sigma_Y\sqrt{\lambda_t}dW^Y_t & (44)
\end{align*}
\]

where \( W^\lambda_t \) and \( W^Y_t \) are uncorrelated Brownian motions; \( \lambda_t \) is the default intensity of a firm, and the short term interest rate, \( r_t \), is given by

\[
r_t = Y_t - \delta \lambda_t. \quad (45)
\]

This setup (with \( \delta > 0 \)) captures the negative correlation between \( r_t \) and \( \lambda_t \) as documented by Duffee (1998).

The recovery value \( \varphi \) of a bond issued by the firm can depend on the default intensity, short rate, and other macro and firm-specific variables. For example, Altman et al. (2005) document significant

---

\(^9\)The completely monotone class requires that each order derivative is either always positive or negative. This precludes, for example, an inflection point as in the Cauchy example we consider. Importantly, the Fourier transform provides for direct inversion formulas not available in the completely monotonic case.
negative correlation between aggregate default rates and recovery rates.\footnote{See Altman et al. (2005) for a review of earlier studies on the relationship between recovery rates and default rates.} Chen (2008) shows that macro variables such as GDP growth and riskfree rate are correlated with the aggregate recovery rates and default rates. Carey and Gordy (2007) show that firm-level recovery rate increases with the share of bank loans in total debt.

Another important property for the recovery function $\varphi$ is that it should only take values from $[0, 1]$. One specification for $\varphi$ that satisfies this requirement and still works with the DPS formulation is $\varphi(X) = e^{\beta X} 1_{\{\beta X < 0\}} + 1_{\{\beta X > 0\}}$. Bakshi et al. (2006) study such a setting. More generally, the cumulative distribution function of any distribution will have this property. Common examples include:

- Logit Model:
  $$
  \varphi(X) = \frac{1}{1 + e^{-\beta_0 - \beta_1 X}}
  $$

- Probit Model:
  $$
  \varphi(X) = \int_{-\infty}^{\beta_0 + \beta_1 X} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds.
  $$

- Cauchy Model:
  $$
  \varphi(X) = \int_{-\infty}^{\beta_0 + \beta_1 X} \frac{\gamma}{\pi((s^2 - s_0^2) + \gamma^2)} ds.
  $$

Modeling $\varphi$ with CDFs has the additional benefit that they have nice Fourier transform properties. For example, the integrands of the Probit and Cauchy model have closed-form Fourier transform. Since Fourier transform has the property that $\hat{f}'(t) = t \hat{f}(t)$, it is very easy to obtain the Fourier transform of $\varphi$ in those cases.

For simplicity, we assume that $\varphi$ only depends on the default intensity, and we adopt a small variation of the Cauchy model:

$$
\varphi(\lambda) = \frac{a}{1 + b(\lambda - \lambda_0)^2} + c. \tag{46}
$$

Its Fourier transform (excluding the constant $c$) is:

$$
\hat{\varphi}(t) = \frac{a\pi}{\sqrt{b}} e^{\lambda_0 it - \frac{1}{\sqrt{b}} |t|}. \tag{47}
$$

We consider two calibrations of $\varphi(\lambda)$ in (46). First, using data from Altman et al. (2005), we calibrate $a = 0.5$, $b = 1415$, $c = 0.25$, and $\lambda_0 = -0.01$. The fitted function is “model I” in Figure 10.
3. The fitted curve is downward sloping and convex. The recovery rate is close to 70% when the probability of default is very low. When annual default probability rises to 10%, the recovery rate drops to 30%. The parametrization of model I is likely too conservative: the recovery rate is bounded from below at 25%, and it treats the recovery rates in the data as the risk-neutral recovery rates. Thus, we also study a second calibration, where $a = 0.8$, $b = 800$, $c = 0$, and $\lambda_0 = -0.08$. The fitted function is “model II” in Figure 3, which has similar recovery rates to “model I” when default intensity is low, but has a sharper decline in recovery rates than “model I” when default intensity rises.

Several features of the recovery curve will matter for bond pricing: how fast (slope) and how far (right tail) the recovery rate drops with default rate, and how much curvature the recovery function has. We will investigate how each of these features affects pricing.

The key step in computing the value of the defaultable zero-coupon bond is to compute the expectation

$$E_0^Q \left[ \exp \left( -\int_0^t (r_u + \lambda_u)du \right) \lambda_t \varphi (\lambda_t) \right],$$

Figure 3: A Cauchy Model of Aggregate Recovery Rates. This figure plots the aggregate recovery rates and default rates from Altman et al. (2005). The blue line fits a Cauchy curve through the data. The functional form of the curve is given by equation (46).
Table 1: Model Calibration

<table>
<thead>
<tr>
<th>$\kappa_\lambda$</th>
<th>$\theta_\lambda$</th>
<th>$\sigma_\lambda$</th>
<th>$\kappa_Y$</th>
<th>$\theta_Y$</th>
<th>$\sigma_Y$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.06</td>
<td>0.01</td>
<td>0.02</td>
<td>0.12</td>
<td>0.015</td>
<td>0.1</td>
</tr>
</tbody>
</table>

which is mapped into the generalized transform (6) by choosing:

\[
 f(\alpha \cdot X) = \xi_1 \cdot X \\
 g(\beta \cdot X) = \frac{a}{1 + b(\xi_1 \cdot X - \lambda_0)^2}
\]

where $\xi_1 = [1 \ 0]^t$.

It is straightforward to use the recovery model $\varphi(X)$ to price other credit products, such as credit default swaps or recovery locks. In addition, our model can be generalized to allow for violations of the doubly-stochastic assumption or no-jump conditions.\(^\text{11}\) Thus, it can be used in models with flight-to-quality, default contagion, systematic jump risk, or other features that violate the no-jump condition.

4.1.2 Results

We use the dynamics of default intensity $\lambda$ and riskfree rate $r$ implied by (43-45) and the recovery model (46) to price 5-year defaultable zero-coupon bonds. The parameter values for the $\lambda$ and $Y$ process are reported in Table 1.

The results are reported in Figures 4. Panel A plots the credit spreads (the yield spreads between the defaultable bond and a default-free bond with the same maturity) as functions of the default intensity. We report the results for (1) the stochastic recovery model I; (2) constant recovery with $\varphi$ set to the unconditional mean recovery rate implied by model I, which is 31%; (3) constant recovery with $\varphi = 25\%$, a popular assumption for the recovery rate in both academic analysis and industry practice (see e.g., Pan and Singleton (2008)).

Naturally, credit spreads increase with default intensity in all cases. When current default intensity is low, the stochastic recovery model implies higher-than-average recovery rates, thus generating lower credit spreads than the constant recovery models. As default intensity rises, the spreads from the stochastic recovery model become higher, and exceeds the spreads from the

\(^{11}\)We can either explicitly make the correction for jumps as in Duffie et al. (1996), or use the change-of-measure method in Collin-Dufresne et al. (2004).
constant recovery model where $\varphi$ is fixed at the Model I-implied average (31%). Since the lower bound for the stochastic recovery model model is 25%, the spreads of the stochastic recovery model is always below that of the constant recovery model with $\varphi = 25\%$.

