Generalized Transform Analysis of Affine Processes and Applications in Finance

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1093/rfs/hhs065">http://dx.doi.org/10.1093/rfs/hhs065</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Oxford University Press</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Mon Feb 25 21:47:09 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/74650">http://hdl.handle.net/1721.1/74650</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike 3.0</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/3.0/">http://creativecommons.org/licenses/by-nc-sa/3.0/</a></td>
</tr>
</tbody>
</table>
Generalized Transform Analysis of Affine Processes and Applications in Finance*

Hui Chen† Scott Joslin‡

September 21, 2011

Abstract

Non-linearity is an important consideration in many problems of finance and economics, such as pricing securities and solving equilibrium models. This paper provides analytical treatment of a general class of nonlinear transforms for processes with tractable conditional characteristic functions, which extends existing results on characteristic function based transforms to a substantially wider class of nonlinear functions while maintaining low dimensionality by avoiding the need to compute the density function. We illustrate the applications of the generalized transform in pricing defaultable bonds with stochastic recovery. We also use the method to analytically solve a class of general equilibrium models with multiple goods and apply this model to study the effects of time-varying labor income risk on the equity premium.

*We thank Peter Feldhutter, Damir Filipovic, Xavier Gabaix, Chris Jones, Martin Lettau, Ian Martin, Jun Pan, Tano Santos, Ken Singleton, and seminar participants at Boston University, Duke Fuqua, MIT Sloan, Penn State Smeal, USC Marshall, the 2009 Adam Smith Asset Pricing Conference, the 2009 European Finance Association (Bergen), and the Winter 2010 Econometric Society Meetings (Atlanta) and the for comments. Tran Ngoc-Khanh and Yu Xu provided excellent research assistance. All the remaining errors are our own.

†MIT Sloan School of Management and NBER, huichen@mit.edu. Tel. 617-324-3896.
‡USC Marshall School of Business, sjoslin@usc.edu. Tel. 213-740-7137.
1 Introduction

In this paper, we provide analytical treatment of a class of transforms for processes with tractable characteristic functions. These transforms bring analytical and computational tractability to a large class of nonlinear moments, and can be applied in option pricing, structural estimation, or solving equilibrium asset pricing models. We demonstrate the utility of our method with two examples, one on pricing defaultable bonds with stochastic recovery, the other on solving a general equilibrium model with stochastic labor income risk.

Consider a state variable $X_t$ with transition dynamics under a certain probability measure summarized by a tractable conditional characteristic function. Many popular stochastic processes in economics have simple characteristic functions, such as the affine jump-diffusions, Lévy processes, and Markov-switching affine processes. We provide closed-form expression (up to an integral) for the following transform:

$$E_t^m \left[ \exp \left( - \int_t^T R(X_s, s) \, ds \right) f(X_T) g(\beta \cdot X_T) \right],$$  \hspace{1cm} (1)

where $f$ can be 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). We use $E^m$ to denote the expectation under an arbitrary measure $m$. The substantial flexibility in choosing $f$ and $g$ in (1) as well as the process for $X_t$ makes the above transform useful in dealing with generic non-linearity problems in asset pricing (nonlinear stochastic discount factors or payoffs), estimation (nonlinear moments), and other areas.

Our method utilizes knowledge of the conditional characteristic function of the state variable $X_t$ (under certain forward measures) jointly with a Fourier decomposition of the non-linearity in $g$. This allows us to replace any non-linearities with an average of log-linear functions, for which the conditional characteristic function can be used directly to compute the expectations. This combination brings tractability to our generalized transform by avoiding intermediate Fourier inversions. Our method allows for a large class of nonlinear functions (tempered distributions) which include discontinuous and non-differentiable functions as well as unbounded and non-integrable
functions for which a standard Fourier transform might not exist. Moreover, a large variety of stochastic processes have tractable characteristic functions (such as affine jump-diffusions or Lévy processes). As a result, our method is applicable to a wide range of problems.

One area where the generalized transform in (1) can be a powerful tool is risk-neutral pricing. We can value a large class of nonlinear payoffs analytically provided we know the forward conditional characteristic function of the underlying state variables.

The generalized transform can also be useful in economic modeling. For example, suppose we want to value an asset under the historical measure (\( \mathbb{P} \)) using the stochastic discount factor \( m_t \). The value at time \( t \) of a stochastic payoff \( y_T \) at time \( T \) is

\[
P_t = \frac{1}{m(t, X_t)} E_t^{\mathbb{P}} \left[ m(T, X_T) y(X_T) \right].
\]

In the background, there is an equilibrium model that endogenously determine the stochastic discount factor \( m \) and payoff \( y \) as functions of the state variables \( X \). In order to maintain tractability, we are often forced to adopt special utility functions (e.g., logarithmic utility), impose strong restrictions on the state variable process (e.g., i.i.d. or conditionally Gaussian), or log-linearize the model to obtain approximate solutions. The forward conditional characteristic function of \( X \) is typically quite complicated for general stochastic discount factors, which means changing to the risk-neutral probability measure will not help simplify the problem.\(^1\)

However, the above pricing equation bears resemblance to the generalized transform (1). The difficulties in pricing are typically due to the non-linearity in the discount factor \( m \) or payoff \( y \), which can be addressed using the new tools provided in this paper. Through the generalized transform, we can (i) price assets with payoffs that are potentially discontinuous or non-differentiable; (ii) allow for more general preferences; (iii) have more flexibility when introducing heterogeneity across agents, firms, or countries (in models of international finance); and (iv) significantly enrich the underlying stochastic uncertainties governing the economy by introducing features such as time-varying growth rates, stochastic volatility, jumps, or cointegration restrictions.

\(^1\)More precisely, the characteristic function under \( Q \) may be known only up to an integral. Computing the expected nonlinear payoffs by risk-neutral pricing and the transform would then require a double integral, substantially increasing the computational difficulty.
To illustrate the application in risk-neutral pricing, we study the pricing of defaultable bonds with stochastic recovery rates. The recovery rate of a defaulted security can depend on firm characteristics and macroeconomic conditions. We provide analytic pricing for defaultable bonds under these conditions. We show that the negative correlation between recovery rates and default rates found in the data implies substantial non-linearities in credit spreads. The comparison between the stochastic recovery model and widely used constant recovery models shows that ignoring stochastic recovery can lead to economically significant pricing errors.

In a second example, we study a general equilibrium model with time-varying labor income risk. We build upon the work of Santos and Veronesi (2006), who find that the share of labor income to consumption predicts future excess returns of the market portfolio. They attribute this result with the “composition effect”: a higher labor share implies a lower covariance between consumption and dividends, which lowers the equity premium. Motivated by empirical evidence that volatilities of labor income and dividends as well as the correlation between the two change over time, we explore a model with time-varying covariance between labor income and dividends. We obtain analytical solutions of the model via the generalized transform.

In the calibrated model, we show that the equity premium depends on both the labor income share and the covariance between labor income and dividends. As in Santos and Veronesi (2006), we find a negative relationship between labor income share and the equity premium provided that their covariance is not too low. However, when the covariance between dividends and labor income is low, the labor share is no longer negatively related to the equity premium. Thus, stochastic covariance between labor income and dividends can help explain the changing predictive power of labor share in the data. In addition, the model has interesting implications for the comovement between the risk premium on financial wealth and human capital. This example illustrates the power of our method in economic modeling. Variations of this model can be used to study areas such as the cross section of stock returns or international asset pricing.

Relation to the literature

Thanks to its tractability and flexibility, affine processes have been widely used in term structure models, reduced-form credit risk models, and option pricing. In particular, the transform analysis
of general affine jump-diffusions in Duffie, Pan, and Singleton (2000) (hereafter DPS) makes it easy to compute certain moments arising from asset pricing, estimation, and forecasting. Example applications include Singleton (2001), Pan (2002), Piazzesi (2005), and Joslin (2010), among others.\(^2\)

When the moment functions do not conform to the basic DPS transform, one possible solution is to first recover the conditional density of the 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. Then, one can evaluate the nonlinear moments by directly integrating over the density. Through this method, DPS obtain the extended transform for affine jump-diffusions, which they apply to option pricing.\(^3\)

Alternative methods to compute nonlinear moments include simulations or solving numerically the partial differential equations arising from the expectations via the Feynman-Kac methodology, which can be time-consuming and lacking accuracy, especially in high dimensional cases. In our approach, we consider affine jump-diffusions and indeed any process with known conditional characteristic functions. Moreover, our method allows direct computation of a large class of nonlinear moments without the need to compute the (forward) density of the underlying state variable.

**Bakshi and Madan** (2000) connect the pricing of a class of derivative securities to the characteristic functions for a general family of Markov processes. In addition, they propose to approximate a nonlinear moment function with a polynomial basis, provided the function is entire, which in turn can be computed via the conditional characteristic function and its derivatives. Our method applies to more general nonlinear moment functions through the Fourier transform. We also extend the results to multivariate settings.

A few earlier studies have considered related Fourier methods. **Carr and Madan** (1999) address the non-linearity in a European option payoff by taking the Fourier transform of the payoff function with respect to the strike price. **Martin** (2011) takes the Fourier transform of a nonlinear pricing

\(^2\)Gabaix (2009) considers a class of linearity generating processes, where particular moments (or accumulated moments) can have a very simple linear form given minor deviations from the assumption of affine dynamics of the state variable. This feature makes it very convenient to obtain simple formulas for the prices of stocks, bonds, and other assets. His work generalizes the model of Menzly, Santos, and Veronesi (2004).

