Essays on Risk Sharing and Pricing

by
Ngoc-Khanh Tran
B.S. Physics, Université Paris-Sud, Orsay, France (2000)
Ph.D. Physics, University of Virginia (2005)
Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2012

Author ....
Sloan School of Management
April 30, 2012

Certified by ....
Leonid Kogan
Nippon Telegraph and Telephone Professor of Management, Professor of Finance
Thesis Supervisor

Certified by ....
Hui Chen
Jon D. Gruber Career Development Professor in Finance, Assistant Professor of Finance
Thesis Supervisor

Certified by ....
Adrien Verdelhan
Assistant Professor of Finance
Thesis Supervisor

Accepted by ....
Ezra Zuckerman Sivan
Nanyang Technological University Professor, Professor of Technological Innovation, Entrepreneurship, and Strategic Management
Director, Ph.D. Program, Sloan School of Management

Essays on Risk Sharing and Pricing

by

Ngoc-Khanh Tran

Submitted to the Sloan School of Management
on April 30, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis consists of three chapters in asset pricing.

Chapter 1 considers an international asset pricing setting with traded and nontraded outputs. It shows that output fluctuations in nontraded industries are a central risk factor driving asset prices in all countries. This is because nontraded industries entail a growth risk that is mostly non-diversifiable, and constitute the largest component of gross domestic product (GDP) of a country. Supportive empirical evidences include; (i) the effect of an industry's growth volatility on the interest rate increases significantly with its nontradability and (ii) carry trade strategies employing currency portfolios sorted on nontraded output growth volatility earns a sizable mean return and Sharpe ratio for US investors.

Chapter 2 considers heterogeneous-agent setting in which agents differ in risk preference, time preference and/or expectations. It shows that, because of equilibrium risk sharing, the precautionary savings motive in the aggregate can vastly exceed that of even the most prudent actual agent in the economy. Consequently, a low real interest rate, resulting from large aggregate savings, can prevail with reasonable risk aversions for all agents. However, as savings rates become extremely sensitive to output fluctuation when savings motive is large, the same mechanism that produces realistically low interest rates tends to make them unrealistically volatile. A powerful isomorphism allows differences in time preference and expectations to be swept away in the analysis, yielding an equivalent economy whose agents differ merely in risk aversion.

Chapter 3 considers a novel tractable and structural pricing framework. It shows that any risk-neutral statistical distribution of state variables can be consistently tied to the economic contents of the underlying pricing model. It establishes this structural linkage by requiring that the economy's stochastic discount factor (SDF) be a proper but unspecified function of the state variables. Consequently, the structural content of the economy as characterized by the SDF can be determined from state variables dynamics through a simple linear differential equation. As a result, state variables' distribution in physical measure can also be recovered.

Thesis Supervisor: Leonid Kogan
Title: Nippon Telegraph and Telephone Professor of Management, Professor of Finance

Thesis Supervisor: Hui Chen
Title: Jon D. Gruber Career Development Professor in Finance, Assistant Professor of Finance

Thesis Supervisor: Adrien Verdelhan
Title: Assistant Professor of Finance

Acknowledgments

I am deeply grateful to my advisors Hui Chen, Leonid Kogan and Adrien Verdelhan for their insights and advices. I am specially indebted to Scott Joslin and Richard Zeckhauser for their insights and supports. I would like to thank Kan Huang, Apur Jain and Indrajit Mitra for many discussions.

I owe the deepest gratitude to my family members, my parents Tran Ung and Luong Thuy, my wife Le Thanh Tu and my brother Tran Ngoc Anh for their love and encouragements. My daughters Tran Thuy Chau and Tran Chau An offer the greatest source of joy and motivation that "drags" me, in one direction and the other, through this graduate study.

Contents

List of Figures

List of Tables

Chapter 1

Growth Risk of Nontraded Industries and Asset Pricing

1.1 Abstract

This paper shows that output fluctuations in nontraded industries are a central risk factor driving asset prices in all countries. This is because nontraded industries entail a growth risk that is mostly non-diversifiable, and constitute the largest component of gross domestic product (GDP) of a country. In interest rate markets, movements in the growth of industries with higher nontradability feed greater risk to the economy, and therefore, stronger downward pressure on the interest rate. Empirically, the effect of an industry's growth volatility on the interest rate increases significantly with its nontradability. In currency markets, this risk factor generates carry trade profits because it induces co-movement of the investor's marginal utility and the exchange rate. Empirically, a carry trade strategy employing currency portfolios sorted on nontraded output growth volatility earns a sizable mean return and Sharpe ratio for US investors. Trade frictions do not alter these mechanisms, although incomplete markets may reverse carry trade profits.

1.2 Introduction

The rational theory and practice of asset pricing center around three fundamental principles: the tradeoff between risk and return, diversification, and no arbitrage. Movements in an economy's nontraded-sector output should play a key role in the determination of domestic asset prices and their differentials across economies, because these are risks that are not easily diversified even in an arbitrage-free international market. This paper shows that the nontraded output growth risk is indeed an important determinant of international asset prices. We adopt a canonical consumption-based exchange economy setting, with multiple countries, multiple traded and nontraded goods, trade costs, and with either complete or incomplete financial markets. A new feature of our model centers on its ability to accommodate *partially traded* goods and services as they actually are in reality. This property allows us to estimate the effects of nontraded output risk that are robust to the possible classification errors in macro data employed. We verify new implications of nontraded output growth risk for the interest rates and carry trade returns using data from the Organisation for Economic Co-operation and Development (OECD) economies.

The main insight of this paper is that the nontradability of an output amplifies the impact of its growth risk on the host economy. From this insight follow all our key conceptual results, which are also verified empirically in the paper. First, at the country level, the fluctuations in gross domestic product (GDP) growth of less open-to-trade economies pose greater risk, incite higher precautionary savings motives, and thus induce relatively lower home interest rates in the cross section of economies. Second, at the industry level, the fluctuations in the output growth of less traded industries also place stronger downward pressure on interest rates. Third, in the currency market, the carry trade strategies that expose investors to larger nontraded output growth risk offer higher returns on average. Fourth, the nontraded output growth risk regulates consumption allocation, moves investors' marginal utility and exchange rates in the same direction, breaks the uncovered interest rate parity, and generates currency forward premia. In contrast, *country-specific* traded output growth risk is much less prominent, because it is subject to diversification via international trades.

The nontraded sector produces goods and services that cannot be consumed outside of the home country. It includes wholesale and retail trade, hotels and restaurants, real estate, financial intermediation, and business activities. Two stylized features of nontraded output stand out. First, nontraded outputs feed the lion's share to the GDP and national aggregate consumption in all countries. Figure 1-1 shows that the ratio of real nontraded output over GDP is substantial among the OECD economies, ranging from 0.5 for Iceland to 0.7 for the United States (US). Second, the *tradabilities*, measured as

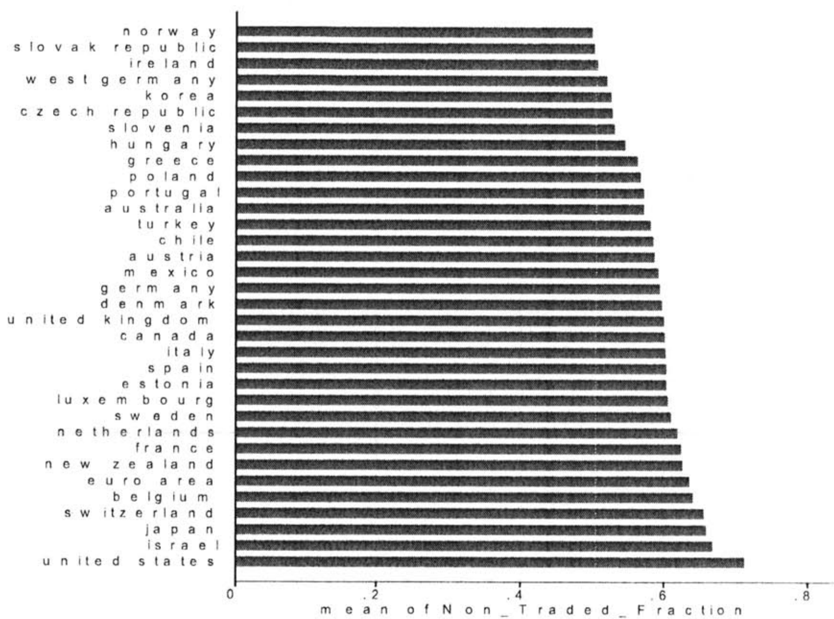


Figure 1-1: Mean of nontraded output-over-GDP ratio, 1971-2010, for OECD countries

the ratio of total import plus export over output, of key nontraded industries are indeed very low. In particular, Table 1.1 shows that the tradabilities in Financial Services, Construction Services, and Other Services rarely exceed 5% across a host of countries.

These

Table 1.1: Services' tradabilities, 1971-2010

Country	Measure	Financial services	Other services	Construction services
Australia	Tradability (%)	0.36	2.02	0.09
	Fraction of GDP (%)	22.31	16.07	6.28
Canada	Tradability (%)	0.69	3.94	0.34
	Fraction of GDP (%)	21.40	20.72	5.67
Czech Republic	Tradability (%)	2.68	18.88	4.47
	Fraction of GDP (%)	14.72	14.44	6.45
Denmark	Tradability (%)	0.67	12.41	2.43
	Fraction of GDP (%)	18.58	24.38	5.58
Hungary	Tradability (%)	1.70	17.01	7.05
	Fraction of GDP (%)	18.15	18.43	3.98
Japan	Tradability (%)	0.21	2.51	1.87
	Fraction of GDP (%)	23.51	23.72	9.78
New Zealand	Tradability (%)	0.22	5.82	0.67
	Fraction of GDP (%)	26.17	17.13	4.83
Norway	Tradability (%)	0.89	9.28	1.29
	Fraction of GDP (%)	14.76	19.71	4.47
Poland	Tradability (%)	0.72	6.95	5.36
	Fraction of GDP (%)	15.94	16.43	06.37
Sweden	Tradability (%)	0.99	14.41	10.16
	Fraction of GDP (%)	20.52	25.17	4.58
Switzerland	Tradability (%)	7.46	2.69	N/A
	Fraction of GDP (%)	19.31	24.88	N/A
United Kingdom	Tradability (%)	2.93	7.77	1.17
	Fraction of GDP (%)	20.39	20.74	5.30
United States	Tradability (%)	0.46	1.43	0.23
	Fraction of GDP (%)	27.65	26.96	5.23

Notes: This table lists the mean of country-specific tradabilities and sizes of financial, construction, and other services for a representative set of 13 OECD countries, 1971-2010. Tradability of services is (one half of) the ratio of total export and import over total output of these services by a country (see (1.26)). Fraction of GDP (or size) of services is the ratio of total output of these services over the GDP of a country. See section 1.7.1 and data appendix for further details.

stylized facts imply that the nontraded output growth volatilities should pose a major source of risk to national economies which should be reflected in the level of domestic interest rates, stock market returns, and real exchange rates. Indeed, figures 1-2 and 1-3 depict a notable inverse relationship between real interest rate and volatility of nontraded (services) output growth across OECD countries. Structurally, this pattern is precisely implied by investors'

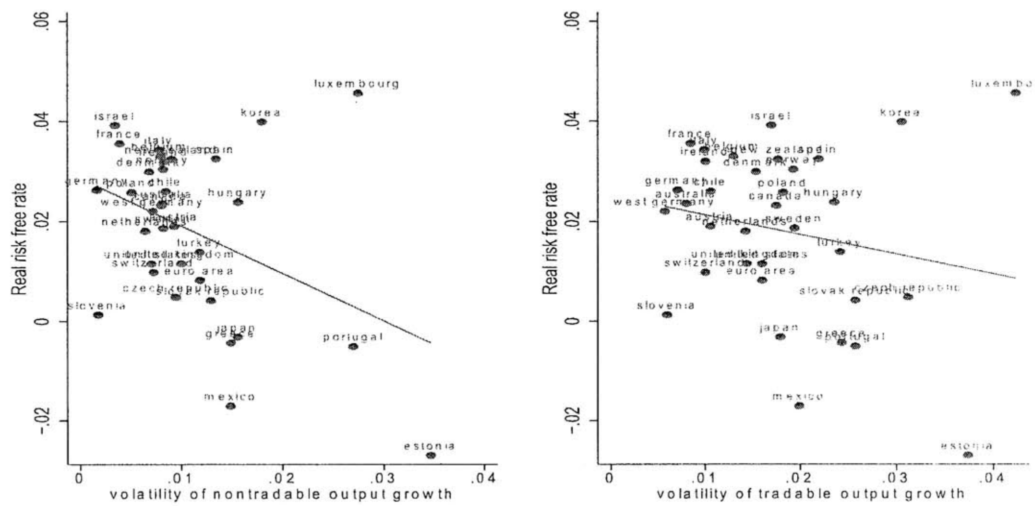


Figure 1-2: Interest rate vs. nontraded and traded per-capita output growth volatility, 1971-2010, for OECD countries

precautionary savings motives. These figures also exhibit a much weaker relationship between interest rate and country-specific traded output growth volatility. This pattern is entirely consistent with the diversification story of traded goods at *global* market scale.

The case of Japan illustrates the insight of nontraded output growth risk. Japan's low real interest rate and the yen's status as a favorite choice for the short currency leg in profitable carry trade strategies are well-known and perplexing issues in international finance. Interestingly, these facts fit neatly with the nontraded risk story proposed here. Among all OECD economies, Japan possesses, in relative terms (i) one of the largest nontraded sectors (figure 1-1), (ii) one of the most volatile nontraded sectors (figure 1-4), and (iii) the most "closed" economy in term of trade-to-GDP ratio (figure 1-5).

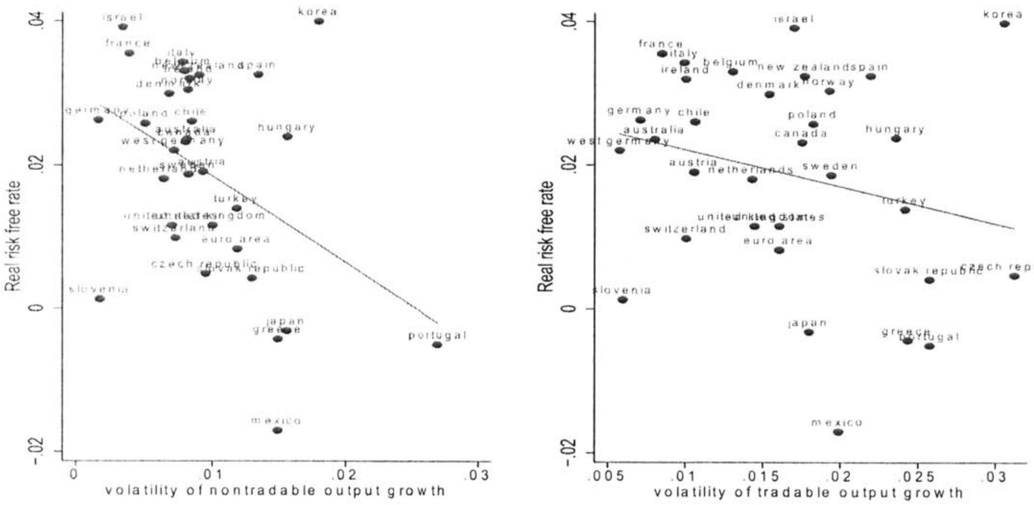


Figure 1-3: Interest rate vs. nontraded and traded per-capita output growth volatility, 1971-2010, for OECD countries excluding Estonia and Luxembourg as outliers

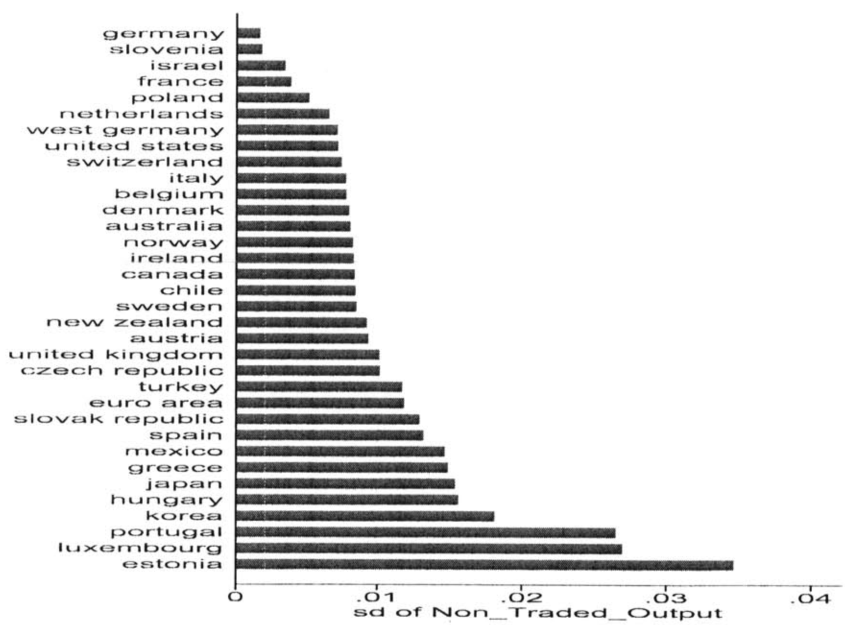


Figure 1-4: Volatility of per-capita nontraded output growth, 1971-2010, for OECD countries

All these empirical regularities suggest that the nontraded output growth risk is more severe in Japan than anywhere else in the OECD. As a result, Japanese risk-free bonds are

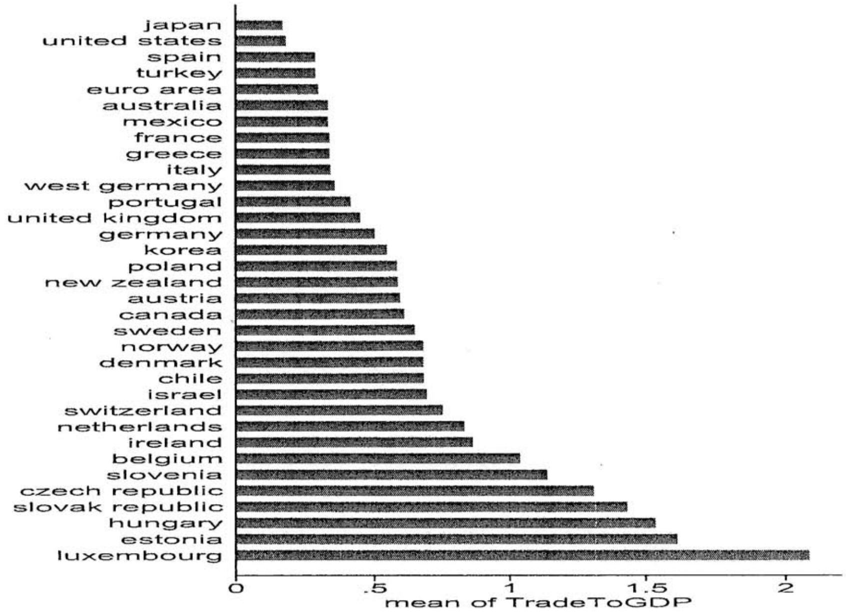


Figure 1-5: Mean of Trade-to-GDP ratio (i.e., openness), 1971-2010, for OECD countries. Trade is defined as the sum of export and import of the country.

highly valuable as a safe hedge against this country-specific risk, and therefore offer both a low yield and are a profitable asset to short in currency investment strategies.

Beyond their dominant impacts at home, nontraded output fluctuations are an important source of risk because they also matter for *all* trade partners of the home country. In the rational framework of this paper, this inter-countries effect underlies the risk and profits of international investment strategies, including currency trades. The transmission of nontraded output shocks is facilitated by two distinctive mechanisms. The first is the *substitution effect*, in which countries can substitute their traded and nontraded consumptions to smooth their overall consumption over time. The second is the *trade effect*, in which a country's traded consumption adjustment influences the traded consumptions of its trade partners by the force of market clearing in traded goods. An example illustrates. Suppose country H receives a windfall of nontraded endowment, which makes nontraded goods relatively cheaper than traded goods. When H's elasticity of intertemporal substitution is lower than that of the traded-nontraded consumption substitution, as documented for many economies (see, e.g., Obstfeld and Rogoff (2001)), H reduces its traded consumption, and its trade partners increase traded consumptions to clear the market and accommodate this adjustment. In other words, the nontraded output risk of a country is actually priced by trade partner countries because it influences partners' consumptions and thus their marginal utilities (or pricing kernel).

We now discuss in depth the specific implications of nontraded risk on interest rates and carry trade returns. In light of the standard precautionary savings motives, volatilities of home nontraded output, trade partners' nontraded output, and global (aggregate) traded output all act to depress home interest rates because these three types of shocks are able to perturb home consumption. However, as mentioned above, although nontraded output risk is primarily internalized, the country-specific traded risk is largely internationalized and thus neutralized in the global pool of traded goods. Consequently, nontraded output volatility should influence home interest rates more strongly than does the home-specific traded output volatility. We discuss aspects of testing this intuitive result below after rigorously formulating the concept of (partial) tradability.

Nontraded output risk is an equally important factor behind carry trade profits. Why do certain currency pairs tend to generate profits, whereas others incur losses in the currency market? Let us consider a strategy of borrowing home currency and lending foreign currency. An adverse foreign nontraded shock simultaneously causes foreign currency to appreciate and home traded consumption to drop (by virtue of the substitution and trade effects mentioned above). That is, with respect to foreign nontraded risk, this strategy pays well when home investors value consumption highly, and vice versa. From the perspective of home investors, such carry trade is a good hedge against foreign nontraded output shocks, and it commands low, possibly negative, expected return to home investors with respect to this risk. By a similar argument, the same carry trade is not a good hedge against home nontraded output growth risk, and thus commands high expected returns to home investors in that regard. The overall expected profit (or loss) of the carry trade is determined by whether home (or foreign) nontraded output growth risk dominates in this process. More specifically, when home nontraded output sector is sufficiently more volatile than that of the foreign trade partner, shorting home and longing foreign currency tend to generate positive expected returns to compensate home investors[1] for bearing the dominating home nontraded risk embedded in the carry trades, and vice versa.

Nontraded output risk then presents a rational cause behind the violation of uncovered interest rate parity (UIP), i.e., the empirical regularity in which increasing-interest-rate currencies tend to appreciate. Lustig and Verdelhan (2007) document that the exchange rates (with respect to the US dollar) of high-interest-rate currencies tend to positively correlate with US consumption growth, and therefore longing high-interest-rate foreign currency and shorting US dollars pose a risk to US investors. These authors consequently attribute this positive correlation pattern to a force that breaks UIP. Movements in nontraded output sectors offer a natural way to rationalize this positive correlation. In our setting, countries having stable nontraded output sectors tend to be associated with high-interest-rate currencies. Thus, for the carry trades that pair US dollars with these currencies, US nontraded output risk dominates its foreign counterpart. As explained above, the dominating US nontraded output shocks generate both a positive correlation between endowment rates and US consumption growth, as well as positive expected profits for the respective carry trade. In contrast, US nontraded output risk does not dominate the carry trade formed between US dollars and low-interest-rate currencies, and as a result these carry trades are not profitable to US investors in the expectation.

In this paper, we devised empirical tests for the effects of nontraded growth risk on interest rates and carry trade returns for OECD economies. The first test concerned interest rates and output growth risk at the industry level. We regressed real interest rates on output growth volatilities of various industries, their tradabilities, and the interaction term, while controlling for other variables. Table 1.5 shows that across OECD economies and on average, the effect of output growth risk on real interest rates increases by 12% when the output's classification moves from traded to nontraded. Another test showed a similar result: the

[1] Carry trade profits to home investors are determined after the carry trade proceeds are converted back into home currency.

volatility of GDP has greater effect on home interest rates when the economy is less open to trades (i.e., having lower ratio of national trade over GDP). The next test concerned profits of investment strategies in currency markets. In particular, sorting currencies based on nontraded output risk and forming carry trade strategies accordingly yield sizable mean returns. Figure 1-6 shows that the long-short strategy on currency portfolios sorted on the volatility of nontraded output growth earns US investors a mean annual real return of almost 3%, and Sharpe ratio of around 20%. Though these strategies are not as profitable as the investment strategy in the US equity index,[2] this figure clearly demonstrates the consistency of the nontraded output risk rationale with the carry trade profits.

Our analysis naturally suggests two-factor pricing model for each country. The factors are nontraded and traded *consumption* growths. We note that in the current setting of exchange economies, the nontraded output is essentially the nontraded consumption and thus is largely internalized within the country. Consequently, shocks in nontraded consumption are always perceived as risk and the corresponding factor price is unambiguously positive.[3] Using carry trade portfolios as test assets and two different data sets, a two-stage GMM procedure gave a statistically significant positive estimate of 32 basis points for nontraded consumption factor price, from the US investors' perspective.

We extended our theoretical analysis to the incomplete asset market setting, where financial assets that are contingent on the nontraded outputs of certain (emerging) economies are not marketable and thus absent from markets. In this incomplete financial market, the nontraded output risk originating from developed economies can still be shared quite efficiently. However, nontraded risk from emerging countries' cannot be shared optimally because of the absence of appropriate assets contingent on these countries' nontraded outputs. In the pooling equilibrium, countries choose to spread this risk evenly within the group of developed countries, and within the group of emerging countries (although not evenly across these two

[2]Based on historical data, the strategy of longing S&P500 index earns real return of 7% and Sharpe ratio of 40% approximately, see e.g., Mehra and Prescott (2008).

[3]In contrast, movements in home traded *consumption* are not necessarily a risk factor to home investors because this consumption is endogenous in the model. Consequently, the factor price associated with traded consumption growth volatility is not necessarily positive.

groups). As a result, in the pooling equilibrium, all of the above results concerning the effects of developed economies' nontraded output risk on other developed economies remain qualitatively intact. However, the effects of nontraded output shocks from emerging economies on other economies are much weaker (because of pooling), or are even reversed, compared to those obtained in the basic setting. To illustrate, a positive shock in an emerging economy's nontraded sector may *decrease* the traded consumption at home *and* in other emerging countries. Consequently, we expect that UIP violation to be more pronounced among currency pairs of developed economies. Bansal and Dahlquist (2000) empirically observe this asymmetry in a mixed data set of developed and emerging economies. In retrospect, the mechanism of an incomplete market thus lends theoretical support to their findings.[4]

The current paper contributes to an important asset pricing literature that attempts to pin down the determinants of asset returns.[5] Different factors have been proposed and found to have statistically significant power in pricing assets in different markets. Nevertheless, many of them are ad-hoc factors that do not necessarily have clear economic intuitions. The nontraded output growth risk that this paper pursues is fully motivated from and thus backed by economic rationales. The concept and modeling of traded and nontraded goods have been widely employed in international economics and international trades. The current study instead brings this keen intuition of output nontradability to the pricing of financial assets. In this aspect our paper builds on the early leads of Stulz (1987), Stockman and Dellas (1989), Backus and Smith (1993), and Zapatero (1995). We extend these analyses by concentrating on the concept of partial nontradability and its dynamic role on prices, in particular the carry trade returns and the underlying risk. While the majority of models in international finance build on the simplified two-country two-good paradigm, the model of this paper works with multiple-country multiple-good setting with the possibility of incomplete financial markets, which is more realistic and promising as advocated by Pavlova and

[4]Bansal and Dahlquist (2000)'s empirical analysis also concern the differential of inflation level in these countries.

[5]This literature expands on the earlier influential Capital Asset Pricing Model (Lintner (1965), Mossin (1966), Sharpe (1964)), Intertemporal Capital Asset Pricing Model (Merton (1973)), Arbitrage Pricing Model (Ross (1976)), and more recent factor pricing model (Fama and French (1993)).

Rigobon (2010). In the presence of multiple economic players who face nontraded risk, we are able to derive explicit and identify the structural factors that contribute to the diversification benefits in both assets and goods markets. In previous literature concerning currency investment strategies, the international diversification benefits are studied mostly under the mean-variance efficiency and reduced-form perspectives, as in Burnside et al. (2008) and Campbell et al. (2010). Other international asset pricing puzzles concerning real exchange rate and stochastic discount factor movement, and possible solutions based on recursive utility (together with a long-run risk component), and habit formation are discussed in the work by Brandt et al. (2006), Colacito and Croce (2011), and Stathopoulos (2011) respectively. Closest to our paper is Hassan (2010)'s, who is the first to analyze the effect of economy's size on carry trade returns. The current paper instead focuses on the role of nontraded risk and makes clear that the economy's size only enter the international pricing dynamics under two premises: (i) size is always coupled with the nontraded output of the host economy, *and* (ii) size's influence is always transmitted by means of international trade. To illustrate, we consider two extreme cases in which we turn off completely one of these two premises: (i) all goods are traded (no nontraded goods), and (ii) all goods are nontraded (countries as isolated islands). In both cases, under the assumption that countries have homogeneous preferences, the sizes of economies do not contribute to the interest rate differentials across countries.

The paper is structured as follows. Section 1.3 presents the basic international asset pricing model with a single traded good and symmetric consumption tastes across countries. Section 1.4 analyzes interest rates and derives testable implications on the relationship between interest rates and nontraded output risk, both with and without trade frictions. Section 1.5 analyzes carry trade strategies and the associated returns, and derives their testable implications. Section 1.6 presents and develops a much more general international asset pricing model with multiple traded goods, arbitrary trade configuration and incomplete financial markets. Section 1.7 conducts empirical tests concerning the pricing of nontraded and traded risk in interest rates and carry trade strategies. Section 1.8 summarizes the main findings. Appendix 1.9.1 presents a short description of data and lists their original sources.

Appendices 1.9.2, 1.9.3 and 1.9.4 present derivations and proofs of technical results.

1.3 Basic model

The basic model of the world economy consists of K countries, engaged in trade with one another and with a single consumption good. Each country also has its country-specific nontraded consumption good, which can be consumed only in that country. We concentrate on the consumption risk in this paper and thus abstract our findings from production aspects of the economy. The countries are endowed with country-specific streams of these traded and respective nontraded goods. Specifically, the endowments (or interchangeably, outputs) $\{\Delta_T^H, \Delta_N^H\}$ are stochastic and follow the country-specific general[6] diffusion processes

$$d\log\Delta_T^H = \mu_T^H dt + \sigma_T^H dZ_T^H; \qquad d\log\Delta_N^H = \mu_N^H dt + \sigma_N^H dZ_N^H; \qquad H = 1\ldots K.$$

where, throughout, the superscript H denotes the country and the subscripts T, N denote the traded and nontraded goods, respectively. In the above equations, Z_T^H and Z_N^H are standard (possibly multi-dimensional) Brownian motions characterizing the country-specific supply shocks of the traded and nontraded sectors. For simplicity, we also omit time index t whenever this omission does not create confusion. Let us first assume that the traded good is shipped without friction around the globe.[7] The market clearing mechanism then simply enforces that traded good outputs from all countries are bundled together, and only the *global* (aggregate) traded endowment Δ_T enters the dynamic

$$\Delta_T \equiv \sum_{H=1}^{K} \Delta_T^H; \qquad d\log\Delta_T \equiv \mu_T dt + \sigma_T dZ_T.$$

In this section, we also assume that investors can trade at least as many financial assets, i.e., contingent claims on these stochastic outputs and risk-free bonds denominated in countries'

[6]That is, the constant moments $\mu_N^H, \mu_T^H, \sigma_N^H, \sigma_T^H$ are not essential for the model's implication, although the geometric Brownian motion specification considerably eases the exposition.

[7]We reinstate the transportation cost in the next section.

currencies, as needed to complete the world market. Incomplete markets are the topic of section 1.6.2. Each country features a representative agent who maximizes the expected utility weighted over traded and nontraded consumptions $C \equiv \{C_T, C_N\}$. It is important to note that in this representative-agent approach, individual investors in each country are assumed to be identical,[8] thus, these are consumptions *per capita*. The period utilities have the following standard form

$$U^H(C^H, t) = e^{-\rho t}\frac{(C^H)^{1-\gamma}}{1-\gamma} = e^{-\rho t}\frac{1}{1-\gamma}\left[\omega_T (C_T^H)^{1-\epsilon} + \omega_N (C_N^H)^{1-\epsilon}\right]^{\frac{1-\gamma}{1-\epsilon}}; \qquad \omega_T + \omega_N = 1, \tag{1.1}$$

where ρ denotes the subjective discount factor. Utility is a power function of the consumption aggregator C^H, which in turn is a function of traded and nontraded consumptions with constant elasticity of substitution (CES). Countries may have different tastes $\{\omega_T, \omega_N\}$ for traded and non traded goods to model the possible effect of home biases in consumption. Their normalization is purely conventional. In this setting, the intertemporal elasticity of consumption is $\frac{1}{\gamma}$, and the elasticity of substitution between traded and nontraded goods is $\frac{1}{\epsilon}$. They satisfy the conditions $\gamma > 0$, $\epsilon > 0$. The interaction between these two substitution effects drives many of the model's implications, as presented below.

Equilibrium consumption allocation

We consider the competitive equilibrium in which each country's representative takes prices as given and dynamically allocates consumption and savings (i.e., investment in financial assets) to maximize her expected utility subject to the budget constraint. Market clearing then consistently determines goods and assets prices. Because the market is complete, equilibrium consumption allocations across countries can be conveniently characterized by (i) formulating the world's representative agent (see Negishi (1960)), and (ii) constructing the *static* optimization scheme in which the world's representative agent maximizes her period utility subject to the aggregate resource constraint at *each* time and for each state (see Cox

[8] An alternative view is to normalize countries' populations to units.

and Huang (1989)). As a result, the world's static optimization problem reads

$$\max_{\{C_T^H\}_{H=1}^K} \sum_{H=1}^{K} \Lambda^H \frac{e^{-\rho t}}{1-\gamma}\left[\omega_T (C_T^H)^{1-\epsilon} + \omega_N (\Delta_N^H)^{1-\epsilon}\right]^{\frac{1-\gamma}{1-\epsilon}} \quad \text{s.t.} \quad \sum_{H=1}^{K} C_T^H = \Delta_T.$$

Note that the intra-country market clearings allow us to explicitly replace the nontraded consumptions by the respective nontraded endowments. The $\{\Lambda^H\}_H^K$ are the countries' Pareto weights. Because individuals are identical within each country, Λ^H is proportional to the product of country H's populations and per-capita wealth. In other words, Λ^H is a measure of H's gross domestic product (GDP).

The law of one price indeed holds for the traded good because the marginal utilities of this good are necessarily equal across countries in equilibrium

$$\Lambda^H \frac{\partial U^H}{\partial C_T^H} = \Lambda^F \frac{\partial U^F}{\partial C_T^F} \equiv M_T \quad \forall H, F = 1 \dots K. \tag{1.2}$$

In principle, these $K-1$ first-order equations together with the traded good's market clearing condition determine the K equilibrium consumptions $\{C_T^H\}_{H=1}^K$. In practice, because marginal utilities are highly nonlinear functions of consumption, the equilibrium allocation is not known in closed form. Instead, we log-linearize this world optimization problem to obtain an approximate but intuitive solution for the sake of analysis. Detailed derivations can be found in appendix 1.9.2. Let the lower-case letters always denote the respective log quantities; $c \equiv \log C$, $\delta_T \equiv \log \Delta_T$, $\delta_N \equiv \log \Delta_N$. In equilibrium, the log per-capita consumptions are given by (see appendix 1.9.2)

$$c_T^H = \delta_T + \frac{1}{\gamma\omega_T + \epsilon\omega_N}\left\{-\rho t - (\gamma - \epsilon)\omega_N \left(\left[1 - \frac{\Lambda^H}{\Lambda}\right]\delta_N^H - \sum_{F \neq H}^{K} \frac{\Lambda^F}{\Lambda}\delta_N^F\right)\right\}. \tag{1.3}$$

where we recall that δ_T is the log *aggregate* traded output. $\Lambda \equiv \sum_{H=1}^K \Lambda^K$ is a measure of the global GDP, therefore $\frac{\Lambda^H}{\Lambda}$ the relative GDP size of countries. This consumption allocation was first obtained by Hassan (2010), who employs a different construction version involving initial wealth transfers among households. His interpretation centers on the relative GDP size, the hedging and the risk aversion effects. In contrast, we focus on various aspects of the nontraded output growth risk in each economy. In particular, we show that the size of economy matters only because it affects the ability of the host country to mitigate its own nontraded output growth risk through international trades.

First, it is reassuring that only the traded good aggregate endowment, but not their country-specific counterparts, explicitly enters the equilibrium consumption allocation. We note that this internationalization has more to do with the global market clearing in the traded good than with the risk sharing. A deeper and surprising result is that the traded output influences log consumptions uniformly across countries in the log-linearization approximation, regardless of the countries' nontraded endowments and sizes. This is an implication of the perfect sharing in traded output risk (i.e., equalized marginal utilities of traded good) and homogeneous preferences across countries.[9] For all countries, the traded consumption [10] necessarily increases with the global supply of the traded good in the current setting.

Second, when $\gamma > \epsilon$, country H's traded consumption c_T^H increases with its trade partners' nontraded endowments δ_N^F and decreases with its own δ_N^H. The intuition is as follows. When the elasticity of substitution between traded and nontraded goods $\frac{1}{\epsilon}$ is higher than that of intertemporal substitution $\frac{1}{\gamma}$, investors are primarily concerned with smoothing consumption over time, and thus are always eager to adjust their traded-nontraded consumption composition to achieve this smoothing. As a result, traded consumptions c_T^H response strongly to nontraded supply shocks. All else being equal, in times of home nontraded surplus ($dZ_N^H > 0$), investors substitute traded consumption ($dc_T^H < 0$) with home nontraded good that has become relatively cheaper. Similarly, in times of foreign nontraded surplus ($dZ_N^F > 0$), foreign investors demand less, and home investors end up consuming more traded goods ($c_T^H > 0$) by force of global market clearing in the traded good. We accordingly make the following assumption throughout. Various empirical estimates reported in Obstfeld and Rogoff (2001) strongly support this assumption.

[9]The setting of heterogeneous tastes and other extensions are analyzed in section 1.6.

[10]The supply shock dZ in $\frac{d\Delta}{\Delta} = \mu dt + \sigma dZ$ is a shock to both endowment growth and endowment level, and the change in log per-capita consumption concerns the growth rate of the per-capita consumption level. For the sake of brevity, we simply refer to the changes in c (or δ) as changes in consumption (or endowment).

Assumption 1: *The elasticity of substitution between the traded and nontraded goods is higher than that of the intertemporal substitution,* $\frac{1}{\epsilon} > \frac{1}{\gamma}$.

The relationships discussed above are then quantified by the proportional coefficients

$$\frac{(\gamma - \epsilon)}{\gamma\omega_T + \epsilon\omega_N} = \frac{\frac{1}{\epsilon} - \frac{1}{\gamma}}{\frac{\omega_T}{\epsilon} + \frac{\omega_N}{\gamma}} = \alpha(\gamma - \epsilon); \qquad \alpha \equiv \frac{1}{\gamma\omega_T + \epsilon\omega_N}, \tag{1.4}$$

which indeed are measures of the relative difference between elasticities of consumption substitution and a weighted substitution elasticity respectively. Later, we will encounter these measures repeatedly in all generalized versions of the current setting.

Finally, in the above expression of equilibrium log consumption, the size of the economy is coupled only to the nontraded output because the traded output is fully internationalized. A more profound explanation is that trade-partner F's nontraded shock affects country H only through the sharing of the traded good. Because the variation in *per-capita* traded consumption of a larger country F projects a larger impact on the common marginal utility,[11] it is clear that a country's size amplifies its nontraded shock impact on the rest of the world.[12] However, it is equally interesting to see that country H's own nontraded shock has a smaller impact on H's log traded consumption when H is larger. This lessened impact arises because a larger country actually finds increasingly less outside room to share traded consumption with its much smaller trade partners.[13] In the limit where $\frac{\Lambda^H}{\Lambda} \to 1$, the super economy H consumes nearly the entire global supply of traded output, which is exogenous and thus non-responsive to whatever happens to H's nontraded output.

[11] We recall that endowment and consumption are per-capita quantities, and thus the marginal utilities of traded good are equalized up to the size factor; $\frac{\Lambda^H}{\Lambda}\frac{\partial U^H}{\partial C_T^H} = \frac{\Lambda^F}{\Lambda}\frac{\partial U^F}{\partial C_T^F}$ $\forall H, F = 1 \ldots K$.

[12] This observation seems particularly germane in the situation in 2009-2010, when Europe and the United States are suffering significant downward shocks to their nontraded production.

[13] It has long been observed that small nations get more from and are more affected by international trade than are large countries, other factors equal. This observation adds an additional dimension to this dynamic.

Stochastic discount factors

In the current consumption-based setting, a country's currency (i.e., its numeraire) is its consumption basket, which is defined as the lowest-cost consumption bundle that delivers one unit of the respective country's utility. Consequently, the stochastic discount factor (SDF) that prices the assets in units of a country's numeraire is country-specific and equal to the country's marginal utility of its consumption aggregator (see appendix 1.9.2); $M^H = e^{-\rho t}(C^H)^{-\gamma}$. We note especially that because these numeraires are different from the traded good, these country-specific SDFs M^H are not the same as the common marginal utility of the traded consumption $M_T = \Lambda^H \frac{\partial U^H}{\partial C_T^H}$.[14] Because in multiple-good settings, assets returns are *not* invariant with respect to numeraires, the country-specific SDFs M^H are the most appropriate choice to price country-specific assets (bonds and stocks).

The log SDF in the log-linearization approximation reads

$$\begin{aligned} m^H &= -\rho t - \gamma\omega_T\delta_T - \gamma\omega_N\left[\delta_N^H - \alpha(\gamma-\epsilon)\omega_T\delta_N^H + \alpha(\gamma-\epsilon)\omega_T\sum_F^K \frac{\Lambda^F}{\Lambda}\delta_N^F\right] \qquad (1.5)\\ &= -\rho t - \gamma\omega_T\delta_T - \gamma\omega_N\left[\delta_N^H - \alpha(\gamma-\epsilon)\omega_T\left(1-\frac{\Lambda^H}{\Lambda}\right)\delta_N^H + \alpha(\gamma-\epsilon)\omega_T\sum_{F\neq H}^K \frac{\Lambda^F}{\Lambda}\delta_N^F\right]. \end{aligned}$$

where $\alpha \equiv (\gamma\omega_T + \epsilon\omega_N)^{-1}$ is a weighted elasticity of substitution, as defined earlier. First, the SDF of any country decreases with the global supply of the traded good. This effect occurs is because countries' traded consumptions increase with the aggregate endowment δ_T and higher consumptions reduce countries' marginal utilities. Reassuringly, δ_T enters countries' log SDF in a uniform manner because the traded good is globally shared without frictions.

Second, the home nontraded endowment δ_N^H impacts the country's SDF m^H through two channels. As a direct effect (the first term within the square brackets), a surge in nontraded consumption (which equals δ_N^H) simply suppresses H's marginal utility and m^H. However, although H needs to consume its entire nontraded endowment, it still is able to somewhat

[14] When we use the common marginal utility of traded consumption, $M_T = \Lambda^H \frac{\partial U^H}{\partial C_T^H}$, to price the assets, prices are in units of the traded good.

mitigate this shock by adjusting its traded good's intake. Indeed, in equilibrium, c_N^H drops (as we have seen earlier), which boosts the marginal utility and prevents m^H from falling all the way.[15] Therefore, this mechanism is driven by the indirect effect (i.e., through trades) and gives rise to the second term within the square brackets, which is reassuringly manifested by the presence of the taste coefficient ω_T associated with the trade. Altogether, the direct effect dominates the indirect,[16] and m^H unambiguously decreases with its own nontraded supply δ_N^H.

Third, country H's SDF decreases with its trade partners' nontraded endowments δ_N^F. Again, this is a consequence of equilibrium consumption allocation and trade effect. All else being equal, a surplus in F's nontraded supply prompts country F to curb, and country H to boost, its traded consumptions. As a result, H's marginal utility and m^H fall. The dependence of a country's stochastic discount factor on its trade partner's nontraded shock is an indirect relationship that arises only through sharing in the traded good.

Finally, the global supply of traded goods impacts all SDFs uniformly when countries have homogeneous preferences. Similar to the way in which the sizes of economies affect consumption allocations, the foreign nontraded endowment δ_N^F matters more for the home SDF m^H when size Λ^F is larger. The same holds for the home country; δ_N^H has greater impact on its own SDF m^H for the larger host country H because larger countries have less outside room to outsource their own nontraded output growth risk. Furthermore, we note that the coefficient associated with δ_N^H is invariably larger in m^H than in any other m^F, the latter is simply an indirect relationship (through trades). We recapitulate these findings in the following result.

Proposition 1 *In the current setting of the world economy, although the nontraded output shock of a country is priced by all of its trade-partner economies, the home nontraded output risk is always more dominant in the home SDF m^H than it is in foreign m^F;* $\left|\frac{\partial m^H}{\partial \delta_N^H}\right| > \left|\frac{\partial m^F}{\partial \delta_N^H}\right|$.

[15] Recall that we assume $\epsilon < \gamma$, an empirically reasonable relationship among the model's parameters, throughout.

[16] We note that $1 - \alpha(\gamma - \epsilon)\omega_T\left(1 - \frac{\Lambda^H}{\Lambda}\right) = \alpha\epsilon + \alpha(\gamma - \epsilon)\omega_T\frac{\Lambda^H}{\Lambda} > 0$ for all $\gamma > \epsilon > 0$.

An immediate consequence of this proposition is that either a positive home nontraded supply shock $dZ_N^H > 0$ (or an adverse foreign shock $dZ_N^F < 0$) will decrease m^H more (or increase m^F less) than m^F, and thus widen the SDF differential $(m^F - m^H)$, i.e., the real exchange rate (see also (1.10)). Therefore, the asymmetry reported in the above proposition is the key to breaking the uncovered interest rate parity (UIP) and to generating carry trade profits in the model as will be shown in more detail in section 1.5.

1.4 Interest rates

In the current multi-country and multi-goods real setting, a country H's interest rate r^H (referred to hereafter as risk-free rate or short rate) is real and defined as the instantaneous return rate of any traded asset that is risk-free with respect to H's currency (i.e., one unit of consumption basket). A conceptually familiar risk-free asset is the consumption-based zero-coupon bond that delivers with certainty one unit of country's consumption basket at maturity. Before embarking on a formal solution and analysis, intuitions suffice to suggest the key role of nontradability on the magnitude of interest rates in the current model. We study settings with either frictionless or costly trades next.

1.4.1 Trades without frictions

For simplicity, we first assume that traded goods can be shipped worldwide without costs. The precautionary savings effects feature prominently in all consumption-risk aspects of interest rates. All else being equal, when an economy exhibits a higher level of uncertainty, the associated bond offering a sure payoff of one consumption unit becomes more valuable and interest rates drop. However, because the country-specific traded outputs are indifferently lumped together into the global supply of traded outputs, it is this global supply (but not the country-specific supplies of traded output) that matters for every country's interest rate. The more volatile the *global* traded output, the lower interest rates in *all* countries. Thus what causes interest rates to differ across countries must be the nontraded outputs.

According to this logic, the volatility of a country's aggregate output, or GDP, is not wholly compounded in the level of interest rate. Thus, the presence of nontraded goods warrants a proper decomposition of GDP into traded and nontraded components, before deciphering the role of GDP movements on the interest rate and other returns.[17]

Volatile nontraded outputs either at home or abroad act to lower home interest rates. A foreign trade partner F with volatile nontraded output transmits its volatility to home country H by consuming highly uneven amount of traded goods. The larger country F is, the stronger is this impact, and the more aggressively H's interest rate decreases with F's nontraded volatility. In contrast, the larger home country H is, the less trading room it finds to outsource its volatility to its trade partners. Consequently, although r^H decreases with own nontraded volatility, such an inverse relationship is weaker when H is larger. All of these intuitions are confirmed by a more quantitative analysis, as presented below. Formally, the interest rate r^H can be determined from the respective SDF M^H through the pricing of the risk-free bond. This bond pays one unit of country's consumption basket in infinitesimal time dt into the future, and its current price is

$$e^{-r^H dt} = E_t\left[\frac{M^H(t+dt)}{M^H(t)}\right] \Longrightarrow r^H = \frac{1}{dt}\left(-E_t\left[dm^H\right] - \frac{1}{2}\mathrm{Var}_t\left[(dm^H)^2\right]\right),$$

where the time subscript indicates conditional moments (expectation and variance). To simplify the exposition, we assume that countries' nontraded outputs are uncorrelated with one another and with the aggregate (global) traded output. This assumption naturally formalizes the stylized premise that nontraded shocks tend to be of an idiosyncratic nature across countries. The assumption simplifies our analysis considerably by separating and hence clearly identifying the role of nontradability on asset pricing. Section 1.7.2 empirically investigates the merit and implications of the assumption. Using the SDF m^H obtained in

[17] Instead, the country's aggregate consumption and its volatility remain truthful indicators of a country's interest rate.

(1.5) yields an expression for risk free rates in equilibrium

$$\begin{aligned} r^H &= \rho + \gamma\omega_T\mu_T - \frac{1}{2}\gamma^2\omega_T^2\sigma_T^2 + \alpha\gamma(\gamma-\epsilon)\omega_T\omega_N\sum_{F=1}^{K}\frac{\Lambda^F}{\Lambda}\mu_N^F \\ &- \frac{1}{2}\alpha^2\gamma^2(\gamma-\epsilon)^2\omega_T^2\omega_N^2\sum_{F=1}^{K}\frac{(\Lambda^F)^2}{(\Lambda)^2}(\sigma_N^F)^2 \\ &+ \alpha\gamma\epsilon\omega_N\mu_N^H - \frac{1}{2}\alpha^2\gamma^2\epsilon^2\omega_N^2(\sigma_N^H)^2 - \alpha^2\gamma^2\epsilon(\gamma-\epsilon)\omega_T\omega_N^2\frac{\Lambda^H}{\Lambda}(\sigma_N^H)^2. \end{aligned} \tag{1.6}$$

All endowment expected growth rates μ's contribute to raising risk-free rates via intertemporal consumption smoothing effect. Given a fixed EIS $\frac{1}{\gamma}$, steadily growing outputs, either at home or abroad, and in either traded or nontraded sectors, always tend to encourage investors to consume more and save less, which causes risk free rates to surge. All endowment growth volatilities σ's act to suppress risk-free rates through the precautionary savings effect, as discussed intuitively above. In particular, the term $(\sigma_N^F)^2$ clearly shows that, in pricing bond H, home investors H are concerned with the nontraded volatility of the trade partner country F's, knowing a shock in that seemingly unrelated sector will affect the traded consumption of F, and thus H itself. All terms containing coefficients $(\gamma-\epsilon)\omega_T$ arise in traded consumption sharing where ω_T characterizes investors' affection for the traded good (trade effect) and $(\gamma-\epsilon)$ their willingness to let nontraded shocks spill over to the traded sector by substituting these two consumption goods (substitution effect).

Interestingly, the first five terms (i.e., all terms in the first line of (1.5)) of risk-free rates are identical across countries, and what drives wedges between countries' real interest rates must have with country-specific nontraded sectors, as anticipated earlier.[18] Apparently, both the nontraded volatility and the size of the host country affect its own interest rate. However, the size contributes only because it influences in how the host country manages to outsource its nontraded shocks to its trade partners; a larger economy internalizes more

[18] The interest rate differential is

$$\Delta r \equiv r^H - r^F = \alpha\gamma\epsilon\omega_N\Delta\mu_N - \frac{1}{2}\alpha^2\gamma^2\epsilon^2\omega_N^2\Delta(\sigma_N)^2 - \alpha^2\gamma^2\epsilon(\gamma-\epsilon)\omega_T\omega_N^2\left(\frac{\Lambda^H}{\Lambda}(\sigma_N^H)^2 - \frac{\Lambda^F}{\Lambda}(\sigma_N^F)^2\right)$$

of its nontraded shocks, which makes bonds more valuable against these uncertainties and depresses its interest rate. Finally, the interest rate (1.6) is derived by employing country-specific consumption basket as numeraire in each country and hence is different from the one obtained by Hassan (2010), who employs the common traded consumption good as numeraire for all countries.[19] Consequently, Hassan's results truly concern carry trade returns, but not interest rate differentials. Our risk free rate expression is more appropriate in the consumption-based setting and for tests using exclusive data on interest rates, as will be shown in section 1.7.2.

A hypothesis concerning interest rates

All findings presented so far paint two very different pictures for the implication of traded and nontraded growth risk on risk-free rates, which warrant a rigorous empirical investigation. Below, we formulate a testable hypothesis that concerns the distinct impact of nontraded output growth risk on the level of interest rate. The actual tests, which indeed confirm the hypothesis, are presented in section 1.7.2. Because *country-specific* traded output risk is internationalized and diversified by means of trades and aggregation, its impact on asset returns should be relatively weak, and we contend the following.

Hypothesis 1: *All else being equal, the impact of country-specific nontraded output growth risk on home interest rate dominates that of the country-specific traded output growth risk.*

The key intuition underlying this hypothesis is the diversification principle, which is directly relevant to the market for traded goods. To see this, we concentrate on the explicit contributions of country-specific traded output volatilities σ_T^H to the interest rate (i.e., omitting terms unrelated to these volatilities)[20]

$$r^H = \# - \frac{1}{2}\gamma^2\omega_T^2\sigma_T^2 = \# - \frac{1}{2}\gamma^2\omega_T^2\frac{1}{dt}\left(\sum_{H=1}^{K}\frac{\Delta_T^H}{\Delta_T}\sigma_T^H dZ_T^H\right)^2.$$

[19]In particular, country nontraded output volatilities σ_N contribute to both interest rates and their differentials as stand-alone terms (i.e., they are not necessarily coupled to economic sizes).

[20]We recall that global (aggregate) traded output is the sum of the country-specific counterparts $\Delta_T = \sum_{H=1}^{K}\Delta_T^H$, and σ_T, $\{\sigma_T^H\}$ are their growth volatilities, respectively.

Clearly, the contribution of country-specific traded shocks dZ_T^H is suppressed by the share of a country's traded output in the world $\frac{\Delta_T^H}{\Delta_T}$. Therefore, unless (i) the traded output shock of a country correlates almost perfectly with global (i.e., aggregate) traded output, or (ii) a country's traded output absolutely dominates the global traded output, home nontraded output volatility $(\sigma_N^H)^2$ affects home interest rate r^H [21] more strongly than $(\sigma_T^H)^2$ for all countries under a mild home bias (i.e., $\omega_N > \omega_T$) condition.[22] The empirical merit of this hypothesis is verified in section 1.7.2.

In a related study, Tian (2011)'s notes that a country's traded consumption growth should be less volatile than the country's traded output growth due to the diversification in the traded good market. Therefore, if the country-specific traded and nontraded output growths are highly correlated and equally volatile, a country-specific positive (negative) shock to these sectors tends to decrease (increase) the domestic relative value of nontraded goods. Consequently, prices of assets contingent on traded output should be more cyclical than those contingent on nontraded output. In the data, she finds that the earnings of traded-good producers are more volatile than those of nontraded-good producers (as many as five times). This result thus provides indirect evidences for the diversification in global market for traded goods.

1.4.2 Costly trades

The previous section's results are derived based on two assumptions, namely, goods are either perfectly traded or nontraded, and trades are frictionless. Consequently, traded goods can be perfectly aggregated globally, which then weakens the country-specific traded output growth risk and gives rise to Hypothesis 1 above. The introduction of trade costs in this section aims to relax both of these simplifications. In particular, the concept of (partial) *tradability* arises naturally by regulating the trade friction. A traded good can become a nontraded good when trade cost is sufficiently high. The tradability is the key to bringing

[21] The impact is characterized by the coefficients associated with $(\sigma_N^H)^2$ and $(\sigma_T^H)^2$.

[22] This condition is $\alpha\left[\epsilon + (\gamma - \epsilon)\omega_T \Lambda^H/\Lambda\right] > \frac{\omega_T}{\omega_N}\frac{\Delta_T^H}{\Delta_T}$.

our model to the data in section 1.7.2.

To model the frictions in trades, we adopt the "iceberg transport cost" approach and analysis of Samuelson (1954), Dumas (1992) and particularly Sercu et al. (1995). In this modeling approach, the commodity trade is not perfect because only a fraction of $\frac{1}{1+\theta}$ of the original traded good that leaves the exporting country arrives at the importing country, and the remainder disappears along the way as a result of this trade friction. To simplify the exposition, we first consider a single good shared by two countries $\{H, F\}$ of similar sizes.[23] The magnitude of θ directly regulates the amount of the good being exchanged (import and export) between countries, and thus determines the *tradability* of that good.[24] With this simplified setting in place, below we focus on the effect of output shocks on interest rates mediated solely by the varying degree of trade friction, while leaving other factors untouched.

The linearity in transport costs is a key modeling advantage because it keeps market completeness intact without further assumption. Consequently, the equilibrium is obtained by solving the static world optimization subject to appropriate global resource constraints.

$$\max_{\{C_H^H, C_F^H, C_H^F, C_F^F\}} U^H(C^H) + U^F(C^F) \equiv e^{-\rho t}\left[\frac{(C_H^H + C_F^H)^{1-\gamma}}{1-\gamma} + \frac{(C_H^F + C_F^F)^{1-\gamma}}{1-\gamma}\right]$$
$$\text{s.t.} \quad C_H^H + (1+\theta)C_H^F = \Delta^H; \quad C_H^F \geq 0; \quad C_F^F + (1+\theta)C_F^H = \Delta^F; \quad C_F^H \geq 0.$$

where $C^H = \{C_H^H, C_F^H\}$ are home consumption components that originate from home and foreign outputs, respectively (the counterpart notation $C^F = \{C_H^F, C_F^F\}$ is preserved for foreign consumption components). Thus, C_F^H is the import by H, which derives from the original amount $(1+\theta)C_F^H$ exported from F. Similarly, C_H^F, which is the import by H, derives from the original amount $(1+\theta)C_H^F$ exported from H. At all time, countries desire to trade to share risk stemming from their unrelated outputs. However, the transport cost hampers risk sharing. Intuitively, if the cost outweighs the benefit of risk sharing, countries opt not to trade and instead fully internalize their endowment shock; $C_H^F = C_F^H = 0$. To determine

[23]It is straightforward to add the transportation costs to the setting of the previous section to have all perfectly traded, partially traded and nontraded goods. Instead, we choose to work with this simplified setting here to concentrate on the role of partial tradability.

[24]Consequently, we drop the subscripts T, N throughout this subsection.

the conditions for commodity market freezing, we assume these conditions are currently not met and that trades take place. Because the shipping incurs a cost, the imported good is always more expensive than the locally endowed good, and countries always deplete their endowed resource before reaching out to the imported resource if they need it. In other words, conditional on trades taking place, there are two mutually exclusive alternatives:

$$\begin{aligned} &\textbf{case 1}\text{: H imports, F exports,} && C_H^H = \Delta^H;\ C_F^H > 0;\ C_H^F = 0;\ C_F^F < \Delta^F, \\ &\textbf{case 2}\text{: H exports, F imports,} && C_H^H < \Delta^H;\ C_F^H = 0;\ C_H^F > 0;\ C_F^F = \Delta^F. \end{aligned}$$

By symmetry, it suffices to study case 1, in which the two FOCs associated with non-binding constraints and the market clearing condition for the home-endowed good establish the remaining equilibrium consumption allocations (i.e., apart from the binding constraints $C_H^H = \Delta^H$, $C_H^F = 0$)

$$C_F^H = \frac{\Delta^F - (1+\theta)^{\frac{1}{\gamma}}\Delta^H}{(1+\theta) + (1+\theta)^{\frac{1}{\gamma}}}; \qquad C_F^F = \frac{(1+\theta)^{\frac{1}{\gamma}}\left[\Delta^F + (1+\theta)\Delta^H\right]}{(1+\theta) + (1+\theta)^{\frac{1}{\gamma}}}. \tag{1.7}$$

It is apparent that the trades require net positive home import $C_F^H > 0$ and commodity market freezes otherwise. We analyze these two regimes in turn.

No-trade regime: Combining cases 1 and 2 yields the following no-trade condition for the commodity market:

$$\text{No-trade conditions:} \qquad (1+\theta)^{-1} < \left(\frac{\Delta^H}{\Delta^F}\right)^{\gamma} < (1+\theta).$$

Clearly, costly transport (large θ), similar outputs ($\frac{\Delta^H}{\Delta^F} \approx 1$), or low risk aversion (small γ) all discourage countries to share risk, and thus enforce the commodity market freeze. In this case, the single good becomes a legitimate nontraded good in any country. Moreover, each country's bond has no hedge power against others' shocks, and the risk-free rate solely reflects the respective country's output risk, as in the consumption-based CAPM. In other words, for each country, the nontraded output volatility is the only risk that matters here.

Costly trade regime: In contrast with the no-trade regime, when friction is moderate and home and foreign outputs are sufficiently different, countries choose to share output risk, although transport costs and trade flows take place in an appropriate direction. Without loss of generality, we continue with case 1 above, in which home is the importing country (or $C_F^H > 0$). Conditional on this being the case, $(1+\theta)^{-1} > \left(\frac{\Delta^H}{\Delta^F}\right)^\gamma$, the home unambiguously curbs its imports when transaction cost increases (C_F^H decreases in θ).[25] However, interestingly, the inverse holds for the exporting country F for all realistic values of transport cost and risk aversion. Contingent on trades taking place, the foreign country actually boosts its export $(1+\theta)C_F^H$ when θ increases to compensate for the increasing loss in the transition.[26] This is because, when home investors are risk averse, their net import C_F^H decreases less than linearly with the transport cost.

As long as trades take place, regardless of their "iceberg-melting" imperfect nature, marginal utilities are equalized across countries ($\frac{\partial U^H}{\partial C^H} = (1+\theta)\frac{\partial U^F}{\partial C^F}$), as are the interest rates in the current setting with a single good. We concentrate on the precautionary savings effect revealed in the interest rates, in which the interplay between output shocks and transport cost dominates.

$$r^H = r^F = \# - \frac{1}{2}\gamma(\gamma+1)\frac{(1+\theta)^2(\Delta^H)^2(\sigma^H)^2 + (\Delta^F)^2(\sigma^F)^2}{[(1+\theta)\Delta^H + \Delta^F]^2}. \tag{1.8}$$

As the transport cost increases, interest rates become increasingly sensitive to home output shocks and decreasingly sensitive to foreign output shocks; $\frac{\partial^2 |r|}{\partial\theta\partial|(\sigma^H)^2|} > 0$, $\frac{\partial^2 |r|}{\partial\theta\partial|(\sigma^F)^2|} < 0$. These behaviors, when combined with the earlier findings that $\frac{\partial C_F^H}{\partial\theta} < 0$ and $\frac{\partial\left[(1+\theta)C_F^H\right]}{\partial\theta} > 0$, precisely support our key thesis that when shocks are of a more nontraded nature (i.e., θ increases), they matter more to the country's asset prices. From the importing country H's perspective, a surge in trade cost coincides with a reduction in trades as its imports C_H^F drop. At the same time, the impact of the country's own volatility σ^H on its interest rate r^H

[25] This is evident from the expression of C_F^H; conditional on trade taking place ($C_F^H > 0$), the numerator decreases and the denominator increases with τ.

[26] $\frac{\partial}{\partial\theta}\left[(1+\theta)C_F^H\right] = \frac{\gamma-1}{\gamma}(1+\theta)^{\frac{1-2\gamma}{\gamma}}\Delta^F - \frac{\gamma-1}{\gamma}(1+\theta)^{\frac{2-2\gamma}{\gamma}}\Delta^H - \frac{1}{\gamma}(1+\theta)^{\frac{1-\gamma}{\gamma}}\Delta^H - \frac{1}{\gamma}(1+\theta)^{\frac{2-2\gamma}{\gamma}}\Delta^H$. For all realistic values of γ and θ, the last two terms are negligible compared with the second term. Then, the trade condition $C_F^H > 0$ immediately implies that $\frac{\partial}{\partial\theta}\left[(1+\theta)C_F^H\right] > 0$.

increases while the impact of foreign volatility σ^F on r^H decreases, all of which is consistent with a reduction in the import in view of the above thesis. Likewise, from the exporting country F's perspective, a surge in trade cost coincides with a boost in trades as its export $(1+\theta)C_H^F$ increases. At the same time, the impact of its own volatility σ^F on its interest rate r^F decreases, whereas the impact of partner's volatility σ^H on r^F increases, which is also consistent with a surge in the export according to the above thesis.[27]

Overall, by making a realistic and smooth transition between traded and nontraded extremes of goods market, the variation in trade frictions implies a structural relationship between nontradability and domestic asset prices. The former is naturally identified as the ratio of trades (import plus export) over output. A refined version of Hypothesis 1 in section 1.4 is

Hypothesis 1A: *All else being equal, a country-specific output growth volatility impacts the home risk-free rate more when the output is less tradable.*

In section 1.7.2, we will test this hypothesis empirically by employing several measures of nontradability, including countries' trade closedness, country-specific and global nontradability at the industry level. Here, we briefly discuss the generalization of the costly trade mechanism to a setting with arbitrary K countries, where subtleties arise because the import from a country does not unambiguously originate in the export of another. In this situation, conditional on trades taking place, each country H is classified into either an importing (I) or an exporting (E) group. Let C_H^H and C_{-H}^H denote country H's consumption components derived from its own and foreign outputs, respectively. Trades take place when $\{C_H^H < \Delta^H; C_{-H}^H = 0\}\ \forall H \in \mathcal{E}$, and $\{C_H^H = \Delta^H; C_{-H}^H > 0\}\ \forall H \in \mathcal{I}$. Because of the ambiguity mentioned above of global import-export source matching, there is now only a single

[27] Obviously, the interest is in the relationship between a country's risk-free rate and its trade volume (i.e., import and export goods that arrive at or leave a country's border). In contrast, the relationship between a country's risk-free rate and its trade partner's exports and imports is not of interest because a portion of these goods is lost in the transition.

market clearing condition, and the world optimization problem reads:

$$\max_{\{C_H^H, C_{-H}^H\}} \sum_H^K e^{-\rho t} \frac{(C_H^H + C_{-H}^H)^{1-\gamma}}{1-\gamma} \quad \text{s.t.} \quad \sum_{H \in \mathcal{E}} C_H^H + (1+\theta) \sum_{H \in \mathcal{I}} C_{-H}^H = \sum_{H \in \mathcal{E}} \Delta^H.$$

Combining FOCs associated with nonbinding constraints[28] and the market clearing condition yields the equilibrium consumption allocations.[29] Subject to trades taking place, mild conditions on the distribution of trades assure that when transport cost θ increases, country H's import C_{-H}^H decreases and its own output volatility σ^H matters more for the domestic risk-free rate r^H.