Panel B reports the fraction of the spreads due to recovery risk, defined as

$$\frac{\text{stochastic recovery yield} - \text{constant recovery yield}}{\text{stochastic recovery yield} - \text{default-free yield}},$$

which can be interpreted as a measure of the relative pricing errors due to ignoring the state-dependence of recovery rates. Relative to the stochastic recovery model I, the main concern of the constant recovery assumption is underpricing, i.e. they generate credit spreads that are too high. The pricing errors can be large. When default intensity is low ($< 2\%$), the pricing errors are close
Figure 5: Zero coupon yields from generalized transform vs. Monte Carlo method. This figure plots zero coupon yields for 5-year bonds with stochastic recovery and constant recovery, compared to prices obtained using Monte Carlo method.

to 40% for the assumption of 25% recovery rate.

In Panel C and D of Figures 4, we compare the spreads of the stochastic recovery model II with the constant recovery models. Again, when the default intensity is low, credit spreads are lower for the stochastic recovery model, because the conditional recovery rates are higher than under constant recovery assumption. As the default intensity increases, the low recovery rates at times of high default probability start to have more and more significant effects on the bond prices. The spreads rise more rapidly in model II than in model I (Panel A) due to the more rapid decline in recovery rates, and they exceed the spreads from the constant recovery models quickly.

The relative pricing errors in Panel D are quite large when default intensity is low or high. For example, under the 25% constant recovery rate assumption, the spreads are 30%+ higher than the stochastic recovery model for $\lambda < 1\%$, and 18%+ lower than the stochastic recovery model for $\lambda > 6\%$. Another important message of the graph is that simply adjusting the constant recovery rate does not solve the problem, as it amounts to (approximately) parallel-shifting the pricing errors, and will either exacerbate the underpricing for low $\lambda$ or overpricing for high $\lambda$. These results suggest that the negative correlation between default intensity and recovery rate can have
important impact for pricing high grade bonds as well as high yield bonds.

Finally, we check the accuracy of the solutions via simulation. The results are reported in Figure 5. The yields computed from the Monte Carlo method (using 50,000 simulations) are consistent with the spreads computed using the generalized transform. The remaining differences are due to the numerical errors of the simulations.

4.2 Two Non-IID trees

In this section we illustrate how to use the generalized transform to compute the equilibrium of asset pricing models. Cochrane et al. (2008) show that in a Lucas economy (Lucas (1978)) with two trees, the equilibrium conditions imply rich dynamics for stock returns and volatility in the time series and cross section. They provide closed-form solutions in the case of log utility and i.i.d. trees. Martin (2008) extends the analysis to multiple trees and power utility, but also assume that dividend growth of each tree is i.i.d. We show that the model can be extended to allow for mean reversion, conditional heteroscedasticity, and jumps with time-varying intensity in dividend growth.

4.2.1 The Model

Following Cochrane et al. (2008) and Martin (2008), we consider an endowment economy with two stocks (trees). The infinitely-lived representative investor has CRRA utility:

$$U_t = E_t \left[ \int_0^\infty e^{-\rho u} C_{t+1}^{1-\gamma} - 1 \frac{1}{1-\gamma} du \right], \quad (50)$$

where we focus on the case \( \gamma > 1 \).

There are two stocks with dividend streams \( D_{1,t} dt \) and \( D_{2,t} dt \). Our model deviates from Cochrane et al. (2008) and Martin (2008) in that the dividend growth is non-i.i.d. The log dividends \( d_{1,t} = \log D_{1,t} \) and \( d_{2,t} = \log D_{2,t} \) follow the processes:

$$dd_{i,t} = g_{i,t} dt + \sigma_{d,i} dW^d_{i,t} \quad (51)$$

$$dg_{i,t} = \kappa_i (\overline{g}_i - g_{i,t}) dt + \sigma_{g,i} dW^g_{i,t} \quad (52)$$

where \( g_{i,t} \) is the expected growth rate for \( d_{i,t} \), which follows an Ornstein-Uhlenbeck process with long term mean \( \overline{g}_i \). For simplicity, we assume that all the Brownian motions \( W^d_{i,t} \) and \( W^g_{i,t} \) are
uncorrelated with each other. When $\gamma = 1$ and $g_{i,t} \equiv \overline{g}_i$, we recover the model of Cochrane et al. (2008). When $\gamma > 1$ and $g_{i,t} \equiv \overline{g}_i$, we recover the two-tree model (without jumps) of Martin (2008).

In equilibrium, aggregate consumption $C_t = D_{1,t} + D_{2,t}$. Define the dividend share for the first stock as $S_t = D_{1t}/(D_{1t} + D_{2t})$. As in the Lucas economy with a single asset, the instantaneous riskfree rate in this economy is determined by the rate of time preference, the expected growth rate of consumption, and precautionary savings driven by consumption volatility:

$$r_{f,t} = \rho + \gamma \left( S_t(g_{1,t} + \sigma^2_{d,1}/2) + (1-S_t)(g_{2,t} + \sigma^2_{d,2}/2) \right) - \frac{\gamma(\gamma + 1)}{2} \left( S_t^2 \sigma^2_{d,1} + (1-S_t)^2 \sigma^2_{d,2} \right). \quad (53)$$

Under the standard regularity conditions, the price of stock $i$ $(i = 1, 2)$, $P_{i,t}$, is given by:

$$P_{i,t} = E_t \left[ \int_0^\infty e^{-\rho u} \frac{u c(C_{t+u})}{u c(C_t)} D_{i,t+u} du \right] = E_t \left[ \int_0^\infty e^{-\rho u} \frac{(D_{1,t} + D_{2,t})^\gamma D_{i,t+u}^\gamma}{(D_{1,t+u} + D_{2,t+u})^\gamma} du \right] = (D_{1,t} + D_{2,t})^\gamma \int_0^\infty e^{-\rho u} E_t \left[ \frac{D_{i,t+u}}{(D_{1,t+u} + D_{2,t+u})^\gamma} \right] du. \quad (54)$$

To map the expectation in (54) to the generalized transform formula, we first define the state variable $X_t = [d_{1t} \ g_{1t} \ d_{2t} \ g_{2t}]'$. The model (51)-(52) implies that $X$ follows an affine diffusion process,

$$dX_t = (K_0 + K_1 X_t) dt + \sqrt{H_0} dW_t, \quad (55)$$

where

$$K_0 = \begin{pmatrix} 0 \\ \kappa_1 g_1 \\ 0 \\ \kappa_2 g_2 \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & -\kappa_1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\kappa_2 \end{pmatrix}, \quad H_0 = \begin{pmatrix} \sigma^2_{d,1} & 0 & 0 & 0 \\ 0 & \sigma^2_{g,1} & 0 & 0 \\ 0 & 0 & \sigma^2_{d,2} & 0 \\ 0 & 0 & 0 & \sigma^2_{g,2} \end{pmatrix}$$

$$W_t = \begin{pmatrix} W_{1,t}^d \\ W_{1,t}^g \\ W_{2,t}^d \\ W_{2,t}^g \end{pmatrix}'$$

