CHAPTER 3

Non-Risk-based Explanations of the Equity Premium

Rajnish Mehra
University of California, Santa Barbara, and NBER

and

Edward C. Prescott
Arizona State University, and Federal Reserve Bank of Minneapolis

JEL Classification: G10, G12, G28, N42, D9

Keywords: asset pricing, equity risk premium, liquidity, risk free rate, transaction balances, government regulations, taxes, agent heterogeneity, intermediation costs

We thank John Donaldson and Ellen McGrattan for invaluable comments.

INTRODUCTION

A widely accepted tenet of the neoclassical paradigm has been that the observed differences in the rates of return on financial assets, in particular, the large difference between the average returns on corporate equity and T-bills, are a *premium for bearing non-diversifiable aggregate risk*. The findings in our paper, "The Equity Premium: A Puzzle" that only a small part of this difference is a premium for bearing aggregate risk came as a surprise to the profession. Over the last 20 years or so, attempts to successfully account for the equity premium have become a major research impetus in finance and economics. A majority of these explanations have focused on mechanisms to increase the premium for bearing non-diversifiable risk.[1]

In this chapter, in contrast, we review the nascent literature that takes as given the findings in Mehra and Prescott (1985) and tries to account for the observed magnitude of the equity premium by using *factors other than aggregate risk.* Much of this literature re-examines the appropriateness of the abstractions and assumptions made in our original paper. In particular, the appropriateness of using T-bills as a proxy for the intertemporal marginal rate of substitution of consumption, the impact of government regulations, the abstractions from taxes, diversification and intermediation costs, borrowing constraints and agent heterogeneity have been explored. We consider each in turn and examine their impact on the equity premium.

1. THE INAPPROPRIATENESS OF USING T-BILLS AS A PROXY FOR THE INTERTEMPORAL MARGINAL RATE OF SUBSTITUTION OF CONSUMPTION

In the two decades since "The Equity Premium: A Puzzle" was published, the way we think about economic phenomena has undergone a fundamental change, influenced in large measure by the development of Real Business Cycle theory by Kydland and Prescott (1982). One finding of the real business cycle literature is that the real after-tax return on capital, that is the after tax capital income divided by the cost of reproducing the tangible capital stock, averages 4.5 percent, with modest variation, over time. A key question that arises is, what is a good *empirical counterpart* to the household's real interest rate? Mehra and Prescott (1985) used the highly liquid T-bill rate, corrected for expected inflation, as a proxy for the return on the riskless asset that is used by agents to smooth consumption. The appropriateness of this assumption was questioned by McGrattan and Prescott (2003) and more recently by Mehra and Prescott (2007). The remainder of this section closely follows the discussion in these papers.

An assumption implicit in Mehra and Prescott (1985) is that agents use *both* equity and the riskless asset to *intertemporally* smooth *consumption*. This is as a direct consequence of the first-order condition (see ahead) for the representative household in their

[1]Many of the chapters in this volume elaborate on these efforts. The previous chapter (Donaldson and Mehra (2008) surveys this literature.

TABLE 1
Household Assets and Liabilities (Average of 2000 and 2005)

Assets (GDP)		Liabilities (GDP)	
Tangible household	1.65	Liabilities	0.7
Corporate equity	0.85	Net worth	4.15
Non corporate equity	0.5		
Pension and life insurance reserves	1.0		
Debt assets	0.85		
Total	**4.85**	**Total**	**4.85**

model, which saves by optimally allocating resources between equity and risk less debt. Equation (1) is the standard asset pricing equation in macroeconomics and finance.

$$0 = E_t \left[\frac{U_c(c_{t+s})}{U_c(c_t)} (r^e_{t,t+s} - r^d_{t,t+s}) \right], \tag{1}$$

where $U_c(c_{t+s})$ is the marginal utility of consumption at time $t + s$, $r^e_{t,t+s}$ and $r^d_{t,t+s}$ are, respectively, the return on equity and the risk less asset over the period t, $t + s$ and E_t is the expectation conditional on the agent's information set at time t.

Is it reasonable to assume that the return on T-bills is equal to the real interest rate that households use to save for retirement and smooth consumption? Do households actually hold T-bills to finance their retirement? **If this were empirically true it would be reasonable to equate their expected marginal rate of substitution of consumption to the rate of return on T-bills.**

This question *cannot be answered in the abstract*, without reference to the asset holdings of households. A natural next step then is to examine the assets held by households. Table 1 details these holdings for American households. The four big asset-holding categories of households are tangible assets, pension and life insurance holdings, equity (both corporate and non corporate), and debt assets.

