Generalized Transform Analysis of Affine Processes and Applications in Finance*

Hui Chen† Scott Joslin‡

March 18, 2011

Abstract

Nonlinearity is an important consideration in many problems of finance and economics, such as pricing securities, computing equilibrium, and conducting structural estimations. We extend the transform analysis in Duffie, Pan, and Singleton (2000) by providing analytical treatment of a general class of nonlinear transforms for processes with tractable conditional characteristic functions. We illustrate the applications of the generalized transform method in pricing contingent claims and solving general equilibrium models with preference shocks, heterogeneous agents, or multiple goods. We also apply the method to a model of time-varying labor income risk and study the implications of stochastic covariance between labor income and dividends for the dynamics of the risk premiums on financial wealth and human capital.

*We thank Damir Filipovic, Xavier Gabaix, Martin Lettau, Ian Martin, Jun Pan, Tano Santos, Ken Singleton, and seminar participants at Boston University, Duke Fuqua, MIT Sloan, Penn State Smeal, the 2009 Adam Smith Asset Pricing Conference, and the Winter 2010 Econometric Society Meetings (Atlanta) 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.

‡MIT Sloan School of Management, sjoslin@mit.edu. Tel. 617-324-3901.

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.

Consider a state variable X_t that follows an affine jump-diffusion, in the sense that the conditional characteristic function is affine.[1] Duffie, Pan, and Singleton (2000), hereafter DPS, derive closed-form expression for the following transform:

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

where $R(X)$ is an affine function of X, which can be interpreted as a stochastic "discount rate," and $e^{u \cdot X_T} \left(v_0 + v_1 \cdot X_T \right) 1_{\{\beta \cdot X_T < y\}}$ 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 transform:

$$E_t \left[\exp \left(- \int_t^T R(X_s, s)\, ds \right) f(X_T)\, g(\beta \cdot X_T) \right], \tag{2}$$

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 satisfying certain regularity conditions. Moreover, X can be any stochastic process with tractable conditional characteristic functions. When X is an affine jump-diffusion, $f(X) = e^{u \cdot X} \left(v_0 + v_1 \cdot X \right)$ and $g(\beta \cdot X) = 1_{\{\beta \cdot X < y\}}$, we recover the transform of DPS in (1). The flexibility in choosing f and g in (2) as well as the process for X_t makes the generalized transform useful in dealing with generic nonlinearity problems in asset pricing (nonlinear stochastic discount factors or payoffs), estimation (nonlinear moments), and other areas.

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

[1] See Duffie, Filipovic, and Schachermayer (2003) for an elaboration on the characterization via the characteristic function.

tractability to our generalized transform by avoiding intermediate Fourier inversions. The class of functions with Fourier transforms is quite large (tempered distributions), and many popular stochastic processes (such as affine jump-diffusions or Lévy processes) have tractable characteristic functions, which make our method applicable to a wide range of problems.

In addition to being a computational tool, we can also take advantage of the generalized transform in economic modeling. Consider the pricing equation for an asset with stochastic payoff y_T at time T:

$$P_t = E_t \left[m(T, X_T) y(X_T) \right],$$

where m_t is the stochastic discount factor. In the background, there is a model that determines the discount factor m and payoff y as functions of the state variables X. To maintain tractability, one often adopts special utility functions, impose strong restrictions on the processes of the state variables (e.g., *i.i.d.* or Gaussian processes), or log-linearize the model to obtain approximate solutions. By doing so one not only loses certain realistic features of the data, but also can potentially miss important nonlinear implications of the model.

The new tools provided in this paper help relax some of these modeling constraints. By using the generalized transform, we can (i) price assets with more complicated payoffs, (ii) relax certain restrictions (such as restrictions on preferences or production function) for models that produce nonlinear stochastic discount factors, and (iii) significantly enrich the underlying stochastic uncertainties without losing tractability. In some cases these extensions are necessary to improve the quantitative performances of the models (e.g., by relaxing the assumptions of logarithmic preferences). Other times they provide a systematic and convenient way to introduce new ingredients to the existing models, such as time-varying growth rates, stochastic volatility, jumps, or cointegration. We provide several example applications that utilize the generalized transform to deal with nonlinear payoffs, including pricing defaultable bonds with stochastic recovery risk and options with exotic payoffs, and several on dealing with nonlinear stochastic discount factors, such as a model of habit formation and models of heterogeneous agents.

We also apply our method to 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 explain 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 has almost no relationship with 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 also has interesting implications for the comovement between the risk premium on financial wealth and human capital.

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) makes it easy to compute the moments arising from asset pricing, estimation, and forecasting. Examples applications include Singleton (2001), Pan (2002), Piazzesi (2005), and Joslin (2010), among many 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, Duffie, Pan, and Singleton (2000) obtain the extended transform in (1) for affine jump-diffusions, which they apply to option

[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).

pricing. Other papers that take this approach include Heston (1993), Chen and Scott (1995), Bates (1996), Bakshi and Chen (1997), Bakshi, Cao, and Chen (1997), Dumas, Kurshev, and Uppal (2009), Buraschi, Trojani, and Vedolin (2010), among others.[3] In our approach, we consider affine jump-diffusions and indeed any process with known condition characteristic function. Moreover, our method allows direct computation of a large class of nonlinear moments without the need to compute the (forward) density function of the appropriate random 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 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, Longstaff, and Santa-Clara (2008).[4] 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 recovery rate of a defaulted corporate bond. The amount an investor recovers from a corporate bond upon default can depend on many factors,

[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. Both methods can be time-consuming and lacking accuracy, especially in high dimensional cases.

[4] In the N-tree case, $N > 2$, Martin (2008) also provides an $(N-2)$-dimensional integral to compute the associated $(N-1)$-dimensional transform.

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 values from $[0, 1]$.

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

$$\varphi(X_t) = \frac{1}{1 + e^{-\beta_0 - \beta_1 \cdot X_t}}, \tag{3}$$

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 the recovery rate if default occurs at some future time T. That is, we are interested in computing $E_0[\varphi(X_T)]$.[5] To simplify notation, we first define $Y_T \equiv \frac{1}{2}(-\beta_0 - \beta_1 \cdot X_T)$. We then rewrite

$$E_0[\varphi(X_T)] = E_0\left[\frac{1}{1 + e^{2Y_T}}\right] = E_0\left[\frac{1}{2} e^{-Y_T} \frac{1}{\cosh(Y_T)}\right], \tag{4}$$

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 (4), 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 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:

$$CCF(T, u; X_0) = E[e^{iu \cdot X_T} \,|\, X_0], \tag{5}$$

where $i = \sqrt{-1}$. If we could "approximate" the non-linear term $1/\cosh(Y_T)$ inside the expectation

[5] We 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 $\{X_\tau\}_{0 \le \tau \le T}$. This is stronger than the standard doubly stochastic assumption; relaxing this assumption is relatively straightforward but would complicate the example.

of (4) with exponential linear functions of Y_T, then we would be able to use the characteristic function to compute the non-linear expectation. As we elaborate in Section 3, this is achieved using the Fourier inversion of $1/\cosh(y)$,

$$\frac{1}{\cosh(y)} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{g}(s)e^{isy}ds, \tag{6}$$

where $\hat{g}$ is the Fourier transform of $1/\cosh(y)$, which is known analytically,[6]

$$\hat{g}(s) = \frac{\pi}{\cosh(\frac{\pi s}{2})}\,. \tag{7}$$

Thus, we can substitute out $1/\cosh(Y_T)$ from (4) and obtain

$$\begin{aligned} E_0[\varphi(X_T)] &= E_0\left[\frac{1}{2}e^{-Y_T}\frac{1}{2\pi}\int_{-\infty}^{+\infty}\hat{g}(s)e^{is\cdot Y_T}ds\right] \\ &= \frac{1}{4\pi}\int_{-\infty}^{+\infty}\frac{\pi}{\cosh(\frac{\pi s}{2})}E\left[e^{(-1+is)\cdot Y_T}\Big| X_0\right]ds \\ &= \frac{1}{4}\int_{-\infty}^{+\infty}\frac{1}{\cosh(\frac{\pi s}{2})}e^{-\frac{1}{2}(-1+is)\beta_0}CCF\left(T,-\frac{1}{2}(i+s)\beta_1;X_0\right)ds, \end{aligned} \tag{8}$$

where for the second equality we assume that the order of integrals can be exchanged; the third equality follows from applying the result in (5). 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 a single N-dimensional partial differential equation.

[6]See, for example, Abramowitz and Stegun (1964), 6.1.30 and 6.2.1.

3 Generalized transforms

We now present our theoretical results. As in Duffie, Pan, and Singleton (2000), 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

$$dX_t = (K_0 + K_1 X_t)dt + \sigma_t dW_t + dZ_t, \tag{9}$$

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 ν, and $(\sigma_t \sigma_t')_{i,j} = H_{0,ij} + \sum_k H_{1,ijk} X_t^k$ 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 Θ denote the parameters of the process $(K_0, K_1, H_0, H_1, \ell_0, \ell_1, \nu, \rho_0, \rho_1)$. Alternatively, we can define the process in terms of the infinitesimal generator or, as Duffie, Filipovic, and Schachermayer (2003) and Singleton (2001) stress, in terms of the conditional characteristic function.

We focus on 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. However, our results only require that the conditional characteristic functions of X_t are tractable and apply whether the stochastic process is modeled in discrete-time or continuous-time. In Section 3.2 we discuss examples of processes that are not affine jump-diffusions. In Appendix C we present the generalized transform result in discrete time.

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 α 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 α[7] and $\|x\|$ is the Euclidean norm of the vector x. The collection of all Schwartz functions is denoted $\mathcal{S}$. $\mathcal{S}$ is endowed with the topology generated by the family of semi-norms $\|f\|_{P,\alpha}$. The dual of

[7] Thus for $\alpha = (\alpha_1, \ldots, \alpha_N)$, $\partial^\alpha f = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_N}}{\partial x_N^{\alpha_N}} f$.

$\mathcal{S}$, 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

$$\langle g, f\rangle = \int_{\mathbb{R}^N} g(x)f(x)dx. \tag{10}$$

As is standard, we maintain the inner product notation even when a tempered distribution g is does not correspond to a function as in (10). For example, the δ-function is a tempered distribution given by $\langle \delta, f\rangle = f(0)$ which does not arrive from a function.[8]

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

$$\hat{f}(s) = \int_{\mathbb{R}^N} e^{-is\cdot x} f(x)dx. \tag{11}$$

The Fourier transform can be inverted through the relation

$$f(x) = \frac{1}{(2\pi)^N}\int_{\mathbb{R}^N} e^{ix\cdot s}\hat{f}(s)ds, \tag{12}$$

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 (10)), but do not have Fourier transform in the sense of (11) because the integral in is not well-defined. An example is the Heaviside function:

$$f(x) = 1_{\{0\le x\}} \Rightarrow \hat{f}(s) = \pi\delta(s) - \frac{i}{s}, \tag{13}$$

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

[8]Throughout the paper, we use the notation $\delta(s)$ to denote the Dirac delta function, with $\delta_x(s) = \delta(s - x)$.

not decay at infinity and may not even be bounded.

We now state our main result:

Theorem 1. *Suppose that $g \in \mathcal{S}^*$ and (Θ, α, β) satisfies Assumption 1 and Assumption 2 in Appendix A. Then*

$$H(g, \alpha, \beta) = E_0 \left[\exp\left(-\int_0^T R(X_u) du \right) e^{\alpha \cdot X_T} g(\beta \cdot X_T) \right] = \frac{1}{2\pi} \langle \hat{g}, \psi(\alpha + \cdot \beta i) \rangle, \tag{14}$$

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

$$s \mapsto \psi(\alpha + s\beta i) = E_0 \left[e^{-\int_0^T R(X_u) du} e^{(\alpha + is\beta) \cdot X_T} \right]. \tag{15}$$

The function ψ is the transform given in DPS. Recalling their result,

$$\psi(\alpha + is\beta) = e^{A(T; \alpha + is\beta, \Theta) + B(T; \alpha + is\beta, \Theta) \cdot X_0}, \tag{16}$$

where A, B satisfy the ordinary differential equations (ODEs)

$$\dot{B} = K_1^\top B + \frac{1}{2} B^\top H_1 B - \rho_1 + \ell_1(\phi(B) - 1) \quad B(0) = \alpha + is\beta, \tag{17}$$

$$\dot{A} = K_0^\top B + \frac{1}{2} B^\top H_0 B - \rho_0 + \ell_0(\phi(B) - 1) \quad A(0) = 0, \tag{18}$$

where $\phi(c) = E_\nu[e^{c \cdot Z}]$, the moment-generating function of the jump distribution and $(B^\top H_1 B)_k = \sum_{i,j} B_i H_{1,ijk} B_k$. Solving the ODE system (17–18) 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

$$H = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{g}(s) \psi(\alpha + is\beta) ds. \tag{19}$$

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 (19) makes this spanning explicit for the cases where $\hat{g}$ defines a function.

In some cases of interest, Assumption 2 for Theorem 1 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).

There is some flexibility in the choice of α and g in (14). Notice that

$$e^{\alpha \cdot X_T} g(\beta \cdot X_T) = e^{(\alpha - c\beta)\cdot X_T} \tilde{g}(\beta \cdot X_T), \tag{20}$$

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 (19).

3.1 Two Extensions

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

$$f(\alpha, \gamma, p, X) = \sum_i p_i X^{\gamma_i} e^{\alpha_i \cdot X}, \tag{21}$$

where $\{p_i\}$ are arbitrary constants, $\{\alpha_i\}$ are complex vectors[9] and $\{\gamma_i\}$ are arbitrary multi-indices so that $X^\gamma = \prod_j X_j^{\gamma_i}$. 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 *pl-linear* functions.