1.5 Carry trade returns

The underlying risk

Let us consider the typical carry trade strategy from the perspective of country H's investors. (i) at time t borrowing risk-free one unit of base (home) currency H at rate r^H; (ii) immediately converting this into foreign currency F and lending risk-free at rate r^F; and (iii) at time $t+dt$, liquidating the long position in currency F, immediately converting the proceeds into home currency and liquidating the short position in base currency H. It is then obvious that the return on carry trade strategies is beyond the simple difference between the two interest rates involved because the former also concerns the exchange rates. As risk free rates are known at t, in our real and rational setting, the uncertainty rests entirely with the

[28]These FOCs arise from the partial derivatives $\frac{\partial}{\partial C_H^H}$ $\forall H \in \mathcal{E}$ and $\frac{\partial}{\partial C_{-H}^H}$ $\forall H \in \mathcal{I}$.

[29]Conditional on trades taking place, these allocations are

$$\begin{aligned} C_{-H}^H &= \frac{(1+\theta)\sum_{I\in\mathcal{I}}\Delta^I + \sum_{E\in\mathcal{E}}\Delta^E}{(1+\theta)K_I + (1+\theta)^{\frac{1}{\gamma}}K_E} - \Delta^H, \quad \forall H \in \mathcal{I}; \\ C_H^H &= \frac{(1+\theta)^{\frac{1}{\gamma}}\left[(1+\theta)\sum_{I\in\mathcal{I}}\Delta^I + \sum_{E\in\mathcal{E}}\Delta^E\right]}{(1+\theta)K_I + (1+\theta)^{\frac{1}{\gamma}}K_E}, \quad \forall H \in \mathcal{E}. \end{aligned}$$

where K_E and K_I are the numbers of exporting and importing countries, respectively.

exchange rate.[30] In other words, carry trades are bets on exchange rates, and the premia associated with the short-horizon strategies are rewards for bearing the exchange rate risk.

Let S_t denote the spot exchange rate. Our convention is that S_t units of foreign currency F exchange for one unit of home currency H. In the current complete market setting,[31] this exchange rate is $S_t = \frac{M_t^H}{M_t^F}$. The realized excess return (i.e., in excess of the base interest rate r^H) to this carry trade strategy, which shorts bond H and longs bond F, and its expected counterpart, respectively, are

$$\begin{aligned} XR_{t+dt}^{-H,+F} &= \tfrac{1}{dt}\left[\tfrac{M_{t+dt}^F}{M_{t+dt}^H}(1+r_t^F dt)\tfrac{M_t^H}{M_t^F} - (1+r_t^H dt)\right], \\ E_t\left[XR_{t+dt}^{-H,+F}\right] &= -\tfrac{1}{dt}Cov_t\left[dm^H, dm^F - dm^H\right]. \end{aligned} \tag{1.9}$$

Reassuringly, the carry trade expected excess return is the premium associated with the exchange rate risk.[32]

The consumption volatilities contribute to the expected carry trade profits precisely because they perturb both SDFs m^H, m^F. Here our discussion is readily carried over from the previous section's analysis on the SDF. Because traded shocks spread uniformly to all countries, they do not affect exchange rates, and are not counted as risk to be compensated in the carry trades. In fact, they are canceled out in the difference $dm^H - dm^F$. This leaves nontraded volatilities as the sole sources of carry trade risk and return in the current rational setting. Indeed, the log exchange rate follows a simple diffusion process implied structurally

[30] Our settings are real. In practice, there is risk associated with inflation. When we consider short-horizon carry trade strategies, which are rebalanced once every quarter or more frequently with new available risk-free rates, inflation risk is less important in practice.

[31] To illustrate this, we examine the current price (denominated in currency H) of bond H, which delivers one unit of currency H at $t + dt$. The pricing can either be done directly in currency H or in any other currency F with the help of exchange rates. The absence of arbitrage implies the law of one price, and thus

$$E_t\left[\frac{M_{t+dt}^H}{M_t^H}\right] = \frac{1}{S_t}E_t\left[\frac{M_{t+dt}^F}{M_t^F}S_{t+dt}\right] \Rightarrow S_t = \frac{M_t^H}{M_t^F}.$$

[32] Indeed, in a currency long bet, a promised payoff of one unit of foreign currency at $t+dt$ yields S_{t+dt}^{-1} unit of home currency also at $t+dt$. The associated consumption-based Euler equation for this bet, under the perspective of country H's investors, produces identical premia above; $-Cov_t\left[\frac{M_{t+dt}^H}{M_t^H}, dS_{t+dt}^{-1}\right] = E_t\left[XR_{t+dt}^{-H,+F}\right]$. See also footnote 34.

from (1.5) in the model

$$d\log S_t = dm^H - dm^F = \#dt + \gamma\alpha\epsilon\omega_N\left(\sigma_N^H dZ_N^H - \sigma_N^F dZ_N^F\right). \qquad (1.10)$$

On one hand, as a result of proposition 1 above, an adverse foreign nontraded shock $dZ_N^F < 0$ makes F's nontraded good scarce and suppresses the real exchange rate S (i.e., foreign currency appreciates), and therefore $m^F - m^H$ surges. On the other hand, $dZ_N^F < 0$ also forces F to consume more and H to consume less traded goods, and m^H surges. That is, the long bet on foreign currency pays off well when home investors highly value consumption. Therefore this carry trade strategy is a good hedge against foreign nontraded risk, and it commands high price and low expected return $E_t\left[XR^{-H,+F}\right]$ in equilibrium.

In contrast, an adverse home nontraded shock $dZ_N^H < 0$ directly boosts m^H. Moreover, it also leaves its trade partner F with less traded consumptions and thus also increases m^F to a lesser extent. Consequently, $m^F - m^H$ drops because the real exchange rate S increases (i.e., home currency appreciates). That is, the long bet on foreign currency pays off poorly when home investors highly value consumption. Therefore, this carry trade strategy is *not* a good hedge against home nontraded risk, and it carries a low price tag and offers a large expected return $E_t\left[XR^{-H,+F}\right]$ to compensate for the risk it cannot hedge in equilibrium.

The overall expected profit (or loss) of the carry trade is determined by whether home (or foreign) nontraded risk dominates, as seen quantitatively in the following result.

Proposition 2 *The expected carry trade excess return to US investors is*

$$E_t\left[XR_{t+dt}^{-H,+F}\right] = \alpha^2\gamma^2\epsilon\omega_N^2\left\{\left[\epsilon + (\gamma-\epsilon)\omega_T\frac{\Lambda^H}{\Lambda}\right](\sigma_N^H)^2 - (\gamma-\epsilon)\omega_T\frac{\Lambda^F}{\Lambda}(\sigma_N^F)^2\right\}, \qquad (1.11)$$

where $\alpha \equiv (\gamma\omega_T + \epsilon\omega_N)^{-1}$ is a weighted elasticity of consumption substitution (1.4). Consequently, the carry trade strategy offers the expected profit when either home nontraded risk dominates or trade effect is weak,

$$\left[\epsilon + (\gamma-\epsilon)\omega_T\frac{\Lambda^H}{\Lambda}\right](\sigma_N^H)^2 > (\gamma-\epsilon)\omega_T\frac{\Lambda^F}{\Lambda}(\sigma_N^F)^2.$$

The intuitions underlying this result are as follows. First, we recall that the carry trade is a good (bad) hedge against the foreign (home) nontraded output growth risk. When home nontraded risk dominates, $(\sigma_N^H)^2 \gg (\sigma_N^F)^2$, this strategy is risky and necessarily offers high expected returns $E_t\left[XR_{t+dt}^{-H,+F}\right] > 0$, and vice versa. Second, when $(\gamma - \epsilon)\omega_T$ is positive but small, investors are not enthusiastic about substituting nontraded for traded consumption goods. This weakens the trade effect and makes home nontraded output risk even worse to home investors. Therefore, in this case, carry trades are also risky and tend to generate compensating profits in the expectation. A reflection on the behaviors of risk-free rates and carry trade returns reveals that the nontraded consumption risk is a factor behind the violation of uncovered interest rate parity, a prevailing puzzle observed in the international financial market.

Uncovered interest rate parity

The uncovered interest rate parity (UIP) puzzle (a.k.a. forward premium puzzle) is an empirical regularity in which appreciating currencies tend to be also associated with increasing interest rates (Hansen and Hodrick (1980), Fama (1984)). This pattern is puzzling because it appears that the appreciating currencies are more valuable, yet investors require higher premia (i.e., interest rates) to hold them. Carry trades, i.e., borrowing low-interest-rate currencies and lending high-interest-rate currencies, are a popular strategy to reap the profit from this regularity. In the current setting, a nontraded consumption risk offers a rationale behind this profit.

When the home country has volatile nontraded sector by nature (σ_N^H large), home risk-free bonds are very valuable as a safe asset, and home interest rates are low (r^H small). At the same time, carry trades returns tend to be high because these strategies are not a good hedge against this home nontraded volatility as asserted by proposition 2. In contrast, when the foreign nontraded sector is perceived to be of low-risk nature (σ_N^F small), foreign interest rates are high (r^F large), *and* the expected carry trade return to home investors also tends to be high.[33] All in all, the nontraded output risk, originated from either home or abroad, is a culprit behind the violation of the uncovered interest parity.

Examining a large set of countries, Lustig and Verdelhan (2007) document that the exchange rates (base currency being US dollar) of high interest rate currencies tend to positively correlate with the US's consumption growth. The study clearly identifies the interrelationship of the exchange rate risk and the consumption risk as the source of the currency bet's expected profits. Namely, the carry trades of selling US dollar and buying high interest rate currencies are risky to US investors because they pay poorly (i.e., foreign currencies depreciate) when investors value consumption the most (i.e., US consumption drops). Our investigation carries this line of rational reasoning a step further by explaining the positive correlation between home consumption growth and exchange rates, as observed for US by Lustig and Verdelhan (2007); it is the nontraded output risk that can not only perturb the two quantities but also push them in the same direction.

Whereas our analysis lends support for the widely-practiced carry trade strategy of shorting low-interest rate currencies and longing high-interest rate currencies, it also suggests the following novel currency bet, which is directly tied to the nontradability aspects of consumption risk. We examine empirically the merits of this macro-based strategy in section 1.7.3.
Hypothesis 2: *Borrowing currencies of countries with a volatile nontraded sector and lending currencies of countries with a stable nontraded sector generate positive expected returns.*

Linear factor analysis: Theory

Our finding that country-specific traded and nontraded shocks are priced very differently by the international market warrants a simple linear-factor pricing model in which the risk factors are country-specific traded and nontraded consumption growths.

$$f_T^H = \frac{dC_T^H}{C_T^H}; \qquad f_N^H = \frac{dC_N^H}{C_N^H}.$$

[33]See proposition 2. Intuitively, this is because the foreign nontraded risk against which carry trade strategies can hedge are perceived to be small.

The exploration also emphasizes the difference between global (aggregate) traded output risk and the country-specific traded consumption risk. For illustration, carry trade portfolios are used as test assets in the discussion below and in the estimation process in section 1.7.3. As the risk factors are independent of test assets a priori, the discussion carries over to any other financial assets.

We consider the same carry trade return strategy of borrowing home and lending foreign currency. Again, its excess return to investor H and to be realized at $t+dt$ is (1.9): $XR_{t+dt}^{-H,+F} = \frac{1}{dt}\left[\frac{M_{t+dt}^F}{M_{t+dt}^H}(1+r_t^F dt)\frac{M_t^H}{M_t^F} - (1+r^H dt)\right]$. The factor analysis starts with the standard *unconditional*[34] consumption-based Euler equation for this carry trade return

$$E\left[\frac{M_{t+dt}^H}{M_t^H} XR_{t+dt}^{-H,+F}\right] = 0 \Longrightarrow E\left[XR_{t+dt}^{-H,+F}\right] = -\frac{1}{dt}Cov\left[1+dm_{t+dt}^H - E[dm_{t+dt}^H], XR_{t+dt}^{-H,+F}\right].$$

Because home consumption is made of both traded and nontraded components, log-linearized SDF (1.5) immediately implicates that the carry trade is priced by the following linear two-factor model ($f_T^H = \frac{dC_T^H}{C_T^H}$, $f_N^H = \frac{dC_N^H}{C_N^H}$)

$$E\left[XR_{t+dt}^{-H,+F}\right] = -Cov\left[b_T f_{T,t+dt}^H + b_N f_{N,t+dt}^H, XR_{t+dt}^{-H,+F}\right] \tag{1.12}$$

$$\begin{bmatrix} b_T \\ b_N \end{bmatrix} = \begin{bmatrix} -\gamma\omega_T \\ -\gamma\omega_N \end{bmatrix}; \quad \begin{bmatrix} f_T^H \\ f_N^H \end{bmatrix} = \begin{bmatrix} d\delta_T - \alpha(\gamma-\epsilon)\omega_N\left(d\delta_N^H - \sum_{F=1}^K \frac{\Lambda^F}{\Lambda} d\delta_N^F\right) \\ d\delta_N^H \end{bmatrix}.$$

Several observations can be made here. First, this is a country-specific pricing model that prices the assets from the perspective of home investors. Accordingly, the risk factors $\{f_T^H, f_N^H\}$ are *home-specific* traded and nontraded consumption growths, because they are the only risks priced by home SDF m^H. By restricting the pricing to a country-specific perspective, we can conveniently pack other countries' nontraded outputs into a single home traded consumption factor to facilitate the accompanied empirical analysis.[35] Second, this is

[34] In the conditional Euler equation approach, $E_t\left[XR_{t+dt}^{-H,+F}\right] = -\frac{1}{dt}Cov_t\left[1+dm_{t+dt}^H - E_t[dm_{t+dt}^H], XR_{t+dt}^{-H,+F}\right] = -\frac{1}{dt}Cov_t\left[dm_{t+dt}^H, dm_{t+dt}^F - dm_{t+dt}^H\right]$, where the last equality confirms that the result here is indeed identical to the expected excess return computed by a more intuitive approach in the previous section.

[35] We can also construct an international factor model in which the global traded output growth is a

a factor pricing model in which the factor loadings (b's) and risk factors (f's) are structurally determined and explicitly obtained. In particular, the loadings unambiguously increase with the tastes and risk aversion of investors. The factor f_T^H reveals all equilibrium effects established in previous sections, just as aggregate traded, trade partners' and country's own nontraded risk (respectively in δ_T, δ_N^F, δ_N^H) are all compounded in the home traded consumption allocation.

To better discern, both empirically and theoretically, the risk factors from the loadings of carry trade strategies on these risk types, we proceed to the beta-pricing version of the linear factor model.

$$
\begin{aligned}
E\left[XR_{t+dt}^{-H,+F}\right] &= \lambda_T^H \beta_T^{H,F} + \lambda_N^H \beta_N^{H,F}, \\
\begin{bmatrix} \lambda_T^H \\ \lambda_N^H \end{bmatrix} &= \left[Cov(\vec{f}^H, \vec{f}^H)\right] \begin{bmatrix} -b_T \\ -b_N \end{bmatrix}, \\
\begin{bmatrix} \beta_T^{H,F} \\ \beta_N^{H,F} \end{bmatrix} &= \left[Cov(\vec{f}^H, \vec{f}^H)\right]^{-1} \begin{bmatrix} Cov(f_T^H, XR^{-H,+F}) \\ Cov(f_N^H, XR^{-H,+F}) \end{bmatrix},
\end{aligned}
\tag{1.13}
$$

where $\left[Cov(\vec{f}^H, \vec{f}^H)\right]$ denotes the 2×2 variance-covariance matrix of the factors $\{f_T^H, f_N^H\}$. As β are slope coefficients of returns linearly regressed on the risk factors, the magnitude of β quantifies the exposures of investment strategies to the two risk factors. In contrast, factor prices $\{\lambda_T^H, \lambda_N^H\}$ are the rewards (in the form of expected returns) to bear one notional unit of corresponding risk (i.e., as if $\beta = 1$), which are independent of assets.

How exactly is risk embedded in asset payoff priced by the home investors? The basic risk-return tradeoff picture is that any shock that moves asset payoff and home marginal utility (or SDF m^H) in opposite directions is perceived as risk (again, because these assets pay poorly when investors highly value the payoff), and the corresponding reward (factor price) is positive, and vice versa. We begin with the home nontraded consumption growth

stand-alone factor. However, this model inevitably needs to involve all other country-specific nontraded outputs, and it will result in a multiple-factor model that would complicate the empirical analysis, requiring non-traded output data of all countries worldwide.

risk. Substituting the analytical expressions above for factors f's and loadings b's yields the following testable results.

Proposition 3 *The factor price associated with nontraded consumption growth risk is unambiguously positive,*

$$\lambda_N^H = \alpha\gamma\omega_N \left[\epsilon + (\gamma - \epsilon)\omega_T \frac{\Lambda^H}{\Lambda}\right] (\sigma_N^H)^2 > 0 \quad \forall H. \tag{1.14}$$

That is, the uncertainties in domestic nontraded consumption growth always pose as a risk to home investors in all countries.

Because idiosyncratic nontraded outputs can only be consumed domestically, the price of nontraded consumption risk involves only the volatility σ_N^H. As smaller economies can better outsource this risk to their trade partners by flexibly adjusting their traded consumption, this risk is more severe for larger economies. We indeed see that the corresponding factor price λ_N^H is higher for larger size Λ^H. Section 1.7.3 obtains a positive and statistically significant estimate for the US nontraded consumption growth factor price, which thus lends empirical support for the current model. We now turn to the factor price associated with the country-specific traded consumption growth risk,

$$\begin{aligned} \lambda_T^H &= \gamma\omega_T(\sigma_T)^2 + \alpha^2\gamma(\gamma - \epsilon)^2\omega_T\omega_N^2 \sum_{F \neq H}^{K} \frac{(\Lambda^F)^2}{(\Lambda)^2}(\sigma_N^F)^2 \\ &- \alpha^2\gamma(\gamma - \epsilon)\omega_N^2 \left(1 - \frac{\Lambda^H}{\Lambda}\right) \left[\epsilon + (\gamma - \epsilon)\omega_T \frac{\Lambda^H}{\Lambda}\right] (\sigma_N^H)^2. \end{aligned} \tag{1.15}$$

In sharp contrast with λ_N^H, the home traded consumption growth uncertainty is not necessarily a risk to home investors, which is manifested in the ambiguous sign of the associated factor price λ_T^H. This ambiguity arises because a country's traded consumption is endogenous in equilibrium. A surge in home traded consumption can be a consequence of either (i) a surge in global (aggregate) traded output (direct effect), (ii) a surge in trade partners' nontraded outputs (substitution and trade effects), or (iii) a *drop* in home nontraded output (substitution effect). Stating the last result inversely, a surge in home nontraded output acts to lower home traded consumption and boost home marginal utility. Consequently, from the perspective of the *endogenous* home traded consumption, home traded output shocks are not perceived as a risk, whereas shocks of global traded output and trade partners' nontraded outputs are, which explains the signs of all terms in λ_T^H. The overall sign of this home traded consumption growth factors depends on the relative contribution of these terms, and may vary from country to country.

Diversification benefits

Our consumption-risk framework not only delivers closed-form returns to carry trade strategies but also sheds light, both qualitatively and quantitatively, on the diversification benefits of the currency investment. In our setting, the key feature is that nontraded output risk of all countries enters the pricing of the carry trade return between any two countries. Consequently, forming currency portfolios facilitates the diversification among these sources of risk.[36] Previous literature[37] has found that forming equally weighted portfolios of currencies can substantially increase the Sharpe ratio of the carry trade investment strategies, although the underlying mechanism is not explicitly analyzed beyond the law of large number and ad-hoc mean-variance intuition.

Indeed, nontraded output shocks carry different weights, depending on the magnitude of their volatilities and the size of the economies of their origins, in the carry trade returns (1.11). This feature immediately offers a structural recipe that balances the above weights to achieve an optimal currency portfolio with maximal diversification. Let η^H denote market prices of risk from country H's perspective,[38] which is a vector in the face of multiple shocks priced by the H's SDF, M^H. Let us consider a generic carry trade portfolio that borrows home currency and lends several foreign currencies with weights $\{y_t^{HF}\}_F$ and $\sum_F y_t^{HF} = 1$.[39]

[36]As long as the total number of countries K is finite, nontraded risk cannot be entirely diversified and expected returns on currency portfolios preserve spread; see footnote 40.

[37]The partial list includes Burnside et al. (2008), Burnside et al. (2011), Lustig and Verdelhan (2007), and Menkhoff et al. (2011).

[38]That is, $\frac{dM^H}{M^H} = -r^H dt - \eta^H \cdot dZ$ where notation $A \cdot B$ emphatically denotes the scalar product of vectors A and B.

[39]To simplify the notation, our convention is that this sum is over all K countries, including H. However,

The realized and expected excess returns of this portfolio are simply the weighted values of the pairwise carry trade realized excess returns,

$$PR_{t+dt} = \sum_F^K y_t^{HF} XR_{t+dt}^{-H,+F} = \frac{1}{dt}\sum_F^K y_t^{HF}\left[\eta_t^H \cdot \left(\eta_t^H - \eta_t^F\right) dt + \left(\eta_t^H - \eta_t^F\right)\cdot dZ_{t+dt}\right],$$
$$EPR_t \equiv E_t\left[PR_{t+dt}\right] = \sum_F^K y_t^{HF}\eta_t^H \cdot \left(\eta_t^H - \eta_t^F\right) = \eta_t^H \cdot \left(\eta_t^H - \sum_F^K y_t^{HF}\eta_t^F\right).$$

It is apparent that forming a portfolio is not about improving the expected excess returns; the return of a portfolio of high-return currency trades remains high and vice versa.[40] Risk-neutral investors, who care only about expected returns would stay only with the single currency that offers the highest expected carry trade profit. The diversification instead helps reduce the portfolio return fluctuation and thus is slated to generate a Sharpe ratio superior to any single-currency carry trade strategies. From the excess return follows the portfolio's Sharpe ratio (we conventionally set investment horizon $dt = 1$ for ease of exposition),

$$SR_t = \frac{E_t\left[PR_{t+dt}\right]}{\left(Var_t\left[PR_{t+dt}\right]\right)^{1/2}} = \frac{\eta_t^H \cdot \left(\eta_t^H - \sum_F y_t^{HF}\eta_t^F\right)}{\|\eta_t^H - \sum_F y_t^{HF}\eta_t^F\|} = \|\eta_t^H\| \cos\Theta,$$

where Θ is the angle between vectors η_t^H and $(\eta_t^H - \sum_F y_t^{HF}\eta_t^F)$ in the output innovation hyperspace. From the perspective of investor H, prices of risk η_t^H are fixed and the optimal portfolio (of highest Sharpe ratio) is characterized by weights $\{y_t^{HF}\}_F$ that deliver the highest value for $\cos\Theta$ (lowest value for Θ). That is, by forming a portfolio, we can align the price of risk vectors as much as possible. The intuition is simple. Independent noises optimally offset one another when they are of similar magnitude. Pairwise carry trade strategies do not offer this condition simply because nontraded output statistics are heterogeneous across countries and are priced differently by H. This can be seen most lucidly in the analytical

it is possible that investors take opposite positions in some pairwise carry trade strategies; i.e., y_t^{HF} can assume negative values.

[40] This statement holds, given the total number of countries K stays fixed and finite. When the number of countries K increases unbounded, however, all economies become atomistic $\frac{\Lambda^F}{\Lambda} \to 0$, and all pairwise expected carry trade returns converge because nontraded risk becomes less prominent in such a diluted world; see (1.11). This effect is related more to the dilution of economic scales than to the diversification of nontraded risk.

expressions of the prices of risk

$$\forall H: \qquad \eta^H \cdot dZ = \left[dZ_T \;\; dZ_N^H \;\; \{dZ_N^F\}_{F \neq H} \right] \cdot \left[\begin{array}{c} -\gamma\omega_T\sigma_T \\ -\gamma\omega_N \left[1 - \alpha(\gamma - \epsilon)\omega_T \left(1 - \frac{\Lambda^H}{\Lambda} \right) \right] \sigma_N^H \\ \left\{ -\gamma\omega_N\alpha(\gamma - \epsilon)\omega_T \frac{\Lambda^F}{\Lambda} \sigma_N^F \right\}_{F \neq H} \end{array} \right].$$

Accordingly, the optimal portfolio choices $\{y_t^{HF}\}_F$ place appropriate weights on $\{\eta^F\}_F$ to essentially undo these heterogeneities to maximally enhance the noise cancellation in the realized portfolio return. Simple geometric arguments immediately show that the minimum Θ is the angle between vector η_t^H and its projected image on the space generated by all other prices of risk vectors $\{\eta_t^F\}_{F \neq H}$.[41] Straightforward but tedious algebra then identifies analytically the optimal Θ, portfolio weights and the maximum Sharpe ratio.

1.6 Beyond benchmark model

The key intuition, developed alongside the basic setting of international finance in previous sections, is that the country-specific traded output risk should have a smaller impact on asset prices than the country-specific nontraded output risk because of the diversification in the traded good market. However, the basic model possesses several simplifications, including (i) homogeneous consumption taste for a single common traded good and (ii) complete financial markets worldwide. In this section, we relax these assumptions and verify and thus strengthen the above intuition to a more realistic and robust economic setting.

1.6.1 Arbitrary trade configuration

Generalized setup: In the current general setting, there are l varieties of traded goods and K types of nontraded goods, and each of the latter is consumed by one respective country. A particular type h of traded goods can be consumed only by some K_h countries, and similarly,

[41] One can show that the choice $\{y_t^{HF}\}$ that minimizes the angle between η_t^H and $(\eta_t^H - \sum_F y_t^{HF}\eta_t^F)$ also minimizes the angle between η_t^H and $-\sum_{F \neq H} y_t^{HF}\eta_t^F$.

a particular country H trades and consumes only some l^H varieties of traded goods. These features aim to capture the realistic and vastly different trade configurations among countries, as well as the vastly different popularity of different traded goods.[42] Moreover, countries can also have country-specific tastes for the traded goods ($\{\omega_h^H\}$) and nontraded good (ω_N^H) that they consume, subject to the conventional normalization $\omega_N^H + \sum_h^{l^H} \omega_h^H = 1$. We also assume that the financial market is complete because contingent claims on all outputs and countries' risk-free bonds are available investment instruments. Consequently, the world's static optimization problem can be used to study the equilibrium behaviors of consumption allocations and asset prices in this economy.

$$\max_{\{C_{h,T}^H\}} \sum_{H=1}^{K} \Lambda^H \frac{e^{-\rho t}}{1-\gamma} \left[\sum_h^{l^H} \omega_{h,T}^H (C_{h,T}^H)^{1-\epsilon} + \omega_N^H (\Delta_N^H)^{1-\epsilon} \right]^{\frac{1-\gamma}{1-\epsilon}} \quad \text{s.t.} \quad \sum_H^{K_h} C_{h,T}^H = \Delta_{h,T} \ \ \forall h = 1, \ldots, l.$$

Although a country may have different tastes for different goods that they consume, the substitutability between any two varieties, either traded or nontraded, is characterized by the same elasticity coefficient ϵ. It is apparent from the market market clearing conditions that only the aggregate outputs for traded good varieties directly enter the dynamic of the economy. However, the associated output shocks will have different impacts on different countries, depending on their country-specific trade configurations. The current complex setting calls for a quantitative analysis to shed light on the role of these shocks on consumption allocations and prices.

Equilibrium allocations: Combining log-linearization and iteration techniques yield the equilibrium log consumption c_h^H of traded good h by country H,

$$\begin{aligned} c_h^H &= \delta_{h,T} - (\gamma - \epsilon)\alpha^H \omega_N^H \delta_N^H + (\gamma - \epsilon) \sum_J^{K_h} \alpha^J \frac{\Lambda^J}{\Lambda_h} \left(\omega_N^J \delta_N^J + \sum_j^{l^J} \omega_{j,T}^J \delta_{j,T} \right) \\ &- (\gamma - \epsilon)\alpha^H \sum_i^{l^H} \omega_{i,T}^H \left[\delta_{i,T} + (\gamma - \epsilon) \sum_I^{K_i} \alpha^I \frac{\Lambda^I}{\Lambda_i} \left(\omega_N^I \delta_N^I + \sum_k^{k^I} \omega_{k,T}^I \delta_{k,T} \right) \right], \quad (1.16) \end{aligned}$$

[42] Examples include the oil consumed by all countries versus rare earth minerals, which are consumed only by the most advanced economies.

where in the current general setting,

$$\Lambda_i \equiv \sum_I^{K_i} \Lambda^I; \qquad \alpha^H \equiv \frac{1}{(1-\omega_N^H)\gamma + \omega_N^H \epsilon} > 0 \tag{1.17}$$

are the good-specific relative size of the aggregate economies (those that consume good i) and a country-specific measure of weighted elasticity of consumption substitution, respectively. It is plausible that in this entangled trade network, many outputs affect country H's consumption of good h. In leading orders of importance, these include h's global supply ($\delta_{h,T}$); H's nontraded output (δ_N^H); nontraded output (δ_N^J) and traded global supply ($\delta_{j,T}$) consumed by any other country $J \in K^h$ that also consumes h; global supply ($\delta_{i,T}$) of any other traded good $i \in l^H$ consumed by H; and finally, the nontraded output (δ_N^I) and global supply ($\delta_{k,T}$) of traded goods k consumed by any country $I \in K^i$ that also consumes i.

Similar to the simpler setting of section 1.3, a country's traded consumption allocation c_h^H increases with the global supply $\delta_{h,T}$, decreases with the host's nontraded output δ_N^H, and increases with nontraded output δ_N^J of all trade partners J in good h. As country H also consumes other traded variates $\{\delta_{i,T}\}_{i \in l^H}$, H's consumption c_h^H in good h tends to negatively correlate with shocks $dZ_{i,T}|_{i \neq h}$ through the substitution effect between any two traded goods. Furthermore, because the consumptions of all trade partners $J \in K_h$ in good h are tuned to the nontraded δ_N^J and traded global supplies $\{\delta_{j,T}\}_{j \in l^J}$ that they consume, these shocks are also positively compounded into c_h^H, again through trade (market clearing) and substitution effects.

Most interestingly, even in the current general trade network setting, the international transmission of output shocks follows a simple and intuitive quantitative pattern in the leading orders. That is, the transmission process involving trades in a good i with a mediating country I warrants a dampening coefficient,[43]

$$(\gamma - \epsilon)\alpha^I \frac{\Lambda^I}{\Lambda_i} = \frac{\frac{1}{\epsilon} - \frac{1}{\gamma}}{\frac{1-\omega_N^I}{\epsilon} + \frac{\omega_N^I}{\gamma}} \frac{\Lambda^I}{\Lambda_i}.$$

[43] As γ is (substantially) larger than ϵ, mild home bias conditions assure that $(\gamma - \epsilon)\alpha^I \frac{\Lambda^I}{\Lambda_i} < 1$.

Here, $\frac{\Lambda^I}{\Lambda_i}$ characterizes the relative power of mediating country I in setting the global price for traded good i (through FOC), and $\left(\frac{1}{\epsilon}-\frac{1}{\gamma}\right)/\left(\frac{1-\omega_N^I}{\epsilon}+\frac{\omega_N^I}{\gamma}\right)$ quantifies how readily shocks in one consumption sector affect the others in a country.[44] Next, we examine the stochastic discount factors (SDFs) to explore how investors price the risk associated with these output shocks in different countries.

Equilibrium pricing: As shocks affect consumption allocations, they also move equilibrium prices accordingly to clear the market. The country H's log SDF is

$$\begin{aligned} m^H &= -\rho t-\gamma\sum_h^{l^H}\omega_{h,T}^H\delta_{h,T}-\gamma\omega_N^H\left[\delta_N^H-\sum_h^{l^H}(\gamma-\epsilon)\alpha^H\omega_{h,T}^H\left(1-\frac{\Lambda^H}{\Lambda_h}\right)\delta_N^H\right] \\ &- \gamma\sum_h^{l^H}\omega_{h,T}^H\sum_{J\neq H}^{K_h}(\gamma-\epsilon)\alpha^J\frac{\Lambda^J}{\Lambda_h}\left(\omega_N^J\delta_N^J+\sum_j^{l^J}\omega_{j,T}^J\delta_{j,T}\right) \end{aligned} \quad (1.18)$$

Reassuringly, all shocks that affect country H's consumptions are also priced by this stochastic discount factor. In particular, all traded and nontraded consumption shocks of H and any of it trade partners are compounded in m^H. As in the simpler case of section 1.3, up to taste coefficients, the traded shocks are fully internationalized (in the aggregate output $\delta_{h,T}$) and spread uniformly to all countries $I \subset K_h$ that consume good h. As $\omega_{h,T}^H$ generally drops with the number l^H of varieties consumed by H,[45] the country-specific traded shock of a particular variety matters even less to its country of origin in the current setting of multiple traded goods. In contrast, nontraded shocks are internalized, but not fully. As the second term within the square brackets shows, country H can tunnel its own nontraded shock in δ_N^H through trades in all l^H channels in which H participates. The ability to mitigate this shock through a particular channel h clearly decreases with a country's relative size $\frac{\Lambda^H}{\Lambda_h}$ in the world trade market for good h. Under mild home bias condition, country-specific nontraded

[44] Section 1.3 asserts that the difference $\frac{1}{\epsilon}-\frac{1}{\gamma}$ characterizes how willing a country is to substitute traded and nontraded consumptions to smooth its aggregate consumption. When this difference is large and positive as in the data, countries are flexible to make this substitution. As a result, a shock from one consumption sector is readily transmitted to the other sector. In the current setting, each country has one nontraded and several traded sectors, but all have the same pairwise substitution elasticity of ϵ.

[45] This is a consequence of the normalization condition $\omega_N^H+\sum_h^{l^H}\omega_{h,T}^H=1$.

shocks still matter more to the country's pricing than do the traded counterparts. Finally, we also see that traded shocks (in $\delta_{j,T}$) affecting any trade partner J are also factored in m^H. When H does not consume these goods, $j \notin l^H$, their shocks to H are similar the purely nontraded shocks of partners J.

1.6.2 Incomplete market

In equilibrium, the complete financial markets equalize all countries' marginal utilities of the traded consumption and thus enable the optimal international risk sharing and consumption allocation. In reality, however, the financial markets of some countries are more developed than those of others, which should better facilitate these developed countries to manage their own as well as trade partners' output risk. Stylistically, because of either information asymmetry or lack of proper managerial enforcement, the equities associated with nontraded sectors of emerging economies are less marketable worldwide. It is interesting to explore the new qualitative implications of market incompleteness on international risk sharing and contrast them with those of the simplified complete market paradigm. To this end, we now analyze a stylized model in which nontraded output risk is the central factor behind the incompleteness in the financial markets.

Setup: We consider the world economy with perfect trades but an incomplete financial market. In the commodity sector, there are country-specific nontraded goods (one per country) and a single traded good (common to all countries). The traded good can be shipped globally without the friction, and thus only its aggregate output influences the pricing. Accordingly, we assume that the financial assets associated with the traded good sector are perfectly structured. That is, a stock S_T contingent on the aggregate output and a risk-free bond B_T paying one unit of traded good in the next period are available to investors worldwide. In contrast, the financial assets associated with nontraded sectors are incomplete. We assume that countries belong to either the "developed" or the "emerging" group. For any developed economy ($H \in \mathcal{D}$), the stock S_N^H contingent on the H's nontraded output and risk-free bond B_N^H paying one unit of H's nontraded good in the following period are also available to all investors. However, assets associated with nontraded sectors of emerging economies ($H \notin \mathcal{D}$) are not marketable and thus simply do not exist. In this framework, the world financial market is incomplete because there are more shocks than the available financial hedging instruments. To simplify the exposition, we assume a homogeneous size for all economies embedded in a two-period setting $\{t, t+1\}$, but maintain the heterogeneous consumption tastes $\{\omega_T^H, \omega_N^H\}_H$ across countries. Relaxing all of these assumptions is tedious but straightforward.

The most convenient choice for the numeraire in this setting is the traded good, which we adopt hereafter. Thus, in every period, all prices are in (contemporaneous term of) the traded good. Because the market is incomplete, we consider the optimization problem for each country.[46] Let $x_T^{HS}, x_T^{HB}, x_N^{HFS}, x_N^{HFB}$ denote the holdings of H's investor, respectively, in world stock S_T, world bond B_T, F's stock S_N^F, and F's bond B_N^F.

$$\max_{C^H_{T,t}, x^{HS}_{T,t}, x^{HB}_{T,t}, \{x^{HFS}_{N,t}, x^{HFB}_{N,t}\}_{F\in\mathcal{D}}} U^H(C_t^H) + e^{-\rho} E_t\left[U^H(C_{t+1}^H)\right],$$

subject to the market clearing and budget constraints

$$\begin{aligned} &\textstyle\sum_H x_{T,t}^{HS} = \sum_H x_{T,t+1}^{HS} = 1; \quad \sum_H x_{T,t}^{HS} = \sum_H x_{T,t}^{HS} = 0 \\ &\textstyle\sum_H x_{N,t}^{HFS} = \sum_H x_{N,t+1}^{HFS} = 1; \quad \sum_H x_{N,t}^{HFS} = \sum_H x_{N,t+1}^{HFS} = 0 \qquad \forall F \in \mathcal{D} \end{aligned}$$

$$C_{T,t}^H + \Delta_{N,t}^H P_{N,t}^H \mathbb{1}_{H\in\mathcal{D}} + S_{T,t} x_{T,t}^{HS} + B_{T,t} x_{T,t}^{HB} + \sum_{F\in\mathcal{D}} S_{N,t}^F x_{N,t}^{HFS} + \sum_{F\in\mathcal{D}} B_{N,t}^F x_{N,t}^{HFB} \le W_t^H$$

$$C_{T,t+1}^H + \Delta_{N,t+1}^H P_{N,t+1}^H \mathbb{1}_{H\in\mathcal{D}} \le x_{T,t}^{HS} \Delta_{T,t+1} + x_{T,t}^{HB} + \sum_{F\in\mathcal{D}} x_{N,t}^{HFS} \Delta_{N,t+1}^F + \sum_{F\in\mathcal{D}} x_{N,t}^{HFB} P_{N,t+1}^F$$

where $C^H = \{C_T^H, C_N^H\}$ denotes the standard CES consumption aggregator as in section 1.3, U^H denotes the power utility function of C^H, and W_t^H denotes investor H's initial wealth. Identity operator $\mathbb{1}_{F\in\mathcal{D}}$ equals one if F is an developed country and zero otherwise, which simply reflects the fact that investors can invest in financial assets paying nontraded goods

[46] With an incomplete market, the centralized optimization can also be formulated as in Pavlova and Rigobon (2008) using the convex duality technique (Cvitanic and Karatzas (1992)). However, this approach offers an exact and analytical solution only for the special case of log utility.

and convert these payoffs into units of traded good at the respective nontraded price P_N^F, $\forall F \in \mathcal{D}$ for their consumption purpose. In contrast, no assets paying nontraded goods of emerging markets exist, and consequently no investors, domestic or otherwise, ever need to convert these goods into the traded good and back. In other words, in the current incomplete market setting, nontraded shocks are identical to preference shocks. Furthermore, we note that by summing all countries, the above budget constraints and market clearing conditions automatically imply the resource constraints $\sum_H C_{T,t}^H = \Delta_{T,t}$, $\sum_H C_{T,t+1}^H = \Delta_{T,t+1}$ in both periods.

First order conditions corresponding to variations about optimal holding positions x_T^{HS}, x_T^{HB}, x_N^{HFS}, x_N^{HFB}, respectively, generate pricing equations for all available financial assets.

$$S_{T,t} = E_t\left[\frac{M_{T,t+1}^H}{M_{T,t}^H}\Delta_{T,t+1}\right]; \qquad B_{T,t} = E_t\left[\frac{M_{T,t+1}^H}{M_{T,t}^H}\right] \qquad \forall H,$$
$$S_{N,t}^F = E_t\left[\frac{M_{T,t+1}^H}{M_{T,t}^H}\Delta_{N,t+1}^F P_{N,t+1}^F\right]; \qquad B_{N,t}^F = E_t\left[\frac{M_{T,t+1}^H}{M_{T,t}^H}P_{N,t+1}^F\right] \qquad \forall F \in \mathcal{D},\ \forall H.$$

where $M_{T,t}^H = \frac{\partial U^H}{\partial C_{T,t}^H}$ is the *country-specific* marginal utility of the *traded* consumption[47]

In the complete market setting, the marginal utilities are necessarily equalized across countries $\frac{M_{T,t+1}^H}{M_{T,t}^H} = \frac{M_{T,t+1}^F}{M_{T,t}^F}$ $\forall\{H, F\}$, which together with market clearing conditions, then establishes directly the equilibrium consumption allocations. In the incomplete market, the marginal utilities are indirectly connected to one another only through the pricing of available assets. Accordingly, the solution approach here is very different. In sequence, we first conjecture a solution for the consumption allocations, solve for the asset prices, and verify that these prices support the conjectured consumptions in equilibrium. As before, we log-linearize the above first order conditions for all countries H and all developed countries

[47]We recall that the current numeraire is the traded good, and therefore $M_{T,t}^H = e^{-\rho t}\omega_T^H (C_{N,t}^H)^{-\epsilon}\left[\omega_T^H (C_{T,t}^H)^{1-\epsilon} + \omega_N^H (C_{N,t}^H)^{1-\epsilon}\right]^{\frac{-\gamma+\epsilon}{1-\epsilon}}$ is the country H's pricing kernel with respect to this numeraire.

$F \in \mathcal{D}$[18]

$$\log\left(\frac{S_{T,t}}{B_{T,t}}\right) = Cov_t\left[dm^H_{T,t+1}, \delta_{T,t+1}\right]; \qquad \log\left(\frac{S^F_{N,t}}{B^F_{N,t}}\right) = Cov_t\left[dm^H_{T,t+1}, \delta^F_{N,t+1}\right] \quad (1.19)$$
$$\log B_{T,t} = E_t\left[dm^H_{T,t+1}\right] + \frac{1}{2}Var_t\left[dm^H_{T,t+1}\right]; \qquad \log\left(\frac{B^F_{T,t}}{B_{T,t}}\right) = Cov_t\left[dm^H_{T,t+1}, P^F_{N,t+1}\right],$$

where dm^H denotes the log-linearized stochastic discount factor (recall from (1.4) that $\alpha^H \equiv \frac{1}{\gamma\omega^H_T + \epsilon\omega^H_N}$),

$$dm^H_{T,t+1} \equiv m^H_{T,t+1} - m^H_{T,t} = \log\left(\frac{M^H_{T,t+1}}{M^H_{T,t}}\right) = -(\gamma - \epsilon)\omega^H_N d\delta^H_{N,t+1} - \frac{1}{\alpha^H}dc^H_{T,t+1}. \quad (1.20)$$

Equilibrium: Consistent with the log-linearization approximation scheme, we look for the equilibrium consumption allocations in the following most general log-linear form,

$$dc^H_{T,t+1} \equiv \log\left(\frac{C^H_{T,t+1}}{C^H_{T,t}}\right) = g^H + a^H d\delta_{T,t+1} + \sum_F b^{HF} d\delta^F_{N,t+1} \qquad \forall H, \quad (1.21)$$

and g's, a's, b's are constant parameters to be determined, and $d\delta$'s denote the changes in log outputs, i.e., output growths ($dt = 1$)

$$d\delta_{T,t+1} \equiv \delta_{T,t+1} - \delta_{T,t} = \mu_T dt + \sigma_T dZ_T; \qquad d\delta^H_{N,t+1} \equiv \delta^H_{N,t+1} - \delta^H_{N,t} = \mu^H_N dt + \sigma^H_N dZ^H_N$$

This choice renders a log-linear SDF dm^H in the approximation and greatly simplifies the pricing of financial assets in the incomplete market settings (Weil (1994)). Indeed, substituting the above conjectured consumptions and SDFs into the pricing equations and the market clearing conditions readily yields the following consumption allocations (derived in appendix 1.9.4),[19] where we recall that $\alpha^I \equiv \frac{1}{\gamma\omega^I_T + \epsilon\omega^I_N} > 0$ denotes the country-specific

[18]Although the log-linearization technique remains useful to obtain an approximate closed-form solution, it does not address the possible multiplicity and stability of the equilibrium.

[19]Specifically, the pricing equations $\log(S_{T,t}/B_{T,t})$'s determine coefficients $\{a^H\}_{\forall H}$, $\log(S^F_{N,t}/B^F_{N,t})$'s determine $\{b^{HF}\}_{\forall F \in \mathcal{D}, \forall H}$, $\log B_{T,t}$'s determine $\{b^{HF}\}_{\forall F \notin \mathcal{D}, \forall H}$, and $\log(B^F_{T,t}/B_{T,t})$'s determine the nontraded prices of developed countries $\{P^F_{N,t+1}\}_{\forall F \in \mathcal{D}}$, see appendix 1.9.4.

weighted elasticity of consumption substitution.

- incomplete market: H is an emerging economy ($H \notin \mathcal{D}$)

$$c^H_{T,t} = g^H t + \frac{K\alpha^H}{\sum_I \alpha^I}\delta_{T,t} - \alpha^H \sum_{F\notin\mathcal{D}} \delta^F_{N,t} + \frac{(\gamma-\epsilon)\alpha^H}{\sum_I^K \alpha^I} \sum_{F\in\mathcal{D}} \alpha^F \omega^F_N \delta^F_{N,t}. \tag{1.22}$$

- incomplete market: H is a developed economy ($H \in \mathcal{D}$)

$$\begin{aligned} c^H_{T,t} &= g^H t + \frac{K\alpha^H}{\sum_I \alpha^I}\delta_{T,t} + \frac{\alpha^H \sum_{I\notin\mathcal{D}} \alpha^I}{\sum_{J\in\mathcal{D}} \alpha^J} \sum_{F\notin\mathcal{D}} \delta^F_{N,t} \\ &+ \frac{(\gamma-\epsilon)\alpha^H}{\sum_I^K \alpha^I} \sum_{F\in\mathcal{D}} \alpha^F \omega^F_N \delta^F_{N,t} - (\gamma-\epsilon)\alpha^H \omega^H_N \delta^H_{N,t}. \end{aligned} \tag{1.23}$$

where g^H's are country-specific parameters. These parameters help to enforce, and thus can be found from the market clearing conditions (see appendix 1.9.4), but because they are deterministic factors, they do not enter the analysis below. To verify these equilibrium consumptions, we substitute them back into the above pricing equations to compute all available asset prices $\{S_{T,t}, B_{T,t}\}$, $\{S^F_{N,t}, B^F_{N,t}\}_{F\in\mathcal{D}}$, which finance these consumptions by the construction of the solution. This configuration is in equilibrium,[50] because, for each available asset, the associated price is identical under all investors' perspectives in the construction. Compared with the counterpart complete market setting with a single traded good, in which the consumption allocations are[51]

- complete market: $c^H_{T,t} = g^H t + \frac{K\alpha^H}{\sum_I^K \alpha^I}\delta_{T,t} + \frac{(\gamma-\epsilon)\alpha^H}{\sum_I^K \alpha^I} \sum_F \alpha^F \omega^F_N \delta^F_{N,t} - (\gamma-\epsilon)\alpha^H \omega^H_N \delta^H_{N,t}$.

the incomplete market allocations are markedly different in several aspects.[52] First, the

[50] Although this is not necessarily the unique equilibrium.

[51] This is a straightforward generalization of (1.3) (in the basic model) to the setting where countries have heterogeneous consumption tastes (but countries' sizes are homogeneous). In the current case, the log-linearization of FOC implies $m_T = -\rho t + \omega^H_T - (\gamma-\epsilon)\omega^H_N \delta^H_N + \frac{1}{\alpha^T} c^H_T$. Combining this FOC with the (log-linearized) market clearing condition (1.27) for traded good yields this log consumption c^H_T in complete market. See further details in appendix 1.9.2.

[52] In light of the possible existence of other incomplete market settings and multiple equilibria, our discussion here pertains to the specific incomplete market setup and the associated equilibrium presented earlier in this section.

traded shock impacts stay the same in both market configurations. This is because even when the market is incomplete, the equity and bond on the traded output δ_T are available to all investors, who then are able to mitigate these shocks as optimally as possible by trading these financial assets. When combined with the force of cross-country diversification in the traded sector, this result implies that country-specific traded output risks remain relatively less material to countries' risk free rates, compared with the nontraded output risk.

Second, the nontraded output shocks (in δ_N^F) of a developed country $F \in \mathcal{D}$ affect the traded consumption c_T^H of *all* other countries H similarly, regardless of the market's completeness. Because investors can trade the financial assets contingent on these nontraded shocks, their associated risk can be shared effectively. In particular, all else being equal, a surge in developed country F's nontraded output prompts F to trim its traded consumption and boosts other countries' traded consumption by forces of trades and market clearings. Similar to the complete market settings, under a mild degree of home biases, a country's own nontraded shocks matter quantitatively more to a developed country's consumption allocation than do the nontraded shocks of their *developed* trade partners.

Third, the nontraded output shocks (in δ_N^F) of an emerging country $F \notin \mathcal{D}$ are uniformly compounded in the consumptions c_T^H of all developed countries $H \in \mathcal{D}$.[53] This feature is intuitive. In the absence of financial assets in emerging markets, these shocks cannot be properly hedged. The developed investors instead opt to simply pool their consumptions uniformly to cope with the associated risk. Risk sharing is still feasible, albeit imperfect, because it is evident from the equilibrium allocation that a surge in the nontraded output from an *emerging* economy boosts traded consumptions of all *developed* economies. The coefficient characterizing this relationship, $\frac{\sum_{I \notin \mathcal{D}} \alpha^I}{\sum_{J \in \mathcal{D}} \alpha^J}$, increases (decreases) with the number of emerging (developed) economies. That is, the significance of the unhedged risk on consumption allocations is larger when the financial market is less complete in this pooling equilibrium.

Fourth, the incomplete market has a strong and surprising impact on risk sharing between

[53]That is, $\frac{\partial c_T^H}{\partial \delta_N^F}$ is same for all $F \notin \mathcal{D}$, $H \in \mathcal{D}$.

two emerging economies. Possessing no financial assets directly tied to the nontraded output shocks of their own or those of their emerging trade partners, the emerging economies also pool their traded consumption in equilibrium to uniformly share nontraded risk. Emerging country H's traded consumption c_T^H decreases with not only its own nontraded good endowment δ_N^H but also with other emerging countries' nontraded output δ_N^F. The latter behavior, which is the inverse of a perfect financial market, signals that the risk sharing is most severely hampered between emerging trade partners. This is indeed the group of countries whose nontraded output risk is the least hedgeable because of the incompleteness of the market.

The incomplete market setting, as formulated in this section and pertaining to the pooling equilibrium, does not qualitatively change the risk sharing behaviors, and thus prices, among developed economies. Any sizable effects stemming from market incompleteness instead arise in the group of emerging countries whose financial markets are the least developed in the setting.

1.7 Empirical results

The principal assertion of this paper, motivated by theoretical considerations in preceding sections, is that nontraded output risk is a key factor determining asset prices and price differentials in international markets. This section investigates this assertion empirically and provides supportive evidence. We implement various tests on interest rates and carry trade returns. Our empirical analysis involves OECD countries[54] plus Eurozone (i.e., Economic and Monetary Union, available after 1998), which are more developed economies and economic and financial data series of which are reasonably expected to be more complete and of higher quality. Our main empirical tests exclude three possible outlier countries (Estonia, Iceland, and Turkey) for the reasons presented in the next section on stylized facts of nontraded

[54] In our notation, before the German reunification in 1990 (and including that year), the Federal Republic of Germany (FRG) is referred to as West Germany. From 1991 onward, the (reunified) Federal Republic of Germany is referred to as Germany.

output risk. All nominal macroeconomic output series are first transformed into real series and then detrended using Hodrick-Prescott (HP) filter.[55] All employed data series are cited in double quotes, and their original sources and other details are listed in the data appendix.

1.7.1 Stylized facts concerning nontraded output risk

We identify "services" as nontraded sectors in all countries, following the standard classification in the literature (see, e.g., Stockman and Tesar (1995)). Key components of services sectors include wholesale and retail trade, hotels and restaurants, financial intermediation, real estate, business activities and construction services.

To obtain some idea about the size of nontraded sectors in the economies worldwide, figure 1-1 plots the ratio of real services output over real GDP, averaged over the period 1971-2010, for all OECD countries plus Eurozone. Output data are from "Aggregate National Accounts: Gross domestic product," and services are computed as the sum of (i) wholesale and retail trade, repairs, hotels and restaurants, and transport; (ii) financial intermediation, real estate, renting and business activities; (iii) construction; and (iv) other service activities. Figure 1-1 shows that nontraded outputs constitute a substantial fraction of the total GDP in all OECD countries, ranging from 0.5 (Iceland) to 0.7 (US). Among others, this figure thus re-documents a known fact that services sectors carry a huge weight of the US economy.

To justify the identification of services as a nontraded sector, Table 1.1 lists the country-specific tradability and size of financial services, construction services, and other services for a representative set of 13 OECD countries (see data appendix for classification details). Tradabilities and sizes are averaged over the period 1971-2010. The country-tradability of services is (one half of) the ratio of total exports and imports over the total output of these services by the country (see (1.26)). The economic size of services is the ratio of total domestic output of these services over the country's GDP. Countries' export and import series are from OECD's "Trade in Services" data base. Countries' services output series[56] are from OECD's "Aggre-

[55] We use smoothing parameters $\lambda = 1600$ for quarterly time series, as in Hodrick and Prescott (1997), and $\lambda = 6.25$ for annual time series, as in Ravn and Harald (2002).

[56] Specifically, these series are B1GF (Construction), B1GJ_K (Financial intermediation, real estate, renting

gate National Accounts: Gross domestic product." The table shows that, whereas the tradabilities and sizes of the same services vary considerably across OECD economies, their tradabilities are indeed small (in the order of few percentage points, and never exceeding 20%). In particular, financial services are a substantial part of GDP in all countries (ranging from 14.7% for the Czech Republic to 27.7% for the US), yet their tradabilities are very low (ranging from .21% for Japan to 7.5% for Switzerland). Similarly, Table 1.2 lists the 15 most traded industries in the US, along with their two measures of tradability. The US-specific tradability of an industry is computed similarly to the above country-specific tradability (1.26).

Table 1.2: Top-15 (ISIC rev. 3) US traded industries, 1971-2010

	ISIC rev. 3 designation	Industries	US-specific tradability (%)	OECD tradability (%)
1	19	leather. leather products and footwear	379.10	173.16
2	30	office, accounting and computing machinery	188.51	217.59
3	18	wearing apparel, dressing and dying of fur	135.52	105.76
4	34	motor vehicles, trailers and semi-trailers	97.98	128.61
5	272+2732	non-ferrous metals	93.10	149.14
6	32	radio, television and communication equipment	88.05	105.83
7	31	electrical machinery and apparatus, n.e.c.	66.99	82.14
8	33	medical, precision and optical instruments	66.83	106.44
9	29	machinery and equipment, n.e.c.	65.42	83.14
10	353	aircraft and spacecraft	60.00	104.28
11	352+359	railroad equipment and transport equipment, n.e.c.	58.33	111.70
12	17	textiles	56.39	99.83
13	24ex2423	chemicals excluding pharmaceuticals	50.20	108.05
14	23	coke, refined petroleum products and nuclear fuel	44.02	101.03
15	271+2731	iron and steel	41.06	74.31

Notes: This table lists 15 most traded industries in the US, along with their US-specific and OECD tradabilities. The industries are classified by ISIC Revision 3. US-specific tradability is (one half of) the ratio of total export and import over total output by the US of the industry (see (1.26)). OECD tradability for a industry is defined similarly, but with export, import, and output replaced by total-OECD counterparts (see (1.25)). See section 1.7.1 and data appendix for further details.

In the determination of OECD tradability (see (1.25)), export, import and output are OECD-aggregate quantities. These industry-level macro series are from the "OECD Structural Analysis (STAN)" database. Table 1.2 shows that all of the top 15 traded industries

and business activities), and B1GL_P (Other service activities).

in the US belong to the manufacturing sector. In either measure, their tradabilities are substantially higher than those in the services sectors listed in Table 1.1, which justifies the classification of traded and nontraded goods adopted in the literature as well as in the current paper. The table also shows that country-specific tradabilities do not necessarily and quantitatively coincide with their OECD counterparts because countries are heterogeneous in their consumption and production to a certain extent. For the sake of robustness, our tests presented in the next section will employ both of these tradability measures.

To have a sense of the level of nontraded output risk across countries, figure 1-4 plots the volatility of per-capita nontraded output growth for each OECD country. The volatility is computed as the standard deviation of these nontraded output growth series over the entire period of 1971-2010. Per-capita quantities are computed using the World Bank's "Total Population" series. This figure shows that the level of fluctuation of nontraded output varies widely across OECD countries. In particular, Estonia is the second smallest economy among OECD member states (Iceland is the smallest economy),[57] yet its per-capita nontraded output growth is subtantially more volatile than any other country (approximately ten times more volatile than Germany, France and the US). We therefore exclude Estonia and Iceland from empirical tests. When countries' nontraded output volatilities are computed for each ten-year period, Turkey exhibits an extremely unstable volatility pattern over time. We thus also drop Turkey from the tests.

To have a sense of the level of trade "openness" of OECD countries, figure 1-5 plots the ratio of each country's total exports and imports over its GDP (see also (1.24)), averaged over the period 1971-2010. These ratios are from OECD's "Trade-to-GDP ratio" annual series. The figure shows that trade openness is markedly heterogeneous across OECD countries, ranging from 0.17 for Japan to 2.08 for Luxembourg. It is known that this ratio can be biased downward for larger economies, and hence a low value of the openness for a country does not necessarily imply high (tariff or non-tariff) obstacles to foreign trade. Rather, the low value of the openness can be a measure of either weak reliance of domestic producers

[57]Estonia's GDP is approximately 20 Bln USD for the year of 2010, or less than 0.05% of the aggregate GDP of OECD group. Iceland GDP is 12 Bln USD for the same year.

on foreign supplies and markets or of the country's geographic remoteness from potential trading partners. Any of these possible causes are consistent with our notion that the output growth risk of the more closed economies is internalized by home countries to a larger extent.

Finally, as a preliminary and graphical check of the allegedly key role of nontraded output risk on national asset prices, figure 1-2 plots the real risk free rates against the volatilities of per-capita nontraded and traded output growths, in the cross section of OECD countries. Real interest rates are deduced from the nominal "IMF Exchange Rates and Short-term Treasury Bill Rates" and the accompanying price index series, following Obstfeld and Rogoff (2001). Figure 1-2 shows an inverse relationship between risk-free rates and nontraded output volatilities, which in particular is vividly stronger than that between risk-free rates and traded output volatilities. This difference persists even when we drop potential outlier countries in figure 1-3. This simple pattern is consistent with the theoretical finding presented earlier that the asset returns *differentials* across countries are tied principally to the countries' nontraded output risk characteristics.[58]

1.7.2 Interest rates

In reality, no goods are either perfectly nontraded or perfectly traded. Even if some goods were, macro output series are inevitably subject to measurement errors. Furthermore, costs in trades also affect the structural relation between nontraded output risk and asset prices. In this section, we investigate the empirical relationship between nontraded output volatility and the level of real interest rate across OECD countries, taking into account these practical regularities. Specifically, we devise four tests based on the various classifications of nontradability, in order of increasing sophistication. These regression-based tests involve (i) the closedness of an economy, (ii) the brute-force cutoff dummy of nontradability at the industry level, (iii) the global nontradabilities at industry level, and (iv) country-specific nontradabilities at the industry level, respectively.

[58] By means of trades and diversification, in contrast, country-specific traded output risk is pooled together and therefore does not *distinctly* impact the risk-free rates around the world.

Tests using countries' trade closedness

The hypothesis to be examined here is that when an economy is exposed more to international trades, its nontraded risk can be better mitigated through trades and the substitution between traded and nontraded consumption. This assertion is a specific form of Hypothesis 1 (section 1.4) and Hypothesis 1A (section 1.4.2), and is motivated by the structural model with trade friction presented in previous sections. The basic regression test of this relationship reads

$$r_t^H = \alpha + \beta_\sigma (\sigma_t^H)^2 + \beta_C \mathcal{C}_t^H + \beta_{\sigma C} (\sigma_t^H)^2 \mathcal{C}_t^H + \beta_x X_t^H + \epsilon_t^H,$$

where σ^H denotes the per-capita GDP growth volatility and X's the various control variables. We adopt the common definition of a country's trade openness $\mathcal{O}^H$ as trade-to-GDP ratio (trade being the sum of export and import), from which also follows the closedness $\mathcal{C}^H$

$$\mathcal{O}^H = \frac{\mathrm{IM}^H + \mathrm{EX}^H}{\mathrm{GDP}^H}; \qquad \mathcal{C}^H = 1 - \frac{\mathrm{IM}^H + \mathrm{EX}^H}{\mathrm{GDP}^H}. \tag{1.24}$$

Table 1.3 reports the results associated with this regression. National output data are from "Aggregate National Accounts: Gross domestic product" and trade openness from "Trade-to-GDP ratio." We compute the volatility of per-capita GDP growth either over the entire period of 1971-2010 (in which case, the above time index t should be dropped), or over each of four non-overlapping 10-year periods, and the mean of interest rates (dependent variable) over exactly the same periods. Control variables include per-capita GDP mean growth, GDP size (or the ratio of countries' GDP over the aggregate GDP of OECD group), and inflation volatility.[59] The last control variable aims to address the fact that the model is real and thus does not capture the possible effects from inflation risk.

The key observation from table 1.3 is that the slope coefficients associated with the interaction term (variance $\times$ closedness) are always negative. These coefficients are statis-

[59] Inflation is computed as the year-to-year percentage change of the consumer price index, and the latter is sourced from IMF's CPI series. Furthermore, inflation volatility is computed as standard deviation of the inflation growth.

Table 1.3: Trade-closedness regression, 1971-2010

	Panel A: Four 10-year Periods				Panel B: Entire Period			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
growth variance	-36.245	-39.245	-37.266	-36.449	6.769	2.9643	19.305	18.233
	(22.249)	(23.685)	(23.92)	(24.666)	(27.713)	(29.912)	(28.615)	(30.002)
closedness	-.01246	-.0118	-.00818	-.00842	.01565	.02349	.0436**	.03991**
	(.00892)	(.00889)	(.01046)	(.01039)	(.01458)	(.01706)	(.01809)	(.01918)
variance × closedness	-44.26	-43.553	-51.34*	-52.343*	-93.324*	-113.38**	-167.21***	-159.61***
	(30.133)	(29.686)	(30.984)	(32.282)	(45.975)	(52.996)	(53.36)	(56.442)
growth mean		.1529	.13368	.1205		.39029	.34268	.26642
		(.28422)	(.29327)	(.30318)		(.277)	(.28719)	(.3311)
gdp size			-.03433	-.03288			-.08515***	-.08045***
			(.03625)	(.03593)			(.02984)	(.0277)
inflation volatility				-.00054				-.00124
				(.0007)				(.00082)
constant	.02892***	.02537***	.02596***	.02682***	.01781**	.0074	.00492	.00944
	(.00426)	(.00721)	(.00742)	(.0079)	(.00731)	(.01071)	(.01071)	(.01271)
N	98	98	98	98	33	33	33	33
adj. R^2	0.103	0.097	0.093	0.085	0.082	0.120	0.228	0.228

Notes: OLS regressions with robust standard errors in parentheses: $r_t^H = \alpha + \beta_\sigma^H(\sigma_t^H)^2 + \beta_C^H C_t^H + \beta_{\sigma C}^H(\sigma_t^H)^2 C_t^H + \beta_x X_t + \epsilon_t^H$ to examine the effects of output volatility σ^H and trade closedness C^H on interest rate r^H. Panel A reports results when the variance of GDP growth is computed for each of 10-year non-overlapping periods, from 1971 to 2010. Panel B reports results when the variance of GDP growth is computed for the entire period from 1971 to 2010. The sample consists of annual data series for OECD countries 1971-2010, excluding Estonia, Iceland and Turkey. Current members of European Monetary Union are dropped from the sample at the moment they joined the Union, and replaced by a single observation for Eurozone. Growth variance is the annualized variance of growth rate of per-capita real GDP over corresponding period. closedness is one subtracted by the ratio of country's total trade over country's GDP (see (1.24)). Growth mean is the annualized mean of growth rate of per-capita real GDP over corresponding period. GDP size is the ratio of country's real GDP over total real GDP of OECD member states. Inflation volatility is the standard deviation of country's consumption price index over the corresponding period. See data appendix for further details.

tically significant when we take into account the GDP growth (which contributes through the intertemporal smoothing desires of investors), economy size, and inflation risk effects, for either the entire period (i.e., in the cross sectional data) or for four 10-year periods (i.e., in the panel data). This negative sign is consistent with the model's central economic rationale that when a country is less open to trade and all else is equal, the country's output shock tends to be more internalized, and to have stronger impacts on lowering country's real interest rate through the precautionary savings mechanism.