Next, from (54),

$$E_t \left[ \frac{D_{1,s}}{(D_{1,s} + D_{2,s})^\gamma} \right] = E_t \left[ \frac{e^{(1-\gamma/2)d_{1,s} - \gamma/2d_{2,s}}}{2 \cosh \frac{d_{1,s} - d_{2,s}}{2}} \right] = E_t \left[ f(\alpha \cdot X_s) g(\beta \cdot X_s) \right], \quad (56)$$

25
where

\[ f(x) = e^x \]  \hspace{1cm} (57)  \\
\[ g(x) = \frac{1}{(2 \cosh(x))^\gamma} \]  \hspace{1cm} (58)

and

\[ \alpha = \begin{bmatrix} (1 - \frac{\gamma}{2}) & 0 & -\frac{\gamma}{2} & 0 \end{bmatrix}^T, \quad \beta = \begin{bmatrix} 1/2 & 0 & -1/2 & 0 \end{bmatrix}^T. \]

Up to this point, our procedure is the same as in Martin (2008). When the increments of \( X \) are \( i.i.d. \), the conditional characteristic function for \( X \) is known explicitly, which Martin uses to compute (56) following a Fourier transform for \( g \). We generalize his method to the case where the increments of \( X \) are \( non-i.i.d. \) by exploiting the properties of the conditional characteristic function for general affine processes.

This model can be further extended in several dimensions. First, while here we only consider time-varying expected growth rates, \( X \) can be any affine process, which allows us to introduce stochastic volatility and jumps with time-varying intensity in consumption growth. One attractive feature of our approach is that these new elements do not increase the dimension of the Fourier transform, which maintains the tractability of the model. Second, we can generalize the utility function, e.g., by making aggregate consumption \( C_t \) a CES aggregator of \( D_{1,t} \) and \( D_{2,t} \), as in Piazzesi et al. (2007), where the two trees are interpreted as nonhousing consumption and housing services. We can also impose cointegration between the two dividend processes and allow for stochastic volatility, as do Piazzesi et al. (2007). Third, it is also convenient to add preference shocks that are \( pl-linear \) in the state variable. Finally, the model can allow for multiple trees using the multi-dimensional version of the generalized transform in Proposition 2.

4.2.2 Price-Dividend Ratios and Expected Excess Returns

To illustrate the quantitative implications of the model, we choose the following parameters. For preferences, \( \rho = 0.1 \) and \( \gamma = 10 \). For the two dividend processes, we assume that the first stock initially has smaller dividend than the second stock, \( D_{1,t} = 1/5 D_{2,t} \), but has higher growth rate in the long run, \( \bar{g}_1 = 0.03 > \bar{g}_2 = 0.01 \). The other parameters are the same for the two stocks: the speed of mean reversion for the growth rate \( \kappa_1 = \kappa_2 = 0.3 \), volatility of dividend \( \sigma_{1d} = \sigma_{2d} = 0.1 \), and volatility of dividend growth \( \sigma_{1g} = \sigma_{2g} = 0.01 \). A possible interpretation for this parametrization is that a young, fast-growing industry is slowly taking over the economy from an established,
Figure 6: **Price-dividend ratios and expected excess returns.** This figure plots the price-dividend ratios and expected excess returns of the two stocks \(\frac{D_1}{D_2} = 1/5\) as a function of their own dividend growth rate and the other stock’s growth rate.

The top two panels of Figure 6 plot the price-dividend ratios for the two stocks against the conditional expected growth rates of the two stocks. The price-dividend ratio the first stock is higher than the second, and it is decreasing in the expected growth rate of both stocks, \(g_{1,t}\) and \(g_{2,t}\), although the decline is faster with \(g_{2,t}\). A rise in the expected growth rate has two effects on the price-dividend ratio of the stock. There is a *cash-flow effect*: higher expected growth rate implies higher future cash flows, which tends to increase the price-dividend ratio. There is also a *discount rate effect*: a rise in the expected growth rate will increase the discount rate for stocks due to the “substitution effects”: when the expected future consumption is high, consumers want to borrow to increase consumption today, which raises the interest rate (see equation (53)). With the CRRA utility and the given parameters, the discount rate effect dominates, causing price-dividend ratio to decrease with the growth rates.
Due to the large size of the second stock, an increase in $g_{2,t}$ has a strong effect on the aggregate discount rate, which causes the price of the first stock to drop. This effect is stronger when $g_{1,t}$ is small. In contrast, changes in $g_{1,t}$ have a small effect on the discount rates, but it has the additional cash flow effect on the small stock. As the share of the first stock gets smaller in this economy, the cash flow effect eventually dominates, resulting in the price-dividend ratio $P_{1,t}/D_{1,t}$ increasing in $g_{1,t}$. As for the large stock, when $g_{2,t}$ is large, changes in $g_{1,t}$ have almost no effect on the price-dividend ratio $P_{2,t}/D_{2,t}$. But when $g_{2,t}$ is small, the effect of $g_{1,t}$ on the price of the larger stock can become sizable.

In order to analyze the expected excess returns, we consider a stock as a portfolio of zero-coupon equities. The risk premium of the stock is then the value-weighted average of the risk premium for these zero-coupon equities, which are easy to compute. In general, the instantaneous expected excess return for any asset is determined by its exposure to the primitive risk sources, $X_t = (d_{1t}, g_{1t}, d_{2t}, g_{2t})$, and the risk premia that these risks demand through their covariance with the pricing kernel. In this case, the pricing kernel, $M_t$, takes the form $M_t = e^{-\rho t}(D_{1,t} + D_{2,t})^{-\gamma}$. By Itô’s Lemma, the expected excess returns for any asset with state-dependent price $P_t = P(X_t, t)$ are then given by

$$ER_t = \frac{1}{P} \frac{E^P[dP]}{dt} - r_t$$

$$= -\frac{1}{P} \nabla_X P(X_t) \cdot H_0 \nabla_X M(X, t) / M(X, t)$$

$$= \frac{1}{P} \frac{\partial P}{\partial d_{1}} \gamma S_t \sigma_{d,1}^2 + \frac{1}{P} \frac{\partial P}{\partial d_{2}} \gamma (1 - S_t) \sigma_{d,2}^2,$$

(59)

where $M(X, t) = e^{-\rho t}(e^{X_1} + e^{X_3})^{-\gamma}$ is the stochastic discount factor and $S_t = D_{1t}/(D_{1t} + D_{2t})$ is the dividend share. Since the innovations in dividends and growth rates are uncorrelated, there is no premium for growth risk in this model. Notice that the risk premium of a stock does not necessarily go to zero when its share approaches 0. Even though the stock’s own dividend shocks become uncorrelated with the pricing kernel as its share drops to 0, the stock can still be exposed to discount rate rate risks. For example, a shock to the second stock’s dividend will change the share and the diversification in the economy, which affect the discount rates, and in turn, the price of the first stock.