Proposition 1. *Suppose that $g \in \mathcal{S}^*$, and $(\Theta, \alpha, \beta, \gamma)$ satisfies Assumption 1' and Assumption 2'*

[9] Allowing complex eigenvalues allows one to have oscillatory *sine* and *cosine* terms.

in Appendix B. Then

$$H(g, \alpha, \beta, \gamma) = E_0 \left[\exp\left(-\int_0^T R(X_u) du \right) X_T^\gamma e^{\alpha \cdot X_T} g(\beta \cdot X_T) \right]$$
$$= \frac{1}{2\pi} \langle \hat{g}, \psi(\alpha + \cdot \beta i; \gamma) \rangle, \tag{22}$$

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

$$s \mapsto \psi(\alpha + s\beta i; \gamma) = E_0 \left[e^{-\int_0^T R(X_u) du} X_T^\gamma e^{(\alpha + is\beta) \cdot X_T} \right]. \tag{23}$$

The function ψ 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

$$H(f, g, \alpha, \beta) = E_0 \left[\exp\left(-\int_0^T R(X_u) du \right) f(\alpha, \gamma, p, X_T) g(\beta \cdot X_T) \right]. \tag{24}$$

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 and 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 β_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 $g \in \mathcal{S}_M^*$ (an M-dimensional tempered distribution), $\alpha \in \mathbb{R}^N$, $\mathfrak{b} \in \mathbb{R}^{M \times N}$ and $(\Theta, \alpha, \mathfrak{b})$ satisfies Assumption 1 and Assumption 2 in Appendix A. Then*

$$H(g, \alpha, \mathfrak{b}) = E_0 \left[\exp\left(-\int_0^T R(X_s) ds \right) e^{\alpha \cdot X_T} g(\mathfrak{b} X_T) \right]$$
$$= \frac{1}{(2\pi)^M} \langle \hat{g}, \psi_M(\alpha + \cdot \mathfrak{b} i) \rangle, \tag{25}$$

where $\hat{g} \in \mathcal{S}^*$ and $\psi_M(\alpha + \cdot \mathfrak{b} i)$ denotes the function

$$\psi_M : \mathbb{C}^M \to \mathbb{C}\ ,\ \ s \mapsto \psi_M(\alpha + s^\top \mathfrak{b} i) = E_0\left[e^{-\int_0^T r(X_u)du} e^{(\alpha + i s^\top \mathfrak{b})\cdot X_T}\right]. \tag{26}$$

It is immediate to extend the transform in Proposition 2 by replacing $e^{\alpha \cdot X_T}$ with a *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 $\widetilde{g}$ for which the Fourier transform is known in closed form.

Moreover, a convenient feature of the generalized transform is that the Fourier transform of g and the solutions to the ODEs in (17–18) need only be computed once for a given set of parameters. Once computed, the Fourier transforms 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 *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 (15), (23), or (26). 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 (modified) conditional characteristic function can be computed.

An 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. One example of such a process is the variance gamma (VG) process studied by Madan, Carr, and Chang (1998). The VG process consists of a time-changed Brownian motion where the clock follows a gamma process. It is a pure jump process that offers increased kurtosis for increments relative to a standard Brownian motion, which can be useful in modeling the fat tails in financial data.

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 ν given by the density k_{CGMY}, where

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

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 VG process. In general, Y controls the activity of very small jumps with a lesser effect on the amount of large jumps since these are dominated by the exponential terms. The parameter G (M) control the tail behavior of large positive (negative) jumps. The parameter C controls the overall activity of jumps of the process. 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

$$E_0[e^{iu(X_t - X_0)}] = \exp\left(tC\Gamma(-Y)[(M - iu)^Y - M^Y + (G + iu)^Y - G^Y]\right). \tag{27}$$

From the characteristic function, we can then immediately apply the results of Theorem 1 and Proposition 1.

A second example of non-AJD process is the class of Markov-switching affine models. A potentially restrictive assumption of the AJD class of processes is that parameters of the model are static. However, discrete changes in the economic environment may result in significant changes in the underlying parameters of the model. The methods of Dai, Singleton, and Yang (2007) and Ang, Bekaert, and Wei (2008) (both models are in discrete time) can be extended to our framework to incorporate regime shifts for an AJD in the conditional mean, conditional covariance, or the conditional probability of jumps.

Suppose that there are n possible regimes, denoted by $s_t = 1, \cdots, n$. The transitions across different regimes are governed by a continuous-time Markov chain, with the generator matrix $\mathbf{\Gamma} = [\gamma_{ss'}]$. Thus, the probability of the economy moving from regime s to regime s' over a period Δt will be $\gamma_{ss'}\Delta t + o(\Delta t)$.[10] Now suppose that there is a variable X_t with *regime-dependent* dynamics:

$$dX_t = (K_0(s_t) + K_1 X_t)dt + \sigma_t dW_t + dZ_t, \tag{28}$$

where s_t is the current regime at time t, W is an $\mathcal{F}_t$-standard n-dimensional Brownian motion, Z is a pure jump process with arrival intensity $\lambda_t = \ell_0(s_t) + \ell_1 \cdot X_t$ and fixed D-invariant distribution ν, and $(\sigma_t\sigma_t')_{i,j} = H_{0,ij}(s_t) + \sum_k H_{1,ijk}X_t^k$ with $H_0(1), \cdots, H_0(n) \in \mathbb{R}^{N\times N}$ and $H_1 \in \mathbb{R}^{N\times N\times N}$. Now the dynamics of the state variable are regime dependent with the constant term in the drift, covariance, and the jump intensity changing across regimes. The (discounted) conditional characteristic function is given by (15), which will be regime-dependent as well. It follows from (28) that

$$\psi(\alpha + is\beta, s_t) = e^{A(T,s_t;\alpha+is\beta,\Theta)+B(T;\alpha+is\beta,\Theta)\cdot X_0}, \tag{29}$$

where B is still the solution to the ODE (17) (with no dependence on the regime), and $A(t,j)$ solves the following regime-dependent system of ODEs:

$$\dot{A}(j) = K_0(j)^T B + \frac{1}{2}B^T H_0(j)B - \rho_0 + \ell_0(j)(\phi(B) - 1) + \sum_{k\neq j}\gamma_{jk}(e^{A(k)-A(j)} - 1) \quad A(0,j) = 0, \tag{30}$$

for $j = 1, \cdots, n$. With these modifications, the rest of the generalized transform applies directly.

A third example of processes outside the class of AJDs for which our results apply are the discrete time autoregressive processes with gamma-distributed shocks. Following the setup of Bekaert and

[10] We could also assume that the transition probabilities are affine in X_t (which can be easily handled by modifying (17)). Such features of the transition probabilities would be important in the cases where the likelihood of regime shifts is time-varying.

Engstrom (2010) (for simplicity we omit any normally-distributed shocks), suppose X_t follows

$$X_{t+1} = \mu + \Phi X_t + \Sigma \omega_{t+1}, \tag{31}$$

where Σ $(n \times p)$ is the conditional volatility matrix for the gamma-distributed shocks, ω_{t+1} $(p \times 1)$, with $\omega_{t+1}^i \sim \Gamma(k_t^i, 1) - k_t^i$, and $k_t = K X_t$. The conditional characteristic function for X_{t+1} is

$$\begin{aligned} E_t\left[e^{iu \cdot X_{t+1}}\right] &= \exp\left(iu^T(\mu + \Phi X_t) - \left(iu^T\Sigma + \ln(1 - iu^T\Sigma)\right) K X_t\right) \\ &= \exp\left(A(1,u) + B(1,u) \cdot X_t\right), \end{aligned} \tag{32}$$

with

$$\begin{aligned} A(1,u) &= iu^T\mu, \\ B(1,u)^T &= iu^T\Phi - \left(iu^T\Sigma + \ln(1 - iu^T\Sigma)\right) K. \end{aligned}$$

One can then compute the conditional characteristic function for X_{t+n} $(n > 1)$ through recursion of formula (32), which renders a pair of difference equations.

The gamma-distributed shocks in the above process provide a convenient way to generate time-varying skewness and kurtosis. Bekaert and Engstrom (2010) use this process to model aggregate consumption growth, which helps explain the equity premium and the variance premium. One can also generalize this autoregressive process in (31) by replacing ω_{t+1} with other types of shocks that have tractable characteristic functions.

4 Applications in Contingent Claim Pricing

Having presented the theory of the generalized transform, we now illustrate how it can be applied to pricing contingent claims. The examples include defaultable bonds with stochastic recovery and European options with general payoffs.

4.1 Recovery risk

Following up on the illustrating example in Section 2, we first discuss the pricing of credit-risky securities (e.g., defaultable bonds or credit default swaps) with stochastic recovery upon default. Recovery risk refers to the uncertainty about the recovery rate. In particular, investors will demand high risk premium if the recovery of these securities are lower in aggregate bad times. For example, Chen (2010) provides evidence that macro variables such as GDP growth and riskfree rate are correlated with the aggregate recovery rates and default rates and that such correlations can have large effects on the pricing of defaultable bonds.

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 τ with risk-neutral intensity h_t.[11] The recovery value at default φ 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, the price of the bond is:

$$\begin{aligned} V_t &= E_t^{\mathbb{Q}}\left[e^{-\int_t^\tau r_u du} 1_{\{\tau \leq T\}} \varphi_\tau\right] + E_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u du} 1_{\{\tau > T\}}\right] \\ &= E_t^{\mathbb{Q}}\left[\int_t^T e^{-\int_t^s (r_u + h_u) du} h_s \varphi_s ds + e^{-\int_t^T (r_u + h_u) du}\right]. \end{aligned} \tag{33}$$

The second equality follows from the doubly-stochastic assumption and regularity conditions.

Duffie and Singleton (1999) show that under the assumption that the recovery value φ_t is a proportion $1 - L_t$ of the market value of the security immediately before default (referred to as "recovery of market value" (RMV)) and a suitable no-jump condition,[12] 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 $\lambda_t L_t$ and r_t as jointly affine under the risk-neutral measure, then the standard results for affine term structure models can be used to price defaultable bonds.

While analytically appealing, the RMV assumption has some limitations. First, since the default

[11] To be precise, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ 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 $\mathbb{Q}$, τ 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.

[12] The no-jump condition is satisfied here by assuming φ is adapted to $\{\mathcal{F}_t\}$. See also Duffie, Schroder, and Skiadas (1996) and Collin-Dufresne, Goldstein, and Hugonnier (2004) for discussions on the no-jump condition.

intensity h_t and the loss rate L_t enter the default-adjusted discount rate symmetrically, we cannot separately identify the risk-neutral default intensity and recovery rate 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 (which is the same across different bonds) and recovery rates (which depend on seniority). Third, 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 (30 days after the date of default) to par value. The statistical models of recovery rates commonly used by academics and practitioners are also generally based on par values.

Let X_t be the vector of state variables that determine the riskfree rate, the default intensity, and the recovery rate. Assume that $r(X_t) = \rho_0 + \rho_1 \cdot X_t$, and $h(X_t) = \theta_0 + \theta_1 \cdot X_t$. If we model the recovery value $\varphi_t = g(\beta \cdot X_t)$ with a proper function g, then equation (33) clearly fits into the transform in Proposition 1. This allows us to substantially relax the RMV restriction. In Appendix E, we calibrate a recovery model that captures the nonlinear relation between recovery rate and aggregate default intensity in the data. The model shows that ignoring such a relation can lead to large pricing errors for defaultable securities.

In addition to the logistic model considered in Section 2, we can also model the stochastic recovery rate using the probit model, the Cauchy model, or the cumulative distribution functions associated with any other probability distributions.[13] Such flexibility in modeling the recovery rate makes the above framework well suited to connect models of defaultable bond pricing with commonly used statistical models of recovery rates.

4.2 Option pricing

When pricing European options, we want to compute expectations of the form

$$E_t^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds + \alpha \cdot X_T} g_y(\beta \cdot X_T)\right], \tag{34}$$

[13] Bakshi, Madan, and Zhang (2006) study a class of recovery models for which φ 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. The completely monotone class requires that each order derivative is either always positive or negative.

where $g_y(x) = 1_{\{x \leq y\}}$ is nonlinear and non-integrable. For example, for a European put option with strike K, $\beta \cdot X_t$ will be the log stock price, $y = \log K$, and the option price is

$$P_t = E_t^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds + y} g_y(\beta \cdot X_T)\right] - E_t^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds + \beta \cdot X_T} g_y(\beta \cdot X_T)\right]. \tag{35}$$

The Fourier transform of g_y is defined as a distribution:

$$\hat{g}_y(s) = \pi\delta(s) + \frac{ie^{-isy}}{s}, \tag{36}$$

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

$$\begin{aligned} E_t^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds + \alpha \cdot X_T} g_y(\beta \cdot X_T)\right] &= \int_{-\infty}^{\infty} \left(\frac{\delta(s)}{2} - \frac{e^{-isy}}{2\pi is}\right) \psi(\alpha + is\beta) ds \\ &= \frac{\psi(\alpha)}{2} - \int_0^{\infty} \text{Real}\left(\frac{\psi(\alpha + is\beta)e^{-isy}}{\pi is}\right) ds. \end{aligned} \tag{37}$$

In the last equation we use the fact that the real part of the integrand is even and the imaginary part is odd. This replicates the formula given in DPS obtained by Lévy-inversion.[14]

Since the generalized transform only requires the state variable to have a tractable conditional characteristic function, we have a lot of flexibility in modeling the underlying uncertainty. Our method also applies to exotic options whose payoffs are more complicated functions of the state variables (such as the terminal value of the stock), as long as the payoff can be written in the form of a product of a *pl-linear* function and a tempered distribution.

Bakshi and Madan (2000) (henceforth BM) also provide a more general method than DPS 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,[15] and K is a fixed strike price. They propose a power series expansion of H, the expectations of which can then be computed through differentiation of the conditional

[14] 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 be reduced to a 1-dimensional integral as in the standard Lévy inversion formula (without a forward measure).

[15] An entire function is one that is analytic at all finite points on the complex plane.

characteristic functions. Their method applies to non-affine processes as well (see Section 3.2).

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 can be 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 $\mathcal{L}^1$ function (see BM Theorem 1).[16] We relax the growth condition on the underlying payoff; the dual space $\mathcal{S}^*$ is quite large and contains $\mathcal{L}^p$ for any p, as well as functions not in $\mathcal{L}^p$ 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 α. In this case H is not entire (for any choice of X_0, the power series converges to $H(X)$ only for $0 < X < 2X_0$).[17] Nor is it an affine translation of an $\mathcal{L}^1$ function. However, we can still use the fact that H represents a tempered distribution to write

$$\hat{g}(s) = (-is)^{-1-\alpha}\Gamma_{\text{inc}}(-\alpha - 1, -is), \tag{38}$$

where $\Gamma_{\text{inc}}(\cdot,\cdot)$ denotes the incomplete Γ function and $s^{-1-\alpha}$ is interpreted in the sense of a homogeneous tempered distribution.

5 Applications in Economic Modeling

Affine general equilibrium models are attractive due to their tractability. This class of models starts with state variables X_t that are affine and develop (either exogenously or endogenously) the

[16] A function H is of class $\mathcal{L}^p$ if $\left(\int |H(x)|^p dx\right)^{1/p} < \infty$.