Tests using multiple industry outputs and their nontradability dummies

Another form of Hypothesis 1 and Hypothesis 1A (sections 1.4.1, 1.4.2, respectively) to be examined in this section is as follows. Controlling for anything else, a country's output growth risk of nontraded industries tends to have a stronger impact on domestic interest rate than its output growth risk of traded industries. Intuitively, this is because country-specific traded risk can be diversified in the global pool of traded goods before it affects prices in any country. The basic regression testing this relationship employs national output data at the industry level. We use binary dummies to classify the nontradability of the industries.

$$r_{i,t}^{H} = \alpha + \beta_{\sigma}(\sigma_{i,t}^{H})^2 + \beta_d d_{i,t} + \beta_{\sigma d}(\sigma_{i,t}^{H})^2 d_{i,t} + \beta_x X_{i,t}^{H} + \epsilon_{i,t}^{H},$$

where $r_{i,t}^{H} = r_t^{H}$ is country H's interest rate and thus independent of industry type i, $d_{i,t}$ is nontradability dummy ($d_{i,t} = 1$ for nontraded industries and 0 otherwise, as we explain below). Table 1.4 reports the results associated with this regression. Countries' real annual industry-level outputs are constructed from the "OECD Structural Analysis (STAN)" database. An industry i is classified as nontraded ($d_{i,t} = 1$) if it belongs to one of the following ISIC classes[60] (see further details in data appendix): 40-41 (electricity gas and water supply); 45 (construction); 50-55 (wholesale and retail trade, restaurant and hotels); 60-64 (transport storage and communications); 65-74 (finance insurance real estate and business services); 75-99 (community social and personal services). Other industries are taken as

[60] ISIC stands for International Standard Industrial Classification of All Economic Activities

Table 1.4: Multi-industry nontradability-dummy regression, 1971-2010

	Panel A: Pooled OLS Regression				Panel B: Panel Regression			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
growth variance	-.02793***	-.02659***	-.03036***	-.02558***	-.03765	-.03676	-.04416*	-.03697*
	(.00767)	(.00674)	(.00831)	(.0097)	(.02311)	(.02342)	(.02303)	(.02179)
nontradability dummy	.00347	.00339	.00291	.00334	.00471**	.0047**	.0044**	.00494**
	(.00237)	(.00236)	(.00233)	(.00228)	(.00222)	(.00222)	(.00218)	(.00206)
variance × dummy	-.36047**	-.44973***	-.47498***	-.52871***	-1.5183	-1.5324	-1.5814*	-1.6366*
	(.16142)	(.16597)	(.17103)	(.17824)	(.94835)	(.95064)	(.93323)	(.88259)
growth mean		.01232**	.00932*	.00891		.00172	-.00299	-.0045
		(.00564)	(.00562)	(.00558)		(.00716)	(.00707)	(.00669)
gdp size			-.04086***	-.03773***			-.04019***	-.0346***
			(.00479)	(.00459)			(.00675)	(.0064)
inflation volatility				-.00096***				-.00189***
				(7.1e-05)				(.00018)
constant	.02555***	.02535***	.02716***	.02824***	.02621***	.02618***	.02785***	.03006***
	(.0006)	(.00061)	(.00069)	(.00072)	(.00051)	(.00052)	(.00059)	(.00059)
N	2026	2026	2026	2026	2026	2026	2026	2026
adj. R^2	0.000	0.001	0.016	0.031	0.006	0.005	0.041	0.143

Notes: OLS regressions $r^H_{i,t} = \alpha + \beta_\sigma(\sigma^H_{i,t})^2 + \beta_d d_{i,t} + \beta_{\sigma d}(\sigma^H_{i,t})^2 d_{i,t} + \beta_x X^H_{i,t} + \epsilon^H_{i,t}$ to examine the effects of industry-level output volatility σ^H_i and its dummy nontradability d_i on interest rate r^H. Panel A reports results with robust standard errors in parentheses. Panel B reports results with between-effect standard errors in parentheses. Dependent variable is the annualized real interest rate, proxied by the short-term Treasury bill rate, averaged over the corresponding period. The sample consists of annual data series for OECD countries 1971-2010, excluding Estonia, Iceland and Turkey. Current members of European Monetary Union are dropped from the sample at the moment they joined the Union, and replaced by a single observation for Eurozone. Growth variance is the annualized variance of growth rate of per-capita country-specific industries' real output over each of four 10-year periods, from 1971 to 2010. Nontradability dummies are at industry level; they assume value 1 for industries classified as nontraded sectors (Electricity gas and water supply, Construction, Wholesale and retail trade, restaurant and hotels, Transport storage and communications, Finance insurance real estate and business services, Community social and personal services), and 0 otherwise. Growth mean is the annualized mean of growth rate of per-capita country-specific industries' real output over the corresponding period. GDP size is the ratio of country's real GDP over total real GDP of OECD member states. Inflation volatility is the standard deviation of country's consumption price index over the corresponding period. See data appendix for further details.

traded ($d_{i,t} = 0$). We divide the entire time period 1971-2010 into four 10-year periods, and the volatility of per-capita output growth for each industry is computed as the respective standard deviation over each period. As before, the control variables include per-capita GDP mean growth, GDP size, and inflation volatility.

The key observation from table 1.4 is that the slope coefficients associated with the interaction term (variance × dummy) are always negative. When we take into account the GDP growth, economy size, and inflation risk effects, these coefficients are statistically significant either for robust or between-effect standard errors.[61] The negative sign precisely fits the basic economic intuition that the output growth risk is more serious to the economy than that of the traded output. Consequently, the output risk enhances the value of risk-free bonds, and depresses risk-free rate more aggressively when the risk comes from a nontraded industry.

Tests using multiple industry outputs and their global nontradabilities

Some industries are not clear-cut traded or nontraded as depicted by a binary dummy of the above regression. In this section, we use continuous-valued global nontradability at industry level to account for this fine distinction. The hypothesis to be examined here is the same as above, namely all else being equal, output risk of nontraded industries matter more to country's interest rate than that of traded industries. The basic regression testing this relationship reads

$$r_{i,t}^H = \alpha + \beta_\sigma(\sigma_{i,t}^H)^2 + \beta_\tau\tau_{i,t} + \beta_{\sigma\tau}(\sigma_{i,t}^H)^2\tau_{i,t} + \beta_x X_{i,t}^H + \epsilon_{i,t}^H,$$

where τ_i is a global measure of nontradability of industry i. We adopt the standard definition of tradability as the ratio OECD aggregate trade over OECD aggregate output of the industry

[61] Due to limited data, the choice of between-effect model is appropriate.

i, and nontradability is the complement to tradability

$$\tau_i = 1 - \frac{\sum_{\text{OECD countries}} [i\text{'s import} + i\text{'s export}]}{2 \times \sum_{\text{OECD countries}} i\text{'s output}}. \tag{1.25}$$

Table 1.5 reports the results associated with this regression. Data sources are identical to those employed in the above regression. We use country-specific output series to compute country-specific industry i's growth volatility over each of four 10-year periods. We aggregate these series to compute the global tradability and nontradability for each of good i.

The key observation from table 1.5 is that the slope coefficients associated with the interaction term (variance $\times$ nontradability) are always negative. When we take into account the GDP growth, economy size, and inflation risk effects, these coefficients are statistically significant either for robust or between-effect standard errors. The negative sign precisely fits the basic economic intuition that as countries mostly internalize their own nontraded shocks, the fluctuations in nontraded industries are more serious risk to the economy than those of the traded ones. Furthermore, output volatility act to lower risk-free rate. Consequently, risk-free rate is more sensitive (and negatively related) to output risk of industries of higher nontradabilities.

Tests using multiple industry outputs and their country-specific nontradabilities

In some situation, global measure of tradability does not exactly reflect the tradability of an industry at country level. This happens, for e.g., when the trade levels are highly heterogeneous across countries in certain industries. To account for this fine distinction, in this section, we use continuous-valued country-specific nontradability at industry level. The hypothesis to be examined here is the same as above, namely all else being equal, output risk of nontraded industries matter more to country's interest rate than that of traded industries. The basic regression testing this relationship reads

$$r^H_{i,t} = \alpha + \beta_\sigma(\sigma^H_{i,t})^2 + \beta_\tau \tau^H_{i,t} + \beta_{\sigma\tau}(\sigma^H_{i,t})^2 \tau^H_{i,t} + \beta_x X^H_{i,t} + \epsilon^H_{i,t},$$

Table 1.5: Multi-industry global nontradability regression, 1971-2010

	Panel A: Pooled OLS Regression				Panel B: Panel Regression			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
growth variance	-.05941**	-.05469**	-.06901**	-.07466***	-.10134	-.10093	-.12434*	-.14698**
	(.02778)	(.02575)	(.02752)	(.02809)	(.06564)	(.06573)	(.0646)	(.06122)
global nontradability	.00133**	.00128**	.00127**	.001*	.0014**	.0014**	.00137**	.00085
	(.00052)	(.00053)	(.00052)	(.00051)	(.00068)	(.00068)	(.00066)	(.00063)
variance × nontradability	-.00608*	-.0055*	-.00706**	-.00829**	-.01099	-.01101	-.01334	-.01707**
	(.00349)	(.00325)	(.00346)	(.00353)	(.00904)	(.00905)	(.00889)	(.00842)
growth mean		.01139**	.00821	.00787		.00109	-.00371	-.00498
		(.00569)	(.00564)	(.00555)		(.00715)	(.00706)	(.00668)
gdp size			-.04127***	-.03829***			-.04075***	-.03539***
			(.00473)	(.00455)			(.00675)	(.00641)
inflation volatility				-.00094***				-.00188***
				(6.9e-05)				(.00018)
constant	.02555***	.02535***	.02719***	.02833***	.02634***	.02632***	.02803***	.03039***
	(.00057)	(.00058)	(.00066)	(.00069)	(.00052)	(.00054)	(.0006)	(.00061)
N	2026	2026	2026	2026	2026	2026	2026	2026
adj. R^2	0.002	0.002	0.018	0.032	0.006	0.005	0.042	0.141

Notes: OLS regressions $r^H_{i,t} = \alpha + \beta_\sigma(\sigma^H_{i,t})^2 + \beta_\tau\tau_{i,t} + \beta_{\sigma\tau}(\sigma^H_{i,t})^2\tau_{i,t} + \beta_x X^H_{i,t} + \epsilon^H_{i,t}$ to examine the effects of industry-level output volatility σ^H_i and its global nontradability τ_i on interest rate r^H. Panel A reports results with robust standard errors in parentheses. Panel B reports results with between-effect standard errors in parentheses. Dependent variable is the annualized real interest rate, proxied by the short-term Treasury bill rate, averaged over the corresponding period. The sample consists of annual data series for OECD countries 1971-2010, excluding Estonia, Iceland and Turkey. Current members of European Monetary Union are dropped from the sample at the moment they joined the Union, and replaced by a single observation for Eurozone. Growth variance is the annualized variance of growth rate of per-capita country-specific industries' real output over each of four 10-year periods, from 1971 to 2010. Here nontradability is a global measure and at industry level: it is one subtracted by the ratio of global total trade (i.e., import plus export) in an industry over the global total output in that industry (see (1.25)). Growth mean is the annualized mean of growth rate of per-capita country-specific industries' real output over the corresponding period. GDP size is the ratio of country's real GDP over total real GDP of OECD member states. Inflation volatility is the standard deviation of country's consumption price index over the corresponding period. See data appendix for further details.

where τ_i^H is a country-specific measure of nontradability of industry i. We adopt the standard definition of tradability as the ratio of national trade over national output of the industry i, and nontradability is the complement to tradability

$$\tau_i^H = 1 - \frac{[i\text{'s import} + i\text{'s export}] \text{ by country } H}{[2 \times i\text{'s output}] \text{ by country } H}. \tag{1.26}$$

Table 1.6 reports the results associated with this regression. Data sources are identical to those employed in the above regression. We use country-specific output series to compute both country-specific industry i's growth volatility over each of four 10-year periods and i's country-specific tradability and nontradability.

The key observation from table 1.6 is that the slope coefficients associated with the interaction term (variance × nontradability) are always negative. When we take into account the GDP growth, economy size, and inflation risk effects, these coefficients are statistically significant either for robust or between-effect standard errors. The negative sign precisely fits the basic economic intuition that as countries mostly internalize their own nontraded shocks, the fluctuations in nontraded industries are more serious risk to the economy than those of the traded ones. Furthermore, output volatility act to lower risk-free rate. Consequently, risk-free rate is more sensitive (and negatively related) to output risk of industries of higher nontradabilities.

1.7.3 Carry trade returns

The evidences above shows that nontraded risk is a key factor behind national asset returns. This is very intuitive because national asset prices are country-specific measures and nontraded shocks are mostly internalized by countries. Taking a step further, as every *international* investment strategy is exposed to nontraded risk of all countries involved, the associated compensating profits should reflect the interplay of these risk factors. In this section, we investigate the empirical relationship between carry trade expected returns and nontraded output volatilities of the countries involved. Specifically, we devise two tests which

Table 1.6: Multi-industry country-specific nontradability regression, 1971-2010

	Panel A: Pooled OLS Regression				Panel B: Panel Regression			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
growth variance	-.03498	-.03985	-.05086*	-.06099**	-.05515	-.05372	-.07389	-.11651*
	(.0269)	(.02707)	(.02697)	(.028)	(.07436)	(.07465)	(.07329)	(.06955)
nontradability	6.3e-05***	4.5e-05*	5.6e-05**	4.2e-05*	9.4e-05*	9.7e-05*	.0001**	6.1e-05
	(2.0e-05)	(2.5e-05)	(2.4e-05)	(2.4e-05)	(5.0e-05)	(5.2e-05)	(5.1e-05)	(4.8e-05)
variance× nontradability	-6.9e-05**	-5.9e-05*	-7.7e-05**	-8.2e-05***	-.00013	-.00013	-.00016**	-.00016**
	(2.8e-05)	(3.0e-05)	(3.0e-05)	(3.1e-05)	(8.1e-05)	(8.1e-05)	(8.0e-05)	(7.5e-05)
growth mean		.00969	.00597	.0063		-.00172	-.00678	-.00659
		(.00626)	(.00623)	(.00614)		(.00739)	(.0073)	(.00691)
gdp size			-.04162***	-.03852***			-.04109***	-.03557***
			(.0048)	(.0046)			(.00675)	(.00642)
inflation volatility				-.00095***				-.00187***
				(7.0e-05)				(.00018)
constant	.02583***	.02564***	.02751***	.02859***	.02664***	.02667***	.0284***	.03061***
	(.00059)	(.00061)	(.0007)	(.00073)	(.00051)	(.00053)	(.00059)	(.0006)
N	2026	2026	2026	2026	2026	2026	2026	2026
adj. R^2	0.001	0.001	0.017	0.031	0.005	0.004	0.042	0.140

Notes: OLS regressions $r_{i,t}^H = \alpha + \beta_\sigma(\sigma_{i,t}^H)^2 + \beta_\tau \tau_{i,t}^H + \beta_{\sigma\tau}(\sigma_{i,t}^H)^2 \tau_{i,t}^H + \beta_x X_{i,t}^H + \epsilon_{i,t}^H$ to examine the effects of industry-level output volatility σ_i^H and its country-specific nontradability τ_i^H on interest rate r^H. Panel A reports results with robust standard errors in parentheses. Panel B reports results with between-effect standard errors in parentheses. Dependent variable is the annualized real interest rate, proxied by the short-term Treasury bill rate, averaged over the corresponding period. The sample consists of annual data series for OECD countries 1971-2010, excluding Estonia, Iceland and Turkey. Current members of European Monetary Union are dropped from the sample at the moment they joined the Union, and replaced by a single observation for Eurozone. Growth variance is the annualized variance of growth rate of per-capita country-specific industries' real output over each of four 10-year periods, from 1971 to 2010. Nontradability is a country-specific measure and at industry level; it is one subtracted by the ratio of country's trade (i.e., import plus export) in an industry over the country's output in that industry (see (1.26)). Growth mean is the annualized mean of growth rate of per-capita country-specific industries' real output over the corresponding period. GDP size is the ratio of country's real GDP over total real GDP of OECD member states. Inflation volatility is the standard deviation of country's consumption price index over the corresponding period. See data appendix for further details.

involve (i) forming currency portfolios based on countries's nontraded volatility and size, and (ii) constructing nontraded and traded consumption risk factors to price carry trades. The valuation of all carry trades is exclusively from the perspective of US investors, for whom the ultimate profits are in term of US dollars.

Forming portfolios based on the nontraded output growth volatilities and economy sizes

The theoretical analysis of section 1.5 clearly indicates that[62] controlling for all else, carry trades with partner countries of smaller sizes and less volatile nontraded outputs yield higher expected returns to US investors.[63] To directly verify this structural mechanism, stated in Hypothesis 2 (section 1.5), we construct portfolios of currencies based mainly on the volatilities of nontraded output as suggested by the theory. As argued by Lustig and Verdelhan (2007), forming portfolios helps filter out the noises in individual currency returns, and delivers large and stable return spreads between portfolios by means of frequent rebalancing. Burnside et al. (2008) document and the current paper's section 1.5 theoretically shows sizable benefits of diversification in portfolio construction.

We consider carry trade returns from US investors' perspectives. For each country, we identify the nontraded consumption as the expenditure on services (a component of the expenditure on total private consumption in the expenditure approach to GDP). These consumption expenditure series are available only at quarterly (or lower) frequencies, and sourced from OECD's "Quarterly National Accounts" database.[64] At the beginning of each quarter t, countries are sorted into four (quartile) portfolios based on the value of country-specific product of per-capita[65] nontraded consumption growth variance and relative GDP

[62] Expected returns of the carry trades to US investors have been computed in section 1.5, $E_t\left[XR_{t+dt}^{-H,+F}\right] = \alpha^2\gamma^2 c\omega_N^2\left\{\left[c + (\gamma - c)\omega_T\frac{\Lambda^H}{\Lambda}\right](\sigma_N^H)^2 - (\gamma - c)\omega_T\frac{\Lambda^F}{\Lambda}(\sigma_N^F)^2\right\}$.

[63] All carry trades involve shorting US dollars and longing foreign currencies.

[64] To obtain a more extensive historical data, however, US quarterly consumption expenditure series are sourced from US Bureau of Economic Analysis. See data appendix for further details.

[65] Since the population time series are not available at quarterly frequency, they are constructed from the annual population by intrapolation, assuming constant population growth within each year. Annual population data are from World Bank's "Total Population series".

size. For each country, the product is computed over the previous eight-quarter period, and thus the portfolios are quarterly rebalanced on rolling basis. Portfolio 1 contains countries with lowest value of the above product, and portfolio 4 the highest. After portfolios' currency compositions are known at the beginning of quarter t, US investors short US dollars and long equally weighted portfolios P of foreign currencies F to earn the quarterly returns $XR_{t+1}^{-US,P}$ realized at the beginning of quarter $t+1$

$$XR_{t+1}^{-US,+F} = \frac{S_t^F}{S_{t+1}^F}(1+\frac{r_t^F}{4}) - (1+\frac{r_t^{US}}{4}); \qquad XR_{t+1}^{-US,+P} = \sum_{F\in P}^{K_P} \frac{1}{K^P} XR_{t+1}^{-US,+F}.$$

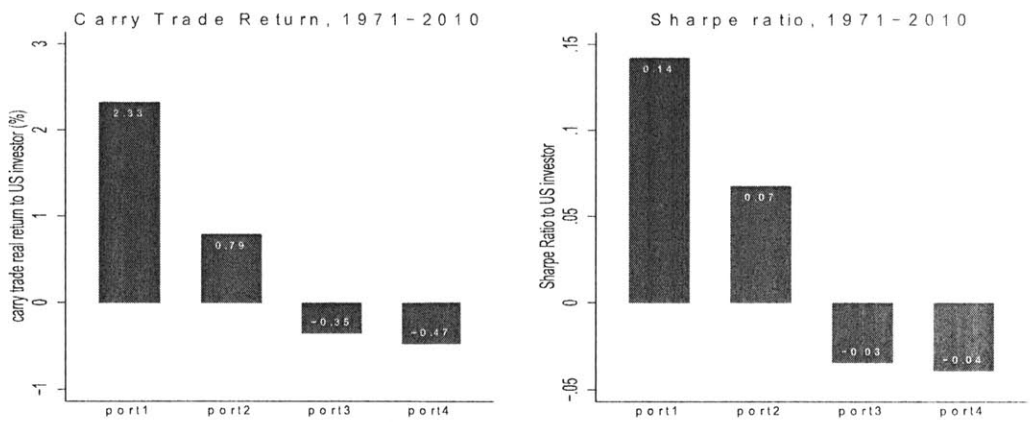
By the convention adopted here, spot exchange rate S_t^F is the number of foreign currency units per US dollar. These spot exchange rates are sampled simultaneously with the above interest rates.[66] To compute the real carry trade returns to US investors, we subtract US inflation from the above nominal returns $XR_{t+1}^{-US,+P}$. The US inflation is constructed as percentage change of "US quarterly consumer price index (CPI) series". Finally, the annualized real carry trade returns for each portfolio are obtained by compounding the quarterly counterpart values.[67] We note that because OECD's "Quarterly National Accounts" database is unbalanced (data start at different times for different countries, see data appendix), when we match it to IMF's IFS dataset, not all OECD countries are available at the same time for the purpose of portfolio sorting.

Figure 1-6 plots the mean annualized returns and Sharpe ratios on four equally weighted carry trade portfolios. The figure shows a monotonically inverse relationship between mean returns and the values of product of nontraded output variance and size across portfolios. Portfolio 1 earns a mean annual return of 2.33% (Sharpe ratio of 14%), and portfolio 4 a return of -.47% (Sharpe ratio of -4%) to US investors. Thus a long-short portfolio strategy (long portfolio 1, short portfolio 4) earns mean annual return of 2.8%, and Sharpe ratio of around 20%. This empirical inverse relationship is supported by our rational theory

[66] Both nominal interest rate series r_t and spot exchange rate series S_t^F are sampled at quarterly frequency from IMF's International Financial Statistics (IFS) database.

[67] Because portfolios are rebalanced quarterly, the currency compositions of portfolios do not necessarily stay fixed over the course of any year.

Figure 1-6: Carry trade excess returns and Sharpe ratios for portfolios sorted on nontraded output risk



This figure presents means and Sharpe ratios of real excess returns on four quarterly rebalanced currency portfolios to US investors. The sample consists of quarterly data series for period 1971-2010. The portfolio are constructed by sorting currencies into four groups at beginning of quarter t based on the value of nontraded variance $\times$ gdp's size over the previous 8 quarters. Portfolio 1 contains currencies with the lowest value of nontraded variance $\times$ gdp's size, portfolio 4 the highest. Due to unbalances in macro-data series, countries' data become available at different times, and number of countries changes over time. See data appendix for further details.

concerning nontraded risk as summarized in Hypothesis 2 (section 1.5). The intuition is that, partner countries' risk-free bonds, as insurance instruments, are relatively less valuable when their domestic economic environments are more stable, and offer larger interest rates to benefit the carry trade investors. However, high-return portfolios' payoffs tend to go up and down together with US nontraded endowment. They thus pose a consumption risk to US investors and necessarily pay superior expected returns to stay attractive in equlibrium. Sorting portfolio based directly on nontraded output volatilities (coupled with sizes) provide direct empirical supports for the key role of nontraded risk in the current rational approach to intrenational asset pricing.

Linear factor analysis: Empirics

The theoretical analysis of section 1.5 suggests another very intuitive way to consider non-traded and traded consumption risk as two key pricing factors. From US investors' perspectives, fluctuations in US traded and nontraded consumption are risk, and payoffs that correlate with these consumptions are priced, and carry risk premium accordingly. In this section we use currency portfolios sorted on interest rates as test assets to estimated the prices of risk associated with these two consumption risk factors. We do not sort currency portfolios on the nontraded output volatilities because doing so amounts to replicating the empirical exercise of the previous section, which already offers evidences that US investors price the nontraded risk of carry-trade partner countries. Instead, the choice of currency portfolios sorted on interest rates aims to relate the consumption risk to the violation of uncovered interest rate parity, which has been most robustly observed in these interest-rate-sorted currency portfolios. Below we discuss, in order, the estimation procedure, the data, and estimation results.

We empirically identify the US traded and nontraded consumption variations as risk factors for US investor: $f_{T,t+1}^{US} = \frac{C_{T,t+1}^{US} - C_{T,t}^{US}}{C_{T,t}^{US}}$, $f_{N,t+1}^{US} = \frac{C_{N,t+1}^{US} - C_{N,t}^{US}}{C_{N,t}^{US}}$. Using carry trade portfolio excess returns $XR_{t+1}^{-US,+P}$ as test assets, the fundamental Euler pricing equation (see section 1.5) can be written as[68]

$$E_t\left[\left\{1 - b_T\left(f_{T,t+1}^{US} - \mu_T^{US}\right) - b_N\left(f_{N,t+1}^{US} - \mu_N^{US}\right)\right\} XR_{t+1}^{-H,+F}\right] = 0,$$

where $\mu_T^{US} \equiv E\left[f_{T,t+1}^{US}\right]$, $\mu_N^{US} \equiv E\left[f_{N,t+1}^{US}\right]$ are *unconditional* means of the factors. The latter form readily suits a GMM process to estimate the factor loadings $\{b_T, b_N\}$. Consequently, follow the factor prices $\{\lambda_T^{US}, \lambda_N^{US}\}$ of the traded and nontraded risk, and the exposures $\{\beta_T^{US,P}, \beta_N^{US,P}\}$ of currency portfolios P to the US traded and nontraded consumption risk

[68]This equation results from the standard Euler equation $E_t\left[\left(1 + dm_{t+1}^{US} - E[dm_{t+1}^{US}]\right) XR_{t+1}^{-US,+P}\right] = 0$ and the linear factor pricing specification $\log \frac{M_{t+1}^{US}}{M_t^{US}} \equiv dm_{t+1}^{US} = b_T f_{T,t+1}^{US} + b_N f_{N,t+1}^{US}$. See section 1.5.

(see section 1.5)

$$\begin{bmatrix} \lambda_T^{US} \\ \lambda_N^{US} \end{bmatrix} = \left[Cov(\vec{f}^{US}, \vec{f}^{US})\right] \begin{bmatrix} b_T \\ b_N \end{bmatrix},$$

$$\begin{bmatrix} \beta_T^{US,P} \\ \beta_N^{US,P} \end{bmatrix} = \left[Cov(\vec{f}^{US}, \vec{f}^{US})\right]^{-1} \begin{bmatrix} Cov(f_T^{US}, XR^{-US,+P}) \\ Cov(f_N^{US}, XR^{-US,+P}) \end{bmatrix}$$

where $\left[Cov(\vec{f}^{US}, \vec{f}^{US})\right]$ is the covariance matrix of risk factors. Thus the GMM procedure employed to estimate factor loading b's also estimates factor prices λ's and portfolio risk exposures β's.

Currencies are sorted into four portfolios based on previous nominal interest rates in a procedure similar to the one presented in the above section. Portfolio 1 contains currencies associated with the lowest interest rates, portfolio 4 the highest rates. For this sorting, we use current quarter's nominal interest rates sourced from IMF. The quarterly carry trade excess returns $XR_{t+1}^{-US,+P}$ to US investors are computed over the next three-month periods. This return computation is identical to that of above section. The risk factors f_T^{US}, f_N^{US} are computed as quarter-to-quarter percentage changes of per-capita real US traded and nontraded consumption respectively. The US consumption and CPI series are from US Bureau of Economic Analysis' "Quarterly US consumption expenditures and price indexes". We identify the personal consumption expenditures on "services" as nontraded consumption, and on "goods" as traded consumption (see data appendix for further details).[69] After having constructed the quarterly series of portfolio returns $XR_t^{-US,+P}$ and factors $f_{T,t}^{US}$, $f_{N,t}^{US}$, we employ a two-stage GMM procedure on the above Euler equation to estimates factor loadings b_T, b_N jointly with the first moments μ_T^{US}, μ_N^{US} of the factors, as detailed in Menkhoff et al. (2011).[70] Finally, traded and nontraded factor prices λ_T^{US}, λ_N^{US} and portfolio risk exposures $\beta_T^{US,P}$, $\beta_N^{US,P}$ are deduced from the above simple matrix operation. Their standard errors are

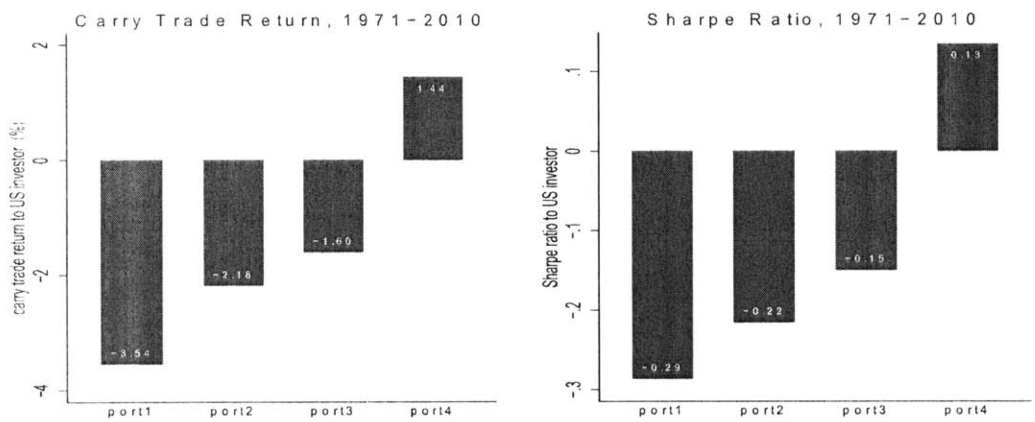
[69] It is important to note that we should not use US output series (in the output approach to GDP) for the current factor analysis. This is because for traded component, due to trades, the US traded output is not the same as US traded consumption. And in the theory being tested, it is the consumption risk that matters for the pricing.

[70] We also use lagged values of the carry trade portfolio returns as instruments.

determined from GMM-generated standard errors of factor loading b's and the delta method, as suggested by Burnside et al. (2011).

Figure 1-7 plots the mean annualized returns and Sharpe ratios on four equally weighted carry trade portfolios. The figures show a monotonic relationship between mean returns

Figure 1-7: Carry trade excess returns and Sharpe ratios for portfolios sorted on nominal interest rates



This figure presents means and Sharpe ratios of real excess returns on four quarterly rebalanced currency portfolios to US investors. The sample consists of quarterly data series for period 1971-2010. The portfolio are constructed by sorting currencies into four groups at beginning of quarter t based on the value of nominal interest rate available then. Portfolio 1 contains currencies with the lowest nominal interest rates, portfolio 4 the highest. Due to unbalances in interest rate and spot exchange rate series, countries' data become available at different times, and number of countries changes over time. See data appendix for further details.

in carry trades and the values of mean interest rates across portfolios. This in essence exhibits the violation of UIP and have been widely documented in the literature.[71] It is this monotone that qualifies these four carry trade portfolios as test assets for the empirical analysis of the current linear factor model. Accordingly, Table 1.7 reports the estimated factor prices. Both factor prices for traded and nontraded risk are positive and significant. Quantitatively, one additional "unit" of exposure to US nontraded consumption risk (i.e., $\Delta\beta_N = 1$) boosts the expected return on the strategy by 32 basis points. The corresponding

[71]For recent related work on UIP violation at protfolio level, see e.g. Burnside et al. (2011), Lustig et al. (2011), Menkhoff et al. (2011).

Table 1.7: Estimation of factor prices in linear factor models

		Nontraded consumption	Traded consumption
Factor prices (%)		.32*** (.02)	.34*** (.07)
beta's	port. 1	-1.92	.49
	port. 2	-1.61	.41
	port. 3	-1.51	.87
	port. 4	-1.87	2.31

Note: Upper panel reports the GMM annualized estimates of the factor prices (in percentage points), lower panel reports the estimates of the portfolios' exposures to risk factors (i.e. beta's) in the carry trade linear factor model using four quarterly rebalanced currency portfolios as test assets. HAC standard errors for the factor prices are obtained by two-stage GMM procedure using constant and lagged carry trade portfolio returns as instruments, and are reported in parenthesis. The currencies are sorted based on interest rates. The sample consists of quarterly data series for the period 1971-2010.

figure for US traded consumption risk is 34 basis points. Most importantly, the positive nontraded factor price well suits the rational implication of nontraded risk.[72] As nontraded output are largely confined and consumed within country's border, fluctuations in nontraded consumption growths are perceived as risk by all host countries. The proposition 3 then asserts that nontraded factor price λ_N^H should always be positive for all countries H. The results reported in table 1.7 thus empirically confirms this assertion from US investors' perspective. Beyond that, the results also show that fluctuations in US traded consumption are perceived as a risk by US investors. Table 1.7 also reports the estimated consumption betas for four currency portfolios. Values of betas vary across portfolios implying that foreign countries with different interest rate levels correlate differently with US traded and nontraded consumption growth. While the current two-factor model most likely leaves out other risk

[72]In the current factor pricing model, the expected excess return on any asset is $E[XR] = \lambda_T \beta_T + \lambda_N \beta_N$. The positive factor price $\lambda_N > 0$ implies that any payoff positively correlated with nontraded consumption growth, $\beta_N > 0$, commands a positive expected return components. In other words, nontraded consumption growth volatility is a risk to investor.

factors,[73] the movements in US traded and nontraded consumption growth are statistically significant sources of risk being priced in the currency market.

1.8 Conclusion

This paper points out the effects of nontraded output growth risk on national asset and international investment returns. Nontraded output growth risk is particularly impactful, because this output makes a large share of GDP and is consumed almost entirely by home population. In contrast, country-specific traded output growth risk can be diversified by means of commodity trades. Hence our analysis calls for a careful decomposition of GDP into traded and nontraded output components before assessing its role on the determination of asset prices.

Nontraded output shocks are nevertheless not entirely internalized by home countries because countries engage in trades in other goods as well. While, to a certain extent, trades weaken the impact of nontraded output risk on the home country, trades also transmit and thus broaden the impact of home nontraded output shocks to all trade partners of the home country. This mechanism is behind the profits of all international strategies, including carry trades. This is because the global traded output risk spreads fairly equally across countries, and thus drops out of strategies involving off-setting positions in different national markets.

The frameworks in which a risk, apparently intrinsic to only one party, actually affects other parties are pervasive in the real world. Examples include any social network settings, financial institutions, or interbank systems featuring counter-party risk. The asset pricing analysis presented here for the international finance setting, especially in regards to transaction costs and incomplete markets, would help shed light on other interesting frameworks just mentioned. We hope to address these frameworks in future work.

[73]We can infer from table 1.7 that these two risk factors account for about 15% of the expected carry trade return to US investors.

1.9 Appendices

1.9.1 Data sources

The empirical part of the current paper concerns only countries that belong to the Organisation for Economic Co-operation and Development (OECD) principally because we reasonably expect that data quality for these developed economies should be higher than the rest of the world.

OECD countries: currently, there are 34 OECD member states listed as follows; Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States.

Eurozone countries: among OECD states, the following 15 belong to the Economic and Monetary Union (a.k.a., Eurozone or Euro area) with respective adopting date in the parenthesis;[74] Austria (01/01/1999), Belgium (01/01/1999), Estonia (01/01/2011), Finland (01/01/1999), France (01/01/1999), Germany (01/01/1999), Greece (01/01/2001), Ireland (01/01/1999), Italy (01/01/1999), Luxembourg (01/01/1999), Netherlands (01/01/1999), Portugal (01/01/1999), Slovak Republic (01/01/2009), Slovenia (01/01/2007), Spain (01/01/1999).

"Aggregate National Accounts: Gross domestic product" contains the following annual real output series available either in national currency or USD, constant prices of OECD base year 2000 (output approach to GDP): Gross domestic product (B1_GA); Wholesale and retail trade, repairs, hotels and restaurants, transport (B1GG_I); Financial intermediation, real estate, renting and business activities (B1GJ_K); Construction (B1GF); Other service activities (B1GL_P), sourced from OECD.org, downloadable via "OECD.Stat Extracts".

"Total Population series" contains annual data on population, sourced from World Bank World Development Indicators (WDI).

"IMF Exchange Rates and short-term Treasury Bill Rates" provide spot exchange rates and nominal interest rates sourced from IMF International Financial Statistics (IFS) at both quarterly and annually frequencies. Treasury Bill Rates are associated with maturities varying from one to three months. For those countries where the short-term Treasury Bill Rates are not available, we use Money Market Rates from the same sources. Consumer price index (CPI) series is also provided by IFS (at quarterly and annual frequencies). Inflation then is computed as the period-to-period percentage change of the consumer price

[74] Only two other Eurozone states are Cyprus and Malta, but they do not belong to OECD and are not considered in the empirical analysis of the current paper.

index.

"Trade-to-GDP ratio" (i.e., trade openness) contains the value of ratio of nominal national total import plus export over national GDP, sourced from OECD Trade Indicators database, downloadable via "OECD.Stat Extracts".

"OECD exchange rate series" contain the exchange rates, in national units per USD (USD monthly average), for all OECD countries. The series is sourced from OECD Main Economic Indicators (MEI), downloadable via "OECD.Stat Extracts".

"Three-month nominal interest rate series" of OECD countries are provided by Data Stream. These original daily series consist of the bid, ask (i.e., offered) and mid quotes for 3-month Eurocurrency-deposit interest rates (end-of-day quotes from London market). This dataset is unbalanced; Australia's series starts in 1997, Greece in 1994, New Zealand in 1997, Norway in 1997, Portugal in 1993, Spain in 1992, Sweden in 1997. Other OCED countries's series start earlier (before 1984, the date when the spot exchange rate series start, and hence this date 1984 does not pose further data limit constraints for the computation of carry trade returns).

"US quarterly consumer price index (CPI) series" is sourced from OECD Main Economic Indicators (MEI), downloadable via "OECD.Stat Extracts".

"Quarterly National Accounts" database contains quarterly series on expenditure on services ("P314B: Services") for individual OECD countries. This is a component of the expenditure on total private consumption, in the expenditure approach to GDP. For those OECD countries where these series on services expenditure are not available, we substitute them by the quarterly services output series ("B1GG_P-Services"). These quarterly dataset is quite unbalanced, namely available data of different countries start at quite different time. Quarterly US consumption data series are very limited, being available only from 1995 onward. Consequently, the Quarterly US consumption data will be sourced from the US Bureau of Economic Analysis (see next).

"Quarterly US consumption expenditures and price indexes" are series from US Bureau of Economic Analysis. Table 2.3.5. therein contains "Personal Consumption Expenditures by Major Type of Product". Table 2.3.4. contains "Price Indexes for Personal Consumption Expenditures by Major Type of Product". We identify the personal consumption expenditures on services (i.e., the component "Services" listed in these tables) as the US nontraded consumption. We identify the personal consumption expenditures on other goods (i.e., the component "Goods" listed in these tables) as the US traded consumption. These quarterly series start well before 1971 (all our empirical studies in the current paper concern periods starting in 1971 or later).

"Trade in Services" is from OECD's International Trade and Balances of Payments database. This dataset includes the export and import series (in the transactions between residents and non-residents), in unit of countries' currencies and at annual frequency, of financial services, construction services and other services. Financial services cover financial intermediary and auxiliary services (except those of insurance enterprises and pension funds) conducted between residents and non-residents. Included are intermediary service fees, such as those associated with letters of credit, bankers' acceptances, lines of credit, financial leasing, and foreign exchange transactions. Construction services cover work performed on construction projects and installations by employees of an enterprise in locations outside the economic territory of the enterprise. Other business services cover various categories of service transactions between residents and non-residents. They include (i) merchanting and other trade-related services, (ii) operational leasing services (rental) without operators, (iii) legal, accounting, management consulting, and public relation services, (iv) advertising, market research and public opinion polling services transacted between residents and non-residents (v) research and development services, (vi) architectural, engineering and other technical services, (vii) agricultural, mining and on-site processing services, (viii) other miscellaneous business, professional and technical services. See original data source for further details.

"OECD Structural Analysis (STAN)" database provides, for each OECD country, the annual nominal output series (in national currency) and the corresponding deflator series (of OECD base year 2000) for various industries. It also provides country-specific annual nominal import and export series (in national currency) and the corresponding deflator series (of OECD base year 2000) for these industries. We construct the real output series by dividing the nominal series by the respective deflator series. The constructed real output series are thus in national currency, constant price of base year 2000. All real output series are detrended using Hodrick-Prescott filter. The following non-nested industries are listed in STAN, with International Standard Industrial Classification of All Economic Activities (ISIC) Rev. 3 identification given in parenthesis: Agriculture hunting and related service activities (01); Forestry logging and related service activities (02); Fishing; fish hatcheries; fish farms and related services (05); Mining of coal and lignite extraction, of peat (10); Extraction of crude petroleum and natural gas and related services (11); Mining of uranium and thorium ores (12); Mining of metal ores (13); Other mining and quarrying (14); Food products and beverages (15); Tobacco products (16); Textiles (17); Wearing apparel, dressing and dying of fur (18); Leather, leather products and footwear (19); Wood and products of wood and cork (20); Pulp, paper and paper products (21); Printing and publishing (22); Coke, refined petroleum products and nuclear fuel (23); Chemicals and chemical products (24ex2423); Pharmaceuticals (2423); Rubber and plastics products (25); Other non-metallic mineral products (26); Iron and steel (271+2731); Non-ferrous metals (271+2732); Fabricated metal products, except machinery and equipment (28); Machinery and equipment n.e.c. (29); Office, accounting and computing machinery (30); Electrical machinery and apparatus n.e.c. (31); Radio, television and communication equipment (32); Medical, precision and optical instruments (33); Motor vehicles, trailers and semi-trailers (34); Building and repairing of ships and boats (351); Aircraft and spacecraft (353); Railroad equipment and transport equipment n.e.c. (352+359); Manufacturing nec (36); Recycling (37); Electricity, gas, steam and hot water supply (40); Collection, purification and distribution of water (41); Construction (45); Sale, maintenance and repair of motor vehicles; retail sale of fuel (50); Wholesale, trade & commission excl. motor vehicles (51); Retail trade excl. motor vehicles; repair of household goods (52); Hotels and restaurants (55); Land transport, transport via pipelines (60); Water transport (61); Air transport (62); Supporting and auxiliary transport activities (63); Post and telecommunications (64); Financial intermediation except insurance and pension funding (65); Insurance and pension funding, except compulsory social security (66); Activities related to financial intermediation (67); Real estate activities (70); Renting of machinery and equipment (71); Computer and related activities (72); Research and development (73); Other business activities (74); Public administration and defense compulsory social security (75); Education (80); Health and social work (85); Sewage and refuse disposal sanitation and similar activities (90); Activities of membership organization n.e.c. (91); Recreational cultural and sporting activities (92); Other service activities (93); Private households with employed persons (95); Extra-territorial organizations and bodies (99); High technology manufactures (N/A); Medium-high technology manufactures (N/A); Medium-low technology manufactures (N/A); Low technology manufactures (N/A).

1.9.2 Derivations and proofs: Basic model

In the basic model with complete market and no trade friction, in equilibrium the marginal utilities of traded consumption equal across countries, which give K FOCs; $M_T = \frac{\partial U^H}{\partial C_T^H}$ $\forall H = 1, \ldots, K$. The market clearing condition for traded good presents another equation to solve for $K+1$ unknowns; $\{C_T^H\}_{H=1}^K$ and M_T. We log-linearize the system to obtain approximative solution in closed form.

Equilibrium log consumption (1.3): Plugging the expression (1.1) for U^H into the FOC (1.2), and log-linearizing this FOC around the steady state corresponding to the symmetric configuration $\{\delta_T^H = \delta_N^H;\ \delta_T^H/\Lambda^H = \delta_T^F/\Lambda^F\}$ yield an approximate equation[75]

$$m_T \approx \lambda^H - \rho t + (\epsilon - \gamma)(\omega_T c_T^H + \omega_N \delta_N^H) - \epsilon c_T^H + \log \omega_T.$$

Similarly, log-linearizing the traded good market clearing equation yields (where $\lambda = \log \Lambda = \log \sum_H^K \Lambda^H$)

$$\sum_H^K \frac{\Lambda^H}{\Lambda} c_T^H = \delta_T + \sum_H^K \frac{\Lambda^H}{\Lambda} \lambda^H - \lambda. \tag{1.27}$$

Substituting c_T^H from the first equation above into the second equation gives m_T, and then c_T^H in (1.3).

Country-specific stochastic discount factor (1.5): In pricing country-specific financial assets, the appropriate measures are country-specific consumption baskets (i.e., national currencies in the current consumption-based setting). A country-specific consumption basket is the lowest-cost bundle of traded and nontraded consumption that delivers a unit of country's utility, given the consumption goods' prices $\{P_T^H \equiv 1, P_N^H\}$ (in term of traded goods). The basket's composition $\{C_T^H, C_N^H\}$ and value P^H solve $\min_{C_T^H, C_N^H} P^H \equiv C_T^H + C_N^H P_N^H$ subject to $\left[\omega_T (C_T^H)^{1-\epsilon} + \omega_N (C_N^H)^{1-\epsilon}\right]^{\frac{1}{1-\epsilon}} = 1$. Then follows the value of consumption basket in term of traded good

$$P_t^H = \left[\omega_T^{\frac{1}{\epsilon}} + \omega_N^{\frac{1}{\epsilon}} (P_N^H)^{\frac{1-\epsilon}{-\epsilon}}\right]^{\frac{-\epsilon}{1-\epsilon}};$$

From this and M_T above follows the identity in equilibrium $M_t P_t^H = M_t^H$, where $M_t^H \equiv \frac{\partial U^H}{\partial C^H} = e^{-\rho t}(C^H)^{-\gamma}$ and C^H is the country-specific consumption aggregator.[76] The current price of the country-specific risk-free bond (that pays one unit of country-specific consumption basket at time s) is

$$B_{t,s}^H = \frac{1}{P_t^H} E_t \left[\frac{M_s}{M_t} P_s^H\right] = E_t \left[\frac{M_s^H}{M_t^H}\right].$$

[75] We recall that lower-case letters always denote logarithms; $m \equiv \log M$, $\lambda \equiv \log \Lambda$, $c = \log C$, $\delta = \log \Delta$ and so on.

[76] In contrast with the country-specific M^H, M_T is the marginal utility with respect to traded good and is same for all countries in complete market settings.

It is this pricing equation that establishes the above M_t^H as the country-specific SDF of country H. That is, prices computed using this SDF are in unit of country-specific consumption basket. Log-linearizing $m^H = \log M^H = -\rho t - \gamma C^H$ and using log equilibrium traded consumption c_T^H in (1.3) yield country-specific log SDF (1.5).

Costly trades: Suppose that home country is an importer (case 1) and trades take place, the variation of social planner's Lagrangian with respect to non-binding consumptions $\frac{\partial}{\partial C_F^H}$, $\frac{\partial}{\partial C_F^F}$ produces FOC $\left(C_H^H + C_F^H\right)^{-\gamma} = (1+\theta)\left(C_H^F + C_F^F\right)^{-\gamma}$. Combining this with binding consumption $C_H^H = \Delta^H$, $C_F^H = 0$, and market clearing condition $C_F^F + (1+\theta)C_F^H = \Delta^F$ yields (1.7). From this we can also find home SDP $M^H = e^{-\rho t}\left(\Delta^H + C_F^H\right)^{-\gamma}$. The risk-free rate r^H is the opposite to expected growth rate of M^H; $r^H = -\frac{1}{dt}E_t\left[\frac{dM^H}{M^H}\right]$. Plugging equilibrium consumption solutions (1.7) into M^H, and an application of Ito lemma yields (assuming independent endowments Δ^H, Δ^F)

$$r^H = \rho + \gamma\frac{(1+\theta)\mu^H\Delta^H + \mu^F\Delta^F}{(1+\theta)\Delta^H + \Delta^F} - \frac{1}{2}\gamma(\gamma+1)\frac{(1+\theta)^2(\sigma^H)^2(\Delta^H)^2 + (\sigma^F)^2(\Delta^F)^2}{[(1+\theta)\Delta^H + \Delta^F]^2}$$

which is a more explicit version of (1.8).

Proof of Proposition 1. From (1.5) follow the partial derivatives

$$\begin{aligned}\frac{\partial m^H}{\partial \delta_N^H} &= -\gamma\omega_N\left[1 - \alpha(\gamma-\epsilon)\omega_T\left(1 - \frac{\Lambda^H}{\Lambda}\right)\right],\\ \frac{\partial m^F}{\partial \delta_N^H} &= -\gamma\omega_N\left[\alpha(\gamma-\epsilon)\omega_T\frac{\Lambda^H}{\Lambda}\right].\end{aligned}$$

Evidently, $\left|\frac{\partial m^H}{\partial \delta_N^H}\right| > \left|\frac{\partial m^F}{\partial \delta_N^H}\right|$ because $\gamma - \epsilon > 0$ (assumption 1, section 1.3). ■

Proof of eq. (1.9) **and Proposition 2.** We start with the differential representation for SDF M^H

$$\frac{dM^H}{M^H} = -r^H dt - \eta^H dZ^H; \quad m^H = \log m^H \Longrightarrow dm^H = -\left(r^H + \frac{1}{2}(\eta^H)^2\right) - \eta^H dZ^H.$$

where η^H is the home market price of risk. Similar relations hold for M^F and m^F. Plugging these into the realized carry trade excess return $XR_{t+dt}^{-H,+F}$ (upper equation in (1.9)), applying Ito's lemma and taking the conditional expectation yield

$$
\begin{aligned}
E_t\left[XR_{t+dt}^{-H,+F}\right] &= E_t\left[\frac{1+\frac{dM^F}{M^F}}{1+\frac{dM^H}{M^H}}(1+r^F dt)-(1+r^H dt)\right] \\
&= E_t\left[\left(1+dm^F+\frac{1}{2}(dm^F)^2\right)\left(1-dm^H+\frac{1}{2}(dm^H)^2\right)(1+r^F dt)-(1+r^H dt)\right] \\
&= E_t\left[dm^F+\frac{1}{2}(dm^F)^2-dm^H+\frac{1}{2}(dm^H)^2-dm^H dm^F+r^F dt-r^H dt\right] \\
&= (\eta^H)^2-\eta^H\eta^F=-Cov_t\left[dm^H, dm^F-dm^H\right],
\end{aligned}
$$

which is (1.9). Next, combining (1.5) and (1.10) implies the key expression for expected carry trade excess return (1.11) of Proposition 2. ■

Proof of Proposition 3. We first develop (1.13) to obtain more explicit expressions for λ_T and λ_N

$$
\lambda_T^H = Var(f_T^H)b_T + Cov(f_T^H, f_N^H)b_N; \qquad \lambda_N^H = Cov(f_T^H, f_N^H)b_T + Var(f_N^H)b_N.
$$

Plugging $\{b_T, b_N\}$ and $\{f_T^H, f_N^H\}$ from (1.12) into above expressions yields (1.14) of Proposition 3 and (1.15) for factor prices associated with nontraded and traded consumption growth risk respectively. ■

1.9.3 Derivations and proofs: Arbitrary trade configurations

This appendix presents technical derivations of the results concerning arbitrary trade configurations of section 1.6.1. Here, there are K countries and l different types of traded goods. A (generic) traded good of type h is consumed by some subset of K_h countries, and a (generic) country H consume l^H types of traded goods (apart from the country's intrinsic nontraded good). Consumption tastes $\left\{\{\omega_{h,T}^H\}_{h=1,\dots,K_h}, \omega_N^H\right\}$ (with normalization $\sum_h^{l^H} \omega_{h,T}^H + \omega_N^H$) are heterogeneous across countries. Country-good count and good-country count are necessarily identical

$$\sum_h^l K_h = \sum_H^K l^H. \tag{1.28}$$

The assumption of complete financial market is maintained here and implies that marginal utilities of a traded good are equalized across counries that consume this traded good in equilibrium (this is a FOC in the social planner's optimization problem). Furthermore, the physical market for this traded good h is also cleared among K_h countries,

$$M_h = \Lambda^H \frac{\partial U^H}{\partial C_h^H}; \qquad \sum_{H \in K_h} C_h^H = \Delta_{h,T}, \qquad \forall h;\ \forall H \in K_h$$

Thus, in total we have $\sum_h K^h$ equations[77] and $\sum_H l^H$ unknowns consumptions $\{C_h^H\}_{h \in l^H}^{H=1,\dots,K}$. By virtue of (1.28), in principle, the social planner's optimization alone is sufficient to determine all equilibrium traded consumption allocations $\{c_h^H\}$. In practice, however, the above system is highly nonlinear for CES utilities (1.1). To obtain approximate solution we log-linearize above system, which yields a set of $\sum_h K^h$ linear equations and that same number of unknowns,

$$\begin{cases} m_h = \lambda^H - \rho t + \log \omega_{h,T}^H + (\epsilon - \gamma)\left(\sum_j^{l^H} \omega_{j,T}^H c_j^H + \omega_N^H \delta_N^H\right) - \epsilon c_h^H, \\ \sum_{H \in K_h} \frac{\Lambda^H}{\Lambda_h} c_h^H = \delta_{h,T} + \sum_{H \in K_h} \frac{\Lambda^H}{\Lambda_h} \lambda^H - \lambda_h; \qquad \Lambda_h \equiv \sum_H^{K_h} \Lambda^H;\ \lambda^H \equiv \log \Lambda^H;\ \lambda_h \equiv \log \Lambda_h; \end{cases} \tag{1.29}$$

for all h, and $H \in K_h$. Albeit linearity, this system is (almost arbitrarily) large due to arbitrary trade configuration. We first note that we can always reduce this system to l equations and l unknowns. Multiplying both sides of above eq for m_h by ω_h^H, then summing

[77] For each traded good h, we have one market clearing equation and $(K_h - 1)$ FOCs (because M_h is not known a priori).

over $h \in l^H$ (while keeping H fixed) generate a relation between $\sum \omega_{h,T}^H m_h$ and $\sum \omega_{h,T}^H c_h^H$.

$$\begin{aligned}\sum_h^{l^H} \omega_{h,T}^H m_h = & \quad \left(\lambda^H - \rho t + \log \omega_{h,T}^H\right)(1-\omega_N^H) + (\epsilon - \gamma)(1-\omega_N^H)\omega_N^H \delta_N^H \\ & - \left[\epsilon \omega_N^H + \gamma(1-\omega_N^H)\right] \sum_h^{l^H} \omega_{h,T}^H c_h^H \\ = & \quad \left(\lambda^H - \rho t + \log \omega_{h,T}^H\right)(1-\omega_N^H) + (\epsilon - \gamma)(1-\omega_N^H)\omega_N^H \delta_N^H - \frac{1}{\alpha^H} \sum_h^{l^H} \omega_{h,T}^H c_h^H \quad (1.30)\end{aligned}$$

where we have used the consumption tastes normalization, $\sum_h^{l^H} \omega_{h,T}^H + \omega_N^H = 1$ and the definition (1.17) of weighted elasticity of substitution

$$\alpha^H \equiv \frac{1}{\epsilon \omega_N^H + \gamma(1-\omega_N^H)}.$$

This relation is the key bridge that connects the country-specific SDF M^H (or marginal utilities of consumption aggregator) to the marginal utilities of traded goods M_h. Indeed, by log-linearizing $m^H \equiv \log M_t^H = \log \frac{\partial U^H}{\partial C^H}$ we obtain

$$m^H = -\rho t - \gamma \left(\sum_h^{l^H} \omega_{h,T}^H c_h^H + \omega_N^H \delta_N^H \right) = \# - \epsilon \gamma \alpha^H \omega_N^H \delta_N^H + \gamma \alpha^H \sum_h^{l^H} \omega_{h,T}^H m_h \quad (1.31)$$

where we have omitted the deterministic terms (which are independent of stochastic endowments δ's). Backing out $\sum \omega_{h,T}^H c_h^H$ in term of $\sum \omega_{h,T}^H m_h$ from (1.30) and substituting it into upper equation of (1.29) give an consumption allocation c_h^H in term of $\{m_j\}$,

$$c_h^H = \alpha^H \left(\lambda^H - \rho t + \log \omega_{h,T}^H\right) - (\gamma - \epsilon)\alpha^H \omega_N^H \delta_N^H - \frac{m_h}{\epsilon} + \frac{(\gamma - \epsilon)\alpha^H}{\epsilon} \sum_j^{l^H} \omega_{j,T}^H m_j. \quad (1.32)$$

Multiplying both sides of this equation by $\frac{\Lambda^H}{\Lambda_h}$, summing over H, and plugging it into market clearing conditions (lower equation of (1.29)) indeed yield l linear equations (i.e., $h = 1, \ldots, l$)

$$\begin{aligned} m_h = & \quad -\epsilon \left(\sum_{H \in K_h} \frac{\Lambda^H}{\Lambda_h} \lambda^H - \lambda_h \right) + \epsilon \sum_{H \in K_h} \frac{\Lambda^H \alpha^H}{\Lambda_h} \left(\lambda^H - \rho t + \log \omega_h^H \right) \\ & -\epsilon(\gamma - \epsilon) \sum_{H \in K_h} \frac{\Lambda^H}{\Lambda_h} \alpha^H \omega_N^H \delta_N^H - \epsilon \delta_{h,T} + (\gamma - \epsilon) \sum_{H \in K_h} \frac{\Lambda^H}{\Lambda_h} \alpha^H \sum_j^{l^H} \omega_{j,T}^H m_j, \quad \forall h = 1, \end{aligned} \quad (1.33)$$

for l unknowns $\{m_h\}$. We next solve this system and the equilibrium consumption (approximately) by iteration method. The procedure consists of 4 steps.

step 1: (Zeroth order of m_h) We conjecture that the global (aggregate) endowment $\delta_{h,T}$ of traded good type h dominates other endowment $\{\delta_{j,T}\}_{j \neq h}$ in the contribution to m_h. We then can decouple the above system and solve for each m_h separately in zeroth order. We also note that the term m_h on the right-hand side of above equation is negligible compared to term m_h on the left-hand side. Thus, in zeroth order, $\forall h = 1, \ldots, l$,

$$\begin{aligned} m_h^{(0)} = & \quad -\epsilon \left(\sum_{H \in K_h} \frac{\Lambda^H}{\Lambda_h} \lambda^H - \lambda_h \right) + \epsilon \sum_{H \in K_h} \frac{\Lambda^H \alpha^H}{\Lambda_h} \left(\lambda^H - \rho t + \log \omega_h^H \right) \\ & -\epsilon(\gamma - \epsilon) \sum_{H \in K_h} \frac{\Lambda^H}{\Lambda_h} \alpha^H \omega_N^H \delta_N^H - \epsilon \delta_{h,T} \end{aligned}$$

step 2: (First order of m_h) We substitute the zeroth-order $m_h^{(0)}$ above into right-hand size of (1.33) to obtain first-order expression for m_h (we again omit all deterministic terms, which are independent of stochastic endowments δ's)

$$\begin{aligned} m_h^{(1)} = \# - \epsilon \left(\delta_{h,T} + (\gamma - \epsilon) \textstyle\sum_{H \in K_h} \frac{\Lambda^H}{\Lambda_h} \alpha^H \sum_j^{l^H} \omega_{j,T}^H \delta_{j,T} \right) \\ -\epsilon(\gamma - \epsilon) \textstyle\sum_{H \in K_h} \frac{\Lambda^H}{\Lambda_h} \alpha^H \left(\omega_N^H \delta_N^H + (\gamma - \epsilon) \sum_j^{l^H} \omega_{j,T}^H \sum_{J \in K_j} \frac{\Lambda^J}{\Lambda_j} \alpha^J \omega_N^J \delta_N^J \right) \end{aligned}$$

The coefficient associated with $\delta_{j,T}|_{j \neq h}$ is $(\gamma - \epsilon) \sum_{H \in K_h \cap K_j} \frac{\Lambda^H}{\Lambda_h} \alpha^H \omega_{j,T}^H$, so endowment of traded good of type j contributes more the marginal utility m_h of good h when there are more countries H that consume both goods. To be consistent with the log-linearization approximation, we do not need to go beyond the iteration's first order.

step 3: (Traded consumption allocation c_h^H) Substituting the first-order $m_h^{(1)}$ above into (1.32) yields equilibrium consumption[78] c_h^H in (1.16).

step 4: (Country-specific log SDF m^H) Substituting the first-order $m_h^{(1)}$ above into (1.31) yields equilibrium consumption c_h^H in (1.18).

1.9.4 Derivations and proofs: Incomplete markets

This appendix presents technical derivations of results concerning incomplete financial markets of section 1.6.2, in particular the equilibrium consumptions (1.22), (1.23). Substituting the conjectured consumption allocation (1.21) into (1.20) yields a more explicit expression for the log-linearized SDF

$$dm_T^H = -\frac{g^H}{\alpha^H} - \frac{a^H}{\alpha^H}d\delta_T - \left[(\gamma - \epsilon)\omega_N^H + \frac{b^{HH}}{\alpha^H}\right] d\delta_N^H - \sum_{F \neq H} \frac{b^{HF}}{\alpha^H} d\delta_N^F. \tag{1.34}$$

In the current setting (K countries of homogeneous size with a single traded good), the market clearing condition in log-linearized form is a special case of (1.27) (where all Λ^H are identical) and reads $\sum_{H=1}^K c_T^H = \sum_{H \in \mathcal{D}} c_T^H + \sum_{H \notin \mathcal{D}} c_T^H = K\delta_T - K \log K$, which implies

$$\sum_{H=1}^K dc_T^H = \sum_{H \in \mathcal{D}} dc_T^H + \sum_{H \notin \mathcal{D}} dc_T^H = K d\delta_T.$$

where d denotes the difference operator acting between t and $t+1$. Substituting the conjectured equilibrium consumption allocations (1.21) in above equation yields the a set of constraints for the solution parameters

$$\sum_H a^H = K; \qquad \sum_H g^H = 0; \qquad \sum_H b^{HF} = 0, \;\; \forall F. \tag{1.35}$$

[78]We note that the log-linearization approximation is accurate up to terms of order $\mathcal{O}(\omega_N)$, $\mathcal{O}(\omega_T)$. Consequently, we disregard all terms of order $\mathcal{O}((\omega_N)^2)$, $\mathcal{O}((\omega_T)^2)$, $\mathcal{O}(\omega_T\omega_N)$.

On the other hand, substituting dm^H in (1.34) into the law of one price (1.19) for $\frac{S_{T,t}}{B_{T,t}}$ implies that the following expression is the same for all H,

$$\log\left(\frac{S_{T,t}}{B_{T,t}}\right) = Cov_t\left[dm^H_{T,t+1}, \delta_{T,t+1}\right] = -\frac{a^H}{\alpha^H}\sigma_T, \quad \forall H.$$

Combining above two equations immediately yields

$$a^H = \frac{K\alpha^H}{\sum_{H=1}^K \alpha^H}; \qquad \alpha^H \equiv \frac{1}{\gamma\omega^H_T + \epsilon\omega^H_N}. \tag{1.36}$$

Similarly, the law of one price (1.19) for $\frac{S^F_{T,t}}{B^F_{T,t}}$ implies that $\log\left(\frac{S^F_{N,t}}{B^F_{N,t}}\right) = Cov_t\left[dm^H_{T,t+1}, \delta^F_{N,t+1}\right]$ is identical for each developed country $F \in \mathcal{D}$ and all countries H. Using (1.34), we have

$$(\gamma-\epsilon)\omega^F_N + \frac{b^{FF}}{\alpha^F} = \frac{b^{HF}}{\alpha^H} \qquad \text{for each } F \in \mathcal{D}, \ \text{ for all } H \neq F.$$

When combined with the constraint (1.35) above, this yields

$$\begin{cases} b^{FF} = -(\gamma-\epsilon)\omega^F_N\alpha^F\left(1 - \frac{\alpha^F}{\sum_{I=1}^K \alpha^I}\right), & \forall F \in \mathcal{D} \\ b^{HF} = (\gamma-\epsilon)\omega^F_N\alpha^F\alpha^H\frac{1}{\sum_{I=1}^K \alpha^I}, & \forall F \in \mathcal{D} \quad \forall H \neq F \end{cases} \tag{1.37}$$

In particular, given a choice of $F \in \mathcal{D}$, we note that $\frac{b^{HF}}{\alpha^H}$ is the same for all $H \neq F$.

Next, first substituting conjectured solution (1.21) into (1.34), and then into the law of one price (1.19) for bond $B_{T,t}$ imply that

$$\begin{aligned} A \equiv \quad & \frac{g^H}{\alpha^H} + (\gamma-\epsilon)\omega^H_N\mu^H_N + \frac{1}{\alpha^H}\sum_{F=1}^K b^{HF}\mu^F_N - \frac{1}{2}\frac{1}{(\alpha^H)^2}\sum_{F=1}^K \left(b^{HF}\right)^2(\sigma^F_N)^2 \\ & -\frac{1}{2}(\gamma-\epsilon)^2(\omega^H_N)^2(\sigma^H_N)^2 - (\gamma-\epsilon)\omega^H_N\frac{b^{HH}}{\alpha^H}(\sigma^H_N)^2, \end{aligned}$$

is the same for all H. Using (1.37), we separately rewrite the above expression for emerging and developed economies,

$$H \notin \mathcal{D}: \quad A = \frac{g^H}{\alpha^H} + (\gamma-\epsilon)\omega_N^H \mu_N^H + \sum_{F\notin\mathcal{D}} \frac{b^{HF}}{\alpha^H}\mu_N^F - \frac{1}{2}\sum_{F\notin\mathcal{D}}\left(\frac{b^{HF}}{\alpha^H}\right)^2 (\sigma_N^F)^2 \tag{1.38}$$

$$- \frac{1}{2}(\gamma-\epsilon)^2(\omega_N^H)^2(\sigma_N^H)^2 - (\gamma-\epsilon)\omega_N^H \frac{b^{HH}}{\alpha^H}(\sigma_N^H)^2$$

$$+ \left[\sum_{F\in\mathcal{D}} \frac{(\gamma-\epsilon)\omega_N^F \alpha^F}{\sum_{I=1}^K \alpha^I}\mu_N^F - \frac{1}{2}\sum_{F\in\mathcal{D}}\left(\frac{(\gamma-\epsilon)\omega_N^F\alpha^F}{\sum_{I=1}^K\alpha^I}\right)^2 (\sigma_N^F)^2\right]$$

$$H \in \mathcal{D}: \quad A = \frac{g^H}{\alpha^H} + \sum_{F\notin\mathcal{D}} \frac{b^{HF}}{\alpha^H}\mu_N^F - \frac{1}{2}\sum_{F\notin\mathcal{D}}\left(\frac{b^{HF}}{\alpha^H}\right)^2 (\sigma_N^F)^2 \tag{1.39}$$

$$+ \left[\sum_{F\in\mathcal{D}} \frac{(\gamma-\epsilon)\omega_N^F \alpha^F}{\sum_{I=1}^K \alpha^I}\mu_N^F - \frac{1}{2}\sum_{F\in\mathcal{D}}\left(\frac{(\gamma-\epsilon)\omega_N^F\alpha^F}{\sum_{I=1}^K\alpha^I}\right)^2 (\sigma_N^F)^2\right].$$

We note that the expressions within the square brackets are identical (i.e., independent) for all countries H (either $H \in D$ or $H \notin D$), and thus can be disregarded. The above requirement imposed by the law of one price on bond $B_{T,t}$ thus becomes

$$\left\{\frac{g^H}{\alpha^H} + (\gamma-\epsilon)\omega_N^H\mu_N^H - \frac{1}{2}(\gamma-\epsilon)^2(\omega_N^H)^2(\sigma_N^H)^2 - (\gamma-\epsilon)\omega_N^H\frac{b^{HH}}{\alpha^H}(\sigma_N^H)^2 \right.$$
$$\left. + \sum_{F\notin\mathcal{D}}\frac{b^{HF}}{\alpha^H}\mu_N^F - \frac{1}{2}\sum_{F\notin\mathcal{D}}\left(\frac{b^{HF}}{\alpha^H}\right)^2(\sigma_N^F)^2\right\}\Bigg|_{\forall H\notin\mathcal{D}}$$
$$= \left\{\frac{g^H}{\alpha^H} + \sum_{F\notin\mathcal{D}}\frac{b^{HF}}{\alpha^H}\mu_N^F - \frac{1}{2}\sum_{F\notin\mathcal{D}}\left(\frac{b^{HF}}{\alpha^H}\right)^2(\sigma_N^F)^2\right\}\Bigg|_{\forall H\in\mathcal{D}}. \tag{1.40}$$

This system has the following simple solution (of pooling type within developed economies, and within emerging economies), that also satisfies the constraint $\sum_H b^{HF} = 0$ in (1.35),

$$\begin{cases} b^{HF} = -\alpha^H, & \forall H \notin \mathcal{D}, F \notin \mathcal{D} \\ b^{HF} = \frac{\sum_{I\notin D}\alpha^I}{\sum_{J\in D}\alpha^J}\alpha^H, & \forall H \in \mathcal{D}, F \notin \mathcal{D} \end{cases} \tag{1.41}$$

and the appropriate country-specific parameters g^H to assure all equalities in 1.40. Finally, substituting the solution parameters in (1.36), (1.37), (1.41) into (1.21) we obtain the equilibrium consumption allocations (1.22), (1.23) for emerging and developed economies, respectively.

