NORTHWESTERN UNIVERSITY

Essays in Asset Pricing and Macroeconomics

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Finance

By

Yizhaq Kleshchelski

EVANSTON, ILLINOIS

June 2008

© Copyright by Yizhaq Kleshchelski 2008

All Rights Reserved

ABSTRACT

Essays in Asset Pricing and Macroeconomics

Yizhaq Kleshchelski

In this dissertation we analyze the decision process of firms and individuals along two dimensions which are central to the field of asset pricing and macroeconomics. In the first chapter, we study the pricing decision of the firm in a framework where customer base matters. Surveys of managers show that the main reason why firms keep prices stable is that they are concerned about losing customers or market share. We construct a model in which firms care about the size of their customer base. Firms and customers form long-term relationships because consumers incur costs to switch sellers. In this environment, firms view customers as long-lived assets. We use a general equilibrium framework where industries and firms are buffeted by idiosyncratic marginal cost shocks. We obtain three main results. First, cost pass-through into prices is incomplete. Second, the degree of pass-through is an increasing function of the persistence of cost shocks. Third, there is a non-monotonic relationship between the size of switching costs and the rate of pass-through. In addition, we characterize the heterogenous response across

industries to marginal cost shocks. The implications of our model are consistent with empirical evidence. We also show an application to the field of international economics.

In the second chapter we study the quantitative implications of the interaction between robust control and stochastic volatility for key asset pricing phenomena. We present an equilibrium term structure model in which output growth is conditionally heteroskedastic. The agent does not know the true model of the economy and chooses optimal policies that are robust to model misspecification. The choice of robust policies greatly amplifies the effect of conditional heteroskedasticity in consumption growth, improving the model's ability to explain asset prices. In a robust control framework, stochastic volatility in consumption growth generates both a state-dependent market price of model uncertainty and a stochastic market price of risk. We estimate the model and show that the model is consistent with key empirical regularities that characterize the bond and equity markets. We also characterize empirically the set of models the robust representative agent entertains, and show that this set is statistically 'small'.

Acknowledgements

I would like to thank my advisors, Martin Eichenbaum and Sergio Rebelo for their continuous guidance, encouragement and support. I am also grateful to Janice Eberly, Lars Hansen, Zhiguo He, Ravi Jagannathan, Arvind Krishnamurthy, and especially to my coauthor, Nicolas Vincent, with whom I hope to collaborate on many more challenging projects in the future.

I wish to thank my parents, Batya and Zvi Kleshchelski, for their love and encouragements throughout the last 33 years. Always supportive of my dreams and choices. Constantly providing unwavering belief in my abilities, even when I could not help but doubt myself. I feel like they share a large part of my achievements.

Finally, I would like to thank my wife Hadas, son Adam and daughter Lia. You followed me patiently in this voyage, lived through the ups and downs. You invested in this dissertation at least as much as I did and therefore it is dedicated to you.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

The behavior of prices is central to the field of macroeconomics. The behavior of asset prices is central to the field of finance. This dissertation aims at furthering our understanding of these topics by examining the decisions firms and individuals make..

How prices respond to shocks, how strongly and how fast they react to changes in the economic environment are all questions which have long interested macroeconomists. The answers are crucial for understanding, among others, the role of monetary policy and the transmission of shocks across countries and sectors. Over the last decades, many theories have been put forward to explain why prices appear sluggish or "sticky". Yet, when surveyed directly about their pricing strategies, managers' actual concerns rarely coincide with the mechanisms most commonly used in macro models. The objective of the first part of this dissertation is to characterize and analyze a framework where price dynamics are affected by the existence of ongoing relationships between the firm and its customers. This environment rationalizes the finding that managers identify customer relations as one of the main reasons for keeping prices stable.

The second part of this dissertation studies the implications of the interaction between robust control and stochastic volatility for key asset pricing phenomena. We quantitatively show that robustness, or fear of model misspecification, coupled with state-dependent volatility provides an empirically plausible characterization of the level and volatility of the equity premium, the risk free rate, and the cross-section of yields on treasury bonds.

We also show that robustness offers a novel way of reconciling the shape of the term structure of interest rates with the persistence of yields. Finally, we quantify the level of robustness encoded in agents' behavior.

CHAPTER 2

Market Share and Price Rigidity (joint with Nicolas Vincent)

2.1. Introduction

This paper analyzes real rigidities in firms' pricing decisions. We focus on the following phenomenon: pass-through from marginal cost to prices is often incomplete. The most obvious example of "incomplete pass-through" is the relatively small impact of exchange rate changes on the retail price of imported goods. There is also evidence of incomplete pass-through from wholesale to retail prices.[1] Using aggregate time-series data, Bils (1987), Rotemberg and Woodford (1999) and Altig, Christiano, Eichenbaum and Linde (2005) argue that prices are less volatile than marginal cost.

There are many theoretical reasons proposed as to why prices are more stable than marginal cost.[2] In surveys, firms report that the main reason they wish to keep prices stable is that they are concerned about losing customers or market share. In contrast, firms give much less weight to factors such as menu costs and costly information which are often emphasized as explanations for price rigidity.

[1]Examples of incomplete pass-through exist in a variety of contexts: see Campa and Goldberg (2002) and Burstein, Eichenbaum and Rebelo (2005) for the case of exchange rates; Besanko, Dubé and Gupta (2005) on the relationship between wholesale and retail prices; Borenstein, Cameron, Gilbert (1992) for gas prices; Neumark and Sharpe (1992) for interest rates; and Peltzman (2000) for a variety of sectors.

[2]See for example Ball and Romer (1990) and the references therein, modern DSGE models with nominal rigidities (Christiano, Eichenbaum and Evans (2005) and Rotemberg and Woodford (1997)), non-constant elasticities of consumer demand (e.g. Dotsey and King (2005)), or costly information (Wiederholt and Mackowiak (2006)).

The interaction between firms and customers has received surprisingly little attention in the macroeconomic literature.[3] The standard framework of monopolistic competition used in macro models is the one developed by Dixit and Stiglitz (1977). Despite its many virtues, it cannot generate incomplete cost pass-through in the absence of nominal frictions. Moreover, in this model there is no distinction between the extensive margin of sales (the number of customers) and the intensive margin (the quantity sold per customer).

We construct a model in which firms care about the size of their customer base. Consumers decide how much of a good to consume and which firm to buy it from. Firms and customers form long-term relationships because consumers incur costs to switch sellers. In this environment firms view customers as long-lived assets. Consequently, they face an intertemporal tradeoff between increasing current profits and building market share for the future.

We embed our model of imperfect competition into a general equilibrium framework where industries and firms are buffeted by idiosyncratic marginal cost shocks. We obtain three main results. First, pass-through is incomplete. Second, the degree of pass-through is an increasing function of the persistence of cost shocks. Third, there is a non-monotonic relationship between the size of switching costs and the rate of pass-through. When switching costs are low, customers are likely to leave in the future and are therefore of little value to the firm. Consequently, firms pass-through a large fraction of marginal cost changes into their prices. As switching costs increase, customers become more attached

[3]A notable exception is Rotemberg (2005). In that paper firms are reluctant to raise prices if they fear that consumers will view the new price as “unfair”. Schmitt-Grohé, Uribe and Ravn (2006) study a model with habit persistence at the good level which can also be related to ours. Other studies include Amano and Hendry (2003) and Ireland (1998). They focus respectively on aggregate inflation persistence and the markup patterns over the business cycle.

and valuable, and pass-through falls. However, when switching costs are so high that customers never switch, the extensive margin is irrelevant and prices move one for one with marginal costs.

The third result implies that there is interesting heterogeneity in the price response across industries following marginal cost shocks. We argue that the model's predictions are in line with the available empirical evidence. Price-setting surveys show that firms which are most concerned about customer relations and with the highest proportion of repeat customers report more stable prices.

Our results are of interest to macroeconomists for at least two reasons. First, to understand how firms respond to idiosyncratic shocks is inherently interesting given the prevalence of such shocks. Second, it is well known that nominal frictions must be combined with real rigidities in order for nominal shocks to have significant and persistent real effects. We conjecture that a combination of our model with Lucas-style imperfect information about the nature of shocks is a promising research avenue.

The outline of this chapter is as follows. Section 2.2 provides an overview of the motivating evidence. Sections 2.3 and 2.4 describe the economic environment as well as the maximization problems faced by households and firms, and the predictions of the model in a static environment. Section 2.5 presents our findings for the dynamic environment and explains the intuition behind the results, while Section 2.6 explores two extensions to the basic framework. Section 2.7 illustrates an application in the context of international economics and Section 2.8 concludes.

2.2. Motivating Evidence

As the list of candidate theories for price rigidity expands, some researchers took to the task of asking firms directly about their pricing behavior. In these studies, managers are asked to rank or assign scores to a number of popular economic theories which are expained to them in non-technical terms. While one might suspect that wording and interpretation issues could hinder the usefulness of such exercise, there is in fact remarkable homogeneity in findings across countries.

Table 2.1 reports some evidence from Fabiani *et al.* (2005). It gathers and summarizes the results from a number of price-setting surveys regarding the relative importance of various theories of price rigidity. The striking feature behind this evidence is the importance that firms attach to factors linked to "customer relations", despite the fact that the actual theory this category refers to may differ across surveys. For example, it includes the implicit contract theory of Okun (1981) where firms keep prices stable in order to build long-term relationships with their customers; the desire of sellers to maintain market share; or their fear of antagonizing customers. Blinder *et al.* (1998) observe that firms often volunteer similar explanations when asked open-ended questions on price rigidity. While it might be difficult to determine which of these variants is most relevant, our emphasis on factors related to customer base and market share appears clearly in line with firms' actual concerns.

Paradoxically, the two mechanisms which have probably garnered the most attention in the state-dependent literature on price stickiness are considered less important by firms. When managers are asked whether price rigidity might be the product of menu costs or costly information gathering, they invariably rank such theories very low. This result is

: Theories Behind Price Rigidity

	Euro	US	CA	SW	UK	BE	ES	FR	NL	AT	PT
Customer relations	1	4	2	1	5	1	1	4	1	1	1
Menu costs	8	6	10	11	11	9	6	6	7	8	7
Costly information	9	-	10	13	-	8	7	-	-	7	-
# of theories	10	12	11	13	11	10	9	7	8	10	9

Note: Rank of different theories based on firm surveys.
Source: Fabiani et al. (2005)

in line with the case study of Zbaracki, Ritson, Levy, Dutta and Bergen (2004): they find that physical menu costs are very small, while customer costs represent 75% of the cost of changing prices. However, this is not to say that those two theories are irrelevant: Ball and Romer (1990) have shown that even small menu costs coupled with some real rigidity, in the spirit of the one we are studying in this chapter, can generate significant nominal price rigidity. Nonetheless, from the perspective of price setters, they do not appear to be the main impediments to price flexibility.

There is also evidence that the degree of price rigidity is related to customer base concerns. The survey on price-setting conducted in Canada by Amirault, Kwan and Wilkinson (2006) offers evidence that there is a significant correlation between the importance of customer relations and price stickiness. They report that "customer relations costs have a very high level of acknowledgement among firms with the stickiest prices. Seventy-six per cent of firms who change their prices only once or not at all during the year recognize this factor as a source of price rigidity" compared with 37% who adjust prices more than 52 times a year. This difference is statistically significant.

Not surprisingly, firms with a higher fraction of repeat customers are also those who are more concerned about factors linked to customer relations. For example, in the survey of Apel, Friberg and Hallsten (2005), "the mean score given to the implicit contract theory is 3.06 [on a scale of 1 to 4] for firms with at least 90% of sales to regular customers, whereas the mean score is 1.94 for firms with less than 10% of their sales to regular customers." Similar findings emerge from the studies by Hall, Walsh and Yates (1997) for the UK and Kwapil, Baumgartner and Scharler (2005) for Austria. In addition, there is evidence that firms with a higher proportion of repeat customers tend to have more rigid prices. Aucremanne and Druant (2005) find that 43% of sticky-price firms have more than 50% of repeat customers, versus 28% for flexible-price firms. Similarly, Hall *et al.* (1997) report "that companies with a greater proportion of long-term customer relationships reviewed and changed prices less frequently than the others."

Recent laboratory studies have also found evidence that price rigidity is more pronounced in a customer market than in an anonymous market. Cason and Friedman (2002) report that in their experiment, when sellers and buyers enter long-term relationships (here because customers face some costs of switching supplier), sellers will often absorb a portion of their cost changes in order to preserve their customer base. Similarly, Renner and Tyran (2004) find that "many sellers do not respond to the cost shock by increasing prices [...] because they hope to reap the gains from trading with loyal customers in the remaining periods of the game."

A number of studies have recently looked into the behavior of individual prices. Bils and Klenow (2004) analyze a dataset of prices collected by the BLS for the U.S. economy, and similar research has been conducted in a number of European economies (Dhyne *et* *al.* (2005)) and other countries (Gagnon (2006)). Despite differences across datasets, some robust findings emerge. First, there is overwhelming evidence that most products exhibit a significant degree of price stickiness: the average monthly frequency of price adjustment is 25 percent in the US and 15 percent in the Euro area. There is, however, considerable heterogeneity in price rigidity along various dimensions. Across categories, services invariably display the stickiest prices, whereas energy goods and unprocessed food prices are the most flexible. Even within categories, there are large differences across products (see Table 2.2). Within services, for example, prices are substantially more rigid in sectors which are typically characterized by long-term relationships between firms and customers (e.g. barbers, beauty services, legal and medical services, etc.). In addition, studies find that traditional corner shops, which arguably have more stable and longer-lived business relationships with their customers, display a significantly higher degree of price rigidity than supermarkets, even after controlling for the type of good.[4] We develop a theory consistent with such findings.

We analyze firms' pricing decisions following sector- and firm-level marginal cost shocks. As pointed out by Golosov and Lucas (2003) and Klenow and Willis (2006), datasets of individual prices show little economy-wide synchronisation of price changes, significant fluctuations in relative prices, as well as price drops which are almost as common as price increases, suggesting a predominant role for non-aggregate shocks. Fabiani *et al.* (2005) present evidence from a number of European countries which suggests that there is also little synchronization within sectors. They use a statistical measure which

[4]See for example Baudry, Le Bihan, Sevestre and Tarrieu (2004). They find that for their reference product, supermarkets are on average twice as likely to change their prices each month compared to traditional corner shops or service outlets.

Table 2.2: Monthly Frequency of Price Changes

Services		*Durable goods*		*Non-durable goods*	
Barber shops	3.9	Plumbing supplies	6.0	Magazines	8.6
Medical services	4.5	Eyeglasses	8.9	Snacks	9.5
Nursing and home care	9.2	Garden supplies	15.5	Wine at home	19.3
Gardening services	11.4	Kitchen furniture	24.1	Cola drinks	38.8
Repair of appliances	16.9	Televisions	31.0	Potatoes	47.3
Automotive repairs	18.5	New cars	39.1	Regular gasoline	78.9

Source: U.S. data, Bils and Klenow (2004)

ranges between 0 and 1, with 1 indicating perfect synchronisation, and find that the median value across sectors ranges from 0.13 to 0.48 depending on the country. However, Veronese, Fabiani, Gattulli and Sabbatini (2005) using Italian data show that this conclusion is highly dependent on the treatment of geography. When product categories also take into account the geographical location of price quotes (e.g. milk in Rome, milk in Milan, etc.), they find that prices are substantially more synchronised: the median synchronisation ratio rises from 0.24 to 0.46. Their finding is consistent with the observation that synchronisation is higher for smaller countries, where markets are more geographically integrated. This evidence suggests a significant role for sectoral shocks in addition to firm-specific disturbances. Also, using factor-augmented vector autoregressions and disaggregated price data, Boivin, Giannoni and Mihov (2007) find that "most of the fluctuations in sectorial inflation rates are due to sector-specific factors."

There are a number of conclusions we draw from the evidence in this section. First, a wide range of surveys find that firms consider factors linked to their customer base to be the main rationale behind keeping prices stable. They also reveal that there is a strong relation between the importance of customer relations, the proportion of repeat customers, and the degree of price stickiness. We show that services, and in particular those sectors where buyer/seller relationships are important, display the most rigid prices. In the next section, we describe a model that can rationalize these findings.

2.3. A macro model with market share dynamics

We develop a tractable model based on micro-foundations in which firms are rationally concerned about their market share position. Our model builds on the work of Ball and Romer (1990) and extends it to a dynamic version based on the standard imperfect competition framework.[5] As such, it collapses to the well-known Dixit-Stiglitz model in certain special cases. The central mechanism is related to the customer market literature (e.g. Phelps and Winter (1970)) under imperfect information (see Stiglitz (1979) and Woglom (1982)).

The environment is comprised of households who consume and provide labor, and firms who produce consumption goods. However, unlike a standard model, the consumption decision here is two-dimensional: households decide not only how much of a particular good to consume, but also which firm to buy it from. The decision to switch supplier is a function of the relative price and a switching cost. The ensuing customer base dynamics render the firm's problem intertemporal.

[5]The static version of Ball and Romer (1990 is used to investigate the interaction of real and nominal rigidities. See also Ireland (1998) for a related extension based on a one-good economy. The objective there is to study the impact of customer flows on the cyclical behavior of markups.

2.3.1. Households

The economy is composed of a continuum of sectors, each producing a good indexed by $i \in [0,1]$. In each sector, there is an infinite number of firms, each selling a distinct brand $k \in [0,1]$.[6] While goods are imperfect substitutes, brands are homogenous and perfectly substituable.

Households are infinitely lived and denoted by $j \in [0,1] \times [0,1]$. Each household j consumes only one brand k of good i.[7] It derives disutility from labor l^j and utility from a basket of goods $\tilde{c}^j$, and solves the following problem:

$$\max U_0^j = E_0 \sum_{t=0}^{\infty} \beta^t u\left(\tilde{c}_t^j, l_t^j\right)$$

$$u\left(\tilde{c}_t^j, l_t^j\right) = \frac{\left(\tilde{c}_t^j\right)^{1-\sigma}}{1-\sigma} - \eta \frac{\left(l_t^j\right)^{1+\epsilon}}{1+\epsilon}$$

$$\tilde{c}_t^j = \left\{ \int_0^1 \left[\left(\delta_{it}^j\right)^{-s_{it}^j} c_{it}^j \right]^{\frac{\gamma-1}{\gamma}} di \right\}^{\frac{\gamma}{\gamma-1}} \tag{2.1}$$

subject to

$$\int_0^1 p_{it}^j c_{it}^j di + E_0 r_{t+1} b_{t+1}^j = b_t^j + w_t l_t^j + \Pi_t$$

[6]Throughout the paper we use the terms "supplier", "producer", "firm" and "seller" interchangeably. Also, we sometimes refer to a "sector" when talking about the set of firms which produce a similar good i.

[7]This is an assumption of the model. However, because the brands are perfect substitutes, the introduction of an infinitesimal cost of consuming a given brand would make it an optimal choice for the household.

where E is the expectation operator, γ is the elasticity of substitution between varieties, and σ is the inverse of the elasticity of intertemporal substitution, or risk aversion parameter. The household supplies homogenous labor and earns the economy-wide nominal wage rate w_t. Households also have access to complete state-contingent claims markets. The stochastic discount factor is given by r_{t+1} such that $E_t r_{t+1} b^j_{t+1}$ is the price at time 0 of a random payment b^j_{t+1} in period $t+1$ (we also impose a no-Ponzi-game constraint). Each household receives an equal share of the period t profits from the firms, Π_t. To avoid confusion, we denote by $\tilde{x}$ any variable x which refers to the aggregate basket of goods.

Our consumption aggregator (2.1) takes into account the switching decision of the household: we write $s^j_{zt} = 1$ if household j switches seller for good z at time t and 0 otherwise. Clearly, the endogenous choice by the consumer to leave his current seller will be a function of the parameter δ^j_{it}, which quantifies the utility implications for household j of changing the brand of good i at time t: *ceteris paribus*, a higher δ^j_{it} reduces the incentive of the consumer to switch brands. We will refer to δ as a *switching cost*. At time t, the household draws a new independently and identically distributed idiosyncratic switching cost $\delta^j_{it} \in \left[\underline{\delta}, \overline{\delta}\right]$, $\overline{\delta} \geq \underline{\delta} > 0$ from a known time-invariant continuous distribution with a cumulative distribution function F and probability density function f.[8]

We do not rule out $\delta < 1$: there are instances when a customer will find it optimal to leave his current seller even if the relative price is low. That brand switching occurs for non-price reasons is widely acknowledged in the marketing literature (see for example Ganesh, Arnold and Reynolds (2000) and Keaveney (1995)). Reasons may include poor product and service quality, inconvenience, relationship quality, etc. We model these

[8]An alternative interpretation would be that the switching cost is constant over time and common across households, but that consumers are hit by i.i.d. taste shocks. The sum of the two would correspond to δ.

exogenous factors by allowing for low values of the switching cost δ. Consequently, our model implies that in steady state a non-zero mass of customers switches suppliers every period.

The timing of household j's sequence of decisions for the purchase of a typical good i is as follows: In period $t-1$, household j bought good i from one, and only one, supplier k which we call his "home seller". At time t, after drawing a switching cost, δ_{it}^{j}, the household observes the price $p_{it}(k)$ set by his home seller as well as the distribution of prices of other brands of good i over the unit interval. We denote the continuum of all sector prices as $\{p_{it}(l)\}_{l\in[0,1]}$. The consumer can then decide to remain with his home seller and pay $p_{it}(k)$, in which case we denote his decision by $s_{it}^{j}=0$. Conversely, he can opt to switch and be *randomly* assigned to a different seller ($s_{it}^{j}=1$). Random matching is consistent with our assumption of imperfect information (households only know the distribution of sector prices). A consumer can only switch once per period. Finally, he decides the quantity of good i to buy, c_{it}^{j}.

From the household's problem, the optimality conditions with respect to l_t^j and b_{t+1}^j are standard:

$$\eta \left(l_t^j\right)^{\epsilon} = \mu_t^j w_t \tag{2.2}$$

$$\mu_t^j E_t r_{t+1} = \beta E_t \mu_{t+1}^j. \tag{2.3}$$

The first-order condition with respect to good i yields:

$$\left(\tilde{c}_t^j\right)^{\frac{1}{\gamma}-\sigma}\left(\delta_{it}^j\right)^{\frac{s_{it}^j(1-\gamma)}{\gamma}}\left(c_{it}^j\right)^{-\frac{1}{\gamma}}=\mu_t^j p_{it}^j \tag{2.4}$$

where μ_t^j is the multiplier on the household's budget constraint. We can rewrite the budget constraint as:

$$\tilde{p}_t^j \tilde{c}_t^j + E_0 r_{t+1} b_{t+1}^j = b_t^j + w_t l_t^j + \Pi_t$$

where the price index for the basket of goods $\tilde{p}_t^j$ is household-specific. The first-order condition with respect to $\tilde{c}_t^j$ yields:

$$\left(\tilde{c}_t^j\right)^{-\sigma}=\mu_t^j \tilde{p}_t^j. \tag{2.5}$$

Using the optimality conditions (2.4) and (2.5), we get a general demand function of household j for good i as a function of the switching decision s_{it}^j:

$$c_{it}^j=\begin{cases}\left(\frac{p_{it}^j}{\tilde{p}_t}\right)^{-\gamma}\tilde{c}_t^j & \text{if } s_{it}^j=0\\ \left(\delta_{it}^j\right)^{1-\gamma}\left(\frac{p_{it}^j}{\tilde{p}_t}\right)^{-\gamma}\tilde{c}_t^j & \text{if } s_{it}^j=1.\end{cases} \tag{2.6}$$

If firm k is the home seller, then the relevant price when the household decides to stay is $p_{it}^j=p_{it}(k)$, whereas it is a random draw from the set of prices $\{p_{it}(l)\}_{l\in[0,1]}$ in the event of a switch. As each consumer faces different prices, the aggregate price index $\tilde{p}_t^j$ is household specific. However, in the symmetric equilibrium, this will no longer be the case.

2.3.1.1. Switching decision. In order to facilitate the exposition of the switching decision of the consumer, we consider a recursive representation of the household's problem.

Since there is a continuum of goods we can focus on the choice to switch in one sector i in isolation and disregard other variables which are invariant to the switching decision.

We denote the sequence of future prices charged by firm k as $p_i^t(k) = \{p_{it+z}(k)\}_{z=0}^{\infty}$. As we need to keep track of the distribution of prices, we write the collection of price sequences for good i as $\{p_i^t(l)\}_{l\in[0,1]}$. We can then define the value for a consumer of staying ($s_{it} = 0$) with his home seller k as:[9]

$$V_0\left[\left\{p_i^t\right\}, p_i^t(k)\right] = U\left[c_{it}\left(p_{it}(k)\right)\right] + \beta E \max\left[\begin{array}{c} V_0\left[\left\{p_i^{t+1}\right\}, p_i^{t+1}(k)\right], \\ V_1\left[\left\{p_i^{t+1}\right\}, \delta_{it+1}\right] \end{array}\right]$$

Recall that when making the switching decision, the household has already observed the price of its current supplier, hence the instantaneous utility at time t is known. The expression for the continuation value indicates that the consumer will face a similar choice tomorrow. The expected value of leaving ($s_{it} = 1$) the home seller is given by:

$$V_1\left[\left\{p_i^t\right\}, \delta_{it}\right] = \int_0^1 M_{it-1}(l)\left\{U\left[c_{it}\left(p_{it}(l), \delta_{it}\right)\right] dl + \beta E \max\left[\begin{array}{c} V_0\left[\left\{p_i^{t+1}\right\}, p_i^{t+1}(l)\right], \\ V_1\left[\left\{p_i^{t+1}\right\}, \delta_{it+1}\right] \end{array}\right]\right\} dl.$$

The expression corresponds to an *expected* value because the consumer only knows the distribution of sectoral prices at the time of switching. Once he decides to switch, we assume that the probability of being matched with seller l is proportional to its previous period's market share, which we denote as $M_{it-1}(l)$. This is similar to Phelps and Winter (1970), and simply implies that big firms will get a larger fraction of the mass of

[9] While $p_i^t(k)$ is technically part of $\{p_i^t\}$, we write it separately to emphasize that the consumer knows only the price charged by his home seller, as well as the distribution of prices within the sector.

switchers.[10] In addition, the realized switching cost is now an inherent part of the value function since it determines the utility at time t.