\(^3\)Other papers that take this approach include Heston (1993), Chen and Scott (1995), Bates (1996), Bakshi and Chen (1997), Bakshi, Cao, and Chen (1997), Chacko and Das (2002), Dumas, Kurshev, and Uppal (2009), Buraschi, Trojani, and Vedolin (2010), among others.
kernel that arises in the two tree model of Cochrane, Longstaff, and Santa-Clara (2008). In both studies the state variables have i.i.d. increments. To the best our knowledge, this paper is the first to generalize the above approach both in terms of the moment function (to the class of tempered distributions) and the process of underlying state variables (including affine processes and Lévy processes).

2 Illustrative example

Before presenting the main result of the paper, we first illustrate the idea behind the generalized transform using an example of forecasting the average recovery rate of defaulted corporate bonds. The amount an investor recovers from a corporate bond upon default can depend on many factors, such as firm specific variables (debt seniority, asset tangibility, accounting information), industry variables (asset specificity, industry-level distress), and macroeconomic variables (aggregate default rates, business cycle indicators). In addition, the recovery rate as a fraction of face value should in principle only take value from $[0, 1]$.

A simple way to capture these features is to model the recovery rate using the logistic model:

$$\varphi(X_t) = \frac{1}{1 + e^{-\beta_0 - \beta_1 \cdot X_t}}, \quad (2)$$

where $X_t$ is a vector of the relevant explanatory variables observable at time $t$. For example, Altman, Brady, Resti, and Sironi (2005) model the aggregate recovery rate as a logistic function of the aggregate default rate, total amount of high-yield bonds outstanding, GDP growth, market return, and other covariates.

Investors may be interested in forecasting future average recovery rate in the economy. That is, we are interested in computing $E_0[\varphi(X_T)]$. To simplify notation, we first define $Y_T \equiv \frac{1}{2}(-\beta_0 - \beta_1 \cdot X_T)$.

---

4In the $N$-tree case, $N > 2$, Martin (2011) also provides an $(N - 2)$-dimensional integral to compute the associated $(N - 1)$-dimensional transform.

5We suppress the conditioning on the default event occurring at time $T$ and further suppose that default occurring at time $T$ is independent of the path $\{X_t\}_{0 \leq t \leq T}$. This is stronger than the standard doubly stochastic assumption; relaxing this assumption is relatively straightforward but would complicate the example.
We then rewrite

\[
E_0[\varphi (X_T)] = E_0 \left[ \frac{1}{1 + e^{2Y_T}} \right] = E_0 \left[ \frac{1}{2} e^{-Y_T} \frac{1}{\cosh(Y_T)} \right],
\]

(3)

where we use the hyperbolic cosine function, \( \cosh(y) = \frac{1}{2} (e^y + e^{-y}) \).

Although only a single variable \( Y_T \) appears in (3), its conditional distribution may depend on the current values of each individual covariate, that is, \( Y_t \) itself may not be Markov. Even if the covariates \( X_t \) follow a relatively simple process, direct evaluation of this expectation requires computing a multi-dimensional integral, which can be difficult when the number of covariates is large.

Suppose, however, that the conditional characteristic function (CCF) of \( X_T \) is known:

\[
CCF(T, u; X_0) = E[e^{iu \cdot X_T} \mid X_0],
\]

(4)

where \( i = \sqrt{-1} \). If we could “approximate” the nonlinear term \( 1/\cosh(Y_T) \) inside the expectation of (3) with exponential linear functions of \( Y_T \), then we would be able to use the characteristic function to compute the nonlinear expectation. As we elaborate in Section 3, this is achieved using the Fourier inversion of \( 1/\cosh(y) \),

\[
\frac{1}{\cosh(y)} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{g}(s)e^{isy} ds,
\]

(5)

where \( \hat{g} \) is the Fourier transform of \( 1/\cosh(y) \), which is known analytically (see, for example, Abramowitz and Stegun (1964) 6.1.30 and 6.2.1), \( \hat{g}(s) = \pi/\cosh(\frac{\pi s}{2}) \).

Thus, we can substitute out \( 1/\cosh(Y_T) \) from (3) and obtain

\[
E_0[\varphi (X_T)] = E_0 \left[ \frac{1}{2} e^{-Y_T} \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{g}(s)e^{isy} ds \right]
\]

\[
= \frac{1}{4\pi} \int_{-\infty}^{+\infty} \frac{\pi}{\cosh(\frac{\pi s}{2})} E \left[ e^{(-1+i)sY_T} \mid X_0 \right] ds
\]

\[
= \frac{1}{4} \int_{-\infty}^{+\infty} \frac{1}{\cosh(\frac{\pi s}{2})} e^{\frac{1}{2}(-1+is)\beta_0} CCF \left( T, -\frac{1}{2} (i + s) \beta_1; X_0 \right) ds,
\]

(6)
where for the second equality we assume that the order of integrals can be exchanged; the third equality follows from applying the result in (4). Now all that remains for computing the expected recovery rate is to evaluate a 1-dimensional integral (regardless of the dimension of $X_t$), which is a significant simplification compared to the direct approach.

If $X_t$ follows an affine process, its conditional characteristic function takes a particularly simple form: it is an exponential affine function of $X_0$. In some cases the exact form is known in closed form. Even in the general case where no closed-form solution is available, the affine coefficients are simple to compute as the solutions to differential equations. In those cases, our approach again offers a great deal of simplicity in the face of a possible curse of dimensionality: the linear scaling involved in solving $N$ ordinary differential equations is dramatically easier than solving an $N$-dimensional partial differential equation.

### 3 Generalized transforms

We now present our theoretical results. Our results only require that the conditional characteristic functions of the underlying state variables are tractable and our results apply whether the stochastic process is modeled in discrete-time or continuous-time. We will start with the case of continuous-time affine jump-diffusion (AJD), because its conditional characteristic function is particularly easy to compute, and because AJDs have been widely used in economics and finance. In Section 3.2 we discuss examples of processes that are not continuous time affine jump-diffusions.

We begin by fixing a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an information filtration $\{\mathcal{F}_t\}$, satisfying the usual conditions (see e.g., Protter (2004)), and suppose that $X$ is a Markov process in some state space $D \subset \mathbb{R}^N$ satisfying the stochastic differential equation

$$dX_t = (K_0 + K_1 X_t)dt + \sigma_t dW_t + dZ_t,$$

where $W$ is an $\mathcal{F}_t$-standard $n$-dimensional Brownian motion, $Z$ is a pure jump process with arrival intensity $\lambda_t = \ell_0 + \ell_1 \cdot X_t$ and fixed $D$-invariant distribution $\nu$, and $(\sigma_t \sigma'_t)_{i,j} = H_0_{i,j} + \sum_k H_1_{i,j,k} X^k_t$ with $H_0 \in \mathbb{R}^{N \times N}$ and $H_1 \in \mathbb{R}^{N \times N \times N}$. 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 parameters of the process \((K_0, K_1, H_0, H_1, \ell_0, \ell_1, \nu, \rho_0, \rho_1)\).

In order to establish our main result, let us first review some basic concepts from distribution theory. A function \( f : \mathbb{R}^N \to \mathbb{R} \) which is smooth and rapidly decreasing in the sense that for any multi-index \( \alpha \) and any \( P \in \mathbb{N} \), \( \|f\|_{P,\alpha} \equiv \sup_x |\partial^\alpha f(x)|(1 + \|x\|)^P < \infty \) is referred to as a Schwartz function. Here, \( \partial^\alpha f \) denotes the higher-order mixed partial of \( f \) associated with the multi-index \( \alpha = (\alpha_1, \ldots, \alpha_N) \) (that is, \( \partial^\alpha f = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_N}}{\partial x_N^{\alpha_N}} f \)), and \( \|x\| \) is the Euclidean norm of the vector \( x \). The collection of all Schwartz functions is denoted \( \mathcal{S} \), and \( \mathcal{S} \) is endowed with the topology generated by the family of semi-norms \( \|f\|_{P,\alpha} \). The dual of \( \mathcal{S} \), denoted \( \mathcal{S}^* \) and also called the set of tempered distributions, is the set of continuous linear functionals on \( \mathcal{S} \). Any continuous function \( g \) 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 \( f \mapsto \langle g, f \rangle \), where we use the inner-product notation

\[
\langle g, f \rangle = \int_{\mathbb{R}^N} g(x)f(x)dx. \tag{8}
\]

As is standard, we maintain the inner product notation even when a tempered distribution \( g \) does not correspond to a function as in \((8)\). For example, the \( \delta \)-function is a tempered distribution given by \( \langle \delta, f \rangle = f(0) \) which does not arrive from a function. Throughout the paper, we use the notation \( \delta(s) \) to denote the Dirac delta function, with \( \delta_x(s) = \delta(s-x) \).