In the year 2000, privately held government debt was only 0.30 GDP, a third of which was held by foreigners. The amount of interest bearing government debt with maturity less than a year was only 0.085 GDP, which is a small fraction of the total household net worth. Virtually no T-bills are directly owned by households.[2] Approximately one-third of the outstanding T-bills are held by foreigners, in particular foreign central banks, and two-thirds by American financial institutions.

Although there are large amounts of debt assets held, most of these are in the form of pension fund and life insurance reserves. Some are in the form of demand deposits for which free services are provided. Most of the government debt is held indirectly; however, some is held in the form of savings bonds that people gift to their grandchildren.

[2]Table B-89, Economic Report of the President 2005.

Thus, much of intertemporal saving is in debt assets such as annuities and mortgage debt, held in retirement accounts and as pension fund reserves. *Assets other than T-bills, are held to finance consumption when old.*

Two natural questions then are who holds T-bills and do the holders use them to intertemporally smooth consumption? A large fraction of the stock of short-term U.S. T-bills are held by foreign central banks and are used to smooth exchange rates. These central banks are *not* using these assets to equate the marginal utility of consumption today to the expected discounted marginal utility of consumption tomorrow.

U.S. households do hold a sizable quantity of close substitutes for T-bills. These including money market accounts, and bank CDs with maturities of one year, or less. They are of the order of magnitude of 0.50 times GDP. However, these short-term liquid assets with low expected returns are held, in most part, for reasons *other than saving for retirement*. These include liquidity and the maintence of transaction balances.

1.1. Liquidity

Part of these assets are held for liquidity purposes and as precautionary balances which serve as a substitute for insurance against idiosyncratic risk. The latter is very costly to insure against because of the associated moral hazard and administrative costs.

When this is the case, in the household's maximization problem there are Lagrange multipliers, which are not zero, and as a result the expected returns are less than that predicted by a theory that abstracts from the costs of insuring against idiosyncratic risk.

One of the first attempts to quantify this liquidity premium was the work of Bansal and Coleman (1996). In their model, some assets other than money play a key feature by facilitating transactions. This affects the rate of return they offer in equilibrium. Considering the role of a variety of assets in facilitating transactions, they argue that, on the margin, the transaction service return of money relative to interest-bearing checking accounts should be the interest rate paid on these accounts. They estimate this to be 6 percent, based on the rate offered on NOW accounts for the period they analyze. Since this is a substantial number, they suggest that other money-like assets may also implicitly include a transaction service component to their return. Insofar as T-bills and equity have a different service component built into their returns, this may offer an explanation for the observed equity premium.

1.2. Transaction Balances

Large transaction costs are associated with moving into and out of high-yielding assets, with a major component being record keeping. With assets whose prices vary, there are capital gains or losses associated with *every sale* of these assets which must be reported on the annual tax form. This is why money market accounts hold debt securities until they mature. Thus, there are no capital gains or losses associated with the transaction when an individual buys or sells shares in his money market account. All that the money market mutual fund need report to its shareholders is an annual statement of the interest they received on the 1099 form. This is a major reason that sizable quantities of currency and non-interest-bearing demand deposits are held. We note that Ml is about 15 percent of GDP.

Brokerage charges were (and perhaps continue to be for some) another significant cost associated with getting into and out of higher-yielding securities such as stocks and bonds. It is true that discount brokers now exist, but gaining the prerequisite knowledge is costly in terms of time. If the holding period is short, even for a risk-neutral investor it makes no sense for the investor to buy and later sell stocks unless the expected return on stocks is huge.

Based on the above insights, both McGrattan and Prescott (2003) and Mehra and Prescott (2007) conclude that returns on T-bills and short-term debt are not a reasonable empirical counterpart of the household interest rate in Eq. (1). Hence, it would be inappropriate to equate the return on these assets to the marginal rate of substitution for an important group of agents.

McGrattan and Prescott (2003), who are concerned with the difference in the returns on debt and equity held to finance retirement, use long term high grade municipal bonds as their measure of the long term household interest rate during the gold standard period of 1880–1934. They note that as these bonds were tax exempt and were held directly by individuals, no further adjustment for taxes and intermediation costs is required. They do not adjust for inflationary expectations, arguing that prices were expected to be stable during this period. For the post World War II and Korean War period (1960–2002), they use "the highest grade corporate bonds which were held primarily in tax-deferred pension funds." To account for inflationary expectations they subtract the inflation rate for the previous 10 years. Since government regulations during the interwar years affected bond returns (see ahead), they used NIPA capital returns for these years. Their results are plotted in Figure 1. The returns for the entire period are seen to be remarkably close to 4 percent, considerably higher than the 0.8 percent return for T-bills reported by Mehra and Prescott (1985).