Since the price of zero coupon equity is available in closed form through the transform analysis, its gradient can also be computed in closed form. Then we can compute the risk premium for zero coupon equities at different maturities using (59), and compute the risk premium for the stock
as a value-weighted average of the premia for zero coupon equities. The results are plotted in the bottom panels of Figure 6. Because the second stock is larger and thus more correlated with the pricing kernel, it demands higher risk premium on average. Interestingly, the conditional risk premia of both stocks are decreasing in the growth rate of the larger stock \( g_{2,t} \), but increasing in the growth rate of the smaller stock \( g_{1,t} \).

Equation (59) implies that the expected growth rates of dividend only affect the risk premia by changing the sensitivity of the log stock price to dividend shocks. A positive shock to \( d_{1,t} \) increases expected future dividends for the first stock, which tends to raise its stock price, and moves the share \( S_t \) closer to 0.5, which increases the interest rate and lowers the stock price. A positive shock to \( d_{2,t} \) has no impact on future dividends of the first stock, but moves the share \( S_t \) closer to 0, which decreases the interest rate and raises the stock prices. The net risk premium of the smaller stock is more affected by the exposure to the dividend shock of the larger stock, resulting in positive risk premium. A higher \( g_{1,t} \) will amplify the effects of both dividend shocks on the interest rates, but since the second dividend shock bears a higher premium, the net risk premium increases in \( g_{1,t} \). A higher \( g_{2,t} \) reduces the effects of both dividend shocks on the interest rates and prices, which causes the risk premium to drop.

### 4.3 Differences of Opinions

Models of heterogeneity of beliefs, or equivalently of preferences, can generate rich implications for trade and affect asset prices in equilibrium (see Basak (2005) for a recent survey). In studying such economies, aggregation often leads to difficulty in computing equilibrium outcomes. In this example, we illustrate the use of our main result in solving economies where there is heterogeneity among agents regarding beliefs (and higher order beliefs) about fundamentals.

#### 4.3.1 General Setup

Suppose there are two agents (A, B) who possess heterogeneous beliefs. There is a state variable \( X_t \) which Agent A believes follows an affine jump diffusion:

\[
    dX_t = \mu^A_t \, dt + \sigma^A_t \, dW^A_t + dZ^A_t, 
\]

where \( \mu^A_t = K^A_0 + K^A_1 X_t \), \( \sigma^A_t (\sigma^A_t)^\top = H^A_0 + H^A_1 \cdot X_t \), and jumps are believed to arrive with intensity \( \lambda^A_t = \lambda^A_0 + \lambda^A_1 \cdot X_t \) and have distribution \( \nu^A \) (with moment generating function \( \phi^A \)). As elaborated
in the examples below, the variable $X_t$ encompass all uncertainty in the economy, including any
time-variation in the heterogeneity of beliefs. For simplicity, we suppose that Agent A’s beliefs are
correct. The method is easily modified to the case where neither agent is correct.

Agent B has heterogeneous beliefs which we shall suppose are equivalent. A broad class\(^{12}\) of
such equivalent beliefs can be characterized as follows. There exists some vector $a$ such that Agent
B believes $X_t$ follows an affine jump diffusion satisfying

$$dX_t = \mu^B_t dt + \sigma^B_t dW^B_t + dZ^B_t,$$

(61)

where

1. $\mu^B_t = \mu^A_t + \sigma^A_t (\sigma^A_t)^T a$
2. $\sigma^B_t = \sigma^A_t$
3. $d\nu^B/d\nu^A(Z) = e^{a \cdot Z}/E_{\nu^A}[e^{a \cdot Z}]$ or $\phi^B(c) = \phi^A(c + a)/\phi^A(a)$
4. $\lambda^B_t = \lambda^A_t \times E_{\nu^A}[e^{a \cdot Z}]$

This difference in beliefs generates a disagreement about not only the drifts of the state variables,
but also the jump frequency and the distribution of jump size.

This structure implies that the two beliefs define equivalent probability measures which may be
related through the Radon-Nikodym derivative $dP^B/dP^A$:

$$\eta_t = E_t \left[ \frac{dP^B}{dP^A} \right] = \exp \left( a \cdot X_t - \int_0^t \left( a \cdot \mu^A_s + \frac{1}{2} \|\sigma^A_s a\|^2 + \lambda^A_s (\phi^A(a) - 1) \right) ds \right).$$

(62)

The variable $\eta_t$ expresses Agents B’s differences in opinion in that when $\eta_t$ is high, Agent B believes
an event is more likely than Agent A believes. We refer to $\eta_t$ as the \textit{db-density} (‘db’ stands for
“difference in beliefs”) process, which differs from the density defining the risk-neutral measure.

Notice that the integral term in the exponent above follows an affine process. Thus, by redefining
$X$ to include the integral term and augmenting $a$ accordingly, we have

$$\eta_t = e^{a \cdot X_t}.$$ 

(63)

\(^{12}\)More generally, we could consider beliefs of the form $e^{h(x_t) - \int_0^t e^{-h(x_s)} dP^1 e^{h(x_s)} ds}$. Provided the integral term
remains tractable, the same analysis applies. Compare also the discussion of essentially affine difference of opinions.
We assume that the agents have time separable preferences:

\[ U^i(c) = E^0_0 \left[ \int_0^\infty u^i(c_t, t) dt \right], \quad i = A, B \]  

(64)

Suppose also that

1. markets are complete;
2. log of aggregate consumption, \( c_t = \log(C_t) \), is linear in \( X_t \) (\( c_t = c \cdot X_t \));
3. agents are endowed with some fixed fraction (\( \theta_A, \theta_B = 1 - \theta_A \)) of aggregate consumption.

Let \( \xi_t \) denote the stochastic discount factor with respect to Agent A’s beliefs. As in Cox and Huang (1989), we impose the lifetime budget constraint and equate state prices to marginal utilities to solve

\[
\begin{align*}
    u^A_c(C^A_t, t) &= \zeta^A \xi_t \\
    u^B_c(C^B_t, t) &= \zeta^B \eta_t^{-1} \xi_t
\end{align*}
\]  

(65)  

(66)

where \( C^i_t \) is Agent \( i \)’s equilibrium consumption at time \( t \), \( \zeta^i \) is the Lagrange multiplier for Agent \( i \)’s budget constraint.