The threshold switching cost, denoted by $\widehat{\delta}_{it}$, is the one which makes the consumer indifferent between switching and staying:

$$V_0\left[\left\{p_i^t\right\}, p_i^t(k)\right] = V_1\left[\left\{p_i^t\right\}, \widehat{\delta}_{it}\right]. \tag{2.7}$$

That is, all customers for which $\delta_{it}^j > \widehat{\delta}_{it}$ will remain with their home supplier of good i while all those with $\delta_{it}^j \leq \widehat{\delta}_{it}$ will find it optimal to switch.

2.3.2. Firms

A firm in this environment is indexed by a pair $g \in G$ indicating the good and the brand, where $G \equiv \{(i,k) : i \in [0,1], k \in [0,1]\}$. Clearly, the firm is atomistic and will take the aggregate variables as well as the decisions of its competitors as given. $M_{it}(k)$ denotes the mass of customers of firm (i,k) at time t. We refer to $M_{it}(k)$ as the "market share" or "customer base".

Consider the problem of a seller k of good i who comes into period t with a market share $M_{it-1}(k)$. The firm observes the realization of a sector-specific productivity shock at time t which is common to all producers in sector i and ponders the possibility of changing its price $p_{it}(k)$.[11] Based on its pricing decision, the firm's current customers then optimally decide between staying or leaving their home seller. When changing its relative price supplier k affects the threshold switching cost $\widehat{\delta}_{it}(k)$: if it increases its price more customers will now find it optimal to switch brand, which raises $\widehat{\delta}_{it}(k)$. This will

[10] It is easy to verify that the condition $\int_0^1 M_{it-1}(l)\,dl = 1$ is satisfied every period.

[11] In Section 2.5.2, we also consider firm-specific shocks under a special case.

lead to a depletion of the firm's customer base available next period. Note that this dimension is entirely missing from Ball and Romer (1990): in their framework, sellers and buyers are randomly matched every period, with the consequence that any change in the mass of customers today has no impact on future profits.

To determine the evolution of market share, we define two groups of customers over which the firm is not allowed to price discriminate. The first group corresponds to *repeat customers*: it consists of consumers who bought from firm k at $t-1$ and who, after observing firm k's price as well as the distribution of prices within sector i, have decided against switching. Their mass corresponds to the portion of customers from last period, $M_{it-1}(k)$, who draw a switching cost larger than $\widehat{\delta}_{it}(k)$:

$$M_{it}^{R}(k) = M_{it-1}(k)\left[1 - F\left(\widehat{\delta}_{it}(k)\right)\right].$$

The assumption that customers have to draw a new i.i.d. switching cost every period is crucial here: if we did not impose this assumption, we would need to keep track of the *distribution* of current customers, indexed by their respective δ. Instead, the mass of customers that the firm keeps from one period to the next is distributed according to the time-invariant distribution F.

The second group is composed of new customers who randomly arrive from other sellers. As in Phelps and Winter (1970), we assume that the rate at which a firm attracts new customers is proportional to its previous period's market share. In other words, big firms will get a larger fraction of the mass of switchers.[12] Since consumers are randomly matched and are not allowed to switch more than once per period, firm k's actions today

[12] Such an assumption implies that the growth rate of the market share is independent of the size of the firm. This is also known as Gibrat's law, and it has received empirical support in the literature.

have no impact on the arrival rate of customers in period t. The mass of new customers is given by:

$$M_{it}^{N}(k) = M_{it-1}(k) \int_{0}^{1} M_{it-1}(l) F\left(\widehat{\delta}_{it}(l)\right) dl$$

and the law of motion of the customer base at time t is:[13]

$$\begin{aligned} M_{it}(k) &= M_{it}^{R}(k) + M_{it}^{N}(k) \\ &= M_{it-1}(k)\left[1 - F\left(\widehat{\delta}_{it}(k)\right) + \int_{0}^{1} M_{it-1}(l) F\left(\widehat{\delta}_{it}(l)\right) dl\right]. \end{aligned} \tag{2.8}$$

Next, we turn our attention to the demand schedule faced by seller (i, k). The quantity sold to repeat customers is:

$$c_{it}^{R}(k) = M_{it-1}(k) \int_{\widehat{\delta}_{it}(k)}^{\infty} \left[\int_{j:\delta_{it}^{j}=\delta} \left(\frac{p_{it}(k)}{\tilde{p}_{t}^{j}}\right)^{-\gamma} \tilde{c}_{t}^{j} dj\right] d\delta \tag{2.9}$$

Since the aggregate price and consumption indexes are household specific, we need to explicitly integrate over each household. Notice that the switching costs drawn do not enter this expression, since by opting to stay with their home seller, repeat customers avoid suffering any utility penalty from switching. The pricing decision has two effects on the demand of repeat customers. First, it impacts the intensive margin through the term $\left(\frac{p_{it}(k)}{\tilde{p}_{t}^{j}}\right)^{-\gamma} \tilde{c}_{t}^{j}$. Second, it influences the extensive margin by affecting the lower bound of

[13]It is easy to verify that given the initial condition $\int_0^1 M_{i,-1}(k)\, dk = 1$, it must be that (1) the mass of switchers is equal to the mass of new customers, and that (2) $\int_0^1 M_{i,t}(k)\, dk = 1$, $\forall t \geq 0$.

the outer integral, $\widehat{\delta}_{it}(k)$. Next, we can write the demand from new customers

$$c_{it}^{N}(k) = M_{it-1}(k)\int_{0}^{1} c_{it}^{L}(l)\,dl \tag{2.10}$$

where $c_{it}^{L}(l)$ is the consumption of customers leaving seller l

$$c_{it}^{L}(l) = M_{it-1}(l)\int_{\underline{\delta}}^{\widehat{\delta}_{it}(l)}\left[\int_{j:\delta_{it}^{j}=\delta}\left(\delta_{it}^{j}\right)^{1-\gamma}\left(\frac{p_{it}(k)}{\tilde{p}_{t}^{j}}\right)^{-\gamma}\tilde{c}_{t}^{j}dj\right]d\delta.$$

Unlike the case of repeat customers the price-setting decision here only impacts the intensive margin. The total demand schedule faced by firm (i,k) is then simply the sum of expressions (2.9) and (2.10):

$$c_{it}(k) = c_{it}^{R}(k) + c_{it}^{N}(k)\,. \tag{2.11}$$

The dynamic problem of a supplier k of good i is to solve the following problem:

$$\widehat{\Pi}_{i0}(k) = \max\sum_{t=0}^{\infty}\beta^{t}E_{0}\mu_{t}\left\{c_{it}(k)\left[p_{it}(k) - \frac{w_{t}}{z_{it}}\right]\right\} \tag{2.12}$$

subject to (2.8) and (2.11), the linear production function $c_{it}(k) \leq z_{it}l_{it}(k)$ and the initial condition $M_{i0}(k) = 1$. $\frac{w_t}{z_{it}}$ corresponds to the marginal cost at time t. The firm discounts profits using the marginal value of a dollar to the households (and owners), μ_t, which varies over time in the general equilibrium version of the model.[14]

[14]Technically, μ_t is an average of the individual μ_t^j.

When the firm is setting its price, $p_{it}(k)$, it takes into account four effects. First, the firm considers the impact on profit per unit sold, i.e. $\left(p_{it}(k) - \frac{w_t}{z_{it}}\right)$; second, the effect on the intensive margin for all customers; third, the consequence on the extensive margin for repeat customers through the impact on $\widehat{\delta}_{it}(k)$; and fourth, the indirect effect on future market share. In the case of a rise in the price relative to the other brands, the first effect is positive (raising $p_{it}(k)$ increases per-unit profit) while all the others are negative.

Once we rewrite the problem in Lagrangean form, the first-order condition with respect to $c_{it}(k)$ is given by:

$$p_{it}(k) - \frac{w_t}{z_{it}} = \lambda_{it}(k) \tag{2.13}$$

where $\lambda_{it}(k)$ is the Lagrange multiplier on the demand faced by the firm. Equation (2.13) simply equates the value for firm (i, k) of selling one more unit of the good, $\lambda_{it}(k)$, to the per-unit profit, $p_{it}(k) - \frac{w_t}{z_{it}}$.

The optimality condition with respect to $p_{it}(k)$ yields

$$c_{it}(k) = -\lambda_{it}(k)\left[\frac{\partial c_{it}^{N}(k)}{\partial p_{it}(k)} + \frac{\partial c_{it}^{R}(k)}{\partial p_{it}(k)}\right] - \upsilon_{it}(k)\frac{\partial M_{it}^{R}(k)}{\partial p_{it}(k)} \tag{2.14}$$

where $\upsilon_{it}(k)$ is the Lagrange multiplier associated with the law of motion of the customer base (2.8). To gain intuition for (2.14), consider the case of an increase of one unit in the price $p_{it}(k)$. The left-hand side gives the benefit of such action: it raises revenues by the quantity sold. The right-hand side defines the costs. First, raising the price means that demand from both new and repeat customers will fall, through the quantity consumed of each customer for both groups as well as the extensive margin for repeat customers. The last term identifies the negative impact on the mass of customers which will be available for the future: it multiplies the marginal value of one more unit of customer base, $v_{it}(k)$, by the change in market share following the price increase.

Finally, the derivative of the Lagrangean with respect to the market share $M_{it}(k)$ is:

$$\begin{aligned} v_{it}(k) &= \beta E_t \frac{\mu_{t+1}}{\mu_t} \lambda_{it+1}(k) \left[\frac{\partial c_{it+1}^N(k)}{\partial M_{it}(k)} + \frac{\partial c_{it+1}^R(k)}{\partial M_{it}(k)} \right] \\ &\quad + \beta E_t \frac{\mu_{t+1}}{\mu_t} v_{it+1}(k) \left[\frac{\partial M_{it+1}^N(k)}{\partial M_{it}(k)} + \frac{\partial M_{it+1}^R(k)}{\partial M_{it}(k)} \right]. \end{aligned} \tag{2.15}$$

Equation (2.15) describes the composition of the marginal value of the market share, which is purely forward-looking.[15] First, raising the customer base increases sales tomorrow by having a larger mass of repeat customers as well as attracting more new consumers (since the firm is now larger). Both effects are evaluated by the marginal value to the firm of selling one more unit, $\lambda_{it+1}(k)$. Second, it boosts the mass of customers available in the future, which has an expected value of $v_{it+1}(k)$. On the basis of these first-order conditions, it becomes clear that it is the dynamic nature of the market share, through the presence of $v_{it}(k)$, which renders the firm's problem intertemporal.

2.3.3. Reaction of the extensive margin

We still need an expression for the impact of the pricing decision on the extensive margin of repeat customers. Consider for example this term from equation (2.14):

$$\frac{\partial M_{it}^R(k)}{\partial p_{it}(k)} = -M_{it-1}(k) f\left(\widehat{\delta}_{it}(k)\right) \frac{\partial \widehat{\delta}_{it}(k)}{\partial p_{it}(k)}.$$

[15] Today's market share, $M_{it}(k)$, does not enter the demand schedule at time t. Instead, it only affects the firm's problem because it corresponds to the mass of customers which the firm will start with at $t+1$.

The expression $\partial\widehat{\delta}_{it}(k)/\partial p_{it}(k)$ determines the reaction of the threshold switching cost to price changes. $\widehat{\delta}_{it}(k)$ is only implicitly defined by the following relation, which is a function of all future prices:

$$V_0\left[\left\{p_i^t\right\}, p_i^t(k)\right] = V_1\left[\left\{p_i^t\right\}, \widehat{\delta}_{it}(k)\right]. \tag{2.16}$$

This setup exhibits a time-consistency problem: customers will be less inclined to switch away from home seller k if it promises to charge low prices in the future. Hence, firms have an incentive at time t to announce low future prices, but later renege on their promises. To deal with this problem, we assume that firms cannot commit to future prices. Instead, all agents in the model form expectations about $p_i^{t+1}(k)$ by solving firm k's problem sequentially.

The problem is further complicated by another issue. Consider a firm (i,k) which raises its relative price at time t. Through its action, the seller will affect the customer base in the future. As the state of the firm at $t+1$ has changed, it should impact consumers' expectations about future prices. This, in turn, has an effect on the forward-looking switching decision of customers at time t. In other words, the object $\partial\widehat{\delta}_{it}(k)/\partial p_{it}(k)$ affects the firm's pricing decision, and vice-versa.

A characteristic of our model allows us to circumvent this recursion. First rewrite the firm's problem at time t as:

$$\widehat{\Pi}_{it}(k) = \max\sum_{\tau=t}^{\infty}\beta^{\tau-t}E_t\mu_\tau\pi_{i\tau}(k)$$

where

$$\pi_{i\tau}(k) = c_{i\tau}(k)\left[p_{i\tau}(k) - \frac{w_\tau}{z_{i\tau}}\right].$$

It is evident from (2.9), (2.10), and (2.11) that the demand schedule faced by the firm is proportional to last period's market share, which allows us to rewrite instantaneous profits as

$$\pi_{i\tau}(k) = M_{i\tau-1}(k)\,\hat{c}_{i\tau}(k)\left[p_{i\tau}(k) - mc_{i\tau}\right]$$

where $\hat{c}_{i\tau}(k) = c_{i\tau}(k)/M_{i\tau-1}(k)$ is not a function of $M_{i\tau-1}(k)$ anymore. We know from the law of motion of the customer base, (2.8), that the market share at any point in the future is proportional to the initial condition

$$M_{i\tau}(k) = M_{it-1}(k)\prod_{y=t}^{\tau}\left[1 - F\left(\widehat{\delta}_{it}(k)\right) + \int_0^1 M_{it-1}(l)\,F\left(\widehat{\delta}_{it}(l)\right)dl\right]$$

which allows us to express profits in any period $\tau > t$ as

$$\pi_{i\tau}(k) = M_{it-1}(k)\,\Psi_{i\tau}(k)$$

where

$$\Psi_{i\tau}(k) = \prod_{y=t}^{\tau-1}\left[1 - F\left(\widehat{\delta}_{it}(k)\right) + \int_0^1 M_{it-1}(l)\,F\left(\widehat{\delta}_{it}(l)\right)dl\right]\hat{c}_{i\tau}(k)\left[p_{i\tau}(k) - mc_{i\tau}\right].$$

Finally, the maximization problem of the firm at time t becomes:

$$\widehat{\Pi}_{it}(k) = M_{it-1}(k)\max_{\{p_{it}(k)\}}\sum_{\tau=t}^{\infty}\beta^{\tau-t}E_t\mu_\tau\Psi_{i\tau}(k) \tag{2.17}$$

where μ_τ and $\Psi_{i\tau}(k)$ are not functions of $M_{it-1}(k)$. As the mass of customers the firm is starting with acts only as a scale factor in (2.17), we know that it has no effect on the pricing decision. This is a crucial finding: it implies that if seller k changes its price at time t, it will not affect customers' expectations about future prices, even if the market share is perturbed.[16] Formally, from the point of view of any consumer j we now know that:

$$\frac{\partial E_t^j p_{it+s}(k)}{\partial p_{it}(k)} = 0, \ \ \forall \ s \geq 1.$$

This result allows us to find the derivative of the threshold switching cost with respect to price from (2.16) by applying the implicit function theorem. We refer the reader to Appendix 1 for the details and note that the object of interest is given by:

$$\frac{\partial \widehat{\delta}_{it}(k)}{\partial p_{it}(k)} = \left(\widehat{\delta}_{it}(k)\right)^{\gamma} \left[\int_0^1 \left(\frac{p_{it}(k)}{p_{it}(l)} \right)^{\gamma} p_{it}(l)\, dl \right]^{-1}. \tag{2.18}$$

2.3.4. Equilibrium and steady state

Since our focus will be on sector-specific shocks, we consider a symmetric equilibrium where all firms *within each sector* start in the first period with equal mass of customers $M_{i,-1}(k) = 1$ and set the same price $p_{it}(k) = p_{it}$ $\forall k, t$. This implies that $M_{it}(k) = 1$, $c_{it}(k) = c_{it}$, $\nu_{it}(k) = \nu_{it}$ and $\lambda_{it}(k) = \lambda_{it}$ for all k, t. Households are not perfectly identical in equilibrium: for a particular good, some switch at time t while others stay with their home seller. However, the relevant aggregate variables $\tilde{p}_t^j$, $\tilde{c}_t^j$ and μ_t^j are not household-specific anymore, even if we do not impose symmetry across sectors. We prove this result

[16] In other words, this property allows us to focus on Markov perfect equilibria where the firm's pricing decision is not a function of its past actions.

in Appendix 2 and derive the expression for $\tilde{p}_t$. We also prove that under symmetry:

$$c_{it} = A(1)\left(\frac{p_{it}}{\tilde{p}_t}\right)^{-\gamma}\tilde{c}_t$$

where

$$A(1) = 1 - F(1) + \int_{\underline{\delta}}^{1} \delta^{1-\gamma} dF(\delta)$$

and the first-order conditions become

$$\lambda_{it} = p_{it} - \frac{w_t}{z_{it}} \tag{2.19}$$

$$c_{it} = \lambda_{it}\gamma A(1)\left(\frac{p_{it}}{\tilde{p}_t}\right)^{-\gamma-1}\frac{\tilde{c}_t}{\tilde{p}_t} + \lambda_{it}\frac{1}{p_{it}}f(1)\left(\frac{p_{it}}{\tilde{p}_t}\right)^{-\gamma}\tilde{c}_t + v_{it}f(1)\frac{1}{p_{it}} \tag{2.20}$$

$$v_{it} = \beta E_t\frac{\mu_{t+1}}{\mu_t}\lambda_{it+1}A(1)\left(\frac{p_{it+1}}{\tilde{p}_{t+1}}\right)^{-\gamma}\tilde{c}_{t+1} + \beta E_t\frac{\mu_{t+1}}{\mu_t}v_{it+1}. \tag{2.21}$$

As a side note, the short-run elasticity of the demand for a single producer around the symmetric equilibrium is given by

$$\left.\epsilon_{c_{it}(k),p_{it}(k)}\right|_{\frac{p_{it}(k)}{p_{it}}=1} = \gamma + \frac{f(1)}{A(1)}. \tag{2.22}$$

In this framework the elasticity faced by a particular seller is greater than in the Dixit-Stiglitz case, where it simply equals γ: $f(1)$ represents the marginal movement in the extensive margin due to the price change, and it is deflated by the size of the taste-adjusted mass of customers. One can also show that a firm which charges $p_{it}(k) > p_{it}$, permanently, will eventually see its market share vanish. Therefore, the long-run elasticity faced by the firm is infinite, i.e. for any $p_{it}(k) = p_{it} + \varepsilon$ where $\varepsilon > 0$, we have that $\lim_{t \longrightarrow \infty} c_{it}(k) = 0$.

The steady state values of the control variables are:

$$p_i = \frac{w}{z_i}\left[\gamma + \frac{f(1)}{A(1)} + \frac{\beta f(1)}{1-\beta}\right]\left[\gamma + \frac{f(1)}{A(1)} + \frac{\beta f(1)}{1-\beta} - 1\right]^{-1}$$
$$\lambda_i = p_i - \frac{w}{z_i}$$
$$c_i = A(1)\left(\frac{p_i}{\tilde{p}}\right)^{-\gamma}\tilde{c}$$
$$v_i = \frac{\beta}{1-\beta}\lambda_i c_i \tag{2.23}$$

Despite the fact that the long-run elasticity is infinite, the steady state gross markup is larger than one. This is because firms discount future profits by $\beta < 1$, which means that in the limit they put no weight on the possibility that sales will eventually vanish. However, it is clear that the markup goes to 1 as $\beta \to 1$. In most parameterizations, we find that the steady state markup is indeed very small because of the competition from other sellers of the same good.

2.4. Analytics under the static case

As is typically the case with this type of forward-looking model, we cannot derive closed-form solutions for the various endogenous variables. An exception is when the firm's problem is static: when $\beta = 0$ firms do not care about the future, and hence only consider the impact of their pricing decisions on the current period's mass of customers and their level of consumption. This can be seen directly from (2.21), which implies that $v_{it} = 0$ when $\beta = 0$ and thus breaks the intertemporal aspect of the firm's problem. As we later argue that it is the dynamic elements that arise from our model which deliver the important results, we first show that if the firm's problem is static, the equilibrium is indeed one where the standard predictions of the Dixit-Stiglitz model hold.

Since $v_{it} = 0$, we can express the price p_{it} as only a function of some parameters and the current marginal cost $mc_{it} = \frac{w_t}{z_{it}}$:

$$p_{it} = \left(\frac{\gamma A(1) + f(1)}{(\gamma - 1) A(1) + f(1)} \right) mc_{it}. \tag{2.24}$$

The assumptions about the distribution of the switching cost, δ, have an impact on the level of the markup. In the special case where the distribution is such that $A(1) = 1$ and $f(1) = 0$, the markup is simply equal to $\frac{\gamma}{(\gamma-1)}$, similar to the Dixit-Stiglitz case.

Most importantly, (2.24) implies that the cost pass-through is complete in the static version of our model: a rise of 5% in the marginal cost will translate into a 5% increase in prices. This is similar to the result in a standard Dixit-Stiglitz framework. In Appendix 3, we prove that such a strategy is the unique symmetric equilibrium, that is independent of our parameterization, and in particular holds for any distribution of δ.

2.5. Pricing when customer base matters

In this section we solve for the equilibrium prices and quantities and discuss the properties of the model.

2.5.1. Sectoral shocks and pricing

We focus our attention on a setting where a sector i is hit by idiosyncratic productivity shocks, common to all the firms within that sector, and study the pricing behavior of a typical seller k. The atomistic nature of sector i implies that the aggregate consumption and price levels, $\tilde{c}_t$ and $\tilde{p}_t$, the marginal utility, μ_t, as well as the wage rate, w_t, are all time-invariant, and we therefore drop their time subscripts.[17] The law of motion of sectoral productivity is given by:

$$\ln(z_{it}) = \rho_z \ln(z_{it-1}) + \varepsilon_{it}^z$$

where ε_{it}^z is a shock specific to sector i. We need to solve for p_{it}, c_{it}, λ_{it}, and ν_{it}.

For our benchmark parameterization we pick values which are comparable to those found in the literature and later show the impact of each of them on our results. We set $\beta = 0.99$ and $\gamma = 5$, which is in line with the parameters estimated by Christiano *et al.* (2005) for a quarterly model of the U.S. economy. We have no strong prior on the distribution of δ; the evidence on switching costs is generally limited to very specific industries, and it is unclear how these estimates could be directly linked to our framework. In our benchmark, we assume that the switching cost δ is distributed lognormally, with support over the interval $(0, \infty)$, mean $\mu_{\ln\delta} = 0.15$ (or $\mu_\delta \approx 1.16$, so that on average switching implies a utility reduction of 16%) and variance $\sigma_{\ln\delta} = 0.1$. This parameterization implies that in equilibrium about 7% of customers in a particular sector will want to switch supplier in a given period ($F(1) \simeq 0.07$) and that the expected duration of a match is about 3 years. The role of the distributional assumptions is carefully analyzed in a later section. We use Dynare to solve the model by linearization techniques and compute the impulse responses. We start by considering a negative productivity shock which raises the marginal cost of

[17] The values chosen for those variables do not affect the results in this section.

all producers in sector i, $\frac{w}{z_{it}}$, by 1% in the first period. As a benchmark, we consider i.i.d. fluctuations in the marginal cost by setting $\rho_z = 0$.

As is well known, in a standard Dixit-Stiglitz model such shocks imply a reaction of the price of good i, p_{it}, perfectly proportional to the movement in marginal cost. In other words, the markup remains constant. As we show in Section 2.4, this is also true in the static version of our model. However, Figure 2.1 makes it clear that the addition of intertemporal market share considerations to the standard model breaks this one-for-one relationship between prices and marginal cost (unless otherwise stated, the values on the y-axis correspond to percentage deviations from steady state).

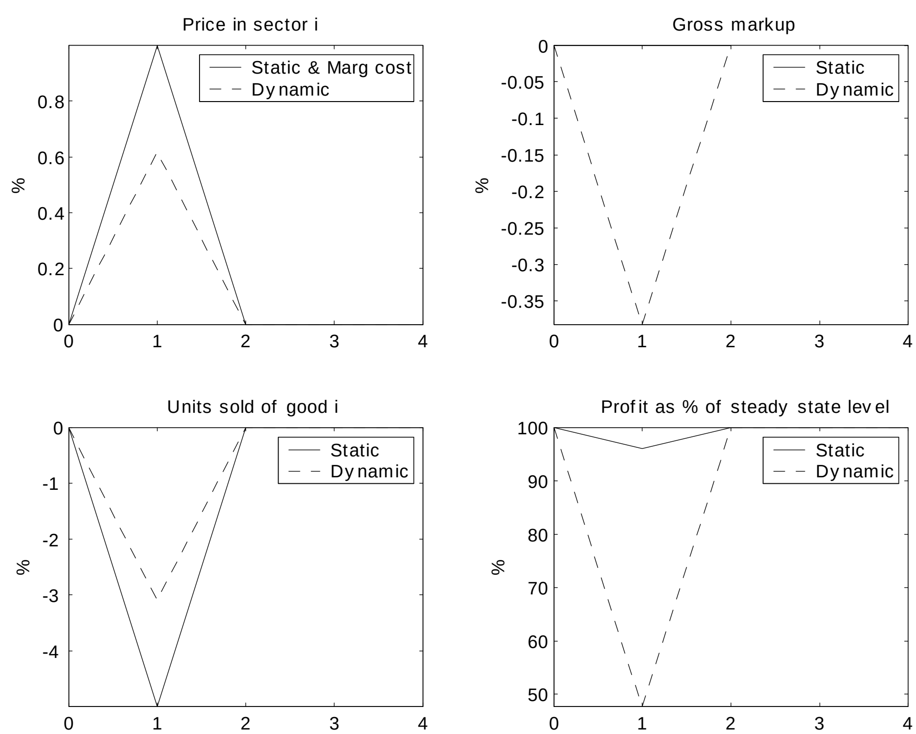

Figure 2.1: Response to a 1% increase in the marginal cost of sector i

In response to a 1% increase in their marginal cost in period 1, firms decide to raise their prices in order to mitigate the negative impact on their profit margin. However, unlike the Dixit-Stiglitz case, the pass-through of marginal cost to price is only about 60%. Hence, in an environment where sectors or industries are hit by idiosyncratic marginal cost shocks, our model yields price rigidity and a time-varying markup. Since the price of good i rises relative to the price of other goods, c_{it} falls as expected. In Appendix 4, we confirm analytically that full pass-through cannot be an equilibrium in this environment.