References

Bansal, R., M. Dahlquist (2000), "The forward premium puzzle: different tales from developed and emerging economies," *Journal of International Economics*, 51, 115-144.

Burnside C., M. Eichenbaum and S. Rebelo (2008), "Carry trade: The gains of diversification," *Journal of the European Economic Association*, 6, 581-588.

Burnside C., M. Eichenbaum, I. Kleshchelski and S. Rebelo (2011), "Do Peso Problems Explain the Returns to the Carry Trade," *Review of Financial Studies*, 24, 853-891.

Backus, D., G. Smith (1993), "Consumption and real exchange rates in dynamic economies with non-traded goods", *Journal of International Economics*, 35, 297-316.

Brandt, M., J. Cochrane and P. Santa-Clara (2006), "International risk sharing is better than you think, or exchange rates are too smooth," *Journal of Monetary Economics*, 53, 671-698.

Campbell, J., K. Mederios and L. Viceira (2010), "Global currency hedging," *Journal of Finance*, 65, 581-598.

Colacito, R. and M. Croce (2011), "Risks for the Long-Run and the Real Exchange Rate," *Journal of Political Economy*, 119, 153-181.

Cox, J. C. and C.-F. Huang (1989), "Optimal consumption and portfolio policies when asset prices follow a diffusion process," *Journal of Economic Theory*, 49, 3383.

Cvitanic, J. and I. Karatzas (1992), "Convex duality in constrained portfolio optimization," *Annals of Applied Probability*, 2, 767818.

Dumas, B. (1992), "Dynamic Equilibrium and the Real Exchange Rate in a Spatially Separated World," *Review of Financial Studies*, 5, 15380.

Fama, E. (1984), "Forward and spot exchange rates," *Journal of Monetary Economics*, 14, 319338.

Fama, E. and K. French (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 356.

Hassan, T. (2010), "Country Size, Currency Unions, and International Asset Returns," working paper, Booth School of Business, University of Chicago.

Hansen, L. P. and R. J. Hodrick (1980), "Forward exchange rates as optimal predictors of future spot rates: An econometric analysis," *Journal of Political Economy*, 88, 829853.

Hodrick, R., and E.P. Prescott (1997), "Postwar Business Cycles: An Empirical Investigation." *Journal of Money, Credit, and Banking*, 29, 116

Lintner, J. (1965), "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets," *Review of Economics and Statistics*, 47, 13-37.

Lustig, H. and A. Verdelhan (2007), "The Cross-section of Foreign Currency Risk Premia and US Consumption Growth Risk," *American Economic Review*, 97, 89117.

Lustig, H., Rousanov, N., and A. Verdelhan (2011), "Common Risk Factors in Currency Markets," *Review of Financial Studies*, forthcoming.

Merton, R. (1973), "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867-887.

Mehra, R. and E. Prescott (2008), Editors, *Handbook of the Equity Risk Premium* (Elsevier B.V.).

Menkhoff, L., L. Sarno, M. Schmeling and A. Schrimpf (2011), "Carry Trades and Global Foreign Exchange Volatility," *Journal of Finance*, forthcoming.

Mossin, J. (1966), "Equilibrium in a Capital Asset Market," *Econometrica*, 34, 768783.

Negishi, Takashi (1960), "Welfare Economics and Existence of an Equilibrium for a Competitive Economy," *Metroeconomica*, 12, 92-97.

Obstfeld, M. and K. Rogoff (2001), "The six major puzzles in international macroeconomics:

Is there a common cause?" in B. Bernanke and K. Rogoff, eds.: 2000 National Bureau of Economic Research Macroeconomics Annual (MIT Press, Cambridge, MA).

Pavlova, A. and R. Rigobon (2008), "The Role of Portfolio Constraints in the International Propagation of Shocks," *Review of Economic Studies*, 75, 121556.

Pavlova, A. and R. Rigobon (2010), "International Macro-Finance," *survey working paper*, LBS and MIT.

Ravn, M., and U. Harald (2002), "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations," *The Review of Economics and Statistics*, 84, 371375.

Ross, S. (1976), "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13, 341360.

Samuelson, P. (1954), "The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments," *Economic Journal*, 64, 264-289.

Sercu, P., R. Uppal and C. Van Hulle (1995), "The Exchange Rate in the Presence of Transaction Costs: Implications for Tests of Purchasing Power Parity," *Journal of Finance*, 50, 1309-1319.

Sharpe, W. (1964), "Capital asset prices: A theory of market equilibrium under conditions of risk," *Journal of Finance*, 19, 425-442.

Stathopoulos, A. (2011), "Asset Prices and Risk Sharing in Open Economies," *working paper*, Marshall School of Business, USC.

Stockman, A. and H. Dellas (1989), "International portfolio nondiversification and exchange rate variability," *Journal of International Economics*, 26, 271-289.

Stockman, A., and L. Tesar (1995), "Tastes and technology in a two-country model of the business cycle: Explaining international comovements," *American Economic Review*, 85, 168-185.

Stulz, R. (1987), "An equilibrium model of exchange rate determination and asset pricing with non-traded goods and imperfect information," *Journal of Political Economy*, 95, 1024-

1040.

Tesar, L. (1993), "International risk sharing and nontraded goods," *Journal of International Economics*, 35, 6989.

Tian, M. (2011) "Tradability of Output, Business Cycles, and Asset Prices," *Job market paper*, Sloan School of Management, MIT.

Weil, P. (1994), "Nontraded assets and the CAPM," *European Economic Review*, 38, 913-922.

Zapatero, F. (1995), "Equilibrium Asset Prices and Exchange Rates," *Journal of Economic Dynamics and Control*, 19, 787-811.

Chapter 2

The Behavior of Savings and Asset Prices When Preferences and Beliefs are Heterogeneous

(in collaboration with Richard Zeckhauser)

2.1 Abstract

This paper establishes new asset pricing results when agents differ in risk preference, time preference and/or expectations. It shows that *risk tolerance* is a critical concept driving savings decisions, consumption allocations, prices and return volatilities. Surprisingly, due to the equilibrium risk sharing, the precautionary savings motive in the aggregate can vastly exceed that of even the most prudent actual agent in the economy. Consequently, a low real interest rate, resulting from large aggregate savings, can prevail with reasonable risk aversions for all agents. One downside of a large aggregate savings motive is that savings rates become extremely sensitive to output fluctuation. Thus, the same mechanism that produces *realistically low* interest rates tends to make them *unrealistically volatile.* A powerful isomorphism allows differences in time preference and expectations to be swept away in the

analysis, yielding an equivalent economy whose agents differ merely in risk aversion. These results hold great potential to simplify the analysis of heterogeneous-agent economies, as we demonstrate in quantifying how asset prices move and bounding their volatilities. All results are obtained in closed form for any number of agents possessing additively separable preferences in an endowment economy.

2.2 Introduction

The genius of the market is its ability to transform the holdings of agents with heterogeneous preferences and endowments into outcomes that are superior for all. When time and subjective beliefs enter the picture, agents' claims shift across time and state in patterns that reflect both aggregate shocks and their beliefs, and time and risk preferences. Aggregate measures in the economy, such as interest rates and saving rates, reflect the outcome of agents who trade within such dynamic market processes.

We assume, as is common in the consumption-based equilibrium asset pricing literature, that agents start with birthright endowments of a risky asset, i.e., the contingent claim on its stochastic dividend stream. The dividend is interchangeably referred to as endowment, output or supply hereafter. In addition there is a riskless asset created by the agents of zero net supply. The price of the risky asset and the interest rate are determined by the supply and demand of the market participants. Those participants possess additively separable utility functions. As the world unfolds, they allocate their available funds - asset values plus asset returns - among consumption and holdings of the two types of assets so as to maximize their discounted expected utility. Thus agents continually shift their portfolios as asset prices rise and fall in response to the economy (endowment). Such shifting would not take place if agents held identical preferences. Note agents are better off in this heterogeneous world. They could mimic a homogeneous world by just refusing to trade.

Our attention to heterogeneity in preferences is intended to capture real world richness, and to study the evolving patterns when diverse agents interact. Most prior analyses have eschewed heterogeneity, thereby sacrificing relevance to escape the technical intractability that normally accompanies attempts to allow for significant agent differences. We were able to define a new but straightforward construct that characterizes the dynamic contribution of individual agents to the demand for assets, and also identifies how current asset returns influence agents' optimal allocations.

We build on our analysis of differences in preferences to examine how disparate subjective beliefs about the economy's uncertain fundamentals also affect outcomes. Whatever the sources of differences, the risk-averse agents share the unavoidably variable aggregate output in a manner that smooths out their personal consumptions. Naturally, more risk averse and impatient consumers respectively get smoother and earlier consumption, but they get less and ultimately much less later consumption.

All of our results are obtained in closed form. We show that all aggregate quantities of interest can be expressed as functions of agents' equilibrium consumptions, which in turn respond to those aggregates. Agents whose consumptions are most sensitive to shocks, not surprisingly, contribute predominantly to influence the behavior of the economy as output fluctuates.

The risk tolerance measure that we advocate in the current paper captures this intuition of risk-sharing mechanism. It is defined as individual i's marginal propensity $\left\{\frac{\partial c^i}{\partial w}\right\}_i$ to consume c^i out of the aggregate endowment w. It is proportional to individual risk tolerance, and shows that more risk tolerant agents embrace more volatile consumption paths (i.e., larger response of $\frac{\partial c^i}{\partial w}$ to an output shock) in return for greater shares of the endowment when times are good. It proves both convenient and reassuring that the economy's implied aggregate (i.e., market-revealed) behavior toward uncertainty, such as the risk premium and precautionary savings behavior inferred from the market prices, and the volatilities of its bond and stock returns can be readily expressed in terms of means and variances under this measure. For this reason, throughout this paper we will interchangeably refer to these aggregate behaviors as market-revealed, and market-equivalent characteristics of a fictitious *equivalent* single individual representing the entire body of agents. This aggregation is feasible in our complete-market economy.

In the special case of heterogeneous CRRA agents, it is well known that aggregate risk aversion decreases with aggregate consumption. Similarly, given that more risk-tolerant agents invest relatively more in the risky stock, a positive shock boosts their relative position in the economy, thereby making them more influential. Observed risk tolerance thus increases in good times, and vice versa, due to ownership shifts.

Our risk tolerance measure makes available many parallel and intuitive results for the economy as a whole on time preference, precautionary savings motive, and the response of aggregate savings to aggregate shocks. A simple decomposition identity illuminates the way. Market-revealed (aggregate) risk aversion is a weighted average of the individual risk aversion in risk tolerance measure, implying that its response to shocks is merely the average of individuals' responses plus the response of the risk tolerance measure itself to such movements. This latter term arises from the equilibrium risk sharing among agents, and is responsible for many noteworthy effects in the aggregation dynamics presented below. If, as is usually assumed, there is a long-term upward drift in endowments, risk tolerance, despite bouncing around with output shocks, will drift upwards as well.

Like risk aversion, the market-revealed time discount factor is the weighted average value of individual counterparts in the risk tolerance measure. As time rolls forward, more patient agents - who have smaller discount factors - are more willing to defer consumption. Assets shift to their hands, which drives down the aggregate discount rate. This phenomenon exerts downward pressure over time on market-revealed time preference in the economy. Of course, the interaction with aggregate shocks and risk preference can amplify or dampen the pressure.[1] Our decomposition identity yields simple expressions for how the discount rate moves with time and supply shocks.

Our story is a story of risk sharing and wealth re-distribution as uncertainties resolve and time passes. Surprisingly, these shifts allow market-revealed characteristics for the equivalent

[1]If more patient agents are more (less) risk tolerant, positive shocks will amplify (dampen) the pressure, and vice versa for negative shocks.

agent to lie outside the range of values held by the agents in the economy. That is, if one were to posit that the observed outcome came from a population of homogeneous agents, the hypothetical representative agent could have values for his preferences or actions that lay beyond those for any agent in the true economy of heterogeneous agents.

Precautionary savings illustrate. The equivalent agent may have stronger savings motive than would even the most prudent actual agent in the heterogeneous world. The explanation is straightforward. Agents facing stochastic output save for a rainy day. A world of heterogeneous agents injects an additional layer of dynamic uncertainty in the economy, since the standings of individuals in the economy change stochastically. This additional dynamic behaves as if it raises the demand for precautionary savings. Thus, we point out that, in heterogeneous-agent economies, the large market-revealed precautionary savings motive is not necessarily associated with the dominance of the precautious agents. Rather, the savings motive is *high* when risk-sharing dynamic between agents is *important*, i.e., when agents are sufficiently different in their beliefs, or in risk and time preferences. To illustrate, the risk sharing can push up market-revealed precautionary savings motive even when the mean value of risk aversion in the economy drops. It is well known that precautionary savings powerfully push up bond values and lower interest rates. Then it is possible and natural that the interest rate moves in the same direction with the economy's average risk aversion when agents differ in their characteristics.

In a heterogeneous and temperate[2] world, savings and savings motives are also highly sensitive to endowment fluctuations: they increase when economic prospects dim and endowments shrink. This phenomenon is consistent with the observed extraordinarily low real interest rates observed in most developed economies in the period following the 2008 meltdown. Aggressive monetary policy surely contributed, but savings had also skyrocketed due to precautionary concerns. Another remarkable implication is that when interest rates are low, they tend to be unstable in the current general additive utilities setting. This is precisely because, as discussed above, the large savings motives responsible for low interest

[2]Temperance is a determinant of portfolio choices. It is proportional to the fourth derivative of the utility function. We will characterize this behavior under uncertainty more precisely in a later section.

rate is induced by substantial level of risk sharing and hence is highly sensitive to economic fluctuations. In other words, large savings imply large savings cyclicalities in the models. We establish an analytical and almost universal lower bounds for interest rate volatilities. Within the additive utility framework, our investigation thus uncovers, both qualitatively and quantitatively, the insightful role of savings cyclicality in the long-standing risk-free rate and equity premium puzzles of macroeconomics and finance. In retrospect, it also explains why promising models addressing these puzzles in the literature need to adopt either features beyond additive utility (e.g., habit formation, recursivity) or richer time-series properties for aggregate supply and consumption.

Furthermore, our results on the dynamics of risk aversion and precautionary savings, and their consequences for the movement of savings with the economy, have significant implications for determining the direction and magnitude of volatilities in stock returns. The underlying logic is clear: saving decisions reflect portfolio choices, which are intimately related to the volatility of all asset prices, which in turn are influenced by the sloshing of assets among different classes of agents. This savings dynamic (more specifically, the savings sensitivity to economic fluctuations) plays no role in simple and popular models of the economy that employ a representative agent or two classes of agents holding power utility functions. The critical role of the cyclicality of savings gets obscured in such models. In our models, with a plethora of heterogeneous agents, the cyclicality of savings stands out for its influential role quite beyond risk aversion and precautionary savings. The extent of heterogeneity, i.e., how greatly agents differ, turns out to be critical.

In any market-exchange economy, prices are determined by both the growth rate and volatility of output (endowments in our models), and by the participants' tastes for risk and tradeoff across time, as well as their beliefs. As far as consumption and risk sharing are concerned, our formulation identifies a simple tradeoff between these two key, but seemingly quite different factors. That is because an interesting duality emerges. An economy whose agents differ on time and risk preferences is isomorphic to another economy whose agents differ merely on risk aversion, though the evolution of the endowment in the second economy will differ from what it is in the first. The isomorphism means that consumption partitions, risk sharing between agents, and market-revealed characteristics are identical in the two economies.

This isomorphism potentially enhances our ability to study economies where agents differ on multiple dimensions. First, the seemingly complex dynamic interactions of market participants in an economy with heterogeneous agents are reduced to those of simpler economy but with a modified output process. In particular, there proves to be an intimate connection between this heterogeneity reduction and the market's "natural" selection (that is, the survival) of agents in the economy. Second, employing this isomorphism may immediately pin down the direction in which additional classes of heterogeneity or expanded heterogeneity (e.g., a mean preserving spread) within an existing class will affect the volatility of asset returns. If the modified volatility of the isomorphic economy's output is lower than that of the original economy, that implies that the expansion in heterogeneity in the original economy tend to shrink the volatility in asset returns. This is simply because the volatility of asset prices increases with output volatility in the first order. The powerful implication of this result is that should endowments change, our bounds on asset return volatilities can be immediately adapted from a world where there are mere differences in risk aversion to one where differences in time preference pile atop those. Our later analysis also allows individuals to differ in their beliefs on how endowments will evolve, what might be thought of as their levels of optimism. Moreover, the isomorphism extends. That is, we can add differences in beliefs to those of time preference and risk aversion, and still find another equivalent economy whose agents differ merely in risk aversion. In other words, the disparities in time preference and optimism can be rotated away by a transformation in the evolution of the output process. Market-revealed characteristics toward risk taking and savings will be identical in the two economies.

The paper is structured as follows. Section 2.3.1 reports briefly the empirical statistical moments (means and volatilities) of interest rates and equity market returns, which have been extensively documented in literature. We also discuss recent estimates of distributions of risk aversion and time preference in the population. Not surprisingly, these show substantial degrees of heterogeneity among individuals. Section 2.3.2 positions our work and findings with respect to the related literature. Section 2.4 derives various equivalent forms of the risk tolerance measure and discusses their merits in the aggregation analysis of the economy with heterogeneous agents. Section 2.5 analyzes the effect of savings behaviors on interest rate volatility and identifies substantial lower bounds given the premise of large savings. Section 2.6 carries out similar analysis on equity return volatilities and derives a sufficient condition for excess equity return volatilities, as long observed in data. Section 2.7 shows and analyzes the equivalence between the effect of heterogeneities in time preferences and beliefs, and an appropriate modification in the output statistics. Section 2.8 concludes. All proofs and derivations are given in the appendices.

2.3 Empirical facts and related literature

This section provides factual material to motivate our study of the linkage between risk sharing and equilibrium asset prices given heterogeneous preferences. First, we recount the observed behaviors of returns on key asset (risk-free bond and stocks). Next, we provide recent evidence from literature surveys showing sizable heterogeneity of market participants' preferences. Models employing homogeneous agents do not capture the richness of the world in which we live. Finally, we discuss the literature most relevant to the current work.

2.3.1 Estimates of asset returns' moments and preferences

Returns on equities and risk-free assets are among the most documented quantities in the empirical finance literature. The behaviors of these returns expose stylized facts that can be "puzzling" from the consumption-based asset pricing perspective.

Table 2.1: Consumption growth, and real return on equity and short-term risk-free debt (annual %): recent history

Quantities[a]		Japan (1970.2-1999.1)	UK (1970.1-1999.2)	US (1970.1-1998.4)
consumption growth	mean	3.20	2.20	1.81
	stddev	2.56	2.51	0.91
real return on equity	mean	4.72	8.16	6.93
	stddev	21.91	21.19	17.56
real return on bills	mean	1.39	1.30	1.49
	stddev	2.30	2.96	1.69
Equity premium		3.33	6.86	5.44

[a]*Source: Campbell (2003)*

Risk-free rate and return on equity

Table 2.1 reports the recent historical means and standard deviations of aggregate consumption growth, returns on equity and short-term risk-free assets (bills), for Japan, UK and US. All returns are real and in annualized percentage values. For further illustration, table 2.2

Table 2.2: Equity premia (annual %): long history

		Japan (1900-2005)	UK (1900-2005)	US (1900-2005)
Equity premium[a]	mean	9.84	6.14	7.41
	stddev	27.82	19.84	19.64

[a]*Sources: Dimson, Marsh and Staunton (2008)*

also reports long historical equity risk premia for these countries. In all three countries, for both recent and long histories, real risk-free rates are both low and stable, compared to much higher and more volatile returns on equities. This is the risk-free rate puzzle (Weil (1989)). Similarly, equity premia are also large and volatile vis-a-vis low and stable aggregate consumption growth.[3] This is the closely related equity premium puzzle (Mehra and Prescott (1985)).

[3]Dividend growths are also much less volatile than returns on equities.

Heterogeneity in risk and time preferences

Our analysis includes heterogeneity in both risk and time preferences, thus it is important to determine whether there is heterogeneity in such dimensions in the real world. Table

Table 2.3: Heterogeneity in Individuals' relative risk aversion R

Country	Method	RRA R	Standard deviation
US[a]	Surveys	12.07	16.58
US[b]	Surveys	8.2	6.8
Norway[c]	Surveys	3.92	2.94
US[d]	Actual financial decisions	2.85	3.62

[a]*Sources: Barsky, Juster, Kimball and Shapiro (1997)*
[b]*Kimball, Sahm and Shapiro (2008)*
[c]*Aarbu and Schroyen (2009)*
[d]*Paravisini, Rappoport and Ravina (2010)*

2.3 reports the results of some recent studies on the distribution of individuals' relative risk aversion, $R = \frac{-c\partial^2 U/\partial c^2}{\partial U/\partial c}$, which have been conducted on the US and Norway populations. The first three estimates are obtained from responses to different surveys, over different periods. The surveys employed various forms of hypothetical gambles. The last estimate is inferred from *actual* financial decisions of investors in an online person-to-person lending platform. Readers should consult the original sources for details. Clearly, all four studies show substantial heterogeneity in the level of relative risk aversion reported by either survey respondents or actual investors. Table 2.4 reports estimates for the distribution of individuals' discount factor $\delta = -\frac{1}{U}\frac{\partial U}{\partial t}$ Both studies found differences in time preference reported by the respondents.

The sizable dispersions in preferences found in these studies motivate our current study of the impacts of heterogeneity on equilibrium asset prices.

Table 2.4: Heterogeneity in individuals' time discount rate δ (annual %): Estimates from surveys

Country	Method	Number of observation	Mean disc. rate δ	Standard deviation
US[a]	Surveys	138	10.6	16.58
US[b]	Surveys	> 8000	7.5	2.4

[a]*Sources: Chesson and Viscusi (2000)*
[b]*Alan and Browning (2010)*

2.3.2 Related literature

Our paper is most closely related to heterogeneous-agent equilibrium models addressing price anomalies in financial economics literature. The interest on price puzzles has skyrocketed since the seminal papers by Mehra and Prescott (1985) and Weil (1989). Mehra and Prescott (2008)'s dedicated handbook offers the most extensive single source of up-to-date references on this important and vibrant topic. The current paper does not attempt to provide new solutions; it instead contributes to a deeper understanding about the nature of risk-free rate and equity premium behaviors within the *classic* additive utility setting, a setting in which these phenomena are most puzzling. First and conceptually, we shed *new* light on the crucial role of the *cyclicality* of precautionary savings in shaping equity and bond return dynamics. Second and analytically, we identify substantial lower bounds on interest rate volatility when interest rates are desirably low. Together, these demonstrate the hard-to-reconcile nature of low and smooth interest rates observed in real-world economies.

In the finance literature, the heterogeneous-agent formulation appeared early on in Benninga and Mayshar (2000), Dumas (1989), Wang (1996) and others, where agents differ in their risk aversions. Heterogeneity in market participants' characteristics has evolved into an attractive topic of active research, which now also incorporates differences in time preferences (Gollier and Zeckhauser (2005), Jouini and Napp (2007), Lengwiler (2005)), beliefs in the fundamentals (Basak (2005), Detemple and Murthy (1994)), or all of the above (Bhamra and Uppal (2010), Lengwiler et al. (2005), Sandroni (2000), Yan (2008)). Heterogeneity generates non-trivial risk sharing patterns and consequently, has rich implications for price dynamics (Bhamra and Uppal (2009), Dumas et al. (2009), Chan and Kogan (2005), Zapatero (1998)), portfolio choices and trading (Gallmeyer and Hollifield (2008), Longstaff and Wang (2008)), and market selection (Blume and Easley (2006), Kogan et al. (2006) and (2009)). In contrast with these works, our paper points out an intuitive tradeoff between agent-based heterogeneities and macroeconomic conditions, which is helpful in analyzing agents' equilibrium interaction and the resulting price dynamics mentioned above.

The degree of heterogeneity in the economy is plausibly the key determinant of the magnitude of heterogeneity's impact. In particular, Chen, Joslin and Tran (2010) study the impact of heterogeneous beliefs in the likelihood and severity of rare events (e.g., crises, disasters and alike) on asset prices. They point out that the risk premium in the economy may drop even when the average level of pessimism among agents surges. This is because there, the driving force is the dynamic dispersion of beliefs and the associated risk sharing, but not just the mean value of the belief distribution. By showing that subject to sufficient heterogeneity in risk aversion in the economy, the equilibrium interest rate may even increase when the average level of precautionary savings motives among agents surges, the current study complements their results in identifying another setting where the risk sharing induced by heterogeneity yields spectacular effects.

2.4 Risk tolerance measure and aggregation

In any economy, be it one of homogeneous or heterogeneous agents, risk taking and savings are determined by the behavior of individual agents. In a heterogeneous world, the dynamic competitive interactions among such agents play a major role in determining aggregate outcomes. To address the interactions that are determined by risk taking propensities, and the ultimate consequences for various aggregates, the concept of risk tolerance proves to be both extremely powerful and convenient. It precisely measures how agents' consumptions move with changes in the aggregate endowment. This section uses risk tolerance measures to derive key market-revealed quantities, including risk aversion, time preference and precautionary savings. The approach neatly separate the contributions of agents' characteristics from their interactions. Many interesting aggregate behaviors of the economy, some known others new, then can be readily elucidated.

2.4.1 The setting

To develop intuitive results on aggregation, we first investigate a general endowment economy with many classes of agents. Within each class, agents have identical preferences,[4] but across classes agent risk aversions and time preferences differ. Throughout the paper, the superscript i denotes quantities associated with agent i. Agents maximize their general time-separable utilities, which are increasing, concave and three-time continuously differentiable. Agent i's relative risk aversion (RRA) $R^i(t, c^i)$ and subjective discount factor $\delta^i(t, c^i)$ generally can be functions of consumption c^i and time t. Alternatively, we will also study the canonical settings with power utilities to make precise the model's key results. For that case, agents' RRAs are constant and simply denoted γ^i, instead of $R^i(t, c^i)$ reserved for more general (non-CRRA) settings. At the outset, each agent i is endowed with a fraction $\theta^i_S(0)$ of a risky stock paying a stochastic dividend stream $w(t)$. The dividend, which reflects the state of the economy, follows a geometric Brownian process (GBM)

$$\frac{dw(t)}{w(t)} = \mu^w dt + \sigma^w dZ(t) \Rightarrow w(t) = w(0)e^{(\mu^w-(\sigma^w)^2/2)t}e^{\sigma^w Z(t)}. \tag{2.1}$$

When $(\mu^w - (\sigma^w)^2/2) > 0$ the economy is growing in the long term ($\lim_{t\to\infty} E_0[w(t)/w(0)] \to \infty$ a.s.). A single share of the risky stock is available in the economy for agents to trade. In addition, there is a zero net supply of a riskless asset (money market account, also loosely referred to as bond below) created by the agents. Agents trade these two assets and choose

[4] For this reason, to simplify notation, hereafter we simply use agent (being representative of her own homogeneous class) in place of class (of identical agents).

consumption levels to maximize their expected utilities subject to a budget constraint[5] and market clearing

$$
\begin{aligned}
&\max_{\{c^i,\theta^i\}} E_0 \int_0^T u^i(c^i(t),t)dt, \\
&\text{s.t. } c^i(t)dt = \theta_S^i(t)[w(t)dt + dS(t)] + \theta_B^i(t)B(t)r(t)dt - dw^i(t), \qquad (2.2)\\
&\text{and } \sum_i \theta_S^i(t) = 1; \qquad \sum_i \theta_B^i(t) = 0 \quad \forall t,
\end{aligned}
$$

where $S(t)$, $B(t) = \exp\left(\int_0^t r dt\right)$ and $w^i = \theta_B^i(t)B(t) + \theta_S^i(t)S(t)$ respectively denote stock price, bond price and wealth processes.[6] Since the market is complete, there exists a set of positive constant utility weights $\{\lambda^i\}$ such that the above optimal individual consumption plans also solve the equivalent-agent optimization (see Negishi (1960))

$$
V^\lambda(\{w\}) \equiv \max_{\{c^i\}} E_0 \sum_i \frac{1}{\lambda^i} \int_0^T u^i(c^i(t),t)dt \quad \text{s.t.} \quad \sum_i c^i(t) = w(t) \ \ \forall t. \qquad (2.3)
$$

As the aggregate constraint holds at all time and states, the optimization problem (2.3) can be equivalently cast in a static formulation at each time and state (Karatzas et al. (1987), Cox and Huang (1989))

$$
v^\lambda(w(t),t) \equiv \max_{\{c^i\}} \sum_i \frac{1}{\lambda^i} u^i(c^i(t),t) \quad \text{s.t.} \quad \sum_i c^i(t) = w(t). \qquad (2.4)
$$

[5]Aggregating the budget constraint (2.2) over all agents we obtain $\sum_i dw^i(t) = dS(t)$, i.e., the total change in agents' wealths equals the change in value of the single share of stock, which is the net asset of the economy.

[6]Given the infinite time horizon $T \to \infty$, Lengwiler, Malamud and Trubowitz (2005) shows that this economy's necessary and sufficient condition for equilibrium existence is precisely the boundedness of every agent's expected utility of aggregate endowment

$$
E_0 \left[\int_0^\infty u^i(w(t),t)dt \right] < \infty \qquad \forall i.
$$

Note that this condition also assures that the stock price is finite.

Combining the first order equations with the envelope theorem we obtain the following system of equations satisfied by optimal consumption plans

$$\frac{1}{\lambda^i} u_c^i(c^i(t), t) = v_w(w(t), t) \qquad \forall i, \tag{2.5}$$

Throughout the paper, subscripts denote partial derivatives. Thus, $f_x(x, y) \equiv \frac{\partial f(x,y)}{\partial x}$.

2.4.2 Risk tolerance measure

In the economics of uncertainty, the ways agents optimally allocate their consumptions across states and time are determined respectively by their relative risk aversion (RRA) and pure time preference (a.k.a. subjective discount factor). It is convenient to adopt these standard characteristics for an equivalent agent of the aggregate economy. Given a complete market, these characteristics are revealed unambiguously from observed prices, and are attributed to this equivalent agent as if there were only one class of agents in the economy. For this reason, hereafter R, δ and T are respectively referred to as risk aversion, discount factor and risk tolerance of the *market-revealed equivalent* agent (hereafter, *equivalent agent*).

$$\begin{aligned}
R^i(c^i, t) &\equiv \frac{-c^i u_{cc}^i(c^i, t)}{u_c^i(c^i, t)} \quad \longleftrightarrow \quad R(w, t) \equiv \frac{-w v_{ww}(w, t)}{v_w(w, t)}, \\
\delta^i(c^i, t) &\equiv \frac{-u_{ct}^i(c^i, t)}{u_c^i(c^i, t)} \quad \longleftrightarrow \quad \delta(w, t) \equiv \frac{-v_{wt}(w, t)}{v_w(w, t)}, \\
T^i(c^i, t) &\equiv \frac{-u_c^i(c^i, t)}{u_{cc}^i(c^i, t)} \quad \longleftrightarrow \quad T(w, t) \equiv \frac{-v_w(w, t)}{v_{ww}(w, t)}.
\end{aligned} \tag{2.6}$$

The apparent analogy of these market-revealed characteristics with those of single-agent economy aims to capture the whole economy's attitudes, such as discount factor δ, risk aversion R and utility function $v(w)$, as of a single equivalent (representative) agent's. In particular, in the aggregate the above definitions implies $T = \frac{w}{R}$, a relation that also holds at individual level.

Following Wilson (1968), there exists a simple aggregation relation on risk tolerance (see also proposition 4)

$$T(w,t) = \sum^{i} T^i(c^i, t).$$

which motivates the choice of the *risk tolerance measure* $\{p^i\}$ as micro-economic building blocks of all these market-revealed characteristics

$$p^i(c^i(w,t),t) \equiv \frac{T^i(c^i,t)}{T(w,t)} = \frac{T^i(c^i,t)}{\sum^i T^i(w,t)} \Rightarrow \sum_i p^i = 1.$$

This implied normalization together with $p^i \in [0,1]$, which holds when all agents are risk averse ($T^i > 0 \ \ \forall i$), qualify $\{p^i\}$ as a standard measure.

This measure is formulated to precisely capture a key concept that risk tolerant agents play predominant role in consumption and wealth distribution dynamics. To see this point, we note the following very interesting and intuitive relation

$$p^i(c^i(w,t),t) = c^i_w(w,t), \tag{2.7}$$

This identity shows that risk tolerance measure exactly characterizes the individual optimal consumption responses to an aggregate endowment shock. In equilibrium, more risk-tolerant (i.e., larger $\frac{T^i}{T}$) agents embrace relatively less smooth consumption paths (i.e., larger c^i_w), and necessarily contribute more to economy's reactions to output fluctuations. In comparison, we note that neither the least risk averse agent ($\min\{R^i\}$) nor the one who consumes most ($\max\left\{\frac{c^i}{w}\right\}$) invariably put up strongest response to the aggregate shocks. This signifies the unique role of risk tolerance measure in determining the risk sharing and consumption partition among agents. As agents save and trade accordingly to realize their optimal consumption plan, asset prices and their volatilities necessarily are contingent on this measure. Establishing this link more quantitatively is a central theme of our subsequent analysis.

Being functions of equilibrium consumptions, $\{p^i(c^i,t)\}$ entirely capture both aggregate fluctuation effects and the dynamics of the competitive interaction between agents. The mere fact that $p^i \geq 0 \ \ \forall i$ (when all agent are risk averse) immediately implies a known and important result that no agents cut their optimal consumption when the aggregate endowment increases, $dw > 0$. Furthermore, agents whose optimal consumptions respond most strongly to an aggregate endowment shock will dominate in this measure,[7] as the following concise result implies.

Proposition 4 *The equivalent RRA, discount factor and risk tolerance of the entire economy are related to their single-agent counterparts as follows*

$$\begin{aligned} R(w,t) &= \sum_i \frac{T^i(c^i,t)}{T(w,t)} R^i(c^i,t) = E_{\{p^i\}}[R^i], \\ \delta(w,t) &= \sum_i \frac{T^i(c^i,t)}{T(w,t)} \delta^i(c^i,t) = E_{\{p^i\}}[\delta^i], \\ T(w,t) &= \sum_i T^i(c^i,t), \end{aligned} \tag{2.8}$$

where $E_{\{p^i\}}[\ldots]$ denotes the expectation under risk tolerance measure $\{p^i = \frac{T^i}{T}\}$. This result generalizes the time preference aggregation obtained in Gollier and Zeckhauser (2005) to stochastic settings. (See also Lengwiler, Malamud and Trubowitz (2005) for a discrete-time formulation of the results). Both market-revealed RRA and discount factor are expressed succinctly as averages in risk tolerance measure.[8] These representations elucidate many important properties of this economy. Indeed, (2.8) indicates $R, T > 0$, and then $v_w > 0$, $v_{ww} < 0$ respectively by virtue of eqs. (2.5), (2.6), guaranteeing the desired risk-averse and increasing utility for the equivalent agent.

In the stochastic and complete market, agents perfectly share their risks by taking stochastic positions in both stock and bonds. The optimal consumption plans thus are necessarily stochastic, and so are their risk tolerance measures (also referred to as weights), $p^i = c^i_w$. The resulting equivalent preference characteristics e.g., R, δ, are stochastic, not necessarily because their agent-based counterparts e.g., R^i, δ^i are stochastic, but rather because their dynamics weights $\{p^i\}$ bounce stochastically. Indeed, in a CRRA utilities setting, the

[7] The most widely-used heterogeneity measure in literature is consumption share $\left\{\frac{c^i(w,t)}{w(t)}\right\}$, which is less expressive with respect to the rich dynamics of equilibrium consumption's *changes* under supply shocks.

[8] That is, weighted averages, with weights being the risk tolerance measures $\frac{T^i}{T}$.

individual R^i, δ^i are constant, yet R, δ in (2.8) are not so, obviously. To understand those dynamics more precisely, it is best to see how the risk tolerance measure changes under aggregate supply shocks

$$\frac{dp^i(w,t)}{dw} = c^i_{ww}(w,t) = \frac{T^i(c^i,t)}{(T(w,t))^2}\left(T^i_c(c^i,t) - T_w(w,t)\right). \tag{2.9}$$

which simplifies in the CRRA utilities setting to

$$c^i_{ww} = \frac{T^i(c^i,t)}{(T(w,t))^2}\left(\frac{1}{\gamma^i} - E_{\{p^i\}}\left[\frac{1}{\gamma^i}\right]\right); \qquad \frac{c^i_{ww}}{c^i_w} = \frac{1}{T(w,t)}\left(\frac{1}{\gamma^i} - E_{\{p^i\}}\left[\frac{1}{\gamma^i}\right]\right). \tag{2.10}$$

These imply that the least risk averse agent i ($\gamma^i = \gamma^{min}$) has convex consumption $c^i_{ww} > 0$ and her weight c^i_w unambiguously increases with aggregate endowment. The converse holds for the most risk averse agent (γ^{max}). In between, the transition is monotonic: percentage changes in less risk averse agents' weights c^i_w are more dramatic than those of more risk averse ones. The stochastic nature of risk tolerance measures is induced by risk sharing mechanism and has profound implications for the volatilities of *all* market-revealed characteristics, as the latter are some form of weighted averages in this measure. This observation is reflected in the following result, which provides the basis for many findings presented below.

Proposition 5 *Suppose $\{a^i\}$ are some agent-based characteristics. The response of the resulting risk-tolerance aggregate $E_{\{p^i\}}[a^i]$ to an aggregate supply shock dw can be decomposed into two components*

$$\frac{\partial E_{\{p^i\}}[a^i]}{\partial w} = E_{\{p^i\}}[a^i_w] + Cov_{\{p^i\}}\left(\frac{p^i_w}{p^i}, a^i\right). \tag{2.11}$$

Of special interest, the second component is exclusively associated with the dynamic behavior $p^i_w \equiv \frac{\partial p^i}{\partial w}$ of individual risk tolerance $p^i(w,t)$.

To a lesser degree, the first component is also related to risk-tolerance measures, because $a^i_w = a^i_c c^i_w = a^i_c p^i$. But it is primarily associated with the dependence $a^i(c^i, t)$ at the agent-specific level at the onset. The mechanism underlying this decomposition is very intuitive.

For a simple illustration, let us continue with eq. (2.10) and assume that all the a^i are constant. Dividing both sides of (2.10) by $p^i = T^i/T$ yields

$$\frac{p_w^i}{p^i} = \frac{1}{T(w,t)}\left(\frac{1}{\gamma^i} - T_w(w,t)\right).$$

Clearly, $\gamma^i < \gamma^j \Rightarrow \frac{p_w^i}{p^i} > \frac{p_w^j}{p^j}$, or percentage changes in weights p^i are greatest for agents with lesser risk aversion γ^i. This is because under a positive shock $dw > 0$ to the aggregate endowment, less risk averse agents, who invest disproportionally in the risky contingent claim on aggregate wealth (stock) become relatively better off, and contribute more to the welfare. Indeed, in this CRRA framework, (2.11) simplifies to

$$\frac{\partial E_{\{p^i\}}[a^i]}{\partial w} = \frac{1}{T(w,t)} Cov_{\{p^i\}}\left(\frac{1}{\gamma^i}, a^i\right).$$

The situations when $a^i > a^j$ for $\gamma^i > \gamma^j$ and vice versa are referred to as comonotone. Similarly, anti-comonotonicity means $a^i > a^j$ if $\gamma^i < \gamma^j$ and vice versa. To illustrate, when a^i is the discount rate δ^i, comonotone relations represent the normal case where less risk averse agents also tend to be more patient. We see that when $\{a^i\}$ and $\{\gamma^i\}$ are comonotone, the mean value $E_{\{p^i\}}[a^i]$ decreases unambiguously with aggregate endowment w. This is precisely because smaller values of a^i (associated with smaller γ^i by co-monotonicity) have relatively larger weights after a positive shock increases w as we argued above, and thus drive down the mean value. The opposite holds when $\{a^i\}$ and $\{\gamma^i\}$ are anti-comonotone; larger a^i (associated with smaller γ^i by anti-comonotonicity) have relatively larger weights after a positive shock increases w, which makes mean value $E_{\{p^i\}}[a^i]$ increases unambiguously with aggregate endowment w.

Two immediate applications concern the market-revealed risk aversion R and discount rate δ of proposition 4, specialized to the CRRA utilities setting[9]

$$R_w(w,t) = \frac{1}{T(w,t)} Cov_{\{p^i\}}\left(\gamma^i, \frac{1}{\gamma^i}\right) < 0, \quad (2.12)$$

$$\delta_w(w,t) = \frac{1}{T} Cov_{\{p^i\}}\left(\frac{1}{\gamma^i}, \delta^i\right). \quad (2.13)$$

The first equation demonstrates a well-known result of decreasing market-revealed risk aversion (see e.g., Wang (1996)). The second formalizes the wealth effect on market-revealed time preference first obtained in Gollier and Zeckhauser (2005). We recast these known and important results in connection with the risk tolerance measure to capture the key intuitions underlying this measure's dynamics.

The above market-revealed characteristics also yields the equivalent hyperbolic discounting behavior of the economy (Gollier and Zeckhauser (2005)). Taking the derivative with respect to time, $\delta_t \equiv \frac{\partial \delta}{\partial t}$, again within the CRRA setting yields

$$\delta_t(w,t) = -\sum_i p^i(w,t) \frac{(\delta(w,t) - \delta^i)^2}{\gamma^i} < 0. \quad (2.14)$$

The intuition again can be distilled from competitive interaction in equilibrium. More patient agents are more willing to defer their consumptions, and thus will increase their dominance as time rolls forward. Given that being more patient means having smaller δ^i, this competitive behavior simply decreases the weighted average discount factor $\delta(w,t)$ over time. This in turn has interesting and direct effects on the term structure of interest rates (Lengwiler(2005)).

When heterogeneities are present in both risk and time preference, either a low risk aversion or a small discount rate will lead an individual to play a greater role in the long run We will analyze quantitatively the tradeoff between these characteristics in conjunction with agents' long-run survival in section 2.7.

[9] Corresponding expression for non CRRA setting is $R_w = \frac{1}{T} Cov_{\{p^i\}}\left(\gamma^i, T_c^i\right)$, see (2.67).

2.4.3 Market-revealed precautionary savings

Prudence (see Leland (1968) and Sandmo (1970)) is a key characteristic determining precautionary savings, and thus both interest rate and returns on other assets. Kimball (1990) shows that the prudence, defined in analogy with relative risk aversion (2.6) as

$$P^i(c^i,t) \equiv \frac{-c^i u^i_{ccc}(c^i,t)}{u^i_{cc}(c^i,t)} \longleftrightarrow P(w,t) \equiv \frac{-w v_{www}(w,t)}{v_{ww}(w,t)},$$

provides an analytical measure of the intensity of the precautionary savings motive. Other factors being equal, an agent i who is more prudent (larger P^i) will save relatively more under the prospect of future income uncertainties. For a heterogeneous-agent economy with general additive utilities, we can differentiate the FOC (2.3) twice to obtain the explicit aggregation relation

$$P(w,t) = E_{\{p^i\}}[P^i(c^i,t)] - \mathrm{Cov}_{\{p^i\}}\left(R^i(c^i,t), \frac{1}{R^i(c^i,t)}\right) + \mathrm{Cov}_{\{p^i\}}\left(R^i(c^i,t), \frac{c^i R^i_c(c^i,t)}{(R^i(c^i,t))^2}\right), \tag{2.15}$$

where the moments again are defined in the risk tolerance measure. The key observation is that while market-revealed risk aversion (2.8) has value bounded within the spectrum of agents' RRA ($R^{min} \leq R(w,t) \leq R^{max}$), such bounding need not apply for market-revealed prudence $P(w,t)$. The market-revealed precautionary savings motive contains a weighted average $E_{\{p^i\}}[P^i]$ over individual agents, which plausibly results from a simple aggregation. More profoundly, it also contains additional components which arise from the dynamics of the risk sharing, and thus the risk tolerance measure itself, much in the spirit of the mechanism underlying propososition 5. To illustrate this insight, let us employ the class of power utilities, wherein (2.15) becomes

$$P(w,t) = E_{\{p^i\}}[P^i(c^i,t)] - \mathrm{Cov}_{\{p^i\}}\left(\gamma^i, \frac{1}{\gamma^i}\right) = E_{\{p^i\}}[P^i] - T\sum_i \gamma^i c^i_{ww}. \tag{2.16}$$

As $p^i = c^i_w$ defines the risk tolerance measure, c^i_{ww} clearly characterizes the dynamics of this measure under changes in aggregate endowment w. Individual agents' savings are not made independently as naive intuition about aggregation might suggest. That is because the economy's precautionary savings reflect both agents' average precautionary savings motive and the response to stochastic wealth distribution.[10] This second factor inflates the market-revealed precautionary savings motive because the term $\mathrm{Cov}_{\{p^i\}}\left(\gamma^i, \frac{1}{\gamma^i}\right)$ is invariably negative. The more risk averse agents have concave consumptions ($c^i_{ww} < 0$, see (2.10)), and they contribute positively to this induced prudence due to their larger γ^i. When agents are sufficiently different in their risk preferences, this covariance tends to be large (and negative) and it can inflate economy's savings motive greatly beyond that of even the most prudent agent in the economy. The proposition 6 and figure 2-1 below confirm this extraordinary effect stemming from risk sharing between agents.

Before turning to the main results of this section, we note that there exists another relation involving prudence $P(w,t)$, directly obtained from the definitions of R and P (derived in appendix 2.9.1)

$$R_w(w,t) = \frac{R(w,t)}{w}(1 + R(w,t) - P(w,t)). \tag{2.17}$$

This equality does not rely on any aggregation mechanism, and hence holds at both the agent and aggregate level. (2.17) implies that high market-revealed precautionary savings are related to the countercyclicality in market-revealed risk aversion. We will discuss this cyclicality and its implication for interest rate volatility in more detail in section 2.5. Many important properties related to risk sharing between agents emerge in a world with merely two classes of agents. We find it very helpful in various places to present these results in a two-agent economy.

---

[10] We may also see this quantitatively in the equivalent agent's optimization problem in a simple two-period model. The equivalent agent optimally chooses current savings X subject to initial wealth constraint W and future uncertain income $\tilde{Y}$

$$\max_X \left[v(W - X, t) + E_t v(X + \tilde{Y}, t+1)\right] =$$
$$\max_X \left\{ \max_{\sum c^i(t) = W - X;\ \sum c^i(t+1) = X + \tilde{Y}} \sum_i \frac{1}{\lambda^i}\left[u^i(c^i, t) + E_t u^i(c^i, t+1)\right] \right\}.$$

Evidently, equivalent agent's precautionary savings optimization composes of two-stage optimization over agents', subject to market clearings in each period. This subtle constraints constitute additional sensitivity of social utility to future uncertainty that equivalent agent should be wary of.

Proposition 6 *1. In the multiple-CRRA-agent economy, market-revealed precautionary savings are*

$$P(w,t) = E_{\{p^i\}}[P^i(c^i,t)] - Cov_{\{p^i\}}\left(\gamma^i, \frac{1}{\gamma^i}\right). \quad (2.18)$$

and thus is always larger than or equal to the average individual precautionary savings $E_{\{p^i\}}[P^i(c^i,t)]$ *in risk tolerance measure* $\{p^i\}$.

2. The market-revealed precautionary savings in the two-CRRA-agent economy are a concave quadratic function of p^A

$$P(w,t) = (p^A(w,t)\gamma^A + p^B(w,t)\gamma^B)\left(1 + \frac{p^A(w,t)}{\gamma^A} + \frac{p^B(w,t)}{\gamma^B}\right). \quad (2.19)$$

When individual RRA γ^A, γ^B *satisfy* $\frac{\gamma^B}{\gamma^B+1} > \gamma^A$, *there exists a region of consumption distribution between the two agents where the market-revealed precautionary savings are higher than that of either agent*

$$P^* > \max\{P^A = \gamma^A + 1; P^B = \gamma^B + 1\}.$$

To illustrate the results of proposition 6, Figure 2-1 plots the market-revealed prudence in a two-CRRA-agent economy with $\gamma^A = 0.1$ and $\gamma^B = 15$. In this case, P is a function of first agent's risk tolerance weight $p^A \equiv \frac{c^A\gamma^B}{c^A\gamma^B + c^B\gamma^A}$. Following the pattern of eq. (2.16), we decompose this aggregate into two components; the weighted average prudence and the dynamics-induced prudence. We see that the maximum market-revealed prudence $P \sim 30$ is reached at $p^A = \frac{c^A\gamma^B}{c^A\gamma^B + c^B\gamma^A} \sim 0.6$. This value far exceeds either individual prudence level, $P^A = 1.1$, $P^B = 16$. The excess stems from the risk sharing mechanism, and is quantified by the risk tolerance measure dynamic. The latter tends to zero in both homogeneous limits ($p^A = 0$ or 1) where the risk sharing possibility between the agents vanishes. Collectively, the agents may keep up this high market-revealed precautionary savings motive for an extended period of time because they differ as well in time preference.[11] We will study in detail how

[11] Yan (2008) shows that no agent dominates the others in the long run when they have similar "survival

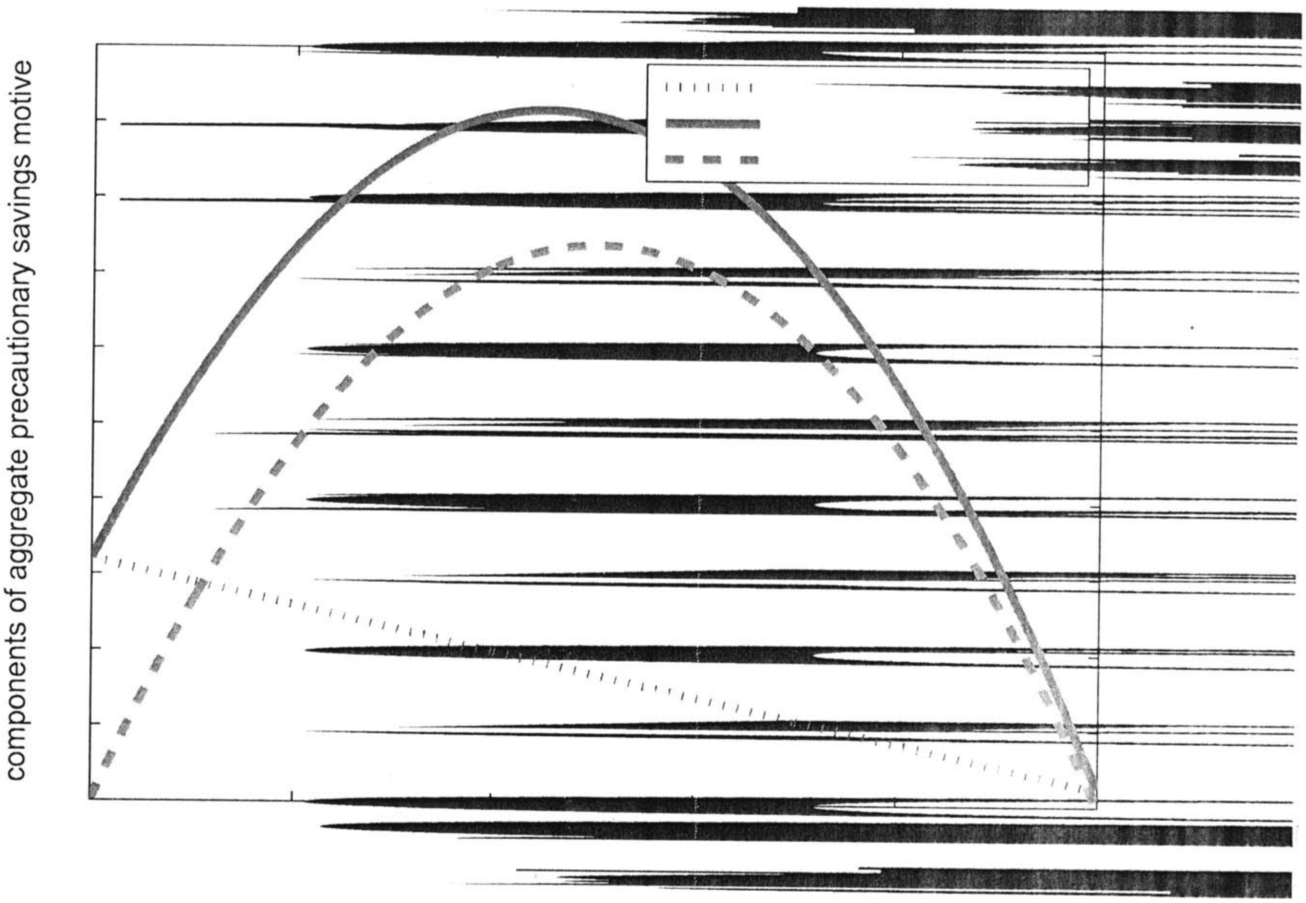


Figure 2-1: Two-CRRA-agent economy: $\gamma^A = 0.1$, $\gamma^B = 15$. Market-revealed (aggregate) prudence $P(p^A)$ and its components (2.18): weighted average (w.a.) $E_{\{p\}}[P^i]$ and dynamics-induced (d.i.) prudence $-\text{Cov}_{\{p^i\}}\left(\gamma^i, \frac{1}{\gamma^i}\right)$. These are plotted against agent A's risk tolerance weight $p^A = \frac{T^A}{T^A+T^B} = \frac{c^A\gamma^B}{c^A\gamma^B+c^B\gamma^A}$.

precautionary savings affect both the levels and volatilities of asset returns in later sections.

2.4.4 Cyclicality of market-revealed precautionary savings

We now delve deeper into the microeconomic foundations of asset pricing to see how the cyclicality of precautionary savings motive moves with consumption and wealth. This anal-

index" values $\delta^A + \gamma^A\left(\mu^w - \frac{(\sigma^w)^2}{2}\right) \approx \delta^B + \gamma^B\left(\mu^w - \frac{(\sigma^w)^2}{2}\right)$. For current parameters $\gamma^A = 0.1$, $\gamma^B = 16$, this co-survival condition holds, e.g., when subjective discount rates are $e^{-\delta^A} \approx 0.8$, $e^{-\delta^B} \approx 1$.

ysis provides rigorous grounds to study the key effects of savings cyclicality on equilibrium price behaviors in later sections.

Central to our analysis is a simple and strong relation between precautionary savings motive $P(w,t)$ and its cyclicality $P_w(w,t)$ that holds for any general time separable utility.

$$P_w(w,t) = \frac{P(w,t)}{w}(1 + P(w,t) - Q(w,t)), \tag{2.20}$$

where $Q(w,t)$ is referred to as temperance

$$Q^i(c^i,t) \equiv \frac{-c^i u^i_{cccc}(c^i,t)}{u^i_{ccc}(c^i,t)} \longleftrightarrow Q(w,t) \equiv \frac{-w v_{wwww}(w,t)}{v_{www}(w,t)}.$$

Kimball (1992, 1993) shows that in a partial equilibrium setting with multiple sources of risks, temperance affects the allocation of savings between safe and risky assets, i.e., portfolio choice. First, in light of the relation (2.20), temperance $Q(w,t)$ contributes decisively to the cyclicality of savings. This savings adjustment in turn is reflected in asset return volatilities and asset (bond and stock) holdings.[12] In the current general equilibrium settings, our observation in (2.20) thus reinforces Kimball's partial equilibrium results.

Second and more important, equation (2.20) constitutes a new and keen relation between savings and savings cyclicality in general heterogeneous-agent settings; savings behaviors tend to be more volatile when savings motives are higher! Indeed, all else being equal, the intensity of cyclicality P_w increases more than linearly with P.[13] in (2.20) This finding is somewhat unexpected since a priori savings and volatility of savings may not necessarily be tightly bound. A counter-example illustrates this point. When the representative agent conventionally has CRRA utility of the form $U(C,t) \sim \frac{C^{1-\gamma}}{1-\gamma}$, the precautionary savings motive $P = \gamma + 1$ is constant, and thus savings cyclicality is null, regardless of how big this savings motive P is. In contrast, the intuition behind our observation (2.20) highlights the risk sharing dynamics in an environment with heterogeneous agents. As we saw in the

[12] Given complete market hedging, portfolio choices are one-to-one with asset return volatilities; the position in the stock is the ratio of wealth volatility to stock price volatility.

[13] Q may also change with P. But in a setting with many agents, this dependence is rather weak.

last section, in such setting the aggregate savings motive P is high not because the most precautious agent dominates the economy. Rather, large P arises when risk sharing dynamics are important, which are possible on the premise that agents sufficiently differ in their characteristics, as illustrated by figure 2-1. Precisely because of this marked heterogeneity in agents' risk preferences, shocks to the output induce considerable amount of assets and wealth changing hands among investors. As a result, economy's savings behavior is then highly sensitive to output fluctuation.

To illustrate, we establish the aggregation relations concerning temperance, along the lines similar to our analysis of market-revealed precautionary savings. For simplicity, we consider again the power utilities setting.[14] Differentiating the FOC (2.5) repeatedly yields the analytical expression of market-revealed temperance $Q(w,t)$

$$Q(w,t) = E_{\{p^i\}}[Q^i] - 2Cov_{\{p^i\}}\left(\gamma^i, \frac{1}{\gamma^i}\right) - \frac{R^2(w,t)}{P(w,t)} Var_{\{p^i\}}\left(\frac{1}{\gamma^i}\right). \tag{2.21}$$

Given that market-revealed temperance arises from the third order derivative of the FOC, the dynamics of risk sharing, and thus risk tolerance measure, contribute two terms beyond the naive weighted average of individual temperance. This basic intuition also emerges from proposition 5. In the difference with prudence, for temperance the contribution of risk tolerance measure dynamics is both strong and ambiguous. The market-revealed Q can either be larger than the largest Q^i, or smaller than the smallest Q^i. In analogy with proposition 6, when specializing to the two-CRRA-agent economy, we can specifically assess the market-revealed tolerance $P(w,t)$ and temperance $Q(w,t)$ on a comparative basis. This comparison is important since both direction and quantitative behavior of savings cyclicality P_w (2.20) are determined by the relative importance of P and Q.

Proposition 7 *The market-revealed temperance in the two-CRRA-agent economy is a sim-*

[14]We derive general results for any additive utilities in the appendix 2.9.1.

ple rational-polynomial function of first agent's risk tolerance weight p^A (note: $p^B = 1 - p^A$)

$$Q(w,t) = \left(p^A(w,t)\gamma^A + p^B(w,t)\gamma^B\right)\left[3\left(\frac{p^A(w,t)}{\gamma^A} + \frac{p^B(w,t)}{\gamma^B}\right) + \frac{1 - \frac{p^A(w,t)}{(\gamma^A)^2} - \frac{p^B(w,t)}{(\gamma^B)^2}}{1 + \frac{p^A(w,t)}{\gamma^A} + \frac{p^B(w,t)}{\gamma^B}}\right], \tag{2.22}$$

which can be either positive or negative. There always exists a consumption region determined by

$$p^A(c^A,t) > \max\left\{0, \frac{1}{2} + \frac{1}{2}\frac{\gamma^A\gamma^B}{(\gamma^A - \gamma^B)}\right\},$$

within which market-revealed precautionary savings motive is countercyclical; $P_w < 0$.

As mentioned above, the cyclicality of P should influence interest rate smoothness. Hence this proposition provides an important precursor to assessing the volatilities of asset returns in this economy. Those results will be reported in proposition 8. To illustrate, Figure 2-2 plots the market-revealed temperance $Q(p^A)$ together with its two components: the weighted average temperance (first term of (2.21)) and the dynamics-induced temperance (last two terms of (2.21)). Each is a function of the first agent's risk tolerance weight $p^A = \frac{T^A}{T} = \frac{c^A\gamma^B}{c^A\gamma^B + c^B\gamma^A}$ in the illustrative two-CRRA-agent economy (with $\gamma^A = 0.1$, $\gamma^B = 10$). Clearly, unlike market-revealed RRA $R(w,t)$, $Q(w,t)$ is not bounded by individual CRRA temperances $Q^i = \gamma^i + 2$. For a certain range of consumption partition, the dynamics-induced temperance is so strong that market-revealed $Q(w,t)$ falls negative albeit all individual Q^i's are positive. Again in homogeneous limits ($p^A = 0, 1$), the sharing dynamics vanish and so does the dynamics-induced temperance.

Interestingly, with three agents or more in the economy, the market-revealed characteristics $R(w,t)$, $P(w,t)$, $Q(w,t)$ are largely independent of each other, allowing more flexibility to estimate the model in accordance with empirical patterns. This shows the rich outcome of genuine heterogeneities, beyond that of the customary but rigid assumption of a CRRA-representative agent in the literature.

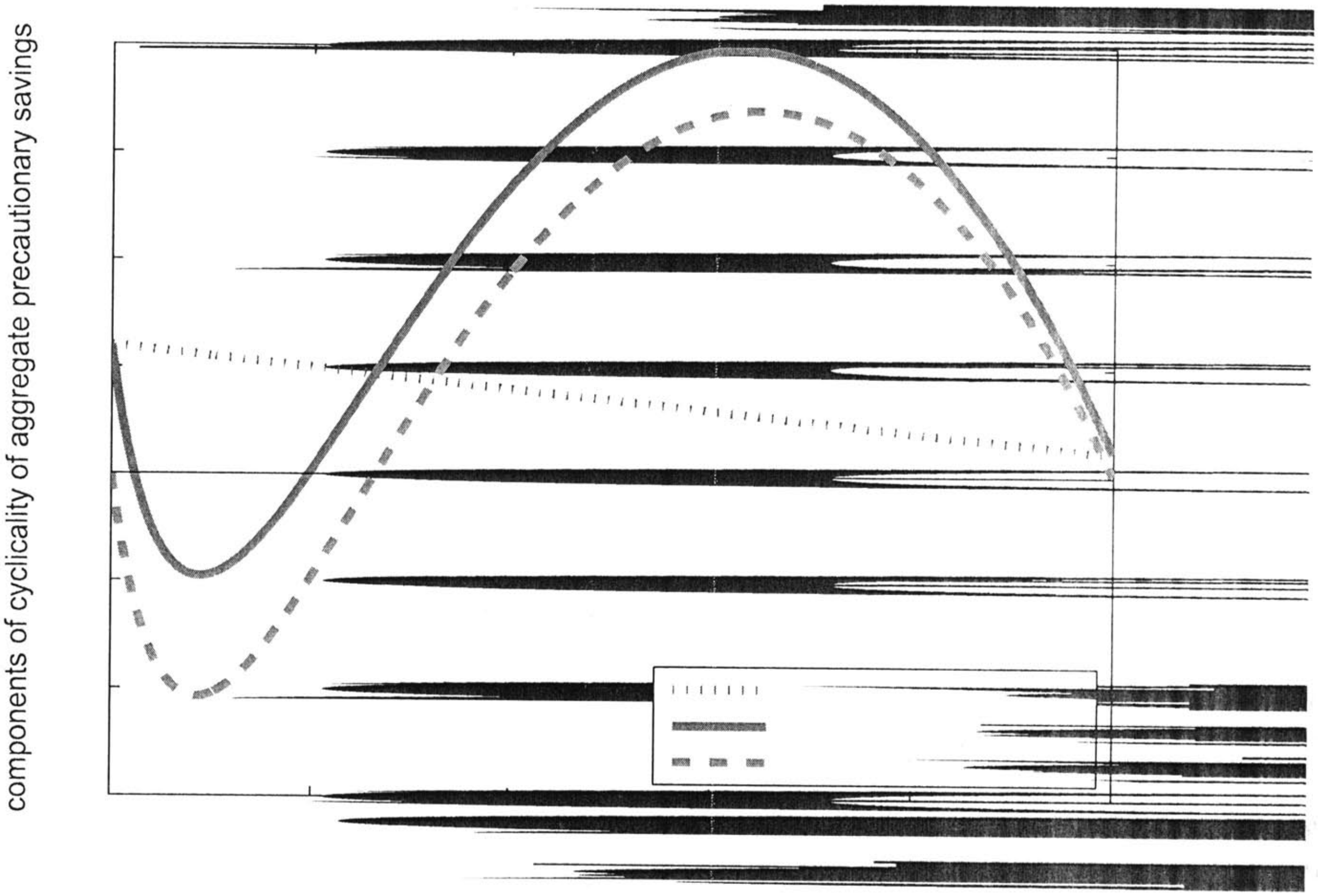


Figure 2-2: Two-CRRA-agent economy: $\gamma^A = 0.1$, $\gamma^B = 10$. Market-revealed (aggregate) temperance (savings cyclicality) Q and its components (2.21): weighted average (w.a.) $E_{\{p\}}[Q^i]$ and dynamics-induced (d.i.) savings cyclicality $Q - E_{\{p\}}[Q^i]$ (eq. (2.21)). These are plotted against agent A's risk tolerance weight $p^A = \frac{T^A}{T^A+T^B} = \frac{c^A\gamma^B}{c^A\gamma^B+c^B\gamma^A}$.