Market clearing then implies

\[ C_t = (u^A_c)^{-1}(\zeta^A \xi_t) + (u^B_c)^{-1}(\zeta^B \eta_t^{-1} \xi_t), \]  

(67)

which implies \( \xi_t = h(c_t, \eta_t) \) for some \( h \). With the additional assumption that \( u^i(c, t) = e^{-\rho t} c^{1-\gamma}/(1-\gamma) \), this simplifies to

\[ \xi_t = e^{-\rho t} \left[ \left( \frac{1}{\zeta^A} \right)^{1/\gamma} + \left( \frac{\eta_t}{\zeta^B} \right)^{1/\gamma} \right]^{\gamma} C_t^{-\gamma}. \]  

(68)

Using \( g(x) = \left[ \left( \frac{1}{\zeta^A} \right)^{1/\gamma} + \left( \frac{\eta_t}{\zeta^B} \right)^{1/\gamma} \right]^{\gamma} \) and \( C_t = e^{c \cdot X_t} \), we finally have

\[ \xi_t = e^{-\rho t} g(a \cdot X_t)e^{-\gamma c \cdot X_t}. \]  

(69)

With the stochastic discount factor in this form, we may price any asset with \( pl \)-linear payoffs, such
as bonds and dividend claims, using Theorem 1. Our method also applies when the two agents have different risk aversion (γA and γB). In that case, we can still express h(c, η) in the separable form as in (69), and proceed the same way.

In some cases, the mapping of a difference-of-opinion model to the standard setting (62) is not immediate, and requires a careful choice of the state variable X_t. Consider the setting where agents believe the state of the economy is summarized by the N-dimensional Gaussian state variables, X_t. According to agents A and B the dynamics of X_t are

\[
\begin{align*}
dX_t & = (K_0^A + K_1^A X_t)dt + \sqrt{H_0}dW_t^A \\
dX_t & = (K_0^B + K_1^B X_t)dt + \sqrt{H_0}dW_t^B
\end{align*}
\]

(70)

In this case, the difference in beliefs cannot be directly expressed as in (62) directly. However, by considering an augmented state variable we can return to the form of (62). Define \( \hat{X}_t = \langle X_t, \text{vech}(X_tX_t^\top) \rangle \). It is easy to verify that \( \hat{X}_t \) is an affine process (with covariance \( \hat{H}_0 + \hat{H}_1 \cdot \hat{X}_t \)) and that by appropriate choice of a,

\[
(K_0^B + K_1^B X_t) - (K_0^A + K_1^A X_t) = (\hat{H}_0 + \hat{H}_1 \cdot \hat{X}_t)a
\]

(71)

Similarly, the integral term in (62) has affine drift in \( \hat{X}_t \), and so we return to our standard setting.\(^{14}\)

Such techniques are common in the term structure literature with respect to affine and quadratic term structure models. The procedure generalizes to accommodate models with stochastic volatility (\( AM(N) \) in the parlance of Dai and Singleton (2000)). Following Duffee (2002), we refer to this as essentially affine difference of beliefs.

An alternative but less general characterization is to consider the “market price of belief risk”, \( \lambda_t \), in analogy to the usual market price of risk. By defining

\[
\begin{align*}
\lambda_t & = \sqrt{H_0^{-1}(\mu_t^B - \mu_t^A)} \\
\eta_t & = e^{-\int_0^t \lambda_s dW_s^A - \frac{1}{2} \int_0^t \lambda_s^2 ds}
\end{align*}
\]

(72)

(73)

\(^{13}\)The function \( g \) is not bounded and in fact does not even define a tempered function. Thus, our theory does not directly apply. One option is to write \( g(x) = g_- (x)e^{-x} + g_+(x)e^{+x} \) where \( g_\pm (x) = g(x)1_{\pm x > 0}e^{\mp x} \). Here \( g_\pm \) are bounded functions whose Fourier transforms can be computed in terms of incomplete Beta functions. Another option is to write \( g(x) = g(x)^{\gamma}/g(x)^{\gamma+1} \). In this case, the first functional is \text{pl-linear} and the second is bounded with Fourier transform known in terms of Beta functions.\(^{14}\) Usually, only part of the elements in the extended state vector is needed to maintain the Markov structure.
where $\eta_t$ is exponential affine in $\hat{X}_t$ and defines an appropriate Radon-Nikodym derivative.

### 4.3.2 Special Cases

The framework above can accommodate a wide range of specifications with heterogeneity of beliefs regarding expected changes in fundamentals, likelihood of jumps, distribution of jumps, and divergence in higher order beliefs. We now provide some examples.

**Disagreement about stochastic growth rates.** This is the model studied in Dumas et al. (2009), hereafter DKU. In their model, there is a single dividend process $C_t$ with time-varying growth rate, but agents A and B have different beliefs regarding the growth rate of the tree, $\hat{f}^A_t$ and $\hat{f}^B_t$, and $\hat{g}_t = \hat{f}^B_t - \hat{f}^A_t$ represents the amount of disagreement between B and A.

This model can be mapped into the essentially affine difference in beliefs specification, and our results can simplify the calculations for the most general model that they consider. First, under Agent B’s probability measure,

\[
\begin{pmatrix}
  c_t \\
  \hat{f}^B_t \\
  \hat{g}_t
\end{pmatrix}
= \begin{pmatrix}
  \frac{1}{2} \sigma^2_c \\
  \hat{g}_t \\
  \kappa (\hat{f}^B_t - \hat{f}^B_t) \\
  -\psi \hat{g}_t
\end{pmatrix} dt + \begin{pmatrix}
  \sigma_c & 0 \\
  0 & \sigma_c \\
  \sigma_{\hat{g},c} & \sigma_{\hat{g},s}
\end{pmatrix}
\begin{pmatrix}
  \frac{\mu}{\sigma_c} \\
  0 \\
  -\sigma_{\hat{g},s} \hat{g}_t
\end{pmatrix} dW^B_t. \tag{74}
\]

Next, in order to map the model to our standard setting, we define the augmented state variable as $X_t = (c_t, \log \eta_t, \hat{f}^B_t, \hat{g}_t, \hat{g}^2_t)$, where $\eta_t$ gives the density process: $\eta_t = E_t[dP^A/dP^B]$. The dynamics of $X_t$ are given by the stochastic differential equation:

\[
dX_t = (K_0 + K_1 X_t) dt + \Sigma_t dW^B_t
\]

where

\[
K_0 = \begin{bmatrix}
-\frac{1}{2} \sigma^2_c \\
\hat{g}_t \\
\kappa \hat{f}^B_t \\
0 \\
\sigma^2_{\hat{g},c} + \sigma^2_{\hat{g},s}
\end{bmatrix}, \quad K_1 = \begin{bmatrix}
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & -1/\sigma_c & 0 \\
0 & 0 & -\kappa & 0 & 0 \\
0 & 0 & 0 & -\psi & 0 \\
0 & 0 & 0 & 0 & -2\psi
\end{bmatrix}, \quad \Sigma_t = \begin{bmatrix}
\sigma_c & 0 \\
0 & \frac{\mu}{\sigma_c} \\
\gamma_B/\sigma_c & 0 \\
\sigma_{\hat{g},c} & \sigma_{\hat{g},s} \\
2\sigma_{\hat{g},c}\hat{g}_t & 2\sigma_{\hat{g},s}\hat{g}_t
\end{bmatrix}.
\]

It is easy to check that the local conditional variance of $X_t$, $\Sigma_t \Sigma_t^T$, is affine in $X_t$ so this represents an
affine process.\footnote{DKU exploit the fact that in this particular case the ODE determining the conditional characteristic function for some variables can be computed in closed form by standard methods. However, in general there is little additional complication to solve the usual ODE by standard numerical methods.} Then, it is immediate that $\eta_t$ takes the form of (62) with $a = (0,1,0,0,0)$. Finally, we extend the state variable $X$ to include the integral term in (62) and augment $a$ accordingly so that (63) holds.