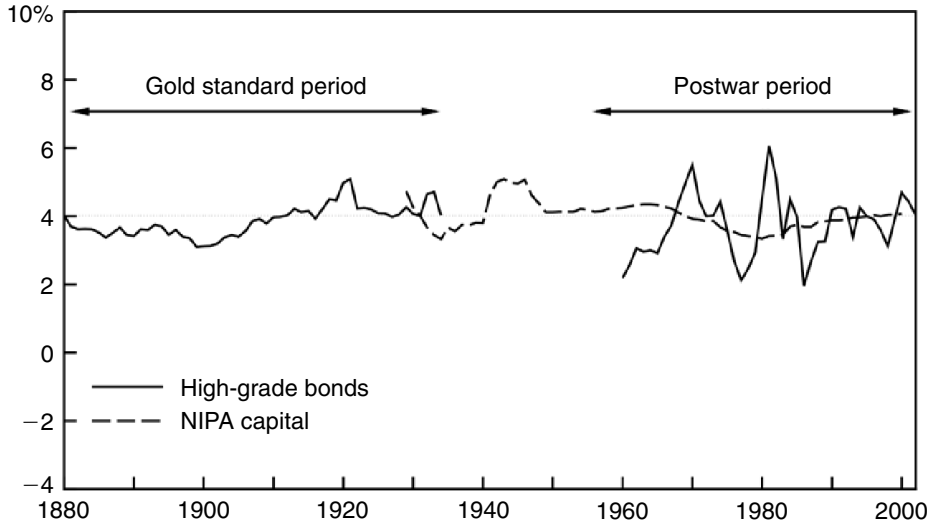

FIGURE 1
Source: McGrattan and Prescott (2003).

Mehra and Prescott (2007) argue that an inflation indexed default-free bond portfolio would essentially be a risk free asset that could be used for savings to finance consumption when old. Since TIPS have only recently (1997) been introduced in the U.S. capital markets, it is difficult to get accurate estimates of the mean return on this asset class. Due to the small quantity supplied, these average returns are not necessarily the same as they would be, had the quantity been larger. The average real return for the 1997–2005 period was 3.7 percent. An alternative would be to use assets such as indexed mortgages guaranteed by Ginnie Mae or issued by Fannie Mae.

Recent work by Krishnamurthy and Vissing-Jorgensen (2007) provides evidence for a convenience yield on *all* treasury securities (not limited to T-bills). They document that the debt/GDP ratio in the U.S. is negatively correlated with the spread between corporate bond yields and Treasury bond yields and claim that the result holds even when controlling for the default risk of corporate bonds. They argue that the corporate bond spread reflects a convenience yield that investors attribute to Treasury debt. This yield, which varies with the debt/GDP ratio, is in the 0.2–1.6 percent range.

This finding, if it holds up, provides additional evidence that returns on even long-dated Treasury securities probably underestimate the marginal rate of intertemporal substitution of consumption and further strengthen the arguments in McGrattan and Prescott (2003).

2. THE EFFECT OF GOVERNMENT REGULATIONS AND RULES

McGrattan and Prescott (2003) argue that the estimated expected return on debt assets in the U.S. (including T-bills) over the period 1926–2000 is biased downwards because of various regulations (in particular, W and X) that helped the Treasury keep nominal rates below 2.5 percent during the 1941–54 period. Table 2 shows that the return on debt securities during that period was considerably lower than their long-term average value. This serves as a reminder that governments can pursue regulatory policies that result in negative interest rates over an extended period of time. Clearly, these rates have little to do with the agents' marginal rate of substitution that would be inferred were there no regulations. Such regulatory periods should be excluded in estimating the long-term average rates on debt securities.

TABLE 2
U.S. Inflation Adjusted Average Return on Debt

Period	1926–2004	1926–40 and 1955–2004	1941–1954
U.S. Treasury bills	0.74%	1.68%	−3.59%
Intermediate-term government bonds	2.48%	3.60%	−2.71%
Long-term government bonds	2.84%	3.86%	−1.88%
Long-term corporate bonds	3.25%	4.36%	−1.89%

Source: Ibbotson 2005.

The third column in Table 2 shows how the conventionally used numbers (in column 2) change when the 1941–54 period is excluded. The estimated average rates increase by about 1 percent for all asset classes.