2.5 Interest rate volatility

In this model's complete-market intertemporal setting, no-arbitrage is enforced by the unique state price density $M(w,t)$. In the current consumption-based framework, this state price density is the marginal utility (2.5) of the equivalent agent

$$M(w,t) = v_w(w,t). \tag{2.23}$$

The risk-free rate (rfr) r and the market price of risk (mpr) (or Sharpe ratio) η are identified with the drift and volatility of the state price density: $\frac{dM(w,t)}{M(w,t)} = -r(w,t)dt - \eta(w,t)dZ(t)$, and thus

$$\begin{aligned} r(w,t) &= \delta(w,t) + R(w,t)\left[\mu^w - \tfrac{1}{2}(\sigma^w)^2 P(w,t)\right], \\ \eta(w,t) &= \tfrac{w\sigma^w}{T(w,t)} = \sigma^w R(w,t). \end{aligned} \tag{2.24}$$

Here $r(w,t)$ is the instantaneous risk-free rate at time t. Throughout this paper, for brevity we also refer to it interchangeably as risk-free rate and interest rate. Both rfr and mpr have forms familiar from a single-agent economy, which justifies the use of the associated characteristics $\{R, P, Q\}$ revealed by market prices as if there were a single equivalent agent representing the current heterogeneous-agent economy. In particular, a strong market-revealed precautionary savings effect is needed to drive down the interest rate's magnitude in (2.24)

$$P(w,t) > \frac{2\mu^w}{(\sigma^w)^2} \sim 100. \tag{2.25}$$

Here the numerical bound is based on the estimates of the aggregate consumption growth moments $\mu^w \sim 2\%$, $\sigma^w \sim 2\%$ (Table 2.1). As we see in proposition 6, the risk-sharing dynamic in heterogeneous-agent economy is able to generate a strong savings motive P out of much smaller individual values P^i, given that agents differ sufficiently in their risk preference. Similarly, for the stock market to be priced by the above state price density $M(w,t)$, mpr η needs to satisfy the Hansen-Jagannathan bound (Hansen and Jagannathan (1991), see also appendix). By virtue of (2.24), this constraint too has a very familiar expression in the current heterogeneous-agent setting

$$\sigma^w R(w,t) = \eta(w,t) \geq \frac{\mu^s(w,t) - r(w,t)}{\sigma_{\mu^s - r}}[1 - r(w,t)], \tag{2.26}$$

where μ^s and σ_{μ^s-r} are respectively the stock market expected return and excess return volatility. In the data, typically the stock market excess return $\mu^s - r \sim 6\%$, the excess volatility $\sigma_{\mu^s-r} \sim 20\%$ and the real rfr $r \sim 2\%$, which imply a conservative lower bound[15] on the aggregate risk aversion

$$R(w,t) \geq \frac{1}{\sigma^w}\frac{\mu^s(w,t)-r(w,t)}{\sigma_{\mu^s-r}}[1-r(w,t)] > 15. \qquad (2.27)$$

The large value for risk aversion implied from the excess stock market return is the well-known main thesis of the equity premium puzzle. In the current section, our main focus is to show analytically that this and *specially* the large precautionary savings bound (2.25) also have profound impact on the interest rate volatility. Intuitively, as hinted by the stochastic natures of $r(w,t)$ and $\eta(w,t)$ in (2.24) as well as the presence of aggregate quantities R, P therein, the heterogeneity among agents necessarily affects the volatilities of asset prices in important ways.

To fix the notation, we adopt the interest rate diffusion process $dr(w,t) = \mu^r(w,t)dt + \sigma^r(w,t)dZ(t)$ where like $r(w,t)$ itself, the μ^r, σ^r are endogenous in the model. Indeed, in analogy with (2.40), the volatility σ^r of the rfr is

$$\sigma^r(w,t) = w\sigma^w r_w(w,t) \equiv \sigma^r_\Gamma(w,t) + \sigma^r_\Delta(w,t), \qquad (2.28)$$

where

$$\sigma^r_\Gamma(w,t) \equiv w\sigma^w\left(\mu^w R_w(w,t) - \tfrac{(\sigma^w)^2}{2}\left[R_w(w,t)P(w,t) + R(w,t)P_w(w,t)\right]\right), \qquad (2.29)$$
$$\sigma^r_\Delta(w,t) \equiv w\sigma^w \delta_w(w,t),$$

are the components of rfr volatility associated primarily with the heterogeneity in risk aversion and time preference, respectively. The expressions for these components are obtained by computing the partial derivative r_w from (2.24). We now analyze the contribution of each type of heterogeneity to rfr volatility.

[15]Both bounds on P (2.25) and R (2.27) are most sensitive to the estimated value of consumption growth volatility σ^w. In the US data (Table 2.1) $\sigma^w \sim 1\%$. Here we adopt $\sigma^w \sim 2\%$ to have very conservative lower values for the aggregate savings motive and risk aversion, while noting that a smaller value of σ^w will lead to larger P, R and thus an even more volatile rfr than what we point out in this section.

Judging from the abundance of the derivatives R_w, P_w in the above expression of σ^r_{Γ}, this component of rfr volatility is necessarily characterized by the response of economy's collective risk preference and savings motive to supply shock dw. A closer look helps to estimate the magnitude of this volatility. Plugging (2.17), (2.20) into (2.29) yields

$$\sigma^r_{\Gamma} = \sigma^w R(w,t)\left[-\mu^w(P(w,t) - R(w,t) - 1) + \frac{(\sigma^w)^2}{2}P(w,t)(Q(w,t) - R(w,t) - 2)\right]. \tag{2.30}$$

Terms on the right-hand side simply express the sensitivity of aggregate intertemporal consumption smoothing and precautionary savings behaviors to output fluctuations, as they are derived directly from the last two terms of (2.24). The most remarkable feature here is that both of these sensitivities are substantial under the afore mentioned premise of large savings motives (2.25) needed for a low real interest rate. Indeed, both terms in (2.30) are dominated by the large factor P, given the realistic values for aggregate consumption moments $\mu^w, \sigma^w \sim 2\%$. This observation then offers a simple but very *drastic* implication for the interest rate of general heterogeneous-agent economies with additive utilities. Namely, in these models, a realistically low interest rate will tend to be excessively volatile. The following proposition quantifies this important observation in analytical terms.

Proposition 8 *Assuming sufficiently large precautionary savings motive (2.25), in a general economy with agents heterogeneous in their time-additive risk preferences, the interest rate volatility is almost always* [16] *bounded from below*

$$|\sigma^r(w,t)| > \mu^w\sigma^w R(w,t)\left|Q(w,t) - \frac{2\mu^w}{(\sigma^w)^2}\right|, \tag{2.31}$$

More specifically,

$$\sigma^r(w,t) > \mu^w\sigma^w R(w,t)\left(Q(w,t) - \frac{2\mu^w}{(\sigma^w)^2}\right) > 0 \ \ if \ \ Q(w,t) > \frac{2\mu^w}{(\sigma^w)^2} + R(w,t) \tag{2.32}$$

$$\sigma^r(w,t) < \mu^w\sigma^w R(w,t)\left(Q(w,t) - \frac{2\mu^w}{(\sigma^w)^2}\right) < 0 \ \ if \ \ Q(w,t) < \frac{2\mu^w}{(\sigma^w)^2} \tag{2.33}$$

[16] That is, the lower bound of interest rate volatility holds for most values of the savings motive cyclicality Q as specified in this proposition.

Qualitatively, a key factor determining the volatility of the rfr is the *cyclicality* P_w of precautionary savings, quantified by market-revealed temperance $Q(w,t)$ in the above expression. This observation identifies a new and interesting factor driving interest volatility, one that is supported by strong intuitions. Here, a critical connection is the relation (2.20), i.e., large precautionary savings P tend to induce strong savings cyclicality $|P_w|$. In turn, for large P (2.25), both the intertemporal consumption smoothing and precautionary savings motives are fiercely sensitive to supply uncertainty as in (2.30), and the resulting interest rate is highly volatile unless these two sensitivities cancel out. Proposition 8 shows that such cancellation holds only within a range of temperance, $Q \in \left(\frac{2\mu^w}{(\sigma^w)^2}, \frac{2\mu^w}{(\sigma^w)^2} + R(w,t)\right)$. Given the small empirical values for the consumption moments $\mu^w, \sigma^w \sim 2\%$, and a non-extreme value of risk aversion ($R \ll \frac{2\mu^w}{(\sigma^w)^2}$), this range is narrow on relative scale, and thus the cancellation is unlikely (see Fig. 2-3 below). As a result, large precautionary savings most likely render the interest rate both low and volatile.

Furthermore, interest rates are potentially volatile regardless of the direction of savings cyclicality. When $Q(w,t) < \frac{2\mu^w}{(\sigma^w)^2}$, the volatility of intertemporal consumption smoothing dominates the precautionary savings term. Given a positive shock to endowment, the aggregate risk aversion decreases and the elasticity of intertemporal substitution increases; agents tend to defer more consumption to later time and the interest rate drops. In other words, the equilibrium interest rate is countercyclical in this case. Conversely, when $Q(w,t) > \frac{2\mu^w}{(\sigma^w)^2} + R(w,t)$, the volatility of precautionary savings dominates the consumption smoothing term. Given a positive shock to endowment, the precautionary savings term decreases and the interest rate surges. In other words, the interest rate is procyclical here.[17] We can also draw parallel results from related literature. Kimball (1992.1993) finds in a partial equilibrium model that sufficiently temperate (large positive Q) investors may invest most of their savings in safe assets. Our findings on the relation between temperance and interest rate volatility echo this link in general equilibrium settings.

[17]Detailed portfolio choice solutions for multiple-agent economies with general additive utilities, as considered in proposition (8), are beyond the scope of this paper. Their closed-form expressions are not known and may not exist.

Quantitatively, the lower bound of interest rate volatility is substantial when Q is not in the vicinity of a knife-edge (critical) value of $\frac{2\mu^w}{(\sigma^w)^2}$. For sufficiently large precautionary savings P (2.25) (to render a low interest rate), when Q is slightly off from the above critical value, the lower bound is several times larger than the observed interest rate volatility of 2% (Table 2.1)

$$\frac{\left|Q(w,t) - \frac{2\mu^w}{(\sigma^w)^2}\right|}{\frac{2\mu^w}{(\sigma^w)^2}} > 0.1 \quad \longrightarrow \quad |\sigma_\Gamma^r| > 0.1R(w,t)\frac{2(\mu^w)^2}{\sigma^w} > 6\%,$$

where the last numerical value is based on a conservative Hansen-Jagannathan bound (2.27). Fig. 2-3 illustrates this bound in a setting with two heterogeneous CRRA agents. The figure plots the volatility of interest rate (upper panel) vis-a-vis the cyclicality of precautionary savings motives as characterized by temperance $Q(p^A)$ (lower panel). The choice of risk aversion parameters $\{\gamma^A, \gamma^B\}$ are dictated by the low empirical interest rate and Hansen-Jagannathan bound (2.25), (2.27). As stated by proposition 8, we clearly see that interest rate volatility is small only when temperance Q assumes values in the immediate vicinity of the critical value $Q^* = \frac{2\mu^w}{(\sigma^w)^2}$ (or $p^A \approx 0.35$). When Q is slightly off this value (by a few percentage points), the interest rate is hugely volatile.[18]

Proposition 8 underlines the rich and complex equilibrium dynamics of the heterogeneous economy. It shows, for e.g., that a standard cure addressing, say, the level of the rfr may adversely increase its volatility. All that said, though large precautionary savings motive has been found very useful in addressing the equity premium and interest rate level in literatures, it is likely to bring about an unrealistically volatile rfr in the heterogeneous-agent economies (with additive utilities). The incompatibility of these canonical exchange economies and the observed equity premium is well known.[19] Our contribution here is to offer a new analytical

[18] Note that $Q(p^A) = Q^* = \frac{2\mu^w}{(\sigma^w)^2}$) in another region in the vicinity of $p^A = 1$, where interest rate is both low and smooth. But in this region the less risk averse agent A dominates the economy, hence Hansen-Jagannathan bound is strongly violated, and stock market is incorrectly priced by the model.

[19] New elements in preferences such as habit formation (Campbell and Cochrane (1999)), catching-up-with-the-Joneses (Chan and Kogan (2002)), or recursive utility together with growth rate long-run predictability (Bansal and Yaron (2004)) have been invoked to tackle these asset price puzzles. In a new hybrid approach, Lettau and Wachter (2009) enlarge the state variable space to include exogenous short rate process while

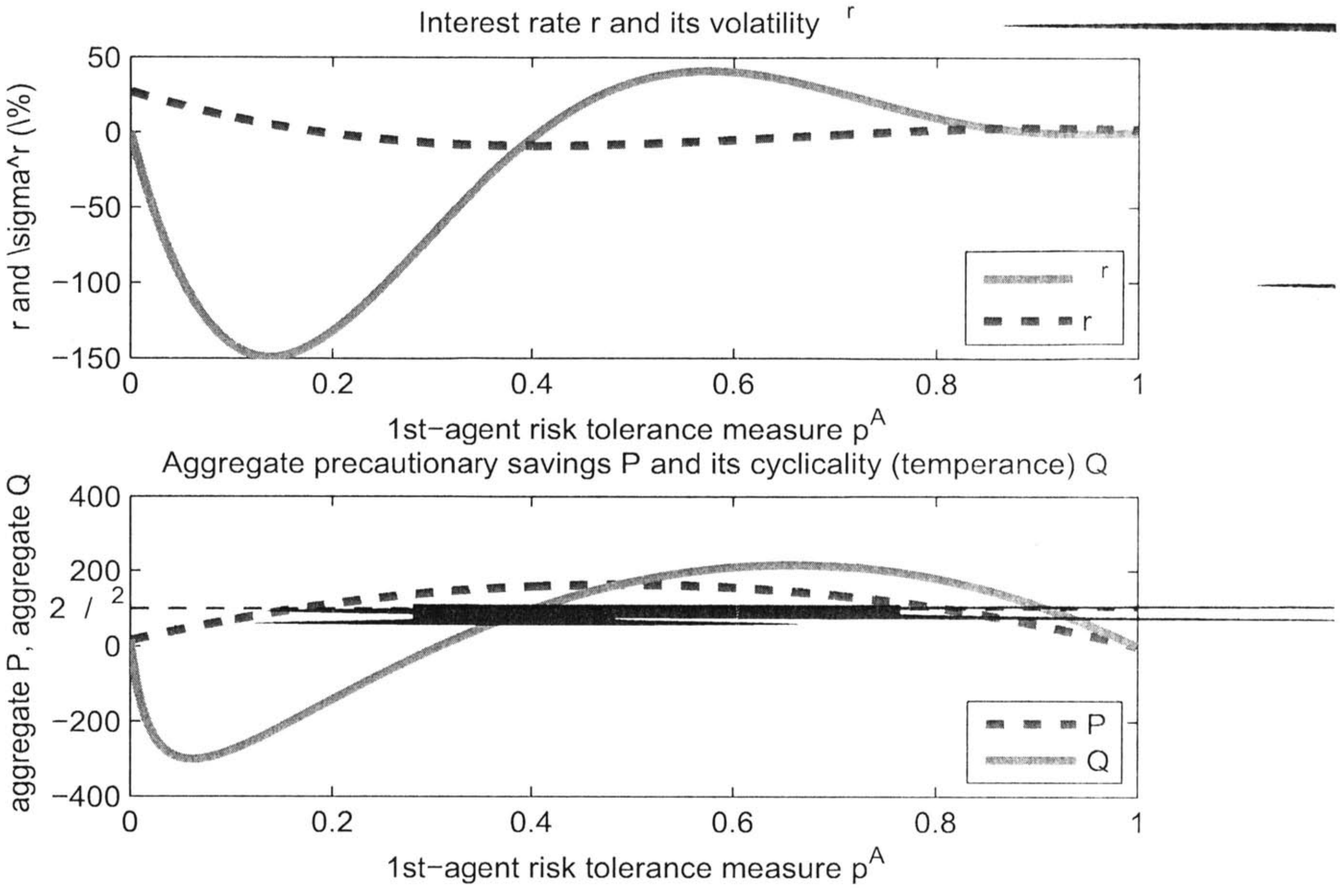


Figure 2-3: Two-CRRA-agent endowment economy: $\gamma^A = 0.01$, $\gamma^B = 15$, $\mu^w = 2\%$, $\sigma^w = 2\%$. The upper panel plots the interest rate $r(p^A)$ and interest rate volatility $\sigma^r(p^A)$ in %, the lower panel plots the market-revealed (aggregate) precautionary savings motive (prudence) $P(p^A)$ (eq. (2.18)) and savings cyclicality (temperance) $Q(p^A)$ (eq. (2.22)), and $Q(p^A) - P(p^A) - 1$. These are ploted against agent A's risk tolerance weight $p^A = \frac{T^A}{T^A+T^B} = \frac{c^A\gamma^B}{c^A\gamma^B+c^B\gamma^A}$.

perspective on this incompatibility, within the standard setting of time separable preferences.

We next consider adding heterogeneity in time preferences to see whether that can ease the puzzles. The contribution of time preference heterogeneity can be computed either directly, as to be performed in this section, or indirectly by first homogenizing this heterogeneity, as explained in section 2.7. The component σ^r_Δ of rfr volatility (2.28) arises from an

maintaining the equilibrium-based relation between the market price of risk and the fundamental dividend process.

interesting interaction between heterogeneities in risk aversion and discount factors

$$\sigma^r_\Delta \equiv w\sigma^w\delta_w(w,t) = \sigma^w R(w,t)\mathrm{Cov}_{\{p^i\}}(T^i_c, \delta^i), \tag{2.34}$$

where the last equality is an application of proposition 5, also derived in appendix 2.9.1 (eq. (2.64)). The covariance structure is rich because both the risk tolerance measure $\{p^i\}$ and marginal risk tolerance T^i_c are dynamic. In a CRRA economy, the latter is the inverse of the risk aversion coefficient. In that setting, the sign of σ^r_Δ depends on the relative orderings (comonotone or anti-comonotone) between risk aversions $\{\gamma^i\}$ and discount factors $\{\delta^i\}$. Under a positive supply shock $dw > 0$, a procyclical discount factor $\delta_w > 0$ increases the time value of consumption, thus encourages consumption and discourages savings. It thereby leads to a surge in the rfr r. Hence, a procyclical discount factor contributes to procyclicality in interest rates and vice versa. The heterogeneity in time preferences can have either positive or negative effect on rfr volatility, and therefore can help temper the extreme nature of the latter's bound.

Indeed, combining (2.28), (2.32) and (2.34) yields more comprehensive bounds on rfr volatility

$$\begin{aligned}\sigma^r(w,t) &> \sigma^w R(w,t)\left(\mu^w Q(w,t) - \tfrac{2(\mu^w)^2}{(\sigma^w)^2} + \mathrm{Cov}_{\{p^i\}}(T^i_c,\delta^i)\right) \text{ if } Q(w,t) > \tfrac{2\mu^w}{(\sigma^w)^2} + R(w,t),\\ \sigma^r(w,t) &< \sigma^w R(w,t)\left(\mu^w Q(w,t) - \tfrac{2(\mu^w)^2}{(\sigma^w)^2} + \mathrm{Cov}_{\{p^i\}}(T^i_c,\delta^i)\right) \text{ if } Q(w,t) < \tfrac{2\mu^w}{(\sigma^w)^2}.\end{aligned} \tag{2.35}$$

Specifically, for countercyclical precautionary savings motive $Q(w,t) < \frac{2\mu^w}{(\sigma^w)^2}$, a time preference ordering such that $\mathrm{Cov}_{\{p^i\}}(T^i_c, \delta^i) > 0$ helps loosen the bound on the volatility of the interest rate.[20] Similar condition holds for the other case where $Q(w,t) > \frac{2\mu^w}{(\sigma^w)^2} + R(w,t)$. Despite being a function of consumption allocations $\{c^i\}$, the covariance term is intimately associated with the discount rate heterogeneity structure, and can be formulated largely independent of the temperance term in (2.35).[21] This makes heterogeneity in time preference

[20] In CRRA settings, $T^i_c = 1/\gamma^i$, so $\mathrm{Cov}_{\{p^i\}}(T^i_c, \delta^i) = \mathrm{Cov}_{\{p^i\}}(\frac{1}{\gamma^i}, \delta^i) > 0$. This means that small γ^i are most likely associated with large δ^i and vice versa (anti-comonotone). These are configurations wherein no agent dominates other in long run (see section 2.7).

[21] In CRRA settings, the covariance term is always negative if RRAs and discount factors are co-monotone

a venue to mitigate the interest rate volatility in the consumption-based pricing models. In an attractive alternative approach, Garleanu and Panageas (2010) show that the combined features of overlapping generations and heterogeneous preferences are able to sustain the long-term survival of groups with different risk aversions, while generating stable risk-free rate.

2.6 Equity return volatility

How do heterogeneities in risk and time preferences affect the volatility of return on stock? The answer is considerably more involved than that for the interest rate because the stock price S is a contingent claim on the entire series of future dividend streams. To pursue this question, we employ the convenient tool of Malliavin calculus, following closely the approach presented in Detemple et al. (2003) and Bhamra and Uppal (2009). We assume that there are just two classes, A and B, of CRRA agents, thus simplifying the exposition while retaining heterogeneity. In such economies, there is a single state variable, which can be chosen as agent A's risk tolerance weight $p^A = \frac{T^A}{T}$. Detailed derivations can be found in the appendix 2.9.2.

In risk-neutral measure $\mathcal{Q}$, all payoffs are discounted at the risk-free rate r. The stock price then is

$$S(w,t) = e^{\int_0^t r(u)du} E_t^{\mathcal{Q}} \left[\int_t^T e^{-\int_0^u r(\tau)d\tau} w(u)du \right]. \tag{2.36}$$

In our Markovian (GBM) setting, the stock price $S(w,t)$ is a function of current endowment w, and thus stock return volatility σ^s can be defined from the associated diffusion process (i.e., gain process)

$$\frac{dS(w,t) + dw}{S(w,t)} = \mu^s(w,t)dt + \sigma^s(w,t)dZ(t). \tag{2.37}$$

$(\gamma^i > \gamma^j \leftrightarrow \delta^i > \delta^j)$, and positive if anti co-monotone $(\gamma^i > \gamma^j \leftrightarrow \delta^j > \delta^i)$, independent of consumption dynamics.

A standard application of Malliavin calculus confirms these relations

$$\sigma^s(w,t) = \sigma^w - \frac{B(w,t)}{S(w,t)} E_t^Q \left[\int_t^T du \, \frac{w(u)}{B(w,u)} \int_t^u d\tau \, \mathcal{D}_t \left(\sigma^w \eta(w,\tau) + r(w,\tau) \right) \right], \tag{2.38}$$

where $B(w,t) = \exp\left(\int_0^t r(w,u)du\right)$ is the numeraire associated with the money market account, and $\mathcal{D}_t$ denotes the Malliavin derivative at time t. This representation of stock return volatility is very intuitive, as it reflects fluctuations both in the fundamental dividend and the discounting process. In the deterministic discounting scheme (r, η are constant), fluctuation in the stock return results entirely and without distortion[22] from stochastic movement in the dividend process $\sigma^s = \sigma^w$. However, in the current general equilibrium settings, both the interest rate and the market price of risk are endogenous and stochastic. They then also contribute to the excess volatility $\sigma^s(w,t) - \sigma^w$ (terms $\sigma^w \mathcal{D}_t \eta$ and $\mathcal{D}_t r$) in (2.38)) of the stock return via the discounting mechanism. Because the Malliavin derivative of a process X is proportional to its volatility σ^X: $\mathcal{D}_t X \sim \sigma^X$ (see (2.80)), we arrive at a simple sufficient condition for stock return excess volatility to be positive, $\sigma^s(w,t) - \sigma^w > 0$, in the current two-CRRA-agent economy

$$\frac{\partial}{\partial p^A(\tau)} \left[r(p^A,\tau) + \sigma^w \eta(p^A,\tau) \right] < 0. \tag{2.39}$$

Empirically, the return excess volatility in stock market is pointed out first by Shiller (1981). Here the above condition allows us to rigorously validate intuitive arguments from the consumption CAPM literature attempting to address this anomaly. In particular, either a countercyclical Sharpe ratio or a countercyclical rfr acts to boost the stock return volatility. We now discuss these two components in more detail.

All else equal, when the interest rate r is countercyclical, r and hence the discount rate decrease with the output. Similarly, when the Sharpe ratio η is countercyclical, the risk premium, and again the discount rate, also tend to move in opposite direction with the supply. Given a positive shock to the endowment, the contingent claim (stock) price

[22] Note that the volatility σ^w of GBM endowment is kept constant by construction.

plausibly increases. However, under either countercyclical r or η, the stock price would increase more than proportionally with the endowment because the discount rate tends to drop in both cases as mentioned above. The opposite holds when the endowment shock is negative. This is why either a countercyclical Sharpe ratio $\sigma^\eta < 0$ or countercyclical interest rate ($\sigma^r < 0$) would contribute directly to positive stock return excess volatility $\sigma^s(w,t) - \sigma^w$, as expressed by each component of (2.39). The countercyclicality is a feature present in many models in the equity premium literature, and is pivotal to producing empirical patterns of predictability in stock returns. Campbell and Cochrane (1999) enlist habit formation to generate a Sharpe ratio that is high when aggregate consumption is low and vice versa. Chan and Kogan (2002) construct a heterogeneous-agent economy with a catching-up-with-the-Joneses feature in preferences, which renders risk premia countercyclical to endowment shocks. Quantitatively, a standard Ito manipulation on η (2.24) yields the following Sharpe ratio volatility (with the convention: $d\eta(w,t) = \mu^\eta(w,t)dt + \sigma^\eta(w,t)dZ(t)$)

$$\sigma^\eta(w,t) = w(\sigma^w)^2 R_w(w,t). \tag{2.40}$$

It follows that the condition $\sigma^\eta(w,t) < 0$ is achieved, as one would expect, when market-revealed risk aversion is decreasing with respect to aggregate consumption, $R_w(w,t) < 0$. This is behaviorally quite reasonable as we would expect agents to be bolder in accommodating risks when they are richer. As viewed intuitively and generically as a direct implication of the risk sharing mechanism (proposition 5), a negative R_w originates from the dynamics of the risk tolerance measure, which favors less risk averse agents after a positive shock to the endowment, and vice versa. It thus arises very naturally in the setting with heterogeneous CRRA agents (see (2.12) and also Wang (1996)). In a more general setting (beyond the CRRA framework), this countercyclicality is easily observed under the premise of large precautionary saving (2.25). Indeed, we can use (2.17) to rewrite $\sigma^\eta(w,t)$ in terms of the aggregate characteristics $R(w,t)$, $P(w,t)$

$$\sigma^\eta(w,t) = (\sigma^w)^2 R(w,t)[1 + R(w,t) - P(w,t)]. \tag{2.41}$$

Unless R assumes unreasonably large values, $R > P > \frac{2\mu^w}{(\sigma^w)^2} \sim 100$, the condition on large savings (2.25) needed for a low interest rate readily assures a countercyclical Sharpe ratio. Alternatively, proposition 9 below provides an agent-based sufficient condition for the countercyclicality beyond CRRA framework.

Proposition 9 *When all agents' risk aversions and precautionary savings motives satisfy the relation $P^i(c^i, t) \geq 1 + R^i(c^i, t)$ on the equilibrium consumption path $\{c^i\}_i$, the counterpart relation must hold at the aggregate level: $P(w, t) \geq 1 + R(w, t)$.*

Intuitively, given a certain degree of uniformity among the heterogeneous agents, this proposition asserts that the individual preference properties, that are central to determining the price volatilities, are preserved under dynamic aggregation. In other words, when all agents possess a large precautionary savings motive, so does the economy as a whole. Proposition 9 confirms and states this intuition as a rigorous sufficient condition. Whereas the risk aversion aggregation is linear (proposition 4), the aggregation on precautionary savings is highly nonlinear. This contrast makes these results far from obvious. It is also interesting to note that, $R_w(t, w) = \frac{R(t,w)}{w}(1 + R(t, w) - P(t, w))$ as in (2.17), proposition 9 simply states that market-revealed risk aversion is decreasing in consumption if that property holds for each individual agent. A known special result of this proposition is obtained when all individual utilities belong to the CRRA class, whence both $R^i = \gamma^i$, $P^i = \gamma^i + 1$ are constant and satisfy the hypothesis of proposition 9. Then

$$P(w, t) = R(w, t) + 1 - Cov_{\{p^i\}}\left(\gamma^i, \frac{1}{\gamma^i}\right) > R(w, t) + 1.$$

Proposition 9, however, holds more generally for any additive expected utilities.

Back to the condition (2.39); combining its two terms yields a more complete insight into the relation between stock price movement and the economy's behavior toward risks. We rewrite this sufficient condition for positive stock return excess volatility in term of aggregate quantities R, P, Q

$$(\sigma^w)^2 \left[\frac{\mu^w}{(\sigma^w)^2}(1 + R(w,t) - P(w,t)) + \frac{P(w,t)(Q(w,t) - R(w,t) - 4)}{2} \right] < -Cov_{\{p^i\}}(\delta^i, \frac{1}{\gamma^i}) \quad (2.42)$$

A few important observations should be made. First, each of risk aversion, precautionary savings and temperance affects stock return volatilities. Intuitively, this is because all three influence savings and portfolio choices. The mechanism at work is as follows. All else being equal, small Q enforces the above sufficient condition, and therefore boosts the excess volatility of the return on stocks. We recall from (2.20) that temperance Q is crucially related to P_w, namely small enough Q is associated with procyclical P. A positive supply shock will increase precautionary savings (as $P_w > 0$), leading to a decrease in both the interest and discount rates (see (2.24)). Thus the stock price increases more than proportionally compared to the endowment, which implies excess volatility in the stock return. (See also Shiller (1981) for a behavioral explanation of this phenomenon.)

Second, the relative orderings between agents' risk aversions and subjective discount factors also influence return volatility, via the term $Cov_{\{p^i\}}(\delta^i, \frac{1}{\gamma^i})$. That is because these orderings determine the dynamics of risk sharing, consumption partition and risk tolerance measure in the economy. These in turn are compounded in the asset price movements due to changes in endowment. We will return to these heterogeneity effects in the next section.

Finally, it is noted that while risk aversion and the precautionary savings motive have enjoyed substantial credence as shapers of asset price patterns in consumption-based pricing models, the cyclical properties of precautionary savings (or equivalently, temperance) are not well studied. Our investigation makes explicit the important link between these cyclical properties and asset (bond and stock) return volatilities. One reason why this very intuitive link has been quite implicit in the literature lies with the heterogeneity structure of the model itself. For a close illustration, we consider the setting of Bhamra and Uppal (2009). They obtain the first sufficient condition for positive stock return excess volatility that involves solely precautionary savings.[23] How can we reconcile this result with our condition (2.42)?

[23]Bhamra and Uppal (2009) investigates an exchange economy with two agents who differ only in risk

The answer is as follows. In the two-CRRA-agent economy, as seen earlier, there is only a single state variable. This can be chosen without loss of generality as the first agent's risk tolerance measure $p^A = \frac{T^A}{T}$. Each and every aggregate quantity R, P and Q then is a simple function of p^A, and thus they pairwise bear a one-to-one relation.[24] The derivation of Bhamra and Uppal's sufficient condition exploits these simple relations, and in doing so inadvertently obscures the role of temperance $Q(w,t)$.[25] In fact, by virtue of (2.24), the derivative of rfr $\frac{dr}{dp^A}$ contains the term $\frac{dP}{dp^A} = \frac{P_w}{p^A_w}$, which is obviously related to the cyclicality P_w of precautionary savings. This example and (2.42) together indicate that in more general multiple-agent settings R and P are important, but far from sufficient statistics to determine stock return volatilities.

It is reassuring that all the above observations and intuitions concerning the cyclicality of precautionary savings, or equivalently temperance, also underlie the parallel results on interest rate volatility, reported in proposition 8.

2.7 Heterogeneities and homogenization of beliefs

The heterogeneous-agent economies we have explored so far address heterogeneities in risk aversion and time preferences. As we have seen, these differences can foster rich and resilient exchanges leading to the equilibrium when agents assume off-setting characteristics in their preferences. While a higher degree of patience (smaller δ^i) favors deferring consumptions, a larger elasticity of intertemporal substitution ψ^i (equivalently lower risk aversion $\gamma^i = \frac{1}{\psi^i}$ in the additive utility framework) produces the same effect. Another practical and important factor in which agents differ is in their subjective beliefs about economic fundamentals.

aversion. Their proposition 2 presents a sufficient condition for positive stock return excess volatility; $P < 1 + \frac{\mu^w}{(\sigma^w)^2}$. This is a stronger version of (2.42), when (2.42) is adapted to the setting of homogeneous time preferences.

[24] In two-CRRA-agent economy, we have $P(w,t) = R(w,t)\left(1 + \frac{\gamma^A + \gamma^B - R(w,t)}{\gamma^A\gamma^B}\right)$.

[25] Since $P = (p^A\gamma^A + p^B\gamma^B)\left(1 + \frac{p^A}{\gamma^A} + \frac{p^B}{\gamma^B}\right)$, we have $\frac{dP}{dp^A} = (\gamma^A - \gamma^B)\left(\frac{P}{R} - \frac{R}{\gamma^A\gamma^B}\right)$ (this relation is needed in the derivation of key condition (2.39), see (2.83)). Thus $\frac{dP}{dp^A}$, and for that matter, sufficient condition (2.39) appear unrelated to temperance Q, while they actually are.

Such beliefs directly affect agents' intertemporal decisions and thus asset prices. In this section we will show that, as far as consumption and risk sharing are concerned, an economy whose agents differ in all time preferences, risk aversions and beliefs may be transformed isomorphically into a far simpler one with heterogeneity only in risk aversion. The required transformation offers new quantitative perspectives on the above-mentioned tradeoff between different dimensions of heterogeneity. The analysis also relates neatly to the survival of market participants (a.k.a market selection) in the long run.

2.7.1 Heterogeneity in time preferences, risk aversions and beliefs

We consider the canonical case, widely studied in literature, of a two-CRRA-agent economy with GBM endowments. The next section addresses the setting with multiple agents. In addition to heterogeneities in discount factors and risk aversion, agents A, B also differ in their beliefs about the growth rates μ^A, μ^B of the endowment process $w(t)$ (2.1). The realizations of $w(t)$ are correctly observed by all parties

$$\mu^{w,A}dt + \sigma^w dZ^A(t) = \frac{dw(t)}{w(t)} = \mu^{w,B}dt + \sigma^w dZ^B(t),$$

where $Z^A(t)$, $Z^B(t)$ are standard Brownian motions under each agent's subjective information set (i.e., belief). We assume agents act on their own persistent beliefs.[26] A comparison with (2.1) yields

$$\begin{aligned} dZ^A(t) &= dZ(t) + \theta^A dt; \qquad \theta^A = \frac{\mu^w - \mu^{w,A}}{\sigma^w}, \\ dZ^B(t) &= dZ(t) + \theta^B dt; \qquad \theta^B = \frac{\mu^w - \mu^{w,B}}{\sigma^w}. \end{aligned} \tag{2.43}$$

Coefficient θ^i in essence characterizes the deviation of agent i's beliefs on the endowment growth rate $\mu^{w,i}$ from the its true value μ^w. When $\theta^i < 0$, agent i is optimistic (with respect to

[26]That is, agents do not draw inferences from the willingness to trade by others. Later, we will extend our framework to accommodate time-varying beliefs, which in turn may arise from learning or other ad-hoc belief adjustment mechanism.

the objective growth rate μ^w) and vice versa. Also, two agents assign different but equivalent probability measures and distributions to the future uncertain endowment process. Since agents are still allowed to trade in the riskless bond and a contingent claim on the aggregate endowment (stock), the market is complete and the equivalent-agent optimization problem can be constructed to explicitly account for different beliefs

$$\max_{\{c^A(t),c^B(t)\}} \frac{1}{\lambda^A} E_0^{(A)} \left[\int_0 e^{-\delta^A t} u^A(c^A) dt \right] + \frac{1}{\lambda^B} E_0^{(B)} \left[\int_0 e^{-\delta^B t} u^B(c^B) dt \right] \quad (2.44)$$
$$\text{s.t.} \quad c^A(t) + c^B(t) = w(t) \quad \forall t.$$

Here $u^i = \frac{(c^i)^{1-\gamma^i}}{1-\gamma^i}$, and $E_t^{(i)}[\ldots]$ denotes the time-t conditional expectation under agent i's belief. There exists a standard approach (see e.g., Detemple and Murthy (1994) and Basak (2005)) to convert the above optimization problem to one under the physical measure

$$\max_{\{c^A(t),c^B(t)\}} E_0 \left[\frac{1}{\lambda^A} \int_0 \xi^A(t) e^{-\delta^A t} u^A(c^A) dt + \frac{1}{\lambda^B} \int_0 \xi^B(t)\theta(t) e^{-\delta^B t} u^B(c^B) dt \right]$$
$$\text{s.t.} \quad c^A(t) + c^B(t) = w(t) \quad \forall t.$$

The above operation involves a change of measure, from subjective $\mathbb{P}^i$ to physical $\mathbb{P}$, using the Radon-Nikodym derivative $\xi^i(t)$

$$\xi^i(t) = \frac{d\mathbb{P}^i}{d\mathbb{P}} = \exp\left(-\frac{1}{2}(\theta^i)^2 t - \theta^i Z(t) \right) \quad i \in \{A, B\}, \quad (2.45)$$

where θ^i is given in (2.43). The dynamics of this heterogeneous-agent economy is captured by the FOC and the market clearing equation

$$\begin{cases} \frac{1}{\lambda^A} e^{-\delta^A t} \xi^A(t)(c^A(t))^{-\gamma^A} = \frac{1}{\lambda^B} e^{-\delta^B t} \xi^B(t)(c^B(t))^{-\gamma^B} \\ c^A(t) + c^B(t) = w(t) \end{cases} . \quad (2.46)$$

Here we clearly see that all three dimensions of heterogeneity - risk aversion, time preference and belief - play roles in shaping the equilibrium. To simplify the analysis, it would be desirable to reduce this economy to one where only risk aversion experiences heterogeneous.

Remarkably, that is possible. Consider the following simple multiplicative transformation (which is derived in the proof of proposition 10, see appendix 2.9.3)

$$\begin{cases} c^A(t) \to \hat{c}^A(t) \equiv \Upsilon(Z(t),t)c^A(t), \\ c^B(t) \to \hat{c}^B(t) \equiv \Upsilon(Z(t),t)c^B(t), \\ w(t) \to \hat{w}(t) \equiv \Upsilon(Z(t),t)w(t), \end{cases} \tag{2.47}$$

where

$$\Upsilon(Z(t),t) = \exp\left(\beta^{\gamma,\delta}t\right)\exp\left(\beta^{\gamma,\theta}Z(t)\right), \tag{2.48}$$
$$\beta^{\gamma,\theta} \equiv \frac{\theta^A-\theta^B}{\gamma^A-\gamma^B}; \qquad \beta^{\gamma,\delta} \equiv \frac{\delta^A+\frac{(\theta^A)^2}{2}-\delta^B-\frac{(\theta^B)^2}{2}}{\gamma^A-\gamma^B} \equiv \frac{\delta^A_{eff}-\delta^B_{eff}}{\gamma^A-\gamma^B}.$$

The coefficients $\delta^A_{eff} \equiv \delta^A + \frac{(\theta^A)^2}{2}$, $\delta^B_{eff} \equiv \delta^B + \frac{(\theta^B)^2}{2}$ are the effective discount rates of agent A and B respectively, with their subjective beliefs being incorporated. The coefficients $\beta^{\gamma,\theta}$ and $\beta^{\gamma,\delta}$ quantify respectively differences in beliefs and in time preferences, normalized with respect to the difference in risk aversions. These coefficients will have a neat interpretation as slopes of a linear projection in characteristics space (δ,γ,θ) when we come to the full multiple-agent settings in the next section. Interestingly, we note that this transformation indeed considerably simplifies the full dynamics (2.46), which now become

$$\begin{cases} \frac{1}{\lambda^A}(\hat{c}^A(t))^{-\gamma^A} = \frac{1}{\lambda^B}(\hat{c}^B(t))^{-\gamma^B} \\ \hat{c}^A(t)+\hat{c}^B(t) = \hat{w}(t) \end{cases}. \tag{2.49}$$

Equation (2.49) represents the familiar dynamics of a two-CRRA-agent economy whose agents differ only in their risk aversions γ^A, γ^B, as studied in Benninga and Mayshar (2000), Dumas (1989) and Wang (1996). Effectively, we have been able to "rotate" the heterogeneities in subjective beliefs and discount factors away by changing the aggregate endowment $w(t)$ to $\Upsilon(Z(t),t)w(t)$. This in turn is equivalent to shifting the growth and volatility rates of the GBM endowment

$$\begin{aligned} \tfrac{d\hat{w}(t)}{\hat{w}(t)} &\equiv \mu^{\hat{w}} dt + \sigma^{\hat{w}} dZ(t), \\ \sigma^{\hat{w}} &= \sigma^{w} + \beta^{\gamma,\theta}, \\ \mu^{\hat{w}} &= \mu^{w} + \beta^{\gamma,\delta} + \beta^{\gamma,\theta}\left(\sigma^{w} + \tfrac{\beta^{\gamma,\theta}}{2}\right). \end{aligned} \tag{2.50}$$

Thus in the dynamics of consumption and risk sharing, the differences in time preferences and beliefs can be taken into account by modifying both the growth and volatility of the supply process. We will refer to $\{\gamma^1, \gamma^2, \delta^1, \delta^2, \theta^1, \theta^2, w(t)\}$ as the original economy, in which two CRRA agents differ in risk aversion, time preference and belief, as specified in (2.43). Similarly, we denote $\{\gamma^1, \gamma^2, \hat{w}(t)\}$ as the reduced economy, whose agents differ only in risk aversion. The defining property of the transformation, that all agents' equilibrium consumptions stay the same up to a (stochastic) multiplicative factor $\Upsilon(Z(t), t)$ in the two economies (2.47), implies a profound relationship between the two respective consumption sharing dynamics. Not only are the consumption shares unchanged ($\frac{\hat{c}^i}{\hat{w}} = \frac{c^i}{w}$ and $\frac{\hat{c}^i}{\hat{c}^j} = \frac{c^i}{c^j}$), but more importantly, the individual marginal propensities to consume out of the aggregate endowment (2.7), our key risk tolerance measure, remain identical in the two economies.

$$\hat{c}^i_{\hat{w}} = \frac{\hat{T}^i(\hat{c}^i, t)}{\hat{T}(\hat{w}, t)} = \frac{\frac{\hat{c}^i}{\gamma^i}}{\sum_i \frac{\hat{c}^i}{\gamma^i}} = \frac{\frac{c^i}{\gamma^i}}{\sum_i \frac{c^i}{\gamma^i}} = \frac{T^i(c^i, t)}{T(w, t)} = c^i_w.$$

And so do the aggregate characteristics built upon this measure in the two economies. The first is the (market-revealed) equivalent risk aversion (2.8)

$$\hat{R}(\hat{w}, t) = \sum_i \frac{\hat{T}^i(\hat{c}^i, t)}{\hat{T}(\hat{w}, t)}\gamma^i = \sum_i \frac{T^i(c^i, t)}{T(w, t)}\gamma^i = R(w, t).$$

Market-revealed precautionary savings $P(w, t)$ and temperance $Q(w, t)$ are also identical in the two economies, which can be directly deduced from their expressions (2.71), (2.72) for CRRA utilities. Because of these relationships, we will refer to this key property generally as preserving *consumption partition dynamics* below. We summarize this precise correspondence in the following proposition.

Proposition 10 *Suppose that the aggregate endowment follows a GBM process $w(t)$ (2.1), and that there are two classes of CRRA agents. In term of consumption partition dynamics at equilibrium, the two economies are isomorphic:*

$$\{\gamma^1, \gamma^2, \delta^1, \delta^2, \theta^1, \theta^2, w(t)\} \longleftrightarrow \{\gamma^1, \gamma^2, \hat{w}(t)\},$$

where the isomorphic endowment $\hat{w}$ is also a GBM process defined in (2.50).

Though this result holds exactly under the specific premise of GBM endowment, it clearly shows the direction and possibility of an interesting and qualitative tradeoff between agent-based characteristics and aggregate supply statistics in more general cases. In this way, the findings in a reduced economy can be adapted to economies with additional dimensions of heterogeneity. Among others, the analytical results on the linkage between risk sharing and the size of endogenous credit markets obtained in Longstaff and Wang (2009) can be immediately generalized to allow agents to differ also in time preference. To fix the convention for the next discussion, we assume without loss of generality that $\gamma^A < \gamma^B$ throughout.

First we note that when $\delta^A_{eff} < \delta^B_{eff}$,[27] $\beta^{\gamma,\delta} > 0$, the modified endowment $\hat{w}$ has an unambiguously higher growth rate (2.50). That is, as agent A is both less risk averse and effectively more patient in the original economy, she would take more risk and be more willingly to defer consumption than would agent B. Then it is necessary to boost the isomorphic economy's endowment growth rate, in which agents are now equally patient,[28] to induce agent A to undertake similar consumption sharing in equilibrium. The opposite holds when $\delta^A_{eff} > \delta^B_{eff}$. Second, when $\theta^A < \theta^B$, $\beta^{\gamma,\theta} > 0$, the modified endowment $\hat{w}$ has both higher growth rate and volatility (2.50). That is, as agent A is both less risk averse and more optimistic[29] in the original economy, she would bear risk more aggressively in this case

[27]Since $\delta^A_{eff} = \delta^A + \frac{(\theta^A)^2}{2}$, $\delta^B_{eff} = \delta^B + \frac{(\theta^B)^2}{2}$, this inequality can be result of $\{\delta^A < \delta^B; \theta^A = \theta^B\}$, or $\{\delta^A = \delta^B; \theta^A < \theta^B\}$, or some of their appropriate mixtures

[28]They are now heterogeneous only in risk aversions

[29]$\theta^A < \theta^B$ and (2.43) imply that agent A believes in a higher growth rate than agent B: $\mu^{w,A} > \mu^{w,B}$

too. Then to preserve equilibrium consumption partition dynamics, it is necessary to boost *both* the isomorphic economy's endowment growth rate and its volatility, given that agents now have identical beliefs. Finally, we also note that while time preference heterogeneity is reflected only in the isomorphic economy's endowment growth rate, belief heterogeneity influences both that growth rate and volatility. This is because a subjective belief relative to truth, as characterized by a Radon-Nikodym change of measure (2.45), is always stochastic, while a discount process $e^{-\delta^i t}$ is deterministic.

Time-varying beliefs

Interestingly, the above isomorphism also exists in the richer class where beliefs vary over time as agents observe the realizations of the endowment process. The analysis can address general forms of time variation of subjective beliefs, for which the perceived growth rates $\mu^{w,A}$, $\mu^{w,B}$ of endowment are bounded, adapted processes.[30] Important special cases would be Bayesian updating and other ad-hoc learning mechanisms. In such settings, in place of (2.45), individual beliefs are characterized by the path-dependent Radon-Nikodym derivatives

$$\xi^i(t) = \frac{d\mathbb{P}^i}{d\mathbb{P}} = \exp\left(-\frac{1}{2}\int^t (\theta^i(w,s))^2 ds - \int^{Z(t)} \theta^i(w,s)dZ(s)\right) \qquad i \in \{A, B\}.$$

The coefficients θ^A, θ^B (2.43) now are bounded, adapted stochastic processes and describe possible evolution patterns of beliefs. To illustrate, let us briefly consider two examples. The first is the Bayesian updating case where agents' priors about the endowment's unobserved growth rate μ^w are normal distributions $N(m^I(t), v^I(t))$, $I \in \{A, B\}$. In this setting, Brennan (1998) obtains the following learning dynamic[31]

$$\begin{cases} dm^I = \frac{v^I(0)}{v^I(0)t+(\sigma^w)^2}\left[(\mu^w - m^I)dt + \sigma^w dZ(t)\right], \\ v^I(t) = \frac{v^I(0)(\sigma^w)^2}{v^I(0)t+(\sigma^w)^2}, \end{cases} \qquad I \in \{A, B\}.$$

[30] These are prerequisites for Girsanov's theorem on change of measure to work. See, e.g., section 3.5 in Karatzas and Shreve (1991).

[31] We assume that agents agree to disagree, and learn only from the observed realizations of endowment.

Evidently, as time lapses, both agents' beliefs converge to truth; $\lim_{t\to\infty} v^I(t) \to 0$, $\lim_{t\to\infty} m^I(t) \to \mu^w$, $I \in \{A, B\}$. In the second example, even if agents eventually learn the truth, their beliefs may diverge incrementally following a negative shocks to the output when relation $\frac{\partial|\theta^A - \theta^B|}{\partial w} < 0$ holds.

The current general belief heterogeneity can be rotated away by modifications in the growth and volatility of endowment process. similar to (2.47). The only difference with (2.48) is that now the transformation parameters $\beta^{\gamma,\theta}$, $\beta^{\gamma,\delta}$ are stochastic. Accordingly, in place of (2.50), the endowment process of the isomorphic economy becomes

$$\begin{aligned} \frac{d\hat{w}(t)}{\hat{w}(t)} &\equiv \mu^{\hat{w}} dt + \sigma^{\hat{w}} dZ(t), \\ \sigma^{\hat{w}}(w,t) &= \sigma^w + \frac{\theta^A(w,t) - \theta^B(w,t)}{\gamma^A - \gamma^B}, \\ \mu^{\hat{w}}(w,t) &= \mu^w + \frac{\delta^A - \delta^B}{\gamma^A - \gamma^B} + \frac{1}{2}\frac{[\theta^A(w,t)]^2 - [\theta^B(w,t)]^2}{\gamma^A - \gamma^B} + \frac{\theta^A(w,t) - \theta^B(w,t)}{\gamma^A - \gamma^B}\left(\sigma^w + \frac{1}{2}\frac{\theta^A(w,t) - \theta^B(w,t)}{\gamma^A - \gamma^B}\right). \end{aligned} \tag{2.51}$$

While the original output $w(t)$ is a pure geometric brownian process, its isomorphic counterpart $\hat{w}(t)$ incorporating the time variance in belief dynamics, generally belongs to richer classes. In particular, when beliefs diverges in bad time ($dw < 0$), the volatility of the isomorphic economy's endowment $\sigma^{\hat{w}}$ gets further away from that of the original economy σ^w, though the former economy does not necessarily become more volatile (i.e., $\sigma^{\hat{w}}$ can either increase or decrease with w). Furthermore, certain time-varying patterns of beliefs in the original economy may transform into a degree of mean reversion in the output of the isomorphic economy so that the risk-sharing dynamic between agents is preserved despite beliefs being homogenized. The mean reversion in the output's growth benefits alternatively one or the other agent when the trend turns.[32] This implies that the original belief heterogeneity acts to compensate agents' difference in risk aversions in a way that sustain their presence in equilibrium, despite market selection. Qualitatively, the isomorphic transformation allows us to see quickly how heterogeneities in beliefs and time preferences affect agent's risk-sharing behaviors in the original economy per se. The dynamic (2.51) of isomorphic economy's output then initiates a quantitative analysis of the risk sharing in the simplified setting of

[32]We will analyze in section 2.7.3 how the output's growth rate affects agents' survival in the long run.

heterogeneity only in risk aversion.

So far our analysis has involved two-CRRA-agent economies, for which case the isomorphism exists. We turn next to the more general setting with multiple CRRA agents and relate it naturally to the important issue of long-run survival of these agents.

2.7.2 Multi-agent setting

We now generalize the findings of the previous section to the case of many CRRA agents, and relegate missing derivations to the appendix 2.9.3. Quantitatively, the consumption dynamics isomorphism between the original (fully heterogeneous) and the reduced (agents heterogeneous only in risk aversions) economy $\{\{\gamma^i, \delta^i, \theta^i\}_i, w(t)\} \longleftrightarrow \{\{\gamma^i\}_i, \hat{w}(t)\}$ is concerned with both FOC and market clearing.

$$\left\{ \begin{array}{c} \frac{1}{\lambda^i} e^{-\delta^i t} \xi^i(t) (c^i(t))^{-\gamma^i} = M(w,t) \quad \forall i \\ \sum_i c^i(t) = w(t) \end{array} \right. \longleftrightarrow \left\{ \begin{array}{c} \frac{1}{\hat{\lambda}^i} (\hat{c}^i(t))^{-\gamma^i} = \hat{M}(\hat{w},t) \quad \forall i \\ \sum_i \hat{c}^i(t) = \hat{w}(t) \end{array} \right. . \tag{2.52}$$

In the above expressions, $M(w,t)$ and $\hat{M}(\hat{w},t)$ are unique state price densities in the respective economies. The key to this isomorphism is the existence of a common multiplicative factor ($\Upsilon(Z(t),t) = \frac{c^i}{\hat{c}} = \frac{w}{\hat{w}}\ \forall i$) that is able to absorb and homogenize all agent-specific time preferences and beliefs

$$[\Upsilon(Z(t),t)]^{\gamma^i} e^{-\delta^i t} \xi^i(t) = \frac{M(w,t)}{\hat{M}(\hat{w},t)} \quad \forall i.$$

Plugging in agent i's belief ξ^i (2.45) for the GBM endowment under current consideration, the above condition is satisfied when two linear (quadratic) relations hold in characteristics space $(\delta^i, \gamma^i, \theta^i)$ (A, B, C, D are some constants, that are identical for all agents)

$$\left\{ \begin{array}{c} \delta^i_{eff} \equiv \delta^i + \frac{(\theta^i)^2}{2} = A + B\gamma^i, \\ \theta^i = C + D\gamma^i, \end{array} \right. \quad \forall i. \tag{2.53}$$

Under these premises, much more meaningful interpretations can be obtained for coefficients A, B, C, D. Namely, they are the slope and intercept coefficients of projections from time preferences $\{\delta^i\}$ and beliefs $\{\theta^i\}$ onto risk aversion $\{\gamma^i\}$ parameter spaces.

$$B = \beta^{\gamma,\delta} = \frac{Cov\left(\gamma^i, \delta^i_{eff}\right)}{Var\left(\gamma^i\right)} = \frac{\frac{1}{N}\sum_i^N \left(\gamma^i\delta^i_{eff}\right) - \frac{1}{N^2}\sum_i^N \gamma^i \sum_j^N \delta^j_{eff}}{\frac{1}{N}\sum_i^N \left(\gamma^i\right)^2 - \frac{1}{N^2}\left(\sum_i^N \gamma^i\right)^2}, \tag{2.54}$$

$$D = \beta^{\gamma,\theta} = \frac{Cov\left(\gamma^i, \theta^i\right)}{Var\left(\gamma^i\right)} = \frac{\frac{1}{N}\sum_i^N \left(\gamma^i\theta^i\right) - \frac{1}{N^2}\sum_i^N \gamma^i \sum_j^N \theta^j}{\frac{1}{N}\sum_i^N \left(\gamma^i\right)^2 - \frac{1}{N^2}\left(\sum_i^N \gamma^i\right)^2}, \tag{2.55}$$

where N is the number of agents in the economy. In this result, heterogeneities in beliefs and time preferences are accounted for by a change in endowment, very much like the setting with two agents

$$\Upsilon(Z(t), t) = \exp\left(\beta^{\gamma,\delta} t\right) \exp\left(\beta^{\gamma,\theta} Z(t)\right), \tag{2.56}$$

$$w(t) \longrightarrow \hat{w}(t) = \Upsilon(Z(t), t) w(t) \equiv \exp\left[\left(\mu^{\hat{w}} - \frac{(\sigma^{\hat{w}})^2}{2}\right) t + \sigma^{\hat{w}} Z(t)\right].$$

$$\sigma^{\hat{w}} = \sigma^w + \beta^{\gamma,\theta}, \tag{2.57}$$

$$\mu^{\hat{w}} = \mu^w + \beta^{\gamma,\delta} + \beta^{\gamma,\theta}\left(\sigma^w + \frac{\beta^{\gamma,\theta}}{2}\right).$$

In particular, when either γ^i and θ^i (or γ^i and δ^i_{eff}) are co-monotone, the slope coefficients $\beta^{\gamma,\theta}$ (or $\beta^{\gamma,\delta}$) are positive. Then the growth rate $\mu^{\hat{w}}$ and volatility $\sigma^{\hat{w}}$ of the isomorphic endowment $\hat{w}$ are unambiguously larger than their original counterparts μ^w, σ^w. This is because the co-monotonicity in γ^i and θ^i means agents are highly polarized; less risk averse agents are also likely more optimistic ones and vice versa. To induce agents to preserve their consumption sharing dynamics, it is necessary to boost both the growth rate and volatility of the endowment in the reduced economy, in which agents by construction have homogeneous time preferences and beliefs (that is, they are less polarized). The same applies for co-monotonicity in γ^i and δ^i_{eff}. These general intuitions, when combined with the regression-based interpretation of the coefficients B, D in (2.54), (2.55), point again to the interesting tradeoff between microscopic (agent-based) characteristics and macroscopic (aggregate) supply statistics in the multiple-agent economy. When the the linearities (2.53) in characteristics space $(\delta^i, \gamma^i, \theta^i)$ do not hold, no exact isomorphism can be found between the original $\{\{\gamma^i, \delta^i, \theta^i\}_i, w(t)\}$ and the reduced $\{\{\gamma^i\}_i, \hat{w}(t)\}$ economies. Nevertheless, the latter can always be explicitly constructed about the linear projections (2.54), (2.55) from time preferences $\{\delta^i\}$ and beliefs $\{\theta^i\}$ onto risk aversion $\{\gamma^i\}$, as we see in (2.57). We reasonably expect that the consumption partition dynamics in the reduced economy, heterogeneous only in risk aversions, would most closely match that of the original economy, heterogeneous in all three dimensions of risk aversion, time preference and beliefs.

So far in this section, our strategy for analyzing heterogeneous-agent economies has been to deform the aggregate supply process to the point that it fully (or best) accounts and thus compensates for agents' heterogeneities in time preferences and beliefs. In certain aspects, this pairs well with a popular strategy in the literature to substitute different dimensions of heterogeneity, either at the individual agent or representative agent level. The latter strategy addresses whether the risk loving, patience and optimism of each agent or the whole economy (market-revealed agent) are equivalent and mutually substitutable given observed risk sharing and price dynamics. In the single-generation settings under current consideration, a specific but central question is on the domination and survival of some agents over the others in the long run. Working in the context of the market selection, we now formally relate these two strategies.

2.7.3 Agent survival

Following Sandroni (2000) and Yan (2008) we use original economy's FOC (2.52) to examine the scaled equilibrium consumption ratio of any two agents i, j

$$\frac{\left[\frac{c^i(w,t)}{w(t)}\right]^{\gamma^i}}{\left[\frac{c^j(w,t)}{w(t)}\right]^{\gamma^j}} = \frac{\frac{1}{\lambda^i} e^{-\delta^i t} \xi^i(t) [w(t)]^{-\gamma^i}}{\frac{1}{\lambda^j} e^{-\delta^j t} \xi^j(t) [w(t)]^{-\gamma^j}}$$

$$= \frac{\lambda^j w_0^{-\gamma^i}}{\lambda^i w_0^{-\gamma^j}} \exp\left[(I^j - I^i)t\right] \exp\left[\left(\theta^j + \gamma^j \sigma^w - \theta^i - \gamma^i \sigma^w\right) Z(t)\right], \qquad (2.58)$$

where w_0 is the initial value of endowment, and

$$I^i \equiv \delta^i + \frac{(\theta^i)^2}{2} + \gamma^i \left(\mu^w - \frac{(\sigma^w)^2}{2}\right) \quad \forall i. \qquad (2.59)$$

Consider the case $\mu^w > \frac{(\sigma^w)^2}{2}$ so that the economy is growing statistically. When $I^i < I^j$, Yan (2008) notes that the above scaled equilibrium consumption ratio (2.58) grows to infinity almost surely as $t \to \infty$. As the consumption ratio $\frac{c^j}{w} \in [0, 1]$ is bounded, this necessarily implies that $\frac{c^j(w,t)}{w(t)} \to 0$ almost surely, or agent j will fail to survive in the long run.[33] For this reason, parameters I^i are referred to as survival indices. By performing this pairwise comparative analysis for all agents in this growing economy, Yan (2008) obtains a necessary condition for long-run survival in this economy.

$$\lim_{t\to\infty} \frac{c^i(w,t)}{w(t)} \neq 0 \Longrightarrow i \in \arg\min_j \{I^j\}. \qquad (2.60)$$

Any agent i who survives in the long run must have minimum survival index among all agents. Clearly, either high risk aversion (large γ^i), impatience (large δ^i) or pessimism (large θ^i) will contribute negatively to the market selection of an agent. On top of these, the economy's strong growth (large positive $\mu^w - \frac{(\sigma^w)^2}{2}$) also fastens the extinction process for those who are not fit to survive. This is because, the statistically growing economies do not reward these characteristics of "reservation" nature in the long run.[34] We note that this condition however is not strictly sufficient for survival. Consider the case where there are several agents i, j all having minimum index $I^i = I^j = I_{min}$. In the limit of $t \to \infty$, standard Brownian motion $Z(t) \to \pm\infty$ with equal probability (a well-known non-stationarity problem). (2.58) then implies additionally that only agents having extremum (minimum or maximum) value

[33]Here any agent i's long-run survival definition is that his consumption ratio $\frac{c^i(w,t)}{w(t)}$ does not tend to zero in the limit of large t.

[34]For example, more risk-loving agents have lower EIS, defer more consumption and invest more in risky equity relatively. When economy grows steadfastly, the stock market pays off well, and these agents quickly dominate the economy. The rate of their ascent increases with the economy's growth rate.

of $\theta^j + \gamma^j \sigma^w$ (among agents with minimum survival index) survive. This observation allows us to deduce a more elaborated set of necessary conditions, that also connect well with our analysis of the isomorphic economy. Namely, common to all agents i who survive, there exist two constants K, L such that

$$\begin{cases} \delta^i + \frac{(\theta^i)^2}{2} + \gamma^i \left(\mu^w - \frac{(\sigma^w)^2}{2}\right) = K, \\ \theta^i + \gamma^i \sigma^w = L, \end{cases} \qquad \forall i. \tag{2.61}$$

These necessary conditions are none other than the linearity sufficient conditions for the existence of the reduced economy. The immediate conclusion is that the set of survival agents implies the existence of the exact isomorphic economy. To put it in another way, ultimately all heterogeneous-agent economies specified in this section can be exactly reduced to its simpler isomorphic version, when all agents differ only in their risk aversion.[35] Furthermore, in this case the reduced economy's supply $\hat{w}$ turns out to be constant, which makes the analysis of co-surviving agents even simpler. In the not-so-long run, the isomorphism does not hold exactly because other agents (who ultimately perish) hang on. Nevertheless, in the current setting with additive utilities, Kogan et al. (2009) show that these agents leave no lingering traces on price dynamics after their consumption shares become negligible. Then as discussed earlier, the linear projection construction (2.54), (2.55) will determine qualitatively the time preference and heterogeneous belief contributions, as well as significantly simplifying the analysis on consumption partition and perhaps the asset price dynamics of the original economy.

We thus show that agent survival implies the existence of an isomorphic economy. But is the converse true, i.e., does isomorphism also imply survival? We recall that isomorphism just requires that the original economy can be reduced to a simpler economy heterogeneous only in risk aversion. Obviously, the latter generally does not imply survival, because both (i) agents are still heterogeneous in risk aversion and (ii) its aggregate endowment $\hat{w}$ can be

[35] In this regard, the special case when only one agent survives is trivial, because he eventually consumes the whole aggregate endowment. For time separable utilities under consideration, the economy will converge to a single-agent economy in all aspects as shown by Kogan et al. (2009).

either growing or shrinking steadily. Thus survival is the stronger concept, and the existence of isomorphic economy does not imply the survival of different agents in general. Only in a special case where the transform $\Upsilon(Z(t), t)$ assumes some particular functional forms, does the isomorphism imply the survival of all agents.

2.8 Conclusion

Finance, and economics more generally, has made great progress utilizing the representative agent model. However, real world agents differ significantly in risk aversion, time preference and beliefs. Moreover, such differences strongly motivate the trades that are made on financial markets, and therefore the behaviors of asset prices.

We analyzed the savings and consumption choices for agents who differ in preferences and beliefs within an economy with a GBM endowment. These choices translate into aggregates, which in turn determine asset price behavior. The most significant results are two remarkable isomorphisms, which may greatly facilitate the study of economies composed of heterogeneous agents. First, when agents differ only in risk aversion, the economy behaves as if all agents were identical to a single *market-equivalent agent* with a derived level of risk aversion. Second, when agents differ in all of risk preferences, time preferences and beliefs about the future growth of the economy, the economy is equivalent to one where all agents differ merely in risk aversion. Combining these two results, despite three dimensions of heterogeneity, the economy operates as if it were homogeneous and composed only of the *market-equivalent agent.*

Surprisingly, the aggregates in the heterogeneous economy, such as the "observed" precautionary savings motive, can lie well outside the behaviors that would be observed were the economy composed of any possible one of its constituent types of agents. That is because the dynamic risk sharing and trading of assets among types as the economy incurs shocks are of a stochastic nature. Low real interest rates, equivalent to those observed in the real world, can be achieved with reasonable risk aversions for all individual agents, given that large aggregate precautionary savings motives are feasible in equilibrium. However, such large savings motives tend to imply large savings cyclicality, which in turn generates unrealistic levels of interest rate volatility. (We show that such volatility can be dampened by heterogeneity in time preference.) Savings cyclicality also influences stock prices and volatility, as is demonstrated.

To move from the heterogeneity in all of risk aversion, time preference and beliefs to those merely on risk aversion, that is to dramatically reduce the dimensions of the problem, requires merely modifying the mean and volatility of the endowment process. We expect this insight to make future investigations of heterogeneities much more tractable.

The risk tolerance measure proves to be an extraordinarily versatile tool quantifying how individuals share risk and how resulting aggregate behaviors response to growth shocks. The sensitivities to these shocks (i.e., derivatives) of risk tolerance reveal how agents are jostled in their weightings within the economy as uncertainties unfold. Conveniently, these derivatives prove to be simple functions of individuals' risk aversion, prudence and temperance. This property allows us to obtain interesting and analytical bounds on asset return volatilities.

The principal risk that we face in the modern economy, as we witnessed in recent years, is the movement of asset prices within the economy. This analysis traced how agents who differ on preferences and beliefs trade amongst themselves to simultaneously hedge against, capitalize on and generate such movements. Most important, it showed that those tracings prove tractable.

2.9 Appendices

We recall that subscripts always denote partial derivatives; $f_x \equiv \frac{\partial f}{\partial x}$ throughout the paper.

2.9.1 Proofs concerning risk tolerance measure

Preliminary derivations

Derivation of key eq. (2.7): Using FOC (2.5) we have

$$v_{ww} = \frac{1}{\lambda^i} u^i_{cc} c^i_w.$$

Plugging FOC (2.5) and above eq. into the expression for market-revealed risk tolerance (2.6)

$$T \equiv \frac{-v_w}{v_{ww}} = \frac{-1}{\lambda^i} \frac{u^i_c}{v_{ww}} = \frac{-u^i_c}{u^i_{cc} c^i_w} \Rightarrow c^i_w = \frac{-u^i_c / u^i_{cc}}{-v_w / v_{ww}} = \frac{T^i}{T},$$

which is (2.7).