For our considerations, the key property is that the set of tempered distributions is suitable for Fourier analysis. For any Schwartz function \( f \), the Fourier transform of \( f \) is another Schwartz function, denoted \( \hat{f} \), and is defined by

\[
\hat{f}(s) = \int_{\mathbb{R}^N} e^{-is \cdot x} f(x)dx. \tag{9}
\]

The Fourier transform can be inverted through the relation

\[
f(x) = \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} e^{i x \cdot s} \hat{f}(s)ds, \tag{10}
\]

which holds pointwise for any Schwartz function. The Fourier transform is extended to apply to
tempered distributions through the definition \( \langle \hat{g}, f \rangle = \langle g, \hat{f} \rangle \). This extension is useful because many functions define tempered distributions (through (8)), but do not have Fourier transform in the sense of (9) because the integral in is not well-defined. An example is the Heaviside function:

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

where integrating against \( 1/s \) is to be interpreted as the principal 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 one main result:

**Theorem 1.** Suppose that \( g \in S^* \) and \((\Theta, \alpha, \beta)\) satisfies Assumption 1 and Assumption 2 in Appendix A. Then

\[
H(g, \alpha, \beta) = \mathcal{E}_0 \left[ \exp \left( - \int_0^T R(X_u)du \right) e^{\alpha \cdot X_T} g(\beta \cdot X_T) \right]
= \frac{1}{2\pi} \langle \hat{g}, \psi(\alpha + \beta i) \rangle,
\]

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

\[
s \mapsto \psi(\alpha + s\beta i) = \mathcal{E}_0 \left[ e^{-\int_0^T R(X_u)du} e^{(\alpha + is\beta) \cdot X_T} \right].
\]

In the case where \( X \) follows an affine jump-diffusion, the discounted conditional characteristic function \( \psi \) is given in DPS,

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

and \( A, B \) are solutions to a system of ordinary differential equations (ODEs) that can generally be computed easily (see Appendix A for more details).
In the special case that \( \hat{g} \) defines a function, we can write the result as

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

(15)

Bakshi and Madan (2000) show that option-like payoffs (affine translations of integrable functions) can be spanned by a continuum of characteristic functions. Theorem 1 above shows that the characteristic functions span a much larger set of functions which includes all tempered distributions. Equation (15) makes this spanning explicit for the cases where \( \hat{g} \) defines a function.

There is some flexibility in the choice of \( \alpha \) and \( g \) in (12). Notice that

\[
e^{\alpha \cdot X_T} g(\beta \cdot X_T) = e^{(\alpha - c\beta) \cdot X_T} \tilde{g}(\beta \cdot X_T),
\]

where \( \tilde{g}(s) = e^{cs} g(s) \). This property can be useful in the case where \( g \) is not integrable but decreases rapidly as \( s \) approaches either positive or negative infinity (e.g., the logit function). In this case, such a transformation of \( g \) makes it possible to apply (15).

### 3.1 Two Extensions

The result of Theorem 1 can be extended in a number of ways. First, we introduce a class of \textit{pl-linear} (polynomial-log-linear) functions:

\[
f(\alpha, \gamma, p, X) = \sum_i p_i X^{\gamma_i} e^{\alpha_i \cdot X},
\]

(17)

where \( \{p_i\} \) are arbitrary constants, \( \{\alpha_i\} \) are complex vectors and \( \{\gamma_i\} \) are arbitrary multi-indices so that \( X^\gamma = \prod_j X_j^{\gamma_j} \). For example with \( N = 3 \) and \( \gamma = (1, 2, 1) \), \( X^\gamma = X_1^1 X_2^2 X_3^1 \). The following proposition extends Theorem 1 to work with any \textit{pl-linear} functions.

**Proposition 1.** Suppose that \( g \in S^* \), and \( (\Theta, \alpha, \beta, \gamma) \) satisfies Assumption 1’ and Assumption 2’
in Appendix B. Then

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

(18)

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

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

(19)

The function \( \psi \) is computed by solving the associated ODE in Appendix B.

It is immediate from Proposition 1 that we can now compute expectations of the form

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

(20)

The assumption that the function \( g \) in the generalized transform be a tempered distribution might appear restrictive at first sight, since \( g \) cannot have exponential growth (see our earlier discussions of Schwartz functions). However, as Proposition 1 demonstrates, by specifying \( f \) and \( g \) appropriately, we can let \( f \) “absorb” any exponential or polynomial growth in a moment function, rendering \( g \) admissible to the transform. We will demonstrate this feature in several examples.

The transform in Theorem 1 assumes that \( g \) can only depend on \( X \) through the linear combination \( \beta \cdot X \). Thus, the marginal impact of \( X_i \) on \( g \) will be proportional to \( \beta_i \), which might be too restrictive in some cases. The following proposition relaxes this restriction by considering \( g(\beta_1 \cdot X, \cdots, \beta_M \cdot X) \) for \( M \in \mathbb{N} \).

**Proposition 2.** Suppose that \( g \in S_M^* \) (an \( M \)-dimensional tempered distribution), \( \alpha \in \mathbb{R}^N \), \( b \in \mathbb{R}^{M \times N} \) and \( (\Theta, \alpha, b) \) satisfies Assumption 1 and Assumption 2 in Appendix A. Then

\[
H(g, \alpha, b) = E_0 \left[ \exp \left( - \int_0^T R(X_u) du \right) e^{\alpha \cdot X_T} g(bX_T) \right]
= \frac{1}{(2\pi)^M} \langle \hat{g}, \psi_M(\alpha + \cdot b i) \rangle,
\]

(21)
where $\hat{g} \in S^*$ and $\psi_M(\alpha + \cdot b_i)$ denotes the function

$$\psi_M : \mathbb{C}^M \to \mathbb{C}, \ s \mapsto \psi_M(\alpha + s^\top b_i) = E_0 \left[ e^{-\int_0^T r(X_u)du} e^{(\alpha + is^\top b_i)^\top X_T} \right]. \quad (22)$$

It is immediate to extend the transform in Proposition 2 by replacing $e^{\alpha \cdot X_T}$ with a \textit{pl-linear} function as in Proposition 1.

Fourier transforms of many functions are known in closed form (see for example, Folland (1984)). Additionally, standard rules allow for differentiation, integration, product, convolution and other operations to be conducted while maintaining closed-form expressions. Even if the function $\hat{g}$ is not known in closed form, including those cases where $g$ itself is given as an implicit function, it is straightforward to compute numerically (a 1-dimensional integral in the case of Theorem 1 or Proposition 1, an M-dimensional integral in Proposition 2). Alternatively, one might consider approximating $g$ with a function $\tilde{g}$ for which the Fourier transform is known in closed form.

For a given set of parameters, the Fourier transform of $g$ and the coefficients $A$ and $B$ in (14) need only be computed once. Once computed, the Fourier transform and the differential equation solutions can be used repeatedly to compute moments with different initial values of the state variable $X_0$ or horizon $T$. When the moment function takes the form of $f(\alpha, \gamma, p, X) g(\beta \cdot X)$ as in Proposition 1, the same Fourier transforms and differential equation solutions can also be used to compute moments with different \textit{pl-linear} function $f$.

### 3.2 Beyond Affine Jump Diffusions

The key aspects of Theorem 1 and the two extensions are the ability to compute the transform given in (13), (19), or (22). These transforms are very tractable for affine jump-diffusions. However, other stochastic processes can also be suitable for the generalized transform, provided that the appropriate (forward) conditional characteristic function can be computed. One example is the discrete time affine processes. Appendix C presents the generalized transform result in discrete time.

Another important example is the class of Lévy processes (see for example, Protter (2004)). Lévy processes allow for both finite and infinite activity jumps, though in some contexts the assumption
of independent increments may be restrictive. To be concrete, consider for example the process of Carr, Geman, Madan, and Yor (2002). They specify a pure jump Lévy process with Lévy measure $\nu$ given by the density $k_{CGMY}$, where

$$k_{CGMY}(x) = \begin{cases} Ce^{-G|x| - 1 - Y} & \text{if } x > 0 \\ Ce^{-M|x| - 1 - Y} & \text{otherwise} \end{cases}$$

where $(C,G,M,Y)$ are constants. When $Y = -1$, this reduces to i.i.d. jump arrivals with an exponential distribution. When $Y = 0$, we recover the variance gamma process studied by Madan, Carr, and Chang (1998). Generally, the CGMY process allows for flexibility in modeling the activity of small and large jumps as well as in the tail properties for large jumps.

The Lévy-Khintchine formula allows us to recover the conditional characteristic function from an arbitrary Lévy measure. Carr, Geman, Madan, and Yor (2002) show that if $X_t$ is CGMY process, then the characteristic function is

$$E_0[e^{iu(X_t - X_0)}] = \exp \left( tC\Gamma(-Y)((M - iu)^Y - M^Y + (G + iu)^Y - G^Y) \right), \quad (23)$$

where $\Gamma$ denotes the standard Gamma function: $\Gamma(t) = \int_0^\infty s^{t-1}e^{-s}ds$. Using the above characteristic function, we can then apply the results of Theorem 1 and Proposition 1 to compute nonlinear moments of Lévy processes.

Other examples of non-AJD processes with tractable conditional characteristic functions include the Markov-switching affine process and the discrete time autoregressive process with gamma-distributed shocks. As shown in Dai, Singleton, and Yang (2007) and Ang, Bekaert, and Wei (2008), one can incorporate regime shifts in the conditional mean, conditional covariance, or the conditional probability of jumps into standard AJDs. Bekaert and Engstrom (2010) show that the autoregressive processes with gamma-distributed shocks provide a convenient way to generate time-varying skewness and kurtosis. In both examples, the conditional characteristic functions can be computed easily.

Having presented the theory of the generalized transform, next we illustrate its power in pricing
contingent claims and in solving equilibrium asset pricing models.

4 Applications in Contingent Claim Pricing

The primary example we study in this section is pricing defaultable bonds with stochastic recovery rates. Recovery rates may depend nonlinearly on the state of the economy which potentially makes the recovery function non-integrable or non-smooth. Our method can easily handle such models. We show that stochastic variations in the recovery rates not only can have large effects on the average level of credit spreads, but also can lead to economically important non-linearities in credit spreads as default intensity changes, which are difficult to capture using standard models with constant recovery rate. We complete the section by comparing our method for risk-neutral pricing of contingent claims with some alternative methods.