In the case of T-bills, a further adjustment needs to be made to the returns in the 1930s. During that period, in some states, T-bills were exempt from personal property taxes while cash in the form of bank notes was not. This created an additional demand for the T-bills, suggesting they were sold at a premium. Again, these rates on return have little to do with the marginal rate of substitution of consumption over time. The effect of these adjustments is to further reduce the magnitude of the equity premium relative to T-bills.

To summarize: using the return on a relative risk-free asset that is used for life cycle saving as a proxy for the intertemporal marginal rate of substitution of consumption (instead of a T-bill return) can significantly reduce the equity premium. Adjusting debt returns for government regulations further reduces the premium by 1 percent *irrespective* of the debt asset used as a benchmark.

3. TAXES

McGrattan and Prescott (2003) point out that although taxes on equity income were low until the mid-1930s, they rose dramatically during the war years. They argue that equity returns should be adjusted to reflect this, especially because of regulations that discouraged insurance companies and pension funds from holding equity in tax deferred accounts. They use IRS data to compute the marginal tax rates on equity income distributions. Figure 2 is a plot of the marginal tax rates on equity income distributions. These rates peaked at about 50 percent.

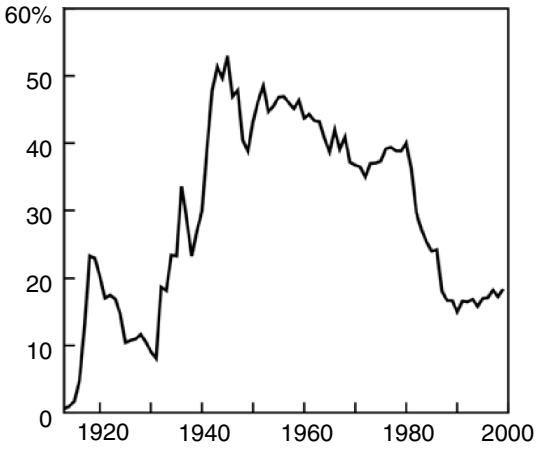

FIGURE 2
Source: McGrattan and Prescott (2003).

Consider the case where the tax on corporate distributions is increased from 0 percent to 50 percent and the additional revenue collected is returned to households in a lump sum. If all, or almost all investments are financed by retained earnings, the value of the stock would be twice as high when the distribution tax rate is zero as when this tax rate is 50 percent. A consequence of this is that the return on equity computed in the standard way will be twice as high when the tax on distribution is 50 percent. This is important because there have been large changes in the tax rate on corporate distributions to owners. Figure 2 plots the distribution tax rate for the United States.

McGrattan and Prescott (2003) use this to compute the after tax equity returns to households after making a number of other adjustments to account for inflation, diversification costs and the fact that some equity was not taxed or was tax deferred. Figure 3 is a plot of their estimate of the after tax return on equity.

An interesting question that arises from their analysis is, why was the return on corporate equity so high in the 1960–2000 period? McGrattan and Prescott (2000, 2005) answer this in the process of estimating the fundamental value of the stock market in 1962 and 2000. They chose these two points in time because after-tax corporate earnings, net corporate debt, and corporate tangible capital stock were approximately the same relative to GDP and the tax system had been stable for a number of years. Further, at neither point in time was there any fear of full or partial expropriation of capital.

What differed was that the value of the stock market relative to GDP in 2000 was nearly twice as large as in 1962.

What changed between 1962 and 2000 were the tax and legal-regulatory systems. The marginal tax rate on corporate distributions was 43 percent in the 1955–1962 period