Derivation of eqs. (2.9), (2.10): Using $p^i = T^i/T = c^i_w$ we have

$$p^i_w = \frac{T^i_w}{T} - \frac{T^i}{T} \frac{T_w}{T} = \frac{1}{T}\left(T^i_c c^i_w - \frac{T^i}{T} T_w \right) = \frac{T^i}{T^2}\left(T^i_c - T_w \right) = \frac{p^i}{T}\left(T^i_c - T_w \right), \tag{2.62}$$

which is (2.9). In the CRRA settings, $T^i = \frac{c^i}{\gamma^i} \Rightarrow T^i_c = \frac{1}{\gamma^i}$, and

$$T_w = \frac{\partial}{\partial_w} \sum_i T^i = \sum_i T^i_c c^i_w = E_{\{p^i\}}[T^i_c] = E_{\{p^i\}}\left[\frac{1}{\gamma^i} \right].$$

now eq. (2.9) becomes (2.10).

Derivation of eqs. (2.12), (2.14): Taking the partial derivative $\frac{\partial}{\partial w}$ of risk aversion $R = \sum_i \frac{\gamma^i T^i}{T}$

$$R_w = \frac{1}{T}\left[\sum_i \gamma^i T^i_c c^i_w - \left(\sum_i \gamma^i \frac{T^i}{T} \right) \left(\sum_i T^i_c c^i_w \right) \right] = \frac{1}{T} Cov_{\{p^i\}}\left(\gamma^i, T^i_c \right) = \frac{1}{T} Cov_{\{p^i\}}\left(\gamma^i, \frac{1}{\gamma^i} \right). \tag{2.63}$$

where we have used (2.7) $c^i_w = \frac{T^i}{T} \equiv p^i$, and in the last equality CRRA utility's property $T^i_c = \frac{\partial T^i}{\partial c^i} = \frac{1}{\gamma^i}$.

Taking Ito differential on both sides of $\delta = \sum_i \frac{T^i \delta^i}{T}$, then identifying diffusion and drift parts gives

$$\delta_w = \frac{1}{T}\sum_i(\delta^i-\delta)T_c^i c_w^i = \frac{1}{T}\sum_i(\delta^i-\delta)T_c^i\frac{T^i}{T} = \frac{1}{T}Cov_{\{p^i\}}(T_c^i,\delta^i), \tag{2.64}$$
$$\delta_t = \frac{1}{T}\sum_i(\delta^i-\delta)T_c^i c_t^i = \frac{-1}{T}\sum_i(\delta^i-\delta)^2 T_c^i T^i = \frac{-1}{T}\sum_i\frac{(\delta^i-\delta)^2}{\gamma^i}T^i,$$

where again the last equality holds for CRRA utilities: $T_c^i = \frac{1}{\gamma^i}$. These concise expressions capture and generalize key results on the behaviors of social discount rate first obtained in Gollier and Zeckhauser (2005) to stochastic environments.

Precautionary savings (prudence) $P \equiv \frac{-wv_{www}}{v_{ww}}$, *temperance* $Q \equiv \frac{-wv_{wwww}}{v_{www}}$ *and their relations:* Taking the partial derivative $\frac{\partial}{\partial w}$ of risk tolerance $T = \frac{-v_w}{v_{ww}}$

$$T_w = -1+\frac{v_w v_{www}}{v_{ww}^2} = -1+\frac{-v_w}{wv_{ww}}\frac{-wv_{www}}{v_{ww}} = -1+\frac{P}{R} \Rightarrow P = (1+T_w)R = \frac{(1+T_w)w}{T}. \tag{2.65}$$

Similarly, since $R = \frac{w}{T}$, and using above expression for P yields a general relation for any time separable utilities (possibly non CRRA)

$$R_w = \frac{1}{T}\left(1-\frac{wT_w}{T}\right) = \frac{R}{w}(1+R-P), \tag{2.66}$$

which together with (2.40) implies (2.17), (2.41). Combining (2.63), (2.17), we have in CRRA setting

$$R_w = \frac{R}{w}(1+R-P) = \frac{1}{T}Cov_{\{p^i\}}\left(\gamma^i, T_c^i\right). \tag{2.67}$$

Very similar to (2.66), we also have in the general case

$$P_w = \frac{\partial}{\partial w}\frac{-wv_{www}}{v_{ww}} = \frac{-v_{www}}{v_{ww}} + \frac{wv_{www}^2}{v_{ww}^2} - \frac{wv_{wwww}}{v_{ww}} = \frac{P}{w}(1+P-Q). \tag{2.68}$$

Next, taking one more time the partial derivative on T_w in (2.65)

$$T_{ww} = \frac{v_{ww}v_{www}}{v_{ww}^2} + \frac{v_w v_{wwww}}{v_{ww}^2} - 2\frac{v_w v_{www}^2}{v_{ww}^3} = \frac{P}{w}\frac{(2P-R-Q)}{R}. \tag{2.69}$$

Plugging R_w (2.66) and P_w (2.68) into σ_Γ^r in (2.29) we obtain

$$\sigma_\Gamma^r = -\sigma^w R\left[\mu^w(P-R-1)+(\sigma^w)^2 P\left(1+\frac{R-Q}{2}\right)\right]. \tag{2.70}$$

which proves (2.30).

Derivation of eq. (2.15), (2.21): The derivation of the key aggregate relation (2.65) $P = R + RT_w$ must also hold at individual level[36] $P^i = R^i + R^i T_c^i$. Computing the latter's mean in risk-tolerance measure (that is, $E_{\{p^i\}}[X] = \sum_i \frac{T^i}{T} X^i$), and taking the difference with the former

$$\begin{aligned} P &= E_{\{p^i\}}[P^i] + RT_w - E_{\{p^i\}}\left[R^i T_c^i\right] = E_{\{p^i\}}\left[P^i\right] + R\sum_i T_c^i c_w^i - E_{\{p^i\}}\left[R^i T_c^i\right] \\ &= E_{\{p^i\}}\left[P^i\right] + E_{\{p^i\}}\left[R^i\right] E_{\{p^i\}}\left[T_c^i\right] - E_{\{p^i\}}\left[R^i T_c^i\right] = E_{\{p^i\}}\left[P^i\right] - Cov_{\{p^i\}}\left(R^i, T_c^i\right) \\ &= E_{\{p^i\}}\left[P^i\right] - Cov_{\{p^i\}}\left(R^i, \frac{1}{R^i} - \frac{c^i R_c^i}{(R^i)^2}\right). \end{aligned}$$

where in the last equality we have used $T^i = \frac{c^i}{R^i}$. This is (2.15). In the special case when all agents have CRRA utilities, $R^i = \gamma^i$, $R_c^i = 0\ \forall i$, $P^i = \gamma^i + 1$, the market-revealed prudence is simplified to

$$P = E_{\{p^i\}}\left[P^i\right] - Cov_{\{p^i\}}\left(\gamma^i, \frac{1}{\gamma^i}\right) = E_{\{p^i\}}\left[\gamma^i\right]\left(1 + E_{\{p^i\}}\left[\frac{1}{\gamma^i}\right]\right) \geq E_{\{p^i\}}\left[P^i\right]. \tag{2.71}$$

Same technique can be used on temperances (see (2.69)) $Q = 2P - R - wT_{ww}\frac{R}{P}$ and $Q^i = 2P^i - R^i - c^i T_{cc}^i \frac{R^i}{P^i}$

$$Q = E_{\{p^i\}}[Q^i] + 2\left(P - E_{\{p^i\}}[P^i]\right) - wT_{ww}\frac{R}{P} + E_{\{p^i\}}\left[c^i T_{cc}^i \frac{R^i}{P^i}\right].$$

[36]We can obtain result at individual level from aggregate result in economy with only a single agent.

First note that we can derive an agent-based sufficient condition for the convexity of market-revealed precautionary savings

$$
\begin{aligned}
T_w = \sum_i T_c^i c_w^i \Rightarrow T_{ww} &= \sum_i T_{cc}^i (c_w^i)^2 + \sum_i T_c^i c_{ww}^i \\
&= \sum_i T_{cc}^i (c_w^i)^2 + \sum_i T_c^i \frac{\partial}{\partial w}\left(\frac{T^i}{T}\right) \\
&= \sum_i T_{cc}^i (c_w^i)^2 + \frac{1}{T}\left[\sum_i (T_c^i)^2 \frac{T^i}{T} - \left(\sum_i T_c^i \frac{T^i}{T}\right)^2\right] \\
&= \sum_i T_{cc}^i (c_w^i)^2 + \frac{1}{T} Var_{\{p^i\}}(T_c^i).
\end{aligned}
$$

Consequently, when $T_{cc}^i \geq 0\ \forall i$, we also have $T_{ww} \geq 0$. This aggregation property echoes a similar result of proposition 9. Now plugging T_{ww} into above Q, we have

$$
\begin{aligned}
Q &= E_{\{p^i\}}[Q^i] + 2\left(P - E_{\{p^i\}}[P^i]\right) - \frac{wR}{P}\sum_i T_{cc}^i \frac{(T^i)^2}{T^2} - \frac{wR}{PT} Var_{\{p^i\}}(T_c^i) + E_{\{p^i\}}\left[c^i T_{cc}^i \frac{R^i}{P^i}\right] \\
&= E_{\{p^i\}}[Q^i] + 2\left(P - E_{\{p^i\}}[P^i]\right) - \frac{wR}{PT} Var_{\{p^i\}}(T_c^i) + E_{\{p^i\}}\left[T^i T_{cc}^i\left(\frac{(R^i)^2}{P^i} - \frac{R^2}{P}\right)\right] \\
&= E_{\{p^i\}}[Q^i] - 2Cov_{\{p^i\}}\left(R^i, \frac{1}{R^i} - \frac{c^i R_c^i}{(R^i)^2}\right) - \frac{R^2}{P} Var_{\{p^i\}}(T_c^i) + E_{\{p^i\}}\left[T^i T_{cc}^i\left(\frac{(R^i)^2}{P^i} - \frac{R^2}{P}\right)\right].
\end{aligned}
$$

In the special case when all agents have CRRA utilities, $R^i = \gamma^i$, $R_c^i = 0$, $T_{cc}^i = 0\ \forall i$, the market-revealed temperance is simplified to (2.21)

$$
Q = E_{\{p^i\}}[Q^i] - 2Cov_{\{p^i\}}\left(\gamma^i, \frac{1}{\gamma^i}\right) - \frac{R^2}{P} Var_{\{p^i\}}\left(\frac{1}{\gamma^i}\right). \tag{2.72}
$$

Derivation of Hansen-Jagannathan bound (2.26): Let $S(w,t)$ be price of the contingent claim (i.e., stock) on the dividend stream,

$$
\begin{aligned}
S(w,t) &= E_t\left[\tfrac{M(t+dt)}{M(t)}\{S(w+dw, t+dt) + dw\}\right] \\
&\Rightarrow E_t\left[\tfrac{M(t+dt)}{M(t)}\tfrac{S(w+dw,t+dt)+dw}{S(w,t)}\right] = 1.
\end{aligned}
$$

Next, since $M(t+dt) = M(t)[1 - r(w,t)dt - \eta(w,t)dZ(t)]$, up to order dt we have

$$E_t\left[\frac{M(t+dt)}{M(t)}\{1 + r(w,t)dt\}\right] = 1.$$

Combining these identities yields

$$E_t\left[\frac{M(t+dt)}{M(t)}\left(\frac{S(w+dw,t+dt)+dw}{S(w,t)} - 1 - r(w,t)dt\right)\right] = 0$$

where $\frac{S(w+dw,t+dt)+dw}{S(w,t)} - 1 - r(w,t)dt$ is simply the stock excess return. Standard argument that the absolute value of correlation between this and the stochastic discount factor $\frac{M(t+dt)}{M(t)}$ is less than unity implies (after plugging in (i) the mean value $1-rdt$ and standard deviation $\eta\sqrt{dt}$ of $\frac{M(t+dt)}{M(t)}$, (ii) the expected stock excess return $E_t\left[\frac{S(w+dw,t+dt)+dw}{S(w,t)} - 1 - r(w,t)dt\right] = (\mu^s - r)dt$ by virtue of gain, and (iii) the notation $\sigma_{\mu^s - r}dt$ for stock excess return volatility)

$$\eta\sqrt{dt} \geq [1 - r(w,t)dt]\frac{|\mu^s(w,t) - r(w,t)|\,dt}{\sigma_{\mu^s-r}dt}$$

Finally, to use annual data, we somewhat coarsely set $dt = 1$. Since the expected stock excess return is positive, this is precisely the bound (2.26).

Proofs of propositions

Proof of proposition 4. Market-revealed risk tolerance: since $\sum_i c^i = w \rightarrow \sum_i c^i_w = 1$.

$$\sum_i T^i = T \quad \text{or} \quad \sum_i p^i \equiv \sum_i \frac{T^i}{T} = 1.$$

Market-revealed risk aversion

$$R \equiv -w\frac{v_{ww}}{v_w} = -w\frac{\frac{1}{\lambda^i}u^i_{cc}c^i_w}{\frac{1}{\lambda^i}u^i_c} = -w\frac{u^i_{cc}}{u^i_c}\frac{T^i}{T} = \sum_i \frac{-c^i u^i_{cc}}{u^i_c}\frac{T^i}{T} = \sum_i \frac{T^i R^i}{T}.$$

Market-revealed discount factor

$$\delta \equiv -\frac{v_{wt}}{v_w} = -\frac{u^i_{ct} + u^i_{cc}c^i_t}{u^i_c} = \sum_i \frac{T^i}{T}\left(\frac{-u^i_{ct}}{u^i_c} + \frac{c^i_t}{T^i}\right) = \sum_i \frac{T^i\delta^i}{T} + \frac{\sum_i c^i_t}{T} = \sum_i \frac{T^i\delta^i}{T},$$

because $\sum_i c^i_t = \frac{\partial(\sum_i c^i)}{\partial t} = \frac{\partial w}{\partial t} = 0$ as aggregate endowment w and time t are two independent variables. ■

Proof of proposition 5.

$$\begin{aligned}
\frac{\partial E_{\{p^i\}}[a^i]}{\partial w} &= \frac{\partial}{\partial w}\sum_i a^i p^i = \sum_i a^i_w p^i + \sum_i a^i p^i_w \\
&= \sum_i a^i_w p^i + \sum_i a^i \frac{p^i_w}{p^i} p^i - \sum_i a^i p^i \sum_j \frac{p^j_w}{p^j} p^j + \sum_i a^i p^i \sum_j \frac{p^j_w}{p^j} p^j \\
&= E_{\{p^i\}}[a^i_w] + Cov_{\{p^i\}}\left(a^i, \frac{p^i_w}{p^i}\right) + \sum_i a^i p^i \sum_j \frac{p^j_w}{p^j} p^j \\
&= E_{\{p^i\}}[a^i_w] + Cov_{\{p^i\}}\left(a^i, \frac{p^i_w}{p^i}\right) + E_{\{p^i\}}[a^i]\frac{\partial}{\partial w}\sum_j p^j = E_{\{p^i\}}[a^i_w] + Cov_{\{p^i\}}\left(a^i, \frac{p^i_w}{p^i}\right).
\end{aligned}$$

The last equality holds because $\sum_j p^j = 1$, and hence term $E_{\{p^i\}}[a^i]\frac{\partial}{\partial w}\sum_j p^j = 0$. ■

Proof of proposition 6. For CRRA utilities, eq. (2.71) shows that market-revealed prudence P is always larger or equal average prudence $E_{\{p^i\}}[P^i]$ under risk tolerance measure $\{p^i = \frac{T^i}{T}\}$. In the case of 2-CRRA economy ($i = A, B$) (and assume without loss of generality throughout that $\gamma^A < \gamma^B$), plugging $P^i = \gamma^i + 1$ into (2.71)

$$\begin{aligned}
P &= E_{\{p^i\}}\left[\gamma^i + 1\right] - 1 + E_{\{p^i\}}\left[\gamma^i\right] E_{\{p^i\}}\left[\frac{1}{\gamma^i}\right] = E_{\{p^i\}}\left[\gamma^i\right]\left(1 + E_{\{p^i\}}\left[\frac{1}{\gamma^i}\right]\right) \\
&= \left(p^A\gamma^A + (1-p^A)\gamma^B\right)\left(1 + \frac{p^A}{\gamma^A} + \frac{1-p^A}{\gamma^B}\right). \qquad (2.73)
\end{aligned}$$

Precautionary savings P is an explicit concave quadratic function of p^A. Theoretically,[37] it obtains maximum value

$$P^* \equiv \max P = \frac{(\gamma^A + \gamma^B + \gamma^A\gamma^B)^2}{4\gamma^A\gamma^B} \quad \text{at} \quad p^{A*} = \frac{1}{2} + \frac{1}{2}\frac{\gamma^A\gamma^B}{\gamma^A - \gamma^B}. \qquad (2.74)$$

[37] This is indeed the legitimate maximum when the corresponding argmax $p^{A*} \in [0, 1]$.

Evidently, when $\frac{\gamma^B}{\gamma^B+1} \geq \gamma^A$, $p^{A*} \in [0, 1]$ and the above value P^* is indeed market-revealed prudence's legitimate maximum. Furthermore in this case, market-revealed prudence $P(p^A)$ is larger than the largest individual prudence (which is agent B's under current convention) $P^B = \gamma^B + 1$ for all $0 \leq p^A \leq 2p^{A*} = 1 + \frac{\gamma^A\gamma^B}{\gamma^A - \gamma^B}$. However, when $\frac{\gamma^B}{\gamma^B+1} < \gamma^A$, $p^{A*} < 0$, the market-revealed prudence's legitimate maximum is $P^* = P^B = \gamma^B + 1$, which is attained at $p^{A*} = 0$. ■

Proof of proposition 7. For CRRA utilities ($Q^i = \gamma^i + 2$), from eqs. (2.72) and (2.71)

$$\begin{aligned} Q &= R + 2RE_{\{p^i\}}\left[\frac{1}{\gamma^i}\right] - \frac{R^2}{P}\left(E_{\{p^i\}}\left[\frac{1}{(\gamma^i)^2}\right] - \left(E_{\{p^i\}}\left[\frac{1}{\gamma^i}\right]\right)^2\right) \\ &= R\left(1 + 2E_{\{p^i\}}\left[\frac{1}{\gamma^i}\right] + \frac{\left(E_{\{p^i\}}\left[\frac{1}{\gamma^i}\right]\right)^2 - E_{\{p^i\}}\left[\frac{1}{(\gamma^i)^2}\right]}{1 + E_{\{p^i\}}\left[\frac{1}{\gamma^i}\right]}\right) \\ &= E_{\{p^i\}}[\gamma^i]\left(3E_{\{p^i\}}\left[\frac{1}{\gamma^i}\right] + \frac{1 - E_{\{p^i\}}\left[\frac{1}{(\gamma^i)^2}\right]}{1 + E_{\{p^i\}}\left[\frac{1}{\gamma^i}\right]}\right). \end{aligned}$$

Next, using (2.68) $P_w = \frac{P(1+P-Q)}{w}$ we see that $Q > P + 1$ if and only if $P_w \leq 0$. Specializing in the 2-CRRA economy, we have

$$P_w = \frac{\partial P}{\partial p^A}\frac{\partial p^A}{\partial w} = (\gamma^A - \gamma^B)\left(1 + \frac{p^A - p^B}{\gamma^A} + \frac{p^B - p^A}{\gamma^B}\right)\frac{T^A T^B}{T^3}\frac{\gamma^B - \gamma^A}{\gamma^A\gamma^B},$$

where we have used the explicit expressions for P (2.73) and p^A_w (2.62). It is now clear that $P_w \leq 0$, or equivalently $Q > P + 1$, if and only if (note that $p^A + p^B = 1$ and we have assumed $\gamma^A < \gamma^B$ throughout)

$$1 + \frac{p^A - p^B}{\gamma^A} + \frac{p^B - p^A}{\gamma^B} \geq 0 \Leftrightarrow p^A \geq p^{A*} \equiv \frac{1}{2} + \frac{1}{2}\frac{\gamma^A\gamma^B}{\gamma^A - \gamma^B}.$$

We note that when $\frac{\gamma^B}{\gamma^B+1} < \gamma^A$, $p^{A*} < 0$. In this case we simply have $Q > P + 1$ for all $p^A > 0$. The value $p^{A*} \equiv \frac{1}{2} + \frac{1}{2}\frac{\gamma^A\gamma^B}{\gamma^A - \gamma^B}$ is also where the market-revealed precautionary savings P attains maximum (see (2.74)). ■

Proof of proposition 8. This proposition holds on the premise of the large precautionary savings $P > \frac{2\mu^w}{(\sigma^w)^2}$ (2.25) needed for the observed low real interest rate.

Case $Q > \frac{2\mu^w}{(\sigma^w)^2} + R + 2$: we first rewrite (2.30) as

$$\sigma_\Gamma^r = \mu^w \sigma^w R \left(\left[\frac{(\sigma^w)^2}{2\mu^w}(Q - R - 2) - 1 \right] P + R + 1 \right).$$

Since the expression inside square brackets is positive in the current case, large precautionary savings (2.25) implies

$$\begin{aligned} \sigma_\Gamma^r &> \mu^w \sigma^w R \left(\left[\frac{(\sigma^w)^2}{2\mu^w}(Q - R - 2) - 1 \right] \frac{2\mu^w}{(\sigma^w)^2} + R + 1 \right) \\ &= \mu^w \sigma^w R \left(Q - 1 - \frac{2\mu^w}{(\sigma^w)^2} \right) \approx \mu^w \sigma^w R \left(Q - \frac{2\mu^w}{(\sigma^w)^2} \right), \end{aligned}$$

which is (2.32) (the last approximation is from the conditions $Q - \frac{2\mu^w}{(\sigma^w)^2} > R + 2$ and bound (2.27)).[38]

Case $Q < \frac{2\mu^w}{(\sigma^w)^2}$: we first rewrite (2.30) as

$$\sigma_\Gamma^r = \sigma^w R \frac{(\sigma^w)^2}{2} \left(P \left[Q - \frac{2\mu^w}{(\sigma^w)^2} \right] + R \left[\frac{2\mu^w}{(\sigma^w)^2} - P \right] + \left[\frac{2\mu^w}{(\sigma^w)^2} - 2P \right] \right).$$

In the current case, all three expressions inside square brackets are negative under large precautionary savings condition (2.25), and thus

$$\sigma_\Gamma^r < \sigma^w R \frac{(\sigma^w)^2}{2} P \left[Q - \frac{2\mu^w}{(\sigma^w)^2} \right] < \mu^u \sigma^w R \left[Q - \frac{2\mu^w}{(\sigma^w)^2} \right],$$

which is (2.33) (the last inequality is again from the conditions (2.25)). ■

Proof of proposition 9. First we note from (2.65) that $P = (1 + T_w)R$, which implies

$$P \geq R + 1 \Leftrightarrow T_w \geq \frac{1}{R} = \frac{T}{w} \Leftrightarrow wT_w \geq T; \quad \text{similarly } P^i \geq R^i + 1 \Leftrightarrow c^i T_c^i \geq T^i. \qquad (2.75)$$

[38] In the same approximation, in the statement of proposition 8 we write $Q > \frac{2\mu^w}{(\sigma^w)^2} + R$ in place of $Q > \frac{2\mu^w}{(\sigma^w)^2} + R + 2$. Practically, the difference is non-material by virtue of empirically large value $\frac{2\mu^w}{(\sigma^w)^2} \sim 100$.

Next, since $T = \sum_i T^i$ and $c_w^i = \frac{T^i}{T}$

$$\begin{aligned} wT_w - T = w \sum_i T_c^i c_w^i - T &= \frac{\left(\sum_i c^i\right)\left(\sum_i T_c^i T^i\right) - T^2}{T} \\ &\geq \frac{\left(\sum_i \sqrt{c^i T_c^i T^i}\right)^2 - T^2}{T} \geq \frac{\left(\sum_i T^i\right)^2 - T^2}{T} = 0, \end{aligned}$$

where the first inequality is an application of Cauchy-Schwarz's, the second arises from the proposition's hypothesis (2.75). Now $wT_w - T \geq 0$ is equivalent to $P \geq R + 1$ again by virtue of (2.75). ■

2.9.2 Proofs concerning asset return volatilities

Preliminaries:

When θ is a continuously differentiable function of the underlying Brownian motion Z, the Malliavin derivative $\mathcal{D}_t\theta$ is the deviation in θ due to change in the path of Z starting at t. The Malliavin calculus is a handy tool to study stock return volatilities. We adopt this tool here along the presentation of Detemple et al. (2003) and Bhamra and Uppal (2009). More extensive exposition of this powerful tool can be found in Nualart (2006). We first state two useful results for our proofs.

Result 1: *Let $\beta(t)$ be a general GBM process with bounded drift and diffusion*

$$\frac{d\beta(t)}{\beta(t)} = \mu(\beta, t)dt + \sigma(\beta, t)dZ(t) \quad \textit{where} \quad |\mu(\beta, t)|,\ |\sigma(\beta, t)| < \infty \quad \textit{almost surely.} \tag{2.76}$$

Then the process $\beta(t)$ never changes its sign

$$\beta(t)\beta(s) \geq 0 \qquad \forall t, s \quad \textit{almost surely.} \tag{2.77}$$

Result 2: *Let $\theta(t)$ be a general diffusion process*

$$d\theta(t) = \mu(\theta, t)dt + \sigma(\theta, t)dZ(t), \tag{2.78}$$

then under regularity conditions the Malliavin derivative $\Theta(\tau) \equiv \mathcal{D}_t\theta(\tau)$ of process $\theta(t)$ is a generalized GBM process with specified initial value

$$\frac{d\Theta(\tau)}{\Theta(\tau)} = \mu_\theta(\theta, \tau)d\tau + \sigma_\theta(\theta, \tau)dZ(\tau); \qquad \Theta(t) = \sigma(\theta, t). \tag{2.79}$$

Note that subscript θ in μ_θ, σ_θ always denotes the partial derivative and Malliavin derivative $\mathcal{D}_t\theta(\tau)$ is a process with respect to the ulterior time τ, and thus is defined only for $\tau \geq t$. This result makes clear the relation between diffusion of a process and its Malliavin derivative. More specifically,

$$\mathcal{D}_t\theta(\tau) = \Theta(\tau) = \sigma(\theta, t)\exp\left\{\int_t^\tau \left(\mu_\theta(\theta, u) - \frac{1}{2}\sigma_\theta^2(\theta, u)\right)du + \int_t^\tau \sigma_\theta(\theta, u)dZ(u)\right\}. \tag{2.80}$$

In particular they are identical when the Malliavin derivative is contemporaneous, $\mathcal{D}_t\theta(t) = \sigma(\theta, t)$.

In case of two-CRRA-agent economies, working with first agent's risk tolerance measure p^A is also convenient for our technical proofs. Applying Ito lemma on $p^A = \frac{T^A}{T}$ yields the dynamics of this state variable Indeed, the general volatility σ^{pA} and drift μ^{pA} of this state variable's diffusion process

$$\begin{aligned} \frac{dp^A(w,t)}{p^A(w,t)} &= \mu^{pA}(p^A)dt + \sigma^{pA}(p^A)dZ(t), \\ \sigma^{pA}(p^A) &= \sigma^w R p^B\left(\frac{1}{\gamma^A} - \frac{1}{\gamma^B}\right), \\ \mu^{pA}(p^A) &= p^B\left[-\frac{R(\delta^A - \delta^B)}{\gamma^A\gamma^B} + R\mu^w\left(\frac{1}{\gamma^A} - \frac{1}{\gamma^B}\right) + (\sigma^w)^2\left(\frac{1}{\gamma^A} - \frac{1}{\gamma^B}\right)\left(\frac{R}{2\gamma^A\gamma^B} - \frac{p^A}{\gamma^A} - \frac{p^B}{\gamma^B}\right)\right]. \end{aligned} \tag{2.81}$$

where $p^B = 1 - p^A$, and $R(p^A) = p^A\gamma^A + p^B\gamma^B$ is the aggregate risk aversion in (2.6). We now proceed to the proofs.

Derivation of mpr volatility (2.41): plugging R_w in (2.17) into (2.40), we immediately obtain (2.41).

Derivation of eq. (2.38): Taking the Malliavin derivative $\mathcal{D}_t$ in measure $\mathcal{Q}$ of both sides of eq. (2.36) yields

$$\sigma^{sQ}(w,t)S(w,t)e^{-\int_0^t r(s)ds} = E_t^Q\left[\mathcal{D}_t G(t,T)\right], \tag{2.82}$$
$$G(t,T) \equiv \int_t^T e^{-\int_0^u r(\tau)d\tau} w(u)du,$$

where σ^{sQ} is the stock return volatility in measure Q. The diffusion invariance principle $\sigma^{sQ} = \sigma^s$ justifies the drop of superscript Q hereafter. Using the explicit aggregate endowment process (2.1) in measure Q

$$w(t) = w(0)\exp\left[\left(\mu^w - \frac{(\sigma^w)^2}{2}\right)t + \sigma^w Z^Q(t) - \sigma^w \int_0^t \eta(w,u)du\right],$$

and the chain rule we obtain Malliavin derivative

$$\mathcal{D}_t G(t,T) = \int_t^T du\, w(u)e^{-\int_0^u r(\tau)d\tau}\left\{\sigma^w - \sigma^w \int_t^u d\tau\, \mathcal{D}_t\eta(w,\tau) - \int_t^u d\tau\, \mathcal{D}_t r(w,\tau)\right\}.$$

Plugging above $\mathcal{D}_t G(t,T)$ into eq. (2.82) we get the excess volatility of stock return (2.38).

Derivation of eq. (2.39): Let's define

$$\theta(w,t) \equiv \sigma^w\eta(w,t) + r(w,t); \qquad d\theta = \mu^\theta dt + \sigma^\theta dZ(t).$$

From (2.38), it is clear that $\mathcal{D}_t\theta(w,\tau) < 0\ \forall\tau \geq t$ implies positive stock return excess volatility $\sigma^s > \sigma^w$. In light of Result 2 above, this Malliavin derivative is a generalized Brownian motion, and Result 1 implies that it will remain negative at all time if all following conditions hold.

1. Diffusion $\sigma^\theta_\theta \equiv \frac{\partial\sigma^\theta}{\partial\theta}$ is bounded. Indeed this is the case. In the current two-CRRA-agent setting, δ, R, P are simple polynomials of p^A, and so are r, η in (2.24), and also θ and $\sigma^\theta = [\partial(r + \sigma^w\eta)/\partial p^A]\sigma^{pA}$ by virtue of (2.81). Then the next-generation partial derivatives $\theta_{p^A} \equiv \frac{\partial\theta}{\partial p^A}$ and $\sigma^\theta_{p^A} \equiv \frac{\partial\sigma^\theta}{\partial p^A}$ are also simple polynomials of p^A. These in turn imply $\sigma^\theta_\theta = \frac{\partial\sigma^\theta}{\partial\theta} = \frac{\sigma^\theta_{pA}}{\theta_{pA}}$ is bounded almost surely because p^A is in $(0,1)$.

2. Drift $\mu_\theta^\theta \equiv \frac{\partial \mu^\theta}{\partial \theta}$ is bounded. This holds by identical reasoning.

3. Initial value $\mathcal{D}_t\theta(w,\tau)|_{\tau=t} < 0$. Note that because $\tau = t$, this Malliavin derivative is simply the volatility $\sigma^\theta = [\partial(r+\sigma^w\eta)/\partial p^A]\sigma^{pA}$. From (2.81), σ^{pA} is always positive for our convention $\gamma^A < \gamma^B$, then this last condition is precisely the required sufficient condition (2.39).

Derivation of eq. (2.42): In 2-agent economy, we can work with risk tolerance measure $p^A \equiv \frac{T^A}{T}$ as key underlying state variable. Using (2.24)

$$\begin{aligned}\frac{\partial(r+\sigma^w\eta)}{\partial p^A} &= \frac{1}{p_w^A}(r_w + \sigma^w\eta_w) = \frac{1}{p_w^A}\left[\delta_w + (\sigma^w)^2\left(\frac{\mu^w}{(\sigma^w)^2}+1-\frac{P}{2}\right)R_w - \frac{1}{2}(\sigma^w)^2 RP_w\right] \quad (2.83)\\ &= \frac{1}{p_w^A}\frac{R}{w}\left[Cov_{\{p^i\}}\left(\delta^i,\frac{1}{\gamma^i}\right) + (\sigma^w)^2\left(\frac{\mu^w}{(\sigma^w)^2}+1-\frac{P}{2}\right)(R+1-P) - \frac{1}{2}(\sigma^w)^2 P(1+P-Q)\right]\\ &= \frac{1}{p_w^A}\frac{R}{w}\left[Cov_{\{p^i\}}\left(\delta^i,\frac{1}{\gamma^i}\right) + (\sigma^w)^2\left(\left\{\frac{\mu^w}{(\sigma^w)^2}+1\right\}\{R+1\} - \frac{\mu^w P}{(\sigma^w)^2} + \frac{P(Q-R-4)}{2}\right)\right].\end{aligned}$$

where the second equality arises from (2.64), (2.66), (2.68). Next, since $\sigma^{pA} = p_w^A\sigma^w$, together with convention $\gamma^A < \gamma^B$ and (2.81) we have $p_w^A > 0$. From (2.83), the derivative $\frac{d(r+\sigma^w\eta)}{dp^A}$ in (2.39) is negative only if the expression in square brackets is negative

$$(\sigma^w)^2\left(\left\{\frac{\mu^w}{(\sigma^w)^2}+1\right\}\{R+1\} - \frac{\mu^w P}{(\sigma^w)^2} + \frac{P(Q-R-4)}{2}\right) < -Cov_{\{p^i\}}\left(\delta^i,\frac{1}{\gamma^i}\right).$$

For empirically reasonable values of aggregate consumption moments $\mu^w \sim 2\%$, $\sigma^w \sim 2\%$, we have $\frac{\mu^w}{(\sigma^w)^2} \geq 1$, above condition becomes (2.42). Thus, (2.42) implies (2.39), so it is also a sufficient condition for positive stock return excess volatility.

2.9.3 Proofs concerning heterogeneity transformations

Proof of proposition 10. The multiplicative factor $\Upsilon(Z(t),t)$ (2.47) is required to be able to reduce FOC (2.46) to a simpler FOC (2.49), thus it satisfies

$$e^{-\delta^A t}\xi^A(\Upsilon)^{\gamma^A} = e^{-\delta^B t}\xi^B(\Upsilon)^{\gamma^B}.$$

Let us look for Υ in the form $\exp\left(\beta^{\gamma,\delta}t\right)\exp\left(\beta^{\gamma,\theta}Z(t)\right)$. Plugging in the Radon-Nikodym derivative $\xi^i = e^{-(\theta^i)^2 t/2}e^{-\theta^i Z(t)}$, above eq. becomes

$$\begin{aligned} & \exp\left[\left(\gamma^A\beta^{\gamma,\delta} - \delta^A - \frac{(\theta^A)^2}{2}\right)t\right]\exp\left[\left(\gamma^A\beta^{\gamma,\theta} - \theta^A\right)Z(t)\right] \\ = \; & \exp\left[\left(\gamma^B\beta^{\gamma,\delta} - \delta^B - \frac{(\theta^B)^2}{2}\right)t\right]\exp\left[\left(\gamma^B\beta^{\gamma,\theta} - \theta^B\right)Z(t)\right]. \end{aligned}$$

Identifying the drift and diffusion parts immediately yields $\beta^{\gamma,\delta}$, $\beta^{\gamma,\theta}$ in (2.48). This transformation implements the isomorphism $\{\gamma^1, \gamma^2, \delta^1, \delta^2, \theta^1, \theta^2, w(t)\} \longleftrightarrow \{\gamma^1, \gamma^2, \hat{w}(t)\}$. ■

References

Aarbu, K. O. and F. Schroyen (2009), "Mapping risk aversion in Norway using hypothetical income gambles," Working paper, Dept. of Economics, Norwegian School of Economics and Business Administration.

Alan S. and M. Browning (2010), "Estimating Intertemporal Allocation Parameters using Synthetic Residual Estimation," *Review of Economic Studies*, 77, 12311261.

Bansal, R., and A. Yaron (2004), "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 14811509.

Barsky, R. B., F. T. Juster, M. S. Kimball and M. D. Shapiro (1997), "Preference parameters and behavioral heterogeneity: An experimental approach in the HRS," *Quarterly Journal of Economics*, 112, S537S579.

Basak, Suleyman (2005), "Asset Pricing with Heterogeneous Beliefs," *Journal of Banking and Finance*, 29, 2849-2881.

Brennan, M. J. (1998), "The role of learning in dynamic portfolio decisions," *European Finance Review*, 1, 295-306.

Benninga, S and J. Mayshar (2000), "Heterogeneity and Option Pricing," *Review of Derivatives Research*, 4, 7-27.

Bhamra, Harjoat S. and Raman Uppal (2009), "The Effect of Introducing a Non-Redundant Derivative on the Volatility of Stock-Market Return When Agents Differ in Risk Aversion," *Review of Financial Studies*, 22, 2303-2330.

Bhamra, Harjoat S. and Raman Uppal (2010), "Asset Prices with Heterogeneity in Preferences and Beliefs," Working paper, Sauder School of Business (UBC) and London Business School.

Blume, L. and D. Easley (2006) "If you're so smart, why aren't you rich? Belief selection in complete and incomplete markets", *Econometrica*, 74, 929-966.

Campbell, J. Y. (2003), "Consumption-based Asset Pricing," in Constantinides, G., M. Harris and R. Stulz (eds) *Handbook of the Economics of Finance* (Amsterdam: North-Holland) 803887.

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.

Chan, Y. and L. Kogan (2002), "Catching Up with the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices," *Journal of Political Economy*, 109, 1255-1285.

Chen, H., S. Joslin and N-K. Tran (2011), "Rare Disasters and Risk Sharing with Heterogeneous Beliefs," *Review of Finance Studies*, forthcoming.

Chesson, H. and W. K. Viscusi (2000), "The heterogeneity of time-risk tradeoffs," *Journal of Behavioral Decision Making*, 13, 251-258.

Cox J. C. and C.-F. Huang (1989), "Optimal consumption and portfolio policies when asset prices follow a diffusion process," *Journal of Economic Theory*, 49, 3383.

Dimson, E., P. Marsh and M. Staunton (2008), "The Worldwide Equity Premium: A Smaller Puzzle," in Mehra R. and E. Prescott (eds) *Handbook of the Equity Risk Premium* (Elsevier B.V.) 467-514 (also SSRN Working Paper No. 891620).

Detemple, J. B., R. Garcia and M. Rindisbacher (2003), "A Monte Carlo Method for Optimal Portfolios," *Journal of Finance*, 58, 40146.

Detemple, J. and Murthy, S. (1994), "Intertemporal asset pricing with heterogeneous beliefs", *Journal of Economic Theory*, 62, 294-320.

Dumas, B. (1989), "Two-Person Dynamic Equilibrium in the Capital Market," *Review of Financial Studies*, 2, 157188.

Dumas, B., A. Kurshev and R. Uppal (2009) "Equilibrium Portfolio Strategies in the Presence of Sentiment Risk and Excess Volatility," *Journal of Finance*, 64, 579-629.

Gallmeyer, M. and B. Hollifield (2008) "An Examination of Heterogeneous Beliefs with a Short Sale Constraint," *Review of Finance*, 12, 323-364.

Garleanu, N., and S. Panageas (2010), "Young, Old, Conservative and Bold: The Implications of Heterogeneity and Finite Lives for Asset Pricing," Working Paper, University of California at Berkeley and University of Chicago.

Gollier, Christian, and Richard Zeckhauser (2005), "Aggregation of Heterogeneous Time Preferences," *Journal of Political Economy*, 113, 878-96.

Hansen, L. P., and R. Jagannathan (1991), "Implications of security market data for models of dynamic economies," *Journal of Political Economy*, 99, 225262.

Jouini, E. and C. Napp, (2007), "Consensus Consumer and Intertemporal Asset Pricing with Heterogeneous Beliefs," *Review of Economic Studies*, 74, 1149-1174.

Karatzas, I, J. P. Lehoczky, and S. E. Shreve (1987), "Optimal portfolio and consumption decisions for a "small investor" on a finite horizon," *SIAM Journal on Control and Optimization*, 25, 1557-1586.

Karatzas, I. and S. E. Shreve (1991), "Brownian Motion and Stochastic Calculus," 2nd ed., Springer, New York.

Kimball, M. S. (1990), "Precautionary Saving in the Small and in the Large," *Econometrica*, 58, 53-73.

Kimball, M. S. (1992), "Precautionary Motives for Holding Assets," *The New Palgrave Dictionary of Money and Finance*, Peter Newman, Murray Milgate and John Eatwell (eds.).

Stockton Press, New York, 158-161.

Kimball, M. S. (1993), "Standard Risk Aversion," *Econometrica*, 61, 589-611.

Kimball, M. S., C. R. Sahm and M. D. Shapiro (2008), "Imputing risk tolerance from survey responses," *Journal of the American Statistical Association*, 103, 10281038.

Kogan. L., S. Ross, J. Wang and M. Westerfield (2006), "The Price Impact and Survival of Irrational Traders," *Journal of Finance*, 61, 195229.

Kogan. L., S. Ross, J. Wang and M. Westerfield (2009). "Market Selection," Working paper, MIT Sloan School of Management.

Leland, H. E. (1968) "Saving and Uncertainty: The Precautionary Demand for Saving," *Quarterly Journal of Economics*, 82, 465-473.

Lengwiler, Yvan (2005), "Heterogeneous Patience and the Term Structure of Interest Rates," *American Economic Review*, 95, 890-896.

Lengwiler, Yvan, Semyon Malamud and Eugene Trubowitz (2005), "Asset Pricing in Heterogeneous Economies," Working paper, ETH Zurich.

Lettau, M. and J. Wachter (2009), "The Term Structures of Equity and Interest Rates," forthcoming, *Journal of Financial Economics.*

Longstaff, F. A. and J. Wang (2009), "Asset Pricing and the Credit Market," Working Paper, MIT and University of California at Los Angeles.

Malmendier, U. and S. Nagel (2010), "Depression Babies: Do Macroeconomic Experiences Affect Risk-Taking?," *Quarterly Journal of Economics*, forthcoming.

Mehra, R. and E. Prescott (1985), "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145161.

Mehra, R. and E. Prescott (2008), *Handbook of the Equity Risk Premium* (Elsevier B.V.).

Negishi, Takashi (1960), "Welfare Economics and Existence of an Equilibrium for a Competitive Economy," *Metroeconomica*, 12, 92-97.

Nualart, David (2006), "The Malliavin calculus and related topics. Probability and its

Applications," Second edition ed., Berlin: Springer-Verlag.

Paravisini, D., V. Rappoport, and E. Ravina (2010) "Risk Aversion and Wealth: Evidence from Person-to-Person Lending Portfolios, Working paper, Columbia Business School.

Sandmo, A. (1970), "The Effect of Uncertainty on Saving Decisions," *Review of Economic Studies*, 37, 353-360.

Sandroni, A. (2000), "Do markets favor agents able to make accurate predictions?," *Econometrica*, 68, 1303-1341.

Shiller, R. J. (1981), "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, 71, 421-436.

Wang, Jiang (1996), "The Term Structure of Interest Rates in a Pure Exchange Economy with Heterogeneous Investors," *Journal of Financial Economics*, 41 (1), 75-110.

Weil, P. (1989), "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401421.

Wilson, R. B. (1968), "The Theory of Syndicates," *Econometrica*, 36, 119-132.

Yan, Hongjun (2008), "Natural Selection in Financial Markets: Does It Work?," *Management Science*, 54 (11), 1935-1950.

Zapatero, F. (1998), "Effect of Financial Innovations on Market Volatility when Beliefs are Heterogeneous," *Journal of Economic Dynamics and Control*, 22, 597-626.

Chapter 3

Bringing Structures to Reduced-form Asset Pricing Models: the Functional Stochastic Discount Factor

3.1 Abstract

Any risk-neutral statistical distribution of state variables, either reliably inferred from prices observed in the market or exogenously formulated to generate closed-form prices, can be consistently and neatly tied to the economic contents of the underlying pricing model. We establish this structural linkage by requiring that the economy's marginal utility, or the stochastic discount factor, be a proper but *unspecified* function of the state variables. In this *functional stochastic discount factor* approach, the most general economic structures, being consistent with any state dynamic of choice, are identified to accommodate investors' rich behaviors. As a further result, state variables' distribution in physical measure can also be recovered. We illustrate the construction with an explicit real business cycle model in which (i) interest rates have affine term structures and (ii) the forward premium puzzle is consistent with consumption-risk rationale, the two key asset pricing features previously deemed conceptually incompatible. More generally, our approach offers novel flexibilities

that serve to extend several existing asset pricing frameworks: affine, quadratic, quotient interest rate models, as well as the models built on the linearity-generating processes.

3.2 Introduction

Arguably, the stochastic discount factor (SDF, also referred to as state price density, or pricing kernel in literature) is one of the most fundamental objects in asset pricing theory and modeling. From the arbitrage pricing perspective, the existence of SDF is equivalent to the absence of arbitrage, as asserted by a fundamental theorem of asset pricing. From the general equilibrium pricing perspective, SDF is the marginal utility of investors in the economy, as derived in the first-order condition of optimality. Indeed it is so fundamental that many asset pricing models just set out with the definition of the stochastic discount factors.

In the current paper, we propose a novel, general and tractable asset pricing construction in which pricing kernel is a proper function of underlying state variables. We refer to all variants of the construction as *functional stochastic discount factor.* Also throughout, by abusing terminology perhaps, pricing kernel and stochastic discount factor are used interchangeably.

Equilibrium asset pricing models, which center on stochastic discount factors, can be broadly classified into two groups. In the first group, the stochastic discount factor is identified structurally with a representative agent's marginal utility of consumption and thus is motivated fundamentally from rational time and risk preferences of market participants. Known models here include endowment and production economies with additive utilities, recursive utilities, and habit formations. Though being richly enhanced with economic intuitions, associated asset (bonds, stocks, options) prices do not necessarily have simple expressions. As a results, models' estimation processes vis-a-vis price data can be cumbersome in practice, even with ever increasingly powerful computational aids. In the second group, the stochastic discount factor is imposed at the onset in reduced form,[1], together with some

[1]This is usually done by picking a special short rate process $r(t)$. We recall that in risk neutral measure,

specification of the underlying state variable dynamic, under risk neutral measure Q. The setting is desirable thanks principally to the conveniences of the resulting closed-form asset prices. Known models here include Black-Scholes option pricing paradigm, and affine and quadratic term structure of interest rate modeling. Though being highly tractable, the associated models do not necessarily have structural economic intuitions.

Our basic construction is motivated to fill the gap between these two groups. Functional stochastic discount factors, when appropriately constructed, can have both first group's economics consideration and second group's pricing tractability. Being a proper function of underlying state variables, SDF is a structural object that can be mapped to rational intuitions. To enforce tractability, we can rely on the same risk-neutral state dynamic of known reduced-form models.

Specifically, our construction is based primarily on very simple observations. First, in many settings the statistical distribution properties of state variables in reduced-form asset pricing models are loosely connected to, and thus are (vaguely) compatible with exceedingly large economic modeling class of investors rational preferences. This can be a consequence of incomplete market or otherwise. Second, once it is imposed that SDF in physical measure P be a proper, but unspecified, function of state variables, it can be consistently linked to, and thus determined from the given risk neutral dynamic of state variables and short rate process via a standard linear differential equation. In other words, our construction proceeds consistently from the state dynamic in risk neutral measure to the endogenous SDF function in physical measure. And then follow market prices of risk, physical state dynamic and all other quantities of interest.

This construction theme fits very well into the practice and theory of asset pricing. First, in many settings the risk neutral probability of state dynamic is observable thanks to (i) the prices observed in the market and (ii) the tractability of the risk-neutral pricing apparatus.[2]

the stochastic discount factor (or more precisely, the pricing kernel), $\exp\left(-\int^t r(s)ds\right)$, is determined solely by the short rate process $r(t)$.

[2]One classic setting is the option market. Breeden and Litzenberger (1978) show that the risk-neutral conditional probability distribution of the underlying stock price can be determined from the prices of European call options of various strikes and maturities.

The simplest version of our construction takes the risk-neutral dynamic as given, while other versions considerably relax this assumption. Second, being able to consistently reconstruct the stochastic discount factor in physical measure simultaneously determine both investors' preferences and the physical probability of state dynamic. In particular, the latter has been embarrassingly difficult and in this regard the literature has to resort to direct but rather noisy responses to surveys from investors. Given that the stochastic discount factor is a proper function of state variable, our construction first consistently pins down this function, and then all other quantities of pricing interest, including the market prices of risk and the physical probability distribution.

The advantages of the proposed construction are illustrated and employed to rationally embrace a key stylized regularity in international finance. We construct a model in which the consumption risk explicitly accounts for the elusive forward premium puzzle (FPP, also known as the violation of uncovered interest rate parity). FPP is an empirically observed puzzling pattern in international market, that high interest rate currencies tend to appreciate.[3]

In the proposed international asset pricing model, our novel functional SDF is implied from a hybrid of power and exponential utilities. When coupled with the general affine interest rate and consumption dynamic, the resulting price of consumption risk correlates negatively with interest rate. Consequently, changes in exchange rates move in the opposite direction of interest rates' differential, and therefore, in consistence with the forward premium puzzle. Intuitively, when home market and consumption surge, home risk-free bonds lose their appeals as of insurance instruments, become cheaper and home interest rate increases. In other words, investors in a bull (say, home) market tend to consume more and confidently reduce precautionary savings, which boosts the interest rate in the associated country (and vice versa). At the same time, investors also perceive lower risk in home market, loosen their risk-based discounting aggressively[4] (which depresses the price of risk)

[3]This is puzzling because it appears that appreciating currencies are more valuable, yet investors require higher premia (i.e., interest rates) to hold them.

[4]albeit a surge in risk-free rate and risk-free discounting.

and end up valuing home currency more favorably. Altogether, these consistently render a highly-desirable affine term structure of interest rates and a rational explanation for the forward premium puzzle: home currency appreciates relatively while home interest increases. This hybrid functional form of the SDF is pinned down naturally and unambiguously in our construction from the imposition of two requirements; (i) SDF be a proper function of state variables (consumption, in particular) and (ii) state variables have affine dynamic. The two requirements are customary, but usually and respectively imposed in different literatures of structural and reduced-form asset pricing modeling. Our hybrid construction is essential for *both* FPP consistency *and* the extremely tractable (affine) term structure of interest rates. In comparison, the SDF in simple exponential-affine form does not deliver completely affine term structure dynamic, while the SDF in simple power form does not accommodate the forward premium puzzle.

In the literature of dynamic term structure modeling (DTSM) of interest rate, several analytical and convenient settings, those featuring affine and quadratic yields, have been proposed and widely employed. Conditional on functional SDF, we establish a unified framework for all these analytical settings. The key ideas are as follows. First, analytical bond pricing can be implemented outside the risk-neutral measure using transforms inspired by characteristic function techniques. This then motivates us to start out with a canonical tractable state dynamic in *any* equivalent measure R, which does not have to be either risk-neutral Q or physical P. Yet after rotations back to these meaningful measures, the associated Q- and P-dynamics are highly non-trivial. In particular, employing this change-of-measure flexibility, we show that quadratic DTSM, quotient DTSM and many other non-linear DTSM can all be derived from an affine DTSM in some spurious but equivalent measure R. This approach not only preserves the desirable bond pricing tractability, it also accommodates non-linear interest rates and rich structures of market prices, and most importantly, explicit economics motivations conveyed by the functional stochastic discount factor. In this regard, our construction shows that any tractable pricing setting can be greatly generalized by embedding original model in a new, appropriately chosen, equivalent measure R.

Furthermore, the recently proposed tractable asset pricing class based on linearity-generating (LG) processes neatly fits into our construction, since in the former setting the stochastic discount factor is a (linear) function of state variable. In term of modeling, this LG class of asset pricing models possesses a strong measure-invariant property (namely, if the model is LG in a measure, it essentially remains LG in any other equivalent measure), which hence cannot be generalized by using above change of measure. Yet, built upon our differential approach, we are able to construct a more general version of LG pricing models that does not have to set out with a strictly LG process.

Our paper contributes to the unceasing interest of building the structural economic models with tractable underpinning dynamics in no-arbitrage asset pricing literature, all revolving about the object of stochastic discount factor.[5] The fundamental properties of the stochastic discount factor; its existence, its relation to no-arbitrage and its pricing implications, are obtained first in Cox and Ross (1976), Ross (1976), and Harrison and Kreps (1979). In structural pricing literature, the properties of stochastic discount factors are formulated based on investors' rational (usually, utility-maximizing) behaviors. Consumption-based capital asset pricing models (C-CAPM) developed by Rubinstein (1976), Lucas (1978) and Breeden (1979), and a large subsequent equilibrium pricing literature follow this utilitarian line, with richer added features ranging from habit formation, recursive utility to heterogeneous-agent setting. The current paper partially adopts this approach in the sense that we explicitly constrain the stochastic discount factor to be proper function of a subset of state variables. This restriction facilitates the structural interpretation of our construction by clearly identifying the necessary characteristics of the utilitarian investors, whose preferences imply the SDF. To take a shortcut, Constantinides (1992) and Rogers (1997) directly formulate the stochastic discount factor[6] in physical measure without invoking investors' utilities. However in their work the SDF is exogenously specified. A key innovation that differentiates our

[5]It is impossible to thoroughly review the ever-growing literature on this subject within the scope of a paper. Here, instead we opt to briefly discuss only works that are most directly related to our proposed construction of functional stochastic discount factor. Interested readers are referred to Cochrane (2005), who gives an extensive account of the merits of the discount factor approach in a much more general setting.

[6]Rogers (1997) refers to the stochastic discount factor as state price density.

functional SDF construction from this (and consumption-based CAPM) literature is that in our approach, once the SDF is set to be a proper function of state variable, the functional form can be implied endogenously and consistently from the statistical distribution governing state variables. In this regard, our construction is non-parametric and thus flexible enough to accommodate rational behaviors of investors.

In tractable pricing literature. Duffie and Kan (1996) construct a general class of models with affine dynamic in risk-neutral measure Q that encompasses several previous classic reduced-form bond pricing models. The current paper generalizes this modeling paradigm by flexibly introducing affine dynamic in any equivalent measure R as the starting point. Dai and Singleton (2000) construct a general scheme to classify and analyze all affine term structure models of interest rate based on the numbers of relevant factors driving the volatility dynamic and the entire model. The current paper also attempts to classify term structure models, but based on a very different dimension. In our scheme, different models are related if their dynamics can be rotated from one to the other by a change of measure. In this classification, affine, quadratic and quotient term structure models can be connected as all may stem from an affine dynamic in some spurious equivalent measure. More recently, Gabaix (2009) introduces the class of linearity-generating processes in conjunction with a linear stochastic discount factor (and dividend) specification that allow for tractable bond and equity pricing. The current paper generalizes his construction by incorporating arbitrary (non linearity-generating) dynamics of the underlying state variables.

Our paper is most closely related to, but independent of and simultaneous with Ross (2011), who also presents a procedure to reconstruct physical dynamic and preferences from the risk-neutral dynamic. In that paper's setting, the state space is discrete and thus the approach therein is of algebraic (matrix) nature. The current paper's construction of functional stochastic discount factor is in continuous state spaces. We show that it is possible to reformulate both papers' approaches using a unified martingale method.

The paper is structured as follows. Section 3.3 introduces motivations for the most basic construction in which functional SDF is derived endogenously from state variables' dynamic in risk neutral measure. In particular, subsection 3.3.3 presents a comparative analysis of our functional SDF construction vis-a-vis Ross (2011)'s recovery theorem. Section 3.4 generalizes the basic construction by introducing a new equivalent measure R and derives many surprisingly close relations between classic models of dynamic term structure of interest rate. Section 3.5 constructs a more general version of linearity-generating processes and shows that they all are special cases of the functional SDF approach. Section 3.6 proposes an explicit equilibrium pricing model in which the forward premium anomaly is consistent with consumption risk and interest rate's term structure is affine. Section 3.7 demonstrates our construction at work in multi-factor settings, and sketches the maximum likelihood estimation procedure. Section 3.8 concludes. Appendices present proofs to all results as well as a table summarizing key technical notations employed in the main text.

3.3 Endogenous construction of stochastic discount factor

In this section, we present a novel approach to the construction of stochastic discount factors. The approach crucially hinges on the assumption 1 below that stochastic discount factor be a proper function of model's underlying state variables. We begin with formal presentation of this construction concept, and then proceed to in-depth discussion and various motivations for this assumption.

3.3.1 Set-up

To set the notation, we first consider a basic asset pricing setting driven by a state variable $X(t)$, which follows standard diffusion process in either physical measure (always denoted by P) or risk-neutral measure (always denoted by Q)

$$
\begin{aligned}
dX(t) &= \mu^{X,P}(X,t)dt + \sigma^X(X,t)dZ^P(t) \\
&= \mu^{X,Q}(X,t)dt + \sigma^X(X,t)dZ^Q(t),
\end{aligned}
\tag{3.1}
$$

where $Z(t)$'s are standard Brownian motions in respective measures, drifts μ^X's and diffusion σ^X are well-defined measurable processes (we note that the diffusion is independent of measure). For the sake of clarity, here we assume that $X(t)$ is a scalar (one-dimensional) process. The multi-dimensional case will be studied in a later section. The inclusion of jump processes is also possible.

Let r and M^P denote the risk free rate (rfr) and SDF in physical measure P respectively. Note that our definition of the stochastic discount factor $M^P(t)$ in the current paper is different from the period SDF commonly used in the literature, which in our convention is $\frac{M^P(t+dt)}{M^P(t)}$. Assuming no arbitrages throughout and standard regularity conditions, the martingale pricing of any contingent payoff $D(X,T)$ generates the identity

$$
E_t^P\left[\frac{M^P(T)}{M^P(t)}D(X.T)\right] = E_t^Q\left[\frac{\exp\left(-\int^T r(X,s)ds\right)}{\exp\left(-\int^t r(X,s)ds\right)}D(X,T)\right], \tag{3.2}
$$

and the following relation for all t

$$
\begin{aligned}
M^P(t) &= \exp\left(-\int^t r(X,s)ds\right)\xi^{QP}(t), \\
\xi^{QP}(t) &\equiv \exp\left(\int^t -\frac{1}{2}(\eta^{QP})^2(X,s)ds - \eta^{QP}(X,s)dZ^P(s)\right),
\end{aligned}
\tag{3.3}
$$

where ξ^{QP} is the Radon-Nikodym derivative[7] chararacterizing the change of measure from

[7] More rigorously, this object is defined by stochastic exponential martingale operator, see e.g. Rogers and Williams (1987). When P and Q are two equivalent measures, $E_t^Q[D(T)] = E_t^P\left[\frac{\xi^{QP}(T)}{\xi^{QP}(t)}D(T)\right]$ where ξ^{QP} is the Radon-Nikodym derivative chararacterizing the change of measure from Q to P.

Q to P, and η^{QP} the associated market price of risk

$$dZ^Q(t) = dZ^P(t) + \eta^{QP}(X,t)dt.$$

Note that ξ^{QP} is a martingale under measure P, and its reciprocal $\xi^{PQ} \equiv \frac{1}{\xi^{QP}}$ is a martingale[8] under Q. Combining relations in (3.3) yields a very useful and known differential representation of SDF that we will repeatedly invoke in later sections

$$\frac{dM^P(t)}{M^P(t)} = -r(X,t)dt - \eta^{QP}(X,t)dZ^P(t). \tag{3.4}$$

We note specially that, when state variable X is traded, Cox and Ross (1976)'s original no-arbitrage argument immediately fixes the its growth rate in risk neutral measure to be the short rate[9]

$$\mu^{X,Q}(X,t) = Xr(X,t). \tag{3.5}$$

In the rest of paper, however, we assume that state variables are not traded, and thus abstract from this explicit requirement on growth rate in risk neutral measure. In other words, there is no relation between $\mu^{X,Q}(X,t)$ and $r(X,t)$ in the setting a priori. We will address this case in a new version of the paper.

3.3.2 Construction: Basic version

The construction is motivated by risk-neutral pricing methodology and starts with the state $Q-$dynamic $\mu^{X,Q}(X,t)$, $\sigma^X(X,t)$ and short rate process $r(X,t)$. This framework is customary in leading dynamic asset pricing models of interest rate term structures. Furthermore, risk-neutral state dynamic can be inferred from observed prices. In a recent work, Carr and Wu (2007) design a procedure to estimate the risk-neutral distribution of currency returns from currency option prices recorded in over-the-counter market, which enables them to

[8] As ξ^{QP} is the Radon-Nikodym derivative associated with the change of measure from Q to P, ξ^{PQ} is the one associated with measure change from P to Q.

[9] The author is very grateful to John Cochrane for pointing out this case of traded state variables.

demonstrate the importance of the skew dynamic in that market. Furthermore, our construction does not impose any specific functional form on the physical state dynamic $\mu^{X,P}(X,t)$, $\sigma^X(X,t)$, which, in contrast to risk-neutral dynamic, is hard to estimate in practice. Instead, we make the following defining assumption for the our endogenous construction of stochastic discount factor.

Assumption 1: *In physical measure, stochastic discount factor M^P is a proper function of state variable X and time t of the form $M^P(X,t) = e^{-\rho t}M^P(X)$, where ρ is the standard subjective time discount factor* [10].

This assumption immediately yields a differential representation for SDF in risk neutral measure simply because Ito's lemma is applicable herein

$$\begin{aligned} dM^P(X,t) &= \left(\frac{1}{2}(\sigma^X(X,t))^2 M^P_{XX}(X,t) + \mu^{X,Q}(X,t)M^P_X(X,t) - \rho M^P(X,t)\right) dt \\ &+ \sigma^X(X,t)M^P_X(X,t)dZ^Q(t), \end{aligned} \tag{3.6}$$

where the subscripts always denote corresponding partial derivatives. From other direction, we have yet another representation for SDF (see also (3.4)), again by applying Ito's lemma on (3.3)

$$dM^P(X,t) = -M^P(X,t)\left[\left\{r(X,t) - (\eta^{QP}(X,t))^2\right\}dt + \eta^{QP}(X,t)dZ^Q(t)\right].$$

Under regularities, the uniqueness of this stochastic differential equation's solution allows to identify the drift and diffusion parts of SDF (recall that $M^P(X,t) = e^{-\rho t}M^P(X)$)

$$\frac{1}{2}(\sigma^X(X,t))^2 M^P_{XX}(X) + \mu^{X,Q}(X,t)M^P_X(X) + \left[r(X,t) - (\eta^{QP}(X,t))^2 - \rho\right] M^P(X) = 0, \tag{3.7}$$

$$\sigma^X(X,t)M^P_X(X) + \eta^{QP}(X,t)M^P(X) = 0. \tag{3.8}$$

[10] The assumption of constant subjective time discount factor ρ is convenient but nonessential for our analysis. Extension to the case where ρ is some function of time is possible, but conceptually contributes little to the construction.

Consistent with given underlying state Q-dynamic $\{\mu^{X,Q}(X,t), \sigma^X(X,t)\}$ and short rate process $r(X,t)$, the assumption 1 determines both SDF M^P and mpr η^{QP} jointly in (3.7). (3.8). To see this more clearly, we can also combine these to produce a single differential equation[11]

$$\begin{aligned} &\frac{(\sigma^X(X,t))^2}{2} M^P_{XX}(X) + \mu^{X,Q}(X,t) M^P_X(X) \\ &+ \left(r(X,t) - \frac{(\sigma^X(X,t))^2 (M^P_X(X))^2}{(M^P(X))^2} - \rho \right) M^P(X) = 0, \end{aligned} \tag{3.9}$$

from which indeed SDF M^P and then the physical probability follow endogenously as desired. However, a very serious technical obstacle in this construction is that this differential equation is highly non-linear and its solutions can be very elusive.

The situation is not that all dreadful. Interestingly, a careful observation offers a hint to meet this challenge. So far, the gist of our approach has been to solve for P-measure SDF consistent with Q-measure state dynamic. In this change of measures, the Radon-Nikodym derivative $\xi^{PQ} = \frac{\exp\left(-\int^t r(X,s)ds\right)}{M^P(t)}$ is necessarily a Q-martingale. That is, if $M^{P,1}$ and $M^{P,2}$ are two consistent solutions of the construction, their linear combinations $k^1 M^{P,1} + k^2 M^{P,2}$ are not [12]. This explains why the construction is not linear in M^P. Rather, it should be linear in $\frac{1}{M^P}$. This prompts us to a change of variable

$$\phi^P(X) \equiv \frac{1}{M^P(X)} = \frac{e^{-\rho t}}{M^P(X,t)}, \tag{3.10}$$

after which key eq. (3.9) becomes a homogeneous second order *linear* differential equation (HSOLDE) in $\phi^P(X)$

$$\frac{1}{2}(\sigma^X(X,t))^2 \phi^P_{XX}(X) + \mu^{X,Q}(X,t)\phi^P_X(X) + [\rho - r(X,t)]\phi^P(X) = 0. \tag{3.11}$$

Note that we furthermore need to impose appropriate condition $\phi^P(X) \geq 0 \ \forall X$, and a

[11] Again we implicitly assume that state variable X is not traded, and thus there is no relation between $\mu^{X,Q}(X,t)$ and $r(X,t)$ a priori. See the discussion below eq. (3.5).

[12] This because while $\frac{\exp\left(-\int^t rds\right)}{M^{P,1}}$, $\frac{\exp\left(-\int^t rds\right)}{M^{P,2}}$ are Q-martingales, $\frac{\exp\left(-\int^t rds\right)}{k^1 M^{P,1} + k^2 M^{P,2}}$ is not.

conventional initial condition $\phi^P(X(0)) = 1$ to qualify M^P as a proper stochastic discount factor.[13] From (3.4), (3.11) follows consistently the market price of risk η^{QP} and the physical dynamic $\{\mu^{X,P}(X,t)\}$

$$\eta^{QP}(X,t) = \frac{-M_X^P(X,t)}{M^P(X,t)}\sigma^X(X,t) = \frac{\phi_X^P(X,t)}{\phi^P(X,t)}\sigma^X(X,t); \qquad (3.12)$$
$$\mu^{X,P}(X,t) = \mu^{X,Q}(X,t) + \eta^{QP}(X,t)\sigma^X(X,t),$$

and any other quantities of interest, that can be used to estimate the model.[14] This is a key result of our functional SDF construction, so we formally recapitulate it in the following proposition before proceeding with further analysis.

Proposition 11 *Given the state dynamic $\{\mu^{X,Q}(X,t), \sigma^X(X,t)\}$ under risk-neutral measure Q and short rate $r(X,t)$, the stochastic discount factor in physical measure P is a proper and endogenous function of state variable $M^P = e^{-\rho t}M^P(X)$ iff $\phi^P(X) \equiv \frac{1}{M^P(X)}$ solves the second-order linear differential (3.11). Under this condition, physical dynamic $\mu^{X,P}(X,t)$ and market price of risk η^{QP} can be consistently inferred as in* (3.12).