DKU show that in their setting a number of equity and fixed income security prices take the form $E_0[e^{\alpha \cdot X_t}g(\beta \cdot X_t)]$ where $g(x) = (1 - e^{ax})^b$ for some $(\alpha, \beta, a, b)$. They use two methods to compute this moment. First, when $b \in \mathbb{N}$, $g$ can be expanded directly and reduced to log-linear functionals. Then the moments can be computed by well-known methods. For more general cases, they compute the moment in two steps: first recover the forward density of $\beta \cdot X$ through a Fourier inversion of the conditional characteristic function, and then evaluate the expectation using the density. The formula they use (A58-A61) is essentially

\[ E_0[e^{\alpha \cdot X_t}g(\beta \cdot X_t)] = \frac{1}{2\pi} \int_{b \in \mathbb{R}} \hat{g}(b) \int_{s \in \mathbb{R}} e^{ibs} E_0[e^{(\alpha - is\beta) \cdot X_t}] ds \, db. \] (75)

This formula requires a double integral, thus increasing the dimensionality of the problem. As Theorem 1 shows, our generalized transform method will only require a single integral to compute this moment. If we consider the generalization $g(\beta_1 \cdot X_t, \beta_2 \cdot X_t)$, the trade-off becomes a somewhat tractable 2-dimensional integral with our method versus a highly intractable 4-dimensional integral by using an extension of the DPS method.

\textbf{Disagreement about volatility.} Suppose that dividends have stochastic volatility. Under Agent A’s beliefs:

\[ d \begin{bmatrix} c_t \\ V_t \end{bmatrix} = \begin{bmatrix} \hat{g} \\ -\kappa V_t \end{bmatrix} dt + \begin{bmatrix} \sigma_d & 0 \\ 0 & 0 \end{bmatrix} dW^A_t + \begin{bmatrix} \sigma_c V_t & 0 \\ 0 & \sigma_{VV} V_t \end{bmatrix} dW^B_t \] (76)

Here $\sigma_d$ is the lowest conditional variance of log dividends, while $V_t$ represents the degree to which volatility is above the lowest level.

Agent B disagrees about the dynamics of volatility. According to his beliefs:

\[ d \begin{bmatrix} c_t \\ V_t \end{bmatrix} = \begin{bmatrix} \hat{g} \\ -((\kappa V - b) V_t) \end{bmatrix} dt + \begin{bmatrix} \sigma_d & 0 \\ 0 & 0 \end{bmatrix} dW^A_t + \begin{bmatrix} \sigma_c V_t & 0 \\ 0 & \sigma_{VV} V_t \end{bmatrix} dW^B_t \] (77)
For example, when \( b > 0 \), Agent B believe that volatility mean reverts more slowly. Using \( a = \langle 0, b/\sigma^2 \rangle \) we get the *db-density* as in Equation (62).

**Disagreement about momentum.** Consider a model with stochastic growth in consumption. Let \( c_t \) be the log consumption, \( g_t \) be the expected growth rate. Also, let \( e_t \) be an exponential weighted moving average of past growth rates:

\[
e_t = \int_{-\infty}^{t} e^{-b(t-s)} g_s ds.
\] (78)

Agent A correctly believes that the expected growth rate of log consumption is \( g_t \). Under her beliefs:

\[
d \begin{bmatrix} c_t \\ g_t \\ e_t \end{bmatrix} = \begin{bmatrix} g_t \\ \kappa(\bar{g} - g_t) \\ g_t - be_t \end{bmatrix} dt + \begin{bmatrix} \sigma_c & 0 \\ 0 & \sigma_g \end{bmatrix} dW_t^A
\] (79)

Agent B believes that growth is due to two components: (1) a mean-reverting component, \( g_t \) and (2) a counteracting momentum component through \( e_t \).

\[
d \begin{bmatrix} c_t \\ g_t \\ e_t \end{bmatrix} = \begin{bmatrix} g_t + ce_t \\ \kappa(\bar{g} - g_t) \\ g_t - be_t \end{bmatrix} dt + \begin{bmatrix} \sigma_c & 0 \\ 0 & \sigma_g \end{bmatrix} dW_t^B.
\] (80)

Fixing the past, for large enough deviations from the steady-state, the mean-reverting component will dominate. However, for small deviation from the steady state, Agent B will believe that past deviations from the steady state lead to larger future deviations from the state steady. In this way we can view Agent B as possessing a conservatism or “law of small numbers” bias.

This example represents a special case of the essentially affine difference of beliefs.

**Disagreement about higher order beliefs.** Heterogeneity in higher order beliefs can affect asset prices as well. We can inductively proceed in defining beliefs:

\[
\tilde{g}^{i} = \text{Agent } i\text{'s beliefs about the growth rate of consumption}
\]

\[
\tilde{g}^{ij} = \text{Agent } i\text{'s beliefs about Agent } j\text{'s belief about the growth rate of consumption}
\]
We can consider the state variable $X_t = [c_t, \hat{g}^A_t, \hat{g}^B_t, \hat{g}^{AB}_t, \hat{g}^{BA}_t]$. Suppose that $X_t$ follows a Gaussian process under both agents beliefs. Agent A’s beliefs are such that

$$
\begin{bmatrix}
    c_t \\
    \hat{g}^A_t \\
    \hat{g}^B_t \\
    \hat{g}^{AB}_t \\
    \hat{g}^{BA}_t
\end{bmatrix}
\begin{bmatrix}
    \kappa_A(\theta - \hat{g}^A) \\
    \kappa_B(\theta - \hat{g}^B) \\
    \kappa_{AB}(\hat{g}^B - \hat{g}^{AB}) \\
    \kappa_{BA}(\hat{g}^A - \hat{g}^{BA})
\end{bmatrix}
\begin{array}{c}
    dt + \Sigma dW_t^A
\end{array}
\tag{81}
$$

Here, the fourth and fifth components of the drift say that Agent A believes that the higher order beliefs (both his beliefs about Agent B and Agent B’s beliefs about him) are correct in the long run, but may have short run deviations.

Again, this model represents a special case of the essentially affine disagreement.