[17] In practice, one may still compute option prices through the Taylor expansion of an non-entire payoff function with reasonable accuracy for special choices of X_0 and the order of Taylor expansion, although it is not obvious what choices would be optimal ex ante.

stochastic discount factor M_t as an exponential affine function of X_t:

$$M_t = m(t, X_t) = \exp(-\rho t - a \cdot X_t). \tag{39}$$

For example, in an endowment economy, the SDF is the marginal utility of the representative agent, which depends on the log endowment and potentially other state variables that are jointly affine.[18] Under these assumptions, it is easy to show that the riskfree rate r_t is affine in X_t, and X is still affine under the risk-neutral probability measure $\mathbb{Q}$. As a result, the value a security with payoff $Y_T = y(X_T)$ at time T can be computed as

$$P_t = E_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u du} y(X_T)\right], \tag{40}$$

which can be solved analytically using the DPS transform provided that $y(X)$ is a *pl-linear* function of X, or using the generalized transform if $y(X)$ can be decomposed into the product of a *pl-linear* function $f(X)$ and a function with a Fourier transform $g(\beta \cdot X)$ as in Proposition 1.

However, the benefit of the generalized transform is not limited to dealing with nonlinear payoffs. Consider the case where $m(t, X)$ is not exponential affine (but still strictly positive). The value of the security above can also be computed as

$$P_t = \frac{1}{m(t, X_t)} E_t\left[m(T, X_T) y(X_T)\right]. \tag{41}$$

This expectation can be computed using the generalized transform provided that $m(T, X)y(X) = f(X)g(\beta \cdot X)$ as in Proposition 1. Then, the riskfree rate r_t is no longer required to be affine in X_t, and X is generally no longer an affine process under $\mathbb{Q}$.

To see this, suppose $m(t, X_t) = e^{\int_0^t R(X_s) ds} g(\beta \cdot X_t)$, where $R(X_t) = \rho_0 + \rho_1 \cdot X_t$, and X follows an affine diffusion (it is also easy to add jumps). Then, the Ito's lemma implies that the riskfree

[18] Examples include Bakshi and Chen (1997), Bekaert and Grenadier (2002), Mamaysky (2002), Longstaff and Piazzesi (2004), among others.

rate is

$$r_t = -(\rho_0 + \rho_1 \cdot X_t) - \frac{g'(\beta \cdot X_t)}{g(\beta \cdot X_t)} \beta^T (K_0 + K_1 X_t) - \frac{1}{2} \frac{g''(\beta \cdot X_t)}{g(\beta \cdot X_t)} tr(\beta \beta^T \sigma_t \sigma_t'),$$

and under the risk-neutral measure X follows

$$dX_t = \left(K_0 + K_1 X_t + \frac{g'(\beta \cdot X_t)}{g(\beta \cdot X_t)} \sigma_t \sigma_t' \beta \right) dt + \sigma_t dW_t^{\mathbb{Q}}.$$

Both the risk free rate and the conditional mean of X_t under $\mathbb{Q}$ are nonlinear, making it unsuitable to do valuation directly under $\mathbb{Q}$. Instead, valuation under $\mathbb{P}$ using (40) can still be very tractable. Again, as discussed in Section 3.2, X does not need to be an affine jump-diffusion. All that is required is that its conditional characteristic function is tractable.

Next, we discuss some specific examples where non-exponential affine stochastic discount factors arise in realistic settings.

5.1 An Affine Model of External Habit

The external habit model of Campbell and Cochrane (1999) is a workhorse in asset pricing that helps generate a high and time-varying equity premium even though consumption growth is *i.i.d.* and has low volatility. Solving this model as well as estimation and forecasting can be challenging due to the complicated dynamics of the external habit process. In this example, we construct a habit process based on affine state variables that captures the desired features of the habit model.

Our construction is based on the continuous-time version of the external habit model in Santos and Veronesi (2010). In an endowment economy, the representative agent's utility over consumption stream $\{C_t\}$ is

$$E\left[\int_0^\infty e^{-\rho t} \frac{(C_t - H_t)^{1-\gamma}}{1-\gamma} dt\right], \tag{42}$$

where H_t is the habit level that is positive and strictly below C_t. The log aggregate endowment

$c_t = log(C_t)$ follows the process

$$dc_t = \mu_c dt + \sigma_c dW_t, \tag{43}$$

with constant expected growth rate μ_c and volatility σ_c, and we specify the process for H_t later.

The stochastic discount factor is obtained from the marginal utility of consumption for the representative agent,

$$m_t = e^{-\rho t}(C_t - H_t)^{-\gamma} = e^{-\rho t - \gamma c_t}\left(\frac{C_t - H_t}{C_t}\right)^{-\gamma}, \tag{44}$$

where we rewrite the SDF as the product of the standard SDF for CRRA utility ($e^{-\rho t - \gamma c_t}$) multiplied by a function of the surplus-consumption ratio.

Campbell and Cochrane (1999) specify a heteroskedastic AR(1) process for the surplus-consumption ratio. Santos and Veronesi (2010) directly model $G_t \equiv \left(\frac{C_t - H_t}{C_t}\right)^{-\gamma}$ with a non-affine process that is mean-reverting with stochastic volatility. Our approach is different. Instead of modeling the surplus-consumption ratio or G_t directly, we assume that G_t is a function of a variable y_t, which is stationary and jointly affine with c_t. An example for the process of y_t is:

$$dy_t = \kappa(\bar{y} - y_t)dt - \sigma_y dW_t, \tag{45}$$

where y_t and c_t are instantaneously perfectly negatively correlated, which captures the property that the habit level is solely driven by consumption shocks. The SDF then becomes

$$m(t, X_t) = e^{-\rho t - \gamma c_t} g(y_t). \tag{46}$$

One can apply several criteria in choosing the functional form for g, which in turn implies the dynamics of G_t. First, we need $g(y) > 1$, because H_t should be between 0 and C_t. Second, we also need $g'(y) \geq 0$, because a negative shock to consumption raises the consumption-surplus ratio and hence the marginal utility of consumption. Third, a negative shock to consumption ought to reduce the habit level, that is, $\text{Cov}_t(dC_t, dH_t) > 0$. Fourth, to generate counter-cyclical risk premium, we would like negative shocks to consumption to raise the conditional Sharpe ratio of the market portfolio (the price of risk for consumption shocks):

$$SR(y_t) = \gamma\sigma_c + \frac{g'(y_t)}{g(y_t)}\sigma_y. \tag{47}$$

Thus, we need $\frac{d}{dy}SR(y) = \frac{d}{dy}log(g(y)) \geq 0$.

Finding such a function g is straightforward, and it is clear that it cannot be exponential-affine in y.[19] Suppose the desired range of Sharpe ratio is between $\gamma\sigma_c$ and $\gamma\sigma_c + \alpha$ for some $\alpha > 0$. Then, we can assume

$$\frac{g'(y)}{g(y)} = \frac{\alpha}{\sigma_y}F(y), \tag{48}$$

where F is any monotone non-decreasing function with

$$\lim_{y\to-\infty} F(y) = 0, \quad \lim_{y\to+\infty} F(y) = 1.$$

For example, F can be the cumulative distribution function of any real univariate random variable. Thus, we have a lot of flexibility in choosing the desired shape of F, which in turn decides the distribution of the conditional Sharpe ratio.

It follows that

$$g(y) = \exp\left(b + \frac{\alpha}{\sigma_y}\int_{-\infty}^{y} F(t)dt\right), \tag{49}$$

which satisfies the criteria we discussed earlier for g provided the constant b is sufficiently large.[20] The SDF in (46) fits the moment function in Theorem 1 (with an appropriate choice of factorization as in (20)). This example highlights the power of the generalized transform: rather than specifying a complicated process for g directly, we can utilize the flexibility in choosing $g(y)$ for some tractable process y (such as an affine process).

[19] In the special case where $m(t, X_t)$ follows the exponential-affine form in (39), one can show that if consumption shocks are *i.i.d.* and X_t is affine, the price of consumption shock has to be constant.

[20] To satisfy the condition $\mathrm{Cov}_t(dC_t, dH_t) > 0$, it suffices to have $b > \gamma \ln\left(1 + \frac{\alpha}{\gamma\sigma_c}\right)$.

5.2 Heterogeneous agents economy

The generalized transform can also be used to solve models with heterogeneous agents. The heterogeneity could be about preferences or beliefs. Earlier work include Dumas (1989) and Wang (1996), among others. The stochastic discount factors in the general form of these models will be implicit functions of stochastic state variables. However, even in these general cases, the generalized transform can still be applied.

For illustration, we assume that there is a single risky asset in an endowment economy. Two infinitely-lived agents A and B have time-separable preferences over consumption stream $\{c_t\}$:

$$U_i(c) = E_0\left[\int_0^{\infty} u_i\left(c_t, t\right) dt\right], \quad i = A, B \tag{50}$$

We model the dividend process D from the risky asset as part of a general affine jump-diffusion. Specifically, suppose log dividend $d_t = \iota_1 \cdot X_t$, where X_t is a vector that follows the process (9). This model can easily capture features such as predictability in dividend growth, stochastic volatility, or time-varying probability of jumps. For example, to get predictability of dividend growth, we can assume

$$dd_t = g_t dt + \sigma_d dW_t^d \tag{51a}$$

$$dg_t = \kappa_g\left(\overline{g} - g_t\right) dt + \sigma_g dW_t^g \tag{51b}$$

where g_t is the expected growth rate of dividends, which follows an Ornstein-Uhlenbeck process with long run mean $\overline{g}$. The state variable is then given by $X_t = [d_t \ g_t]'$, and it is straightforward to verify that X is affine with coefficients implied by the dynamics of d_t and g_t. Finally, as discussed in Section 3.2, other non-AJD processes with tractable conditional characteristic functions can be used for X_t as well.

Assuming markets are complete, we can first solve for the optimal allocations through the social planner's problem

$$\max_{\{C_A, C_B\}} E_0\left[\int_0^{\infty} \left\{u_A\left(C_{A,t}, t\right) + \lambda u_B\left(C_{B,t}, t\right)\right\} dt\right], \tag{52}$$

which has constant relative Pareto weight λ and is subject to the market clearing condition $C_{A,t} + C_{B,t} = C_t$. For concreteness, consider the case where the agents have power utility and differ only in their relative risk aversion, $u_i(c,t) = e^{-\rho t} c^{1-\gamma_i}/(1-\gamma_i)$, with $\gamma_A \neq \gamma_B$. The optimal allocations can be solved through the planner's first order conditions and the market clearing condition. Except for a few special cases,[21] agent A's equilibrium consumption is an implicit function of aggregate endowment

$$C_{A,t} = f\left(D_t^{1-\frac{\gamma_B}{\gamma_A}}\right) D_t. \tag{53}$$

Then, the unique stochastic discount factor in this economy is given by

$$\xi_t = e^{-\rho t} f\left(D_t^{1-\frac{\gamma_B}{\gamma_A}}\right)^{-\gamma_A} D_t^{-\gamma_A} = e^{-\rho t} g(d_t) e^{-\gamma_A d_t}. \tag{54}$$

where g is an implicit function that is smooth and bounded. The generalized transform can now be applied when we use the discount factor to price claims. For example, the price of a zero coupon bond that pays one unit of consumption at time T is

$$B(t,T) = E_t\left[\frac{\xi_T}{\xi_t}\right] = \frac{e^{-\rho(T-t)}}{\hat{f}(d_t) e^{-\gamma_A d_t}} E_t\left[e^{-\gamma_A \iota_1 \cdot X_T} g(\iota_1 \cdot X_T)\right], \tag{55}$$

where Theorem 1 can be applied to evaluate the expectation in (55). In case where there is an explicit formula for g, the transform may be known in closed form. However, even when g does not have a closed form solution, one will only need to compute $\hat{g}$ numerically once: this single calculation can then be used for the valuation of a variety of securities and need not be re-computed for different horizons.

Besides heterogeneous preferences, the above model framework can also be used to study heterogeneity in beliefs across agents, provided that the beliefs of the agents satisfy the "affine-disagreement framework" (Chen, Joslin, and Tran (AER 2010)).[22] We can also extend the model to $N > 2$ agents, which will only change the functional form of g for the SDF in (54) (the transform

[21] Wang (1996) shows that the model can be solved in closed form when $\gamma_A = n\gamma_B$ where $n = 2, 3, 4$.

[22] See Appendix D for more discussion on affine differences in beliefs. Bhamra and Uppal (2010) provide a recent treatment of heterogeneous preferences and beliefs when the underlying uncertainties are *i.i.d.*

remains one-dimensional), or to a model of international finance with multiple goods and multiple countries as in Pavlova and Rigobon (2007).

5.3 Multiple goods economy

Our next example is to solve for the equilibrium in an endowment economy with two assets. Several recent papers have studied the general equilibrium effects of multiple sources of consumption in an endowment economy. 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 (2008) (analytical results for N *i.i.d.* trees and power utility).

There are two assets (or two types of consumption goods) in the economy, both in unit supply, with dividend streams $D_{1,t}dt$ and $D_{2,t}dt$. As in the previous section, 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 ($d_{1,t} = \iota_1 \cdot X_t, d_{2,t} = \iota_2 \cdot X_t$), which follows an affine jump-diffusion (9). Again, 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.

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

$$U(c) = E_0 \left[\int_0^\infty e^{-\rho t} \frac{C_t^{1-\gamma} - 1}{1-\gamma} dt \right], \tag{56}$$

where aggregate consumption is a CES aggregator of the two goods $D_{1,t}$ and $D_{2,t}$,

$$C_t = \left(D_{1,t}^{(\epsilon-1)/\epsilon} + \omega D_{2,t}^{(\epsilon-1)/\epsilon} \right)^{\epsilon/(\epsilon-1)}. \tag{57}$$

The parameter ϵ is the elasticity of intratemporal substitution between the two goods, and ω 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.*, 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 (2008), both of which assume *i.i.d.* dividend growth.