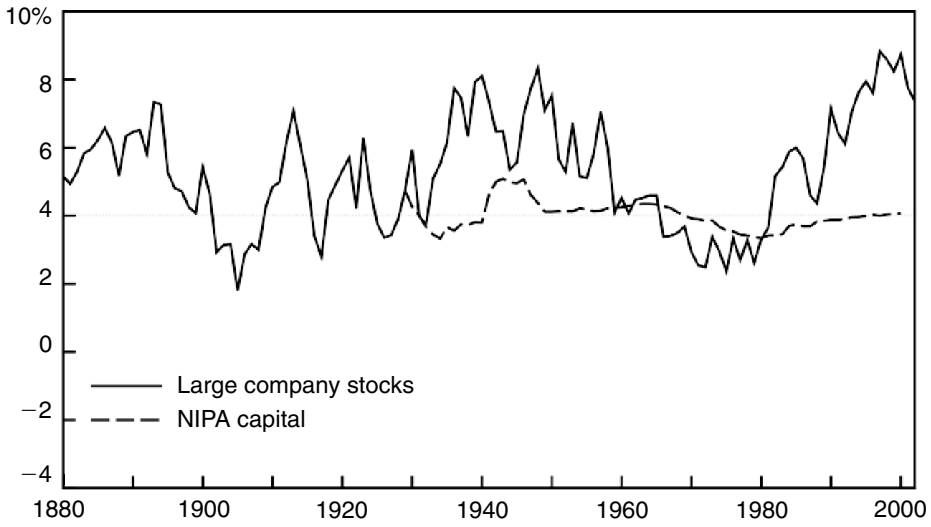

FIGURE 3
Source: McGrattan and Prescott (2003).

and only 17 percent in the 1987–2000. This marginal tax rate on dividends does not have consequences for steady-state after-tax earnings or steady-state corporate capital, if tax revenues are returned lump sum to households. This tax rate *does,* however, have consequences for the value of corporate equity.

The important changes in the legal-regulatory system, most of which occurred in the late 1970s and early 1980s, were that corporate equity was permitted to be held as pension fund reserves and that individuals could invest on a before-tax basis in individual retirement accounts that include equity. The threat of a lawsuit is why debt assets and not equity with higher returns were held as pension fund reserves prior to the late 1970s. At that time, little equity was held in defined contribution retirement accounts because the total assets in these accounts were then a small number. Thus, debt and not equity could, and was, held tax free in 1962. In 2000, both could be held tax-free in defined benefit and defined contribution pension funds and in individual retirement accounts. Not surprisingly, the assets held in tax deferred retirement accounts were large in 2000, being approximately 1.3 GDP (McGrattan and Prescott (2000)).

In determining whether the stock market was over or undervalued vis-à-vis standard growth theory, McGrattan and Prescott (2000, 2005) exploit the fact that the value of a set of real assets is the sum of the values of the individual assets in the set. They develop a method for estimating the value of intangible corporate capital, something that is not reported on balance sheets and, like tangible capital, adds to the value of corporations. Their method uses only national account data and the equilibrium condition that after-tax returns are equated across assets. They also incorporate the most important features of the U.S. and U.K. tax systems into the model, in particular, the fact that capital gains are only taxed upon realization.

The formula they develop for the fundamental value of corporate equities V is

$$V = (1 - \tau_d)K'_T + (1 - \tau_d)\,(1 - \tau_c)K'_I \tag{2}$$

where τ_d is the tax rate on distributions, τ_c is the tax rate on corporate income, K'_T is the end-of-period tangible corporate capital stock at reproduction cost, and K'_I is the end-of-period intangible corporate capital stock at reproduction cost.

The reasons for the tax factors are as follows. Corporate earnings significantly exceed corporate investment, and, as a result, aggregate corporate distributions are large and positive. Historically, these distributions have been in the form of dividends though buybacks have come to be an increasingly important form of distribution in recent years. Therefore, the cost of a unit of tangible capital on margin is only $1 - \tau_d$ units of forgone consumption. In the case of intangible capital, the consumption cost of a unit of capital is even smaller because investments in intangible capital reduce corporate tax liabilities.[3]

The tricky part of the calculation is in constructing a measure of intangible capital. These investments reduce current accounting profits and the returns on the existing stock

[3]In fact, formula (1) must be adjusted if economic depreciation and accounting depreciation are not equal and if there is an investment tax credit. See McGrattan and Prescott (2005).

of intangible capital but increase future accounting profits. The formula for steady-state before tax accounting profits is

$$\pi = \frac{i}{1-\tau_c} K_T + iK_I - gK_I, \tag{3}$$

where g is the steady-state growth rate of the economy and i the steady-state after-tax real interest rate. Note that gK_I is the steady-state net investment in intangible capital, which reduces accounting profits because it is expensed. Note also that all the variables in formula (3) are reported in the system of national accounts with the exception of i and K_I.

McGrattan and Prescott (2005) estimate i using national income data. Their estimate of i is the after-tax real return on capital in the non-corporate sector, which has as much capital as does the corporate sector. They find that the stock market was neither overvalued nor undervalued in 1962 and 2000. The primary reason for the low valuation in 1962 relative to GDP and high valuation in 2000 relative to GDP is that τ_d was much higher in 1962 than it was in 2000. A second reason is that the value of foreign subsidiaries of U.S. corporations grew in the period. An increase in the size of the corporate intangible capital stock was also a contributing factor.