**Disagreement about the likelihood of disasters.** Suppose that log consumption, $c_t$, has constant growth with IID innovations with time-varying probability, $\lambda_t$, of rare disaster. Let $X_t = [c_t, \lambda_t]$. Under Agent A’s beliefs,

$$
dX_t =
\begin{bmatrix}
    g_A \\
    -\kappa_\lambda \lambda_t
\end{bmatrix}
dt +
\begin{bmatrix}
    \sigma_c & 0 \\
    0 & \sigma_\lambda \sqrt{\lambda_t}
\end{bmatrix}
\begin{array}{c}
    dW_t^A + dZ_t^A
\end{array}
\tag{82}
$$

where $Z_t^A$ are jumps in $c_t$ which occur with intensity $\lambda_0 + \lambda_t$ and distribution $\nu$. Suppose that Agent B’s beliefs are specified by the db-density of form (62) with $a = (b, 0)$. Then, Agent B’s beliefs will be

$$
dX_t =
\begin{bmatrix}
    g_A + b\sigma_c^2 \\
    -\kappa_\lambda \lambda_t
\end{bmatrix}
dt +
\begin{bmatrix}
    \sigma_c & 0 \\
    0 & \sigma_\lambda \sqrt{\lambda_t}
\end{bmatrix}
\begin{array}{c}
    dW_t^B + dZ_t^B
\end{array}
\tag{83}
$$

where jumps arrive with intensity $\lambda_t^B = E_{\nu^A}[e^{a\cdot Z}(\lambda_0 + \lambda_t)]$ and have distribution $\nu^B$ with Radon-Nikodym derivative $d\nu^B/d\nu^A(Z) = e^{a\cdot Z}/E_{\nu^A}[e^{a\cdot Z}]$.

In this sense, Agent B is more optimistic about the future growth both in terms of (1) higher expected growth rates, (2) lower likelihood of disasters, (3) less severe losses conditional on there being a disaster.

For illustration, we consider the case of disagreement about the likelihood of disasters. The
Figure 7: Price-Dividend Ratios with Heterogeneity in Beliefs. These figures plots the Price-Dividend ratio in an economy where agents hold different beliefs regarding growth rates and disaster probabilities.

The model is calibrated as follows:

- \( g_A = 2\% , \sigma_c = 1.5\% , \kappa_\lambda = 0.1 , \lambda_0 = 1.5\% , \sigma_\lambda^\infty = 1\% \), (the volatility of the stationary distribution of \( \lambda \)), \( \gamma = 2 \). Each jump causes a 8% drop in consumption (\( Z = -8\% \)).

- Agent B’s beliefs are generated with \( a = (b, 0) \) where \( b \) is chosen so that the agent is either more pessimistic or optimistic (both with regard to growth rates and the likelihood of disasters).

- Lagrange multiplier for agent A, \( \zeta_A \) is adjusted to vary the fraction of wealth endowed to agent A. (\( \zeta_B \) is normalized to 1.)

We vary the amount of disagreement between the two agents by changing \( b \). A more negative value for \( b \) simultaneously lowers Agent B’s perceived growth rate \( g_B \), and increases his perceived likelihood of a disaster. More specifically, the growth rate of consumption (excluding jumps) under Agent B’s belief is \( g_B = g_A + b \sigma_c^2 \), and the jump intensity under B’s belief is \( e^{bZ} \) times the intensity under A’s belief.

Figure 7 plots the price-dividend ratio of the aggregate consumption claim. Agent A believes the expected growth rate of consumption is \( g_A = 2\% \). Agent B can be pessimistic or optimistic relative to Agent A. The horizontal axis gives the fraction of total wealth in the economy owned
by Agent A at $t = 0$. The left panel considers the special case without disaster risk ($Z = 0$), and the right panel considers the case with disasters. First, in an economy where only Agent A lives, the price-dividend ratio is 6.9 (the dashed line). When we introduce Agent B, who is more pessimistic than A (B believes in lower expected growth rate and higher chances of disasters), the equilibrium price of the consumption claim drops. Interestingly, as the wealth of Agent B (A) increases (decreases), the price first decreases and then increases. Finally, when Agent B owns all the wealth, the price-dividend ratio rises to 7.8. The results are qualitatively similar when Agent B is more optimistic.

The eventual increase of asset prices with the wealth of the pessimistic agent is due to the low elasticity of intertemporal substitution of a CRRA-utility agent with $\gamma > 1$. The agent wants to smooth consumption by saving more, which lowers the interest rate and raises the price of consumption claims. The reason for the initial decline of asset prices is as follows. Consider Agent A, who has more optimistic beliefs. He sells Agent B his share on aggregate consumption in the “bad states” (where aggregate consumption is low) in exchange for a bigger share on aggregate consumption in the “good states” (where aggregate consumption is high). In equilibrium, his consumption has higher expected growth rate and volatility than the aggregate endowment. As a result, the interest rate rises, and so does the Sharpe ratio under his beliefs. For large enough disagreement in beliefs, this effect will dominate, causing prices to fall.

When there is disagreement about the probability of disasters, the impact of heterogeneous beliefs on prices becomes more striking. As shown in the right panel, when Agent B is more pessimistic, the equilibrium price of consumption claims falls by over 14% even when Agent B only owns 10% of the total wealth in the economy. This result is consistent with the findings of Kogan et al. (2006) that irrational traders can still have nontrivial price impacts when their wealth is small.

5 Concluding Remarks

We extend the transform analysis in Duffie et al. (2000) to compute a general class of nonlinear moments for affine jump diffusions. Through a Fourier decomposition of the nonlinear moments, we can directly utilize the properties of the conditional characteristic functions for affine processes and compute the moments analytically.

We demonstrate the power of this method with examples from several areas, including option
pricing, term structure, credit risk modeling, and GMM estimations. We also illustrate the application of the generalized transform method in three in-depth examples: a model of defaultable bond pricing, where the recovery rates are conditionally correlated with default intensities; an equilibrium model of a Lucas economy with non-i.i.d. trees; and a general class of difference-of-opinion models.
A Proofs

A.1 Proof of Theorem 1

Throughout, we maintain the following assumptions:

**Assumption 1:** In the terminology of DPS, \((\Theta, \alpha, \beta)\) is well-behaved at \((s, T)\) for all \(s \in \mathbb{R}\). That is,

(a) \(E\left(\int_{0}^{T} |\gamma_t| dt\right) < \infty\) where \(\gamma_t = \Psi_t(\phi(B(T - t)) - 1)(\lambda_0 + \lambda_1(X_t))\)

(b) \(E[(\int_{0}^{T} \|\eta_t\|^2 dt)] < \infty\) where \(\|\eta_t\|^2 = \Psi_t^2B(T - t)^\top(H_0 + H_1 \cdot X_t)B(T - t)\)

(c) \(E[|\Psi_T|] < \infty\)

where \(\Psi_t = e^{-\int_{0}^{t} r_\tau d\tau} e^{A(T-t)+B(T-t)\cdot X_t}\) and \(A, B\) solve the ODE/BVP given in (8-9).