Profits, which are small in steady state, are heavily affected by the reduction in markup.[18] In fact, an interesting implication of our model is that firms may willingly and optimally decide to sustain instantaneous negative profits for a certain period of time in order to preserve their market share, without having any incentive to exit. This happens because in our environment, exiting in a single period has severe consequences for the future: it implies that firms lose a customer base which will later be difficult to rebuild.[19] It clearly discourages sellers from shutting down operations temporarily, a feature which we believe is realistic and desirable.

To better understand why we obtain a time-varying markup in our model, we first define a new variable which corresponds to the marginal value of the extensive margin (or

[18]Small steady-state profits are due to the low markup and the fact that we focus on a technology with constant returns to scale. Allowing for decreasing returns would raise the level of profits and make the response in Figure 2.1 look less pronounced.

[19]In fact, under our assumption that the mass of switchers in a particular period is distributed across sellers in proportion to their market share, firms could never regain back their customer base after exiting. But even under less extreme environments, the market share dynamics would create a strong disincentive to exit the market only for a few periods. Also, the same rationale explains why there is no incentive for outsider firms to enter the market.

of an extra customer) at time t:

$$\bar{v}_{it} = \lambda_{it} \left(\frac{p_{it}}{\tilde{p}} \right)^{-\gamma} \tilde{c} + v_{it}. \tag{2.25}$$

The second term on the right-hand side corresponds to the forward-looking value of the customer base from (2.15). In addition, repeat customers lead to additional sales at time t, an effect captured by the first term in (2.25). In Appendix 5, we show that under our benchmark parameterization we can derive the following approximate relation

$$\widehat{mk}_{it} \approx -\frac{\bar{v}_i}{mk_i} \left[\beta E_t \widehat{\bar{v}}_{it+1} - \widehat{\bar{v}}_{it} \right] \tag{2.26}$$

where $mk_{it} = p_{it}/mc_{it}$ is the gross markup and the hatted variables refer to percentage deviations from steady state. Equation (2.26) is central to the intuition behind our results: it shows that the optimal markup today is directly linked to the expected movements in the value of the extensive margin. In other words, when the marginal value of the mass of customers is relatively high in the future, we expect the firm to lower its markup.

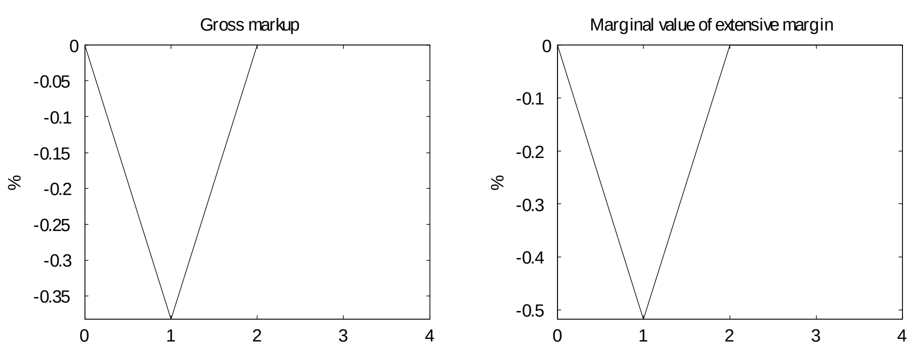

Figure 2.2: Markup and marginal value of extensive margin

The impulse responses in Figure 2.2 confirm this analytical relation. The solid lines reproduce the reaction of the markup and our new variable, $\bar{v}_{it}$, to the shock described earlier (an increase in the sectoral marginal cost of 1% in period 1). When facing a rise in marginal cost, the firm realizes that maintaining its profit margin requires raising the price proportionately. However, a higher price means that each customer now consumes less of the good at the intensive margin, and the value of the marginal customer, $\bar{v}_{it}$, is consequently diminished. As the shock is transitory, the seller expects the environment to go back to steady state and the value of the customer base to rise in later periods. For a firm facing a dynamic problem, this in turn affects its pricing decision: it now becomes optimal for the seller to absorb today a portion of the rising marginal cost into its markup in order to attract customers who are expected to be more valuable in the future. This mechanism results in price fluctuations which are muted relative to the standard Dixit-Stiglitz model.

The intuition for the case of a fall in the marginal cost is simply reversed: there, firms would have an incentive to raise their markup today since customers are not as valuable in the future. This result is related to Klemperer's (1995) observation that in an environment with switching costs, firms will most likely respond to positive demand conditions by raising prices because they "prefer to take profits in the current period rather than in the relatively less attractive future." In our simulations, demand conditions are endogenously changing through the intensive margin following price fluctuations.

Based on the intuition behind the benchmark results, one might expect that firms would react differently based on their expectations about the persistence of the marginal cost shock. Figure 2.3 confirms this conjecture: as the persistence of the shock decreases, the relative price in sector i becomes more rigid. To understand why pass-through is complete when $\rho_z = 1$, recall equation (2.26): because the change in marginal cost is permanent, the marginal value of the extensive margin is lower not only today but also in all future periods. Consequently, the intertemporal substitution motive of the firm is irrelevant: with customers having a permanently lower value, the seller has no incentive to deviate from the full pass-through equilibrium to invest in its market share. However, as the shock becomes more temporary, the firm is less and less willing to maintain its profit margin intact because this would imply losing a customer base which will become valuable again very soon.

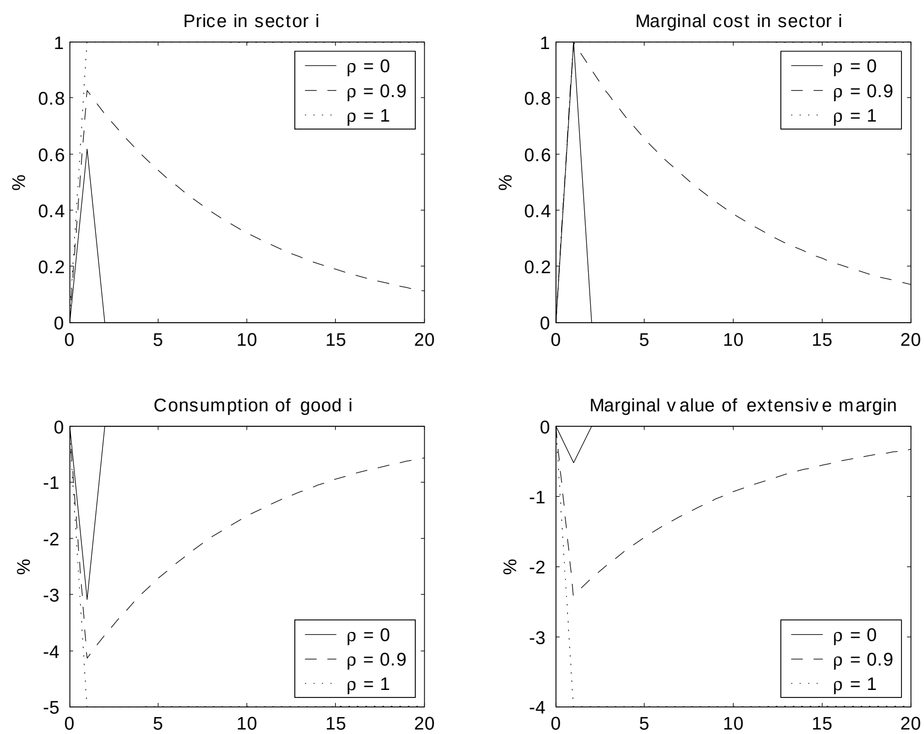

Figure 2.3: Dynamic response and persistence of the shock

2.5.2. Scope of the shock and cost pass-through

So far, we have focused on the response of prices to sector-specific marginal cost shocks. Under symmetry, all sellers of the same good behave similarly in equilibrium, and there is no price dispersion at the sectoral level.

Arguably, disturbances may also be firm-specific in nature. The complication in this case arises from the fact that we do not have a closed-form expression for the threshold switching cost $\widehat{\delta}_{it}(k)$, which will not be equal to 1 anymore as under the symmetric equilibrium. However, there is a special case for which we can solve for the optimal response: a temporary ($\rho_z = 0$) shock to the marginal cost of seller k around the symmetric equilibrium where all current and future prices within sector i are the same. Since we know that a price change by seller k does not affect agents' expectations of its future prices, the continuation values of V_0 and V_1 remain equal and we can solve for the threshold switching cost:

$$\widehat{\delta}_{it}(k) = \frac{p_{it}(k)}{p_{it}}.$$

This allows us to simulate the model, using the first-order conditions, (2.13)-(2.15), and the expression for the derivative of the threshold switching cost, (2.18). Unlike the symmetric case, the model here is not stationary: our assumption that the arrival rate of new customers is proportional to the size of the firm implies that the market share, $M_{it}(k)$, and the consumption level, $c_{it}(k)$, do not go back to their initial levels following a temporary shock. Accordingly, we rescale the model by dividing the equilibrium conditions by $M_{it-1}(k)$ and define the variables $\breve{c}_{it}(k) = c_{it}(k)/M_{it-1}(k)$ and $\breve{M}_{it}(k) = M_{it}(k)/M_{it-1}(k)$.

Figure 2.4 plots the response of a single seller to a 1% increase in its marginal cost, around the symmetric equilibrium. The atomistic nature of the seller implies that the sector- and aggregate-level variables are not affected by the shock. We also reproduce the sector-specific case for comparison purposes.

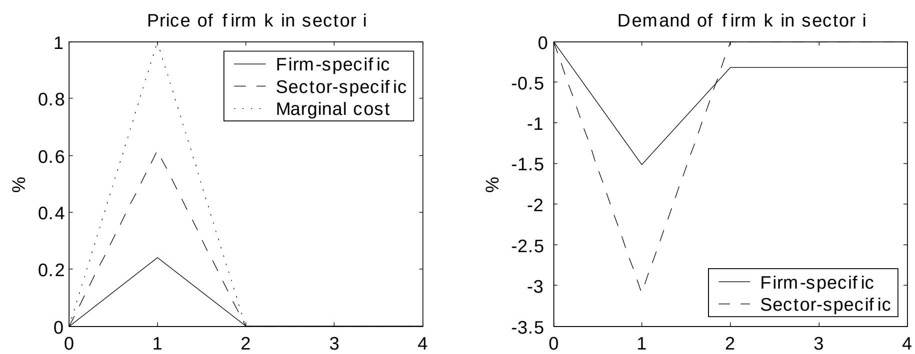

Figure 2.4: Firm- and sector-specific marginal cost shocks

Under the firm-specific shock, the degree of price rigidity is much higher: the seller passes-through only 24% of the rise in the marginal cost in its price, compared to 62% when the entire sector is hit. In the model, a temporary rise in the price results in a permanent fall in both the mass of customers and the demand. However, because the price change is so muted, the responses remain small: consumption falls by 1.5% in the period of the shock, and settles around 0.3% below its initial level.

Our finding is intuitive and sensible: when a seller is the only one hit by a rise in its marginal cost, it knows that its direct competitors have no incentive to raise their prices. Therefore, the firm is particularly wary of increasing its own price, for fear of losing a portion of its market share and future profits. Gron and Swenson (2000) find that this prediction holds in the context of the U.S. automobile market: they report that "measured pass-through is higher for cost shocks experienced by all products than for model-specific shocks".

It is important to clarify that the rigidity under firm-specific shocks is not only a consequence of the dynamic nature of the firms's problem, unlike the sector-specific case. Here, the real rigidity is both intratemporal and intertemporal: the fluctuation in $\widehat{\delta}_{it}(k)$ in itself discourages the firm from passing-through completely the rise in the marginal cost, even if it does not care about the future. However, it remains true that customer base dynamics significantly amplify the rigidity of prices: when $\beta = 0$, the pass-through rises from 24% to 61%.

Next we consider the case of a productivity shock which affects half the sectors in the economy. This requires solving for a general equilibrium version of the model, since the aggregate prices and quantities will now be affected by the shock. We set the other parameters of the model to values in line with those in the literature: $\sigma = 1, \epsilon = 2$ and $\eta = 1.2$. Figure 2.5 plots the price response following a 1% increase in the marginal cost for a typical sector. Not surprisingly, a shock which is common to a subset of sectors results in a higher marginal cost pass-through.

As we showed earlier, the incomplete marginal cost pass-through stems from intertemporal fluctuations in the value of the customer base coming from the intensive margin. What we find throughout our simulations is that as goods become better substitutes, that is as γ increases, firms become more reluctant to raise prices. This is because when goods are more substituable, an increase in the sectoral price relative to the aggregate price level results in a larger drop in the quantity consumed by each customer. In turn, the response

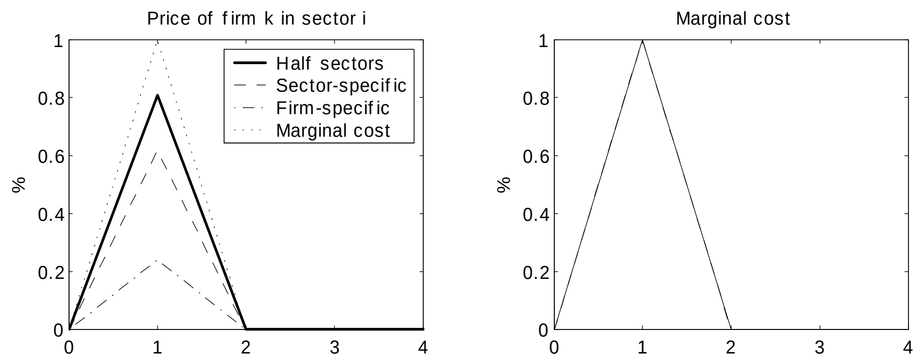

Figure 2.5: Marginal cost shock to half the sectors

of the value of the extensive margin is exacerbated, leading to a stronger response of the markup and a more muted price response.

The same rationale explains why our mechanism cannot in itself generate price rigidity in the wake of full aggregate shocks. If all sectors are hit at the exact same time and firms have full information, then they will fully pass-through marginal changes into their prices.[20] This is in fact a standard result for models based on real rigidities. To obtain price stickiness following a monetary expansion, one would need to interact our mechanism with nominal rigidities or make firms uncertain about the scope of the shock, an exercise which we do not pursue here.[21]

[20] In reality, under certain parameterizations the markup can be time-varying due to some general equilibrium effects, but we find the fluctuations to be very small. The properties under economy-wide shocks are briefly discussed in Appendix 4.

[21] The main objective of Ball and Romer (1990) was to study that interaction. They found that real rigidities could amplify the stickiness from mechanisms relying on nominal rigidities, such as menu costs. This would also be true in our setup, with an additional effect coming from the intertemporal dimension. For a recent treatment based on the Kimball (1995) aggregator, see Klenow and Willis (2006).

2.5.3. Switching costs and price rigidity

We present evidence in Section 2.2 that not only do firms point to customer-related factors as the main reason for keeping prices stable, but those concerns appear to be correlated with the degree of price rigidity. In our model, customer base dynamics arise because households face some costs of switching suppliers. Next, we describe how the marginal cost pass-through is affected by the distribution of the switching costs δ, and how this relates to the empirical evidence presented in Section 2.2. We focus on sector-specific shocks, but the results are qualitatively similar in the firm-specific case studied in the previous section.

The distributional assumptions impact the first-order conditions (2.19)-(2.21) in two ways: through the probability density function at the equilibrium relative price of 1, $f(1)$, and the taste-adjusted demand parameter $A(1)$. The latter plays only a very marginal role, and we do not discuss it further. It is easy to show from the law of motion of the market share (2.8) that the object $f(1)$ corresponds to the price elasticity of the customer base in equilibrium. As such, a change in the distribution of the switching costs will modify the model properties as long as it affects the value of $f(1)$.

To investigate the relation between price rigidity and switching costs, we simulate the price response to a sector-specific marginal cost shock ($\rho_z = 0.9$) under different values of μ_δ. A distribution mean of $\mu_\delta = 1$ indicates that, on average, there is no penalty to switching suppliers. Figure 2.6 reports the results along a few dimensions. The first plot offers a visual description of the relationship between the three objects we are interested in: as the average switching cost (μ_δ) increases, the elasticity of the customer base ($f(1)$)

falls and the proportion of repeat customers $(1 - F(1))$ rises.[22] The upper-right plot illustrates how the degree of pass-through is affected by the average size of the switching cost, μ_δ, while the lower graphs describe how price rigidity depends on $f(1)$ and $1 - F(1)$.

Consider an extreme case where switching costs are very high: this correponds to the rightmost distribution on the first plot. In such a scenario, the market share is inelastic $(f(1) = 0)$, customers are strongly attached to their current supplier, firms' market power and markup are high, there is no one switching in equilibrium and the model reverts to the standard Dixit-Stiglitz framework without an extensive margin. Consequently, firms completely pass-through any change of the sectoral marginal cost into their prices. As switching costs decrease, the elasticity of the customer base around the equilibrium rises and, not surprisingly, firms become more reluctant to pass-through cost changes. The surprising result, however, is that the relation is non-monotonic: while sector prices initially become more rigid, pass-through reaches a minimum around $\mu_\delta = 1.3$, or a very low elasticity value of $f(1) = 0.017$. After that point, and for most of the $f(1)$ parameter space, prices become *more* flexible as the elasticity of the market share rises. Notice that we are not considering extreme market share elasticities: in Figure 2.6, the fall in the mass of customers following a 1% rise in the relative price does not exceed 4%.

Our findings indicate that, somewhat paradoxically, firms are generally more willing to pass-through marginal cost fluctuations when their customer base is very sensitive to variations in the relative price. The reason behind this result is in fact intuitive. When the market share is highly elastic, customers are not loyal and can easily switch to other

[22] The link between μ_δ $f(1)$, and $F(1)$ is clear under a unimodal distribution for the switching costs such as the lognormal distribution we use here. However, the intuition is not obvious under bi-modal distributions, for example. Similarly, the case where switching costs are uniformly distributed would be uninteresting, given that $f(1)$ could then take only two extreme values.

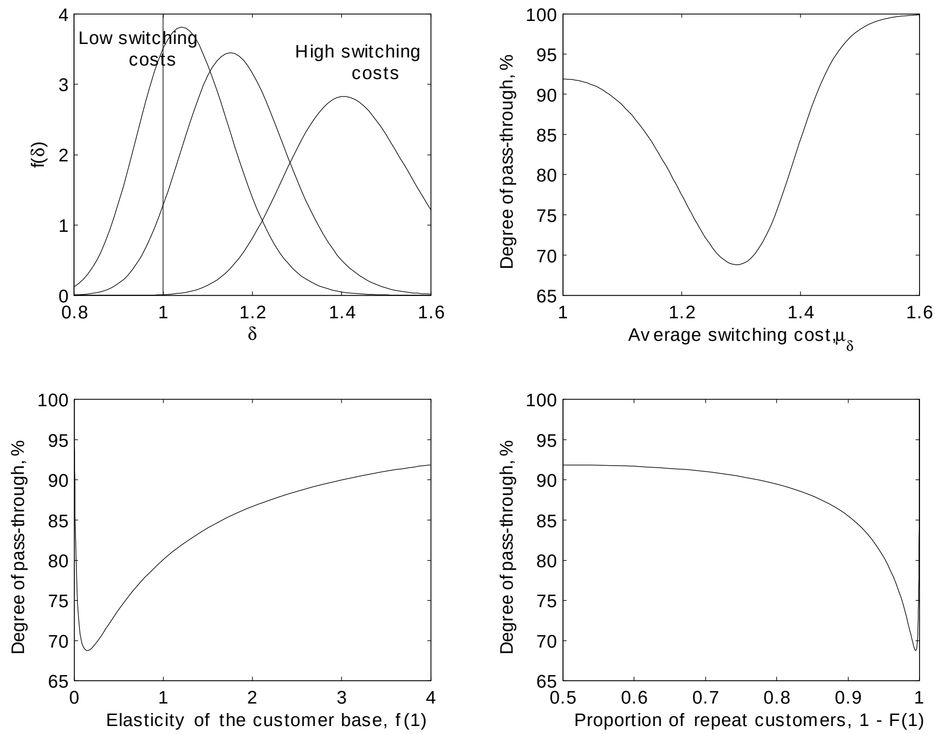

Figure 2.6: Pass-through and distribution of switching costs

suppliers. This translates into low market power for the firm and low equilibrum markups, a well-known result in the industrial organization literature (Klemperer (1995)). The customer base is therefore not valued as much by the seller, which explains the negative relation between the marginal value of the market share, v_i, and $f(1)$ (see the steady state equations (2.23)). The value of the customer base in turns interacts with the pricing decision of the firm: the seller will be less inclined to cut its profit margin to preserve its market share when the marginal value of customers is low. This is what we call the "loyalty effect", and its interaction with the "elasticity effect" determines the degree of price rigidity in the model. For very low values of $f(1)$, the elasticity effect is the most potent, while the loyalty effect quickly takes over as switching costs fall. This explains the non-monotonic relationship between switching cost and pass-through from Figure 2.6.

As an illustration, consider an industry in which individuals value highly a close and continuous business relationship with their provider and where switching is infrequent, such as in the case of barbers, for example. One would expect that the customer switching costs are relatively high in such sectors, since they incorporate "costs related to the loss of capitalized value of relationships previously established" (Kim, Kliger and Vale (2001)). Klemperer (1995) also notes that "markets for professional services of doctors, consultants, accountants, etc. involve switching costs of several, and perhaps all, kinds."

On the one hand, a barber facing a rise in his marginal cost might be more inclined to capitalize on his captive clientele by raising prices: since the elasticity of the customer base is low, any price deviation is not expected to affect the market share much. However, a price increase has the potential to encourage some valuable customers to switch, with the consequence of losing a significant revenue stream for the future. In the model, this second force generally dominates, and industries with less elastic customer base have more rigid prices. Now consider the opposite case: when switching costs are on average very small, customers often switch for exogenous reasons and the expected length of a match is short. There is, therefore, little incentive for the firm to sacrifice current profits in order to preserve its market share, since there is a high probability that any customer retained will switch supplier in the near future.

We also show in Section 2.2 that firms with a higher fraction of repeat customers empirically tend to have more rigid prices. We find a similar link in our model: the last plot in Figure 2.6 shows that the degree of marginal cost pass-though is a decreasing function of the proportion of repeat customers, $1-F(1)$, except for extremely high values. Fabiani *et al.* (2005) report for a set of European countries that the average proportion of repeat customer is 70%.

Interestingly, our mechanism can be linked to the notion of "customer lifetime value" or CLV, a popular concept in the recent marketing literature. According to Bauer, Hammerschmidt and Braehler (2003), "the CLV measures the profit streams of a customer across the entire customer life cycle". The authors notice that the economic reality "is marked by customer migration and a strong tendency to switch vendors" and highlight the key role of the retention rate, which "refers to the probability that an individual customer remains loyal to a particular supplier." In our model, the CLV corresponds to the value of an extra customer, $\bar{v}_{it}$, while the retention rate is closely related to the elasticity of the customer base, $f(1)$, and the proportion of repeat customers, $1-F(1)$. Our contribution is to show how fluctuations in the CLV, or $\bar{v}_{it}$, influence the firm's pricing decision.

We believe that the findings from this section can potentially explain why service prices are in general more rigid, particularly in sectors where buyer/seller relationships are important. Empirically, categories such as haircuts and beauty parlor services, legal services, home care, pet and veterinarian services, medical and dental services, etc. display the highest degree of price rigidity among all products (see Bils and Klenow (2004)).

2.6. Extensions

In this section we consider two extensions to the basic framework: learning about the persistence of the shock and allowing for an elastic arrival rate of new customers. We analyze how such modifications affect the properties of the model.

2.6.1. Learning about the persistence of the shock

It is conceivable that when firms experience a productivity shock, they find it difficult to recognize whether the nature of the shock is temporary or highly persistent. While in many models such a distinction is inconsequential or secondary, in our setup the *perceived* persistence of the shock is important for the dynamics of the price response. We now briefly illustrate how uncertainty about the persistence of the productivity shock affects the optimal pricing decision of a firm.

One possible way to model such uncertainty would be to let the persistence parameter in the marginal cost process be unknown, and allow Bayesian agents to learn about this parameter. Agents, endowed with prior beliefs about ρ, would then use Bayes' law to optimally update their beliefs and derive their posterior distribution for the persistence parameter. Here, however, we pursue a different approach to capture uncertainty. We model the learning process of the firm by using a simple linear Kalman filtering framework. Agents do not observe the true marginal cost shock. Instead, they observe a noisy signal from which they try to infer what is the true state of the underlying marginal cost process is. Using the Kalman filter algorithm, agents generate recursive forecasts of the true underlying state process.