McGrattan and Prescott (2005) find that in the economically and politically stable 1960–2000 period, the after-tax real return on holding corporate equity was as predicted by theory if the changes in the tax and regulatory system were *not* anticipated. These unanticipated changes led to a large unanticipated capital gain on holding corporate equity. Evidence of the importance of these changes is that the share of corporate equity held in retirement accounts and as pension fund reserves increased, from essentially zero in 1962 to slightly over 50 percent in 2000. This share is significantly higher for publicly traded securities as Schedule S corporations and small private Schedule C corporations are not held in retirement accounts. This is important because it means that half of corporate dividends are now subject to zero taxation.

In periods of economic uncertainty, such as those that prevailed in the 1930–1955 period with the Great Depression, World War II, and the fear of another great depression, the survival of the capitalistic system was in doubt. In such times, low equity prices and high real returns on holding equity are not surprising. This is the Brown, Goetzmann, and Ross (1995) explanation of the equity premium. By 1960, the fears of another great depression and of an abandonment of the capitalistic system in the United States had vanished, and clearly other factors gave rise to the high return on equity in the 1960–2000 period.

4. BORROWING CONSTRAINTS[4]

In models with borrowing constraints and transaction costs, the effect is to force investors to hold an inventory of bonds (precautionary demand) to smooth consumption.

[4]The reader is also referred to the section on Market Incompleteness and Trading Frictions in Chapter 2 in this volume.

Hence, in infinite horizon models without borrowing constraints, agents come close to equalizing their marginal rates of substitution with little effect on the equity premium.[5] Some recent attempts to resolve the puzzle incorporating both borrowing constraints *and* consumer heterogeneity appear promising. One approach, which departs from the representative agent model, has been proposed in Constantinides, Donaldson and Mehra (2002).

In order to systematically illustrate these ideas, the authors construct an overlapping-generations (OLG) exchange economy in which consumers live for three periods. In the first period, a period of human capital acquisition, the consumer receives a relatively low endowment income. In the second period, the consumer is employed and receives wage income subject to large uncertainty. In the third period, the consumer retires and consumes the assets accumulated in the second period.

The authors explore the implications of a borrowing constraint by deriving and contrasting the stationary equilibria in two versions of the economy. In the *borrowing-constrained* version, the young are prohibited from borrowing and from selling equity short. The *borrowing-unconstrained* economy differs from the borrowing-constrained one only in that the borrowing constraint and the short-sale constraint are absent.

An *unconstrained* representative agent's maximization problem is formulated as follows. An agent born in period t solves:

$$\max_{\{z^e_{t,i}, z^b_{t,i}\}} E\left(\sum_{i=0}^{2} \beta^i U(C_{t,i})\right) \tag{4}$$

subject to

$$c_{t,0} + q^e_t z^e_{t,0} + q^b_t z^b_{t,0} \leq w^0 \tag{5}$$

$$c_{t,1} + q^e_{t+1} z^e_{t,1} + q^b_{t+1} z^b_{t,1} \leq \left(q^e_{t+1} + d_{t+1}\right) z^e_{t,0} + \left(q^b_{t+1} + b\right) z^b_{t,0} + w^1_{t+1}$$

$$c_{t,2} \leq \left(q^e_{t+2} + d_{t+2}\right) z^e_{t,1} + \left(q^b_{t+2} + b\right) z^b_{t,1} \tag{6}$$

$c_{t,j}$ is the consumption in period $t + j (j = 0, 1, 2)$ of a consumer born in period t. There are two types of securities in the model, *bonds,* and *equity* with *ex-coupon* and *ex-dividend* prices q^b_t and q^e_t respectively. Bonds are a claim to a coupon payment b every period, while the equity is a claim to the dividend stream $\{d_t\}$. The consumer born in period t receives deterministic wage income $w^0 > 0$ in period t, when young; stochastic wage income $w^1_{t+1} > 0$ in period $t + 1$, when middle-aged; and zero wage income in period $t + 2$, when old. The consumer purchases $z^e_{t,0}$ shares of stock and $z^b_{t,0}$ bonds when young. The consumer adjusts these holdings to $z^e_{t,1}$ and $z^b_{t,1}$, respectively, when middle-aged. The consumer liquidates his/her entire portfolio when old. Thus $z^e_{t,2} = 0$ and $z^b_{t,2} = 0$.

[5]This is true unless the supply of bonds is unrealistically low. See Aiyagari and Gertler (1991).

When considering the borrowing constrained equilibrium the following additional constraints are imposed: $z^e_{t,0} = 0$ and $z^b_{t,0} = 0$.