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

$$
\begin{aligned}
P_{i,t} &= E_t \left[\int_0^\infty \frac{\xi_{t+u}}{\xi_t} D_{i,t+u} du \right] \\
&= \left(D_{1,t}^{(\epsilon-1)/\epsilon} + \omega D_{2,t}^{(\epsilon-1)/\epsilon} \right)^{\gamma\epsilon/(\epsilon-1)} \int_0^\infty e^{-\rho u} E_t \left[\frac{D_{i,t+u}}{\left(D_{1,t+u}^{(\epsilon-1)/\epsilon} + \omega D_{2,t+u}^{(\epsilon-1)/\epsilon} \right)^{\gamma\epsilon/(\epsilon-1)}} \right] du. \quad (58)
\end{aligned}
$$

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. To map the expectation in (58) into the generalized transform, we rewrite the expectation inside the integral of (58) as:

$$
\begin{aligned}
E_t \left[\frac{D_{1,s}}{\left(D_{1,s}^{(\epsilon-1)/\epsilon} + \omega D_{2,s}^{(\epsilon-1)/\epsilon} \right)^{\gamma\epsilon/(\epsilon-1)}} \right] &= E_t \left[\frac{e^{(1-\gamma/2)d_{1,s} - \gamma/2 d_{2,s}}}{\left(2\cosh\left(\frac{\epsilon-1}{\epsilon} \frac{d_{1,s} - d_{2,s}}{2} \right) \right)^\gamma} \right] \\
&= E_t \left[f\left(\alpha \cdot X_s\right) g\left(\beta \cdot X_s\right) \right], \quad (59)
\end{aligned}
$$

where

$$
\begin{aligned}
f(x) &= e^x \\
g(x) &= \frac{1}{(2\cosh(x))^\gamma}
\end{aligned}
$$

and

$$
\alpha = \left(1 - \frac{\gamma}{2}\right)\iota_1 - \frac{\gamma}{2}\iota_2, \quad \beta = \frac{\epsilon - 1}{2\epsilon}(\iota_1 - \iota_2).
$$

Since X is affine and $g \in \mathcal{S}^*$, Theorem 1 readily applies to (59). When the increments of X are *i.i.d.*, the conditional characteristic function for X is known explicitly, which Martin (2008) uses to compute (59) following a Fourier transform for g. In this i.i.d. increment case, the process for X

can be generalized to any Lévy process as in Section 3.2.

Several observations are in order. First, when we introduce additional state variables to X within the affine framework, due to the time-separable utility function, these additional state variables do not directly enter into the pricing equation (58). Hence, adding these new state variables do not increase the dimension of the Fourier transform. 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, Schneider, and Tuzel (2007), where the two trees are interpreted as nonhousing consumption and housing services. It is also convenient to add preference shocks that are *pl-linear* in the state variables. Third, the model can allow for multiple trees using the multi-dimensional version of the generalized transform in Proposition 2. Finally, although our method computes stock values as an average of exponential affine functions of the state, when pricing options on stocks, the exercise regions will typically be a non-linear function of all covariates. In this case, we can proceed as in Singleton and Umantsev (2002) to approximate the option value by linearizing the exercise region and using the closed form valuation of the payoff.

5.4 A concrete 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 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

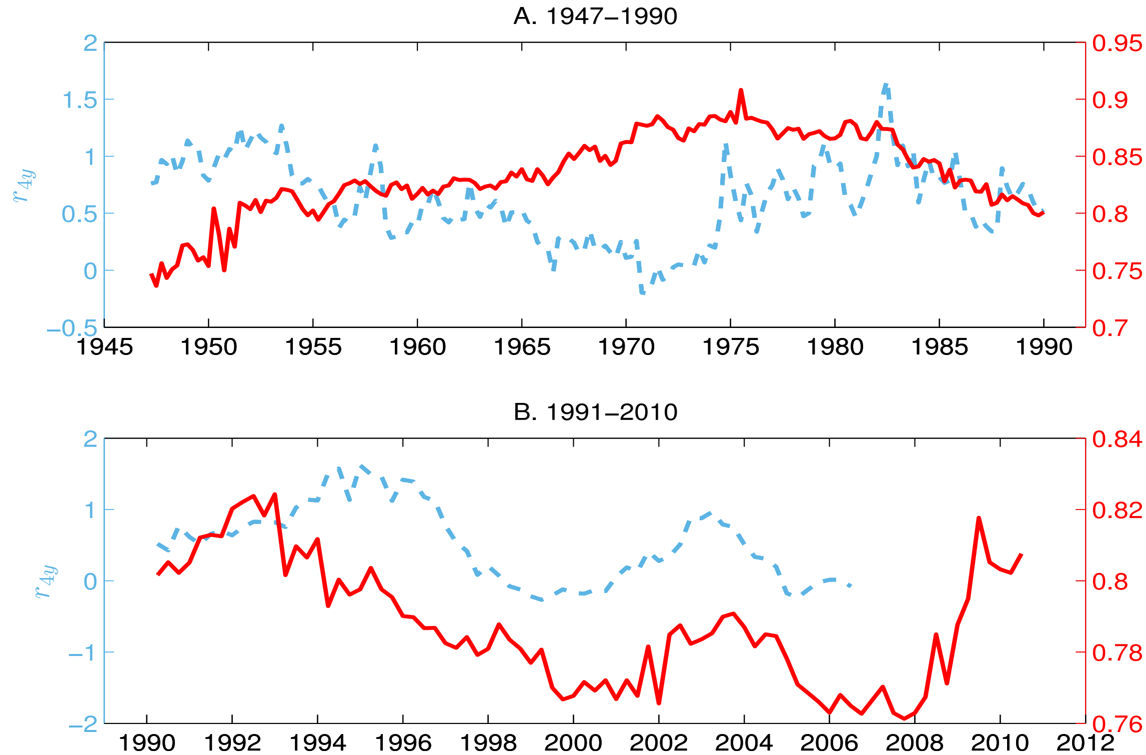

Figure 1: **Long-term returns and labor share pre and post-1990.** Data is quarterly and the sample period is 1947Q1 to 2010Q2. L/C is the share of labor income to consumption. r_{4y} is the lagged four-year cumulative returns of the market portfolio. We use per-capita consumption (nondurables and services) from the BEA and labor income series constructed following Lettau and Ludvigson (2001a).

and multiple financial assets, which provides very convenient closed-form solutions for asset prices.

Following SV, we plot in Figure 1 the share of labor income and lagged four-year cumulative returns of the CRSP value-weighted market index. The labor share is defined as the ratio of labor income to consumption. Consistent with the findings of SV, the labor share and lagged market return in Panel A of Figure 1 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 the composition effect.[23] 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

[23] In fact, to some extent this effect is already documented in Santos and Veronesi (2006) (Table 2), where they show that the predictive power of the labor share is stronger in the sample 1948-1994 than in 1948-2001.

(2008) argue that declining macroeconomic volatility since the 1980s has played a key role in the decline of the equity premium and of dividend yields.[24] In Figure 2, we plot the volatility of consumption growth for non-overlapping 5-year periods using quarterly data from 1952 to 2010. In addition, we also plot 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 also 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.[25]

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 in Section 5.3, 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 (2008) 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 to ensure 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. The affine family of processes allow us to achieve all of the objectives as follows.

Let log dividends and log labor income be d_t and ℓ_t, and let V_t be a volatility factor. Suppose

[24] They estimate a structural break occurred in consumption volatility in 1991, which motivates our choice of the two subsamples in Figure 1.

[25] Different economic mechanisms may have lead to the decline in consumption volatility and the decline in correlation between labor income and dividends. A number of authors have attributed the great moderation to transparency in monetary policy. On the other hand, Piketty and Saez (2003) demonstrate that in the mid-1980s, income inequality began to rise. This "Great Divergence" is consistent with a increasingly negative correlation between labor income and dividends and may have been linked to structural changes associated with legal changes and the declining role of labor unions (see, for example, Card (2001)).

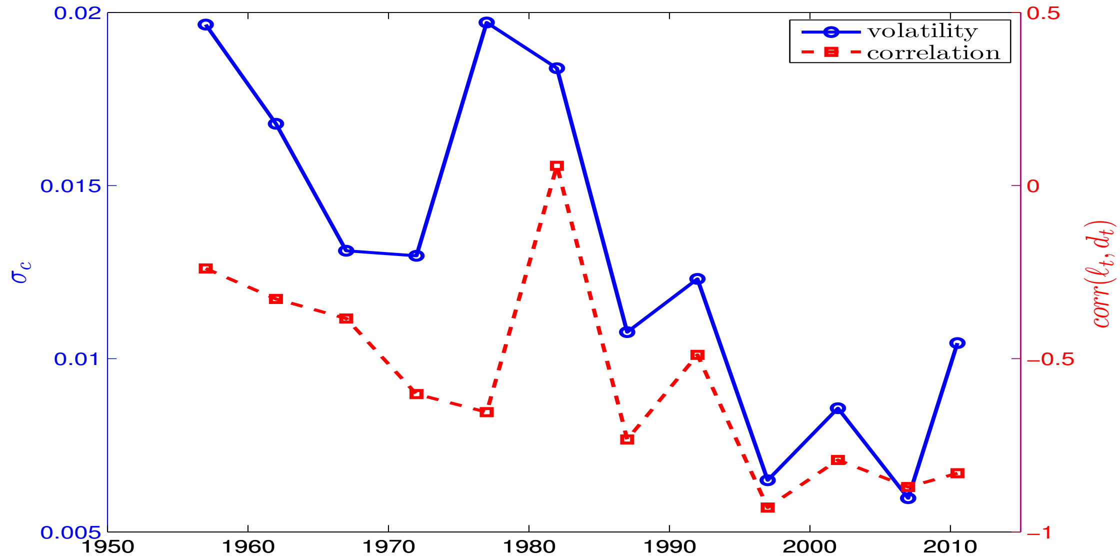

Figure 2: **Consumption volatility and correlation between labor income and dividends.** This figure plots the (annualized) standard deviation of aggregate consumption growth and the correlation between the growth rates of labor income and dividends in 5-year windows. The data are quarterly from 1952Q1 to 2010Q2. At the end of each series we add the estimate for the most recent 5 years.

$X_t = (d_t, \ell_t, V_t)'$ follows an affine process

$$dX_t = \mu_t dt + \sigma_t dW_t. \tag{60}$$

We assume that the conditional drift μ_t is given by

$$\mu_t = \begin{pmatrix} \overline{g} - \quad a\,\kappa_s(\overline{s} - (\ell_t - d_t)) \\ \overline{g} + (1-a)\,\kappa_s(\overline{s} - (\ell_t - d_t)) \\ \kappa_V(\overline{V} - V_t) \end{pmatrix}. \tag{61}$$

This formulation gives the same average growth rate of $\overline{g}$ 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 (61) implies that the drift of s_t will be $\kappa_s(\overline{s} - s_t)$, with $\overline{s}$ 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 (61) also implies that the volatility factor V_t is stationary, with long-run mean $\overline{V}$ and speed of mean reversion κ_V.

The conditional covariance of the factors is given by

$$\sigma_t \sigma_t' = \Sigma_0 + \Sigma_1 V_t, \tag{62}$$

where Σ_0 and Σ_1 can be any positive semi-definite matrices with the the restriction $\Sigma_{0,33} = 0$ so that the volatility factor always remains positive. According to the terminology of Dai and Singleton (2000), this is an $A_1(3)$ model of dividends and labor income. As V_t increases, the volatilities of dividends and labor income will increase. Moreover, the instantaneous correlation between dividends and labor income, ρ_t, also varies with the volatility factor V_t:

$$\rho_t = \frac{\Sigma_0(1,2) + \Sigma_1(1,2)V_t}{\sqrt{\Sigma_0(1,1) + \Sigma_1(1,1)V_t}\sqrt{\Sigma_0(2,2) + \Sigma_1(2,2)V_t}}\,. \tag{63}$$

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 2, and it helps keep the number of state variables small. It is possible to substantially relax the covariance structure using the Wishart process, which allows for fully stochastic covariance between labor income and dividends.

Section 5.3 already explains 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, each of which with one "dividend" payment D_t or L_t. 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

$$E_t[R^e] = (\nabla_X \log M_t)' \, \sigma_t \sigma_t' \, (\nabla_X \log P_t) \,, \tag{64}$$

where $\sigma_t \sigma_t'$, the time t covariance of the factors, is given in (62). Here $(\nabla_X \log M_t)' \, \sigma_t$ in (64) 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 be priced in the equilibrium. The assumption that the volatility factor V_t is independent of dividends and labor income then implies that shocks to dividends and labor income are the only ones with nonzero risk prices in our model. Finally, since the covariance among the factors is time-varying in our model, both the prices of risk and risk exposures can change over time.

Next, we 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.[26] 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 (as 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 (61) (with $a > 0$). The solution of our model allows us to incorporate all of these mechanisms to determine the risk premium for the dividend claim.

Figure 3 plots the conditional risk premium on financial wealth and human capital as functions of the labor share and correlation. We leave the details of model calibration to Appendix F. The parameters are summarized in Table 2. We focus 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 contributor to the risk premium. When volatilities and correlations are high, the

[26]The value of the dividends tree will also be sensitive to covariance shocks (V_t). However, these shocks are not priced in our model by the assumption that they are not correlated with consumption.

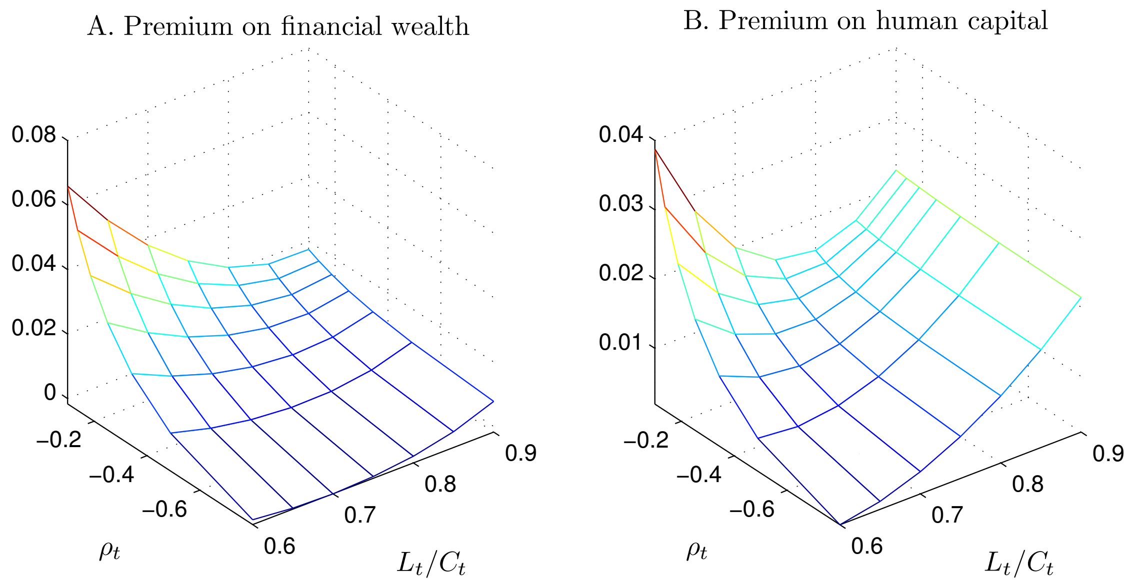

Figure 3: **Conditional risk premium on financial wealth and human capital.** The left panel plots the conditional risk premium on the stock (financial wealth), and the right panel plots the conditional risk premium on human capital as function of labor share L_t/C_t and correlation ρ_t.

model generates significant composition effect: When the correlation $\rho_t = -0.1$, 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 0 for the same rise in labor share. What's more, consistent with the prediction on the price of dividend risk, 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 becoming 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 low, which is again consistent with our earlier analysis of the price of labor income risk. 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.[27] 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 variations in the volatilities and correlation become the main driver of variations in risk premium, the two risk premiums will become positively correlated.