4.1 Stochastic Recovery

Following up on the illustrating example in Section 2, we now examine the pricing of credit-risky securities (e.g., defaultable bonds or credit default swaps) with stochastic recovery upon default. Investors will demand a recovery risk premium if the recovery rate of defaulted securities tend to be lower during aggregate bad times, which has been documented in several studies. For example, Altman, Brady, Resti, and Sironi (2005) document significant negative correlation between aggregate default rates and recovery rates. Acharya, Bharath, and Srinivasan (2007) and Chen (2010) also find evidence that recovery rates and default rates of corporate bonds are significantly related to industry and macroeconomic conditions.

Consider a $T$ year defaultable zero-coupon bond with face value normalized to 1. Following the literature of reduced-form credit risk models, the default time is assumed to be a stopping time $\tau$ with risk-neutral intensity $\lambda_t$. The risk-neutral recovery rate at default $\tilde{\varphi}_t$ is a bounded predictable process that is adapted to the filtration $\{\mathcal{F}_t : t \geq 0\}$. The instantaneous riskfree rate is $r_t$. Then, to be precise, we fix a probability space $(\Omega, \mathcal{F}, P)$ and two filtrations $\{\mathcal{F}_t : t \geq 0\}, \{\mathcal{G}_t : t \geq 0\}$. The default time is a totally inaccessible $\mathcal{G}$-stopping time $\tau : \Omega \to (0, +\infty]$. We assume that under the risk neutral measure $Q$, $\tau$ is doubly-stochastic driven by the filtration $\{\mathcal{F}_t : t \geq 0\}$. See Duffie (2005) for a survey on the reduced form approach for modeling credit risk and the doubly-stochastic property.
the price of the bond is:

\[
V_t = E_t^Q \left[ e^{-\int_t^T r_u du} 1_{\{\tau \leq T\}} \tilde{\varphi}_\tau \right] + E_t^Q \left[ e^{-\int_t^T r_u du} 1_{\{\tau > T\}} \right] + E_t^Q \left[ \int_t^T e^{-\int_s^T (r_u + \lambda_u) du} \lambda_s \tilde{\varphi}_s ds + e^{-\int_t^T (r_u + \lambda_u) du} \right].
\]

The second equality follows from the doubly-stochastic assumption and regularity conditions. Let \( X_t \) be the vector of state variables that determine the riskfree rate, the default intensity, and the recovery rate. Suppose that both \( r_t \) and \( \lambda_t \) are affine in \( X_t \). In addition, we model the risk-neutral recovery rate as \( \tilde{\varphi}_t = g(\beta \cdot X_t) \) for some proper function \( g \), which ensures that the recovery rate is between 0 and 1 in addition to satisfying a suitable no-jump condition.\(^7\)

To investigate the quantitative impact of stochastic recovery on the pricing of defaultable bonds, we directly specify the dynamics of state variables \( X_t = [\lambda_t \ Y_t]' \) under the risk neutral measure \( Q \) as follows:

\[
d\lambda_t = \kappa_\lambda (\theta_\lambda - \lambda_t) dt + \sigma_\lambda \sqrt{\lambda_t} dW_\lambda^\lambda,
\]

\[
dY_t = \kappa_Y (\theta_Y - Y_t) dt + \sigma_Y \sqrt{\lambda_t} dW_Y^Y,
\]

where \( W_\lambda^\lambda \) and \( W_Y^Y \) are uncorrelated Brownian motions. The riskfree rate is given by

\[
r_t = Y_t - \delta \lambda_t.
\]

This simple setup (with \( \delta > 0 \)) captures the negative correlation between \( r_t \) and \( \lambda_t \) in the data.

With the help of the generalized transform, we now have a lot of flexibility in choosing the recovery function \( \tilde{\varphi}_t \) and still maintain tractability for pricing. For example, the cumulative distribution function (CDF) of any probability distribution, such as the logit or probit model, will take values in \([0,1]\) and can be used to model the recovery rate. Modeling \( \tilde{\varphi} \) with CDFs has the added benefit of having nice Fourier transform properties. For example, the integrands of the Gaussian and Cauchy model have closed-form Fourier transform. Since Fourier transform has the property that

\(^7\)The no-jump condition is satisfied here by assuming \( \varphi \) is adapted to \( \{F_t\} \). See also Duffie, Schroder, and Skiadas (1996) and Collin-Dufresne, Goldstein, and Hugonnier (2004) for discussions on the no-jump condition.
\( \hat{f}'(t) = t \hat{f}(t) \), it is very easy to obtain the Fourier transform of the CDF in such cases.\(^8\)

For simplicity, we assume that \( \tilde{\varphi} \) only depends on the default intensity, and we adopt a modified Cauchy model:

\[
\tilde{\varphi}(\lambda) = \frac{a}{1 + b(\lambda - \lambda_0)^2} + c. \tag{28}
\]

The constant term \( c \in [0, 1] \) sets a lower bound for \( \tilde{\varphi} \) that is potentially above 0, which gives us more flexibility in matching the empirical distribution of recovery rates. The Fourier transform of \( \tilde{\varphi} \) (excluding the constant \( c \)) is

\[
\hat{\tilde{\varphi}}(t) = \frac{a \pi}{\sqrt{b}} e^{\lambda_0 it - \frac{1}{\sqrt{b}} |t|}. \tag{29}
\]

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

\[
E_Q^0 \left[ \exp \left( - \int_0^t (r_u + \lambda_u)du \right) \lambda_t \tilde{\varphi}(\lambda_t) \right],
\]

which is mapped into the generalized transform of Theorem 1 by choosing

\[
f(\alpha \cdot X) = \iota_1 \cdot X,
\]

\[
g(\beta \cdot X) = \frac{a}{1 + b(\iota_1 \cdot X - \lambda_0)^2} + c,
\]

where \( \iota_1 = [1 \ 0]' \). Notice that one can introduce additional state variables in \( X \) to capture richer dynamics of the term structure, default risk, recovery rate, and macroeconomic conditions, which does no add any complication to pricing as long as the risk-neutral recovery rate is still given by \( \tilde{\varphi}_t = g(\beta \cdot X_t) \).

We now use the processes of default intensity \( \lambda_t \) and riskfree rate \( r_t \) (25–27) and the recovery model (28) to price a 5-year defaultable zero-coupon bond. We calibrate the process of \( \lambda_t \) and \( Y_t \) under the risk-neutral measure following Duffee (1999). The parameter values are reported in Table 1. Notice that \( \kappa_\lambda < 0 \), which is consistent with Duffee’s finding that the default intensity of a typical firm is nonstationary under the risk-neutral measure.

---

\(^8\)There are also specifications where the existing methods apply, for example, \( \tilde{\varphi}(X) = e^{\beta \cdot X} 1_{\{\beta \cdot X < 0\}} + 1_{\{\beta \cdot X > 0\}} \). Bakshi, Madan, and Zhang (2006) study such a setting. However, our method is more general. For example, a power law specification such as \( \tilde{\varphi}(X) = (\beta \cdot X)^{-\alpha} + c \) for \( \alpha, c > 0 \) falls under our theory but wouldn’t directly be solvable with existing methods.
We consider two calibrations of $\tilde{\varphi}(\lambda)$ in (28). First, we directly calibrate the risk-neutral recovery rate to the actual recovery rates. Using Moody’s data on aggregate annual default rates and recovery rates, we calibrate $a = 0.68$, $b = 2000$, $c = 0.25$, and $\lambda_0 = -0.014$. The fitted function is “Model I” in Figure 1. Fitting the risk-neutral recovery rate to the actual recovery rate amounts to assuming no recovery risk premium, and by treating the physical and risk-neutral default intensity as the same we are also assuming there is no jump-to-default risk premium. The fitted curve is downward sloping and convex. The recovery rate is close to 70% when the default intensity is low. When the default intensity rises to 10%, the recovery rate drops to 30%. In the second calibration, we assume $a = 0.9$, $b = 1200$, $c = 0$, and $\lambda_0 = -0.014$. The fitted function is “Model II” in Figure 1, which has very similar recovery rates to Model I when the default intensity is low, but has a sharper decline in recovery rates than Model I as default intensity rises. The widening gap between the two models implies that the recovery risk premium in Model II is increasing with the aggregate default probability.

A popular assumption for default recovery in both academic analysis and industry practice is that the risk-neutral recovery rate is constant, and an often-used value is 25%, see e.g., Pan and Singleton (2008). This value is lower than the historical mean recovery rate, which is a parsimonious way to capture the recovery risk premium. We compare our stochastic recovery model with a model with 25% constant recovery rate. In addition, we consider another constant recovery rate which is calibrated to best fit the stochastic recovery models (by minimizing the mean square error computed based on the empirical distribution). Finally, we have also considered the analogous recovery of market value models, but as in Duffie and Singleton (1999) we found that in our calibration the credit spreads in the constant recovery of market value models closely match those in the constant recovery of face value models, so we omit them form our comparison.