Proof. Alternative to the intuitive derivation above, we sketch here a direct and very short proof to this proposition. The Radon-Nikodym derivative $\xi^{PQ} = \frac{\exp\left(-\int^t r(X,s)ds\right)}{M^P(X,t)}$ can be written as $\exp\left(-\int^t [r(X,s) - \rho]ds\right)\phi^P(X)$. It is a Q-martingale, and so is driftless under risk neutral measure. Assumption 1 then allows us to obtain an explicit expression of ξ^{PQ}'s drift under Q-measure in term of $\{\mu^{X,Q}(X,t), \sigma^X(X,t)\}$. Identifying this drift with zero

[13] Since these boundary conditions are on case-by-case basis, we omit further details in the current section's general discussion. Cheridito et al. (2007) explicitly treat the regularity conditions for the class of extended affine DTSM. The Feller's admissibility condition for the square-root process (a.k.a., Cox-Ingersoll-Ross or CIR) is discussed in section 3.6.2.

[14] There is another way to see why the introduction of $\phi^P(X) \equiv \frac{1}{M^P(X)}$ comes in handy in the current construction. Namely, once we make the assumption $M^P = e^{-\rho t}M^P(X)$, we rightfully have a *linear* differential equation by staying within measure P (see (3.7))

$$\frac{1}{2}(\sigma^X(X,t))^2 M_{XX}^P(X) + \mu^{X,P}(X,t)M_X^P(X) + [r(X,t) - \rho]\, M^P(X) = 0.$$

The problem, however, is that the construction does not wish to impose rigidly any specific functional form for $\mu^{X,P}$ at the onset. As a result, above linear differential equation is unspecified, and cannot be used. Conveniently, this simple change of variable also works for the setting of multi-dimensional state variables, as seen in section 3.7.

immediately gives rise to equation (3.11), which underlies the above proposition. ■

While subsequent sections will present the most general, analytical solutions to the fundamental differential equation (3.11) under many configurations, the most remarkable feature of the construction is readily conveyed by proposition 11. Starting out with the observable risk-neutral dynamic $\{\mu^{X,Q}(X,t), \sigma^X(X,t)\}$ (e.g., by inferring from option prices) and short rate $r(X,t)$, we can reconstruct consistently the marginal utility M^P, the market price of risk η^{QP} and the physical dynamic $\{\mu^{X,P}(X,t), \sigma^X(X,t)\}$. In our current continuous-state approach, any specific and qualified[15] solution of (3.11) may constitute a possible stochastic discount factor consistent with the same observable risk-neutral dynamic $\{r(X,t), \mu^{X,Q}(X,t), \sigma^X(X,t)\}$. This substantially simplifies the application and thus empowers the functional SDF approach. Instead of solving this differential equation in earnest generality, we may much simpler construct a specific solution with appropriate properties motivated by economics considerations. The obtained SDF should have both the consistency with the prescribed Q-dynamic and equilibrium economics appeals. Interestingly, it is this feature that also renders practical uses for our multi-factor functional SDF construction of section 3.7.

3.3.3 In relation to the recently-proposed "recovery theorem"

In an independent and simultaneous work,[16] Ross (2011) formulates a theorem, named "The Recovery Theorem", to recover the physical probability distribution (and preferences) from risk-neutral probability distribution. In term of their goals, thus, Ross (2011)'s theorem and the proposition 11 above are very similar. Whereas the recovery theorem employs algebraic (matrix) approach, the proposition 11 employs analytical (differential equation) approach. This section reconciles the two approaches using a martingale formulation.

We again start with a Q-martingale property $E_t^Q\left[\xi^{PQ}(T)\right] = \xi^{PQ}(t)$ for the Radon-

[15]Stochastic discount factors need be positive to enforce no arbitrage. Other properties may also be motivated and imposed out of economics considerations.

[16]The author is very grateful to John Cochrane and Steve Ross for the introduction to and many discussions on Ross (2011)'s paper.

Nikodym derivative $\xi^{PQ}(t) = \exp\left(-\int^t r ds\right)/M^P(X,t) = \exp\left(-\int^t [r-\rho]ds\right)\phi^P(X)$, which implies

$$E_t^Q\left[e^{-\int_t^T (r-\rho)ds}\phi^P(Y)\right] = \phi^P(X).$$

To bring the above martingale condition to the formulation of the recovery theorem, we consider an infinitesimal period $T = t + dt$ and denote $p(X,t;Y,T)$ the transition probability density from (X,t) to (Y,T) in *risk-neutral* measure. The above equation then reads

$$\int_Y e^{-\int_t^T (r-\rho)ds} p(X,t;Y,T)\phi^P(Y)dY = \phi^P(X) \tag{3.13}$$

Since this holds for any initial state X, on one hand, in discrete-state setting this equation is identical to the characteristic root equation in Ross (2011), which in turn gives rise to the recovery theorem therein.[17] On the other hand, in the continuous-state space, by virtue of the Kolmogorov backward equation on the transition probability density $p(X,t;Y,T)$ in the risk-neutral measure,

$$\frac{\partial}{\partial t}p(X,t;Y,T) = \mu^{X,Q}(X,t)\frac{\partial}{\partial X}p(X,t;Y,T) + \frac{1}{2}(\sigma^X(X,t))^2\frac{\partial^2}{\partial X^2}p(X,t;Y,T)$$

the same equation (3.13) immediately implies the key differential equation (3.11), which underlies our proposition 11. Thus, both the recovery theorem and the current paper's construction trace their roots back to the fundamental change-of-measures martingale, the premier apparatus of modern asset pricing theory. It is worthwhile to note that the current construction works with continuous-state, continuous-time setting and does not produce strong results in the uniqueness as in the case of the recovery theorem. Additionally, we do not fix investors' preferences a priori. Our approach instead systematically and endogenously reconstructs a set of possible preferences for investors that are consistent with dynamic in risk neutral measure. It is this flexibility that will help us to construct equilibrium models consistent with rational economic intuitions. The physical probability distribution is

[17] In the one-period discrete-state setting, $T = t + 1$, (3.13) becomes $\sum_j e^{-r_i} p_{ij}\phi_j = \delta\phi_i$, or the characteristic root equation $P\phi = \delta\phi$, where $P_{ij} = e^{-r_i}p_{ij}$; $\delta = e^{-\rho}$; $\phi = \{\phi_i\}_i$.

identified once the preferences have been pinned down. Section 3.6 present an international asset pricing model constructed precisely along this theme that is consistent with the forward premium puzzle.

3.3.4 Motivations and discussion

Several further thoughts on this construction approach are in order here.

First, that the stochastic discount factor is a function of state variable features predominantly in consumption-based equilibrium asset pricing models. Therein SDF is representative agent's marginal utility $M^P = \frac{\partial U(C,H)}{\partial C} = U_C(C, H)$, and thus is proper function of aggregate consumption C, and possibly other state variables such as consumption surplus H (in habit formation setting), and so forth. In fact, this is one of prime motivations of our construction and aims to explore the possibility to place certain no-arbitrage pricing models, for e.g., those in dynamic term structure of interest rates literature, on explicit utilitarian framework. In this regard, although the construction restricts the choice of market price risk η^{QP} as we seen above, we have the freedom in modeling the short rate function $r(X, t)$ as desired. In turn, the resulting P-dynamic $\mu^{X,P}(X, t)$ in (3.12) is very rich. Specifically, in the next section we construct a class of tractable bond pricing models, wherein short rate $r(X, t)$, dynamic $\mu^{X,P}(X, t)$ and even $\mu^{X,Q}(X, t)$ do not have to be linear in X. The class thus is beyond affine dynamic term structure framework.

Second, the assumption 1 that stochastic discount factor be some proper function $M^P(X, t)$ of underlying state variable appears similar to imposing a Markovian structure on it. In single-factor setting and under standard conditions, diffusion dynamic (3.1) implies that $X(t)$ is Markovian and so is $M^P(X, t)$ given this function being regular enough. Any SDF has the following integral representations

$$M^P = \exp - \int^t \left[\left\{ \frac{1}{2}\eta^{QP}(X, s)^2 + r(X, s) \right\} ds + \eta^{QP}(X, s) dZ^P(s) \right], \qquad (3.14)$$

which implies that in general (outside assumption 1) M^P depends on the history path of state process $\{X(s)\}_0^t$ (or path-dependent). Hence beginning with two exogenously given and *arbitrary* functions $r(X,t)$ and $\eta^{QP}(X,t)$, the SDF may not always be a simple function of only current state (X,t). It is only when $r(X,t)$ and $\eta^{QP}(X,t)$ jointly are restricted [18] by system of equations (3.7), (3.8), the SDF can be proper function of current state variable. In other words, $\eta^{QP}(X,t)$ is implied from the fundamentals $\{\mu^{X,Q}(X,t), \sigma^X(X,t), r(t,X)\}$ via assumption 1, and in the same process a consistent function $M^P(X,t)$ is endogenously determined.

Yet interestingly, the functional requirement placed on SDF does not rule out a design in which M^P retains the path-dependent feature. A simple counterexample is obtained when the state $X(t)$ itself depends on the entire path $\{Z^P(s)\}_0^t$ of Brownian motion $Z^P(t)$, and so does $M^P(X,t)$. As suggested in Chen and Joslin (2011), we may augment the state space to absorb the path dependence in one variable into another new state variable. Consider an overly simple setting in which $F(\{X(s)\}_0^t) = F\left(X(t), \int^t f(X,s)dZ^P(s)\right)$ is a path-dependent object. After defining a new state variable $Y \equiv \int^t f(X,s)dZ^P(s)$, $F = F(X,Y)$ becomes a proper function in new augmented state space (X,Y). Now the new state vector (X,Y) has the dynamic similar to (3.1), but generalized to a multi-dimensional framework, the task that we take up in section 3.7. Alternatively, we can also embed our basic functional SDF construction in any other equivalent measure R (which is not necessarily P or Q, see construction 2 below). As a result, when we get back to physical measure P, the SDF $M^P = M^P(\{X(s)\}_0^t, t)$ now depends on the entire history path of state variable. We present now yet a more specific construction of path-dependent SDF in our approach. Consider the following specification

$$\mathcal{M}^P(\{X\},t) = \exp\left(\int^t f(X(s))dZ^Q(s)\right)M^P(X,t),$$

where $f(X)$ is some general and given function of state variable, and $M^P(X,t) = e^{-\rho t}M^P(X)$ is another function to be solved endogenously in our construction. This path-dependent spec-

[18] Indeed, beginning with given and unrelated functions $r(t,X)$ and $\eta^{PQ}(t,X)$, the system (3.7), (3.8) will generally have no solution $\phi^P(X)$ (or P-SDF $M^P(t,X) = \frac{e^{-\rho t}}{\phi^P(X)}$) (because either $r(t,X)$ or $\eta^{PQ}(t,X)$ alone is sufficient to yield a solution $\phi^P(X)$ up to constants of integration).

ification $\int^t f(X)dZ^Q(s)$ is an Ito's integral under measure Q to facilitate the determination of path-independent factor $M^P(X)$. This feature is not essential and we will restore the full endogenous specification of $\mathcal{M}^P$ under measure P later. Indeed, by the identical reasoning, $\phi^P(X) \equiv \frac{1}{M^P(X)}$ can be determined for a linear differential equation analogous to (3.11)

$$\frac{1}{2}[\sigma^X(X)]^2\phi^P_{XX}(X) + \left[\mu^{X,Q}(X) - f(X)\sigma^X(X)\right]\phi^P_X(X)$$
$$+ \left[\tfrac{1}{2}f^2(X) + \rho - r(X,t)\right]\phi^P(X)(X) = 0.$$

Path-dependent factor also enriches the market price of risk η^{QP} (compared with (3.12)), which can be found from (3.4)

$$\eta^{QP}(X) = \frac{\phi_X(X)}{\phi(X)}\sigma^X(X) - f(X).$$

In this example, the ex-post SDF in physical measure is path-dependent and reads

$$\mathcal{M}^P(\{X\},t) = \exp\left(\int^t \left[-\rho + f(X(s))\eta^{QP}(X(s))\right]ds + \int^t f(X(s))dZ^P(s)\right)\frac{1}{\phi^P(X)}.$$

Third, the SDF in explicit functional form $M^P(X,t)$ can facilitate testing and estimation via generalized method of moments (GMM). Specially, when underlying state variable X is observable, the associated Euler equation can be estimated in discrete time following the standard procedure of Hansen and Singleton (1982)

$$E^P_t\left[\frac{M^P(X(t+1).t+1)}{M^P(X(t).t)}R(X(t+1),t+1)\right] = 1,$$

where R is a gross return on any traded asset. Alternatively, the resulting P-dynamic $\{\mu^{X,P}(X,t), \sigma^X(X,t)\}$ explicitly obtained in this construction is sufficient statistics to carry out an approximate but efficient maximum likelihood estimation as proposed by Ait-Sahalia (2002). The section 3.7 below provides key steps for this procedure.

Finally, we will analytically solve for key equation (3.11) for many important functional configurations of $\mu^{X,Q}$, $\sigma^{X,Q}$ and r in the next section. For now we just note that there exists a very standard, convenient and simple solution method that works for arbitrary time-homogeneous functions $\mu^{X,Q}(X)$, $\sigma^{X,Q}(X)$, and $r(X)$. In that setting, a simple change of variable $\psi^P(X) \equiv \frac{\phi_X^P(X)}{\phi^P(X)}$ transforms second order differential eq. (3.11) into a (first-order) Riccati differential equation

$$\psi_X^P(X) + (\psi^P(X))^2 + \frac{2\mu^{X,Q}(X)}{(\sigma^X(X))^2}\psi^P(X) + \frac{2[\rho - r(X)]}{(\sigma^X(X))^2} = 0,$$

which can be numerically solved very quickly. The nice feature of this transformation is that by virtue of (3.8) and (3.10), $\psi^P(X) \equiv \frac{\phi_X^P(X)}{\phi^P(X)} = \frac{\eta^{QP}(X)}{\sigma^X(X)}$, and above Riccati equation directly determines mpr $\eta^{QP}(X)$. In other words, the mpr η^{QP} satisfies a simple Riccati different equation in this construction.

In the next section we apply this construction to modeling dynamic term structure of interest rate, and explore various generalizations of the current basic configuration in section 3.7.

3.4 Affine term structure modeling and beyond

While affine and other term structure models are reduced-form models motivated by remarkable fixed-income derivatives pricing tractabilities, the proposed construction is motivated by a closed-form SDF and thus has the appeal of structural models. A possible connection between these two approaches will place them on firmer footings, either from pricing or equilibrium consumption perspectives. The principal question here is on equilibrium modeling side: how we can build a functional SDF that also possess tractabilities of leading models of dynamic term structure.

We first note that the premise of interest rate affine term structure models suits particularly well the basic construction of previous section. In particular, both require specifying Q-dynamic $\{\mu^{X,Q}(X,t), \sigma^X(X,t)\}$ and short rate process $r(X,t)$. But this is just the starting point of the current comparative exploration. We will substantially generalize the basic construction by initiating it in any equivalent measure R, and recover tractable bond pricing frameworks with very rich P- and Q-dynamics as well as highly non-linear short rate. Subject to assumption on a functional SDF, this approach thus provides a single new framework for many dynamic term structure models (DTSM) from affine, quadratic, quotient and other classes.

3.4.1 Construction 1: basic Q-dynamic term structure modeling

We recall the key ingredients of affine term structure models (Vasicek (1977), Cox, Ross and Ingersoll (1985), Duffie and Kan (1996)), that render tractable bond prices and yields. They are affine Q-dynamic ($\mu^{X,Q}$ and $(\sigma^X(X))^2$ linear in X) and linear short rate (r linear in X). Then follows price of zero-coupon bond of maturity T in closed form

$$ZCB(t, t+T) = E_t^Q\left[\frac{\exp - \int^{t+T} r(X(s))ds}{\exp - \int^{t} r(X(s))ds}\right] = \exp\left[A(T) + B(T)X(t)\right].$$

where $A(T)$, $B(T)$ satisfy a system of Riccati equations. Evidently, the term structure is linear in these settings. Although leading affine DTSM models, such as completely affine (Dai and Singleton (2000)), essentially affine (Duffee (2002)), extended affine (Cheridito et al. (2007)) also impose affine dynamic in physical measure P out of econometrics conveniences, affine Q-dynamic is the key for bond pricing tractabilities.

Our first construction is built on above ingredients of Q-dynamic term structure models in order to retain the fixed-income derivatives pricing tractabilities, together with the assumption on functional SDF.

Construction 1:

- *in P-measure, SDF is proper, but* unspecified, *function of state variable* [19] *:* $M^P(X,t) = e^{-\rho t}M^P(X)$
- *affine Q-dynamic:* $\mu^{X,Q}(X) = K_0^Q + K_1^Q X$; $(\sigma^X(X))^2 = H_0 + H_1 X$

[19] This ingredient also contains an implicit requirement that time discount rate in the economy be ρ.

- *linear short rate:* $r(X) = a + bX$

Implied SDF

The specified Q-dynamic immediately yields a differential equation on $\phi^P = \frac{1}{M^P}$, as a special case of eq. (3.11)

$$\frac{1}{2}\left[H_0 + H_1 X\right]\phi^P_{XX}(X) + \left[K_0^Q + K_1^Q X\right]\phi^P_X(X) + \left[\rho - a - bX\right]\phi^P(X) = 0. \qquad (3.15)$$

Appendix 3.9.2 derives analytical solution to this equation. We summarize the general resulting characteristics of construction 1 in the following proposition.

Proposition 12 *The most general functional stochastic discount factor M^P consistent with construction 1 is*

$$M^{P,\{\lambda_1,\lambda_2\}}(X,t) = \frac{e^{-\rho t}e^{-\frac{\alpha}{\beta}z}}{\lambda_1\Phi(\delta,\gamma;z) + \lambda_2 z^{1-\gamma}\Phi(\delta-\gamma+1.2-\gamma;z)}, \qquad (3.16)$$

$$z \equiv \beta(H_0 + H_1 X),$$

where λ_1, λ_2 are two constants of integration associated with differential equation (3.15), $\alpha, \beta, \gamma, \delta$ are constant coefficients related to model's parameters given in appendix 3.9.2 and $\Phi(\cdot,\cdot;z)$ is the confluent hypergeometric function of argument z.

We note that there may be many functional SDFs consistent with the same construction 1, each is characterized by a constant pair $\{\lambda_1, \lambda_2\}$. However, $\{\lambda_1, \lambda_2\}$ are not arbitrary. They should be chosen to assure the positivity of $M^P(X)$ in admissible domain of X and the normalization [20] $M^P(X(0)) = 1$.

In general solution (3.16), a very convenient property of confluent hypergeometric function $\Phi(a, a; z) = e^z \; \forall a, z$ gives rise to the following two interesting special cases.

[20] Stochastic discount factor can be determined only up to a multiplicative constant.

1. $\lambda_2 = 0, \delta = \gamma$: in this case,

$$M^P(X,t) = e^{-\rho t} e^{-(\alpha+\beta)(H_0+H_1X)}, \tag{3.17}$$

which is well-known in term structure modeling literature as exponential affine (see e.g., Duffie et al. (2000)). In particular, this is a completely affine configuration, because the resulting mpr is $\eta^{QP} \sim \sigma^X = \sqrt{H_0 + H_1 X}$, and P-dynamic $\mu^{X,P}$ is affine in X.

2. $\lambda_1 = 0, \delta = 1$: in this case,

$$M^P(X,t) = e^{-\rho t} e^{-(\alpha+\beta)(H_0+H_1X)} (H_0 + H_1 X)^{\gamma-1}. \tag{3.18}$$

which is a new and richer SDF form that also contains a polynomial factor in X (referred to as polynomial-exponential-affine hereafter). *Remarkably*, the P-dynamic implied by this SDF is *also* affine, even though market price of risk η^{QP} associated with M^P does not have to be proportional to state variable's volatility σ^X. We will derive and study this special SDF in much more details in sections 3.4.4 and 3.6. There we show that, even in one-factor settings, its richness pays off a desirable negative correlation between changes in exchange rate and interest rate differentials. This is the forward premium puzzle (FPP) in international finance that, in comparison, cannot be accommodated by above exponential affine configuration, as noted by Backus et al. (2001).

Relations to affine DTSMs

In our construction, the market price of risk is readily implied from (3.8) and P-dynamic drift from (3.12)

$$\begin{aligned} \eta^{QP}(X) &= -\tfrac{M_X^P(X)}{M^P(X)}\sigma^X(X) \\ \mu^{X,P}(X) &= \mu^{X,Q}(X) + \eta^{QP}(X)\sigma^X(X) = K_0^Q + K_1^Q X - \tfrac{M_X^P(X)}{M^P(X)}(H_0 + H_1 X) \end{aligned}.$$

Then it is clear from the solution (3.16) that this construction is able to accommodate non-affine P-dynamic $\mu^{X,P}(X)$. In comparison with leading affine dynamic term structure models [21] where both P and Q dynamics are affine, the tradeoff is evident. We have rich (non-linear) P dynamic at the price of more restrictive choice of market price of risk (needed to enforce proper functional SDF in our construction). We now characterize this tradeoff more quantitatively in the following proposition.

Proposition 13 *In 1-factor settings with linear short rate $r = a + bX$:*
(i) The functional-SDF construction 1 with additional specifications

$$K_0^P - K_0^Q = 0, \qquad \frac{1}{2}\left[(K_1^P)^2 - (K_1^Q)^2\right] = b,$$

reduces to a model in the completely (and essentially) affine DTSM class.
(ii) The functional-SDF construction 1 with additional specifications

$$\left(K_0^P - K_0^Q\right)\left(K_0^P + K_0^Q - 1\right) = 0, \qquad \frac{1}{2}\left[(K_1^P)^2 - (K_1^Q)^2\right] = b,$$

reduces to a model in the extended affine DTSM class.

Other term structure models with non-affine state dynamic in data generating measure have been proposed in the literature, all with linear short rate. Duarte (2004) constructs semi affine square-root (SAS-R) model in which state variables have affine dynamic in Q, but non-affine in P. He shows that the SAS-R model outperforms known DTSMs in matching the time variability of the term premium. Most recently, Le et al. (2010) propose a class of discrete-time dynamic pricing models with a very general functional market prices of risk, which in our notation is $\eta(X,t) = \Lambda(X,t)\sqrt{X}$, where $\Lambda(X,t)$ is some general exogenous function of state variables. Their models then imply non-linear physical dynamic $\mu^P \sim \mu^Q + \Lambda(X,t)X$ via the corresponding Q-martingale Radon-Nikodym derivative $\frac{\xi^{PQ}(X(t+1),t+1)}{\xi^{PQ}(X(t),t)} = \frac{e^{\Lambda(X(t+1),t+1)X(t+1)}}{E_t^Q[e^{\Lambda(X(t+1),t+1)X(t+1)}]}$. In comparison, market price of risk function is implied endogenously

[21] Completely affine models by Dai and Singleton (2000), essentially affine models by Duffee (2002), and extended affine models by Cheridito et al. (2007).

in our construction. In later sections, we generalize our construction further to allow for non-linearity specification on both Q-drift and short rate processes. We turn now to a new key in implementing these generalities via the change of measures.

3.4.2 Construction 2: introducing equivalent measure R

A simple observation generalizes our basic construction substantially. In a nutshell, the introduction of (any) equivalent measure R (which is not necessarily risk-neutral Q or physical P) to construction 2 naturally and richly renders (i) non-affine Q- *and* P- dynamics $\mu^{X,Q}(X)$, $\mu^{X,P}(X)$, (ii) non-linear $r(X)$, (iii) general non-Markovian (path-dependent) SDF in measure P, while (iv) keeping bond pricing tractable. Though, in the difference with the risk-neutral probability, there is no apparent link, and thus constraint, on R dynamic directly from price data.

In fact, no-arbitrage pricing may be performed in P, Q or any equivalent measure R. For a contingent payoff $D(X,T)$, an extension of (3.2) reads

$$E_t^Q\left[\frac{\exp\left(-\int^T r(X,s)ds\right)}{\exp\left(-\int^t r(X,s)ds\right)}D(X,T)\right] = E_t^R\left[\frac{M^R(X,T)}{M^R(X,t)}D(X,T)\right],$$

where M^R is the stochastic discount factor associated with equivalent measure R by construction. In particular, the tractability of bond pricing is extended to R-affine framework using a transform technique (see e.g., Chen and Joslin (2011) and Cuchiero et al. (2009)). All that is needed here is the existence of the Fourier transform [22] $\hat{M}^R$ of SDF M^R

$$\hat{M}^R(v,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-ivX}M^R(X,t)dX \longleftrightarrow M^R(X,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{ivX}\hat{M}^R(v,t)dv. \tag{3.19}$$

Assuming the existence of $\hat{M}^R(v)$, affine zero-coupon bond pricing in equivalent measure

[22] Laplace transform can also be employed.

proceeds as usual

$$
\begin{aligned}
ZCB_{t,t+T} &= E_t^R\left[\frac{M^R(X(t+T),t+T)}{M^R(X(t),t)}\right] \\
&= \frac{e^{-\rho T}}{\sqrt{2\pi}}\frac{1}{M^R(X(t))}\int_{-\infty}^{\infty}\hat{M}^R(v)E_t^R\left[e^{ivX(t+T)}\right]dv \\
&= \frac{e^{-\rho T}}{\sqrt{2\pi}}\frac{1}{M^R(X(t))}\int_{-\infty}^{\infty}\hat{M}^R(v)e^{A(v,T)+B(v,T)X(t)}dv.
\end{aligned} \tag{3.20}
$$

Alternatively, when M^R has certain functional forms, e.g., product of polynomial and exponential functions $M^R = e^{\alpha X}X^n$, the bond pricing can be performed in measure R by repeated differentiating as suggested by the technique of moment generating function

$$
ZCB_{t,t+T} = E_t^R\left[\frac{M^R(X(t+T),t+T)}{M^R(X(t),t)}\right] \sim \left.\frac{\partial^n}{\partial\alpha^n}\right|_{\alpha=0} E_t\left[e^{\alpha X(t+T)}\right] = \left.\frac{\partial^n}{\partial\alpha^n}\right|_{\alpha=0} e^{A(\alpha,T)+B(\alpha,T)X(t)}.
$$

where the last equality is obtained because state dynamic is affine under measure R.

These flexibilities in turn allow for non-linear short rates, and thus relate our construction to more general DTSMs such as quadratic, quotient, and other models.

Construction 2: Let R be an (any) equivalent measure

- *SDF is proper, but* unspecified, *function of state variable in measures P and R: $M^P(X,t) = e^{-\rho t}M^P(X)$, $M^R(X,t) = e^{-\rho t}M^R(X)$.*[23] *Furthermore, $M^R(X)$ is bounded function.*

- *affine R-dynamic: $\mu^{X,R}(X) = K_0^R + K_1^R X$; $(\sigma^X(X))^2 = H_0 + H_1 X$*

- *Q-dynamic drift $\mu^{X,Q}(X)$ is some given*[24]*, but arbitrary function.*

[23] It is straightforward to incorporate the more general configurations where rates ρ^P and ρ^R are different. This arises for e.g., in models where measure R characterizes representative agent's subjective belief and true discount rate ρ^P is mixed up with belief's drift term to produce an effective discount rate ρ^R, see e.g., Yan (2008). However, this flexibility do not present new construction concept, and will be omitted for the sake of simple exposition.

[24] The choice of Q-dynamic $\mu^{X,Q}(X)$ is either dictated by price data, as in Breeden and Litzenberger (1978)'s formula for option market, or exogenously specified as in models of affine term structure of interest rate.

The main advantage of this construction's generalization is the substantial modeling flexibilities to compensate for functional SDF restriction motivated by general equilibrium principle, and at the same time R-dynamic is not directly constrained by observed prices. The boundedness of $M^R(X)$ is sufficient to assure the existence of its Fourier transform $\hat{M}^R(v)$, and the subsequent tractable bond pricing. R-SDF $M^R(X,t)$ is implied in the construction by integrating out the first-order differential equation (see (3.22))

$$\begin{aligned} & (\sigma^X(X))^2 M_X^R(X) + [\mu^{X,R}(X) - \mu^{X,Q}(X)] M^R(X) = 0 \\ \Longrightarrow \quad & M^R(X) = \lambda \exp\left\{ -\int^X \frac{\mu^{X,R}(x) - \mu^{X,Q}(x)}{(\sigma^X(x))^2} dx \right\}, \end{aligned} \tag{3.21}$$

where λ is the constant of integration. As constructed, this is the *most general* functional form of possible SDFs in measure R that are consistent with the given dynamic $\{\mu^{X,Q}(X), \mu^{X,R}(X), \sigma^X(X)\}$. As a check, however, we need to verify the boundedness on this $M^R(X,t)$ afterward. Before proceeding further, two quick observations concerning this function are in oder here. First when the dynamic follows Ornstein-Uhlenbeck mean-reverting (constant diffusion, $H_1 = 0$) process in both measures,

$$M^R(X) \sim e^{A+BX+CX^2},$$

which has the exponential-quadratic functional form studied by Constantinides (1992). We will study this configuration in the next section. Second, when the dynamic follows CIR (i.e., $H_1 \neq 0$) process in both measures,

$$M^R(X) \sim e^{A+BX}(H_0 + H_1 X)^C,$$

which has the polynomial-exponential-affine functional form. We thus reconfirm the form (3.18), which was derived under similar assumption of CIR square-root Q- and P- dynamics.

Now back to the construction 2, a specification for $\mu^{X,Q}(X)$ is still needed as we do not wish to impose any functional form on SDFs $M^P(X)$, $M^R(X)$. This construction does not fall into the jurisdiction of proposition 11 because short rate $r(X)$ is not given at the onset here. Alternatives to this configuration will be considered in construction 3.

Non-linear interest rate

The first property of this construction is the non-affine dynamic $\mu^{X,P}(X)$, $\mu^{X,Q}(X)$ in *both* measures P and Q. This is the essential consequence of shifting the bond pricing task (and the required affine dynamic) to an (any) equivalent measure R, and thus setting loose the dynamics on Q and P measures. As a result, interest rate is non linear in general. Combining Ito's lemma and no-arbitrage principle [25] on function $M^R(X,t)$

$$-M^R(X,t)\left[r(X)dt + \frac{\mu^{X,R}(X)-\mu^{X,Q}(X)}{\sigma^X(X)}dZ^R\right] = dM^R(X,t) = \tag{3.22}$$
$$\left[\frac{1}{2}(\sigma^X(X))^2 M^R_{XX}(X,t) + \mu^{X,R}(X)M^R_X(X,t) - \rho M^R(X,t)\right]dt + \sigma^X M^R_X(X,t)dZ^R,$$

and plugging in construction 2's specified R-dynamic yields the interest rate

$$r(X) = \rho + \frac{-[\mu^{X,Q}(X)]^2 - \mu^{X,Q}_X(X) + \mu^{X,Q}(X)H_1 + [K^R_0 + K^R_1 X]^2 + K^R_1 H_0 - K^R_0 H_1}{H_0 + H_1 X}. \tag{3.23}$$

Flexibility in the choice of $\mu^{X,Q}$ translates into the very flexible form of short rate. The resulting non-linearities here can be useful in interest rate modeling practice. In particular, the non-parametric empirical studies of Ait-Sahalia (1996) and Stanton (1997) point to a diffusion term of power $\delta \approx 1.5$ in the short rate process $dr = \mu^r(r)dt + \sigma r^\delta dZ^P$. Meanwhile, one-factor affine dynamic setting with linear interest rate can only generate either $\delta = 0$ or $\delta = 0.5$. Appropriate specification of $\mu^{X,Q}(X)$ in our construction (3.23) in contrast can give rise to wider choice for δ. We might have started with some exogenously specified and non-linear $r(X)$ and proceed to P-SDF $M^P(X,t)$ along the line of proposition 11. This specification, however, would not lead to tractable bond pricing in general.

Note that up to this point we have not made use of the assumption on proper functional form $M^P(X,t)$ in measure P. It will be needed if we wish to pin it down along the strategy

[25] Note that interest rate is the opposite growth (drift) rate of stochastic discount factor in any equivalent measure.

of construction 1 above. This is because by now we have obtained the explicit functional short rate $r(X)$ in (3.23). The only difference is that r is specified exogenously as a linear function of the state variable in construction 1. SDF $M^P(X,t)$ in physical measure follows from linear differential equation (3.11). We will carry out this procedure explicitly next.

Quadratic DTSM

We consider now the first simple and special specification of construction 2, wherein state variable has affine dynamic also in risk neutral measure, with constant volatility (i.e., mean-reverting).

$$\mu^{X,Q}(X) = K_0^Q + K_1^Q X; \qquad [\sigma^X(X)]^2 = H_0.$$

The short rate in this model is necessarily quadratic in X as implied by general formula (3.23)

$$r(X) = \rho_0 + \rho_1 X + \rho_2 X^2, \tag{3.24}$$
$$\rho_0 \equiv \rho + \frac{(K_0^R)^2-(K_0^Q)^2-K_1^Q-K_1^R H_0}{H_0}; \quad \rho_1 \equiv \frac{2(K_0^R K_1^R-K_0^Q K_1^Q)}{H_0}; \quad \rho_2 \equiv \frac{(K_1^R)^2-(K_1^Q)^2}{H_0}.$$

It is interesting to see that this is the quadratic DTSM developed in Ahn et al. (2002) and Leippold and Wu (2002) from the exogenous Ornstein-Uhlenbeck (OU) mean-reverting state dynamic and quadratic short rate. [26] Thus the current construction 2, when specialized to mean-reverting Q-dynamic [27] , relates to this quadratic DTSM framework in the literature. However, since we also assume that P-SDF is proper function of stare variable $M^P(X,t) = e^{-\rho t}M^P(X) = \frac{e^{-\rho t}}{\phi^P(X)}$, we can construct its governing linear differential equation. Feeding quadratic short rate (3.24) into proposition 11 yields

$$\frac{1}{2}H_0\phi^P_{XX}(X) + (K_0^Q + K_1^Q X)\phi^P_X(X) + (\rho - \rho_0 - \rho_1 X - \rho_2 X^2)\phi^P(X) = 0. \tag{3.25}$$

[26] These quadratic DTSMs specify OU state dynamic in measure P, and impose an affine market price of risk. These together imply OU dynamic in measure Q, which is all needed for risk-neutral tractable bond pricing with given quadratic discount rate $r(X)$.

[27] Recall that we need to specify the Q-dynamic $\mu^Q(X,t)$ in construction 2, though this function can be quite arbitrary.

This equation pins down all possible qualified SDF functions $M^P(X,t)$.

Proposition 14 *The most general functional stochastic discount factor M^P consistent with construction 2, when the latter is specialized to OU mean-reverting Q-dynamic (or equivalently, quadratic DTSM), is*

$$M^{P,\{\lambda_1,\lambda_2\}}(X,t) = \tag{3.26}$$

$$\frac{e^{-\rho t}e^{-MX-NX^2}}{\lambda_1 e^{\frac{-(m+nX)^2}{4}}\Phi\left(\frac{\nu}{2}+\frac{1}{4},\frac{1}{2};\frac{(m+nX)^2}{2}\right)+\lambda_2(m+nX)e^{\frac{-(m+nX)^2}{4}}\Phi\left(\frac{\nu}{2}+\frac{3}{4},\frac{3}{2};\frac{(m+nX)^2}{2}\right)},$$

where λ_1, λ_2 are two constants of integration, ν, m, n, M, N are constant coefficients related to model's parameters given in appendix 3.9.2, z is linear in X, and again $\Phi(.,.;z)$ is confluent hypergeometric function of argument z.

Two particularly simple cases obtain when ν assumes special values.[28] (This amounts to imposing a constraint on model's parameters. See appendix 3.9.2 for the expression of ν.)

1. $\nu = \frac{3}{2}$, $\lambda_1 = 0$: in this case,

$$M^P(X,t) = e^{-\rho t}\frac{1}{m+nX}\exp\left(-MX - NX^2 - \frac{1}{4}(m+nX)^2\right).$$

   The resulting P-dynamic $\mu^P(X,t)$ as determined by (3.12) has the form $\frac{A}{X}+B+CX$ which is more general than affine.

2. $\nu = \frac{1}{2}$, $\lambda_2 = 0$: in this case,

$$M^P(X,t) = e^{-\rho t}\exp\left(-MX - NX^2 - \frac{1}{4}(m+nX)^2\right),$$

   is an exponential-quadratic SDF under physical measure P. The resulting P-dynamic is affine (just as the given Q and R-dynamics in the current setting). This is a strong reminiscence of the SAINTS [29] model introduced in Constantinides (1992). We will

[28] Again, the property $\Phi(a,a;z) = e^z$ $\forall a, z$ of confluent hypergeometric function is behind these results.

[29] Squared autoregressive independent state variable nominal term structure.

formally analyze this model in connection with the framework of our construction 3 below.

Back to general cases, for any values of ν and λ_1, λ_2, we can straightforwardly determine R-SDF $M^R(X,t)$ from (3.21), which in the current quadratic setting is exponential quadratic function of state variable.

In light of proposition 14, we can start out with either affine R-dynamic (then coefficients $\delta_0, \delta_1, \delta_2$ are necessarily related to construction's original parameters by (3.24)), or a new quadratic short rate $r(X) = \rho_0 + \rho_1 X + \rho_2 X^2$ (then $\delta_0, \delta_1, \delta_2$ are exogenous, and only subject to the positivity of $r(X)$ in admissible domain of X). In comparison with the quadratic DTSM of Ahn et al. (2002) and Leippold and Wu (2002), our construction has rich P-dynamic (non-affine drift $\mu^P(X,t)$ and mpr η^{QP}) at the price of a restrictive (functional) SDF $M(X,t)$, yet both give quadratic forward rate and equally tractable bond pricing.

Quotient DTSM

In another simple and special specification of construction 2, Q-dynamic is also affine (a square-root process)

$$\mu^{X,Q}(X) = K_0^Q + K_1^Q X; \qquad H_0 = 0.$$

From (3.23) follows the quotient short rate process

$$r(X) = \theta_{-1}\tfrac{1}{X} + \theta_0 + \theta_1 X, \tag{3.27}$$

$$\theta_{-1} \equiv K_0^Q - K_0^R + \tfrac{(K_0^R)^2 - (K_0^Q)^2 - K_1^Q}{H_1}; \;\; \theta_0 \equiv \rho + \tfrac{2(K_0^R K_1^R - K_0^Q K_1^Q)}{H_1} + K_1^Q; \;\; \theta_1 \equiv \tfrac{(K_1^R)^2 - (K_1^Q)^2}{H_1}.$$

R-SPD is found from the general formula (3.21)

$$M^R(X,t) = e^{-\rho t} X^{(K_0^Q - K_0^R)/H_1} \exp\left(\frac{K_1^Q - K_1^R}{H_1} X\right). \tag{3.28}$$

Not surprisingly, $M^R(X,t)$ has the same polynomial-exponential-affine form as $M^P(X,t)$ in (3.18), because this is the most general functional form of SDF consistent with square-root dynamic in both respective (R and Q, or P and Q) measures, as discussed below (3.21). However, the approaches to these functional SDFs are quite different. Putting this in a simplified way, measure P of the setting leading to (3.18) is similar to measure R of this setting. The current interest rate of the quotient form (3.27) is implied and more general than the linear function $r(X) = a + bX$ of construction 1. Consequently, when we go to the data generating measure, the current SDF $M^P(X,t) \equiv e^{-\rho t}\frac{1}{\phi^P(X)}$ is determined by a special version of different equation (3.11)

$$\frac{1}{2}H_1 X\phi^P_{XX}(X) + (K_0^Q + K_1^Q X)\phi^P_X(X) + (\rho - \theta_{-1}\frac{1}{X} - \theta_0 - \theta_1 X)\phi^P(X) = 0, \qquad (3.29)$$

Consequently, $M^P(X,t)$ is different from (3.18) of previous construction 1, confirmed by the following result.

Proposition 15 *The most general functional stochastic discount factor M^P consistent with construction 2, when the latter is specialized to CIR square-root Q-dynamic (or equivalently, quotient short rate), is*

$$M^{P,\{\lambda_1,\lambda_2\}}(X,t) = e^{-\rho t}\frac{e^{-\alpha X}X^{-\beta}}{\lambda_1\Phi(\delta,\gamma;z) + \lambda_2 z^{1-\gamma}\Phi(\delta-\gamma+1, 2-\gamma; z)}, \qquad (3.30)$$

$$z \equiv \frac{2\left[(K_1^Q)^2 + 2H_1\theta_1\right]^{1/2}}{H_1}X$$

where λ_1, λ_2 are two constants of integration, $\alpha, \beta, \gamma, \delta$ are constant coefficients related to model's parameters given in appendix 3.9.2, and $\Phi(\cdot,\cdot;z)$ is the confluent hypergeometric function of argument z.

Then follow consistently from (3.12) the mpr $\eta^{QP} = \frac{-\sigma^X M^P_X}{M^P}$ and P-dynamic $\mu^{X,P} = \mu^{X,Q} + \sigma^X\eta^{QP}$, which is generally non-affine. Bond pricing is tractable by (3.20), or other transforms presented in Duffie et al. (2000). Again we note that result of proposition 15 holds regardless whether we begin with the given quotient short rate or affine dynamic in some equivalent R. In the latter case, coefficients $\theta_{-1}, \theta_0, \theta_1$ are given in (3.27).

3.4.3 Construction 3: bypassing Q-measure

All our constructions so far have required the specification of Q-dynamic $\{\mu^{X,Q}, \sigma^{X,Q}\}$. Traditionally, this is because pricing in interest rate models is very conveniently (and explicitly) performed in risk-neutral measure as in (3.2), and hence state variables' Q-dynamic can either be directly inferred, for e.g. by observing stock prices, or exogenously chosen to produce closed-form prices. Nevertheless as we see in the previous sections, not only pricing can be done in other equivalent measure (see (3.20)), but also r can be implied (not specified) from functional SDF. We then can generalize the construction by replacing risk-neutral specification by that in any equivalent measure, in place of the risk neutral Q. As a result, we obtain as a special case the class of squared autoregressive independent state variable nominal term structure models of Constantinides (1992).

Construction 3: Let R be an (any) equivalent measure

- *SDF is proper function of state variable in measures P and R: $M^P(X,t) = e^{-\rho t}M^P(X)$, $M^R(X,t) = e^{-\rho t}M^R(X)$. Function $M^R(X)$ is exogenously specified and bounded, whereas $M^P(X)$ is implied.*
- *affine R-dynamic: $\mu^{X,R}(X) = K_0^R + K_1^R X$; $(\sigma^X(X))^2 = H_0 + H_1 X$*

This construction clearly does not rely on any specification involving measure Q. Similar to construction 2, Fourier transform of bounded $M^R(X,t)$ exists and renders tractable zero-coupon bond price (3.20). Also in this bond pricing process, the specification of SDF $M^R(X,t)$ in equivalent measure is a natural replacement [30] for short rate specification $r(X,t)$ in construction 1. The advantage of this construction lies in the arbitrary and infinite choice

[30] From the view of martingale pricing (3.19), $\exp\left(-\int^t r(X,s)ds\right)$ can be defined as SDF in measure Q.

of equivalent measure R, compared to a single and rigid choice of risk-neutral measure Q in no-arbitrage complete markets. As always, the assumption on functional P-SDF helps to pin it down endogenously and consistently. However, as both $r(X,t)$ and $\mu^{X,Q}$ are not given at the onset here, we first have to generalize (3.11) to the current setting of construction 3.

We begin with another generalized version of martingale pricing identity (3.2) for measures P and R

$$E_t^P\left[\frac{M^P(X,T)}{M^P(X,t)}D(X,T)\right] = E_t^R\left[\frac{M^R(X,T)}{M^R(X,t)}D(X,T)\right].$$

The change of measure $P \leftrightarrow R$ is implemented by the Radon-Nikodym derivative $\xi^{RP} = \frac{M^P}{M^R}$ and associated mpr η^{RP}

$$\eta^{RP}(X) = \frac{\mu^{X,P}(X) - \mu^{X,R}(X)}{\sigma^X(X)}; \quad d\xi^{RP}(X,t) = -\xi^{RP}(X,t)\eta^{RP}(X)dZ^P(t).$$

Similar to the derivation of eq. 3.11, by combining Ito's lemma with the same key change of variable $M^P(X,t) = e^{-\rho t}\frac{1}{\phi^P(X)}$ in (3.10), we obtain a differential equation

$$\begin{aligned} &\tfrac{[\sigma^X(X)]^2}{2}\phi^P_{XX}(X) + \left(\mu^{X,R}(X) + \tfrac{[\sigma^X(X)]^2 M^R_X(X)}{M^R(X)}\right)\phi^P_X(X) \\ &+ \left(\tfrac{\mu^{X,R}(X)M^R_X(X)}{M^R(X)} + \tfrac{[\sigma^X(X)]^2 M^R_{XX}(X)}{2M^R(X)} + \tfrac{[\sigma^X(X)]^2[M^R_X(X)]^2}{2[M^R(X)]^2}\right)\phi^P(X) \quad = 0. \end{aligned} \tag{3.31}$$

Since $M^R(X,t)$ and R-dynamic $\{\mu^{X,R}(X), \sigma^X(X)\}$ are all specified in construction 3, this is well-specified second order linear differential equation determining all possible P-SDF $M^P(X,t)$ that are consistent with the construction. Also follow endogenously all other quantities of interest in order: r and η^{RQ} from drift and volatility [31] of M^R respectively, Q-dynamic $\mu^{X,Q} = \mu^{X,R} - \eta^{QR}\sigma^X$, P-dynamic $\mu^{X,P} = \mu^{X,R} + \eta^{RP}\sigma^X$, and finally mpr η^{QP} form volatility of above M^P. In other words, assumptions of construction 3 are indeed self-sufficient. Moreover, the implied interest rate and dynamics in canonical measures P and Q all are non-linear and rich. For a simple illustration of this approach we next consider

[31] That is, $\frac{dM^R}{M^R} = -rdt - \eta^{QR}dZ^R(t)$, where $dZ^Q(t) = dZ^R(t) + \eta^{QR}dt$ and $\eta^{QR} = \frac{\mu^{X,R} - \mu^{X,Q}}{\sigma^X}$.

the class of the squared autoregressive independent state variable nominal term structure (SAINTS) models.

SAINTS models in any equivalent measure

SAINTS models of Constantinides (1992) fit into our construction framework directly, because they are built upon specifying a proper functional SDF. This class of models is constructed originally in physical measure P with the following ingredients: (i) OU mean-reverting P-dynamic $\{\mu^{X,P} = K_0^P + K_1^P X; (\sigma^X)^2 = H_0 = \text{constant}\}$ (ii) exponential-quadratic P-SDF $M^P = e^{-\rho t} e^{(X-a)^2}$. The zero-coupon bond price is tractable because state variable X is a (conditional) Gaussian, and thus M^P is log χ^2 process. The (no-arbitrage) differential representation $\frac{dM^P}{M^P} = -rdt - \eta^{QP} dZ^P(t)$ for P-SDF immediately implies quadratic short rate $r(X)$, and linear mpr $\eta^{QP}(X) = 2\sqrt{H_0}(a - X)$. The latter in turn implies that Q-dynamic is also affine

$$\mu^{X,Q}(X) = \mu^{X,P}(X) - \sigma^X \eta^{QP}(X) = K_0^P - 2aH_0 + (K_1^P + 2H_0)X.$$

We note that the featuring exponential-quadratic form of SDF in SAINTS models can also be precisely established the other way around. Once proper functional (but unspecified) SDF $M^P(X,t)$ is assumed (our key assumption in this paper), above O-U Q-dynamic unambiguously implies an exponential-quadratic SDF as explained in the discussion following (3.21). Hence the two approaches to SAINTS models by specifying either (i) SDF $M^P(X,t)$ (as in Constantinides (1992), and construction 3 more generally) or (ii) risk-neutral dynamic $\mu^{X,Q}(X)$ (as in construction 2) are equivalent.

Our current pursuit of introducing equivalent measure to modeling scheme can extend SAINTS models in a very simple way. Let us specify SAINTS dynamic in an (any) equivalent measure R (instead of P). As a result, bond pricing is equally tractable while physical dynamic $\mu^{X,P}(X)$ and SDF $M^P(X,t)$ are much richer after this extension. Specifically, we consider the following scheme in line with construction 3.

- OU R-dynamic: $\{\mu^{X,R} = K_0^R + K_1^R X; (\sigma^X)^2 = H_0\}$

- Specified R-SDF: $M^R = e^{-\rho t} e^{(X-a)^2}$

- Functional P-SDF: $M^P(X,t) = e^{-\rho t} M^P(X)$ with unspecified (implied) $M^P(X)$

Bond price ZCB

$$ZCB_{t,t+T} = E_t^P \left[\frac{M^P(X(t+T), t+T)}{M^P(X(t), t)} \right] = E_t^R \left[\frac{M^R(X(t+T), t+T)}{M^R(X(t), t)} \right],$$

and the resulting forward rate dynamic are obviously as analytical as in original SAINTS setting, by the interchange $P \leftrightarrow R$. The short rate $r(X,t)$ is also quadratic in X by the same reason. The functional assumption $M^P(X,t)$ pins it down from a differential equation (which is a particular version of the general equation (3.31) adapted to the current exponential-quadratic $M^R(X,t)$). Not surprisingly, as the current short rate is quadratic, this second order linear differential equation has identical form as eq. (3.25). An application of Proposition 14 then immediately yields the P-SDF in our extended SAINTS framework.

$$M^P(X,t) = \frac{e^{-\rho t} e^{-MX - NX^2}}{\lambda_1 e^{\frac{-(m+nX)^2}{4}} \Phi\left(\frac{\nu}{2} + \frac{1}{4}, \frac{1}{2}; \frac{(m+nX)^2}{2}\right) + \lambda_2 (m+nX) e^{\frac{-(m+nX)^2}{4}} \Phi\left(\frac{\nu}{2} + \frac{3}{4}, \frac{3}{2}; \frac{(m+nX)^2}{2}\right)},$$

where λ_1, λ_2 are constants of integration, ν, m, n, M, N are constant coefficients related to model's parameters (appendix 3.9.2). Evidently, this SDF $M^P(X,t)$ is more general than exponential-quadratic function of the original SAINTS[32]. Consequently, both market price of risk $\eta^{QP}(X)$ and state dynamic $\mu^{X,P}(X)$ in physical measure are not confined to linear form in out extension.

3.4.4 Summary

We now briefly summarize the main connections between our functional SDF approach and the key interest rate term structure models in literature, for later uses. The fundamental

[32] This $M^P(X,t)$ becomes an exponential-quadratic function in the special case where $\nu = \frac{1}{2}$ and $\lambda_2 = 0$ (see discussion following Proposition 14), or when equivalent measure R simply coincides the physical P.

assumption here is that stochastic discount factor is a proper (either specified or implied) function of underlying state variable.

When state dynamic follows OU mean-reverting processes under risk-neutral measure Q and another equivalent measure R (which can be P as a special case), $(\sigma^X)^2 = H_0 =$ constant, the functional R-SDF is necessarily a exponential-quadratic function $M^R(X,t) = e^{-\rho t}e^{A+BX+CX^2}$, and the resulting short rate is quadratic function of state variable $r(X) = \rho_0 + \rho_1 X + \rho_1 X^2$. Various special forms of this result are covered in (3.21), (3.24) and SAINTS models.

When state dynamic follows CIR square-root processes under risk-neutral measure Q and another equivalent measure R (which can be P as a special case), $(\sigma^X)^2 = H_0 + H_1 X$, the functional R-SDF is necessarily a polynomial-exponential-affine function

$$M^R(X,t) = e^{-\rho t}e^{A+BX}(H_0 + H_1 X)^C, \tag{3.32}$$

and the resulting short rate is generally quotient function [33] of state variable $r(X) = \theta_{-1}X^{-1} + \theta_0 + \theta_1 X$. Various special forms of this result are covered in (3.17), (3.18), (3.21) and (3.28) [34]. In particular, our above polynomial-exponential-affine function is the most general functional form of SDFs that are compatible with the canonical complete-affine DTSM studied by Dai and Singleton (2000). This form is more general than the exponential-affine SDF widely considered in literature. It is this new generality that will prove to be a very useful feature of affine dynamic models to address the forward premium puzzle in international finance.

[33] In special cases, when model's parameters satisfy certain relations, the quotient short rate can be reduced to linear function, as in affine DTSM underlying eq. (3.17).

[34] Indeed, as SDF can be determined only up to a multiplicative constant, the forms (3.18) and (3.32) are essentially the same, while (3.17) is a special case of (3.32) with $C = 0$.

3.5 Linearity-generating dynamic and beyond

In this section we study the functional stochastic discount factor approach in conjunction with the linearity-generating (LG) processes recently proposed by Gabaix (2009). Asset pricing models based on LG processes possess closed-form prices for both bond and equity as hinted in an early model by Menzly et al. (2004) on return predictability. This tractability can be very useful in illustrating economics mechanisms, for e.g., rare disasters effect underlying price anomalies (see Gabaix (2008)).

Interestingly, the class of LG asset pricing models has SDFs as proper (in fact, linear) functions of state variables and thus is directly related to our current construction. We first studies the original LG dynamic using infinitesimal generators of stochastic calculus. This powerful differential tool places LG models in line with our analysis framework, yields Gabaix (2009)'s key results promptly, and specially, points to possible generalizations of LG modeling approach.

3.5.1 Linearity-generating dynamic and infinitesimal generator

Linearity-generating bond pricing models comprise of (i) underlying LG (vector) process $X(t)$ in physical measure such that $E_t^P[dX(t)] = -\Omega X(t)dt$ where Ω is generator matrix and (ii) SDF is a linear in X: $M^P(X) = \lambda^m \cdot X$ where λ^m is a constant vector [35]. For stock pricing, additional specification is $M^P D(X) = \lambda^{md} \cdot X$, where $M^P D$ is the product of P-SDF M^P and dividend process D.

To simplify the exposition, we employ the infinitesimal generator $\mathcal{D}^{X,P}$ associated with difussion process $X(t)$ (3.1) in measure P. This operator acts on appropriate function $f(X,t)$

[35] Both Ω and λ^m can vary with time t, but must be independent of state variable X to assure tractable asset prices of the model.

and is defined as follows

$$\begin{aligned}\mathcal{D}^{X,P} f(X,t) &\equiv \lim_{\Delta t \to 0} \frac{E_t\left[f(X(t+\Delta t),(t+\Delta t)) - f(X(t),t)\right]}{\Delta t} \\ &= f_t(X,t) + \mu^{X,P}(X,t) f_X(X,t) + \frac{1}{2}\text{Tr}\left[\sigma^X(X,t)\sigma^{X,T}(X,t) f_{XX}(X,t)\right],\end{aligned}$$

where Tr and superscript T denote the trace and transpose operator respectively. Practically for well-behaved functions, $\mathcal{D}^{X,P} f$ presents the drift of the associated diffusion process $df(X,t)$. In special case $f(t,X) = X$, $\mathcal{D}^{X,P} X = \mu^{X,P}(X,t)$. The main ingredients of Gabaix (2009)'s LG asset pricing models are

$$\mathcal{D}^{X,P} X(t) = \mu^{X,P}(X,t) = -\Omega X(t); \quad M(X) = \lambda^m \cdot X; \quad MD(X) = \lambda^{md} \cdot X.$$

The use of infinitesimal generator combined with state-independent property of Ω greatly simplifies the conditional expectation operation. By induction we have

$$\begin{aligned}E_t^P\left[X(t+T)\right] &= X(t) + E_t^P\left[\int_t^{t+T} dX(s)\right] = X_t - \int_t^{t+T} \Omega E_t^P\left[X(s)\right] ds \\ &= X_t - T\Omega X(t) + \Omega^2 \int_t^{t+T}\int_t^s E_t^P\left[X(\tau)\right] d\tau ds = \ldots = e^{-T\Omega} X(t),\end{aligned}$$

where in the last expression, matrix-exponential notation is the limit of the usual Taylor expansion of an exponential function. A sufficient condition for this convergence is that all eigenvalues of Ω be strictly positive. Zero coupon bond price (maturing at $t+T$) then follows immediately because the stochastic discount factor M^P is in the linear span of vector X

$$ZCB(t,X) = E_t^P\left[\frac{M^P(X,t+T)}{M^P(X,t)}\right] = \frac{\lambda^m \cdot E_t^P\left[X(t+T)\right]}{\lambda^m \cdot X(t)} = \frac{\lambda^m \cdot e^{-T\Omega} X(t)}{\lambda_m \cdot X(t)}.$$

Similarly, stock contingent on dividend stream $\{D(X,t)\}$ also possesses closed-form price [36] (assuming usual regularity conditions to interchange the order of integration and expectation

[36]Explicit bond and stock pricing requires additional step of diagonalizing the generator Ω to implement exponential matrix operation $e^{-T\Omega}$.

operations)

$$P(X,t) = E_t^P\left[\int_t^{t+T} \frac{M^P(X,s)D(X,s)}{M^P(X,t)}ds\right]$$
$$= \frac{\lambda^{md}\cdot\int_t^{t+T} E_t^P[X(s)]ds}{M^P(X,t)} = \frac{\lambda^{md}\cdot\left[\int_t^{t+T} e^{-(s-t)\Omega}ds\right]X(t)}{M^P(X,t)}.$$

LG pricing models versus affine DTSMs

At first glance, linear drift $\mu^{X,P} = -\Omega X(t)$ of LG processes is a strong reminiscence of affine DTSMs. However, the resemblance stops here. First, the short rate in LG pricing models can be computed from the differential representation $\frac{dM^P}{M^P} = -rdt - \eta^{QP}dZ^P(t)$ to be

$$r(X) = -\frac{\mathcal{D}^{X,P}M^P(X,t)}{M^P(X,t)} = -\frac{\lambda_m\cdot\mathcal{D}^{X,P}X(t)}{\lambda_m\cdot X(t)} = \frac{\lambda_m\Omega X(t)}{\lambda_m\cdot X(t)}, \tag{3.33}$$

which is rational, but not linear as in affine DTSMs, in X.

Second, as far as bond pricing is concerned, LG models do not place any restriction on the state variable's diffusion $(\sigma^X)^2$. Whereas affine DTSMs specify a linear structure $H_0 + H_1\cdot X$ on this quantity. But as LG models also aim to price stock analytically, they actually also imply some specification on the diffusion. Technically, this specification follows directly the LG model's requirement $-\lambda^{md}\cdot\Omega X(t) = \mathcal{D}^{X,P}[M^P D(X,t)]$

$$\begin{aligned}-\lambda^{md}\cdot\Omega X(t) &= -\lambda^m\cdot\Omega X(t)D(X,t) + \frac{1}{2}M^P(X,t)\mathrm{Tr}\left(\sigma^X(X,t)\sigma^{X,T}(X,t)D_{XX}(X,t)\right)\\ &+ M^P(X,t)\left[-\Omega X(t) + \sigma^X(X,t)\sigma^{X,T}(X,t)\frac{M_X^P(X,t)}{M^P(X,t)}\right]\cdot D_X(X,t).\end{aligned}$$

Plugging in LG explicit specifications $M^P = \lambda^m\cdot X(t)$, $D = \frac{MD}{M} = \frac{\lambda^{md}\cdot X(t)}{\lambda^m\cdot X(t)}$, this clearly is a key dynamics constraint that partially [37]. specifies diffusion σ^X of the underlying LG process.

The two approaches to tractable bond pricing are quite different. Whereas canonical

[37] In settings with vector state variables, this constraints is not sufficient to pin down matrix $\sigma^X\sigma^{X,T}$ unambiguously.

affine DTSMs specify linear short rate, affine Q-dynamic and imply non-linear SDF. LG pricing models specify linear SDF $M^P(X)$, affine drift on P-dynamic and imply non-linear short rate. As a result, affine DTSMs generates a neatly linear forward rate, while LG models can also give analytical stock prices.

3.5.2 Extension to LG modeling

We now analyze the connection between LG pricing models and functional SDF approach, before exploring possible extensions to the former.

LG pricing models versus functional SDF construction

As noted earlier, LG pricing models can be classified as a functional SDF construction. We now explore deeper relation between these two approaches.

The gist of LG modeling starts with the eigen-problem of infinitesimal generator: $\mathcal{D}^P X = -\Omega X$, and then specifies SDF M^P and SDF-dividend product $M^P D$ on linear span of the eigen-basis $X(t)$. Interestingly, our construction can also be neatly built around this fundamental differential operator.

From martingale pricing perspectives, if M^P is the SDF in physical measure and r is interest rate process, then Radon-Nikodym derivative $\xi^{QP} \equiv e^{\int^t r(X,s)ds} M^P(X,t)$ is P-martingale and $\xi^{PQ} \equiv e^{-\int^t r(X,s)ds} \frac{1}{M^P(X,t)}$ is Q-martingale. These imply null drifts under respective measures

$$\mathcal{D}^{X,P} \xi^{QP} = 0; \qquad \mathcal{D}^{X,Q} \xi^{PQ} = 0.$$

In particular, the same second equation implies both differential equation (3.11) after appropriate change of variable in functional SDF construction, and the short rate (3.33) after plugging in $M^P(X,t) = \lambda^m \cdot X$ in LG modeling. In short, our construction specifies $r(X,t)$ and lets loose $M^P(X,t)$, while LG models specify $M^P(X,t)$ and let loose $r(X,t)$.

Similarly, if M^R is the SDF in an (any) equivalent measure, corresponding Radon-

Nikodym derivatives $\xi^{RP} \equiv \frac{M^P(X,t)}{M^R(X,t)}$ and $\xi^{PR} \equiv \frac{M^R(X,t)}{M^P(X,t)}$ are P- and R-martingale respectively, and

$$\mathcal{D}^{X,P}\xi^{RP} = 0; \qquad \mathcal{D}^{X,R}\xi^{PR} = 0.$$

The last equation is eq. (3.31) in our construction 3, and yet there are no counterpart LG models here. A natural and tempting question then is whether the original LG setting can be generalized along the line of introducing an (any) equivalent R. It turns out that LG specifications are invariant with respect to measure. To see this, we assume therefore in some (any) equivalent measure R the followings:

$$\mathcal{D}^{X,R}X(t) = -\Omega X(t); \quad M^R(X,t) = \lambda^m \cdot X(t); \quad M^R D(X,t) = \lambda^{md} \cdot X(t).$$

After defining a new state variable, $\hat{X}(t) \equiv \xi^{RP}X(t) = \frac{M^P}{M^R}X(t)$, we can bring above specifications to physical measure P (recall that Radon-Nikodym derivative ξ^{RP} is a scalar P-martingale)

$$\begin{aligned}
\mathcal{D}^{X,P}\hat{X}(t) = \mathcal{D}^{X,P}\left[\xi^{RP}X(t)\right] = \xi^{RP}\mathcal{D}^{X,R}\left[X(t)\right] = -\xi^{RP}\Omega X(t) = -\Omega\hat{X}(t),\\
M^P(X,t) = \tfrac{M^P(X,t)}{M^R(X,t)}M^R(X,t) = \xi^{RP}\lambda^m \cdot X(t) = \lambda^m \cdot \hat{X}(t),\\
M^P D(X,t) = \tfrac{M^P(X,t)}{M^R(X,t)}M^R D(X,t) = \xi^{RP}\lambda^{md} \cdot X(t) = \lambda^{md} \cdot \hat{X}(t).
\end{aligned}$$

That is, $X(t)$ is LG process under P if and only if $\hat{X}(t)$ is LG process under R. This measure-invariant property shows that LG pricing dynamic is already most general, and thus neutral, with respect to measure rotation. Generalizations to LG setting within the current differential venue, while keeping its analytical pricing power, is still possible after a simple twist.

Extension

The extension starts out with a non-LG, and thus more general, vector dynamic

$$dX(t) = \mu^{X,P}(X,t)dt + \sigma^X(X,t)dZ^P(t),$$

where $\mu^{X,P}(X,t)$ is not necessarily linear in X. Hence this extension is most handy in setting when state variable dynamics are given beforehand possibly to meet other empirical constraints or pricing/statistical aspects of the model. We then wish to construct a new state variable vector Y as (vector) function of $Y = F(X,t)$ that has desirable LG dynamic, $\mathcal{D}^{X,P}F(X,t) = -\omega F(X,t)$, or for all components F^i of vector F

$$\frac{1}{2}\text{Tr}\left[\sigma^X(X,t)\sigma^{X,T}(X,t)F^i_{XX}(X,t)\right] + \mu^{X,P}(X,t)\cdot F^i_X(X,t) + F^i_t(X,t) = -\sum_j \Omega^{ij}F^j(X,t). \tag{3.34}$$

where F^i_X and F^i_{XX} denote respectively gradient vector and Hessian matrix of scalar component F^i. The final step is to specify P-SDF to be linear in Y, $M^P = \lambda^m \cdot Y$ for bond pricing (and $M^PD = \lambda^{md}\cdot Y$ for stock pricing). In the essence, in (3.34) we are building a functional SDF from the general dynamic of underlying state variable X. This extension scheme fits exactly into our construction approach. In the special case of original LG bond pricing models, $F(X)$ is linear in (in fact, identical to) X, $F_{XX} = 0$, which clarifies the irrelevance of volatility specification $\sigma^X(X,t)$ there.

In practice, given the state dynamic $\{\mu^{X,P}(X,t), \sigma^X(X,t)\}$ and a solution Y of eq. (3.34), any function of the form $M^P = \lambda^m \cdot Y$, subject to non negativity and other regularity conditions, is a consistent stochastic discount factor of equally tractable bond pricing model

$$ZCB_{t,t+T} = E^P_t\left[\frac{M^P(X(t+T),t+T)}{M^P(X(t),t)}\right] = \frac{\lambda^m \cdot e^{-T\Omega}F(X,t)}{\lambda^m \cdot F(X,t)},$$

even though the $X(t)$ is not a LG process. The set of dynamic $\{\mu^{X,P}(X,t), \sigma^X(X,t)\}$, that can render a closed-form solution $F(X,t)$ for eq. (3.34), can be much larger than the linear span of X. Then plausibly follows extra flexibilities for LG modeling.