Uncertainty in the model is captured explicitly in a standard way by the following two processes:

$$mc_t = mc_t^* + \varepsilon_t \ , \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right) \tag{2.27}$$

$$mc_t^* = \rho mc_{t-1}^* + v_t \ , v_t \sim N\left(0, \sigma_v^2\right) \tag{2.28}$$

where mc is observable by agents, while mc^* represents the true unobservable state of the underlying marginal cost process. In other words, v is a fundamental persistent shock, while ε can be described as noise. Here, the structure of the model is known by agents. More explicitly, ρ is known by all agents.[23] Equations (2.27) and (2.28) are the observation and measurement equations, respectively.[24]

For this exercise we set $\rho = 1$. Hence, ε shocks are purely temporary while a v shock has a permanent effect on the marginal cost. However, agents cannot distinguish between the source of the disturbance. The important elements of the filtering process applied to (2.27) and (2.28) can be summarized as follows:

$$mc^*_{t|t} = mc^*_{t|t-1} + K\left(mc_t - mc^*_{t-1|t-1}\right) \tag{2.29}$$

$$K = \frac{P}{P + \sigma^2_\varepsilon} \tag{2.30}$$

$$\sigma^2_v = \frac{P^2}{P + \sigma^2_\varepsilon} \tag{2.31}$$

where equation (2.29) indicates that agents update their beliefs about mc^*_t after they observe mc_t. As is common in the literature, in equation (2.30) we use the steady state level of the Kalman gain process, K. P is the steady state level of the variance of mc^*, which solves the Riccati equation given in (2.31).

We simulate our model under the maintained assumption of uncertainty about the nature of the shock. We use our benchmark parametrization described at the beginning of section 2.5.1, and analyze the responses under the following signal-to-noise ratio values:

---

[23]We assume that all the standard Kalman filter assumptions hold. For a general discussion, see, for example, Hamilton (1994, section 13.1).

[24]Equation (2.27) is only part of the observation system, since agents also observe other variables, aside from mc, such as c, l etc. This point is taken into account when solving for the optimal response in prices.

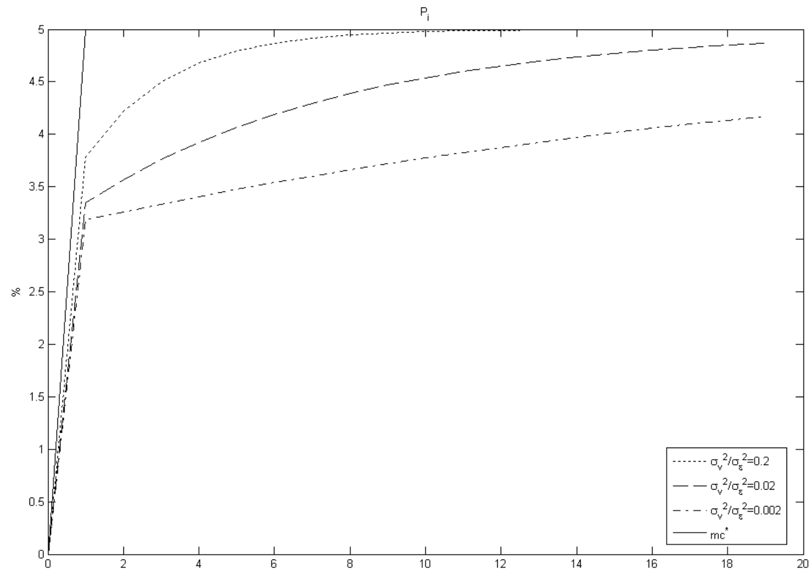

Figure 2.7: Permanent shock to marginal cost under shock uncertainty.

$\frac{\sigma_v^2}{\sigma_\varepsilon^2} = \{0.2, 0.02, 0.002\}$.[25] We simulate a 5% positive fundamental (i.e. permanent) shock to mc^* and present the reaction of p_i.

Under this setup, when the firm observes a jump in its marginal cost, it is unclear about its persistence. Since we assume that $\frac{\sigma_v^2}{\sigma_\varepsilon^2} < 1$, the firm initially puts a higher weight on the possibility that the shock is temporary, and accordingly only partially passes through the observed marginal cost increase into prices. However, as the firm continues period after period to observe a high level of mc, it updates its belief and eventually converges to full pass-through as it becomes more convinced that the shock is permanent. Not surprisingly, the response of prices is more delayed the higher the ratio of relative volatilities $\frac{\sigma_v^2}{\sigma_\varepsilon^2}$.

[25]The solution to (2.30) and (2.31) will depend solely on the signal-to-noise ratio parameter.

2.6.2. Attracting customers

So far, the pricing decision of the firm has no immediate impact on the arrival rate of customers: because there is no active search in the model, a lower relative price does not imply that a firm will be getting a larger share of switchers.[26] It is therefore of interest to know whether our results depend crucially on this assumption.

While adding search to the household problem would impair the tractability of the model, we can use a reduced-form specification which will alter the firm's problem by allowing for an elastic arrival rate. Let the probability of a given switcher being matched with seller k be defined by

$$\phi_{it}(k) = \frac{M_{it-1}(k)\,\phi\left(\frac{p_{it}(k)}{p_{it}}\right)}{\Phi_{it}}$$

where Φ_{it} is such that $\int_0^1 \phi_{it}(k)\,dk = 1$

$$\Phi_{it} = \int M_{it-1}(k)\,\phi\left(\frac{p_{it}(k)}{p_{it}}\right) dk.$$

Along with the condition that $\phi'\left(\frac{p_{it}(k)}{p_{it}}\right) < 0$, this modification simply states that a household is less likely to switch to a firm with a high relative price. Clearly, the firm's problem is affected. It is easy to show that the law of motion of the market share and the consumption from new customers are now respectively given by

[26] There is, however, an indirect effect through the size of the market share: a lower relative price today will imply a larger customer base next period, which in turn raises the inflow of new customers. This is a consequence of our assumption that the probability of a consumer being matched with seller k is proportional to $M_{it-1}(k)$.

$$M_{it}(k) = M_{it-1}(k)\left[1 - F\left(\widehat{\delta}_{it}(k)\right) + \frac{\phi\left(\frac{p_{it}(k)}{p_{it}}\right)}{\Phi_{it}}\int_0^1 M_{it-1}(l)\,F\left(\widehat{\delta}_{it}(l)\right)dl\right].$$

$$c_{it}^{N}(k) = \frac{M_{it-1}(k)\,\phi\left(\frac{p_{it}(k)}{p_{it}}\right)}{\Phi_{it}}\int_0^1 c_{it}^{L}(l)\,dl$$

where the expression for $c_{it}^{L}(l)$ is unchanged. When choosing its price $p_{it}(k)$, the firm will now also take into account the impact on the extensive margin of new customers. Therefore, the first-order condition with respect to $p_{it}(k)$ becomes

$$c_{it} = \lambda_{it}\left(\frac{p_{it}}{\widetilde{p}_t}\right)^{-\gamma}\frac{\widetilde{c}_t}{p_{it}}\left[\gamma A(1) - \phi'(1)\int_{\underline{\delta}}^{1}\delta^{1-\gamma}dF(\delta) + f(1)\right] + \frac{v_{it}}{p_{it}}\left[f(1) - \phi'(1)\,F(1)\right]$$

while all the other equilibrium conditions are unchanged.[27] The first graph in Figure 2.8 plots the relationship between the elasticity of the mass of repeat customers $f(1)$ and the degree of pass-through, for different values of $\phi'(1)$. All the other parameters are kept to their benchmark values. The thick solid line corresponds to our original specification, without active search by switchers ($\phi'(1) = 0$). Notice that while passthrough tends to increase as $\phi'(1)$ becomes larger (the mass of new customers is more responsive to the relative price), the effect is relatively small, except for very high values of the parameter. To understand this property, consider the expression for the elasticity of the mass of new customers

[27]Without loss of generality, we normalize $\phi(1) = 1$ which implies that $\Phi_{it} = 1$.

$$\varepsilon_{M_{it}^{N}(k),p_{it}(k)}\Big|_{s.e.} = \phi'(1)\,F(1)$$

Notice that it does not only depend on the function ϕ, but also the proportion of switchers $F(1)$: when there are only few customers switching every period, there is less potential market share gain from lowering the relative price. In the limit, as switching costs become very large, both $f(1)$ and $F(1)$ equal zero and the market share is inelastic. To illustrate this point, consider the second graph in Figure 2.8. There, we plot the degree of pass-through as a function of the proportion of repeat customers, for different values of $\phi'(1)$. While the effect of $\phi'(1)$ is significant for sectors where there is a lot of switching occuring, for more reasonable parameterizations the impact is quite small.

We reach two conclusions on the basis of those results: qualitatively, our findings do not hinge on the assumption of random matching; and the quantitative effects of incorporating active search are most likely small, at least for markets characterized by a certain significant of buyer/seller relationships.

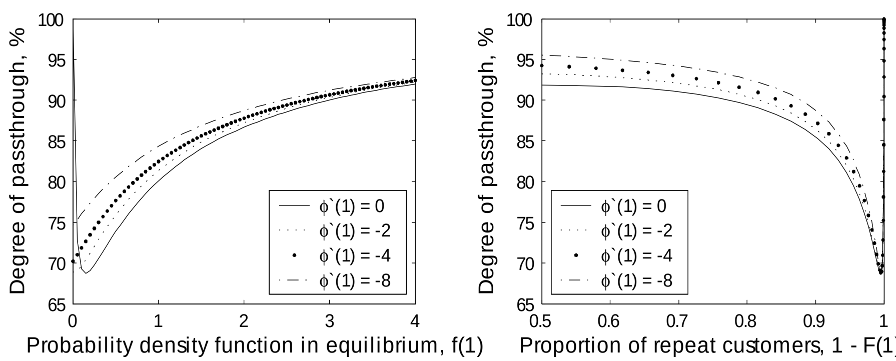

Figure 2.8: Degree of pass-through with elastic arrival rate of customers

2.7. An application: market share and exchange rate pass-through

> "*Although the dollar's exchange rate has been declining since early 2002, increasingly tight competitive conditions in the United States, as elsewhere, in 2002 and 2003 apparently induced exporters to the United States to hold dollar prices to competitive levels to ensure their market share and foothold in the world's largest economy*" Alan Greenspan, Advancing Enterprise 2005 Conference, London, February 4, 2005.

In this section, we turn our attention to what we believe is a natural application for our framework. Despite hedging and other measures, exporters' revenues are highly affected by fluctuations in the exchange rate, particularly when their sales abroad are denominated in the destination currency.[28] But even when firms are pricing in their own currency, they might be unwilling to let the relative price of their product fluctuate excessively in the foreign market because of competitive pressure from native firms.[29] Such concern appears to be a plausible mechanism behind the well-documented evidence that exporters do not fully pass-through exchange rate fluctuations into their prices.

For this application, we make abstraction of many side issues and consider a small open economy H who produces and exports goods, as well as imports from abroad. In particular, home consumers import a certain number of varieties from country F. We assume that the size of the H market is small relative to total sales by F firms, which means that profit fluctuations from this particular market will not influence significantly

[28]See, for example, Dominguez and Tesar (2006) for evidence that the firm value of exporters is affected by exchange rate fluctuations.

[29]See Mair (2005) for an analysis of surveys of exporters conducted by the Bank of Canada. Its main focus is to study the reaction of exporters to the recent appreciation of the Canadian dollars.

total F profits. Similarly, we assume that exports of F firms to country H are very small relative to the size of the home market: this ensures that any change in import prices from country F will not affect the general price level in H.

To simplify the exposition, let's define only two kinds of varieties available in country H: those which are sold by F firms and those produced by sellers from the rest of the world, including home producers. We will denote the second group simply as H. Therefore, for the purpose of our analysis, a household j from country H derives utility from a consumption basket c_t^j which is a combination of H ($c_{H,t}^j$) and F goods ($c_{F,t}^j$). The relevant aggregators are standard in the literature, except for the switching decision which is specific to our framework.

$$
\begin{aligned}
c_{H,t}^j &= \left\{ \int_0^1 \left[\left(\delta_{ht}^j\right)^{-s_{ht}^j} c_{ht}^j \right]^{\frac{\gamma-1}{\gamma}} dh \right\}^{\frac{\gamma}{\gamma-1}} \\
c_{F,t}^j &= \left\{ \int_0^1 \left[\left(\delta_{ft}^j\right)^{-s_{ft}^j} c_{ft}^j \right]^{\frac{\gamma-1}{\gamma}} df \right\}^{\frac{\gamma}{\gamma-1}} \\
c_t^j &= \left[\kappa_H^{\frac{1}{\omega}} \left(c_{H,t}^j\right)^{\frac{\omega-1}{\omega}} + \kappa_F^{\frac{1}{\omega}} \left(c_{F,t}^j\right)^{\frac{\omega-1}{\omega}} \right]^{\frac{\omega}{\omega-1}}
\end{aligned}
$$

The parameter γ denotes the elasticity of subsitution across varieties produced by a particular country, while ω is the elasticity of substitution between H and F varieties. The weights κ_H and κ_F will dictate the relative importance of H and F goods in the consumption basket of a typical home household. Consistent with our earlier model, each F variety is sold by a continuum of unit mass of foreign sellers producing homogenous brands, and similarly for H varieties.[30] The decision problem of household j in country H is very similar to the one from Section 2.3:

$$\max U_t^j = \sum_{\tau=t}^{\infty} \beta^{\tau-t} u\left(c_\tau^j, l_\tau^j\right) \tag{2.32}$$

subject to

$$u\left(c_t^j\right) = \frac{1}{1-\sigma}\left(c_t^j\right)^{1-\sigma} - \eta \frac{\left(l_\tau^j\right)^{1+\epsilon}}{1+\epsilon} \tag{2.33}$$

$$\int_0^1 p_{ht}^j(k)\, c_{ht}^j dh + \int_0^1 p_{ft}^j(k)\, c_{ft}^j df + E_t r_{t+1} b_{t+1}^j = b_t^j + w_t l_t^j + \Pi_t^j$$

The first-order conditions of the household problem are similar to the ones we derived in Section 2.3.1, and we will not repeat them here. Aggregating across customers, the demand function faced by foreign seller k producing good f is given by

$$c_{ft}(k) = \kappa_F A\left(\widehat{\delta}_{ft}(k)\right)\left(\frac{p_{ft}(k)}{\widetilde{p}_{Ft}}\right)^{-\gamma}\left(\frac{\widetilde{p}_{Ft}}{\widetilde{p}_t}\right)^{-\omega} \widetilde{c}_t \tag{2.34}$$

where

$$A\left(\widehat{\delta}_{ft}(k)\right) = M_{ft-1}(k)\left[\int_{\underline{\delta}}^1 \delta^{1-\gamma} dF + 1 - F\left(\widehat{\delta}_{ft}(k)\right)\right]$$

$\widetilde{p}_{Ft}$ is the price index for the basket of F goods, while $\widetilde{p}_t$ and $\widetilde{c}_t$ are the aggregate price and consumption levels in country H respectively. The expressions for the price indexes in the symmetric equilibrium are

[30] This assumption is necessary since our environment requires symmetry across suppliers within a particular sector.

$$
\begin{aligned}
\widetilde{p}_{Ft} &= A(1)^{\frac{1}{1-\gamma}} p_{ft} \\
\widetilde{p}_{Ht} &= A(1)^{\frac{1}{1-\gamma}} p_{ht} \\
\widetilde{p}_{t} &= \left[\kappa_H \left(\widetilde{p}_{Ht}\right)^{1-\omega} + \kappa_F \left(\widetilde{p}_{Ft}\right)^{1-\omega}\right]^{\frac{1}{1-\omega}}
\end{aligned}
$$

where the derivations for $\widetilde{p}_{Ft}$ and $\widetilde{p}_{Ht}$ are similar to those in Appendix 2, and $\widetilde{p}_t$ follows

directly from our aggregator.

Consider the decision problem of a typical F firm. It sells domestically and engages in exports activities, and we assume that profits from these activities are separable. This means that we can focus solely on the firms' export decision to country H, since only the bilateral exchange rate between F and H will be affected in our simulations. The decision problem of a foreign firm (f,k) selling in the home market is therefore[31]

$$
\max_{\{p_{ft}(k)\}} \sum_{t=0}^{\infty} \beta^t E_0 \mu_t^* \left\{ c_{ft}(k) \left[e_t p_{ft}(k) - \frac{w_t^*}{z_{ft}^*} \right] \right\}
$$

subject to the demand schedule (2.34) and the law of motion for the market share. $p_{ft}(k)$ is expressed in the currency of the home market, and e_t is the exchange rage expressed as units of foreign currency per home currency. The law of motion of the exchange rate is given by $\ln(e_t) = \rho \ln(e_{t-1}) + \varepsilon_t$. The other variables have been described in Section 2.3, and asterisks denote their foreign counterparts.

[31] Once again, notice that we abstract from the pricing decision in the F market, simply because the firm's return function is additive across markets. Given the nature of our assumptions, that dimension of the producer's problem will not be affected in our simulations.

The experiment consists in an exogenous temporary depreciation of the home currency (a negative realization of ε_t). We impose that κ_F is very small which, in conjunction with our earlier assumptions about the relative market sizes, ensures that the aggregate economies of H and F are, at a first approximation, unaffected by shocks to the bilateral exchange rate, and that fluctuations in $\widetilde{p}_{Ft}$ and $\widetilde{c}_{Ft}$ have no significant impact on $\widetilde{p}_t$ and $\widetilde{c}_t$. Consequently, the only variables which will not be constant over time in our simulations are e_t, p_{ft}, c_{ft}, $\widetilde{p}_{Ft}$ and $\widetilde{c}_{Ft}$. In particular, the foreign variables w_t^*, μ_t^* and z_{ft}^* will be held constant throughout the exercise. This allows us to focus on the problem faced by the foreign exporter following the exchange rate shock.

Clearly, the first-order conditions to the firm's problem are very similar to the ones we derived in Section 2.3, and we will not repeat them for the sake of conciseness. What is important to realize is that the exchange rate e_t will enter our system of equations and affect the pricing decision.

Figure 2.9 shows what happens to the foreign exporters following a 5% temporary depreciation of the home currency.[32] Remember that these firms are facing competition from home sellers producing competing varieties. In a standard framework without customer flows and constant elasticity of demand, the optimal strategy of the foreign exporter would be to raise export prices proportionately in order to maintain its profit margin. As we have already seen earlier, in our setup this is not the best strategy: when faced by negative fluctuations in the exchange rate, the exporter is reluctant to fully raise prices. The reason is intuitive: while increasing the price allows to improve the markup, it also

[32]For this simulation, we use the following parameter values: $\beta = 0.99$, $\gamma = 5$, $\omega = 3$, $\mu_{\ln \delta} = 0.15$, $\sigma_{\ln \delta} = 0.1$, $\rho = 0.8$.

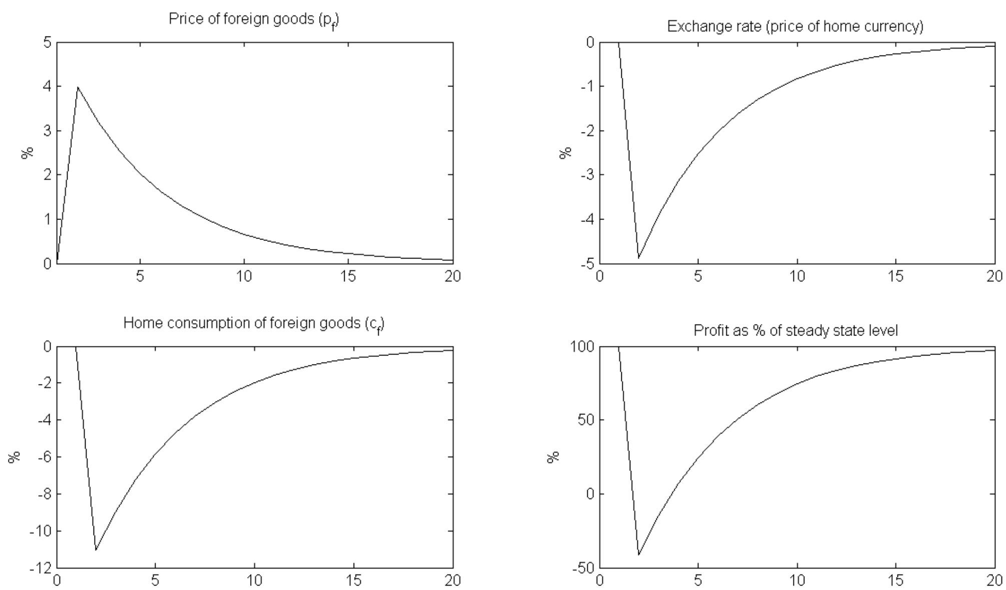

Figure 2.9: Response to exchange rate shock in a small open economy

involves the risk of losing customers who will be hard to regain later, exactly when the exchange rate will have realigned itself and revenues per customer will be higher.

Our customer flow mechanism therefore delivers incomplete exchange rate pass-through, in line with the extensive empirical evidence on the subject. While the literature generally imposes nominal rigidities or demand functions with non-constant elasticity to achieve this result (e.g. Devereux, 2003, and Burstein et al., 2005), our model delivers incomplete pass-through as an endogenous outcome: exporters are reluctant to pass-through completely exchange rate changes to their prices because they care about their market share position.[33]

[33]Atkeson and Burstein (2005) obtain a similar result in a game-theoretic, partial equilibrium framework.

If firms live in an environment where the exchange rate process is very noisy, they are likely to initially put a high weight on the possibility that an exchange rate shock is temporary and only slowly update their beliefs about the nature of the shock. Similar to the results from Section 2.6.1, our model would then deliver a degree of pass-through which is incomplete in the short run, but complete at longer horizons. This would be consistent with the evidence from international prices (see Campa and Goldberg, 2002), and related to the intuition of Froot and Klemperer (1989). Finally, our model can easily be embedded in a general equilibrium open-economy model, as it builds on the standard framework used in the literature.

2.8. Conclusion

In this section we show that a standard macro model in which firms and households form long-term relationships can deliver incomplete pass-through of sector-specific cost shocks. In addition, we find that the degree of price rigidity is inversely related to the persistence of the shock, and that cost pass-through is lower in the case of firm-level disturbances. We also show that in an environment where temporary sector-specific shocks are predominant, sellers will initially respond to a change in their marginal cost by only partially raising prices, and slowly revise their strategy as they update their priors. This learning mechanism naturally leads to a delayed response of prices to persistent shocks. Given that firms view customer-related factors as the main impediments to having more flexible prices, we believe that our mechanism offers a sensible and interesting theory of price rigidity.

In our model, customer base dynamics arise because consumers face costs of switching to a different supplier. We show that cost pass-through is a non-monotonic function of the size of switching costs, and that prices tend to become more stable as the fraction of repeat customers increases and the elasticity of the customer base falls. Based on surveys of firms' pricing behavior, we argue that those results are in line with the empirical evidence and that they offer a potential explanation for some of the heterogeneity in price rigidity observed in the data. However, more empirical work needs to be done on the subject. As switching costs are difficult to quantify, it might prove easier to find a measure of repeated interaction between sellers and customers, and relate it to the degree of price rigidity. Additional firm surveys could also provide useful information.

Another research avenue would be to interact the real rigidity we propose with some nominal rigidities. While the evidence suggests that menu costs *per se* do not play a major role in firms' pricing decisions, it is conceivable that a small amount of menu costs coupled with the mechanism we propose could produce significant price stickiness and real effects from monetary shocks. Also, modelling firm uncertainty about the scope of the shock could be an interesting way to obtain price rigidity following aggregate shocks: if most shocks are idiosyncratic in nature, firms may be slow to recognize a shock as aggregate. Given that prices are rigid under sector- and firm-specific shocks, we conjecture that a combination of our mechanism with Lucas-style imperfect information would give rise to real effects from monetary shocks.

A natural application of our model is in the context of international economics: in markets where both domestic and foreign firms compete, movements in the real exchange rate create a wedge between their marginal costs expressed in a common currency. In our framework, exporters find it optimal to pass-through only a fraction of exchange rate fluctuations in order to stabilize their market share. This prediction is supported by the large empirical literature on exchange rate pass-through.

CHAPTER 3

Robust Equilibrium Yield Curves (joint with Nicolas Vincent)

3.1. Introduction

This paper studies the implications of the interaction between robust control and stochastic volatility for key asset pricing phenomena. We quantitatively show that robustness, or fear of model misspecification, coupled with state-dependent volatility provides an empirically plausible characterization of the level and volatility of the equity premium, the risk free rate, and the cross-section of yields on treasury bonds. We also show that robustness offers a novel way of reconciling the shape of the term structure of interest rates with the persistence of yields. Finally, we quantify the level of robustness encoded in agents' behavior.

We construct a continuous-time, Lucas (1978)-type, asset pricing model in which a representative agent is averse to both risk and ambiguity. The presence of ambiguity stems from the agent's incomplete information about the economy's data generating process (DGP). In other words, the agent does not know which of several possible models is the true representation of the economy. Introducing ambiguity aversion into our framework allows us to reinterpret an important fraction of the market price of risk as the market price of model (or Knightian) uncertainty. We model ambiguity aversion using robust control techniques as in Anderson et al. (2003).[1] In our model, the representative agent

[1]Behavioral puzzles such as the Ellsberg paradox (Ellsberg (1961)) led to the axiomatization of the maxmin decision making by Gilboa and Schmeidler (1989). Robust control is one way of modeling Knightian uncertainty. For a comprehensive treatment of robustness see Hansen and Sargent (2007a).

distrusts the reference model and optimally chooses a distorted DGP. His consumption and portfolio decisions are then based on this distorted distribution. Ambiguity aversion gives rise to endogenous pessimistic assessments of the future.

A key assumption in the model is that the output growth process is conditionally heteroskedastic. The consumption growth process inherits this heteroskedasticity, which gives rise to a stochastic market price of risk.[2] The main contribution of this paper is to show that ambiguity aversion greatly amplifies the effect of stochastic volatility in consumption growth and, therefore, can explain asset prices in an empirically plausible way. In the absence of ambiguity aversion, plausible degrees of stochastic volatility in consumption growth do not generate sufficient variation in the stochastic discount factor.