In Figure 2, the top panels investigate Stochastic Recovery Model I (no recovery risk premium)
Figure 1: A Cauchy Model of Aggregate Recovery Rates. This figure plots the historical aggregate recovery rates and default rates from Altman and Kuehne (2011) for the period 1982-2010. The solid line is a Cauchy recovery model fitted to the historical data. The dash line is a Cauchy recovery model with recovery risk premium.

while the bottom panels consider Stochastic Recovery Model II (risk adjustment for recovery risk). All the results are computed with the riskfree rate fixed at 5%. We see from Panel A and Panel C that the credit curve is almost linear (with a small amount of concavity) in the default intensity both for a constant recovery rate of 25% and for the constant recovery rate calibrated to best match the stochastic recovery models (29.6% and 8.5% for model I and model II, respectively). Changing the recovery rate primarily results in a change in the steepness of the credit curve with respect to the default intensity. In contrast, the stochastic recovery curves exhibit a fair amount of non-linearity, with some convexity in the region of small default probabilities. The convexity is caused by the fact that when default intensities are low (high), recoveries are high (low) so the incremental effect of an increase in default probabilities is small (large); thus the curve becomes steeper as the intensity increases. This non-linearity is more pronounced in Model II due to the faster decline of the recovery rate with default intensity. Thus we see that stochastic recovery introduces non-linearities in the credit curve (as a function of default intensity) which are qualitatively different from the near-linear

9For a recovery of market value model, the yields are exactly affine in the default rate, as shown by Duffie and Singleton (1999).
Figure 2: **Credit spreads for 5-year bonds with constant recovery and Cauchy recovery.** For different values of conditional default intensity, this figure plots the credit spreads of a 5-year zero-coupon defaultable bond, and the pricing errors of the RMV and RFV model with constant recovery rates relative to two versions of the stochastic recovery model. “RMV” stands for “recovery of market value”; “RFV” stands for “recovery of face value”.

Panels B and D assess the economic importance of stochastic recovery by plotting the pricing errors of constant recovery models relative to the stochastic recovery models. We see that there are quite large differences between the fixed recovery rate of 25% specification and the two stochastic recovery models: the root mean square difference across the given range of intensities are 38 bp and 206 bp for Model I and II, respectively. On the one hand, relative to Model I (without recovery risk premium), the 25% constant recovery assumption can lead to overstating the credit spread by up to 50 bps for moderate default intensities. On the other hand, relative to Model II, a 25% constant recovery rate becomes too conservative and leads us to understate the credit spread most of the
time, where the pricing errors can be as large as 400 bps.

Re-calibrating the constant recovery rate for each model produces a closer fit. The optimized recovery rates result in root mean square differences of 26 bp and 51 bp relative to Model I and II. However, the yield differences are still economically significant most of the time, with the constant recovery model typically overstating the credit spread for low default intensity and understating the spread for high intensity. These results highlight the importance of carefully incorporating stochastic recovery rates into credit risk modeling.

4.2 Comparison with Some Alternative Methods

We now discuss how the generalized transform method in this paper differ from the methods developed by Duffie, Pan, and Singleton (2000) and Bakshi and Madan (2000). We will use the example of pricing an European option to illustrate some of the key differences.

Again, we assume that \( X_t \) is the vector of state variables that follows an affine process under the risk-neutral probability measure \( Q \), and that the instantaneous riskfree rate satisfies \( r_t = \rho_0 + \rho_1 \cdot X_t \). Consider the example of an European put option with strike \( K \). Let \( \beta \cdot X_t \) be the log stock price, then the payoff is \((K - e^{\beta \cdot X_T})^+\). Define \( g(\beta \cdot X) \equiv 1_{\{\beta \cdot X \leq \log K\}} \). The option price can be written as

\[
P_t = E_t^Q \left[ e^{-\int_t^T r_s ds + \beta \cdot X_T} \right] - E_t^Q \left[ e^{-\int_t^T r_s ds + \beta \cdot X_T} g(\beta \cdot X_T) \right],
\]

which can be computed by applying Theorem 1. In this case, the Fourier transform of \( g \) is defined as a distribution:

\[
\hat{g}_y(s) = \pi \delta(s) + \frac{i e^{-isy}}{s},
\]

where the second term is interpreted as a principal value integral. It follows that

\[
E_t^Q \left[ e^{-\int_t^T r_s ds + \alpha \cdot X_T} g_y(\beta \cdot X_T) \right] = \int_{-\infty}^{\infty} \left( \frac{\delta(s)}{2} - \frac{e^{-isy}}{2 \pi is} \right) \psi(\alpha + is\beta) ds
\]

\[
= \frac{\psi(\alpha)}{2} - \int_{0}^{\infty} \text{Real} \left( \frac{\psi(\alpha + is\beta) e^{-isy}}{\pi is} \right) ds.
\]

In the last equation we use the fact that the real part of the integrand is even and the imaginary
part is odd.

The above result replicates the formula given in DPS obtained by Lévy-inversion. However, their results are limited to the case of affine jump-diffusion and payoffs that are \textit{pl-linear} in the underlying state variable, which is a special case of \textbf{Theorem 1}. DPS 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 be reduced to a 1-dimensional integral as in the standard Lévy inversion formula (without a forward measure).

\textbf{Bakshi and Madan (2000)} (henceforth BM) provide a more general method for pricing options. It allows for payoffs of the form \((H(X_T) - K)^+\), where \(X\) is a univariate stochastic process with known conditional characteristic function under the risk-neutral measure, \(H\) is a positive and entire function (analytic at all finite points on the complex plane), and \(K\) is a fixed strike price. BM propose a power series expansion of \(H\), the expectations of which can then be computed through differentiation of the conditional characteristic functions. Their method applies to non-affine processes as well.

Our method differs from BM in several aspects. First, the power series expansion approach in BM requires that \(H\) be infinitely differentiable and that the power series converges to \(H(X)\) for all \(X\). However, there are cases where the payoff function has many kinks or is not entire (simple examples include \(\ln(X)\) and \(\sqrt{X}\)). Second, in those cases where the Fourier transform \(\hat{g}\) is known, our method requires only one 1-dimensional integration, whereas the method of BM requires one 1-dimensional integration and one infinite sum. Third, we extend BM’s result on spanning of option payoff with characteristic functions. In BM, the set of payoff functions \(H\) that can be spanned by characteristic functions is an affine translation of an \(L^1\) function (see BM Theorem 1).\footnote{A function \(H\) is of class \(L^p\) if \(\left(\int |H(x)|^p dx\right)^{1/p} < \infty\).} We relax the growth condition on the underlying payoff; the dual space \(S^*\) is quite large and contains \(L^p\) for any \(p\), as well as functions not in \(L^p\) for any \(p\). Finally, our theory extends the analysis to multivariate settings.

To illustrate some of these differences, consider the payoff of the form \(H(X) = X^\alpha\) for a positive non-integer \(\alpha\). In this case \(H\) is not entire (for any choice of the center of the power series, \(X_0\),...
the power series converges to $H(X)$ only for $0 < X < 2X_0$). Nor is it an affine translation of an $L^1$ function. In such cases, if one were to compute option prices through the Taylor expansion of the non-entire payoff function with a particular choice of $X_0$ and the order of Taylor expansion, the error could be very large. However, we can still use the fact that $H$ represents a tempered distribution to write

$$
\hat{g}(s) = (is)^{-1-\alpha} \Gamma_{\text{inc}}(\alpha + 1, isk) - Ke^{-iks}\hat{H}(s),
$$

(32)

where $k = K^{1/\alpha}$, $\hat{H}$ is given by (11), $\Gamma_{\text{inc}}(\cdot, \cdot)$ denotes the incomplete Gamma function and $s^{-1-\alpha}$ is interpreted in the sense of a homogeneous tempered distribution.

5 Applications in Economic Modeling

In addition to risk-neutral pricing, the generalized transform can also be a powerful tool in economic modeling. For example, the need to compute nonlinear moments arises naturally when we price assets using the stochastic discount factor under the physical measure $\mathbb{P}$. Suppose the state variables are affine under $\mathbb{P}$, and the underlying economic model gives rise to a stochastic discount factor $m(t, X)$. Then, the present value of a stochastic payoff $Y_T = y(X_T)$ at time $T$ is

$$
P_t = \frac{1}{m(t, X_t)} \mathbb{E}_t^\mathbb{P}[m(T, X_T)y(X_T)].
$$

(33)

Except for some special cases (when $m_t$ is an exponential affine function of $X_t$), the dynamics of $X_t$ under the risk-neutral measure $\mathbb{Q}$ can be quite complicated, and the riskfree rate $r_t$ may not be affine in $X_t$, making it difficult to do pricing under $\mathbb{Q}$. Instead, prices can be more easily computed under $\mathbb{P}$ using the generalized transform, provided that $m(T, X)y(X)$ can be decomposed into $f(X)g(\beta \cdot X)$ as in Section 3.

Two important classes of models where such non-pl-linear stochastic discount factors arise are in general equilibrium models where: (1) there are heterogeneities in the sources of income or in the cross section of stocks, or (2) there are heterogeneities across agents in terms of preferences or beliefs. We focus on the first case in this section, and leave the analysis of models with heterogeneous agents
Several recent papers have studied the general equilibrium effects of multiple sources of income (consumption) and demonstrated their importance for understanding the time series and cross section of asset prices. See Santos and Veronesi (2006) (analytical result for a model with financial asset and labor income that satisfy special co-integration restrictions), Piazzesi, Schneider, and Tuzel (2007) (numerical solution for a model with housing and non-housing consumption), Cochrane, Longstaff, and Santa-Clara (2008) (analytical result for a model with two i.i.d. trees and log utility), and Martin (2011) (analytical result for $N$ i.i.d. trees and power utility). In the following example, we use the generalized transform to obtain analytical results in a model with power utility and non-i.i.d. trees.