6 Conclusion

We extend the transform analysis in Duffie, Pan, and Singleton (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. By not resorting to an intermediate computation of the (forward) density, this method greatly reduces the dimensionality of such problems, allowing for tractability in a wide range of economic applications.

We demonstrate the power of this method with examples from several areas, including defaultable bond pricing, option pricing, and equilibrium asset pricing models with heterogeneous agents or multiple goods. Underlying all of these examples are the rich dynamics provided by affine processes, allowing for time-varying conditional means and variances as well as jumps occurring with stochastic intensity. Finally, we apply the generalized transform method in a general equilibrium model of

[27] 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).

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

Abramowitz, M., and I. A. Stegun, 1964, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, ninth dover printing, tenth gpo printing edn.

Altman, E. I., B. Brady, A. Resti, and A. Sironi, 2005, "The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications," *Journal of Business*, 78, 2203–2228.

Ang, A., G. Bekaert, and M. Wei, 2008, "The Term Structure of Real Rates and Expected Inflation," *Journal of Finance*, 63, 797–849.

Bakshi, G., C. Cao, and Z. Chen, 1997, "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, 52, 2003–49.

Bakshi, G., and D. Madan, 2000, "Spanning and derivative-security valuation," *Journal of Financial Economics*, 55, 205–238.

Bakshi, G., D. Madan, and F. Zhang, 2006, "Understanding the Role of Recovery in Default Risk Models: Empirical Comparisons and Implied Recovery Rates," Working paper.

Bakshi, G. S., and Z. Chen, 1997, "An alternative valuation model for contingent claims," *Journal of Financial Economics*, 44, 123–165.

Basak, S., 2005, "Asset Prices with Heterogenous Beliefs," *Journal of Banking & Finance*, 29, 28492881.

Bates, D. S., 1996, "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *Review of Financial Studies,*, 9, 69–107.

Bekaert, G., and E. Engstrom, 2010, "Asset Return Dynamics under Bad Environment-Good Environment Fundamentals," Working Paper.

Bekaert, G., and S. Grenadier, 2002, "Stock and Bond Pricing in an Affine Equilibrium," Working Paper.

Bhamra, H. S., and R. Uppal, 2010, "Asset Prices with Heterogeneity in Preferences and Beliefs," Working Paper.

Buraschi, A., F. Trojani, and A. Vedolin, 2010, "Economic Uncertainty, Disagreement, and Credit Markets," Working Paper, Imperial College Business School.

Campbell, J. Y., and J. H. Cochrane, 1999, "By Force of Habit: A Consumption Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

Card, D. E., 2001, "The Effect Of Unions On Wage Inequality In The U.S. Labor Market," *Industrial and Labor Relations Review.*

Carr, P., H. Geman, D. B. Madan, and M. Yor, 2002, "The Fine Structure of Asset Returns: An Empirical Investigation," *Journal of Business*, 75, 305–332.

Carr, P., and D. B. Madan, 1999, "Option valuation using the fast Fourier transform," *Journal of Computational Finance*, 2, 61–73.

Chen, H., 2010, "Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure," *Journal of Finance*, 65, 2171–2212.

Chen, R.-R., and L. Scott, 1995, "Interest Rate Options in Multifactor Cox-Ingersoll-Ross Models of the Term Structure," *Journal of Derivatives*, 3, 53–72.

Cochrane, J. H., F. Longstaff, and P. Santa-Clara, 2008, "Two Trees," *Review of Financial Studies*, 21, 347–385.

Collin-Dufresne, P., R. Goldstein, and J. Hugonnier, 2004, "A General Formula for Valuing Defaultable Securities," *Econometrica*, 72, 1377–1407.

Cox, J. C., and C. Huang, 1989, "Optimum Consumption and Portfolio Policies When Asset Prices Follow a Diffusion Process," *Journal of Economic Theory*, 49, 33–83.

Dai, Q., and K. J. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943–1978.

Dai, Q., K. J. Singleton, and W. Yang, 2007, "Regime shifts in a dynamic term structure model of U.S. Treasury bond yields," *Review of Financial Studies*, 20, 1669–1706.

Duffee, G. R., 1999, "Estimating the Price of Default Risk," *Review of Financial Studies*, 12, 197–226.

Duffee, G. R., 2002, "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57, 405–443.

Duffee, G. R., 2005, "Time Variation in the Covariance between Stock Returns and Consumption Growth," *Journal of Finance*, 60, 1673–1712.

Duffie, J. D., 2005, "Credit Risk Modeling with Affine Processes," *Journal of Banking and Finance*, 29, 2751–2802.

Duffie, J. D., D. Filipovic, and W. Schachermayer, 2003, "Affine Processes and Applications in Finance," *Annals of Applied Probability*, 13, 984–1053.

Duffie, J. D., J. Pan, and K. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

Duffie, J. D., M. Schroder, and C. Skiadas, 1996, "Recursive valuation of defaultable securities and the timing of resolution of uncertainty," *Annals of Applied Probability*, 6, 1075–1090.

Duffie, J. D., and K. J. Singleton, 1999, "Modeling Term Structures of Defaultable Bonds," *Review of Financial Studies*, 12, 687–720.

Dumas, B., 1989, "Two-Person Dynamic Equilibrium in the Capital Market," *Review of Financial Studies*, 2, 157–88.

Dumas, B., A. Kurshev, and R. Uppal, 2009, "Equilibrium Portfolio Strategies in the Presence of Sentiment Risk and Excess Volatility," *Journal of Finance, Forthcoming.*

Folland, G., 1984, *Real Analysis: Modern Techniques and Their Applications*, Wiley-Interscience.

Gabaix, X., 2009, "Linearity-Generating Processes: A Modelling Tool Yielding Closed Forms for Asset Prices," Working Paper.

Hansen, L. P., J. Heaton, J. Lee, and N. Roussanov, 2007, "Intertemporal Substitution and Risk Aversion," in J.J. Heckman, and E.E. Leamer (ed.), *Handbook of Econometrics*, vol. 6, . chap. 61, pp. 3967–4056, Elsevier.

Heston, S. L., 1993, "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327–343.

Joslin, S., 2010, "Pricing and Hedging Volatility in Fixed Income Markets," Working paper, MIT.

Kozhanov, I., 2009, "The Challenge of the Composition Effect," Working Paper.

Le, A., K. Singleton, and Q. Dai, 2010, "Discrete-time Affine^Q Term Structure Models with Generalized Market Prices of Risk," *Review of Financial Studies*, 23, 2184–2227.

Lettau, M., and S. Ludvigson, 2001a, "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56, 815–849.

Lettau, M., and S. Ludvigson, 2001b, "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109, 1238–1287.

Lettau, M., S. C. Ludvigson, and J. A. Wachter, 2008, "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?," *The Review of Financial Studies, Vol. 21, Issue 4, pp. 1653-1687, 2008.*

Longstaff, F. A., and M. Piazzesi, 2004, "Corporate earnings and the equity premium," *Journal of Financial Economics*, 74, 401 – 421.

Lustig, H., and S. Van Nieuwerburgh, 2008, "The Returns on Human Capital: Good News on Wall Street is Bad News on Main Street," *Review of Financial Studies*, 21, 2097–2137.

Madan, D. B., P. P. Carr, and E. C. Chang, 1998, "The Variance Gamma Process and Option Pricing," *European Finance Review*, 2, 79–105.

Mamaysky, H., 2002, "A Model For Pricing Stocks and Bonds," Working Paper.

Martin, I., 2008, "The Lucas Orchard," Working Paper.

Mehra, R., and E. C. Prescott, 1983, "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

Menzly, L., T. Santos, and P. Veronesi, 2004, "Understanding Predictability," *Journal of Political Economy*, 112, 1–47.

Pan, J., 2002, "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics*, 63, 3–50.

Pan, J., and K. J. Singleton, 2008, "Default and Recovery Implicit in the Term Structure of Sovereign "CDS" Spreads," *Journal of Finance*, 63, 2345–2384.

Pavlova, A., and R. Rigobon, 2007, "Asset Prices and Exchange Rates," *Review of Financial Studies*, 20, 1139–1181.

Piazzesi, M., 2005, "Bond Yields and the Federal Reserve," *Journal of Political Economy*, 113, 311–344.

Piazzesi, M., M. Schneider, and S. Tuzel, 2007, "Housing, consumption and asset pricing," *Journal of Financial Economics*, 83, 531–569.

Piketty, T., and E. Saez, 2003, "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, 11, 1–39.

Protter, P. E., 2004, *Stochastic Integration and Differential Equations*, Springer-Verlag, New York, NY, 2nd edn.

Santos, T., and P. Veronesi, 2006, "Labor Income and Predictable Stock Returns," *Review of Financial Studies*, 19, 1–44.

Santos, T., and P. Veronesi, 2010, "Habit formation, the cross section of stock returns and the cash-flow risk puzzle," *Journal of Financial Economics*, 98, 385–413.

Singleton, K. J., 2001, "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111–141.

Singleton, K. J., and L. Umantsev, 2002, "Pricing Coupon-Bond Options And Swaptions In Affine Term Structure Models," *Mathematical Finance*, 12, 427–446.

Wang, J., 1996, "The term structure of interest rates in a pure exchange economy with heterogeneous investors," *Journal of Financial Economics*, 41, 75–110.

Appendix

A Proof of Theorem 1

Throughout, we maintain the following assumptions:

Assumption 1: In the terminology of DPS, (Θ, α, β) 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 $\|\eta_t\|^2 = \Psi_t^2 B(T-t)^\top (H_0 + H_1 \cdot X_t) B(T-t)$,

(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 (17-18).

Assumption 2: The measure F defined by its Radon-Nikodym derivative,

$$\frac{dF}{dP} = \frac{e^{-\int_0^T r_\tau d\tau} e^{\alpha \cdot X_T}}{E_0[e^{-\int_0^T r_\tau d\tau} e^{\alpha \cdot X_T}]}, \tag{65}$$

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

$$E_t\left[e^{-\int_t^T r_\tau d\tau + \alpha + is\beta}\right] e^{-A_{T-t} - B_{T-t}\cdot X_t},$$

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,

$$\begin{aligned}
H &= E_0[e^{-\int_0^T r_\tau d\tau} e^{\alpha \cdot X_T} g(\beta \cdot X_T)] \\
&= F_0\, E_0^F[g(\beta \cdot X_T)] \\
&= F_0 \int g(b) f^F_{\beta \cdot X_T}(b) db \\
&= F_0\, \langle g, f^F_{\beta \cdot X_T} \rangle.
\end{aligned}$$

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

$$\begin{aligned}
H &= F_0\ \langle g, (\hat{f}^F_{\beta \cdot X_T})\check{}\rangle \\
&= F_0\ \langle \hat{g}, \check{f}^F_{\beta \cdot X_T} \rangle.
\end{aligned}$$

This equation holds because of Fourier inversion and the definition of Fourier transform of a tempered distribution. Notice that when both $f_{\beta \cdot X_T}$ and g are in $\mathcal{S}$, 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

$$\begin{aligned}
H &= \langle \hat{g}, F_0 \check{f}^F_{\beta \cdot X_T} \rangle \\
&= \frac{1}{2\pi} \langle \hat{g}, \psi(\alpha + \cdot \beta i) \rangle.
\end{aligned}$$

This last step holds by Assumption 1. This is the desired result. $\square$

B Proof of Proposition 1

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

$$G(\alpha_0; v, n|x, t) = e^{A_t + B_t \cdot x} \sum_{|\xi|=n} \binom{n}{\xi} L(x)^{\xi}, \tag{66}$$

where $L(x)$ is the n-dimensional vector whose ith coordinate is $(\partial_i A + \partial_i B \cdot x)^{1/i}$, ξ is a n-dimensional multi-index, and $(\partial_i A, \partial_i B)_i$ satisfies the ODE

$$\dot{B} = K_1^\top B + \frac{1}{2} B^\top H_1 B - \rho_1 + \lambda_1(\phi(B) - 1), \quad B(0) = \alpha_0, \tag{67}$$

$$\dot{A} = K_0^\top B + \frac{1}{2} B^\top H_0 B - \rho_0 + \lambda_0(\phi(B) - 1), \quad A(0) = 0, \tag{68}$$

$$\partial_1 \dot{B} = K_1^\top \partial_1 B + \partial_1 B^\top H_1 B + \lambda_1 \nabla\phi(B) \cdot \partial_1 B, \quad \partial_1 B(0) = v, \tag{69}$$

$$\partial_1 \dot{A} = K_0^\top \partial_1 B + \partial_1 B^\top H_0 B + \lambda_0 \nabla\phi(B) \cdot \partial_1 B, \quad \partial_1 A(0) = 0, \tag{70}$$

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

$$\partial_m \dot{B} = K_1^\top \partial_1 B + \frac{1}{2} \sum_{i=0}^{m} \binom{m}{i} \partial_i B^\top H_1 \partial_{m-i} B + \partial_{m-1}(\lambda_1 \nabla\phi(B) \cdot \partial_1 B), \ \partial_m B(0) = 0, \tag{71}$$

$$\partial_m \dot{A} = K_0^\top \partial_1 B + \frac{1}{2} \sum_{i=0}^{m} \binom{m}{i} \partial_i B^\top H_1 \partial_{m-i} B + \partial_{m-1}(\lambda_0 \nabla\phi(B) \cdot \partial_1 B), \ \partial_m A(0) = 0. \tag{72}$$

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

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

$$\frac{dF}{dP} = \frac{e^{-\int_0^T r_\tau d\tau} e^{\alpha \cdot X_T} (v \cdot X_T)^n}{E_0[e^{-\int_0^T r_\tau d\tau} e^{\alpha \cdot X_T} (v \cdot X_t)^n]}, \tag{73}$$

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 (9) with

$$\Delta X_t = (K_0 + K_1 X_t) + \epsilon_{t+1}, \tag{74}$$

where ϵ_{t+1} has a conditional distribution which depends on X_t which satisfies

$$E\left[e^{\alpha \cdot \epsilon_{t+1}}\right] = e^{\hat{A}(\alpha) + \hat{B}(\alpha) \cdot X_t}. \tag{75}$$

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 (74–75) is quite flexible and allows for jump-type processes and time-varying covariance (see Le, Singleton, and Dai (2010), for example).