The model introduces two constraints on consumer trades. First, consumers of one generation remove cannot trade claims against their future wage income with consumers of another generation.[6] Second, consumers of one generation cannot trade bonds and equity with consumers of an unborn generation. They build on the observation that absent a complete set of contingent claims, consumer heterogeneity in the form of *uninsurable*, *persistent* and *heteroscedastic* idiosyncratic income shocks, with *counter-cyclical* conditional variance, can potentially resolve empirical difficulties encountered by representative-consumer models.[7]

The novelty of their paper lies in incorporating a life-cycle feature to study asset pricing. The idea is appealingly simple. As discussed earlier, the attractiveness of equity as an asset depends on the correlation between consumption and equity income. If equity pays off in states of high marginal utility of consumption, it will command a higher price, (and consequently a lower rate of return), than if its payoff is in states where marginal utility is low. Since the marginal utility of consumption varies inversely with consumption, equity will command a high rate of return if it pays off in states when consumption is high, and vice versa.[8]

A key insight of their paper is that as the correlation of equity income with consumption *changes* over the life cycle of an individual, so does the attractiveness of equity as an asset. Consumption can be decomposed into the sum of wages and equity income. A young person looking forward at his life has uncertain future wage *and* equity income; furthermore, the correlation of equity income with consumption will not be particularly high, as long as stock and wage income are not highly correlated. This is empirically the case, as documented by Davis and Willen (2000). Equity will thus be a hedge against fluctuations in wages and a "desirable" asset to hold as far as the young are concerned.

The same asset (equity) has a very different characteristic for the middle-aged. Their wage uncertainty has largely been resolved. Their future retirement wage income is either zero or deterministic and the innovations (fluctuations) in their consumption occur from fluctuations in equity income. At this stage of the life cycle, equity income is highly correlated with consumption. Consumption is high when equity income is high, and equity is no longer a hedge against fluctuations in consumption; hence, for this group, it requires a higher rate of return.

The characteristics of equity as an asset therefore change, depending on who the predominant holder of the equity is. Life cycle considerations thus become crucial for asset pricing. If equity is a "desirable" asset for the marginal investor in the economy, then the observed equity premium will be low, relative to an economy where the marginal

[6]Being homogeneous within their generation, consumers have no incentive to trade claims with consumers of their own generation.

[7]See Mankiw (1986) and Constantinides and Duffie (1996).

[8]This is precisely the reason why high-beta stocks in the simple CAPM framework have a high rate of return. In that model, the return on the market is a proxy for consumption. High-beta stocks pay off when the market return is high, i.e. when marginal utility is low, hence their price is (relatively) low and their rate of return high.

investor finds it unattractive to hold equity. The *deus ex machina* is the *stage* in the life cycle of the marginal investor.

The authors argue that the young, who should be holding equity in an economy with markets for all event-contingent commodities, are effectively shut out of this market because of borrowing constraints. The young are characterized by low wages; ideally they would like to smooth lifetime consumption by borrowing against future wage income (consuming a part of the loan and investing the rest in higher return equity). However, they are prevented from doing so because human capital alone does not collateralize major loans in modern economies for reasons of moral hazard and adverse selection.

In the presence of borrowing constraints, equity is thus exclusively priced by the middle-aged investors, since the young are effectively excluded from the equity markets and we observe a high equity premium. If the borrowing constraint is relaxed, the young will borrow to purchase equity, thereby raising the bond yield. The increase in the bond yield induces the middle-aged to shift their portfolio holdings from equity to bonds. The increase in demand for equity by the young and the decrease in the demand for equity by the middle-aged work in opposite directions. On balance, the effect is to increase both the equity and the bond return while simultaneously shrinking the equity premium. Furthermore, the relaxation of the borrowing constraint reduces the net demand for bonds and the risk-free rate puzzle re-emerges.

5. THE IMPACT OF AGENT HETEROGENEITY AND INTERMEDIATION COSTS

A limitation of the *homogenous household* construct is that it precludes the modelling of borrowing and lending among agents. In equilibrium, the shadow price of consumption at date $t + 1$ in terms of consumption at date t is such that the amount of borrowing and lending is zero. Homogenous household models are thus incapable of matching the *quantities* of assets held and *intermediated*.