Suppose there are two assets (or two types of consumption goods) in the economy, both in unit supply, with dividends paying out continuously at rates $D_{1,t}$ and $D_{2,t}$. We assume that the log dividends $d_{1,t} = \log D_{1,t}$ and $d_{2,t} = \log D_{2,t}$ are part of a vector $X_t$ ($d_{1,t} = \iota_1 \cdot X_t, d_{2,t} = \iota_2 \cdot X_t$, with $\iota_1 = [1 \ 0 \ 0 \ \cdots]^\prime$ and $\iota_2 = [0 \ 1 \ 0 \ \cdots]^\prime$), which follows an affine jump-diffusion (7). This model can allow for time variation in the expected dividend growth rates, stochastic volatility, and time variation in the probabilities of jumps. Co-integration restriction can also be imposed to allow for stationary of the shares of the two assets.

There is an infinitely-lived representative investor with CRRA utility over aggregate consumption:

$$U(c) = E^P_0 \left[ \int_0^\infty e^{-\rho t} \frac{C_t^{1-\gamma} - 1}{1-\gamma} dt \right], \quad (34)$$

where $\gamma$ is the coefficient of relative risk aversion and $\rho$ is the time discount rate. Aggregate consumption is a CES aggregator of the two goods $D_{1,t}$ and $D_{2,t}$,

$$C_t = \left( D_{1,t}^{(\epsilon-1)/\epsilon} + \omega D_{2,t}^{(\epsilon-1)/\epsilon} \right)^{\epsilon/(\epsilon-1)}. \quad (35)$$

The parameter $\epsilon$ is the elasticity of intratemporal substitution between the two goods, and $\omega$ determines the relative importance of the two goods.

We can recover the continuous time version of Piazzesi, Schneider, and Tuzel (2007) when we
interpret \( D_{1,t} \) and \( D_{2,t} \) as housing and nonhousing consumption, respectively, and assume that the growth rate of nonhousing consumption is i.i.d. and that the log ratio of the two dividends follows a square-root process. In the case \( \epsilon \to \infty \) and \( \omega = 1 \), the two goods become perfect substitutes, so that \( C_t = D_{1,t} + D_{2,t} \). This is the case considered by Cochrane, Longstaff, and Santa-Clara (2008) and Martin (2011), both of which assume i.i.d. dividend growth.

In equilibrium, there is a unique stochastic discount factor \( m_t = e^{-\rho t}C_t^{-\gamma} \). Under the standard regularity conditions, the price of asset \( i \) \( (i = 1, 2) \), \( P_{i,t} \), is then given by

\[
P_{i,t} = \mathbb{E}^P \left[ \int_0^\infty \frac{m_{t+u}}{m_t} D_{i,t+u} du \right]
\]

\[
= \left( D_{1,t}^{(\epsilon-1)/\epsilon} + \omega D_{2,t}^{(\epsilon-1)/\epsilon} \right)^{\gamma\epsilon/(\epsilon-1)} \int_0^\infty e^{-\rho u} \mathbb{E}^P \left[ \frac{D_{i,t+u}}{D_{1,t+u}^{(\epsilon-1)/\epsilon} + \omega D_{2,t+u}^{(\epsilon-1)/\epsilon}} \right]^{\gamma\epsilon/(\epsilon-1)} du. \tag{36}
\]

The main challenge of computing the stock price comes from the stochastic discount factor, which is non-pl-linear in the state variable \( X_t \). As a result, the riskfree rate is not affine in \( X_t \), and \( X_t \) is not affine under the risk-neutral measure. To map the expectation in (36) into the generalized transform, we rewrite the expectation inside the integral of (36) as

\[
\mathbb{E}^P \left[ \frac{D_{1,s}}{D_{1,s}^{(\epsilon-1)/\epsilon} + \omega D_{2,s}^{(\epsilon-1)/\epsilon}} \right]^{\gamma\epsilon/(\epsilon-1)} = \mathbb{E}^P \left[ \frac{e^{-\gamma/2}d_{1,s} - \gamma/2d_{2,s}}{2 \cosh \left( \frac{\epsilon-1}{\epsilon} d_{1,s} - \frac{\epsilon-1}{2\epsilon} d_{2,s} \right)} \right]^{\gamma}
\]

\[
= \mathbb{E}^P \left[ f (\alpha \cdot X_s) g (\beta \cdot X_s) \right], \tag{37}
\]

where

\[
f (x) = e^x, \quad g (x) = \frac{1}{(2 \cosh(x))^{\gamma}}
\]

and

\[
\alpha = \left( 1 - \frac{\gamma}{2} \right) t_1 - \frac{\gamma}{2} t_2, \quad \beta = \frac{\epsilon - 1}{2\epsilon} (t_1 - t_2).
\]

Since \( X \) is affine and \( g \in S^* \), Theorem 1 readily applies to (37). When the increments of \( X \) are
i.i.d., the conditional characteristic function for $X$ is known explicitly, which Martin (2011) uses to compute (37) following a Fourier transform for $g$.

Several observations are in order. First, introducing additional state variables to $X$ within the affine framework to capture richer dynamics of dividends (such as time-varying conditional moments of dividend growth) is quite straightforward. Due to the time-separable utility function, these additional state variables do not directly enter into the pricing equation (36) and thus will not result in a curse of dimensionality: the dimension of the problem remains exactly the same. Second, one can also further enrich the model by adding preference shocks that are $pl$-linear in the state variables, e.g., see the external habit models in Pástor and Veronesi (2005) or Bekaert and Engstrom (2010). See also Chen and Joslin (2011) for a general specification. Third, we can extend the model to have more than two assets, which can be solved using the multi-dimensional version of the generalized transform in Proposition 2.

5.1 A calibrated example: time-varying labor income risk

In this section, we study a general equilibrium model with time-varying labor income risk. This model not only serves as a concrete example of applying the generalized transform method for economic analysis, but also draws a number of new insights on how the time-varying covariance between labor income and dividends affects asset pricing.

In consumption-based asset pricing models, the covariance between shocks to consumption and cash flows of an asset is often a key determinant of the risk premium for the asset. As this covariance fluctuates over time, so will the implied risk premium. Santos and Veronesi (2006) (hereafter SV) point out a natural source of such time variation in the covariances via a composition effect: as the share of labor income in total consumption varies over time, so will the covariances between consumption and dividends, which in turn generates time-varying equity premium. Intuitively, higher labor income relative to dividends tends to make investors less sensitive to fluctuations in dividend income. Santos and Veronesi illustrate this point in a model with stationary labor share and multiple financial assets, which provides very convenient closed-form solutions for asset prices.

We plot in Figure 3 the share of labor income and lagged four-year cumulative returns of the
CRSP value-weighted market index. We use per-capita consumption (nondurables and services) from the BEA and labor income series constructed following Lettau and Ludvigson (2001). Following SV, the labor share is defined as the ratio of labor income to consumption and thus dividends are defined to be the difference between consumption and labor income. Consistent with the findings of SV, the labor share and lagged market return in Panel A of Figure 3 are negatively correlated, with an average correlation of $-0.35$. However, in the post-1990 period, the two series become positively correlated, which is opposite of what the composition effect implies. These results suggest that other covariates may be playing a role in determining the relationship between the labor share and the equity premium (see also Duffee (2005) and Kozhanov (2009) for related findings).

One example of such covariates is consumption volatility. Lettau, Ludvigson, and Wachter (2008) argue that declining macroeconomic volatility since the 1980s has played a key role in the decline of the equity premium. They estimate that a structural break occurred in consumption
volatility in 1991, which motivates our choice of the two subsamples. In Figure 4, we plot the volatility of consumption growth for non-overlapping 5-year periods using quarterly data from 1952 to 2010, and the correlation between labor income and dividends during the same periods. Interestingly, the correlation between labor income and dividends shows a similar pattern as volatility. It ranges from the peak of 0.1 in early 1980s to the low of $-0.9$ in the late 1990s, and it tends to rise when consumption volatility rises. Like changes in consumption volatility, changes in the correlation between labor income and dividends could also affect the equity premium. All else equal, consumption volatility should indeed rise with the correlation between its two components. However, consumption volatility could also change independently of the correlation, e.g., through time-varying share of labor income in consumption, or variations in the volatilities of labor income and dividends.

Motivated by these findings, we propose a simple model that captures these interesting dynamics of the conditional moments of labor income, dividends and consumption. The model is a special case of the two-asset model discussed previously, with dividends from the two assets interpreted as financial income (dividends) and labor income. We extend the models of Cochrane, Longstaff, and
Santa-Clara (2008) and Martin (2011) in two important dimensions: (1) we add a volatility factor, which simultaneously drives the conditional volatilities of labor income and dividends, as well as the correlation between the two; (2) we impose cointegration between labor income and dividends so that the long run stationarity of labor share. Our model also differs from SV in that labor share and the correlation between labor income and dividends can move independently.