By choosing a distorted DGP, the robust representative agent has biased expectations of future consumption growth. Being pessimistic, the agent tilts his subjective distribution towards states in which marginal utility is high. With stochastic volatility, positive volatility surges result in a more diffuse distribution of future consumption growth. In that case, the objective distribution assigns more probability mass to future 'bad' realizations of consumption growth. The agent seeks policies that can reasonably guard against such 'bad' realizations. Consequently, he increases the distortion to his expectations of

Examples of the use of robust control in economics and finance include Anderson et al. (2003), Cagetti et al. (2002), Gagliardini et al. (2004), Hansen and Sargent (2007b), Hansen et al. (2006), Liu et al. (2005), Maenhout (2004), Routledge and Zin (2001), Uppal and Wang (2003). An alternative approach to modeling ambiguity allows agents to have multiple priors. See, for example, Epstein and Schneider (2003), Epstein and Wang (1994).

[2]Recently, several authors (e.g., Bansal and Yaron (2004), Bansal et al. (2005)) argue that conditionally heteroskedastic consumption growth is potentially important to understand asset prices. We agree that this channel is indeed important, but claim that it is the interaction with ambiguity aversion that is critical.

consumption growth. The interaction between robustness and stochastic volatility introduces a state dependent distortion to the drift of consumption growth, and therefore, to the drift in the agent's intertemporal marginal rate of substitution.[3] This state dependent distortion generates sharp implications for asset pricing phenomena.

We estimate our model and assess its implications using data from the equity and bond markets, as well as consumption data. We exploit cross-equation restrictions across bond and equity markets to improve both the identification of structural parameters in our model and the estimation of the market price of risk and uncertainty.[4] Our main findings are as follows.

First, we show that our model, calibrated with a unitary degree of risk aversion and elasticity of intertemporal substitution (EIS), can reproduce both the high and volatile equity premium and the low and stable risk free rate observed in the data. Previous studies, such as Mehra and Prescott (1985) and Weil (1989), show that explaining the behavior of the equity premium requires implausibly high levels of risk aversion. Ambiguity aversion generates an uncertainty premium that helps to alleviate the difficulties encountered in these previous studies. Since there is no benchmark value for the degree of ambiguity aversion, we use detection error probabilities to show that the degree of robustness required to fit the data is reasonable. In other words, we show empirically that

[3]In the next section and in the empirical section of the paper we provide an additional, more technical, explanation for the link between stochastic volatility and the robust distortion by using the properties of the relative entropy and the size of the set of models the agent entertains.

[4]Campbell (2000) convicingly argues that "it is important to reconcile the characterization of the SDF provided by bond market data with the evidence from stock market data. Term structure models of the SDF are ultimately unsatisfactory unless they can be related to the deeper general-equilibrium structure of the economy. Researchers often calibrate equilibrium models to fit stock market data alone, but this is a mistake because bonds carry equally useful information about the SDF. The short-term real interest rate is closely related to the conditional expected SDF and thus to the expected growth rate of marginal utility; in a representative-agent model with power utility of consumption, this links the real interest rate to expected consumption growth...The risk premium on long-term bonds is also informative."

the set of models the robust representative agent entertains is small. By this we mean that it is statistically difficult to distinguish between models in this set.

Second, our model can account for the means of the cross-section of bond yields. In particular, we can replicate the upward sloping unconditional yield curve observed in the data. This result highlights a novel interpretation of the uncertainty premium generated by robustness. On empirical grounds, we assume that the conditional variance of output growth, and hence consumption growth, is stationary and positively correlated with the consumption growth process. This positive correlation implies that when marginal utility is high the conditional variance of consumption growth is low. Consequently, a downward bias in the subjective conditional expectations of consumption growth induces a negative distortion to the subjective expectations of variance changes. We show that this negative distortion is a linear function of the level of the conditional variance process. Consequently, the unconditional distortion is a linear negative function of the objective steady state of the variance process. Therefore, the subjective steady state of the variance process is lower than the objective steady state. In other words, on average, the agent thinks that the conditional variance of consumption growth should decrease. Since the unconditional level of bond yields and the steady state level of the conditional volatility of consumption growth are inversely related, the agent expects, on average, that yields will increase. Consequently, the unconditional yield curve is upward sloping.

Third, our model can replicate the declining term structure of unconditional volatilities of real yields, and the negative correlation between the level and the spread of the real yield curve. The fact that the robust distortion to the conditional variance process is a linear function of the level of the variance implies that the distorted process retains the mean reversion structure of the objective process. Since shocks to the conditional variance are transitory, the short end of the yield curve is more responsive to volatility shocks relative to the long end. Hence, short maturity yields are more volatile than long maturity yields. Also, our model implies that yields are an affine function of the conditional variance of consumption growth. Therefore, all yields are perfectly positively correlated. Short yields are more responsive to volatility shocks than long yields, but both move in the same direction. So, when yields decrease, the spread between long yields and short yields increases and becomes more positive. As a result, the level and spread of the real yield curve are negatively correlated.

Fourth, the model can reconcile two seemingly contradictory bond market regularities: the strong concavity of the short end of the yield curve and the high degree of serial correlation in bond yields.[5] The intuition for this result is closely linked to the mechanism behind the upward sloping real yield curve. Generally, in a one-factor affine term structure model, the serial correlation of yields is driven by the serial correlation of the state variable implied by the objective DGP. In contrast, in our model the slope of the yield curve is shaped by the degree of mean reversion of the conditional variance process implied by the agent's distorted (i.e., subjective) distribution. The state dependent distortion to the variance process not only changes the perceived steady state of the variance but also its velocity of reversion. With positive correlation between consumption growth and the conditional variance process, we show that the subjective mean reversion is faster than the objective one. Ex ante, the agent expects shocks to the variance process to die out fast, but ex-post these shocks have a longer lasting effect than expected. The slope of the

[5]In a standard one-factor model, it is difficult to separate these two properties, since both observations are directly tied to the persistence of the underlying univariate shock process.

yield curve is a reflection of how fast the agent expects the effect of variance surges to dissipate. The positive slope of the yield curve declines rapidly when the subjective mean reversion is high. The serial correlation of yields is measured ex-post, using realized yields. If the objective persistence of the variance process is high, yields are highly persistent, which is in line with the empirical evidence. We also show that when the agent seeks more robustness, the separation between the ex-ante and ex-post persistence is stronger.

The remainder of the paper is organized as follows. In section 3.2 we present the robust control idea in a simple two-period asset pricing model. The main purpose of this section is to highlight the channels through which uncertainty aversion considerations alter the predictions of a standard asset pricing model. In section 3.3 we present our continuous time model and discuss its implications for the equity market's valuation patterns and the implied risk free rate. In section 3.4 we discuss the model's predictions concerning the bond market. We derive analytical affine term structure pricing rules and discuss the distinction between the market price of risk and uncertainty. In section 3.5 we present empirical evidence that supports our modeling assumptions. We also estimate our complete model and investigate the implied level of uncertainty aversion exhibited by the representative agent. In section 3.6 we offer our concluding remarks and discuss potential avenues for future research.

3.2. Robustness in a Two-Period Example

In this section we introduce the terminology and concepts used throughout the paper using a two-period consumption-saving example. This example helps build our intuition and motivate the modeling assumptions used in our model.

3.2.1. Reference and Distorted Models

The representative agent in our economy uses a *reference* or *approximating* model. However, since he fears that this model is potentially misspecified, he chooses to diverge from it when making his decisions.[6] In the context of this paper, the reference model is assumed to generate the observed data. In contrast with the rational expectations paradigm, the agent entertains alternative DGPs. The size of the set of possible models is implicitly defined by a penalty function (relative entropy) incorporated into the agent's utility function. So the agent chooses an optimal distorted distribution for the exogenous processes. In other words, the agent optimally chooses his set of beliefs simultaneously with the usual consumption and investment decisions. The robust agent distorts the approximating model in a way that allows him to incorporate fear of model misspecification. We will refer to the optimally chosen model as the *distorted* model.[7]

3.2.2. A Two-Period Model

We now discuss a simple two-period example. Our discussion is intentionally informal. Our goal is to illustrate how robustness considerations alter the predictions of a standard asset pricing model. We consider a Lucas-tree type economy in which the agent receives one unit of consumption good in the first period. He decides how many units (α) of a

[6] Another possibility is to claim that for some reason the agent dislikes extreme negative events and wants to take special precautionary measures against these events. If we choose this behavioral interpretation, we can then assume the agent knows the true DGP, but that his marginal utility function is very high in bad states of the world. Low consumption is so costly that the agent requires policies that are robust to these states. Even though there is complete observational equivalence between the two approaches, they are utterly different from a behavioral perspective.

[7] An alternative is to allow for the possibility that a different, unspecified model, is actually the DGP. In this scenario, it is likely that neither the distorted nor the reference model generate the data. The agent must in this case infer which model is more likely to generate the data. See Hansen and Sargent (2007b) for an example.

claim to the stochastic endowment in period one (D_1) to buy. The unit price of a claim to the tree's output is denoted by S_0. We assume that period one output is drawn from a lognormal distribution, which we refer to as $\mathbb{P}$:

$$\ln D_1 \sim N\left(\mu, \sigma^2\right).$$

We denote the agent's subjective distribution by $\mathbb{Q}$. The robust agent solves a max-min problem, where the minimization takes place over $\mathbb{Q}$:[8]

$$\max_{\alpha} \min_{\mathbb{Q}} \left\{u\left(C_0\right) + \beta \mathbb{E}^{\mathbb{Q}}\left[u\left(C_1\right) + \theta \mathcal{R}\left(\mathbb{Q}\right)\right]\right\} \tag{3.1}$$

subject to:

$$C_0 = 1 - \alpha S_0; \quad C_1 = \alpha D_1.$$

Here C_0 and C_1 are the levels of consumption in periods zero and one, respectively. The object $\mathcal{R}\left(\mathbb{Q}\right)$ represents the penalty imposed on the agent whenever he decides to choose a distribution different from $\mathbb{P}$. We assume that this penalty is the relative entropy, or Kullback-Leibler divergence, between the objective ($\mathbb{P}$) and subjective ($\mathbb{Q}$) distributions. The parameter θ is a multiplier which determines the sensitivity of the agent's value function to the relative entropy. Without this penalty, the minimization problem would have a boundary solution in which the agent assigns all the probability mass to the worst possible state (if the support is the entire real line, the agent distorts the mean by setting it to negative infinity). Note also that, when $\theta \to \infty$, problem (3.1) converges to the

[8]This preference specification is referred to in the literature as 'multiplier preferences'. The decision theoretic foundation for the use of multiplier preferences is discussed in Maccheroni et al. (2006) and Strzalecki (2007). These authors also discuss the interpretation of the parameter θ as a measure of the level of ambiguity aversion which the agent exhibits.

conventional time-additive expected utility case. In this case, the penalty for distorting the objective distribution is so large that the agent optimally decides to construct his beliefs using the objective measure $\mathbb{P}$ ($\mathcal{R}(\mathbb{Q}) = 0$ when $\mathbb{P} = \mathbb{Q}$).

In continuous time, one can show that given a conditional normal distribution for the growth rate of output, the distorted distribution is also normal with the same variance and a lower mean. This result follows from the regularity conditions (i.e., absolute continuity) required when using relative entropy.[9] Since our complete model is cast in continuous time, we consider only mean distortions in this two-period example in order to make the transition to the full model more transparent. The following lemma characterizes more generally the distortions that the agent considers in discrete time when the reference distribution is a univariate Normal:

Lemma 1. *Consider the class of Normal distributions. If* $u(C) = \ln(C)$ *then the agent chooses* $\mu_{\mathbb{Q}} = \mu + h$ *and* $\sigma^2_{\mathbb{Q}} = \sigma^2$*. If* $u(C) = C^{1-\gamma}/(1-\gamma), \gamma \neq 1$ *then the agent chooses* $\mu_{\mathbb{Q}} = f(\mu, \sigma^2)$ *and* $\sigma^2_{\mathbb{Q}} = g(\mu, \sigma^2)$ *for some* $f, g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *such that if* $\mu > \mu_{\mathbb{Q}}$ *then* $\frac{\partial \sigma^2_{\mathbb{Q}}}{\partial(\mu - \mu_{\mathbb{Q}})} > 0$

Proof. See Appendix 6. □

We assume that utility is logarithmic. Lemma 1 shows that, in this case, it is optimal for a robust agent to distort only the mean of the distribution. An agent with logarithmic utility derives no benefit from distorting the variance, since he cares only about the first

[9]Relative entropy and Radon-Nikodym are ultimately likelihood ratios. In continuous time, likelihood ratios are extremely sensitive to variance distortions. This is because high frequency observations allow us to estimate the variance fairly accurately. So, the likelihood ratio reveals the difference between models with different variances easily. In the robust control formulation, this imposes a large penalty on the agent. Thus, he optimally chooses not to distort the variance.

moment of the distribution. But, he incurs a cost since an increase in variance raises the relative entropy of the two distributions. In contrast, in the non-logarithmic case ($\gamma \neq 1$), the agent also cares about the second moment of the distribution and therefore distorts the variance. We see that $\partial\sigma_{\mathbb{Q}}^2/\partial\left(\mu - \mu_{\mathbb{Q}}\right) > 0$ since the 'distance' between the distributions is positively related to mean distortions. However, for a given mean distortion, the agent needs to distort the variance in order to maintain a desired distance between the distributions.

Let $\mathbb{Q}$ be $N\left(\mu + h, \sigma^2\right)$, where $h \in \mathbb{R}$ represents the mean distortion chosen by the robust agent. The relative entropy of $\mathbb{P}$ and $\mathbb{Q}$ is given by,

$$\begin{aligned} \mathcal{R}\left(\mathbb{Q}\right) &\equiv \int \ln\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right) d\mathbb{Q}, \\ &= \frac{h^2}{2\sigma^2}. \end{aligned} \tag{3.2}$$

Not surprisingly, the divergence between $\mathbb{P}$ and $\mathbb{Q}$ is a positive function of the distortion to the mean, h. The presence of the variance in the denominator reflects the fact that the distortion in the mean must be measured relative to the degree of volatility associated with the distribution.

Assuming that $u\left(C\right) = \ln C$ and using (3.2) we can rewrite (3.1) as:

$$\max_{\alpha} \min_{h} \left\{ \ln\left(1 - \alpha S_0\right) + \beta \mathbb{E}^{\mathbb{Q}}\left[\ln\left(\alpha D_1\right) + \frac{\theta h^2}{2\sigma^2}\right]\right\}.$$

Note that now the minimization problem is taken over h which serves as a sufficient statistic for the divergence between $\mathbb{P}$ and $\mathbb{Q}$. The first-order condition with respect to h

yields:

$$h = -\frac{\sigma^2}{\theta} \leq 0.$$

This value of h represents the distortion to the mean of the conditional distribution of next period's output. As the penalty parameter θ becomes smaller, the agent seeks more robust policies and the (absolute) size of the distortion increases. A result that is particularly important in our context is that the distortion becomes more pronounced when output growth is expected to be more volatile. The intuition for this result is rather simple: a robust agent is more prone to take precautionary measures against misspecification when bad outcomes are more likely.

Since h is independent of the other controls (i.e., neither consumption nor the investment policy affect the choice of h), the maximization with respect to α and the imposition of the equilibrium condition ($\alpha = 1$) yield the usual pricing formula:

$$S_0 = \frac{\beta}{1+\beta}.$$

The maximization also yields the usual decision rule for consumption. The agent consumes a fraction of his wealth that is independent of the price of the risky asset:

$$C_0 = \frac{1}{1+\beta}.$$

Note that, up to this point, the robust considerations do not alter any of the model's predictions: both consumption and the price of the risky asset are unaffected by the choice of h. To understand this result, note that the optimal level of investment in the risky asset is affected by two forces. First, the robust agent fears exposing his capital to adverse shocks to the return on the risky asset. Second, the agent fears a bad output growth realization that will leave him 'hungry' next period and prefers to save. When momentary utility is logarithmic, these two effects cancel out.[10]

The implications of robustness in our setup take the form of a precautionary savings motive. To see this, first consider the Euler equation used to price a risk free one-period bond which is in equilibrium in zero net supply,

$$1 = \beta \mathbb{E}^{\mathbb{Q}} \frac{1/D_1}{(1+\beta)} \exp(r), \tag{3.3}$$

where r is the continuously compounded risk-free rate. Using our distributional assumptions, one can show that:

$$r = \ln \frac{1+\beta}{\beta} + \mu - \sigma^2 \left(\frac{1}{2} + \frac{1}{\theta} \right).$$

We see that robustness affects the risk free rate: as θ decreases and the agent becomes more robust, there is downward pressure on r due to an increased precautionary savings motive. Notice that this effect is independent of the degree of the EIS: the increased precautionary savings motive allows us to derive a low risk free rate despite a unitary EIS. There is no upwards pressure on the risk free rate due to the usual substitution effect. Note also that when the agent does not seek robust policies ($\theta \to \infty$), we are back to the expected additive utility case.

We now discuss how robustness considerations can alter the model's predictions regarding the equity premium. In this two-period example, the continuously compounded

[10] Miao (2004) also discusses a similar example.

return on the risky asset is given by $\ln(D_1/S_0)$ and the observed equity premium is simply:

$$\mathbb{E}\ln\frac{D_1}{S_0} - r = \sigma^2\left(1+\frac{1}{\theta}\right).$$

This expression shows that it is theoretically possible to generate a high equity premium and a low risk free rate when the robustness parameter, θ, is low enough. This result will be later confirmed and quantified in the context of our complete model.[11]

While the two-period example provides interesting insights into the implications of robustness, it does not allow us to study the term structure of interest rates, which is the main objective of this study. In the next section, we describe a continuous time, infinite horizon model in which we not only embed robustness considerations but also enrich the environment to allow for a time-varying investment opportunity set.

3.3. Robustness in a Continuous Time Model with Stochastic Volatility

In this section we present an infinite horizon, continuous time, general equilibrium model in which a robust representative agent derives optimal policies about consumption

[11]Standard models with time additive expected utility violate the Hansen-Jagannathan (HJ) bound (Hansen and Jagannathan (1991)). Examining the HJ bound can shed more light on how robustness modifies a standard asset pricing framework. The robust HJ bound needs to be modified. This bound takes the following form $\underbrace{\frac{\sigma(m)}{\mathbb{E}^{\mathbb{Q}}(m)} - \frac{h}{\sigma(R^e)}}_{\text{Model implied}} \geq \underbrace{\frac{|\mathbb{E}(R^e)|}{\sigma(R^e)}}_{\text{Data}}$, where $m = \beta u'(C_1)/u'(C_0)$ is the intertemporal marginal rate of substitution (IMRS) and R^e is the excess return on a given asset class relative to the risk free rate (for a more general treatment see, for example, Barillas et al. (2007)). Why do we need to modify the HJ bound when the agent seeks robust policies? Note that the HJ bound links observed data (RHS) to the predictions of a candidate model (LHS). Since we refer to $\mathbb{P}$ as the objective measure, the moments on the RHS are taken with respect to $\mathbb{P}$ (i.e. the data are actually generated under $\mathbb{P}$). However, pricing is done using the subjective distribution $\mathbb{Q}$ since the robust agent's IMRS serves as the pricing kernel. In our two period example, one can show that the LHS takes the following form $\frac{\sigma(m)}{\mathbb{E}^{\mathbb{Q}}(m)} - \frac{h}{\sigma(R^e)} \approx \sigma\left(1+\frac{1}{\theta}\right)$. This expression implies that there is always a value of θ that satisfies the bound. The desire for robustness introduces the term $-h/\sigma(R^e)$ on the LHS. This term can potentially be large.

and investment. For simplicity we assume a Lucas tree type economy with a conditionally heteroskedastic growth rate of output. Our ultimate goal is to analyze the implied equilibrium yield curve in this economy and, in particular, identify the implications of robustness for the term structure of interest rates.[12]

3.3.1. The Economy

There is a single consumption good which serves as the numeraire. We fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a univariate Brownian motion $B = \{B_t : t \geq 0\}$. The diffusion of information is described by the filtration $\{\mathcal{F}_t\}$ on $(\Omega, \mathcal{F})$. All stochastic processes are assumed to be progressively measurable relative to the augmented filtration generated by B. Note, however, that this probability space corresponds to an approximating model and serves only as a reference point for the robust agent. The agent entertains a *set* of possible probability measures on $(\Omega, \mathcal{F})$, denoted by $\mathcal{P}$. The size of this set is determined by a penalty function (relative entropy) which is incorporated into the agent's utility function. Every element in $\mathcal{P}$ is equivalent to $\mathbb{P}$ (i.e., define the same null events as $\mathbb{P}$). We denote the distorted measure which the agent chooses by $\mathbb{Q} \in \mathcal{P}$. The assumption that the penalty function is the relative entropy imposes a lot of structure on the possible distorted measures. By Girsanov's theorem we require the distorted measure

[12]Gagliardini et al. (2004) also study the implications of robust control for the behavior of the term structure of interest rates in a Cox et al. (1985)-type economy. We differ from their analysis along two dimensions. First, they study a two factor model closely related to Longstaff and Schwartz (1992), while we focus on a one factor model. Second, and more importantly, we study the empirical implications of our model and quantify the contribution of the state dependent market price of model uncertainty to our understanding of asset prices both in the equity and bond market. We also present supporting evidence for our key assumption of state dependent volatility in consumption growth. Finally, we estimate the implied degree of uncertainty aversion implied by the data.

to be absolutely continuous with respect to the reference measure. Finally, the conditional expectation operator under $\mathbb{P}$ and $\mathbb{Q}$ is denoted respectively by $\mathbb{E}_t(\cdot) \equiv \mathbb{E}(\cdot|\mathcal{F}_t)$ and $\mathbb{E}_t^{\mathbb{Q}}(\cdot) \equiv \mathbb{E}^{\mathbb{Q}}(\cdot|\mathcal{F}_t)$.

Let D be an exogenous output process that follows a geometric Brownian motion and solves the following stochastic differential equation (SDE),

$$dD_t = D_t \mu dt + D_t \sqrt{v_t} dB_t. \tag{3.4}$$

We can obviously think of D as a general dividend process of the economy. We allow the trading of ownership shares of the output tree. The parameters μ and v are the local expectations (drift) and the local variance of the output growth rate, respectively. We assume that v follows a mean-reverting square-root process:

$$\begin{aligned} dv_t &= (a_0 + a_1 v_t)\, dt + \sqrt{v_t}\sigma_v dB_t, \\ a_0 &> 0, \quad a_1 < 0, \quad \sigma_v \in \mathbb{R}, \quad 2a_0 \geq \sigma_v^2. \end{aligned} \tag{3.5}$$

Note that the same shock (Wiener increments) drives both the output growth and the output growth volatility processes.[13] We impose this assumption to retain the parsimonious description of the economy. The requirement $a_1 < 0$ guarantees that v converges back to its steady state level $-\frac{a_0}{a_1} (= \bar{v})$ at a velocity $-a_1$. The long run level of volatility is positive since $a_0 > 0$. The Feller condition $2a_0 \geq \sigma_v^2$ guarantees that the drift is sufficiently strong to ensure that $v > 0$ a.e. once $v_0 > 0$. The parameter σ_v is constant over

[13] We could also make the expected instantaneous output growth rate, μ, stochastic. By assuming, for example, an affine relation between μ_t and v_t, the model remains tractable and can be solved analytically. For the purpose of this paper, however, we maintain the assumption of a constant drift in the output process.

time. We show below that the sign of σ_v plays an important role in our model since it determines the risk exposure of default free bonds to the source of risk in the economy.

When v is constant over time, the market price of risk is state independent, and the expectations hypothesis of the term structure of interest rates holds. This result stands in sharp contrast to the empirical evidence (e.g., Fama and Bliss (1987), Campbell and Shiller (1991), Backus et al. (1998), Cochrane and Piazzesi (2002)). We discuss in the next section how stochastic volatility interacts with robustness considerations to affect the predictions of our model.

Let dR_t be the instantaneous return process on the ownership of the output process and S_t be the price of ownership at time t. Then, we can write

$$\begin{aligned} dR_t &\equiv \frac{dS_t + D_t dt}{S_t} \\ &= \mu_{R,t} dt + \sigma_{R,t} dB_t, \end{aligned} \tag{3.6}$$

where μ_R and σ_R are determined in equilibrium. We also let r be the short rate process, which is determined in equilibrium.

3.3.2. The Dynamic Program of the Robust Representative Agent

The robust representative agent consumes continuously and invests both in a risk-free and a risky asset. The risky asset corresponds to the ownership on the output process (the tree). The risk free asset is in zero net supply in equilibrium. As discussed in Section 3.2, the agent chooses optimally a distortion to the underlying model in a way that makes his decisions robust to statistically small model misspecification. Formally, the agent has the following objective function

$$\sup_{C,\alpha}\inf_{\mathbb{Q}}\left\{\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{\infty} e^{-\rho(s-t)}u\left(C_s\right)ds\right]+\theta\mathcal{R}_t\left(\mathbb{Q}\right)\right\}, \tag{3.7}$$

subject to his dynamic budget constraint

$$dW_t=\left[r_tW_t+\alpha_tW_t\left(\mu_{R,t}-r_t\right)-C_t\right]dt+\alpha_tW_t\sigma_{R,t}dB_t, \tag{3.8}$$

where $\mathbb{Q}$ is the agent's subjective distribution, W is the agent's wealth, ρ is the subjective discount factor, C is the consumption flow process, α is the portfolio share invested in the risky asset, and θ is the multiplier on the relative entropy penalty $\mathcal{R}$. The level of θ can be interpreted as the magnitude of the desire to be robust. When θ is set to infinity, (3.7) converges to the expected time additive utility case. A lower value of θ means that the agent is more fearful of model misspecification and thus chooses $\mathbb{Q}$ further away from $\mathbb{P}$ in the relative entropy sense. In other words, the set $\mathcal{P}$ is larger the smaller θ is.