**Assumption 2:** The measure \(F\) defined by its Radon-Nikodym derivative,

\[
\frac{dF}{dP} = \frac{e^{-\int_{0}^{T} r_\tau d\tau} e^{A\cdot X_T}}{E_0[e^{-\int_{0}^{T} r_\tau d\tau} e^{A\cdot X_T}]},\tag{84}
\]

is such that the density of \(\beta \cdot X_T\) under \(F\) is a Schwartz function. In particular, the density of \(\beta \cdot X_T\) is smooth and declines faster than any polynomial under \(F\).

Proposition 1 of DPS gives conditions under which Assumption 1 holds. These are integrability conditions which imply that, for every \(s\), the local martingale \(E_t[e^{-\int_{t}^{T} r_\tau d\tau + \alpha + i s \beta} e^{-A_{T-t} - B_{T-t} \cdot X_t}]\) is a fact a martingale.

Assumption 2 is analogous to (2.11) of DPS. However, we require a somewhat stronger assumption to directly apply our theory. This assumption can typically be shown to hold by verifying that the moment generating function (under \(F\)) is finite in a neighborhood of 0.

We now prove Theorem 1. Suppose now that Assumptions 1 and 2 hold. Then,

\[
H = E_0[e^{-\int_{0}^{T} r_\tau d\tau} e^{\alpha \cdot X_T} g(\beta \cdot X_T)]
= F_0 E_0^F [g(\beta \cdot X_T)]
= F_0 \int g(b) F^F_{\beta \cdot X_T}(b) db
= F_0 \langle g, F^F_{\beta \cdot X_T} \rangle.
\]
In the last equation, we interpret \( g \in S' \). By Assumption 2, \( f_{\beta,X_T} \in S \), and so \( \hat{f}_{\beta,X_T} \in S \) also. Thus Fourier inversion holds and \( (\hat{f}_{\beta,X_T}) = \frac{1}{2\pi} f_{\beta,X_T} \) (see Corollary 8.28 in Folland (1984).)  

Applying this,

\[
H = \frac{1}{2\pi} F_0 \langle g, (\hat{f}_{\beta,X_T}) \rangle \\
= \frac{1}{2\pi} F_0 \langle \hat{g}, \hat{f}_{\beta,X_T} \rangle \\
= \frac{1}{2\pi} \langle \hat{g}, F_0 \hat{f}_{\beta,X_T} \rangle \\
= \frac{1}{2\pi} \langle \hat{g}, G(\alpha - \beta i) \rangle.
\]

The second step holds by the definition of the Fourier transform of a tempered distribution and the last step hold by Assumption 1. This is the desired result. \( \Box \)

### A.2 Proof of Proposition 1

In analogy to Duffie et al. (2000) and Pan (2002), define

\[
G(\alpha_0; v, n|x, t) = e^{A t + B \cdot x} \sum_{|\xi| = n} \left( \begin{array}{c} n \\ \xi \end{array} \right) L(x)^{\xi} 
\]

where \( L(x) \) is the \( n \)-dimensional vector whose \( i \)-th coordinate is \( (\partial_i A + \partial_i B \cdot x)^{1/i} \), \( \xi \) is an \( n \)-dimensional multi-index, and \( (\partial_i A, \partial_i B) \) satisfies the ODE/BVP

\[
\dot{B} = K_1^T B + \frac{1}{2} B^T H_1 B - \rho_1 + \lambda_1 (\phi(B) - 1) \quad B(0) = \alpha_0 \tag{86}
\]

\[
\dot{A} = K_0^T B + \frac{1}{2} B^T H_0 B - \rho_0 + \lambda_0 (\phi(B) - 1) \quad A(0) = 0 \tag{87}
\]

\[
\partial_1 \dot{B} = K_1^T \partial_1 B + \partial_1 B^T H_1 B + \lambda_1 \nabla \phi(B) \cdot \partial_1 B \quad \partial_1 B(0) = v \tag{88}
\]

\[
\partial_1 \dot{A} = K_0^T \partial_1 B + \partial_1 B^T H_0 B + \lambda_0 \nabla \phi(B) \cdot \partial_1 B \quad \partial_1 A(0) = 0 \tag{89}
\]

and for \( 2 \leq m \leq n \), \( (\partial_m B, \partial_m A) \) satisfy

\[
\partial_m \dot{B} = K_1^T \partial_1 B + \frac{1}{2} \sum_{i=0}^{m} \left( \begin{array}{c} m \\ i \end{array} \right) \partial_i B^T H_1 \partial_{m-i} B + \partial_{m-1} (\lambda_1 \nabla \phi(B) \cdot \partial_1 B) \quad \partial_m B(0) = 0 \tag{90}
\]

\[
\partial_m \dot{A} = K_0^T \partial_1 B + \frac{1}{2} \sum_{i=0}^{m} \left( \begin{array}{c} m \\ i \end{array} \right) \partial_i B^T H_0 \partial_{m-i} B + \partial_{m-1} (\lambda_0 \nabla \phi(B) \cdot \partial_1 B) \quad \partial_m A(0) = 0 \tag{91}
\]

\[\text{We use the convention that for } f \in L^1, f(s) = \int e^{-i\alpha s} f(x)dx, \text{ f(x) = } \int e^{i\alpha s} f(s)ds \text{ and for } g \in S^*, \langle \hat{g}, f \rangle \equiv \langle g, \hat{f} \rangle.\]
We strengthen Assumptions 1 and 2 as follows:

1. **Assumption 1’**: The moment generating function, $\phi \in C^N(D_0)$ where $D_0$ is an open set containing the image of the solutions to (8) for any initial condition of the form $\alpha_0 = \alpha + is\beta$ for any $s \in \mathbb{R}$. Additionally, for any such a initial condition:

   (a) $E \left( \int_0^T |\gamma_t| dt \right) < \infty$ where

   $$\gamma_t = \lambda_t E_\nu [\Psi^n_t(i_t, X_t + Z) - \Psi^n_t(i_t, X_t)]$$

   and $\Psi^n_t(i, x) = e^{-t}G(\alpha, v, n|x, T-t)$ and $i_t = \int_0^t r_s ds$.

   (b) $E[\left( \int_0^T \|\eta_t\|^2 dt \right)] < \infty$ where

   $$\|\eta_t\|^2 = \nabla_x \Psi^n_t(i_t, X_t)^\top (H_0 + H_1 \cdot X_t) \nabla_x \Psi^n_t(i_t, X_t)$$

   (c) $E[|\Psi_T(i_T, X_T)|] < \infty$

2. **Assumption 2’**: The measure $F$ defined by its Radon-Nikodym derivative,

   $$\frac{dF}{dP} = \frac{e^{-\int_0^T r_s d\tau} e^{\alpha \cdot X_T (v \cdot X_T)^n}}{E_0 [e^{-\int_0^T r_s d\tau} e^{\alpha \cdot X_T (v \cdot X_T)^n}]}$$

   is such that the density of $\beta \cdot X_T$ under $F$ is a Schwartz function.

Given Assumption 1’ and Assumption 2’ hold, the proof follows as before.
References