Let us illustrate this extension approach in a simple example of two independent factors $X = (X_1, X_2)^T$ following a $\frac{3}{2}$-power process (see Ahn and Gao (1999)) [38]

$$dX_1(t) = X_1(t)[a_1 - X_1(t)]dt + \sqrt{2}[X_1(t)]^{3/2}dZ_1^P(t),$$
$$dX_2(t) = X_2(t)[a_2 - X_2(t)]dt + \sqrt{2}[X_2(t)]^{3/2}dZ_2^P(t).$$

We can easily verify that the following transformed state variable $Y(t)$ is a LG process with generator Ω

$$Y(t) = \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} = e^{-\rho t}\begin{pmatrix} X_1(t)[a_1 + X_1(t)] \\ X_2(t)[a_2 + X_2(t)] \end{pmatrix}; \qquad \Omega = \begin{pmatrix} \rho - a_1 & 0 \\ 0 & \rho - a_2 \end{pmatrix}.$$

This suggests, for two constants $\lambda^{m,1}, \lambda^{m,2}$, the SDF

$$M^P(X,t) = e^{-\rho t}\left\{\lambda^{m,1}X_1(t)[a_1 + X_1(t)] + \lambda^{m,2}X_2(t)[a_2 + X_2(t)]\right\},$$

and the resulting bond prices

$$ZCB_{t,t+T} = \frac{\lambda^{m,1}e^{-T(\rho-a_1)}X_1(t)[a_1 + X_1(t)] + \lambda^{m,2}e^{-T(\rho-a_2)}X_2(t)[a_2 + X_2(t)]}{\lambda^{m,1}X_1(t)[a_1 + X_1(t)] + \lambda^{m,2}X_2(t)[a_2 + X_2(t)]}.$$

We note that while there exists closed-form general solution of eq. (3.34) for this specific dynamic, we can be content with some simple and special solutions. This is because any solution, regardless of how special, is consistent with same state X dynamic and has identical LG pricing power by construction. In practice, this feature renders both flexibility and ease to incorporate extension to LG modeling.

[38] This process evidently does not belong to linearity-generating class.

3.6 Application: The forward premium puzzle

We study in this section an application of our functional SDF construction to a pricing anomaly in international finance. This anomaly is commonly known as forward premium puzzle (FPP) (a.k.a., uncovered interest rate parity puzzle): on relative basis appreciating currencies tend to be also associated with increasing interest rates. Generally speaking and assuming complete market, the forward premium puzzle can be very conveniently discussed using the apparatus of stochastic discount factor. In particular, in the same international finance setting, Bakshi et al. (2008) construct exogenous SDFs that accommodate both local and global risk to produce stochastic risk premia consistent with data in currency option market. Our construction instead concentrates on the consumption risk and the general equilibrium aspects of the pricing model by solving the endogenous SDFs as proper function of (consumption) state. We attempt to shed light into the necessary ingredients of investor rational behaviors (preferences) and canonical equilibrium models that are compatible with this international asset pricing anomaly.

3.6.1 Forward premium puzzle and affine dynamic

To set the notation, we use the standard superscripts h and f to denote quantities pertaining to home and foreign countries respectively. Let $S(t)$ be the exchange rate available at time t, namely $S(t)$ units of home currency exchange for one unit of foreign currency then. Consider any payoff $D^f(T)$ available at a future time T and denominated in foreign currency. We can compute its current value in home currency by converting either its current foreign value to home currency, or the payoff to home currency first. By no arbitrage, the two approaches give identical value

$$S(t)E_t^P\left[\frac{M^{f,P}(T)}{M^{f,P}(t)}D^f(T)\right] = E_t^P\left[\frac{M^{h,P}(T)}{M^{h,P}(t)}S(T)D^f(T)\right], \tag{3.35}$$

where $M^{i,P}$ is country i's stochastic discount factor in physical measure

For simplicity we do not assume information asymmetries between countries [39], and consequently all countries use identical prior distribution for pricing. In complete market settings, the SDFs are unique, which implies (omitting multiplicative factor immaterial for the dynamic under investigation) for all time t

$$S(t) = \frac{M^{f,P}(t)}{M^{h,P}(t)} \quad \text{or} \quad s(t) = m^{f,P}(t) - m^{h,P}(t),$$

where lower-case letters denote logarithms of appropriate quantities. The FPP can be quantitatively and succinctly expressed as negative unconditional covariance between the changes ds in log exchange rates and the interest rates' differential (all expectation and covariance in this section are with respect to physical distribution)

$$\sigma^{ds,\Delta r} \equiv Cov\left(ds(t), r^h(t) - r^f(t)\right) < 0,$$

This covariance is indeed proportional to the slope coefficient of Fama (1984)'s forward premium regression, the negative sign of which constitutes a necessary condition of the puzzle. Plugging in the above formula for exchange rate S and applying Ito's lemma yield a more explicit representation for the unconditional covariance (we hereafter omit the factor dt to simplify the exposition)

$$\sigma^{ds,\Delta r} = Cov\left(r^h(t) + \frac{[\eta^{QP,h}]^2}{2} - r^f(t) - \frac{[\eta^{QP,f}]^2}{2}, r^h(t) - r^f(t)\right) < 0, \tag{3.36}$$

where $\eta^{QP,i}$ is country i's market price of risk, see (3.4). This signed relation, observed empirically for majority of countries pairs, implies that foreign currency's appreciation ($ds(t)$ increases) tends to go hand in hand with relative increase in foreign interest rate (interest rate differential $r^h(t) - r^f(t)$ decreases) and vice versa [40]. The directions of these comovements constitute a puzzle, apparently it looks like international investors would demand a

[39] Though this possibility may fit very well into our introduction of an equivalent measure R to pricing model construction in general.

[40] Eq. (3.38) presents a restricted version of the puzzle's counterintuition, namely country's market price of risk necessarily moves in the opposite direction of it's interest rate.

lower premium (i.e., interest rate) for holding a depreciating currency.

Motivated by the power of affine dynamic framework in modeling interest rates and associated term structures, Backus et al. (2001) explore the FPP within given affine dynamic setting. They find that it is difficult, both theoretically and empirically, to accommodate the puzzle with the given dynamic. In the relation with this stalemate, our construction can be best illustrated to be able to overcome these difficulties, and at the same time provides a viable risk-based explanation for the puzzle. Albeit Backus et al. (2001) employ a somewhat special class of affine dynamic detailed next, the flexibility and advantage of our construction manifests itself in that it is bound to the same dynamic restrictions. Before presenting the construction, we briefly sort through Backus et al. (2001)'s arguments.

Backus et al. (2001) consider symmetric, independent and single-factor Q-affine dynamic for each country [41] together with positive interest rates linear in respective state variables

$$\begin{cases} dX^i(t) = (K_0^{i,Q} + K_1^{i,Q} X^i)dt + \sqrt{H_0^i + H_1^i X^i} dZ^{i,Q}(t); \\ r^i(X) = a^i + b^i X(t); \qquad E_t\left[dZ^{h,Q}(t) dZ^{f,Q}(t)\right] = 0; \end{cases} \quad i \in \{h, f\}. \tag{3.37}$$

In particular, they assume the standard completely affine dynamic setting of Dai and Singleton (2000), wherein market price of risk is proportional to the volatility of state variable: $\eta^{QP,i}(X^i, t) \sim \sigma^{X^i} = \sqrt{H_0^i + H_1^i X^i}$, for $i \in \{h, f\}$, so that dynamic in physical measure P is also affine (see discussion below eq. (3.17)). In this setting, the condition (3.36) reads

$$\begin{aligned} \sigma^{ds,\Delta r} &= Var\left(r^h(t)\right) + \frac{1}{2} Cov\left([\eta^{QP,h}]^2, r^h(t)\right) + (h \leftrightarrow f) \\ &= (b^h)^2 Var\left(X^h(t)\right) + \frac{1}{2} H_1^h b^h Var\left(X^h(t)\right) + (h \leftrightarrow f) < 0. \end{aligned} \tag{3.38}$$

where $(h \leftrightarrow f)$ denotes the repetition of terms but with concerning home quantities being replaced by foreign counterparts. This serves both to shorten the notation and to emphasize the current symmetric setting. The key observations of Backus et al. (2001) are as follows. First, for either countries, the admissible domain $H_0^i + H_1^i X^i > 0$ for positive, possibly un-

[41] Note that symmetric factors common to both countries do not contribute in any way to the covariance (3.36) by mutual cancellation.

bounded, square-root process $X^i(t)$ implies that $H_1^i > 0$ (more rigorous discuss of regularity conditions is presented in the next section). Second and by the same reason, almost surely positive interest rate assumptions $a^i + b^i X^i > 0$ require both $b^h, b^f > 0$. These two observations render the inequality (3.38) impossible, and consequently the FPP inconsistent with the given affine dynamic setting. Backus et al. (2001) then relax assumptions on single and independent factors, still they found that this more general setting fares poorly in addressing empirically the puzzle. Their paper also points to a possible modeling solution, which allows interest rates to assume negative values with some positive probabilities. We keep intact all original dynamic restrictions and instead propose a natural generalization of market prices of risk to tackle this deadlock. Conceptually and more importantly, our construction also points to a risk-based story behind the puzzle.

3.6.2 A risk-based class of FPP-consistent models

As summarized in the discussion leading to eq. (3.32) (and also (3.17)), we see that the completely affine dynamic employed in the above study of forward premium puzzle can be implied by the exponential-affine SDF

$$M^P(X,t) = e^{-\rho t} e^{A+BX} \Longrightarrow \eta^{PQ}(X) \sim \sigma^X = \sqrt{H_0 + H_1 X}. \tag{3.39}$$

This and our earlier observations then motivate a simple generalization of the complete-affine setting.

The model

Our FPP-consistent model is based on construction 1 in section 3.4.1. We now start out with a functional SDF of the more general polynomial-exponential-affine form (3.32) for each countries in physical measure P. As a result, the implied market prices of risk have much richer structure[42]

$$M^{i,P}(X^i,t) = e^{-\rho^i t} e^{A^i + B^i X^i} (H_0^i + H_1^i X^i)^{C^i}, \tag{3.40}$$

$$\Longrightarrow \eta^{i,QP}(X^i) = -B^i \sqrt{H_0^i + H_1^i X^i} - \frac{C^i H_1^i}{\sqrt{H_0^i + H_1^i X^i}} \qquad i \in \{h, f\}. \tag{3.41}$$

We note, however, that in one regard this specification is simplistic because currently the above square-root process $X(t)$ and SDF $M^{h,P}$ are stationary. For more realistic construction, we can overcome this shortfall by augmenting the state spaces, and adding (identical) non-stationary multiplicative factors to the SDF of both home and foreign economies (see section 3.7).

For the sake of simple exposition, we hereafter adopt the convention

$$H_0^i = 0, \qquad \forall i \in \{h, f\}.$$

This amounts equivalently to an innocuous change of variable $X^i \to \hat{X}^i \equiv X^i + \frac{H_0^i}{H_1^i}$, and does not affect the validity of our construction in any way. We otherwise retain exactly the specification (3.37) used by Backus et al. (2001), namely (i) linear, almost surely positive short rates and (ii) independent, symmetric Q-affine dynamic for each country. With our choice of polynomial-exponential-affine SDF (3.40), the dynamics are also affine in physical measure P

$$dX^i(t) = (K_0^{i,P} + K_1^{i,P} X^i)dt + \sqrt{H_1^i X^i} dZ^{i,P}(t),$$
$$K_0^{i,P} = K_0^{i,Q} - C^i H_1^i, \qquad K_1^{i,P} = K_1^{i,Q} - B^i H_1^i \qquad i \in \{h, f\}.$$

Hereafter we work exclusively with the dynamic specification $\{K_0^{i,P}, K_1^{i,P}\}$ in measure P. All findings below concerning this specification can be immediately obtained for the risk-neutral $\{K_0^{i,Q}, K_1^{i,Q}\}$ specification by reversing the above linear relations between these two sets.

[42]Cheridito et al. (2007) first obtain this form of the market price of risk in their extended affine DTSM setting. Here we generate it from a polynomial-exponential-affine SDF.

The key FPP necessary condition (3.36) now reads

$$
\begin{aligned}
\sigma^{ds,\Delta r} &= Var\left(r^h(t)\right) + \frac{1}{2}Cov\left([\eta^{QP,h}(t)]^2, r^h(t)\right) + (h \leftrightarrow f) \\
&= (b^h)^2 Var\left(X^h(t)\right) + \frac{1}{2}(B^h)^2 H_1^h b^h Var\left(X^h(t)\right) \\
&+ \frac{1}{2}(C^h)^2 H_1^h b^h Cov\left(X^h(t), \frac{1}{X^h(t)}\right) + (h \leftrightarrow f) < 0.
\end{aligned}
\tag{3.42}
$$

Compared to previous structure (3.38), our construction offers a key new ingredient, namely the covariance terms in (3.42) for the FPP regression coefficient $\sigma^{ds,\Delta r}$. These terms stem from the richer SDF and associated mpr (3.40), and interestingly are invariably negative. Under mild condition (see the proof of proposition 16 in the appendix), we also have a very convenient approximation as an application of delta method

$$
Cov\left(X^i(t), \frac{1}{X^i(t)}\right) \approx -\frac{Var\left(X^i(t)\right)}{\left(E\left[X^i(t)\right]\right)^2} \qquad i \in \{h, f\}. \tag{3.43}
$$

Next, we need to examine whether their values can be small enough to drag the full $\sigma^{ds,\Delta r}$ into the negative-valued domain for FPP to be consistent. The analysis also helps to understand the economics intuition behind working models.

1. Feller's admissibility condition: For the square-root processes (3.37) under consideration, $X^i(t)$ will be strictly positive almost surely when following conditions hold (recall that we currently set $H_0^i = 0\ \forall i$ for simplicity)

$$
2K_0^{i,P} > H_1^i > 0 \qquad i \in \{h, f\}, \tag{3.44}
$$

   where the first inequality is Feller's condition, the second is a regularity to make sure that the square-root operations $\sigma^{X^i} = \sqrt{H_1^i X^i}$ do not generate complex-valued volatilities for all admissible $X^i(t)$.

2. Linear interest rates: As summarized in section 3.4.4, the polynomial-exponential-affine

SDF of the forms (3.32) or (3.40) generally leads to quotient short rates

$$r^i(X^i) = -\frac{1}{dt}E_t^P\left[\frac{dM^{i,P}(X^i,t)}{M^{i,P}(X^i,t)}\right] = C^i\left(\frac{H_1^i(1-C^i)}{2} - K_0^{i,P}\right)\frac{1}{X^i} + \left(\rho^i - K_0^{i,P}B^i - K_1^{i,P}C^i - B^iC^iH_1^i\right) + B^i\left(-K_1^{i,P} - \frac{H_1^iB^i}{2}\right)X^i. \quad (3.45)$$

Hence we impose the following parametric relation to enforce linear short rates (by getting rid of $\frac{1}{X^i}$-term) as a requisite following Backus et al. (2001).

$$K_0^{i,P} = \frac{H_1^i(1-C^i)}{2} \qquad i \in \{h, f\}. \quad (3.46)$$

Note that the same specification gives rise to the linear short rate of our earlier construction 1 of section 3.4.1 (see the discussion following eq. (3.32)). Plugging this specification into immediately above expression for short rate, we indeed have

$$r^i = a^i + b^iX^i \quad \text{with} \quad \begin{cases} a^i = \rho^i - K_0^{i,P}B^i - K_1^{i,P}C^i - B^iC^iH_1^i \\ b^i = -B^i\left(K_1^{i,P} + \frac{H_1^iB^i}{2}\right) \end{cases} \quad i \in \{h, f\}. \quad (3.47)$$

3. Non-negative interest rates: now as $X^i(t)$ are strictly positive (possibly unbounded), like before, the conditions to assure non-linear short rates $r^i = a^i + b^iX^i$ are $b^i > 0\ \forall i$ or

$$B^i\left(K_1^{i,P} + \frac{H_1^iB^i}{2}\right) < 0 \qquad i \in \{h, f\}. \quad (3.48)$$

We are ready to tackle the elusive sign of FPP covariance (3.42). We note that as a prerequisite inherited from last section, the parametric setting is presumably symmetric between countries, so similar parametric conditions are to be enforced in both countries. The most plausible and robust sufficient specifications are to make the covariance terms $Cov\left(X^i, \frac{1}{X^i}\right)$, being always negative, the dominant contribution in (3.42) and thus render a FPP-consistent negative $\sigma^{ds,\Delta r}$

$$(C^i)^2H_1^i \gg b^i \gg (B^i)^2H_1^i \qquad i \in \{h, f\}.$$

Combining this with above conditions (3.44), (3.46), (3.48) we finally arrive at the core specifications (3.49) of the FPP-consistent construction.

Proposition 16 *In an international asset pricing model possessing all of the following properties (where $i \in \{h, f\}$)*

(i) polynomial-exponential-affine functional SDF in physical measure [43]

$$M^{i,P}(X^i, t) \sim e^{-\rho t} e^{B^i X^i} (X^i)^{C^i}$$

(ii) affine, independent and symmetric-across-countries state dynamic (3.37)

(iii) additional parametric specifications

$$B^i > 0; \quad C^i < 0; \quad H_1^i > 0; \quad K_0^{i,P} = \frac{H_1^i(1 - C^i)}{2}; \quad -K_1^{i,P} \gg B^i H_1^i, \tag{3.49}$$

the change in exchange rates correlates negatively with interest rate differential as in (3.36), i.e., the forward premium puzzle holds.

Within linear interest rate class (3.46), (3.47) our construction is both consistent and robust with respect to forward premium anomaly in the sense that the FPP-consistent specifications (3.49) can be easily satisfied for a wide range of parameters. The most decisive resulting constraints among all is on B^i, as long as B^i has small enough values permissible domains of other parameters implied by our construction will widen substantially (see appendix for details). A more important task of uncovering any possible risk-based intuitions underlying the anomaly, within this line of construction, warrants a thorough examination of nature of these restrictions.

The risk story

The independence of risk factors between countries transforms FPP into separate intra-country anomalies in (3.42). That is, as long as (squared) market price of risk $(\eta^{i,QP})^2$

[43] This is (3.40), under the convention $H_0^i \equiv 0$. Consequently, note that though parameters $\{H_1^i\}_{i=h,f}$ are not in these SDFs $\{M^{i,P}\}_{i=h,f}$, they still contribute fundamentally to the forward premium via their role in the state dynamic volatilities.

and interest rate r^i moves in opposite directions *within each* country i, the countries' cross independence will only *add up* these negative correlations and push the (squared) market price of risk differential $(\eta^{h,QP})^2 - (\eta^{f,QP})^2$ changes further *away* from those of interest rate differential $r^h - r^f$. Motivated both by the desire to preserve the original Backus et al. (2001)'s modeling framework and this clear-cut separation, we have also assumed this same independence. Our strategy actually relies more fundamentally on the mean-reverting dynamic of consumption $X(t)$, and establishes FPP-consistent relations between key economic quantities, shown schematically below.

$$\left.\begin{array}{c} M^{h,P} \downarrow\uparrow S, \\ M^{h,P} \downarrow\uparrow \left(\eta^{h,QP}\right)^2 \downarrow\uparrow r^h \Rightarrow M^{h,P} \uparrow\uparrow r^h \end{array}\right\} \Rightarrow r^h \downarrow\uparrow S.$$

Symbols $\downarrow\uparrow$, $\uparrow\uparrow$ respectively denote same and opposite directions of comovements. These movements should be taken only in statistical sense. As such, the concluding thesis $r^h \downarrow\uparrow S$ expresses FPP observed in the data: all else being equal, home interest rate r^h likely drops ($r^h \downarrow$) when home currency depreciates (exchange rate [44] $S \uparrow$), and vice versa. Here is our risk story underlying all of the above linkages.

$\mathbf{M^{h,P} \downarrow\uparrow S}$: This relation is a mechanical consequence of the assumption which keeps foreign economy intact. Nevertheless, the intuition is that, all else being equal, home currency is likely to depreciate when home risk is likely to increase. This can be best inferred from no arbitrage relation (3.35)

$$S(t)E_t^P\left[\frac{M^{f,P}(T)}{M^{f,P}(t)}D^f(T)\right] = E_t^P\left[\frac{M^{h,P}(T)}{M^{h,P}(t)}S(T)D^f(T)\right].$$

Keeping all but $S(T), M^{h,P}(T)$ fixed for a simple exposition, increases in home country's future risk prompt investors to apply more aggressive discount scheme ($M^{h,P}(T) \downarrow$) there. Facing such cloudy prospects at home, investors are to accept a time-T payoff $S(T)D^f(T)$ in *home currency* only when projected exchange ratio is sufficiently attractive ($S(T) \uparrow$) into

[44]Recall that at time t, $S(t)$ units of home currency exchange equivalently for one unit of foreign currency.

the future.

$\mathbf{M^{h,P}} \downarrow\uparrow (\eta^{\mathbf{h,QP}})^{\mathbf{2}}$: The reasoning here is already covered by above argument. Increases in future risk simultaneously boost both required risk premium $((\eta^{h,QP})^2 \uparrow)$ and discounting $(M^{h,P}(T) \downarrow)$. Reassuringly, this is the just a statement of quantitative relation (3.4).

$(\eta^{\mathbf{h,QP}})^{\mathbf{2}} \downarrow\uparrow \mathbf{r^h}$: This key relation is a more subtle, but is also keenly implied by our construction. Nevertheless, without diving into full-blown rigors, the intuition is again very simple. When home market goes up, home riskless bonds lose their appeals (equilibrium interest rate surges $r^h \uparrow$), at the same time risk-averse investors likely wary less about risk (equilibrium risk premium drops $(\eta^{h,QP})^2 \downarrow$), and vice versa. We now substantiate this intuition via canonical consumption risk embedded in state variable $X(t)$ itself: $(\eta^{h,QP})^2 \downarrow\uparrow X(t) \uparrow\uparrow r^h$. We now argue for these relations in turn.

The signs $B^i > 0$, $C^i < 0$ in specification (3.49) fit particularly well into a consumption-based story. For illustration, we consider a setting wherein positive and mean-reverting state variable X^i_t is simplistically identified with consumption[45], with explicit (growth) dynamic (3.37) (recall $H^i_0 \equiv 0$)

$$\frac{dX^i(t)}{X^i(t)} = \frac{\mu^{X^i,P}}{X^i(t)}dt + \frac{\sigma^{X^i}}{X^i(t)}dZ^i(t) = \left(\frac{K_0^{i,P}}{X^i(t)} + K_1^{i,P}\right)dt + \sqrt{\frac{H_1^i}{X^i(t)}}dZ^i(t).$$

Recall that (3.49) also implies $K_0^{i,P} > 0$, $K_1^{i,P} < 0$, or assuringly state variable dynamic is mean-reverting in each country. We conventionally associate *good states* of economy with *large realized values* of endowment $X^i(t)$. As such, both consumption expected growth $\frac{\mu^{X^i,P}}{X^i}$ and growth volatility $\frac{\sigma^{X^i}}{X^i}$ drop in good states. These dynamics are key to of the risk-based explanation of the forward premium puzzle.

In equilibrium, country i's representative agent has (additive) marginal utility of the desired polynomial-exponential-affine form $U^i_X(X^i,t) = M^{i,P}(X^i,t) = e^{-\rho^i t}e^{B^i X^i}(X^i)^{C^i}$. Since

[45]This identification is somewhat simplistic, but X^i can easily be enriched with additional factors by state space augmentation techniques discussed in section 3.7.

$M^{i,P} > 0$, the associated utility [46] is increasing. It also has the hybrid appeal of exponential and power preferences (though $B^i > 0$ here). We may see this alternatively in the implied risk aversion γ^i

$$\gamma^i(X^i, t) \equiv -\frac{X^i U^i_{XX}(X^i, t)}{U^i_X(X^i, t)} = -\frac{X^i M^{i,P}_X(X^i, t)}{M^{i,P}(X^i, t)} = -C^i - B^i X^i.$$

The first constant component is from the power utility factor, and the second linear component from the exponential. For almost power preference, quantified by parameters' choices $B^i K_0^{i,P} \ll C^i K_1^{i,P}$, the consumptions X^i mean-revert about a positive level $\frac{K_0^{i,P}}{-K_1^{i,P}}$. This mean value is well below $\frac{-C^i}{B^i}$ and the above risk aversion coefficient γ^i is always positive. In other words, utilities are convex, and representative agents have decreasing positive risk aversion for all admissible consumptions as rationally desired. This downward-sloping behavior of $\gamma^i(X^i)$ is due to exponential factor of preference. For small enough B^i (which is most relevant for our construction), power behavior dominates the preference, wherein risk-averse representative agent demands lower risk premia in better states of the economy. That is, market price of risk $\eta^{QP,i}(X^i)$ *decreases* with consumption X^i. Indeed, a transparent quantitative analysis confirms this intuition (a direct but less intuitive computation on (3.40) also does the job)

$$\frac{\partial}{\partial X^i}\eta^{QP,i}(X^i) = \frac{\partial}{\partial X^i}\left(\gamma^i(X^i, t)\frac{\sigma^{X^i}(X^i)}{X^i}\right) < 0.$$

The inequality results from observation that both risk aversion coefficient γ^i and market growth's volatility $\frac{\sigma^{X^i}}{X^i} \sim \frac{1}{\sqrt{X^i}}$ drops when market goes up. Consequently, investors demand less premium for holding consumption-contingent asset and market price of risk decreases in good states in our rational model: $\left(\eta^{h,QP}\right)^2 \downarrow\uparrow X(t)$.

In light of consumption risk, the same interest rate (3.45) can also be recast in a very lucid form related to risk aversion coefficient (square-bracket expression) and precautionary

[46] The associated utility is $U^i(X^i, t) \sim e^{-\rho^i t} \int^{X^i} e^{-\rho^i t} e^{B^i Y} Y^{C^i} dY$.

savings ($\theta^i \equiv -X^i \frac{M^{i,P}_{XX}}{M^{i,P}_X}$, curly-bracket expression)

$$r^i(X^i) = \rho^i + \left[\frac{\mu^{X^i,P}(X^i)}{X^i}\gamma^i(X^i)\right] - \left\{\frac{1}{2}\frac{[\sigma^{X^i}(X^i)]^2}{(X^i)^2}\gamma^i(X^i)\theta^i(X^i)\right\}$$

$$= \text{constant} + \left[\frac{-C^i K^{i,P}_0}{X^i} - B^i K^{i,P}_1 X^i\right] - \frac{1}{2}H^i_1\left\{(B^i)^2 X^i + \frac{C^i(C^i-1)}{X^i}\right\}. \qquad (3.50)$$

Interestingly, all C^i-terms originated from power preference load on $\frac{1}{X^i}$, all B^i-terms from exponential preference load on X^i. This signals a very interesting interaction between intertemporal consumption smoothing desire (square-bracket terms), precautionary saving motives (curly-bracket terms) and the mean-reverting consumption dynamic. It is this rich interplay that can give rise to FPP-consistent behaviors of equilibrium interest rates.

- Linear interest rate specification: Under this restriction (3.46), all $\frac{1}{X^i}$-terms cancel out and (3.50) coincides with (3.47). We now concentrates only on non-canceling terms (other terms will be studied next). Positive shocks in X^i boost up elasticity of intertemporal substitution $\frac{1}{\gamma^i(X^i)}$. But as investors face negative expected consumption growth (mean- reverting coefficient $K^{i,P}_1 < 0$ in $\frac{\mu^{X^i,P}}{X^i}$), the intertemporal consumption smoothing (term $-B^i K^{i,P}_1 X^i > 0$) increases in good states (large X^i). This effect contributes to a surge in interest rate. Furthermore, when preference is mostly of power type ($B^i \ll C^i$), we can always disregard the second-order term originated from precautionary saving (term $\frac{1}{2}(B^i)^2 H^i_1 X^i$). Thus, in our linear interest rate specification mimicking Backus et al. (2001)'s, intertemporal consumption smoothing dominates over precautionary saving motives, and interest rate moves in the direction of economy: $r^i \uparrow\uparrow X^i$.

- Quotient interest rate specification: For the sake of completeness, we now venture out of domain of linear interest rate and additionally consider all $\frac{1}{X^i}$-terms in (3.50). Under almost power preferences, consumption expected growth actually drops in good states due to mean reversion, which prompts investors to trim current consumption and save more (term $\frac{-C^i K^{i,P}_0}{X^i}$). As a result, interest rate decreases. But at the same time growth volatility $\frac{\sigma^{X^i}}{X^i}$ also drops in good states. Investors then reduce their precautionary saving motives (term $\frac{1}{2}H_1^i\frac{C^i(C^i-1)}{X^i}$) and consequently interest rate increases. When either investors are risk averse or consumption is variable enough, $H_1^i \geq \frac{K_0^i}{-C^i} > 0$, the precautionary saving effects dominate [47] and interest rate again surges in good states of economy: $r^i \uparrow\uparrow X^i$.

Altogether, our construction tells a risk story on the dominant negative correlation between riskless rate and risk premium in (3.42), as the two moves oppositely with respect to consumption $X^i(i)$. Looking back, Backus et al. (2001) consider a dynamic setting that can be implied from pure exponential preference (3.39). We inherit this structure, but add to the picture a *dominant* factor of power type (3.40), after which the model becomes FPP-consistent. It might appear that power utility is all we need for the story to work here, and in particular it might also be tempting to set $B^i = 0$. However, the remarkable and relevant role of this exponential preference $e^{B^iX^i}$ in SDF is in fostering appropriate degree of interest rate variability: slope coefficient b^i (3.47) is proportional to B^i. Graveline (2006) estimates a two-country pricing model of extended affine class, in which the market price of risk dynamic is exogenously specified to have the form similar to (3.41). He shows that extended affine dynamic is consistent with FPP in the data. This study thus supports our construction empirically. Since these same dynamics are being implied from the consumption model, we have gained further viable risk-based intuitions which drive forward premium anomaly in this setting.

There is neither consumption risk sharing nor trading at international level here, so in such aspects, the construction is a rather simplistic version of real world. Nevertheless, it serves our aim to demonstrate the advantage of functional stochastic discount factor approach, the principal theme of current paper, in constructing economic models to address certain economic and price phenomena. Introducing interdependence between countries' risk factors will certainly enrich the model in many relevant ways. This however calls for a multi-factor generalization of our functional stochastic discount factor construction, the subject of

[47] Note that now real interest rate may have negative values, but also we currently are not strictly bound to linear interest rate specification of Backus et al (2001).

next section.

3.7 Multi-factor settings and estimation

With the exception of section 3.5.2, so far our constructions have been confined within one-factor setting wherein state variable $X(t)$ is unidimensional (scalar) process. Realistic economics problems are usually driven most certainly by multiple state variables. It is always desirable to have analytical frameworks, such as the original affine DTSM or linearity-generating dynamic, that can handle several correlating factors. In this section first we show that our functional stochastic discount factor approach also works in multiple-factor settings, then briefly outline the quasi-maximum likelihood estimation procedure for the construction.

3.7.1 Multi-factor setting

The state variable $X(t)$ now is a vector-valued diffusion process in $\mathbf{R}^n$, driven by m independent standard Brownian motions $Z(t) \in \mathbf{R}^m$. To present key ingredients of the multi-factor generalization, below we work with the basic construction (section 3.3.2). Generalizations to other constructions follow in analogy.

Here the state dynamic specification in risk-neutral measure (n-vector drift $\mu^{X,Q}(X)$, $n \times m$-matrix volatility $\sigma^X(X)$) and scalar short rate process $r(X)$ are given in conjunction with the featuring assumption that SDF in physical measure $M^P(X,t) = e^{-\rho t} M^P(X)$ be proper function of state variable (and time). Under these condition, we can explicitly apply Ito's lemma on general function $M^P(X,t)$ and identify outcomes with martingale differential representation (3.4). This results in multi-factor counterpart of eq. (3.9), which now is a second order partial differential equation (PDE)

$$\begin{aligned} &\tfrac{1}{2}\mathrm{Tr}\left(\sigma^X(X,t)\sigma^{X,T}(X,t)M^P_{XX}(X)\right) + \mu^{X,Q}(X)\cdot M^P_X(X) \\ &+ \left(r(X) - \tfrac{\mathrm{Tr}\left[M^{P,T}_X(X)\sigma^X(X)\sigma^{X,T}(X)M^P_X(X)\right]}{(M^P(X))^2} - \rho\right) M^P(X) = 0, \end{aligned}$$

where n-vector M_X^P and $n \times n$-matrix M_{XX}^P respectively are gradient and Hessian of P-SDF. The convenient change of variable $\phi^P(X) \equiv \frac{1}{M^P(X)}$ nicely transforms the above non-linear differential equation into a *linear* one, just as (3.11)

$$\frac{1}{2}\text{Tr}\left(\sigma^X(X,t)\sigma^{X,T}(X,t)\phi_{XX}^P(X)\right) + \mu^{X,Q}(X) \cdot \phi_X^P(X) + (\rho - r(X))\phi^P(X) = 0. \quad (3.51)$$

At the first look, it is apparent that there is little hope to pin down the multivariate SDF from this PDE, given $\{\mu^{X,Q}(X), \sigma^X(X)\}, r(X)$. A closer examination concurs that this vagueness actually poses a very practical advantage for our current multi-factor construction. First, the gist of this construction (see proposition 11) is that any solution of differential equation (3.51), subject to regularity conditions to rule out arbitrage, can be a SDF consistent with the given dynamic. That is, we do not need to solve this PDE in full generality, a very difficult task in multi-dimensional setting. It turns out that, for the flexible equation (3.51), not only it is much quicker to obtain special and consistent solutions, but also one has more room to impose and accommodate economics-motivated structures on these solutions. Let us illustrate this point in a simple specific example, motivated by separation-of-variable class of special solutions.

We consider now a two-correlated-factor model, and for notational clarity we write $(X, Y)^T$ in place of the 2-vector state variable X above. Similar to construction 1 in section 3.4.1, we specify a linear interest rate and Q-affine dynamic for state variable

$$r = a + b^x X + b^y Y; \qquad \begin{pmatrix} dX \\ dY \end{pmatrix} = \mu^Q dt + \sigma \begin{pmatrix} dZ^{X,Q}(t) \\ dZ^{Y,Q}(t) \end{pmatrix}, \quad (3.52)$$

where $Z^{X,Q}(t)$, $Z^{Y,Q}(t)$ are uncorrelated standard Brownian motions in risk neutral measure Q and

$$\mu^Q \equiv \begin{pmatrix} \mu^{X,Q} \\ \mu^{Y,Q} \end{pmatrix} = \begin{pmatrix} k_0^x + k_1^{xx} X \\ k_0^y + k_1^{yx} X + k_1^{yy} Y \end{pmatrix};$$

$$\sigma\sigma^T \equiv \begin{pmatrix} \sigma^{2xx} & \sigma^{2xy} \\ \sigma^{2yx} & \sigma^{2yy} \end{pmatrix} = \begin{pmatrix} h_0^{xx} + h_1^{xx}X & h_0 + h_1 X \\ h_0 + h_1 X & h_0^{yy} + h_1^{yx}X + h_1^{yy}Y \end{pmatrix}.$$

Plugging these specifications into the key equation (3.51), we can confirm that special solutions of interest can be obtained using standard separation of variables techniques. Specifically, we look for solution of the form

$$\phi(X, Y) \sim e^{BY} G(X).$$

This is indeed a solution of (3.51) if the following relations hold for unknowns B and $G(X)$

$$h_1^{yy}B^2 + k_1^{yy}B - b^y = 0,$$

$$\frac{1}{2}\sigma^{2xx}G_{XX} + \left(B\sigma^{2xy} + \mu^{X,Q}\right)G_X + \left(\frac{B^2}{2}[h_0^{yy} + h_1^{yx}X] + B[k_0^y + k_1^{yx}X] + \rho - a - b^x X\right)G = 0.$$

Solving first quadratic equation yields parameter B. With all coefficients being linear in X, the second equation is identical to (3.15) of construction 1. The most general solution of $G(X)$ is in term confluent hypergeometric functions $\Phi(.,.;\beta X)$, as readily given in proposition 12. Accordingly, in practice we may start out with a functional stochastic discount factor of the class

$$M^P(X, Y, t) = \frac{e^{-\rho t}e^{-BY-\alpha X}}{\lambda_1\Phi(\delta, \gamma; \beta X) + \lambda_2(\beta X)^{1-\gamma}\Phi(\delta - \gamma + 1, 2 - \gamma; \beta X)},$$

which will be consistent with the multi-factor dynamic (3.52). Yet different specific choices within this class yield rich sets of possible equilibrium interpretations, market prices of risk and P-dynamics, as we have seen in our previous re-constructions of various term structure models, linearity-generating dynamic, and specially the forward premium puzzle.

3.7.2 Maximum likelihood estimation procedure

Since zero-coupon bond prices are tractable here, we can also use the maximum likelihood for the model estimation. Singleton (2006) offers an extensive resource for empirical estimation of many dynamic pricing models. For the sake of completeness, in this section we are content with only sketching the maximum likelihood estimation steps entailed specifically to our construction. The following procedure is also drawn upon the works of Ait-Sahalia (2002), Cheridito et al. (2007) and others.

- Step 1: Collect data on more bonds than state variables (Pearson and Sun (1994)): Pick price data of K bonds (with K different maturities. Treat first N bonds' prices as exact (i.e., observed without errors), and the rest $K - N$ prices as noisy (i.e., observed with errors). Here N is number of state variables. Assume that observation error vector is i.i.d. Gaussian multivariate with parameters set $\mathcal{T}_{err}$

- Step 2: Pick a model to be estimated, e.g., as in Construction 3. Choose some specific numerical value set for input parameters of this construction. The input parameters are

$$\mathcal{T}_{total} = \{H_0; H_1; \tilde{K}_0; \tilde{K}_1; \text{parameters inside SPD } \xi^P; \mathcal{T}_{err}\}.$$

- Step 3: Because zero-coupon bond prices are tractable (Chen and Joslin (2011)'s method) in the model. from the data (prices) of first N bonds, back out the implied (latent) variable vector $X(t) = \{X^1(t); \ldots; X^N(t)\}$ (corresponding specifically to the above numerical set of parameters). (In our one-factor model, $N = 1$)

- Step 4: Now because our construction allows for explicitly (solved) $P-$measure dynamic μ^P, σ, we can (approximately) construct the transition probability $L_X(X(t)|X(t-1))$ in $P-$measure using Ait-Sahalia (2002)'s approximation (Hermite polynomial expansion).

- Step 5: Since zero coupon bond prices is tractable, there is a tractable relation between latent variable vector $X(t)$ and yield vector $y(t)$: $y = y(X)$. Jacobian of this relation allows us to convert state-variable transition probability $L_X(X(t)|X(t-1))$ into yield transition probability $L_y(y(t)|y(t-1))$

- Step 6: We compute the implied error vector (in the observed data) of the rest $K - N$

bonds:

$$\epsilon \equiv \ (K-N) \text{ yields' error vector } =$$
$$\text{observed } (K-N) \text{ yields } - \text{ theoretical } (K-N) \text{ yields}$$

where theoretical yields are computed by feeding the implied state variables (obtained in Step 3) into the model's tractable zero-coupon bond price

- Step 7: As error vector ϵ is assumed Gaussian multivariate, its likelihood is known by placing ϵ into normal multivariate density $\mathcal{N}(\epsilon, \mathcal{T}_{err})$. Then the total likelihood function is the product

$$L(\mathcal{T}_{total}) = L_y(y(t)|y(t-1)) \times \mathcal{N}(\epsilon, \mathcal{T}_{err}).$$

- Step 8: By maximizing this total likelihood function $L(\mathcal{T}_{total})$ by changing the numerical value of the parameter set $\mathcal{T}_{total}$ in Step 2, we will arrive at the best-fit parameter set $\mathcal{T}^*{}_{total}$.

3.8 Conclusion

This paper starts with a key and simple observation that when stochastic discount factor is proper function of underlying state variables, it can be determined from the risk-neutral state dynamic via a simple linear differential equation. Consequently, state dynamic in physical measure can also be consistently pinned down. Accordingly, we propose a novel, tractable and most general asset pricing model of functional stochastic discount factor (SDF). The construction is motivated by and provides structural foundation for many popular reduced-form pricing models, which currently might lack of economic intuitions.

As an application, we construct a functional stochastic discount factor that sheds light into viable consumption risk underlying the forward premium anomaly. Intuitively, when home market and consumption go up, home risk-free bonds lose their appeals, become cheaper and home interest rate increases. At the same time, risk-averse international investors perceive lower risk in bull home market, and value home currency more favorably. Altogether, these consistently render a rational explanation for forward premium puzzle: home currency relatively appreciates while home interest increases.

As a function of state variables, stochastic discount factor also offers new framework to unify diverse existing asset pricing models. To illustrate, we establish simple conditions under which many classic settings of dynamic term structure modeling (such as affine, quadratic and quotient interest rate models), as well as pricing models based on recently-proposed linearity-generating processes, all can be derived from functional stochastic discount factors.

3.9 Appendices

3.9.1 Table of notations

The following table lists all key quantities and their notations employed in the main text.

3.9.2 Proofs

We recall that subscripts always denote derivatives or partial derivatives (when appropriate): e.g., $f_X \equiv \frac{\partial f}{\partial X}$ throughout the paper. To simplify the notation, we also omit the explicit state and time contingency (X, t) from general function $f(X, t)$ wherever the omission does not create possible ambiguity.

Proof of proposition 12. Our construction of functional stochastic discount factor $M^P(t, X)$ out of given processes governing the state variables is based on the key second-order linear differential equation (SOLDE) (3.11), and thus benefits greatly from established mathematical results. A recent comprehensive resource on differential equations and their special function solutions is the NIST handbook (2010) edited by Olver et al.

In particular, equation (3.15) can be solved analytically by transforming it into the following standard confluent hypergeometric differential equation (CHGDE) (recall that subscripts denote the derivatives)

$$g_{zz} + \left(\frac{\gamma}{z} - 1\right) g_z - \frac{\delta}{z} g = 0, \tag{3.53}$$

whose two fundamental independent solutions are expressed in term of confluent hypergeo-

Notation	Description
P	physical (data-generating) measure
Q	risk-neutral measure
R	any general measure equivalent to P and Q
Z^P, Z^Q, Z^R	standard Brownian motions under respective measure
X	(vector) state variable
$\mu^{X,P}$, $\mu^{X,Q}$, $\mu^{X,R}$	dynamic (drift) of state variable X in measure P, Q, R respectively
σ^X	dynamic (volatility) of state variable X (identical in any equivalent measure)
M^P, M^Q, M^R	stochastic discount factor (SDF) under respective measure
$M^P(X,t)$	SDF as proper function of (X,t)
$M^{P,\{\lambda_1,\lambda_2\}}(X,t)$	SDF as proper function of (X,t), parametrized by λ_1, λ_2
$\hat{M}(v,t)$	Fourier transform of $M(X,t)$ (in variable X)
ρ	subjective discount factor
$\phi^P(X) = \frac{1}{M^P(X)} = \frac{e^{-\rho t}}{M^P(X,t)}$	reciprocal of SDF $M^P(X)$
$\xi^{QP} = \frac{dP^Q}{dP^P}$	Radon-Nikodym derivative to change measure from Q to P
$r(X,t)$	instantaneously risk-free rate (short rate) process
$\eta^{QP} = \lim_{dt \to 0} \frac{dZ^Q - dZ^P}{dt}$	market price of risk associated with measure change from Q to P
$\mathcal{D}^{X,P}$	infinitesimal operator associated with (diffusion) process X in measure P

metric functions $\mathbf{\Phi}(.,.;z)$

$$\mathbf{\Phi}(\delta, \gamma; z); \qquad z^{1-\gamma}\mathbf{\Phi}(\delta - \gamma + 1, 2 - \gamma; z).$$

That is, any solution $g^{\{\lambda_1,\lambda_2\}}(z)$ to (3.53) is a linear combination of the two independent solutions

$$g^{\{\lambda_1,\lambda_2\}}(z) = \lambda_1\mathbf{\Phi}(\delta, \gamma; z) + \lambda_2 z^{1-\gamma}\mathbf{\Phi}(\delta - \gamma + 1, 2 - \gamma; z),$$

where λ_1, λ_2 are constants of integration. Specific steps to bring (3.15) into (3.53) are as follows.

First, after a change of variable

$$y = H_0 + H_1 X \longleftrightarrow X = \frac{y}{H_1} - \frac{H_0}{H_1},$$

equation (3.16) becomes

$$\phi^P_{yy} + \frac{2(K_0^Q H_1 - K_1^Q H_0 + K_1^Q y)}{H_1^2 y}\phi^P_y + \frac{2\left(\rho + \frac{bH_0}{H_1} - a\right) - 2\frac{b}{H_1}y}{H_1^2 y}\phi^P = 0. \tag{3.54}$$

Next, we make the following transformation and another change of variable

$$\phi(y) \equiv e^{\alpha y} g(\beta y); \qquad z \equiv \beta y,$$

where α and β are two constants of choice to be determined below. Differential equation of $g(z)$ then follows from (3.54)

$$\beta^2 g_{zz} + 2\beta\left[\alpha + \frac{K_1^Q}{H_1^2} + \frac{\beta(K_0^Q H_1 - K_1^Q H_0)}{H_1^2 z}\right] g_z$$
$$+ \left[\left\{\alpha^2 + \frac{2\alpha K_1^Q}{H_1^2} - \frac{2b}{H_1^3}\right\} + \frac{2\beta\left(\alpha[K_0^Q H_1 - K_1^Q H_0] + \rho + \frac{bH_0}{H_1} - a\right)}{H_1^2 z}\right] g = 0.$$

To bring this equation into the standard CHGDE (3.53) we choose parameter α such that the expression inside the curly brackets vanishes[48]

$$\alpha^2 + \frac{2\alpha K_1^Q}{H_1^2} - \frac{2b}{H_1^3} = 0 \Rightarrow \alpha = -\frac{K_1^Q}{H_1^2} \pm \left(\frac{(K_1^Q)^2}{H_1^4} + \frac{2b}{H_1^3}\right)^{\frac{1}{2}}. \tag{3.55}$$

Dividing both sides of above DE then yields

$$g_{zz} + \frac{2}{\beta}\left[\frac{2(K_0^Q H_1 - K_1^Q H_0)}{H_1^2 z} + \frac{2\left(\alpha + \frac{K_1^Q}{H_1^2}\right)}{\beta}\right] g_z + \left[\frac{2\left(\alpha[K_0^Q H_1 - K_1^Q H_0] + \rho + \frac{bH_0}{H_1} - a\right)}{\beta H_1^2 z}\right] g = 0.$$

Evidently, this equation is identical to standard CHGDE (3.53) by the following parameter identifications

$$\gamma = \frac{2(K_0^Q H_1 - K_1^Q H_0)}{H_1^2}; \quad \beta = -2\left(\alpha + \frac{K_1^Q}{H_1^2}\right) = \mp 2\left(\frac{(K_1^Q)^2}{H_1^4} + \frac{2b}{H_1^3}\right)^{\frac{1}{2}};$$

$$\delta = \frac{-2\left(\alpha[K_0^Q H_1 - K_1^Q H_0] + \rho + \frac{bH_0}{H_1} - a\right)}{\beta H_1^2} = \frac{\alpha[K_0^Q H_1 - K_1^Q H_0] + \rho + \frac{bH_0}{H_1} - a}{\alpha H_1^2 + K_1^Q},$$

where α is given by (3.55). Undoing previous transformation and changes of variables we obtain the most general solution of (3.15)

$$\varphi^{P,\{\lambda_1,\lambda_2\}}(X) = e^{\frac{\alpha}{\beta}z}\left[\lambda_1 \Phi(\delta, \gamma; z) + \lambda_2 z^{1-\gamma}\Phi(\delta - \gamma + 1, 2 - \gamma; z)\right],$$

with $z = \beta y = \beta(H_0 + H_1 X)$. Finally, using definition (3.10) yields (3.16). ■

Proof of proposition 13. The specification of completely affine DTSM with one factor $X \in R^+$ (Dai and Singleton (2000)) can be written as

$$\mu^{X,P} = K_0^P + K_1^P X; \quad (\sigma^X)^2 = X; \quad r = a + bX; \quad \eta^{QP} = \lambda_{11}\sqrt{X},$$

[48] Which root of α to be chosen should be dictated by economic consideration, such as how stochastic discount factor $M^P(X)$ varies (increases or decreases) with state variable X. See section 3.6 for an illustration.

where λ_{11} is a constant. This specification implies that the dynamic is also affine under Q: $\mu^{X,Q} = \mu^{X,P} - \sigma^X \eta^{QP} = K_0^Q + K_1^Q X$ with

$$K_0^Q = K_0^P; \qquad K_1^Q = K_1^P - \lambda_{11}. \tag{3.56}$$

In this setting, SDF $M^P(X,t)$ satisfies a special version of (3.7), (3.8)

$$\begin{gathered}\tfrac{1}{2} X M_{XX}^P + [K_0^P + K_1^P X] M_X^P + [a + bX - \rho] M^P(X) = 0, \\ M_X^P + \lambda_{11} M^P = 0 \Rightarrow M_{XX}^P = \lambda_{11}^2 M^P.\end{gathered}$$

Plugging second equation into the first, and identifying terms of order X^0 (constants) and X^1 in both sides respectively yield

$$\begin{aligned} a &= \rho + \lambda_{11} K_0^P = \rho + K_0^P (K_1^P - K_1^Q), \\ b &= \lambda_{11} \left(K_1^P - \frac{\lambda_{11}}{2} \right) = \frac{1}{2} \left[(K_1^P)^2 - (K_1^Q)^2 \right]. \end{aligned}$$

Using λ_{11} from (3.56) we obtain first set of identities in proposition 13. (The first identity above can be attributed to a choice of discount factor ρ, and was omitted in the proposition.) The specification of completely affine DTSM with one factor $X \in R^+$ (Cheridito et al. (2007)) can be written as

$$\mu^{X,P} = K_0^P + K_1^P X; \quad (\sigma^X)^2 = X; \quad r = a + bX; \quad \eta^{QP} = \frac{\lambda_{01}}{\sqrt{X}} + \lambda_{11}\sqrt{X}.$$

where λ_{01}, λ_{11} are constants. This specification implies that the dynamic is also affine under Q: $\mu^{X,Q} = \mu^{X,P} - \sigma^X \eta^{QP} = K_0^Q + K_1^Q X$ with

$$K_0^Q = K_0^P - \lambda_{01}; \qquad K_1^Q = K_1^P - \lambda_{11}. \tag{3.57}$$

In this setting, SDF $M^P(X,t)$ satisfies a special version of (3.7), (3.8)

$$\tfrac{1}{2}XM^P_{XX} + [K^P_0 + K^P_1 X]M^P_X + [a + bX - \rho]M^P(X) = 0 \qquad ,$$
$$M^P_X + \left(\tfrac{\lambda_{01}}{X} + \lambda_{11}\right)M^P = 0 \Rightarrow M^P_{XX} = \left(\tfrac{\lambda_{01}+\lambda_{01}^2}{X^2} + \tfrac{2\lambda_{01}\lambda_{11}}{X} + \lambda_{11}^2\right)M^P.$$

Plugging second equation into the first, and identifying terms of order X^{-1}, X^0 and X^1 in both sides respectively yield

$$\begin{aligned} 0 &= \lambda_{01}\left(K^P_0 - \frac{\lambda_{01}+1}{2}\right), \\ a &= \rho + \lambda_{11}K^P_0 + \lambda_{01}K^P_1 - \lambda_{01}\lambda_{11}, \\ b &= \lambda_{11}\left(K^P_1 - \frac{\lambda_{11}}{2}\right). \end{aligned}$$

Finally, using λ_{01}, λ_{11} from (3.57) we obtain the second set of identities in proposition 13. (The second identity above can be attributed to a choice of discount factor ρ, and was omitted in the proposition.)

Generally, the functional SDF in construction 1 does not necessary imply $P-$affine dynamic. Conversely, completely (or extended) affine DTSM does not necessarily imply a proper functional SDF $M^P(X,t)$. Only with these additional parameter restrictions, the functional SDF $M^P(X,t)$ in construction 1 generates a completely (or extended) affine DTSM. ■

Proof of proposition 14. Equation (3.25) can be solved analytically by transforming it into a form of standard Weber differential equation (WDE)

$$g_{zz} - \left(\frac{z^2}{4} + \nu\right)g = 0, \qquad (3.58)$$

whose two fundamental independent solutions are expressed in term of confluent hypergeometric functions $\mathbf{\Phi}(.,.;z)$

$$e^{\frac{-z^2}{4}}\mathbf{\Phi}\left(\frac{\nu}{2}+\frac{1}{4},\frac{1}{2};\frac{z^2}{2}\right); \qquad ze^{\frac{-z^2}{4}}\mathbf{\Phi}\left(\frac{\nu}{2}+\frac{3}{4},\frac{3}{2};\frac{z^2}{2}\right).$$

That is, any solution $g^{\{\lambda_1,\lambda_2\}}(z)$ to (3.58) is a linear combination of the two independent solutions

$$g^{\{\lambda_1,\lambda_2\}}(z) = \lambda_1 e^{\frac{-z^2}{4}} \mathbf{\Phi}\left(\frac{\nu}{2}+\frac{1}{4},\frac{1}{2};\frac{z^2}{2}\right) + \lambda_2 z e^{\frac{-z^2}{4}} \mathbf{\Phi}\left(\frac{\nu}{2}+\frac{3}{4},\frac{3}{2};\frac{z^2}{2}\right),$$

where λ_1, λ_2 are constants of integration. Specific steps to bring (3.25) into (3.58) are as follows.

First, after a transformation

$$\phi^P(X) \equiv e^{MX+NX^2} g(X)$$

where M, N are constants of choice to be determined below, equation (3.25) becomes

$$\frac{1}{2}H_0 g_{XX} + \left[(2H_0N + K_1^Q)X + (H_0M + K_0^Q)\right] g_X - \left[AX^2 + BX + C\right] g = 0. \qquad (3.59)$$

where parameters A, B, C are related to M, N and are deferred till after the latter are determined. Evidently, to bring (3.59) into Weber form (3.58) we choose M, N such that term g_X vanishes

$$H_0 M + K_0^Q = 0 \;\Rightarrow\; M = -\frac{K_0^Q}{H_0};$$
$$2H_0 N + K_1^Q = 0 \;\Rightarrow\; N = -\frac{K_1^Q}{2H_0}.$$

These choices then pin down A, B, C in equation (3.59)

$$A = -\frac{4H_0N^2 + 4K_1^Q N - 2\rho_2}{H_0} = \frac{(K_1^Q)^2 + 2\rho_2 H_0}{H_0^2};$$

$$B = -\frac{4H_0N + 4K_0^Q N + 2K_1^Q M - 2\rho_1}{H_0} = \frac{2(K_1^Q H_0 + 2K_0^Q K_1^Q + \rho_1 H_0)}{H_0^2};$$

$$C = -\frac{H_0M^2 + 2K_0^Q M + 2(\rho - \rho_0)}{H_0} = \frac{(K_0^Q)^2 - 2H_0(\rho - \rho_0)}{H_0^2}.$$

Next, the change of variable[49]

$$z = (4A)^{\frac{1}{4}} \left(X + \frac{B}{2A} \right),$$

transforms (3.59) into

$$g_{zz} - \left(\frac{z^2}{4} + \frac{4AC - B^2}{4A^{\frac{3}{2}}} \right) g = 0.$$

Identifying this with standard Weber equation (3.58)

$$\nu \equiv \frac{4AC - B^2}{4A^{\frac{3}{2}}},$$

immediately yield the most general solution of original equation (3.25)

$$\phi^P(X) = e^{MX+NX^2} \left[\lambda_1 e^{\frac{-z^2}{4}} \mathbf{\Phi} \left(\frac{\nu}{2} + \frac{1}{4}, \frac{1}{2}; \frac{z^2}{2} \right) + \lambda_2 z e^{\frac{-z^2}{4}} \mathbf{\Phi} \left(\frac{\nu}{2} + \frac{3}{4}, \frac{3}{2}; \frac{z^2}{2} \right) \right],$$

where λ_1, λ_2 are constants of integration, and $z = (4A)^{\frac{1}{4}} \left(X + \frac{B}{2A} \right)$ is linear in original state variable X. Finally, using definition (3.10) yields (3.26). ■

Proof of proposition 15. This proof is similar to that of proposition 12 in the sense that equation (3.29) can also be transformed into the standard CHGDE (3.53), though detailed steps are a bit different.

First, after a transformation

$$\phi^P(X) \equiv e^{\alpha X} X^{\beta} g(X),$$

where α, β are constants of choice to be determined below, equation (3.29) becomes

$$g_{XX} + \left[\frac{2}{X} \left(\beta + \frac{K_0^Q}{H_1} \right) + 2 \left(\alpha + \frac{K_1^Q}{H_1} \right) \right] g_X +$$
$$\left[\frac{\beta^2 H_1 + (2K_0^Q - H_1)\beta - 2\theta_{-1}}{H_1 X^2} + \frac{2(H_1 \alpha\beta + K_0^Q \alpha + K_1^Q \beta + \rho - \theta_0)}{H_1 X} + \frac{H_1 \alpha^2 + 2K_1^Q \alpha - 2\theta_1}{H_1} \right] g = 0.$$

[49]Real value for z requires $A > 0$. In case of $A < 0$ we can proceed similarly to bring the original equation (3.25) to another form of Weber differential equation $g_{zz} + \left(\frac{z^2}{4} - \nu \right) g = 0$.

To bring this into (3.53) we choose parameters α, β such that coefficients of order X^0 and X^{-2} of term g (last term in the above differential equation) vanish.

$$H_1\alpha^2 + 2K_1^Q\alpha - 2\theta_1 = 0 \Rightarrow \alpha = \frac{-K_1^Q \pm \left[(K_1^Q)^2 + 2H_1\theta_1\right]^{1/2}}{H_1},$$

$$\beta^2 H_1 + (2K_0^Q - H_1)\beta - 2\theta_{-1} = 0 \Rightarrow \beta = \frac{H_1 - 2K_0^Q \pm \left[(2K_0^Q - H_1)^2 + 8H_1\theta_{-1}\right]^{1/2}}{2H_1}.$$

The differential equation for g becomes

$$g_{XX} + \left[\frac{2}{X}\left(\beta + \frac{K_0^Q}{H_1}\right) + 2\left(\alpha + \frac{K_1^Q}{H_1}\right)\right] g_X + \left[\frac{2(H_1\alpha\beta + K_0^Q\alpha + K_1^Q\beta + \rho - \theta_0)}{H_1 X}\right] g = 0.$$

Finally, the change of variable

$$z = -2\left(\alpha + \frac{K_1^Q}{H_1}\right) X = \mp\frac{2\left[(K_1^Q)^2 + 2H_1\theta_1\right]^{1/2}}{H_1} X,$$

precisely transforms the above equation for g into the standard CHGDE (3.53). Analytical solution for $g(z)$ and then $\phi^P(X)$ follow similarly as in the proof of proposition 12. We thus obtain (3.30). ■

Proof of proposition 16. We will show that when relations specified in (3.49) hold, the covariance $\sigma^{ds,\Delta r} < 0$, or high interest rate currencies tend to appreciate. But first we explain how these relations are formulated in the first place. The relation $H_1^i > 0$ is dictated by the Feller's admissibility condition (3.44) to assure the positivity for the volatility of X^i's square-root dynamic. Similarly, $K_0^{i,P} = \frac{H_1^i(1-C^i)}{2}$ is required to generate linear short rate r^i (3.45). The choice condition $C^i < 0$ is motivated by an economic intuition that $-C^i > 0$ characterizes the risk aversion of the representative investor in country $i \in \{h, f\}$ as the SDF M^i is to be identified with the her marginal utility in the structural model of section 3.6.2. The choice condition $B^i > 0$ together with small absolute value $|B^i|$ robustly assure non-negative interest rate (3.48).

Next, using delta method approximation (3.43) we rewrite the key FPP-consistent condition (3.42) as[50]

$$\sigma^{ds,\Delta r} \approx b^h Var\left(X^h\right)\left[b^h + \frac{1}{2}(B^h)^2 H_1^h - \frac{(C^h)^2}{2}\frac{(K_1^{h,P})^2}{(K_0^{h,P})^2} H_1^h\right] + (h \leftrightarrow f) < 0.$$

Since $b^h, b^f > 0$ (see (3.47), (3.48)), this covariance is negative when the last term inside square brackets dominates the first two terms (we explicitly plug in $K_0^{i,P} = \frac{H_1^i(1-C^i)}{2}$ and $b^i = -B^i\left(K_1^{i,P} + \frac{B^i H_1^i}{2}\right)$ (3.47) in what follows)

$$\frac{(C^i)^2(K_1^{i,P})^2}{H_1^i(1-C^i)^2} \gg -B^i\left(K_1^{i,P} + \frac{B^i H_1^i}{2}\right) \gg (B^i)^2 H_1^i \qquad i \in \{h, f\}.$$

Since $B^i, H_1^i > 0$, the last relation of (3.49) $-K_1^{i,P} \gg B^i H_1^i$ clearly implies the above inequalities, and thus also the negative covariance $\sigma^{ds,\Delta r} < 0$, for any risk aversion coefficient $-C^i$ that is strictly positive. The later part of section 3.6.2 justifies all relations in (3.49) we have just derived here from a structural (risk-based) consideration. ■

References

Ahn, D. H., R. F. Dittmar and A. R. Gallant (2002), "Quadratic Term Structure Models: Theory and Evidence," *The Review of Financial Studies*, 15, 243-288.

Ahn, D. H. and Gao Ahn, D-H. and B. Gao (1999), "A Parametric Nonlinear Model of Term Structure Dynamics," *The Review of Financial Studies*, 12, 721762.

Ait-Sahalia, Y. (1996), "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527-560.

Ait-Sahalia, Y. (2002), "Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223-262.

[50]We also replace $E^P[X^i]$ by its long-run mean $\frac{-K_0^{i,P}}{K_1^{i,P}}$ in the approximation.

Backus, D. K., S. Foresi and C. I. Telmer (2001), "Affine Term Structure Models and the Forward Premium Anomaly," *Journal of Finance*, 56, 279304.

Bakshi, G., P. Carr, and L. Wu (2008), "Stochastic Risk Premiums, Stochastic Skewness in Currency Options, and Stochastic Discount Factors in International Economies," *Journal of Financial Economics*, 87, 132156.

Breeden, D. (1979), "An intertemporal asset pricing model with stochastic consumption and investment opportunities," *Journal of Financial Economics*, 7, p. 265296.

Breeden, D. and R. Litzenberger (1978), "Prices of State-Contingent Claims Implicit in Option Prices", *Journal of Business*, 51, 621-651.

Carr, P. and L. Wu (2007), "Stochastic Skew for Currency Options," *Journal of Financial Economics*, 86, 213-247.

Chen, H. and S. Joslin (2011), "Generalized Transform Analysis of Affine Processes and Applications in Finance," to appear in *Review of Finance Studies*.

Cheng, P. and O. Scaillet (2007), "Linear-quadratic jump-diffusion modeling," *Mathematical Finance*, 17, 575598.

Cheridito, P., D. Filipovic and R.L. Kimmel (2007), "Market Price of Risk Specifications for Affine Models," *Journal of Financial Economics*, 83, 123-170.

Cochrane, J.H. (2005), "Asset Pricing," *Princeton University Press*, Princeton, N.J.

Constantinides, G. M. (1992), "A theory of the nominal term structure of interest rates," *Review of Financial Studies*, 5, 531552.

Cox, J. C. and S. A. Ross (1976), "A Survey of Some New Results in Financial Option Pricing Theory," *Journal of Finance*, 31, 383-402.

Cox, J. C., J. E. Ingersoll and S. A. Ross (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385407.

Cuchiero, C., D. Filipovic and J. Teichmann (2009), "Affine models," in R. Cont (Ed.), *Encyclopedia of Quantitative Finance*.

Dai, Q. and K. J. Singleton (2000), "Specification analysis of affine term structure models," *Journal of Finance*, 55, 1943-1978.

Duarte, J. (2004), "Evaluating an Alternative Risk Preference in Affine Term Structure Models," *Review of Financial Studies*, 17, 379-404.

Duffee, G. R. (2002), "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57, 405443.

Duffie, D. and R. Kan (1996), "A Yield Factor Model of Interest Rates," *Mathematical Finance*, 6, 379406.

Duffie, D., J. Pan and K. Singleton (2000), "Transform analysis and asset pricing for affine jump-diffusion," *Econometrica*, 68, 1343-1376.

Fama, E. (1984), "Forward and spot exchange rates," *Journal of Monetary Economics*, 14, 319-338.

Gabaix, X. (2008), "Variable Rare Disasters: A Tractable Theory of Ten Puzzles in Macro-Finance," *American Economic Review*, 98, 64-67.

Gabaix, X. (2009), "Linearity-Generating Process: A Modeling Tool Yielding Closed Forms Asset Prices," New York University, working paper.

Graveline, J. (2006), "Exchange Rate Volatility and the Forward Premium Anomaly," working paper, Carlson School of Management, University of Minnesota.

Harrison, J. M. and D. M. Kreps (1981), "Martingale and Arbitrage in Multiperiods Securities Markets," *Journal of Economic Theory*, 20, 381-408.

Hansen, L. P. and K. J. Singleton (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269-86.

Le, A., K. J. Singleton and Q. Dai (2010), "Discrete-Time Dynamic Term Structure Models with Generalized Market Prices of Risk," *Review of Financial Studies*, 23, 2184-2227.

Leippold, M. and L. Wu (2002), "Asset Pricing Under the Quadratic Class," *Journal of Financial and Quantitative Analysis*, 37, 271-295.

Lucas, R. (1978), "Assets prices in an exchange economy," *Econometrica*, 46, p. 14291445.

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

Olver, F., D. Lozier, R. Boisvert and C. Clark, Editors (2010), "NIST Handbook of Mathematical Functions," *Cambridge University Press.*

Pearson, N. D. and T-S. Sun (1994), "Exploiting the Conditional Density in estimating the Term Structure: An application to the Cox, Ingersoll and Ross model," *Journal of Finance*. 49, 1279-1304.

Rogers, L. C. G. (1997), "The Potential Approach to the Term Structure of Interest Rates and Foreign Exchange Rates," *Mathematical Finance*, 7, 157-176.

Rogers, L. and D. Williams (1987), "Diffusions, Markov Processes, and Martingales." vol. 2, *Wiley, New York.*

Ross, S. A. (1976), "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13. 341-360.

Ross, S. A. (2011), "The Recovery Theorem," NBER Working Paper No. 17323.

Rubinstein, M. (1976), "The valuation of uncertain income streams and the pricing of options," *Bell Journal of Economics and Management Science*, 7, p. 407425.

Singleton, K. J. (2006), "Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment," *Princeton University Press*, Princeton, NJ.

Stanton, R. (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk", *Journal of Finance*, 52, 1973-2002.

Vasicek, O. (1977), "An Equilibrium Characterisation of the Term Structure", *Journal of Financial Economics*, 5, 177188.

Yan, Hongjun (2008), "Natural Selection in Financial Markets: Does It Work?," *Management Science*, 54, 1935-1950.