Let $\mathcal{L}_2$ be the set of all progressively measurable univariate processes h such that $\int_0^\infty h_s^2ds<\infty$ a.s.. Let $\mathcal{H}$ be the set of all $h\in\mathcal{H}\subseteq\mathcal{L}_2$ such that the process $\xi^{\mathbb{Q}}$ defined by

$$\xi_t^{\mathbb{Q}}=\exp\left(\int_0^t h_sdB_s-\frac{1}{2}\int_0^t h_s^2ds\right),\quad t\geq 0, \tag{3.9}$$

and is a $\mathbb{P}$-martingale. Then, h defines the probability $\mathbb{Q}\in\mathcal{P}$ by $\mathbb{Q}\left(F\right)=\lim_{t\to\infty}E\left(1_F\xi_t^{\mathbb{Q}}\right)$ for every $F\in\mathcal{F}$, and $\xi^{\mathbb{Q}}$ is also the conditional density process, or the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$, and satisfies

$$\xi_t^{\mathbb{Q}} = \mathbb{E}_t\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right), \quad t \geq 0.$$

By Girsanov's theorem, for every $h \in \mathcal{H}$ we can define a Brownian motion under $\mathbb{Q}$ as

$$B_t^{\mathbb{Q}} = B_t - \int_0^t h_s ds, \quad t \geq 0. \tag{3.10}$$

Using (3.10) we can also rewrite (3.9) as

$$\xi_t^{\mathbb{Q}} = \exp\left(\int_0^t h_s dB_s^{\mathbb{Q}} + \frac{1}{2}\int_0^t h_s^2 ds\right), \quad t \geq 0. \tag{3.11}$$

Note that $\xi^{\mathbb{Q}}$ is not a $\mathbb{Q}$-martingale.

With this setup at hand, the relative entropy process $\mathcal{R}(\mathbb{Q})$ for some $\mathbb{Q} \in \mathcal{P}$ can be expressed conveniently as[14]

$$\mathcal{R}_t(\mathbb{Q}) = \frac{1}{2}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{\infty} e^{-\rho(s-t)} h_s^2 ds\right], \quad t \geq 0. \tag{3.12}$$

The expression in (3.12) allows us to rewrite (3.7) as

$$\sup_{C,\alpha} \inf_h \left\{\mathbb{E}_t^{\mathbb{Q}} \int_t^{\infty} e^{-\rho(s-t)} \left[u(C_s) + \frac{\theta}{2} h_s^2\right] ds\right\}. \tag{3.13}$$

Note that now the infimization problem is well defined over $\mathcal{H}$.

[14]See, for example, Hansen et al. (2006) and section 3, and especially proposition 4, in Skiadas (2003).

Finally, using (3.10) we write (3.4), (3.5) and (3.8) under the distorted measure $\mathbb{Q}$. For example, the wealth process under the agent's subjective distribution corresponds to

$$dW_t = \left[r_t W_t + \alpha_t W_t \left(\mu_{R,t} - r_t\right) - C_t + \underbrace{h_t \alpha_t W_t \sigma_{R,t}}_{\text{Drift contamination}} \right] dt + \alpha_t W_t \sigma_{R,t} dB_t^{\mathbb{Q}}. \tag{3.14}$$

In the context of the market return, for example, this drift contamination has an obvious interpretation: it is the uncertainty premium the agent requires for bearing the risk of potential model misspecification

$$dR_t = \left[\mu_{R,t} - \underbrace{(-h_t \sigma_{R,t})}_{\text{Uncertainty premium}} \right] dt + \sigma_{R,t} dB_t^{\mathbb{Q}}. \tag{3.15}$$

The process h is the (negative of) the process for the *market price of model uncertainty.* The diffusion part $\sigma_{R,t}$ on the return process is, as usual, the *risk exposure* of the asset. The product $-h_t \sigma_{R,t}$ is the equilibrium *uncertainty premium.* In order to obtain the risk premium in the drift, one needs to rewrite the return process under the risk neutral measure. Let $\varphi \equiv \frac{\mu_R - r}{\sigma_R}$ be the local Sharpe ratio, or the process for the market price of risk in the model. Then, using the same arguments that lead to (3.10) we can link the reference measure to a risk neutral measure, denoted by B^q, as follows

$$B_t^q = B_t + \int_0^t \varphi_s ds, \quad t \geq 0. \tag{3.16}$$

Alternatively, using both (3.10) and (3.16) to relate the risk neutral measure to the distorted measure, we can write

$$B_t^q = B_t^{\mathbb{Q}} + \int_0^t (\varphi_s + h_s)\, ds. \tag{3.17}$$

Then, the return process can be written as

$$dR_t = \left(\mu_{R,t} - \underbrace{\varphi_t \sigma_{R,t}}_{\text{Risk premium}} \right) dt + \sigma_{R,t} dB_t^q.$$

Here, we see that the risk exposure σ_R is identical to the asset's uncertainty exposure. This result implies that there is a perfect correlation of risk and uncertainty premia in our model.

3.3.3. Optimal Policies with Robust Control

In this subsection we solve the robust representative agent's dynamic problem posited in Section 3.3.2. We use dynamic programming to derive closed form solutions for his optimal consumption and investment decisions policies together with the conditional distorted distribution.

Let $J(W_t, v_t)$ denote the agent's value function at time t where W_t and v_t correspond to current wealth and the conditional variance level respectively. The agent's Hamilton-Jacobi-Bellman (HJB) equation is[15,16]

$$\rho J = \log C_t + \frac{\theta}{2} h_t^2 + \mathcal{D}^{\mathbb{Q}} J, \tag{3.18}$$

[15] See also Anderson et al. (2003) and Maenhout (2004) for similar formulations.

[16] See Appendix 7 for a more detailed derivation of the policies and the value function.

where $\mathcal{D}^{\mathbb{Q}}$ is the Dynkin generator under the distorted measure. Informally, $\mathcal{D}^{\mathbb{Q}} J$ is $\mathbb{E}^{\mathbb{Q}}(dJ)/dt$ and is derived by applying Ito's lemma and using (3.14) and the distortion of (3.5) to characterize the dynamics of J. The only difference between this HJB equation and a standard one is the introduction of a *cost* and *benefit* for distorting the objective distribution. The cost is given by the relative entropy term $\frac{\theta}{2}h_t^2$ (pessimism is costly) and the benefit is hidden in the distortion of the Dynkin generator. The drift of the J process is distorted since the state processes are themselves distorted.

The solution for h from the infimization problem is given by

$$h_t = -\frac{1}{\theta}\left(J_{W,t}\sigma_{W,t} + J_{v,t}\sigma_v\sqrt{v_t}\right). \tag{3.19}$$

We can see that the intuition from the 2-period example regarding θ survives in our infinite-horizon, continuous-time setting: lower θ implies higher distortion. However, there are important differences since volatility is stochastic. First and foremost, the robustness correction h is state dependent. The robust agent derives the distorted conditional distribution in such a way that the reference conditional distribution first order stochastically dominates the chosen distorted conditional distribution. If it was not the case then there would be states of the world in which the robust agent would be considered optimistic. Also, the agent wants to maintain the optimal relative entropy penalty constant since θ is constant. In order to achieve this when conditional volatility is stochastic, the distortion has to be stochastic and increase with volatility (see expression (3.2)).

Second, the size of the distortion is inversely proportional to the penalty parameter θ: the distortion vanishes as $\theta \to \infty$. Third, whenever the marginal indirect utility and volatility of wealth (J_W and σ_W) are high, the agent becomes more sensitive to uncertainty and distorts the objective distribution more. Low levels of wealth imply large marginal indirect utility of wealth. These are states in which the agent seeks robustness more strongly. The second term in the parentheses corresponds to the effect of the state v on the distortion h. Since $J_v < 0$ for all reasonable parametrizations, the sign of σ_v dictates the optimal response of the agent. Consider the benchmark case when σ_v is positive. Following a positive shock, marginal utility falls as consumption rises, and volatility v increases. Therefore, the investment opportunity set deteriorates exactly when the agent cares less about it. Since the evolution of v serves as a natural hedge for the agent, he reduces the distortion h. The opposite occurs when $\sigma_v < 0$.

Maximizing (3.18) over α, the optimal portfolio holding of the risky asset at time t can be expressed in two equivalent forms, each emphasizing a different aspect of the intuition. The first one is the myopic demand

$$\alpha_t = \frac{\mu^{\mathbb{Q}}_{R,t} - r_t}{\sigma^2_{R,t}}. \tag{3.20}$$

Equation (3.20) that the demand for the risky asset is myopic: the agent only cares about the current slope of the mean-variance frontier. However, this slope is constructed using his subjective beliefs. From an objective point of view, the agent deviates from the observed mean-variance frontier portfolio due to his (negative) distortion to the mean h: he believes the slope is lower and thus decreases his demand for the risky asset. The second form of the demand for the risky asset captures this idea

$$\alpha_t = \frac{\mu_{R,t} - r_t}{\sigma^2_{R,t}} + \frac{h_t}{\sigma_{R,t}}. \tag{3.21}$$

The first element on the right-hand side of equation (3.21) describes the myopic demand of a log-utility agent who is endowed with the objective measure. However, the pessimistic agent optimally reduces his holdings of the risky asset by $h/\sigma_R < 0$ since he believes the expected return on the risky asset is lower than the one implied by the objective measure.

We posit the guess that the value function is concave (log) in the agent's wealth and affine in the conditional variance

$$J\left(W_t, v_t\right) = \frac{\log W_t}{\rho} + \delta_0 + \delta_1 v_t. \tag{3.22}$$

Now, we can use (3.22) to rewrite (3.19) as

$$h_t = -\frac{1}{\theta}\left(\frac{1}{\rho} + \delta_1 \sigma_v\right)\sqrt{v_t}. \tag{3.23}$$

Here, we see that the distortion, or the (negative of the) market price of model uncertainty is linear in the conditional volatility of the output growth rate. In equilibrium $\sqrt{v}$ is the conditional volatility of the consumption growth rate.[17]

[17] It is possible to assume an exogeneous process for μ separately from v and still maintain a fairly simple closed-form equilibrium. All one needs is to scale the local volatility of μ with the current level of $\sqrt{v}$. One such possible model will assume

$$d\mu_t = \left(x_0 + x_1 \mu_t\right) dt + \sqrt{v_t}\sigma_\mu dB_{2,t}, \quad x_1 < 0$$

Then, the value function for the robust agent is

$$J\left(W_t, \mu_t, v_t\right) = \frac{\log W_t}{\rho} + \delta_0 + \delta_1 \mu_t + \delta_2 v_t$$

and the robust correction is still linear in the conditional stochastic volatility

$$h = -\frac{\sqrt{v}}{\theta}\left(\frac{1}{\rho} + \delta_1 \sigma_\mu + \delta_2 \sigma_v\right)$$

We can also rewrite (3.20) as

$$\alpha_t = \underbrace{\frac{1}{1+\frac{1}{\rho\theta}} \frac{\mu_{R,t} - r_t}{\sigma_{R,t}^2}}_{\text{Myopic demand}} - \underbrace{\frac{\frac{1}{\theta}}{1+\frac{1}{\rho\theta}} J_v \sigma_v}_{\text{Hedging demand}} \quad .$$

The first element on the RHS corresponds to a variant of the usual myopic demand for a risky asset in a log-utility setup. This term simply gives the trade-off between excess return compensation and units of conditional variance. Note that the coefficient is not unitary, as in the usual log problem. The reason is best understood when one keeps in mind the mapping between a robust control agent and an SDU agent with unitary EIS. When introducing robustness, we effectively increase risk aversion, but maintain the unitary EIS. This effect pushes down the demand schedule for the risky asset. The second element is the hedging-type component arising from uncertainty aversion, and it is larger in absolute terms the larger J_v or σ_v, ceteris paribus. The hedging part is positive since $J_v \sigma_v > 0$ due to the intuition given in (3.19).

The consumption policy is unchanged when the agent seeks robust policies: $C = \rho W$. The wealth and substitution effects still cancel out in our setup. When volatility increases, the agent decreases his holdings of the risky asset substantially since he cannot amortize the volatility increase through changes in his consumption. Unitary EIS implies a constant consumption-wealth ratio, thus all volatility changes are channelled through the asset market. In other words, robustness, or pessimism, entails that the agent perceives the local expectations on the risky asset to be lower than the objective drift on the same asset. The substitution effect implies that the agent should invest less since the asset is expected to yield low return in the future. In contrast, the wealth effect predicts that he should consume less today and save instead. In the case of log utility, these two effects cancel each other. Consequently, the effect of robustness on the consumption policy is eliminated. Changing a log-agent's desire to be robust will only affects the risk free rate and the return on the risk free asset.

3.3.4. Robust Equilibrium

In this section we solve for the equilibrium price of the risky asset and the risk free rate. We define and discuss the implications of the robustness assumption on the equilibrium prices. Specifically, we will examine the level and volatility of both the equity premium and the risk free rate. First, we define a robust equilibrium:

Definition 2. *A robust equilibrium is a set of consumption and investment policies/processes* (C, α) *and a set of prices/processes* (S, r) *that support the continuous clearing of both the market for the consumption good and the equity market* $(C = D, \alpha = 1)$ *and (3.13) is solved subject to (3.8), (3.5) and (3.10).*[18]

The only difference between this equilibrium definition and a conventional one is that the agent solves a robust control problem. We will now show that this affects the equilibrium short rate.

In equilibrium, since the agent consumes the output $(C = D)$ the local consumption growth rate and the local output growth rate are the same $(\mu_C = \mu)$. Also, the agent's equilibrium path of wealth is identical to the evolution of the price of the 'tree' since $\alpha = 1$. Therefore, $W = S$. Hence, $D = C = \rho W = \rho S$. As is usually the case with

[18]The same definition also appears in Maenhout (2004). Without stochastic volatility considerations, he also derives the equilibrium risk free rate and equity premium.

a log representative agent, not only the consumption wealth ratio is constant but so is the dividend-price ratio $\left(\frac{C}{W} = \frac{D}{S} = \rho\right)$. We see that, as in the two-period example, the robustness considerations do not affect the consumption policy and the pricing of the 'tree'. In that case, what are the implications of the fact that the agent seeks robust policies? The effect shows up in the risk free rate and the way expectations are formed about growth rates or the return on the risky asset. The equilibrium risk free rate can be derived from (3.21)

$$
\begin{aligned}
r_t &= \rho + \mu_{C,t} + \sqrt{v_t} h_t - v_t \\
&= \rho + \mu_{C,t} - v_t \left[1 + \frac{1}{\theta}\left(\frac{1}{\rho} + \delta_1 \sigma_v\right)\right] \\
&= \rho + \mu - \phi v_t. \qquad (3.24)
\end{aligned}
$$

For the remainder of the paper we define

$$
\phi \equiv 1 + \frac{1}{\theta}\left(\frac{1}{\rho} + \delta_1 \sigma_v\right).
$$

The usual comparative statics arguments apply to this short-rate equation. A higher subjective discount rate preference parameter ρ makes the agent wants to save less, so that the equilibrium real rate must be higher to compensate the agent for saving as much as before. Higher future expected consumption growth makes the agent want to consume more today (substitution effect). The real rate must therefore be higher to prevent him from borrowing. Higher consumption volatility activates a precautionary savings motive, so that the real rate must be lower to prevent the agent from saving. The role of robustness can be interpreted in two ways. First, robustness distorts the expected consumption growth rate. Lower expected consumption growth rate lowers the equilibrium risk free rate since the substitution effect is now weaker. The second interpretation, which may be more intuitive, is that robustness amplifies the effect of the precautionary savings motive in the same direction ($h < 0$ when $\theta < \infty$), and thus lowers the equilibrium level of the short rate. All else equal, the robust agent wants to save more than an expected utility agent and therefore the former needs a stronger equilibrium disincentive to save in the form of lower risk free rate. In this latter interpretation, the distortion is proportional to consumption growth rate volatility and thus can be interpreted as a modification to the precautionary savings motive.

The equilibrium local expected return on the risky asset can immediately be derived from (3.6) and the fact that $S = D/\rho$

$$\begin{aligned} dR_t &= \left(\mu_{D,t} + \rho\right) dt + \sigma_{D,t} dB_t \\ &= \left(\mu_{D,t} + \rho + h_t \sigma_{D,t}\right) dt + \sigma_{D,t} dB_t^{\mathbb{Q}}. \end{aligned}$$

The observed equity premium is[19]

$$\begin{aligned} \mu_{R,t} - r_t &= \phi v_t = \underbrace{v_t}_{\text{Risk Premium}} + \underbrace{(\phi - 1)\, v_t}_{\text{Uncertainty Premium}} \\ &= cov_t\left(\frac{dC_t}{C_t}, dR_t\right) + \frac{1}{\theta}\left(\frac{1}{\rho} + \delta_1 \sigma_v\right) v_t. \end{aligned}$$

[19]We use the qualifier ‘observed’ to emphasize again that what the agent treats as merely a reference model is actually the DGP. Therefore, anything under the reference measure is what the econometrician observe when he has long time series of data.

The equity premium has both a risk premium and an uncertainty premium components. The former is given by the usual relation between the agent's marginal utility and the return on the risky asset. If the correlation between the agent's marginal utility and the asset return is negative, the asset commands a positive risk premium $\left[cov_t\left(\frac{dC_t}{C_t}, dR_t\right) > 0\right]$ and vice versa. The higher the degree of robustness (i.e., the smaller the parameter θ) higher are the uncertainty premium and the market price of model uncertainty. While a decrease in θ increases the equity premium, it also decreases the risk free rate through the precautionary savings motive. The EIS is independent of θ. By lowering θ we are effectively increasing the aversion to model uncertainty but not affecting the intertemporal substitution. Also, the distortion of equilibrium prices is not surprising since the agent believes consumption growth rate is lower than the actual growth rate under the reference model. Hence, his IMRS process is distorted.

We see that robustness can account for both a high observed equity premium and low level of the risk free rate. What about the volatility of the risk free rate? Since we do not change the substitution motive, the only magnification is through the precautionary savings. Empirically v is extremely smooth and, thus, contributes very little to the volatility of r.[20]

Previous studies (e.g., Anderson et al. (2003), Skiadas (2003), Maenhout (2004)) have showed that without wealth effects, a robust control economy is observationally equivalent to a recursive utility economy in the discrete time case (Epstein and Zin (1989), Weil (1990)) or to a stochastic differential utility (SDU) in the continuous time economy as in

[20]If we allow for a stochastic μ with positive correlation with v, fluctuations in v will be countered by movements in μ since they affect the risk free rate with opposite signs. In other words, if we allow the substitution effect and the precautionary motive to vary positively over time, the risk free rate can be very stable.

Duffie and Epstein (1992a) and Duffie and Epstein (1992b). Thus, our combined market price of risk and uncertainty can be viewed as an effective market price of risk in the SDU economy.[21] The difficulty with such approach is that it requires implausibly high degrees of risk aversion. Another difficulty arises in the context of the Ellsberg paradox. Our approach assumes that agents do not necessarily know the physical distribution and want to protect themselves against this uncertainty.

3.4. Pricing the Term Structure of Interest Rates

Denote the intertemporal marginal rate of substitution (IMRS) process by Λ where $\Lambda_t \equiv e^{-\rho t}/C_t$. Using Ito's lemma we characterize the dynamics of Λ as

$$\frac{d\Lambda_t}{\Lambda_t} = -r_t dt - \sqrt{v_t} dB_t^{\mathbb{Q}}, \tag{3.25}$$

where the drift is the (negative of) the short rate and the diffusion part is the market price of risk.

Using (3.25) it is straightforward to price default free bonds.[22,23] We use the following guess for the functional form for the time t default-free zero-coupon bond price (an affine

[21] Even though we do not lose the homotheticity of our problem since our agent has log preferences, Maenhout (2004) discusses the need to rescale the problem in order to obtain an exact mapping from the robust control economy to an SDU economy. We do not incorporate this rescaling since our interpretation focuses solely on an agent who faces Knightian uncertainty and acts as an ambiguity averse agent. Thus, we conduct this study with the intention of studying the behavior of both the market price of risk and uncertainty.

[22] A more detailed derivation of the bond pricing rule, using the PDE approach, is in appendix 8.

[23] Our paper belongs to the vast literature on affine term structure models. The term structure literature is too large to summarize here but studies can be categorized into two strands - equilibrium and arbitrage free models. Our paper belongs to the former strand. The advantage of the equilibrium term structure models is mainly the ability to give meaningful macroeconomic labels to factors that affect asset prices. Dai and Singleton (2003) and Piazzesi (2003), for example, review in depth the term structure literature.

yield structure) that matures at time T. Let $\tau = T - t$ and write

$$p(\tau; v_t) = \exp\left[\beta_0(\tau) + \beta_1(\tau) v_t\right]. \tag{3.26}$$

Start with the fundamental pricing equation where the expected marginal utility weighted price is a martingale

$$\mathbb{E}_t^{\mathbb{Q}}\left[d(\Lambda_t p_t)\right] = 0 \Longrightarrow \mathbb{E}_t^{\mathbb{Q}}\left(\frac{dp_t}{p_t}\right) - r_t dt = -\frac{d\Lambda_t}{\Lambda_t}\frac{dp_t}{p_t}. \tag{3.27}$$

The excess expected return on a bond over the short rate is determined by the conditional covariance of the return on the bond and marginal utility, or alternatively, by the product of the market price of risk and the risk exposure of the bond. As usual, if they covary positively, the asset serves as a hedge against adverse fluctuations in marginal utility and commands a negative risk premium. In times of high volatility, the precautionary savings motive induces the agent to shift his portfolio away from the equity market and towards bonds. Such a shift induces an upward pressure on bond prices (and thus yields decrease). Therefore, bonds pay well in good times, rendering them a risky investment. Note, however, that the expectations are taken over the distorted measure. These distorted expectations affect prices in a systematic way relative to the prices that would have prevailed under the objective measure, introducing an uncertainty premium element into the price of the bond.

From (3.26) one can show that the risk premium on a default free bond is

$$-\frac{d\Lambda_t}{\Lambda_t}\frac{dp_t}{p_t} = \beta_1(\tau)\sigma_v v_t,$$

where β_1 is positive and determines the cross section restrictions amongst different maturity bonds. The sign of the risk premium is determined by the correlation of the output growth rate and the conditional variance. In the next section we discuss the intuition behind the predictions of the model, and especially the role robustness plays in our context.

Moreover, the *observed* excess return that long term bonds earn over the short rate is not completely accounted for by the risk premium component. We derive the dynamics of the return on a bond with arbitrary maturity by applying Ito's lemma to (3.26). Under the objective measure we have,

$$\frac{dp(\tau;v_t)}{p(\tau;v_t)} = \left[r_t + \underbrace{\beta_1(\tau)\,\sigma_v v_t}_{\text{Risk Premium}} + \underbrace{\beta_1(\tau)\,\sigma_v v_t(\phi-1)}_{\text{Uncertainty Premium}} \right] dt + \beta_1(\tau)\,\sigma_v \sqrt{v_t} dB_t.$$

In the presence of uncertainty aversion, there is an uncertainty premium that drives a wedge between the return on a τ-maturity bond and the short rate. The more robust the agent, the larger the market price of uncertainty is in absolute terms (i.e., ϕ is larger so $-h = (\phi-1)\sqrt{v}$ is larger). Also, higher conditional variance increases the uncertainty premium since the agent distorts the mean of the objective model more. In other words, higher σ_v also increases the uncertainty exposure of the asset. We can express the uncertainty premium as

$$\underbrace{-h_t}_{\text{Price of uncertainty}} \times \underbrace{diff\left(\frac{dp(\tau;v_t)}{p(\tau;v_t)}\right)}_{\text{Uncertainty Exposure}} = \underbrace{\beta_1(\tau)\,\sigma_v v_t(\phi-1)}_{\text{Uncertainty premium}},$$

where $diff(\cdot)$ is the diffusion part of the process. The intuition and implication of this result are discussed in the empirical section (3.5).

The yield on a given bond is simply an affine function of the conditional variance

$$\mathcal{Y}(\tau; v_t) = -\frac{1}{\tau} \ln p(\tau; v_t).$$

The two extreme ends of the yield curve are $\lim_{\tau \to 0} \mathcal{Y}(\tau; v_t) = r_t$ and $\lim_{\tau \to \infty} \mathcal{Y}(\tau; v_t) = \rho + \mu - a_0 \bar{\beta}_1$. Thus the spread is

$$\lim_{\tau \to \infty} \mathcal{Y}(\tau; v_t) - \lim_{\tau \to 0} \mathcal{Y}(\tau; v_t) = -a_0 \bar{\beta}_1 + \phi v_t,$$

where the expression for $\bar{\beta}_1$ is given in Appendix 8.

3.4.1. Why Can The Model Explain the Cross Section of Bond Yields? Some Intuition

In this section we explain more intuitively why the model accounts for the cross section regularities of bond yields. More importantly, we focus on the contribution of robustness considerations to the results.

3.4.1.1. Bond Returns and Upward Sloping Yield Curve. A bond price is the conditional expected IMRS. Ceteris paribus, a positive shock to the expected growth rate of consumption lowers the equilibrium bond price, and thus, increases the yield on that bond. The bond price decreases due to a negative substitution effect. If expected consumption growth rate has positive contemporaneous correlation with consumption, or negative marginal utility, the bond is considered a safe asset and therefore commands a negative risk premium. The opposite also holds true. Furthermore, the expected IMRS is also affected by the conditional variance of consumption growth but in the opposite direction. Holding everything else constant, a positive shock to the conditional variance of the growth rate of consumption increases the bond price, and thus, lowers the yield on that bond. Here, people want to save more due the precautionary savings motive and therefore, in equilibrium, bond prices are higher and yields are lower. Again, what determines the sign of the risk premium is the correlation of the conditional variance with marginal utility. If the correlation with marginal utility is negative the bond is considered a risky asset since it pays well in good times. Hence, investors require a positive risk premium on the bond.

Since the distortion h is linear in the conditional volatility of consumption growth (v), it is natural to think of robustness as magnifying the precautionary savings motive. Mean reversion of the conditional variance process coupled with a positive correlation between conditional variance and consumption growth entails a positive risk premium on long term bonds relative to short term bonds. Also, since long term yields are averages of future expected short term yields plus a risk premium, the average yield curve is expected to be upward sloping.