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

$$\begin{aligned} H_D(g, \alpha, \beta) &= E_0\left[\exp\left(-\sum_{u=0}^{T} r(X_u)\right) e^{\alpha \cdot X_T} g(\beta \cdot X_T)\right] \\ &= \frac{1}{2\pi} \langle \hat{g}, \psi(\alpha + \cdot \beta i) \rangle, \end{aligned} \tag{76}$$

where now (17–18) are replaced by

$$\Delta B = K_1^\top B + \hat{B}(B) - \rho_1, \quad B(0) = \alpha + is\beta, \tag{77}$$

$$\Delta A = K_0^\top B + \hat{A}(B) - \rho_0, \quad A(0) = 0. \tag{78}$$

D Affine Differences in Beliefs

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.

D.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:

$$dX_t = \mu_t^A dt + \sigma_t^A dW_t^A + dZ_t^A, \tag{79}$$

where $\mu_t^A = K_0^A + K_1^A X_t$, $\sigma_t^A (\sigma_t^A)^\top = H_0^A + H_1^A \cdot X_t$, and jumps are believed to arrive with intensity $\lambda_t^A = \lambda_0^A + \lambda_1^A \cdot X_t$ and have distribution ν^A (with moment generating function ϕ^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[28] of such equivalent beliefs can be characterized as follows. There exists some vector a such that Agent B believes X follows an affine jump-diffusion satisfying

$$dX_t = \mu_t^B dt + \sigma_t^B dW_t^B + dZ_t^B, \tag{80}$$

[28] More generally, we could consider beliefs of the form $e^{h(x_t) - \int_0^t e^{-h(x_s)} \mathcal{D}^1 e^{h(x_s)} ds}$. Provided the integral term remains tractable, the same analysis applies. Compare also the discussion of essentially affine difference of opinions.

where

1. $\mu_t^B = \mu_t^A + \sigma_t^A(\sigma_t^A)^\top a$

2. $\sigma_t^B = \sigma_t^A$

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. $\lambda_t^B = \lambda_t^A \times E_{\nu^A}[e^{a\cdot Z}]$

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.[29]

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

$$\eta_t = E_t\left[\frac{dP^B}{dP^A}\right] = \exp\left(a\cdot X_t - \int_0^t \left(a\cdot\mu_s^A + \frac{1}{2}\|\sigma_s^A a\|^2 + \lambda_s^A(\phi_\nu(a)-1)\right)ds\right). \tag{81}$$

The variable η_t expresses Agents B's differences in opinion in that when η_t is high, Agent B believes an event is more likely than Agent A believes. We refer to η_t as the *db-density* ('db' stands for "difference in beliefs") process, which differs from the density defining the risk-neutral measure.

While we specify the differences in beliefs exogenously, this does not preclude agents' beliefs from arising through Bayesian updating based on different information sets. For example, when the state variables and signals follow a joint Gaussian process, Bayesian updating can reduce to a difference of beliefs in the form of (81).

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

$$\eta_t = e^{a\cdot X_t}. \tag{82}$$

We assume that the agents have time separable preferences:

$$U^i(c) = E_0^i\left[\int_0^\infty u^i(c_t,t)dt\right], \quad i = A, B. \tag{83}$$

[29] To be precise, as a process $Z^A = Z^B$ (i.e. the functions $Z^i : \Omega\times[0,\infty)\to\mathbb{R}$ are the same). Agents disagree about the probability measures on Ω.

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 ξ_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

$$u_c^A(C_t^A, t) = \zeta^A \xi_t \ , \tag{84}$$

$$u_c^B(C_t^B, t) = \zeta^B \eta_t^{-1} \xi_t \ , \tag{85}$$

where C_t^i is Agent i's equilibrium consumption at time t and ζ^i is the Lagrange multiplier for Agent i's budget constraint.

Market clearing then implies

$$C_t = (u_c^A)^{-1}(\zeta^A \xi_t) + (u_c^B)^{-1}(\zeta^B \eta_t^{-1} \xi_t), \tag{86}$$

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

$$\xi_t = e^{-\rho t}\left[\left(\frac{1}{\zeta^A}\right)^{1/\gamma} + \left(\frac{\eta_t}{\zeta^B}\right)^{1/\gamma}\right]^{\gamma} C_t^{-\gamma}. \tag{87}$$

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

$$\xi_t = e^{-\rho t} g(a \cdot X_t) e^{-\gamma c \cdot X_t}. \tag{88}$$

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.[30] Our method also applies when the two agents

[30]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

have different risk aversion (γ_A and γ_B). In that case, we can still express $h(c_t, \eta_t)$ in the separable form as in (88), and proceed the same way.

In some cases, the mapping of a difference-of-opinion model to the standard setting (81) is not immediate, and requires a careful choice of the state variable X_t. For example, consider the setting where the agents believe that (de-trended) aggregate log consumption, c_t, follows an Ornstein-Uhlenbeck process:

$$dc_t = \kappa_A(\theta_A - c_t)dt + \sigma dW_t^A, \tag{89}$$

$$dc_t = \kappa_B(\theta_B - c_t)dt + \sigma dW_t^B. \tag{90}$$

In this case, the difference in beliefs cannot be dexpressed as in (81) directly. However, by considering an augmented state variable we can return to this form. The state variable $\langle c_t, c_t^2\rangle$ follows the process

$$d\begin{bmatrix} c_t \\ c_t^2 \end{bmatrix} = \begin{bmatrix} \kappa_A(\theta_A - c_t) \\ 2c_t\kappa_A(\theta_A - c_t) + \frac{1}{2}\sigma^2 \end{bmatrix} dt + \begin{bmatrix} \sigma \\ 2c_t\sigma \end{bmatrix} dW_t^A. \tag{91}$$

Since the corresponding 2×2 conditional covariance matrix, $[\sigma, 2c_t\sigma]^\top[\sigma, 2c_t\sigma]$, is affine in $\langle c_t, c_t^2\rangle$, it follows that $\langle c_t, c_t^2\rangle$ is an affine process. Moreover, we return to our standard case since P^B is given by the change of measure as in (81) with

$$a = \sigma^{-2}\begin{bmatrix} \kappa_B\theta_B - \kappa_A\theta_A \\ \frac{1}{2}(\kappa_A - \kappa_B) \end{bmatrix}. \tag{92}$$

More generally, we can have the case where each agent believes that the state of the economy is summarized by the N-dimensional Gaussian state variables, X_t, and each agent believes that X_t satisfies the stochastic differential equation $dX_t = (K_0^i + K_1^i X_t)dt + \sqrt{H_0}dW_t^i$. Again by considering an augmented state variable of the form $\hat{X}_t = \langle X_t, \mathrm{vech}(X_t X_t^\top)\rangle$ we can return to our standard setting.[31] Such techniques are common in the term structure literature with respect to affine and

bounded functions whose Fourier transforms can be computed in terms of incomplete Beta functions. Another option is to write $g(x) = g(x)^{\lceil\gamma\rceil/\gamma} g(x)^{-\lceil\gamma\rceil/\gamma+1}$. In this case, the first functional is *pl-linear* and the second is bounded with Fourier transform known in terms of Beta functions.

[31] For a square matrix M, vech denotes the lower triangular entries written as a vector. Usually, only part of the elements in the extended state vector is needed to maintain the Markov structure.

quadratic term structure models. The procedure generalizes to accommodate models with stochastic volatility ($A_M(N)$ in the parlance of Dai and Singleton (2000)). Following Duffee (2002), we refer to this as *essentially affine difference of beliefs.*

An alternative characterization is to consider the "market price of belief risk", λ_t, in analogy to the usual market price of risk. By defining

$$\lambda_t = \sqrt{H_0^{-1}}(\mu_t^B - \mu_t^A) \ , \tag{93}$$

$$\eta_t = e^{-\int_0^t \lambda_s dW_s^A - \frac{1}{2}\int_0^t \|\lambda_s\|^2 ds} \ . \tag{94}$$

When η_t is exponential affine in X_t, this defines an appropriate Radon-Nikodym derivative for our setting.

D.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, Kurshev, and Uppal (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, $\hat{f}_t^A$ and $\hat{f}_t^B$, and $\hat{g}_t = \hat{f}_t^B - \hat{f}_t^A$ 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,

$$d\begin{bmatrix} c_t \\ \hat{f}_t^B \\ \hat{g}_t \end{bmatrix} = \begin{bmatrix} \hat{f}_t^B - \frac{1}{2}\sigma_c^2 \\ \kappa(\bar{f} - \hat{f}_t^B) \\ -\psi\hat{g}_t \end{bmatrix} dt + \begin{bmatrix} \sigma_c & 0 \\ \frac{\gamma_B}{\sigma_c} & 0 \\ \sigma_{\hat{g},c} & \sigma_{\hat{g},s} \end{bmatrix} dW_t^B. \tag{95}$$

Next, in order to map the model to our standard setting, we define the augmented state variable as $X_t = \langle c_t, \log\eta_t, \hat{f}_t^B, \hat{g}_t, \hat{g}_t^2 \rangle$, where η_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:

$$dX_t = (K_0 + K_1 X_t)dt + \Sigma_t dW_t^B,$$

where

$$K_0 = \begin{bmatrix} -\frac{1}{2}\sigma_c^2 \\ 0 \\ \kappa\bar{f} \\ 0 \\ \sigma_{\hat{g},c}^2 + \sigma_{\hat{g},s}^2 \end{bmatrix}, \; K_1 \quad = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{2\sigma_c^2} \\ 0 & 0 & -\kappa & 0 & 0 \\ 0 & 0 & 0 & -\psi & 0 \\ 0 & 0 & 0 & 0 & -2\psi \end{bmatrix}, \; \Sigma_t = \begin{bmatrix} \sigma_c & 0 \\ -\hat{g}_t/\sigma_c & 0 \\ \gamma_B/\sigma_c & 0 \\ \sigma_{\hat{g},c} & \sigma_{\hat{g},s} \\ 2\sigma_{\hat{g},c}\hat{g}_t & 2\sigma_{\hat{g},s}\hat{g}_t \end{bmatrix}.$$

It is easy to check that the local conditional variance of X_t, $\Sigma_t\Sigma_t^\top$, is affine in X_t so this represents an affine process.[32] Then, it is immediate that η_t takes the form of (81) with $a = \langle 0, 1, 0, 0, 0\rangle$.

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 (α, β, 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 (A58-A61) in DKU is essentially

$$E_0[e^{\alpha\cdot X_t} g(\beta \cdot X_t)] = \frac{1}{2\pi} \int_{b\in\mathbb{R}} \hat{g}(b) \int_{s\in\mathbb{R}} e^{ibs} E_0[e^{(\alpha - is\beta)\cdot X_t}] ds \; db. \tag{96}$$

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.

[32]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.

Disagreement about volatility. Suppose that dividends have stochastic volatility. Under Agent A's beliefs:

$$d\begin{bmatrix} c_t \\ V_t \end{bmatrix} = \begin{bmatrix} \bar{g} \\ -\kappa_V V_t \end{bmatrix} dt + \sqrt{\begin{bmatrix} \sigma_d & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \sigma_{cV} V_t & 0 \\ 0 & \sigma_{VV} V_t \end{bmatrix}} dW_t^A. \tag{97}$$

Here σ_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:

$$d\begin{bmatrix} c_t \\ V_t \end{bmatrix} = \begin{bmatrix} \bar{g} \\ -(\kappa_V - b) V_t \end{bmatrix} dt + \sqrt{\begin{bmatrix} \sigma_d & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \sigma_{cV} V_t & 0 \\ 0 & \sigma_{VV} V_t \end{bmatrix}} dW_t^B. \tag{98}$$

For example, when $b > 0$, Agent B believe that volatility mean reverts more slowly. Using $a = \langle 0, b/\sigma_{VV}^2 \rangle$ we get the *db-density* as in (81).