To address this issue, Mehra and Prescott (2007) construct a model economy that incorporates agent heterogeneity in the form of differences in the strength of the bequest motive. In light of their earlier finding (1985) that the premium for bearing non-diversifiable aggregate risk is small, their analysis abstracts from *aggregate risk*. The only uncertainty that agents face is idiosyncratic risk about the duration of their lifetime after retirement. Agents have identical preferences for consumption; however, they differ with respect to their intensity for bequests. In equilibrium, those with a strong bequest motive accumulate equity assets and, when retired, live off the income of these assets. The households with no bequest motive buy annuities during their working years and consume the annuity benefits over their retirement years. We emphasize that annuities are mostly defined benefit pension plans.

The incorporation of agent heterogeneity allows them to capture a key empirical fact—that there is a large amount of borrowing and lending between households, in

particular, between older households. This borrowing is done either directly, by issuing mortgages to finance owner occupied housing or indirectly, by owning partially debt financed rental properties through direct or limited partnerships or REITS. Mehra and Prescott (2007) abstract from that small amount of borrowing and lending directly between households and assume that all of it is intermediated through financial institutions such as banks and pension funds. For the United States, in 2005 the amount intermediated was approximately 1.6 times the GDP.

Their intermediation technology is constant returns to scale with intermediation costs being proportional to the amount intermediated. To calibrate the constant of proportionality, they use Flow of Funds statistics and data from National Income and Product Accounts. The calibrated value of this parameter equals the net interest income of financial intermediaries, divided by the quantity of intermediated debt and is a little over 2 percent.

In the absence of aggregate uncertainty, the return on equity and the borrowing rate are identical, since the agents who borrow are also marginal in equity markets. In their framework, government debt is not intermediated and thus its return is equal to the lending rate. The equity premium relative to government debt is the intermediation spread. The divergence between borrowing and lending rates gives rise to a 2 percent equity premium even in a world *without aggregate uncertainty.*

6. CONCLUDING COMMENTS

In this essay, we present plausible arguments that using an appropriate benchmark for the households' real interest rate, incorporating the effect of government regulations, taxes and intermediation costs can account for a large part of the observed equity premium. As a result, we have a deeper understanding of the role and importance of the abstractions that contribute to the puzzle. While no single explanation fully resolved the anomaly, considerable progress has been made and the equity premium is a lesser puzzle today than it was twenty years ago.

References

Aiyagari, S. R., and M. Gertler. Asset returns with transactions costs and uninsured individual risk. *Journal of Monetary Economics* 27 (1991): 311–331.

Bansal, R., and J. W. Coleman. A monetary explanation of the equity premium, term premium and risk free rate puzzles. *Journal of Political Economy* 104 (1996): 1135–1171.

Brown, S., W. Goetzmann, and S. Ross. Survival. *Journal of Finance* 50 (1995): 853–873.

Constantinides, G. M., J. B. Donaldson, and R. Mehra. Junior can't borrow: A new perspective on the equity premium puzzle. *Quarterly Journal of Economics* 118 (2002): 269–296.

Constantinides, G. M., and D. Duffie. Asset pricing with heterogeneous consumers. *Journal of Political Economy* 104 (1996): 219–240.

Davis, S. J., and P. Willen. Using financial assets to hedge labor income risk: Estimating the benefits. Working paper, University of Chicago (2000).

Donaldson, J. B., and R. Mehra. Risk based explanations of the equity premium forthcoming in *Handbook of the Equity Risk Premium,* R. Mehra, ed. Amsterdam (2008).

Economic Report of the President (2005). Council of Economic Advisors. United States Government Printing Office. Washington.

Ibbotson Associates. Stocks, bonds, bills and inflation. 2005 yearbook. Ibbotson Associates, Chicago (2006).

Krishnamurthy, A., and A. Vissing-Jorgensen. The demand for treasury debt. Working paper, Northwestern University (2007).

Kydland, F., and E. C. Prescott. Time to build and aggregate fluctuations. Econometrica 50 (1982): 1345–1371.

Mankiw, N. G. The equity premium and the concentration of aggregate shocks. *Journal of Financial Economics* 17 (1986): 211–219.

McGrattan, E. R., and E. C. Prescott. Is the stock market overvalued? *Federal Reserve Bank of Minneapolis Quarterly Review* (2000).

McGrattan, E. R., and E. C. Prescott. Average debt and equity returns: Puzzling? *American Economic Review* 93 (2003): 392–397.

McGrattan, E. R., and E. C. Prescott. Taxes, regulations, and the value of U.S. and U.K. corporations. *Review of Economic Studies* 92 (2005): 767–796.

Mehra, R., and E. C. Prescott. The equity premium: A Puzzle. *Journal of Monetary Economics* 15 (1985): 145–161.

Mehra, R., and E. C. Prescott. Intermediated quantities and returns. Working paper, UCSB (2007).