An alternative way of interpreting the average positive slope of the yield curve is by examining the objective and subjective (endogenous) evolution of the conditional variance of consumption growth rate. The (perceived) evolution of v under the distorted measure $\mathbb{Q}$ is different from the evolution of v under the objective measure $\mathbb{P}$ in two respects. Write (3.5) under both measures

$$\begin{aligned} dv_t &= -\kappa_v \left(v_t - \bar{v}\right) dt + \sigma_v \sqrt{v_t} dB_t \\ &= -\kappa_v^{\mathbb{Q}} \left(v_t - \bar{v}^{\mathbb{Q}}\right) dt + \sigma_v \sqrt{v_t} dB_t^{\mathbb{Q}}. \end{aligned} \tag{3.28}$$

Here, κ_v is the velocity of reversion and $\bar{v}$ is the steady state of v, both under the reference measure. However, the subjective velocity of reversion is

$$\kappa_v^{\mathbb{Q}} = \kappa_v - \sigma_v (1 - \phi) > \kappa_v \tag{3.29}$$

and the subjective steady state is

$$\bar{v}^{\mathbb{Q}} = \frac{\kappa_v}{\kappa_v^{\mathbb{Q}}} \bar{v} < \bar{v}. \tag{3.30}$$

Observation (3.30) is enough to explain the positive slope of the yield curve. Note that pricing is done using the IMRS of the robust agent and he thinks that the steady state of the conditional variance of consumption growth rate is lower than the objective target. By persistently missing the target, the agent on average believes that v is expected to decrease. In other words, he on average thinks that yields are expected to increase due to the effect of the precautionary savings motive on prices.

A different way of interpreting (3.28) is the following. The variance dynamics are characterized by a non-negative mean-reverting process. This process gravitates towards its steady state and the speed of reversion is stronger the further the variance level is from its steady state. Robustness introduces a negative distortion to the drift of the variance process $\left(h_t \sigma_v \sqrt{v_t} = \sigma_v (1 - \phi) v_t < 0\right)$. A negative distortion to the drift that depends linearly on the level of the variance introduces zero as an additional focal point to the variance process. When the variance is above its objective steady state, both the distortion and the pull towards the objective steady state work in the same direction. However, when the variance is below its steady state, both forces work in opposite directions. The distortion always pulls down towards zero while the other force pulls the variance up towards its objective steady state. The point where these two forces are equal is the subjective steady state and it is between the objective steady state (positive) and zero, leading to (3.28).

3.4.1.2. Negative Contemporaneous Correlation Between the Spread and the Level of Yields (Yield Curve Rotation), and the Term Structure of Unconditional Volatilities of Yields. In quarterly data over the sample $52.Q2 - 06.Q4$ the correlation between the level and slope of the real yield curve is -0.5083 with standard errors of 0.0992 (Newey-West corrected with 4 lags). Here, the slope is the difference between the 1-year and 3-months yields. This finding is robust over different time intervals and different frequencies. The model can account for this fact in the following way.[24] Recall that a positive shock to conditional volatility lowers yields. Also note that yields are perfectly (positively) correlated since all of them are an affine function of the same factor. However, short yields are more sensitive to conditional volatility shocks. To understand why, it helps to think about the mean reversion of the conditional variance (the ergodicity of its distribution). The effect of any shock is expected to be transitory. The full impact of the shock happens at impact and then the conditional variance starts reverting back to its steady state. Therefore, the effect of, say, a positive shock is expected to dissipate and yields are expected to start to climb back up. This expected effect is incorporated into long term yields immediately. Short yields in the far future are almost unaffected by the current shock since it is expected that the effect of the shock will disappear eventually.

[24]We explain the intuition through the time variation of the conditional volatility of consumption growth rate. One can alternatively use the substitution channel and focus on time variation in expected consumption growth rate.

Since long term yields are an average of future expected short yields plus expected risk premia, they tend to be smoother than short term yields.

The expected risk premium is also a linear function of the state, and thus inherits its mean reversion. Therefore, the expected risk premium in the far future is also smoother than the risk premium in the short run. This also contributes to the rotation of the yield curve: since the short end of the yield curve is very volatile relative to the long end, whenever yields decrease, the spread increases (or become less negative, depending on the initial state). The opposite also holds true.

3.4.1.3. How Does the Model Account for the Rapidly Declining Slope of the Yield Curve and the High Persistence of Yields? Traditionally, one factor models encounter an inherent difficulty in trying to account simultaneously for the rapidly declining slope of the yield curve (i.e., strong convexity of the slope of the yield curve) and the high persistence of yields. Time-series evidence implies that interest-rate shocks die out much more slowly than what is implied from the rapidly declining slope of the average yield curve (Gibbons and Ramaswamy (1993)).

Even though we present a one factor model, we can still account for these two facts with a single parametrization. The key lies in expression (3.29). The agent believes that the conditional variance reverts to its steady state *faster* than under the objective measure $\left(\kappa_v^{\mathbb{Q}} > \kappa_v\right)$. Since yields are affine functions of the conditional variance of consumption growth, they inherit the velocity of reversion of v under the objective model. In other words, the persistence of yields is measured ex-post and is solely determined by the objective evolution of v without any regard to what the agent actually believes.

At the same time, the slope of the yield curve (or the pricing of bonds) is completely determined by what the agent believes the evolution of v is. If $\kappa_v^{\mathbb{Q}}$ is substantially larger than κ_v, the slope of the yield curve can flatten at relatively short horizons, reflecting the beliefs of the agent that v will quickly revert to its steady state level. Since the agent persistently thinks that $\kappa_v^{\mathbb{Q}} > \kappa_v$ the slope can be on average rapidly declining. When analyzing the results of our estimation we will show that this is indeed the case.

3.4.1.4. Biased Expectations: Pessimism and (the Reverse of) Doubt. Abel (2002) argues that one can potentially account for the equity premium and the risk free rate when modeling pessimism and doubt in an otherwise standard asset pricing (Lucas tree) model. Pessimism is defined as a leftward translation of the objective distribution in a way that the objective distribution first order stochastically dominates the subjective distribution. Doubt is modeled in a way that the subjective distribution is a mean preserving spread of the objective distribution.

There is evidence that people tend to consistently underestimate both market return and the conditional volatility of output growth rate (e.g., Soderlind (2006)). Also, Giordani and Soderlind (2006) confront the Abel (2002) suggestion with survey data and find strong support for the pessimism argument in growth rates of both GDP and consumption. The result is robust over forecasts of different horizon and with both the Livingston survey and the Survey of Professional Forecasters data. However, they also find evidence of overconfidence in the sense that forecasters underestimate uncertainty. Therefore, the evidence suggests the existence of the reverse of doubt.

Our model endogenously predicts both phenomena.[25] First, robustness requirements lead the agent to pessimistic assessments of future economic outcomes (e.g., expression (3.15) in which the agent negatively distorts the expected return on the risky asset). Consequently, the agent persistently underestimates expected growth rates of both the risky asset and consumption. In that sense, robustness endogenizes the pessimism idea of Abel (2002). Our model also predicts biased expectations concerning the dynamics of the conditional variance process v in a way that is consistent with the data. Expressions (3.29) and (3.30) formalize this idea. In the case where $\sigma_v > 0$ (an assumption that we later support empirically), a pessimistic assessment of expected output growth rate leads to what can be interpreted as optimistic beliefs about future output growth volatility. In other words, the model predicts also the reverse of doubt. Note that here the agent knows exactly the current conditional variance but wrongly estimates its future evolution.

3.5. The Empirical Study

In this section we undertake three tasks. First, we provide empirical support for our assumption that the volatility of consumption growth is state dependent. Our discussion complements the analysis of Bansal and Yaron (2004) and Bansal et al. (2005) who argue that there is stochastic volatility in the growth rate of consumption. Second, we estimate our model.[26] There are six parameters in the model, five of which are standard. Third, we interpret the non-standard parameter θ using *detection error probabilities* to map θ.

---

[25]For a decision-theoretic link between ambiguity averse agent and the setup of Abel (2002), see Ludwig and Zimper (2006).

[26]Wachter (2001), for example, studies the effect of consumption externalities (habits) on the term structure of interest rate by drawing empirical restrictions from consumption data and both the equity and bond markets.

Since the model is a description of a real economy, all the data we use are expressed in real terms. The description and discussion of the data are relegated to Appendix 9.[27]

3.5.1. Conditionally Heteroskedastic Consumption Growth

In this subsection we provide direct empirical evidence about the level and behavior of the conditional variance of real aggregate consumption growth. We examine two measures of conditional volatility: realized volatility and series estimated from various GARCH specifications.

3.5.1.1. ARMAX-GARCH Real Consumption Growth Rate. We start with a simple univariate time series parametric estimation. The model we are fitting to the consumption growth process is an ARMAX$(2,2,1)$ model and a GARCH$(1,1)$ to the innovations process:

$$\begin{aligned} A(L)\frac{\Delta C_t}{C_{t-1}} &= c + B(L)R_{t-1} + C(L)\eta_{C,t}, \\ \eta_{C,t+1} &= \sigma_{C,t}\varepsilon_{C,t+1}, \quad \varepsilon_{C,t} \sim N(0,1), \\ D(L)\sigma_{C,t} &= \omega + F(L)\eta^2_{C,t}, \end{aligned} \tag{3.31}$$

where A, B, C, D, F are polynomials of orders $2, 1, 2, 1, 1$ respectively, in lag operators. $\frac{\Delta C_t}{C_{t-1}}$, R_t, η_t are, respectively, the realized real consumption growth rate at time t, the real return on the aggregate market index at time $t-1$, and an innovation process with time-varying variance. In Figure 3.1 we plot the GARCH volatility estimates for both real

[27] A few studies, for example Brown and Schaefer (1994) and Gibbons and Ramaswamy (1993), also use real data to estimate a term structure model. However, they do not draw restrictions from the equity market and consumption data and their preferences assumption is standard which implies that the equity premium and risk free rate puzzles are still present in the models they estimate.

aggregate consumption growth rate and the real return on the aggregate stock market. We also plot a measure of realized volatility for both consumption growth and market return series that we obtain by fitting an ARMA$(2,2)$ to the original data and then use the square innovations to construct the realized variance series. The sample period is $Q2.52 - Q4.06$.

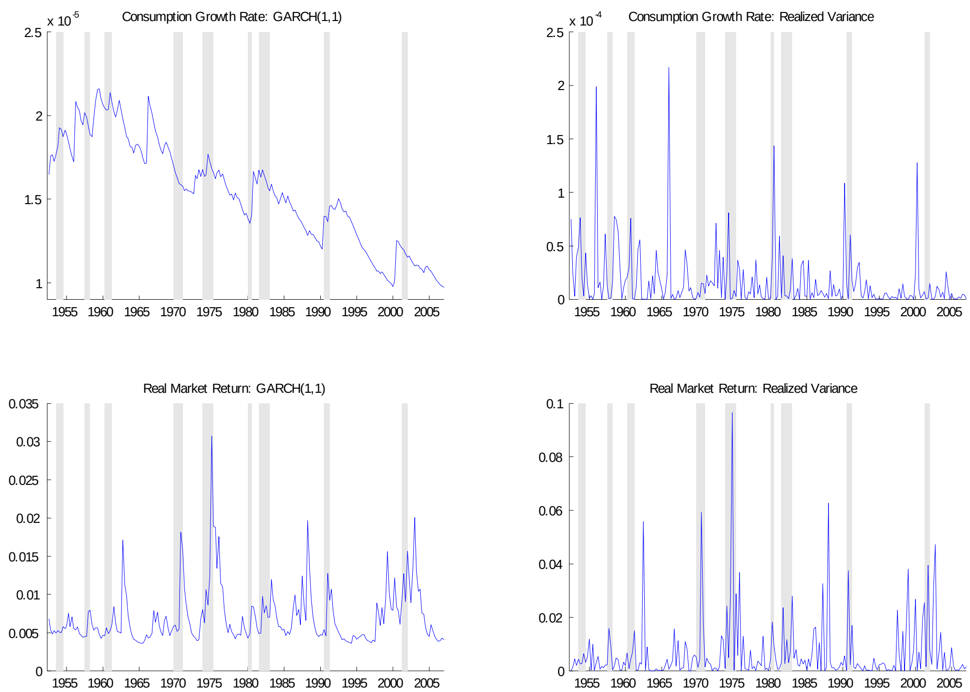

Figure 3.1: ARMAX-GARCH estimation for both real consumption growth rate and real aggregate market return. We fit model (3.31) and present the GARCH estimates for the conditional variance of real consumption growth rate and real aggregate market return in the left panel. The right panel present the square innovations from an ARMX specification to real consumption growth and real aggregate market return. The quarterly data is $Q2.52 - Q4.06$. The gray bars are contraction periods determined by the NBER.

First, there seems to be evidence of what has been dubbed as the 'Great Moderation' (e.g., Stock and Watson (2003)). It is clear that consumption growth volatility has slowly declined over the sample period but the volatility of the market return did not. This pattern is apparent in both measures of conditional volatility. Second, it seems that there are both high frequency (business cycle) fluctuations and a very low frequency stochastic trend in consumption growth volatility. We will show that the estimation procedure mostly identifies these higher frequency movements in the conditional variance and not the very low frequency movements. Our hypothesis is that higher frequency fluctuations are channeled through the asset market while there are other aspects which we do not identify that contribute to the low frequency fluctuations. In other words, when we estimate the full model, the effect of the equity and bond market restrictions is reflected in the implied persistency of the conditional variance process. Here, we use the Hodrick-Prescott filter with parameter 1600 to disentangle these two components of consumption growth volatility. Figure 3.2 presents this result and makes clear that the decline in the low frequency component started in the $'60$, before the Great Moderation.[28]

[28]In our model it is hard to make 'conditional' statements about the economy, mainly because we modeled a constant drift to the consumption growth rate process. It is obviously interesting to think about the correlation structure of expected consumption growth rate and the conditional variance process. Empirically, there is evidence that suggests that interest rates are procyclical (e.g., Donaldson et al. (1990)) and volatility is either countercyclical or at least slightly leads expected growth rates which are believed to be countercyclical (e.g., Whitelaw (1994)). Our conditional variance process is assumed to correlate positively with realized consumption growth rate. Also, the conditional variance correlate negatively with interest rates. In this sense, variance and real interest rates behave as in the data. If, for example, expected growth rate correlate negatively with realized consumption growth rates, they will correlate negatively with the conditional variance. In that case, a positive shock to consumption growth rate will have a double negative effects on real interest rates. Expected growth rates will be low and thus the substitution effect will make equilibrium real interest rates lower. At the same time, conditional variance will be higher and the precautionary savings motive will push the equilibrium real interest rate even lower. Also, Chapman (1997) documented the strong positive correlation of real yields and consumption growth rate when excluding the monetary experiment period of $1979 - 1985$.

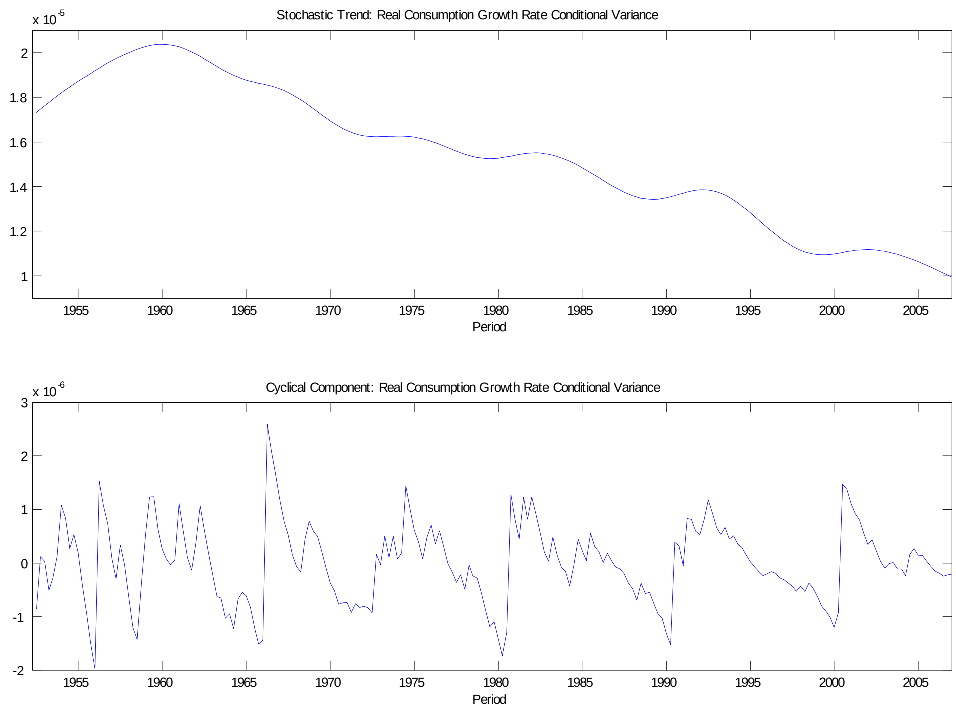

Figure 3.2: HP-flitered conditional variance of real consumption growth rate derived from an ARMAX-GARCH estimation in (3.31). The top panel presents the low frequency trend and bottom panel presents the cyclical component. The HP-filter parameter is 1600. The quarterly data is over the period $Q2.52 - Q4 - 06$.

We also use the volatility estimates to explain asset prices (see also Chapman (1997), Bansal and Yaron (2004), Bansal et al. (2005)). In particular, in figure 3.3 we examine the dynamic cross correlation patterns between consumption growth volatility obtained from the GARCH estimation in (3.31) and the spread between the real 1-year real yield and the real 3-months real yield.

These patterns agree with the model's predictions. We know that shorter maturity yields respond more than longer maturity yields to a volatility shock. This result is mainly

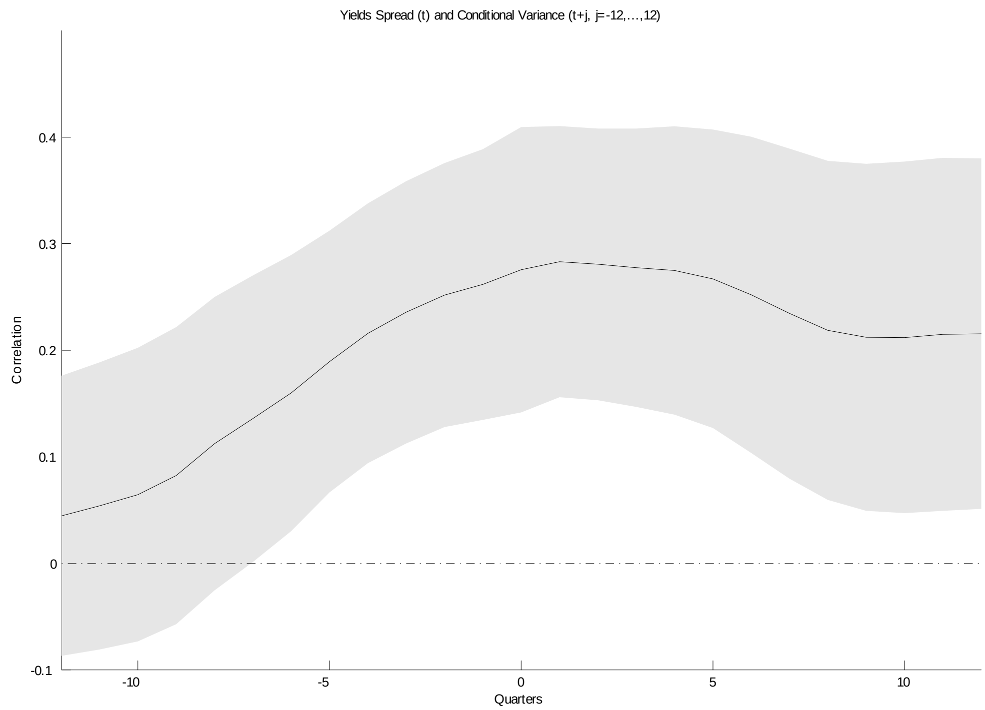

Figure 3.3: Dynamic cross-correlation between real consumption growth rate volatility and the real spread between the 1 year and 3 months yields. The quarterly data covers the period $Q2.52 - Q4.06$.

due to the ergodicity of the state variable that affect yields. If the state is assumed to revert back to a known steady state, we expect the longer yield to have a smaller response to contemporaneous shocks. Note that we do not identify the type of shock in this exercise. We merely observe a shock that happens to affect both consumption growth volatility and the bond market.

The second result is the sign response of the yields to a volatility shock. When conditional volatility increases we see that yields decrease. From the precautionary savings motive effect we do expect such response. Since in our model ambiguity aversion amplifies the precautionary savings motive, we expect this channel to play an important role when linking consumption growth volatility and yields. When combining these two results, we expect the spread to increase with a volatility shock. In other words, on average, the yield curve rotates when a shock to volatility occurs.

There are three caveats to these results. First, the upper left panel in Figure 3.1 depicts the behavior of the conditional variance of real consumption growth. One can argue that the series exhibit a non-stationary behavior. If this is the case, then the GARCH process is potentially misspecified. Given the slow-moving component we identified, it is hard to convincingly argue against such hypothesis. Second, our macro data is sampled at quarterly frequency. Drost and Nijman (1993) have shown that temporal aggregation impedes our ability to detect GARCH effects in the data. Even if our model is not misspecified, the fairly low frequency sampling may suggest it is (see also Bansal and Yaron (2004)). Third, we showed that the (sign of the) correlation between shocks to realized consumption growth and the conditional variance is important in explaining risk and uncertainty premia. The simple GARCH exercise does not help us identify the sign of this correlation. We address this difficulty next.

3.5.1.2. Real Dividends Growth Rate: GJR-GARCH. Since we argue that the sign of σ_v plays an important role in understanding risk premia in our model, we also estimate a GJR-GARCH$(1,1)$ (Glosten et al. (1993)). Originally, this model was constructed to capture 'leverage' effects when examining market returns (i.e., a negative shock to returns means lower prices and more leveraged firms, hence higher volatility of future returns). Here we use it with a different interpretation in mind. We use the leverage coefficient to extract information about the sign of the correlation between consumption/dividends growth rate innovations and conditional variance innovations. Since we argue that the sign of σ_v is positive, as indicated by asset prices behavior, we hope to find the reverse of a leverage effect.[29] We fit the following time series model

$$\begin{aligned} \frac{\Delta C_t}{C_{t-1}} &= c + \eta_{C,t}, \\ \eta_{C,t+1} &= \sigma_{C,t}\varepsilon_{C,t+1}, \quad \varepsilon_{C,t} \sim N(0,1), \\ D(L)\,\sigma_{C,t} &= \omega + F(L)\,\eta_{C,t}^2 + G(L)\,I_{\{\eta_{C,t}<0\}}\eta_{C,t}^2, \end{aligned} \tag{3.32}$$

where the polynomial G captures the leverage effects and

$$I_{\{\eta_{C,t}<0\}} = \begin{cases} 1 & \eta_{C,t} < 0 \\ 0 & \text{otherwise} \end{cases}.$$

We regress the realized consumption growth rate only on a constant (effectively demeaning the growth rate) since we assume in our model that dividends growth rate drifts on a constant. The more negative η is, the larger is η^2. Thus, we expect the leverage effect coefficient to be negative in order to capture the positive correlation between shocks to growth rates and conditional variance. In most lag specifications we estimated, the leverage coefficients in the G polynomial have a negative sign, which suggests that negative shocks to the dividends growth rate implies a negative shock to the conditional variance. However, and perhaps not surprisingly, with quarterly frequency data it is hard to detect these GARCH effects. Leverage effects are especially hard to detect. In most cases we

[29] Even though our interpretation has nothing to do with the leverage effect discussed in Glosten et al. (1993), we still use this term for convenience.

cannot reject the null that leverage effects are not present. In order to investigate the sign of σ_v further, we use real dividends instead of consumption. To alleviate the problem with the GARCH estimation, we use *monthly* data.[30] Figure 3.4 displays the results of a GJR-GARCH$(1,1)$ estimation where c is the unconditional mean of the real growth rate of aggregate dividends

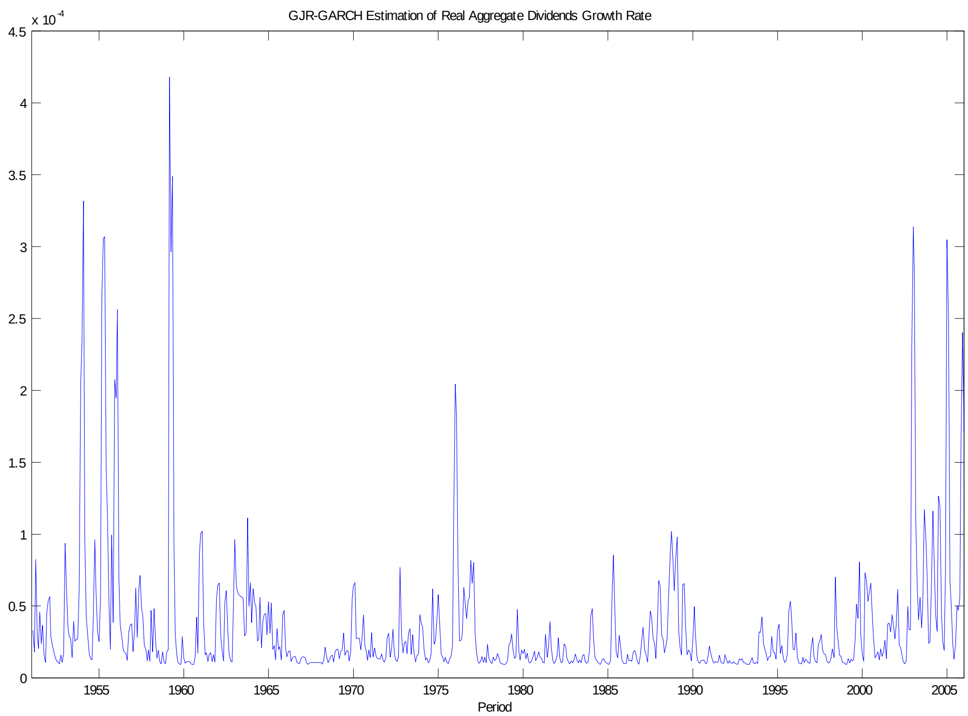

Figure 3.4: GJR-GARCH$(1,1)$ estimation (model 3.32) of the conditional variance of real aggregate dividends growth rate with monthly observations over the period $M1.52 - M4.06$.