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:

$$e_t = \int_{-\infty}^{t} e^{-b(t-s)} g_s ds. \tag{99}$$

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

$$d\begin{bmatrix} c_t \\ g_t \\ e_t \end{bmatrix} = \begin{bmatrix} g_t \\ \kappa(\bar{g} - g_t) \\ g_t - b e_t \end{bmatrix} dt + \begin{bmatrix} \sigma_c & 0 \\ 0 & \sigma_g \\ 0 & 0 \end{bmatrix} dW_t^A. \tag{100}$$

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.

$$d\begin{bmatrix} c_t \\ g_t \\ e_t \end{bmatrix} = \begin{bmatrix} g_t + ce_t \\ \kappa(\bar{g} - g_t) \\ g_t - be_t \end{bmatrix} dt + \begin{bmatrix} \sigma_c & 0 \\ 0 & \sigma_g \\ 0 & 0 \end{bmatrix} dW_t^B. \tag{101}$$

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:

$$\begin{aligned} \hat{g}_t^i &= \text{Agent } i\text{'s beliefs about the growth rate of consumption} \\ \hat{g}_t^{ij} &= \text{Agent } i\text{'s beliefs about Agent } j\text{'s belief about the growth rate of consumption} \end{aligned}$$

We can consider the state variable $X_t = [c_t, \hat{g}_t^A, \hat{g}_t^B, \hat{g}_t^{AB}, \hat{g}_t^{BA}]$. Suppose that X_t follows a Gaussian process under both agents beliefs. Agent A's beliefs are such that

$$d\begin{bmatrix} c_t \\ \hat{g}_t^A \\ \hat{g}_t^B \\ \hat{g}^{AB} \\ \hat{g}^{BA} \end{bmatrix} dt = \begin{bmatrix} \hat{g}_t^A \\ \kappa_A(\theta - \hat{g}^A) \\ \kappa_B(\theta - \hat{g}^B) \\ \kappa_{AB}(\hat{g}^B - \hat{g}^{AB}) \\ \kappa_{BA}(\hat{g}^A - \hat{g}^{BA}) \end{bmatrix} dt + \Sigma dW_t^A. \tag{102}$$

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 represent 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, λ_t, of rare disaster. Let $X_t = [c_t, \lambda_t]$. Under Agent A's beliefs,

$$dX_t = \begin{bmatrix} g_A \\ -\kappa_\lambda \lambda_t \end{bmatrix} dt + \begin{bmatrix} \sigma_c & 0 \\ 0 & \sigma_\lambda \sqrt{\lambda} \end{bmatrix} dW_t^A + dZ_t^A, \tag{103}$$

where Z_t^A are jumps in c_t which occur with intensity $\lambda_0 + \lambda_t$ and distribution ν. Suppose that Agent B's beliefs are specified by the db-density of form (81) with $a = \langle b, 0 \rangle$. Then, Agent B's beliefs will be

$$dX_t = \begin{bmatrix} g_A + b\sigma_c^2 \\ -\kappa_\lambda \lambda_t \end{bmatrix} dt + \begin{bmatrix} \sigma_c & 0 \\ 0 & \sigma_\lambda \sqrt{\lambda} \end{bmatrix} dW_t^B + dZ_t^B, \tag{104}$$

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

E Recovery Risk

In this section, we investigate the quantitative impact of stochastic recovery on the pricing of defaultable bonds. Following the discussions in Section 4.1, we directly specify the dynamics of state variables $X_t = [\lambda_t \; Y_t]'$ under the risk neutral probability measure $\mathbb{Q}$:

$$\begin{aligned} d\lambda_t &= \kappa_\lambda(\theta_\lambda - \lambda_t)dt + \sigma_\lambda \sqrt{\lambda_t} dW_t^\lambda, & (105) \\ dY_t &= \kappa_Y(\theta_Y - Y_t)dt + \sigma_Y \sqrt{\lambda_t} dW_t^Y, & (106) \end{aligned}$$

where W_t^λ and W_t^Y are uncorrelated Brownian motions; λ_t is the default intensity of a firm, and the short term interest rate, r_t, is given by

$$r_t = Y_t - \delta \lambda_t. \tag{107}$$

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

The recovery value φ of a bond issued by the firm can depend on the default intensity, the short rate, and other macro and firm-specific variables. Also, in principle it should only take values from $[0, 1]$. One specification for φ that satisfies this requirement and is compatible with the DPS formulation is

$$\varphi(X) = e^{\beta \cdot X} 1_{\{\beta \cdot X < 0\}} + 1_{\{\beta \cdot X > 0\}}.$$

Bakshi, Madan, and Zhang (2006) study such a setting. More generally, the cumulative distribution function of any distribution will take values in $[0, 1]$. Below are some commonly used examples.

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

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

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

Modeling φ 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 φ in those cases.

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

$$\varphi(\lambda) = \frac{a}{1 + b(\lambda - \lambda_0)^2} + c. \tag{108}$$

The constant term c sets a lower bound for φ. Its Fourier transform (excluding the constant c) is

$$\hat{\varphi}(t) = \frac{a\pi}{\sqrt{b}} e^{\lambda_0 i t - \frac{1}{\sqrt{b}}|t|}. \tag{109}$$

We consider two calibrations of $\varphi(\lambda)$ in (108). First, using data from Altman, Brady, Resti, and Sironi (2005), we calibrate $a = 0.68$, $b = 2000$, $c = 0.25$, and $\lambda_0 = -0.014$. The fitted function is "Model I" in Figure 4. The fitted curve is downward sloping and convex. The recovery rate is close

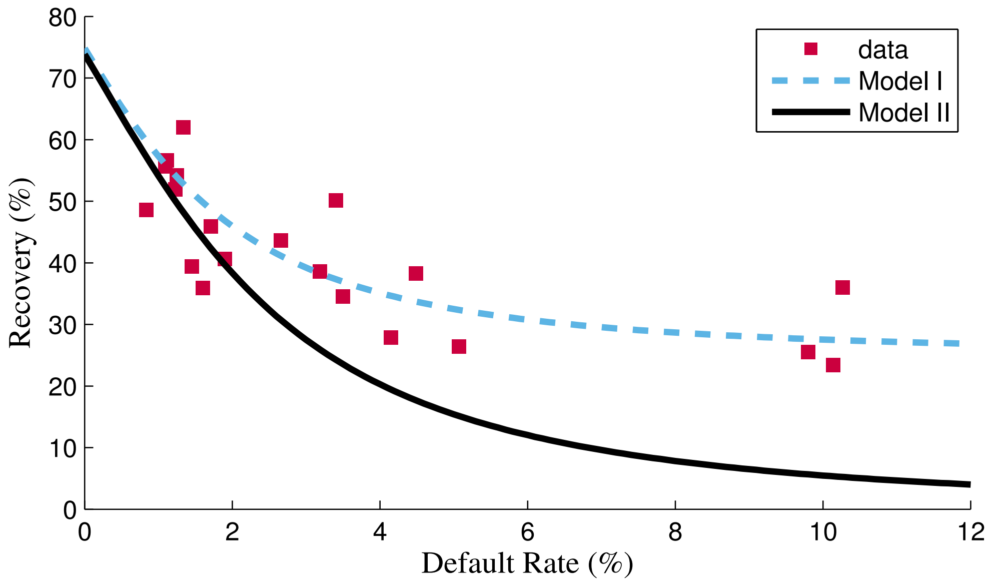

Figure 4: **A Cauchy Model of Aggregate Recovery Rates.** This figure plots the aggregate recovery rates and default rates from Altman, Brady, Resti, and Sironi (2005). The dash line is a Cauchy recovery model fitted to the historical data. The solid line is a Cauchy recovery model with recovery risk premium.

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: it treats the recovery rates in the data the same as the risk-neutral recovery rates, assuming no recovery risk premium. 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 4, which has very 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. The widening gap between the two models implies that the recovery risk premium in Model II is increasing with the aggregate default probabilities.

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

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

Table 1: CALIBRATION OF THE RISK-NEUTRAL DYNAMICS OF λ AND Y

κ_λ	θ_λ	σ_λ	κ_Y	θ_Y	σ_Y	δ
-0.035	-0.08	0.07	0.02	0.10	0.06	0.1

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

$$f(\alpha \cdot X) = \iota_1 \cdot X,$$
$$g(\beta \cdot X) = \frac{a}{1 + b(\iota_1 \cdot X - \lambda_0)^2},$$

where $\iota_1 = [1\ 0]'$.

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 no-jump conditions.[33] 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.

We now use the processes of default intensity λ_t and riskfree rate r_t (105–107) and the recovery model (108) to price a 5-year defaultable zero-coupon bond of a representative firm, which has the same default intensity as the aggregate intensity λ_t. We calibrate the process of λ_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.

The results are reported in Figure 5. The top panels investigate Stochastic Recovery Model I, where the recovery function is fitted to the historical recovery rates (no recovery risk premium); the bottom panels investigate Stochastic Recovery Model II, which assumes that the recovery risk premium increases with aggregate default intensity (see Figure 4). A popular assumption for default recovery in both academic analysis and industry practice is to assume 75% constant loss rate (see e.g., Pan and Singleton (2008)). This value is higher than the historical mean loss rate, which is a parsimonious way to capture the recovery risk premium. Thus, for comparison, we also report the results for two alternative models: (1) recovery of market value (RMV), with constant loss rate

[33]We can either explicitly make the correction for jumps as in Duffie, Schroder, and Skiadas (1996), or use the change-of-measure method in Collin-Dufresne, Goldstein, and Hugonnier (2004).

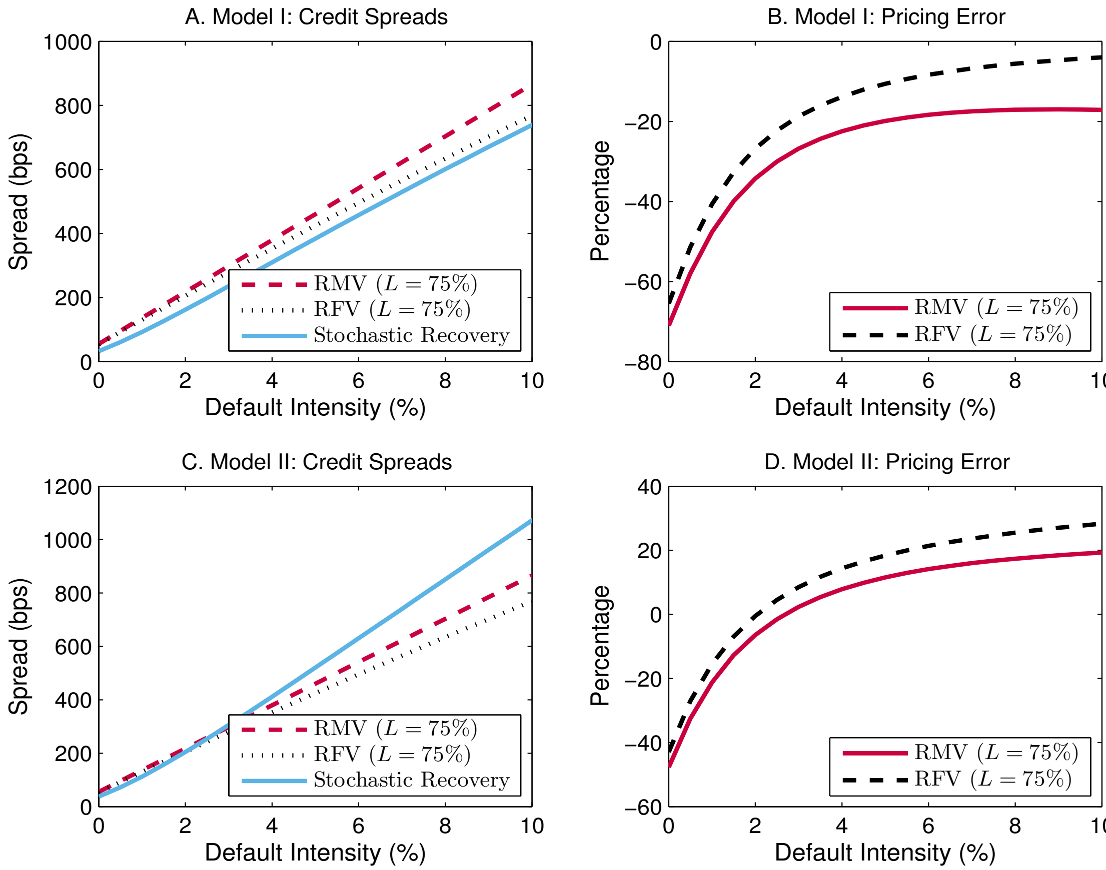

Figure 5: **Credit spreads for 5-year bonds with constant recovery and Cauchy recovery.** For different values of conditional default intensity, this figure plots the credit spreads of a 5-year zero-coupon defaultable bond, and the pricing errors of the RMV and RFV model with constant recovery rates relative to two versions of the stochastic recovery model. "RMV" stands for "recovery of market value"; "RFV" stands for "recovery of face value".

$L = 0.75$; (2) recovery of face value (RFV), with constant loss rate $L = 0.75$.

Panel A shows that, without the recovery risk premium, the credit spreads generated by the stochastic recovery model are lower than the RMV and RFV model with constant recovery rate. Panel B reports the pricing errors of the RMV and RFV model relative to the stochastic recovery model (assuming the latter is the true model), computed as

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

As expected, if there is no recovery risk premium (as in 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 are large. For example, the pricing errors of the RMV model are at least 20%, and can be over 40% when default intensity is low ($< 2\%$).

The results are quite different for Stochastic Recovery Model II. As shown in Panel C, the credit spreads generated by Model II are visibly more convex than the spreads from the RMV or RFV model with constant recovery rate. When default intensity is low, credit spreads are lower for the stochastic recovery model, because the conditional recovery rates are higher than under the constant recovery assumption. As aggregate default intensity rises, the size of the recovery risk premium increases, which lowers the risk-neutral recovery rates and explains the rapid rise in the spreads of Model II.

Panel D gives us a better sense of the size of pricing errors. When the aggregate default intensity is low, the recovery risk premium is small. Thus, the assumption of a 25% constant recovery rate during these times would be too "conservative", which makes the RMV model generate spreads that are too high. The pricing errors can be well over 20% for low values of λ. On the contrary, at times when the aggregate default rates are high, the assumption of 25% constant recovery rate becomes too optimistic. In fact, for $\lambda > 2.5\%$, the spreads in Model II exceed those in the RMV model. The pricing errors are over 10% for $\lambda > 4.5\%$, and can be as high as 20% when the default intensity reaches 10%.

There is another important message in Figure 5. The curvature in the graphs of the spreads and pricing errors suggests that simply adjusting the constant recovery rate in the RMV model does not solve the mispricing problem. Changing the recovery rate amounts to (approximately) parallel-shifting the spreads or pricing errors, and will either exacerbate the underpricing for low λ or overpricing for high λ. These results suggest that it is important to dynamically account for the negative correlation between default intensity and recovery rate when pricing credit-sensitive securities.

F 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

γ	6	$\bar{\rho}_{\ell,d}$	-30.3%
ρ	1%	$\bar{\sigma}_{\ell}$	5.4%
$\overline{g}$	1.5%	$\bar{\sigma}_d$	11.1%
$\overline{S}$	80%	$\sigma_{\infty}(\rho_{\ell,d})$	9.8%
a	$\frac{1}{3}$	$\sigma_{\infty}(\sigma_{\ell})$	0.0018%
κ_s	0.0231	$\sigma_{\infty}(\sigma_d)$	0.017%
κ_V	0.0693	$\sigma_{SS}(V)$	1.07

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 (Σ_0, Σ_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 σ_d. $\bar{V}$ is normalized to be one.

share, $\bar{S}$, to be 75%. As the covariance parameters (Σ_0 and Σ_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 $\bar{\sigma}_\ell = 5.4\%$, $\bar{\sigma}_d = 11.1\%$, and $\bar{\rho}_{\ell,d} = -30.3\%$ 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 $\sigma_\infty(\sigma_d) = 1.7$bp, $\sigma_\infty(\sigma_\ell) = 0.18$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 ℓ or d) fix the free parameters in Σ_0 (3 parameters) and Σ_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 where chosen to qualitatively match the variation found in Figure 2.