Specifically, let log dividends and log labor income be $d_t$ and $\ell_t$, and let $V_t$ be a volatility factor. Suppose $X_t = (d_t, \ell_t, V_t)'$ follows an affine process

$$dX_t = \mu_t dt + \sigma_t dW_t.$$  

(38)

We assume that the conditional drift $\mu_t$ is given by

$$\mu_t = \begin{pmatrix} \bar{g} - a \kappa_s (\bar{\sigma} - (\ell_t - d_t)) \\ \bar{g} + (1 - a) \kappa_s (\bar{\sigma} - (\ell_t - d_t)) \\ \kappa_V (\bar{V} - V_t) \end{pmatrix}.$$

(39)

This formulation gives the same average growth rate of $\bar{g}$ for dividends and labor income. The second term in the growth rates of dividends and labor income allows for the shares of labor income and dividends to be stationary. To see this, consider the dynamics of the log labor income-dividend ratio $s_t = \ell_t - d_t$. Equation (39) implies that the drift of $s_t$ will be $\kappa_s (\bar{s} - s_t)$, with $\bar{s}$ being the long-run average of the log labor income-dividends ratio, and $\kappa_s \geq 0$ the speed of mean reversion. Thus, when the share of labor income relative to dividends is high, the expected growth rate of dividends will be higher than that of labor income, which causes $s_t$ to revert to its long-run mean. The opposite is true when the labor share is low. The parameter $a$ gives us additional flexibility in specifying the degree of time variation in the growth rates of dividends and labor income. For example, $a = 1$ implies that the expected labor income growth is constant over time. Equation (39) also implies that the volatility factor $V_t$ is stationary, with long-run mean $\bar{V}$ and speed of mean reversion $\kappa_V$. 

28
The conditional covariance of the factors is given by

\[ \sigma_t \sigma_t' = \Sigma_0 + \Sigma_1 V_t, \]  

(40)

where \( \Sigma_0 \) and \( \Sigma_1 \) can be any positive semi-definite matrices with the restriction \( \Sigma_{0,33} = 0 \) so that the volatility factor always remains positive. As \( V_t \) increases, the volatilities of dividends and labor income will increase. Moreover, the instantaneous correlation between dividends and labor income, \( \rho_t \), also varies with the volatility factor \( V_t \). The structure of a single volatility factor implies that the correlation and volatilities have to move in lock steps. While this feature is clearly restrictive, it allows us to capture the comovement between consumption volatility and correlation demonstrated in Figure 4. One can substantially relax the covariance structure using the Wishart process, which allows for fully stochastic covariance between labor income and dividends.

We calibrate the model to match moments for dividends and labor income in the data; for full details of model calibration see Appendix D. For simplicity, we assume that the shock to the volatility factor are uncorrelated with dividend and labor income shocks.

We have already shown how to price the stock and human capital using the generalized transform. To compute the expected excess returns and volatilities of the stock and human capital, we can consider them as portfolios of zero-coupon equities. The risk premium of the stock or human capital is then the value-weighted average of the risk premium for these zero-coupon equities. In general, the instantaneous expected excess return for any asset is determined by its exposure to the primitive shocks in the state variable \( X_t \) and the risk prices associated with these shocks, which in turn are determined by their covariances with the stochastic discount factor \( M_t \). Thus, by Itô’s Lemma, the expected excess return for any asset with price \( P_t = P(X_t, t) \) is given by

\[ E_t[R^t] = (\nabla_X \log M_t)' \sigma_t \sigma_t' (\nabla_X \log P_t), \]  

(41)

where \( \sigma_t \sigma_t' \), the time \( t \) covariance of the factors, is given in (40). Here \( (\nabla_X \log M_t)' \sigma_t \) in (41) gives the price of risk for all the shocks in \( X_t \), while \( (\nabla_X \log P_t)' \sigma_t \) gives the exposure that the asset has to these shocks. With power utility, only those shocks that directly affect consumption will
We now examine the risk premium on financial wealth and human capital. The risk premium for financial wealth depends on its exposure to both dividend and labor income shocks. First, the financial wealth claim is directly exposed to dividend shocks via its cash flows (dividends). Second, via the discount factor, it is exposed to both dividend and labor income shocks. The value of the dividends tree will also be sensitive to covariance shocks ($V_t$), but these shocks are not priced in our model by the assumption that they are not correlated with consumption. Both mechanisms will have an effect on the risk premium. For example, positive shocks to labor income will decrease the premium demanded for dividends, increase the risk-free rate (a higher share of the less volatile labor income tends to smooth consumption, reducing the precautionary savings motive), and increase expected future dividends due to the mean reversion in (39) (with $a > 0$). Our model incorporate all of these mechanisms to determine the risk premium for the dividend claim.

Figure 5 plots the conditional risk premium on financial wealth and human capital as functions of the labor share and correlation. The plot focuses on the region that is more relevant based on
the stationary distributions of the two variables. Both the labor share and stochastic covariance are important contributors to the risk premium. When volatilities are high and the correlation is less negative, the model generates significant composition effect. For example, when the conditional correlation between labor income and dividends $\rho_t = -0.1$, the conditional risk premium on financial wealth falls from 6.6% to 1.8% as labor share rises from 0.6 to 0.9. However, when $\rho_t = -0.8$, the risk premium essentially remains at zero for the same rise in labor share. What’s more, the risk premium on financial wealth is more sensitive to changes in volatility and correlation when labor share is low.

Why is the risk premium changing so little with the labor share when volatilities are low? First, as labor share rises, the composition effect tends to drive down the risk premium per unit of dividend risk, but this effect weakens as the volatility falls. At the same time, the price-dividend ratio (P/D) is rising (due to both lower risk free rate and higher expected dividend growth) and becomes more sensitive to changes in labor share, hence also more sensitive to labor income and dividend shocks (with opposite signs). Since the price of labor income risk is rising with higher labor share while the price of dividend risk is decreasing, the net effect via the price-dividend ratio tends to be offsetting the composition effect for sufficiently high labor share. Moreover, when volatility falls, the rise in the price of labor income risk accelerates, which strengthens the P/D effect. Under our parameterizations, when volatilities are sufficiently low, the two effects essentially cancel each other.

As for the risk premium on human capital, the composition effect is stronger when the correlation is more negative. For example, when $\rho_t = -0.8$, the premium on human capital rises from 0 to 1.8% as labor share rises from 0.6 to 0.9. As the correlation rises, the premium flattens and eventually becomes U-shaped in labor share.

Through the lens of our model, we can also analyze the comovement between the risk premium on financial wealth and human wealth. Previous studies have different findings when measuring the sign of this comovement in the data. For example, see Hansen, Heaton, Lee, and Roussanov (2007) and Lustig and Van Nieuwerburgh (2008). Cash flows from the claim on financial and human wealth are negatively correlated most of the time in our model. However, both positive and negative correlation between the risk premium on financial and human wealth can occur. The risk premium
on financial wealth and human wealth will be negatively correlated when the composition effect is the main driver of variations in risk premium over time. However, when changes in the volatilities and correlation become the main driver of variations in risk premium, the two risk premiums can become positively correlated.

This example illustrates the power of our method to generate new economic insights through the analysis of general equilibrium models. The general class of models that we study here of heterogeneous agents and heterogeneous goods can also be used to examine the general equilibrium effects of the cross section of stocks as well as international finance models. Our method makes many of these models tractable for analysis.

6 Conclusion

We provide analytical results for computing 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. By not resorting to an intermediate computation of the (forward) density, this method greatly reduces the dimensionality of such problems. Our method can also be applied to other processes with tractable characteristic functions, such as discrete time affine processes, Lévy processes, and Markov-switching affine processes.

We demonstrate the power of this method with two examples. First, we study the pricing of defaultable bonds with stochastic recovery. We show that not only does the commonly used constant recovery assumption lead to substantial pricing errors in comparison to the stochastic recovery model, but the latter exhibits important non-linearity that cannot be replicated by constant recovery models. In the second example, we apply the generalized transform method in a general equilibrium model of time-varying labor income risk. The model not only helps explain the changing predictive power of labor share with declining volatility, but also shows that the risk premium on financial wealth and human capital can be positively or negatively correlated, depending on whether variations in labor share or covariances are the main driver of risk premium.
References


Appendix

A Proof of Theorem 1

\[ \dot{B} = K_1^T B + \frac{1}{2} B^T H_1 B - \rho_1 + \ell_1(\phi(B) - 1) \quad B(0) = \alpha + i\beta, \quad (A1) \]

\[ \dot{A} = K_0^T B + \frac{1}{2} B^T H_0 B - \rho_0 + \ell_0(\phi(B) - 1) \quad A(0) = 0, \quad (A2) \]

where \( \phi(c) = E_\nu[e^{cZ}] \), the moment-generating function of the jump distribution and \((B^T H_1 B)_k = \sum_{i,j} B_i H_1 B_{jk} B_k\). Solving the ODE system (A1–A2) 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.

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 \( \|\eta_t\|^2 = \Psi_t^2 B(T - t)^T (H_0 + H_1 \cdot X_t) B(T - t) \),

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

where \( \Psi_t = e^{-\int_0^t r_s ds} e^{A(T-t)+B(T-t)\cdot X_t} \) and \( A, B \) solve the ODE given in (A1-A2).

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

\[ \frac{dF}{dP} = \frac{e^{-\int_0^T r_s dt} e^{A \cdot X_T}}{E_0[e^{-\int_0^T r_s dt} e^{A \cdot X_T}]} \quad (A3) \]

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

\[ E_t \left[ e^{-\int_t^T r_s dt + \alpha + i\beta} e^{-A_{T-t} - B_{T-t} \cdot X_t} \right] \]
is in 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. See Duffie, Filipovic, and Schachermayer (2003).