Next, we investigate the price of risk for shocks to dividends and labor income, which are transparent and helpful for our understanding of the risk premiums on financial wealth and human capital. From the stochastic discount factor, we can compute the price of dividend risk, which is

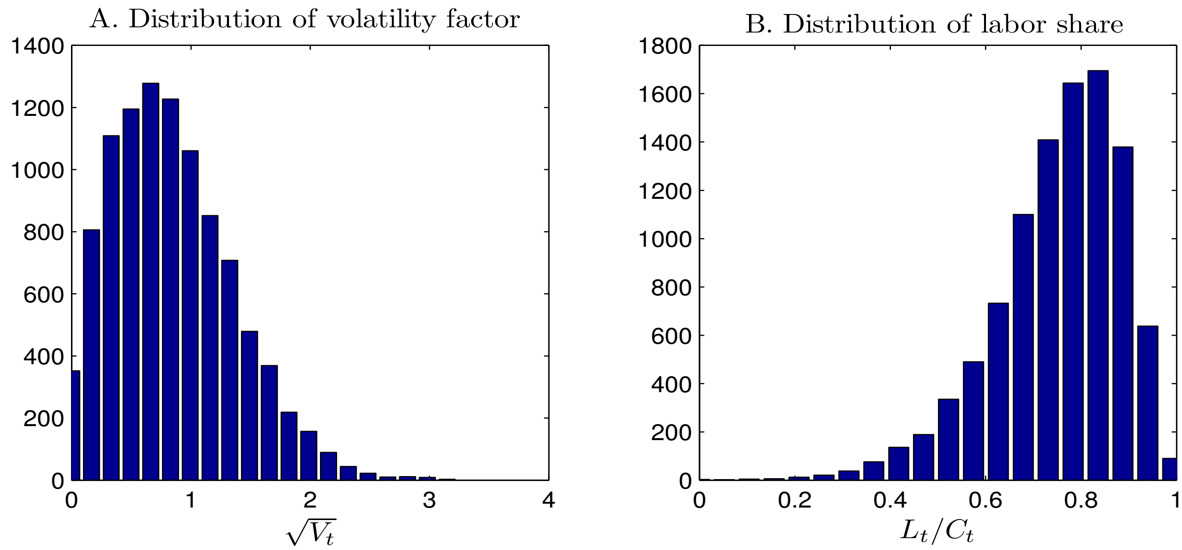

Figure 6: **Simulated distributions.** Panels A and B plot the simulated distributions of the volatility factor $\sqrt{V_t}$ and labor share L_t/C_t.

defined as the risk premium for one unit exposure to dividend shocks:

$$PR_t^d = \frac{\gamma cov_t(c_t, d_t)}{\sigma_{d,t}} = \gamma \left(\frac{L_t}{C_t} \rho_t \sigma_{\ell,t} + \left(1 - \frac{L_t}{C_t}\right) \sigma_{d,t} \right), \tag{110}$$

where c_t is log consumption; $\sigma_{\ell,t}$ and $\sigma_{d,t}$ are the conditional volatilities of labor income and dividends, with their dependence on the volatility factor V_t captured by the time subscripts. On the one hand, holding V_t constant (so that $\sigma_{\ell,t}$, $\sigma_{d,t}$, and ρ_t are all constant), the price of dividend risk will rise as labor share falls as long as $\sigma_{d,t} > \rho_t \sigma_{\ell,t}$ (for example, when dividends are more volatile than labor income or when their correlation is negative), which is the composition effect highlighted in SV. On the other hand, holding labor share fixed, as V_t increases, the volatility of labor income, dividends, and the correlation between the two will increase. All three factors contribute to raise the price of dividend risk. Intuitively, investors become more reluctant to hold financial assets either when there is less labor income to buffer the financial shocks (lower labor share), or when labor income becomes a worse hedge against the financial shocks (higher correlation or higher volatility of labor income).

Finally, if the correlation between labor income and dividends is sufficiently negative and the labor share is sufficiently high, the price of dividend risk can become negative. This is because when the investor is heavily exposed to labor income risk and ρ_t is close to -1, bad news for dividends will tend to be accompanied by good news to labor income, which causes consumption to rise (opposite to dividends).

Similarly, the price of labor income risk measures the risk premium for one unit exposure to labor income shocks:

$$PR_t^\ell = \frac{\gamma cov_t(c_t, \ell_t)}{\sigma_{\ell,t}} = \gamma\left(\frac{L_t}{C_t}\sigma_{\ell,t} + \left(1 - \frac{L_t}{C_t}\right)\rho_t\sigma_{d,t}\right). \tag{111}$$

Provided that $\sigma_{\ell,t} > \rho_t\sigma_{d,t}$ (for example, when the correlation is very small or negative), the price of labor income risk will rise with labor share. Also, holding the labor share constant, the price of labor income risk unambiguously falls as the correlation and volatilities fall.

Figure 7 shows the quantitative effects of the labor share and correlation (volatility) on the price of dividend risk and labor income risk. First, we show how the volatilities of labor income, dividends, and the correlation between the two are tied to the volatility factor V_t. As Panel A shows, both the volatility of labor income and dividends are monotonically increasing in the volatility factor, although dividend volatility varies significantly more than does labor income. This property is an important feature of our model, which is also consistent with the data. In Panel B, the conditional correlation between labor income and dividends is also monotonically increasing in the volatility factor. Since the volatilities and correlation are all monotonic functions of the volatility factor V_t, without loss of generality we use the correlation ρ_t in place of V_t in the plots for the remainder of the paper.

As Panels C and D show, holding the correlation constant, the price of dividend risk is decreasing in labor share, while the price of labor income risk is increasing in the labor share. When holding the labor share constant, both prices of risk increase with the correlation between labor income and dividends (the volatility factor). These features are consistent with our analysis above.

Notice also that while the price of dividend risk is always falling as the labor share rises, the decline is less pronounced when the correlation (volatility) is small. For example, when the correlation is -0.1, the price of dividend risk falls from 0.36 to 0.06 as labor share rises from 0.6 to 0.9; when the correlation is -0.8, price of dividend risk falls from 0.01 to -0.17 for the same increase in labor share. The opposite is true for the price of labor income risk. From (110) and (111) we see that the sensitivity of PR_t^d to L_t/C_t is $\sigma_{d,t} - \rho_t\sigma_{\ell,t}$, which is increasing in V_t under our calibration, whereas the sensitivity PR_t^ℓ to L_t/C_t is $\sigma_{\ell,t} - \rho_t\sigma_{d,t}$, which is decreasing in V_t under our calibration.

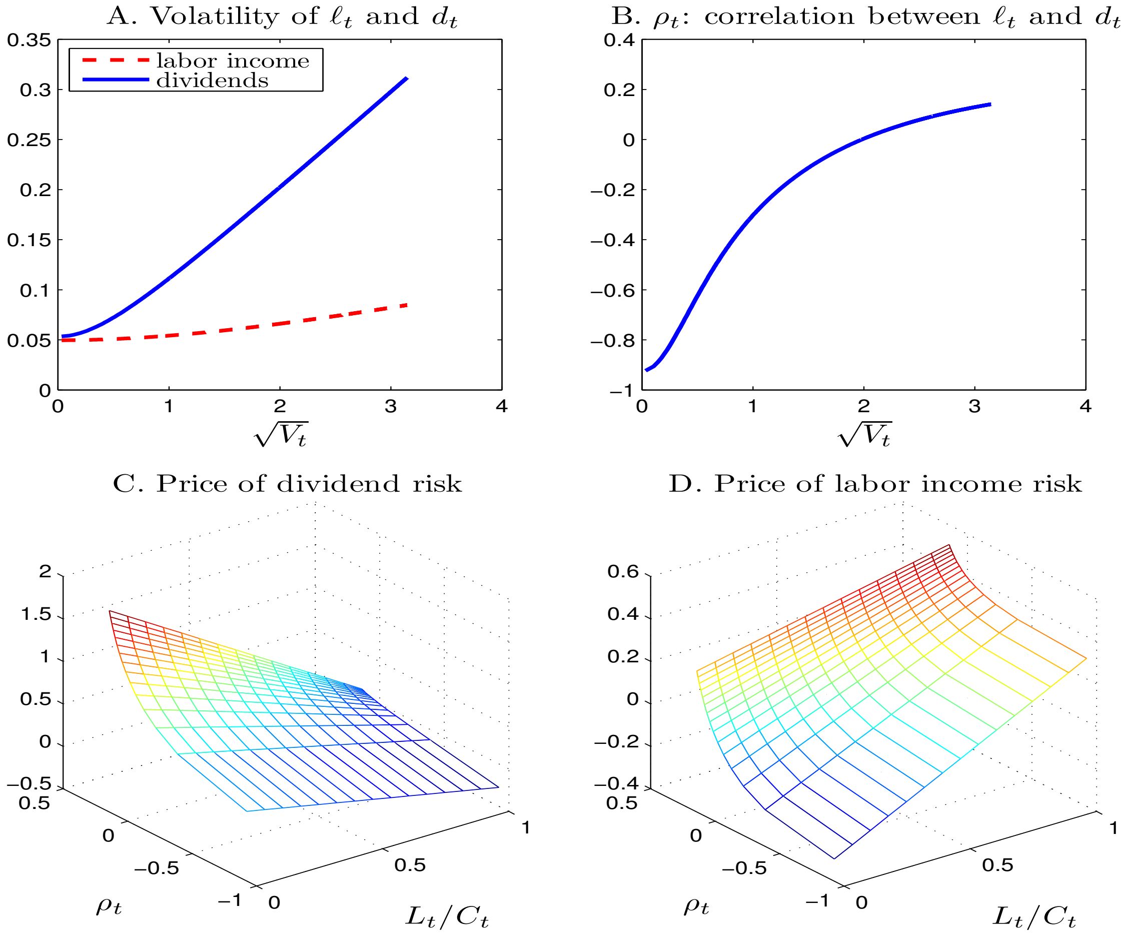

Figure 7: **Price of dividend risk and labor income risk.** Panel A and B plot the conditional volatility of labor income ($\sigma_{\ell,t}$), dividends ($\sigma_{d,t}$), and the conditional correlation between the two (ρ_t), as functions of $\sqrt{V_t}$. Panels C and D plot the conditional price of dividend and labor income risk PR_t^d and PR_t^ℓ as function of labor share L_t/C_t and correlation ρ_t.

Both results follow from the fact that the volatility of dividend income $\sigma_{d,t}$ rises significantly faster with V_t than does the volatility of labor income $\sigma_{\ell,t}$, as illustrated in Panel A of Figure 7.

Panels C and D also show that the decline in the prices of dividend risk and labor income risk with correlation (and volatility) is more pronounced when labor share is lower. Since $\sigma_{d,t}$ rises significantly faster with V_t than $\sigma_{\ell,t}$ and even $\rho_t\sigma_{\ell,t}$, a lower labor share will make the price of dividend risk more sensitive to changes in the volatilities and correlation between labor income and dividends.

Next, we return to assess whether changing covariance can account for the changing relationship between labor income share and expected excess returns in Figure 1. Specifically, our model predicts

Table 3: **Return Forecasting Regressions.** Dependent variable: cumulative excess return on the market over various horizons (quarters). Predictive variable L_t/C_t: share of labor income to consumption. The first set of t-stats is the Newey-West adjusted t-statistics, with the number of lags double the forecasting horizon. The second set of t-stats is the Hansen-Hodrick adjusted t-statistics using the same number of lags as the Newey-West t-stats.

	Panel A: 1947-1990				Panel B: 1991-2010			
	4	8	12	16	4	8	12	16
L_t/C_t	-1.042	-2.440	-4.257	-6.263	3.621	7.233	11.342	15.791
t-stat (NW)	-1.890	-3.664	-4.367	-4.396	1.948	2.185	2.569	3.637
t-stat (HH)	-1.086	-1.401	-1.900	-2.330	1.295	1.406	1.531	1.724
Adj. R^2	0.041	0.125	0.270	0.406	0.101	0.214	0.286	0.363

that the labor income share has little effect on the equity premium when consumption volatility and the correlation between labor income and dividends are low.

In Table 3, we formally examine the difference of return predictability with labor share in the two samples using regressions of long-horizon excess returns on lagged labor shares. The specific form of the regression is

$$r^x_{t,t+K} = \beta_0(K) + \beta_1(K)\frac{L_t}{C_t} + \epsilon_{t+K}, \tag{112}$$

where $r^x_{t,t+K}$ is the cumulative excess return on the market over K quarters, and we consider $K = 4, 8, 12, 16$. For each regression, we report the point estimates of $\beta_1(K)$, the NeweyWest corrected t-statistics (with the number of lags equal to $2K$, as in SV), the Hansen-Hodrick corrected t-statistics and the adjusted R^2.

The regression results are consistent with what we have inferred from Figure 1. Labor share significantly predicts long-horizon returns with a negative sign in the period 1947-1990, which confirms the findings of SV. In the post-1990 period, the regression coefficients on labor share become positive, although they are statistically non-significant at all horizons when using the Hansen-Hodrick standard errors. Based on the moving-window estimates of consumption volatility and correlation between labor income and dividends in Figure 2, volatilities and correlation are both higher pre-1990, which according to our model implies a stronger composition effect (risk premium falls as labor share rises) during this period. As both the volatilities and correlation become smaller in the second sample, the composition effect should indeed become weaker.

On the other hand, in Table 3 the point estimate of the regression coefficients are economically large and significantly positive at longer horizons with respect to the Newey-West standard errors. What could explain such a possible change in the predictive power of labor share? It is possible that some variable that is driving the risk premium on stocks has become correlated with labor share (or happens to be correlated in the relatively short sample). For example, observe that over this period, labor share and covariance are generally positively correlated in Figure 2. In this case, according to Figure 3, decreasing share may be associated with decreasing risk premium due to declining correlation, despite the reverse univariate relationship where covariance is held fixed. Put differently, our covariance variable, V_t, may represent a variable which is correlated with both the regressor and the residual in (112), resulting in biased estimates. Other risks factors may also be at play as well. For example, the correlation between labor share and the consumption-wealth measure CAY (see Lettau and Ludvigson (2001b)) is -0.15 over the period 1952-2010, but jumps to 0.6 in the period 1990-2010. Thus, the fact that CAY predicts the equity risk premium with a positive sign can help explain the regression results in the post-1990 period as well. These possibilities could be explored further by expanding our framework to include the additional covariates.