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 $g \in S^\ast$. By Assumption 2, $f^F_\beta \cdot X_T \in S$, and so $f^F_\beta \cdot X_T \in S$ also. Thus Fourier inversion holds and $(\hat{f}^F_\beta \cdot X_T) = f^F_\beta \cdot X_T$ (see Corollary 8.28 in Folland (1984)), where we denote the inverse fourier transform of a function $h$ by $\hat{h}(s) = \frac{1}{2\pi} \int e^{is \cdot x} h(x) dx$. Applying this,

$$H = F_0 \langle g, (\hat{f}^F_\beta \cdot X_T) \rangle$$

$$= F_0 \langle \hat{g}, \hat{f}^F_\beta \cdot X_T \rangle.$$

This equation holds because of Fourier inversion and the definition of Fourier transform of a tempered distribution. Notice that when both $f^F_\beta \cdot X_T$ and $g$ are in $S$, we can write this last equality as

$$\frac{F_0}{2\pi} \int_x g(x) \int_s \hat{f}^F_\beta \cdot X_T e^{isx} ds dx = \frac{F_0}{2\pi} \int_s \hat{f}^F_\beta \cdot X_T \int_x g(x) e^{isx} dx ds,$$

thus we see that the theory from Fourier analysis of tempered distributions justifies the change of order of integration in a general sense. We can therefore further simply to obtain

$$H = \langle \hat{g}, F_0 \hat{f}^F_\beta \cdot X_T \rangle$$

$$= \frac{1}{2\pi} \langle \hat{g}, \psi(\alpha + \beta i) \rangle.$$

This last step holds by Assumption 1. This is the desired result. ☐

38
In some cases of interest, Assumption 2 may be violated. It could be that $\beta \cdot X_T$ has 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).

B Proof of Proposition 1

In analogy to Duffie, Pan, and Singleton (2000) and Pan (2002), define

$$G(\alpha_0; v, n| x, t) = e^{A_t + B_t \cdot x} \sum_{|\xi|=n} \binom{n}{\xi} L(x)^\xi, \quad (A4)$$

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 a $n$-dimensional multi-index, and $(\partial_i A, \partial_i B)$ satisfies the ODE

$$\dot{B} = K_1^T B + \frac{1}{2} B^\top H_1 B - \rho_1 + \lambda_1 (\phi(B) - 1), \quad B(0) = \alpha_0, \quad (A5)$$

$$\dot{A} = K_0^T B + \frac{1}{2} B^\top H_0 B - \rho_0 + \lambda_0 (\phi(B) - 1), \quad A(0) = 0, \quad (A6)$$

$$\partial_1 \dot{B} = K_1^T \partial_1 B + \partial_1 B^\top H_1 B + \lambda_1 \nabla \phi(B) \cdot \partial_1 B, \quad \partial_1 B(0) = v, \quad (A7)$$

$$\partial_1 \dot{A} = K_0^T \partial_1 B + \partial_1 B^\top H_0 B + \lambda_0 \nabla \phi(B) \cdot \partial_1 B, \quad \partial_1 A(0) = 0, \quad (A8)$$

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

$$\partial_m \dot{B} = K_1^T \partial_m B + \frac{1}{2} \sum_{i=0}^{m} \binom{m}{i} \partial_i B^\top H_1 \partial_{m-i} B + \partial_{m-1} (\lambda_1 \nabla \phi(B) \cdot \partial_1 B), \quad \partial_m B(0) = 0, \quad (A9)$$

$$\partial_m \dot{A} = K_0^T \partial_m B + \frac{1}{2} \sum_{i=0}^{m} \binom{m}{i} \partial_i B^\top H_0 \partial_{m-i} B + \partial_{m-1} (\lambda_0 \nabla \phi(B) \cdot \partial_1 B), \quad \partial_m A(0) = 0. \quad (A10)$$

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 (A1) 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

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

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

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

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

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

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

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

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.

**C Generalized Transform in Discrete Time**

In this appendix, we show how our method applies in a discrete time setting. Here, we replace (7) with

$$\Delta X_t = (K_0 + K_1 X_t) + \epsilon_{t+1},$$

where $\epsilon_{t+1}$ has a conditional distribution which depends on $X_t$ which satisfies

$$E[e^{\alpha \cdot \epsilon_{t+1}}] = e^{\hat{A}(\alpha) + \hat{B}(\alpha) \cdot X_t} .$$

(A13)

For example, if $\hat{A}(\alpha) = \frac{1}{2}\alpha'\Sigma\alpha$ and $\hat{B}(\alpha) = 0$, it follows that $\epsilon_t \sim N(0, \Sigma)$, in which case $X_t$ follows a simple vector autoregression. In general, the discrete time family of processes given by (A12–A13) is quite flexible and allows for jump-type processes and time-varying covariance (see Le, Singleton, and Dai (2010), for example).
Table 2: **Parameters.** This table gives the parameters and moments used to calibrate the model. The left column gives the preference parameters and conditional mean parameters for the process. The right column gives the conditional moments used to calibrate the parameters \( (\Sigma_0, \Sigma_1) \). The first three calibration moments refer to the steady state values. The next three refer to the conditional volatility of the conditional moments evaluated at the long run mean of \( V \). \( \sigma(\sigma_d) \) is the steady state volatility of \( \sigma_d \). \( \bar{V} \) is normalized to be one.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )</td>
<td>6</td>
</tr>
<tr>
<td>( \rho )</td>
<td>1%</td>
</tr>
<tr>
<td>( \bar{\rho} )</td>
<td>1.5%</td>
</tr>
<tr>
<td>( \bar{\sigma} )</td>
<td>80%</td>
</tr>
<tr>
<td>( a )</td>
<td>( \frac{1}{3} )</td>
</tr>
<tr>
<td>( \kappa_s )</td>
<td>0.0231</td>
</tr>
<tr>
<td>( \kappa_V )</td>
<td>0.0693</td>
</tr>
<tr>
<td>( \hat{\rho}_{\ell,d} )</td>
<td>-30.3%</td>
</tr>
<tr>
<td>( \hat{\sigma}_\ell )</td>
<td>5.4%</td>
</tr>
<tr>
<td>( \hat{\sigma}_d )</td>
<td>11.1%</td>
</tr>
<tr>
<td>( \sigma_\infty(\rho_{\ell,d}) )</td>
<td>9.8%</td>
</tr>
<tr>
<td>( \sigma_\infty(\sigma_{\ell}) )</td>
<td>0.0018%</td>
</tr>
<tr>
<td>( \sigma_{\infty}(\sigma_{\ell}) )</td>
<td>0.017%</td>
</tr>
<tr>
<td>( \sigma_{\infty}(V) )</td>
<td>1.07</td>
</tr>
</tbody>
</table>

For the discrete time processes, the analogous version of Theorem 1 gives

\[
H_D(g, \alpha, \beta) = E_0 \left[ \exp \left( -\sum_{u=0}^{T} r(X_u) \right) e^{\alpha \cdot X_T} g(\beta \cdot X_T) \right] = \frac{1}{2\pi} \langle \hat{g}, \psi(\alpha + \beta i) \rangle,
\]

where now \( (A1–A2) \) are replaced by

\[
\Delta B = K_1^T B + \hat{B}(B) - \rho_1, \quad B(0) = \alpha + is\beta, \quad (A15)
\]

\[
\Delta A = K_0^T B + \hat{A}(B) - \rho_0, \quad A(0) = 0. \quad (A16)
\]

**D Labor income risk**

This section provides more details on the calibration and analysis of the model with time-varying labor income risk.

The parameters are summarized in Table 2 and are calibrated as follows. We set the long-run mean growth rate of labor income and dividends to 1.5%. We specify the long run labor income share, \( \bar{S} \), to be 75%. As the covariance parameters \( (\Sigma_0, \Sigma_1) \) are difficult to directly interpret, we calibrate them by considering their effect on the volatility of labor income, the volatility of dividends, and their correlation. We set the parameters so that when \( V_t \) is at its long run mean \( \bar{V} \) (which is normalized to be one), \( (\sigma_{\ell,t}, \sigma_{d,t}, \rho_{\ell,d,t}) \) are given by \( \hat{\sigma}_\ell = 5.4\%, \hat{\sigma}_d = 11.1\%, \) and \( \hat{\rho}_{\ell,d} = -30.3\% \) respectively. Note that due to CRRA utility, our model presents the equity premium-risk free rate.
puzzle (Mehra and Prescott (1983)), and we choose our parameterization to generate higher premium with reasonable risk aversion by slightly overstating the volatility of labor income relative to the data, with the ratio of dividend to labor income volatility qualitatively similar to Lettau, Ludvigson, and Wachter (2008). We also calibrate the volatility of \((\sigma_d, \sigma_\ell, \rho_{\ell,d})\) when \(V_t\) is at its long run mean, which we denote with by \(\sigma_\infty(\sigma_d) = 1.7\text{bp}, \sigma_\infty(\sigma_\ell) = 0.18\text{bp},\) and \(\sigma_\infty(\rho_{\ell,d}) = 9.8\%\). Finally, we calibrate the volatility of \(V\) in the steady state distribution, which we denote by \(\sigma_{SS}(V)\), to be 1.07. Taken together, these 7 moments (along with the simplifying assumption that innovations to \(V\) are uncorrelated with innovations to either \(\ell\) or \(d\)) fix the free parameters in \(\Sigma_0\) (3 parameters) and \(\Sigma_1\) (4 parameters). Under this calibration, when \(V\) is at the highest (lowest) decile, the volatility of labor income is 6% (5%), the volatility of dividends is 16%(6%) and their correlation is -10% (-80%). The volatility parameters were chosen to qualitatively match the variation found in Figure 4.