This figure shows the presence of volatility clustering. The estimation procedure suggests that σ_v is indeed positive since the leverage coefficient is *always* negative and

[30]We obtained the real dividends series from Robert Shiller's website. See also Appendix 9.

statistically significant. On average, when a negative shock hits the dividends growth rate, we tend to see a decline in the conditional variance of the same process. Table 3.5.1.2 summarizes the estimation results for the leverage coefficient over different time intervals.[31]

Table 3.1: Estimating the 'leverage' coefficient over different time intervals.

Period	'Leverage' Coefficient	Standard Errors
$M1.52 - M12.06$	-0.390	0.129
$M1.62 - M12.06$	-0.242	0.175
$M1.72 - M12.06$	-0.312	0.215
$M1.82 - M12.06$	-0.263	0.163
$M1.90 - M12.06$	-0.256	0.195
$M1.52 - M12.81$	-0.509	0.167
$M1.52 - M12.89$	-0.442	0.156

The data is monthly real aggregate dividends over $M1.52 - M12.06$ from Robert Shiller's website. A negative point estimate means that a negative shock to realized dividends growth rate is accompanied by a negative shock to the conditional variance of dividends growth rate.

It is interesting to note that the earlier post-war data supports more strongly the hypothesis that shocks to dividends are positively correlated with shocks to volatility. This covariation measures the risk exposure of default free bonds to risk and uncertainty. If the market prices of these risks and uncertainty did not move in the opposite direction one should, ceteris paribus, expect to observe higher risk premia in the earlier part of the sample.

In summary, the data seems to confirm two things. First, the existence of a small time-varying component in the volatility of growth rates. Second, the correlation of shocks to dividends growth rate and shocks to conditional variance is positive.

[31] This suggestive evidence is also consistent with different time intervals and with EGARCH estimation (see Nelson (1991)) over the same time intervals. Results are available from the authors upon request.

3.5.2. Model Estimation

In this section we present and interpret our complete model estimation results. Since the model permits closed-form expressions for first and second moments we use the generalized method of moments (GMM) in the estimation procedure (Hansen (1982)). Even though conditional variance is not directly observable in the data it is theoretically an affine function of the short rate (or any other real yield with arbitrary maturity). Therefore, we use the short rate as an observable that completely characterizes the behavior of the conditional variance.[32] We also compare the moments implied by the model to their empirical counterparts.

3.5.2.1. Orthogonality Restrictions and Identification. Our procedure is similar to the one used by, for example, Chan et al. (1992). Our approach is to focus mainly on the time series restrictions to estimate the structural parameters. We do not focus on the cross sectional restrictions of the model as in Longstaff and Schwartz (1992) and Gibbons and Ramaswamy (1993). Since we have a single factor model, yields are perfectly correlated. Therefore, including cross sectional restrictions may reduce the power of the overidentifying restrictions in small samples. We use our point estimates to generate the model's implied yield curve and compare it to the empirical yield curve. In that sense, our approach is more ambitious. It is important to note that since our model only makes statements about the real economy, all the data we use is denominated in real

[32]We also used the simulated method of moments (SMM, Duffie and Singleton (1993)) to estimate the model. This method is natural when the model contains unobservables. The results we obtain using SMM support the results we obtain using GMM and are available from the authors upon request. Bansal et al. (2007) also compare their GMM estimates to an SMM estimates and conclude that in the presence of time averaging, using SMM can prove useful.

terms. Other authors have used nominal data to estimate real models (e.g., Brown and Dybvig (1986)).

We need to estimate 6 parameters $\{a_0, a_1, \mu, \rho, \theta, \sigma_v\}$. We form orthogonality conditions implied by the model using the following notation

$$\begin{aligned} Y_{t+1} &\equiv \left[\Delta\mathcal{Y}\left(1; v_{t+1}\right), R_{t+1}, \frac{\Delta C_{t+1}}{C_t}, \mathcal{Y}\left(4; v_t\right) - \mathcal{Y}\left(1; v_t\right)\right], \\ X_t &\equiv \mathcal{Y}\left(1; v_t\right), \\ Z_t &\equiv \left[1, \mathcal{Y}\left(1; v_t\right), R_t, \frac{\Delta C_t}{C_{t-1}}\right], \end{aligned}$$

where Y_{t+1} is observed at time $t+1$ and contains the change in the one-quarter real yield ($\Delta\mathcal{Y}\left(1; v_{t+1}\right)$), the realized real aggregate market return (R_{t+1}), the realized real aggregate consumption growth rate $\left(\frac{\Delta C_{t+1}}{C_t}\right)$ and the real spread between the 1-year and 3-months real yields ($\mathcal{Y}\left(4; v_t\right) - \mathcal{Y}\left(1; v_t\right)$). X_t is the explanatory factor. We use the 3-months yield as a sufficient statistic for the unobserved conditional variance process. Last, we use lagged 3-months, market return and realized consumption growth rate as instruments in the vector Z_t.

The stacked orthogonality conditions are given in m

$$\begin{aligned} u_{1,t+1} &\equiv Y_{t+1} - \mu_{Y,t} | X_t, \\ u_{2,t+1} &\equiv diag\left(u_{1,t+1} u'_{1,t+1} - \sigma_{Y,t} \sigma'_{Y,t} | X_t\right), \\ m_{t+1} &\equiv \left[\begin{matrix} u_{1,t+1} & u_{2,t+1} \end{matrix}\right] \otimes Z_t. \end{aligned}$$

We draw first and second moment restrictions. $\mu_{Y,t}|X_t$ and $\sigma_{Y,t}|X_t$ have the parametric forms implied by the model and are affine in X_t.

What about identification? Note that since robust preferences are observationally equivalent to recursive preferences, disentangling the risk aversion coefficient from the robustness parameter θ is generally not trivial. Since we have log-utility we do need to worry about such a potential identification problem: log preferences restrict to unity the EIS and risk aversion and thus allow us to identify the uncertainty parameter θ. Also, μ is identified through the consumption growth rate restriction. Once μ is identified, we can identify ρ from the aggregate market return condition. The three parameters that govern the dynamics of the conditional variance v can be identified either from the second moment of consumption growth rate or the second moment of the aggregate market return. Also, the bond market contributes important information about v. The fact that identifying the dynamics of v is done through these three channels can potentially create some ambiguity in the interpretation of the level and speed of reversion of the conditional variance. Nevertheless, we believe that these sources of information shed some new light on the dynamics of v in a way that will be clear in our interpretation of the point estimates, a task we undertake next.

3.5.2.2. Point Estimates of Structural Parameters. Table 3.2 presents the point estimates over different time periods. In Table 3.3 we perform the same estimation exercise without including the volatility of consumption growth rate in our set of moments.

Table 3.2: Model estimation with consumption volatility over different time intervals.

Period	T	a_0	a_1	μ	ρ	θ	σ_v	J-stat/P-val	DEP
$Q2.52 - Q4.06$	218	0.0005	−0.1951	0.0052	0.0145	9.5730	0.0189	33.5147	4.41%
		6.7904	−10.1242	29.0936	7.6824	3.7579	7.7160	14.77%	
$Q1.62 - Q4.06$	180	0.0004	−0.1594	0.0051	0.0125	14.9578	0.0175	28.0792	11.01%
		6.9588	−10.7284	28.2098	6.3202	2.7028	5.5972	35.46%	
$Q1.72 - Q4.06$	140	0.0004	−0.1611	0.0048	0.0148	11.8660	0.0149	22.3470	10.86%
		6.3636	−10.3456	28.4321	6.1634	2.5565	5.1858	66.96%	
$Q1.82 - Q4.06$	100	0.0004	−0.1314	0.0054	0.0208	6.6591	0.0082	17.3000	7.31%
		6.2872	−6.6014	34.0290	8.7880	3.7862	7.4882	89.97%	
$Q1.90 - Q4.06$	68	0.0003	−0.0997	0.0050	0.0172	9.0779	0.0064	12.2261	14.49%
		8.1423	−9.7387	36.4052	10.5310	5.7235	10.3818	98.98%	
$Q2.52 - Q4.81$	118	0.0007	−0.2528	0.0050	0.0115	13.4790	0.0431	20.2067	17.11%
		7.1360	−13.8242	20.8878	5.6148	2.7839	6.0434	78.17%	
$Q2.52 - Q4.89$	149	0.0006	−0.2232	0.0053	0.0140	10.6126	0.0290	24.2714	10.10%
		6.2481	−9.5348	24.3649	6.3682	3.2241	6.9336	56.04%	

The data is in quarterly frequency and in quarterly values. 6 parameters are estimated using an iterated GMM. There are 8 moments and 4 instruments that produce 32 orthogonality conditions. T is the number of observation in each estimation. Robust t-statistics are indicated below each point estimate. The standard error are corrected using the Newey-West procedure with 4 lags. p-val is the p-value for the J-test statistic distributed χ^2 with 26 degrees of freedom. The DEP column reports the detection error probabilities.

Table 3.3: Model estimation without consumption volatility over different time intervals.

Period	T	a_0	a_1	μ	ρ	θ	σ_v	J-stat/P-val	DEP
$Q2.52 - Q4.06$	218	0.0011	−0.1903	0.0052	0.0159	23.5142	0.0331	26.4281	15.56%
		6.0765	−9.0950	25.1487	7.0689	3.1259	6.7880	23.38%	
$Q1.62 - Q4.06$	180	0.0009	−0.1666	0.0051	0.0134	39.9255	0.0335	25.7987	24.97%
		5.6809	−10.0265	25.9144	5.9769	2.1620	5.3863	26.04%	
$Q1.72 - Q4.06$	140	0.0010	−0.1719	0.0048	0.0157	31.0175	0.0286	20.3411	24.16%
		4.8876	−9.1749	24.2349	5.9522	2.0961	5.0048	56.17%	
$Q1.82 - Q4.06$	100	0.0008	−0.1336	0.0054	0.0214	15.9299	0.0159	16.4269	18.93%
		5.2730	−6.4625	31.9802	7.9282	2.9953	6.8526	79.42%	
$Q1.90 - Q4.06$	68	0.0006	−0.1006	0.0050	0.0172	22.3478	0.0125	11.7640	27.01%
		7.4532	−7.9747	32.3043	9.4070	4.0868	9.7618	96.21%	
$Q2.52 - Q4.81$	117	0.0013	−0.2332	0.0050	0.0119	34.7236	0.0749	18.0621	30.07%
		6.6457	−10.2125	18.5634	5.4752	2.3172	5.5378	70.23%	
$Q2.52 - Q4.89$	149	0.0012	−0.2069	0.0053	0.0148	26.6319	0.0530	20.8378	23.44%
		5.5588	−8.0626	21.2137	5.9463	2.7119	6.3294	53.08%	

The data is in quarterly frequency and in quarterly values. 6 parameters are estimated using an iterated GMM. There are 8 moments and 4 instruments that produce 32 orthogonality conditions. T is the number of observation in each estimation. Robust t-statistics are indicated below each point estimate. The standard error are corrected using the Newey-West procedure with 4 lags. p-val is the p-value for the J-test statistic distributed χ^2 with 26 degrees of freedom. The DEP column reports the detection error probabilities.

Aside from the robustness parameter θ, all coefficients are immediately interpretable. All parameters are statistically different from zero. Also, the model is not being rejected according to the J-test. We will explain the DEP's column later.

Note that μ is stable and equal to the average real aggregate quarterly consumption growth rate over the sample. Similarly, ρ is stable over different samples and invariant to the consumption volatility restriction.

One obvious finding is that the estimated θ is sensitive to the inclusion of consumption volatility in the estimation. When the volatility of consumption growth is ignored, the procedure is not restricted by the smooth consumption process and thus the implied pricing kernel (SDF) is much more volatile and more robustness is not needed to justify the observed asset prices. In this sense, the implied volatility of the SDF is closer to the Hansen-Jagannathan bound. Also, note that a_0 and σ_v are much larger when we do not impose the consumption volatility restriction. The reason is that the procedure mainly picks up the aggregate market return volatility, which is much larger than the volatility of consumption growth rate. The implied evolution of v is much more volatile when consumption growth volatility is excluded.

Interestingly, the velocity of reversion $(-a_1)$ of v is invariant to the consumption growth rate volatility. What is obvious from Tables 3.2 and 3.3 is that the estimation procedure detects mostly high frequency movements and not the potential slow moving component in consumption growth volatility we identified earlier (Figure 3.2). Hence, it appears that the high-frequency component from the market data dominates in the full-model estimation.

Panel A of Table 3.5.2.2 presents the half life of the volatility shock process implied by the estimation procedure. We also present in that panel the perceived half life by the robust agent. Expression (3.29) shows that the perceived velocity of mean reversion is faster than the physical speed at which shocks to volatility die out. In general, the point estimates imply that shocks to volatility die out relatively fast. For comparison purposes, Panel B of Table 3.5.2.2 presents the implied reversion coefficient and half life derived using the autoregressive coefficient we calculated from the GARCH estimated conditional variance series in (3.31) without adding the market as an explanatory variable to the consumption growth rate.[33] These results confirm that without forcing asset market restrictions on the consumption series, we observe a very slow moving process for conditional variance. At the same time, the conditional variance of the aggregate market return is much less persistent. The general estimation procedure results in panel A are, to some extent, a combination of these two effects.[34]

Interestingly, in our benchmark estimation result we find the half life of the conditional variance process to be 3.553 quarters. The recent 'long run risks' literature usually calibrates asset pricing models with a highly persistent conditional variance process.[35]

[33]Our point estimates correspond to quarterly data. In general, with data sampled at quarterly frequency one can map an autoregressive coefficient to a coefficient governing the speed of reversion as our κ_v. Let $\hat{\alpha}$ denote the autoregressive coefficient. Then, the quarterly speed of reversion coefficient $\kappa_v = -\ln(\hat{\alpha})$ and the half life is $\ln(2)/\kappa_v$.

[34]We conduct this comparison only for the entire period $Q2.52 - Q4.06$ since we want to examine evidence concerning very low frequency components. Even our longest sample is somewhat short to conveniently detect the slow moving component. We believe that shorter samples will make the detection exercise impossible.

[35]Bansal and Yaron (2004) find that introducing a small highly persistent predictable component in consumption growth can attenuate the high risk aversion implications of standard asset pricing models with recursive utility preferences. However, this persistent component is difficult to detect in the data. Croce et al. (2006) present a limited information economy where agents face a signal extraction problem. Their model addresses the identification issues of the long run risk component. Hansen and Sargent (2007b) is another example for the difficulty in identifying the long run risk component. However, in

Table 3.4: Mean Reversion.

Panel A:	$Q2.52 - Q4.06$		$Q1.90 - Q4.06$	
	Estimate	Half Life (Q)	Estimate	Half Life (Q)
Objective	0.1951	3.553	0.0997	6.952
Distorted	0.2994	2.315	0.1343	5.161
Panel B:	$Q2.52 - Q4.06$			
	Estimate		Half Life (Q)	
Consumption	0.010		68.373	
Market	0.4069		1.704	

Panel A: Point estimates of the velocity of reversion coefficient and the implied half life (in quarters) of the conditional variance process. Objective referes to the physical rate in which the conditional variance gravitates to its steady state. Distorted referes to the rate in which the robust agent believes the conditional variance gravitates to its steady state. These point estimates are from the estimation procedure that imposes the volatility of real aggregate consumption growth rate as a moment condition. Panel B: implied reversion coefficients and half lifes (in quarters) for the conditional volatility of consumption growth rate and aggregate market return derived from the GARCH procedure. The consumption growth rate mean is modeled as an ARMAX(2,2,1) and the aggregate market return is modeled as ARMA(2,2).

For example, Bansal and Yaron (2004) assume that the autoregressive coefficient (with monthly frequency data) in the conditional variance of the consumption growth process is 0.987.[36] This number implies a half life of 13.24 quarters, which is almost 4 times higher than the number we obtain in our empirical results. As explained earlier, this difference is driven largely by the inclusion of equity and bond markets in our set of moments. What we show in this paper is that robust decision making coupled with state dependent volatility requires moderate levels of persistence in the conditional variance of the consumption growth process. Recall that we assume a constant drift in consumption growth. If we assume a stochastic and highly persistent μ, as in Bansal and Yaron (2004), we would need

addition to a signal extraction problem, their agent seeks robust policies and consequently his estimation procedure is modified.

[36]See table IV in Bansal and Yaron (2004).

to worry about the volatility of the risk free rate. In other words, if the substitution effect channel is very persistent and the precautionary savings motive is much less persistent, the short rate can potentially be very volatile. If shocks to μ were to die out much slower than shocks to v, the ergodic distribution of the short rate would be very volatile. In that sense, we might be able to reconcile our results with the calibration exercise of Bansal and Yaron (2004) if we assumed an expected consumption growth rate process.

Expressions (3.29) and (3.30) allow us to discuss a mechanism which is central to our results. In Figure 3.5 we plot the objective and perceived impulse response functions for the conditional variance v following a shock. Note that, unlike a rational expectations agent, the robust agent is on average wrong about the future evolution of v. Hence, his biased expectations lead him to believe that the conditional variance will decrease. As mentioned earlier, this should lead to an upward-sloping unconditional sloping yield curve through the precautionary savings channel.

3.5.2.3. Theoretical and Empirical Moments. Table 3.5 presents a comparison of model-implied and empirical moments over different time spans for the equity and consumption data. Table 3.6 presents the same exercise, but without imposing consumption growth rate volatility in the estimation. The model fares well, especially with the aggregate market return and the equity premium.[37] Also, the model is doing a good job in matching the low consumption growth rate. The same conclusion seems to hold over different time horizons. Note, however, that again we see the tension between market return and consumption growth volatility. When imposing consumption growth volatility,

[37]Erbas and Mirakhor (2007) document global evidence (53 emerging and mature markets) that a large part of the equity premium reflects investor aversion to ambiguities resulting from institutional weaknesses.

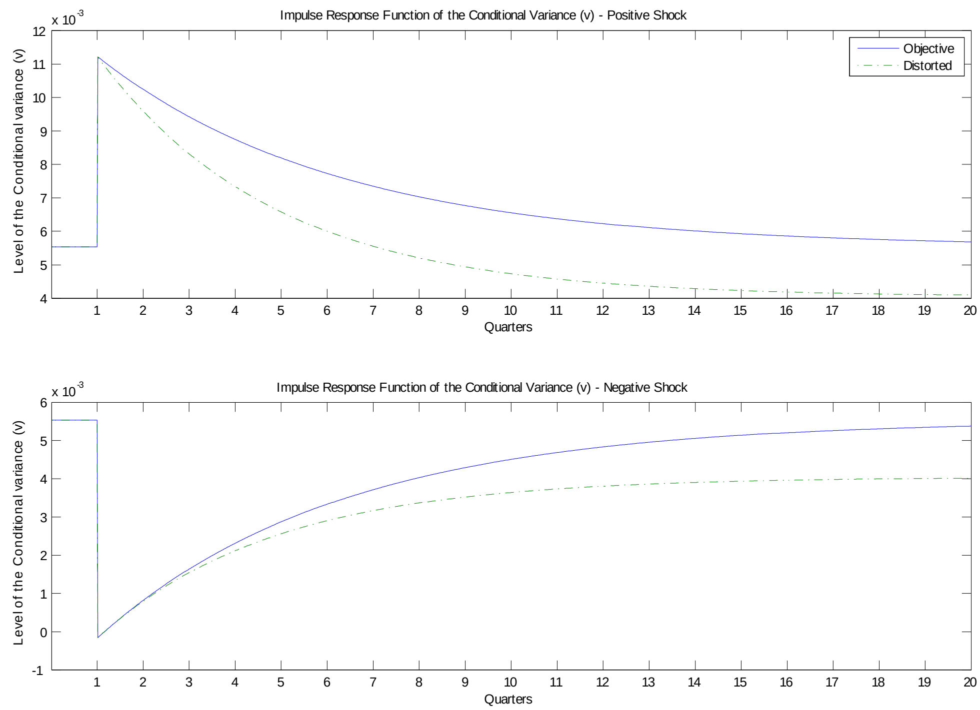

Figure 3.5: Biased expectations. Using the parameters estimated over the entire period $Q2.52 - Q4.06$, the figure shows the impulse response function of the conditional variance to a positive and negative shocks. The solid line represents the objective evolution of v and the dashed line represents what the robust agent believes the evolution of v is going to be.

the model compromises on the implied market return volatility being somewhere between the empirical consumption growth rate volatility and the empirical market return volatility. When ignoring the consumption growth volatility from the estimation procedure, the model easily matches the aggregate market return volatility. This result is obviously not surprising since we have a log-agent that consumes a constant fraction of his wealth. Given that the substitution effect and the income effects cancel each other, the agent absorbs all market fluctuations to his marginal utility.

Table 3.5: Empirical and theoretical equity and goods market moments (with consumption volatility restriction).

Period	T	μ_R		$\mu_R - \mathcal{Y}_{3m}$		σ_R		μ_C	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
$Q2.52 - Q4.06$	218	12.820	13.692	11.289	11.719	34.306	22.385	2.109	2.924
		2.321	0.024	2.350	0.034	2.164	0.038	0.174	0.022
$Q1.62 - Q4.06$	180	11.879	12.586	10.130	10.552	35.499	22.748	2.085	2.940
		2.544	0.028	2.562	0.039	2.487	0.044	0.188	0.025
$Q1.72 - Q4.06$	140	12.611	12.985	10.925	10.879	36.599	23.302	1.937	2.882
		3.021	0.033	3.029	0.044	2.865	0.045	0.204	0.030
$Q1.82 - Q4.06$	100	15.424	14.614	13.234	12.218	35.510	24.372	2.206	3.259
		3.240	0.039	3.206	0.051	2.941	0.041	0.178	0.035
$Q1.90 - Q4.06$	68	13.060	12.298	11.393	10.579	34.075	23.315	2.035	2.971
		3.887	0.044	3.807	0.057	3.753	0.046	0.203	0.040
$Q2.52 - Q4.81$	118	10.749	17.617	9.784	16.053	32.983	24.751	2.004	2.977
		3.155	0.042	3.276	0.062	3.111	0.097	0.279	0.036
$Q2.52 - Q4.89$	149	12.710	15.950	11.241	13.842	34.411	23.964	2.143	3.086
		2.905	0.034	2.976	0.048	2.642	0.063	0.236	0.029

The period column represents the time interval of the data that is used to estimate the model. The data is in quarterly frequency with quarterly values. T is the number of quarterly observations used to estimate the model. Columns with the number (1) present the empirical moments. Empirical moments computed with the data and theoretical moments are implied by the estimated model. Columns with the number (2) present the theoretical moments. The theoretical moments were generated using $1,000$ replications of the economy that was calibrated using the estimated parameters over the corresponding period. Robust standard errors are given below each moment. The standard errors were corrected using the Newey-West procedure with 4 lags. The standard errors for the theoretical moments were computed over the $1,000$ replications. All moments are given in % values. μ_R, μ_C, σ_R, and $\mathcal{Y}_{3m}$ are the real return on the market (including dividends), real growth rate of consumption, volatility of real aggregate market return and real 3 month yield, respectively.

Tables 3.7 and 3.8 report model implied and empirical moments for the bond market, where the second table ignores the volatility of consumption growth in our set of moments. The model is doing a good job in reproducing the levels of the 3-months and 1-year real yields. The results in the last two columns of each table are particularly interesting. The

Table 3.6: Empirical and theoretical equity and goods market moments (without consumption volatility restriction).

Period	T	μ_R		$\mu_R - \mathcal{Y}_{3m}$		σ_R		μ_C	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
$Q2.52 - Q4.06$	218	12.820	19.564	11.289	17.682	34.306	36.566	2.109	4.110
		2.321	0.043	2.350	0.056	2.164	0.080	0.174	0.037
$Q1.62 - Q4.06$	180	11.879	18.579	10.130	16.616	35.499	36.742	2.085	4.117
		2.544	0.050	2.562	0.064	2.487	0.095	0.188	0.043
$Q1.72 - Q4.06$	140	12.611	18.648	10.925	16.713	36.599	37.482	1.937	4.139
		3.021	0.057	3.029	0.072	2.865	0.094	0.204	0.049
$Q1.82 - Q4.06$	100	15.424	18.802	13.234	16.548	35.510	37.124	2.206	4.496
		3.240	0.063	3.206	0.079	2.941	0.081	0.178	0.056
$Q1.90 - Q4.06$	68	13.060	15.531	11.393	13.782	34.075	34.177	2.035	3.992
		3.887	0.070	3.807	0.087	3.753	0.087	0.203	0.062
$Q2.52 - Q4.81$	117	10.592	25.701	9.625	24.142	33.078	41.079	2.025	4.257
		3.204	0.083	3.327	0.109	3.132	0.231	0.278	0.067
$Q2.52 - Q4.89$	149	12.710	23.824	11.241	21.822	34.411	39.497	2.143	4.347
		2.905	0.062	2.976	0.083	2.642	0.148	0.236	0.052

The period column represents the time interval of the data that is used to estimate the model. The data is in quarterly frequency with quarterly values. T is the number of quarterly observations used to estimate the model. Columns with the number (1) present the empirical moments. Empirical moments computed with the data and theoretical moments are implied by the estimated model. Columns with the number (2) present the theoretical moments. The theoretical moments were generated using $1,000$ replications of the economy that was calibrated using the estimated parameters over the corresponding period. Robust standard errors are given below each moment. The standard errors were corrected using the Newey-West procedure with 4 lags. The standard errors for the theoretical moments were computed over the $1,000$ replications. All moments are given in % values. μ_R, μ_C, σ_R and $\mathcal{Y}_{3m}$ are the real return on the market (including dividends), real growth rate of consumption, volatility of real aggregate market return and real 3 month yield, respectively.

second to last column $(\rho(\mathcal{Y}_{3m}))$ reports the first-order autocorrelation of the 3-months yield. Note that we do not impose this restriction in our estimation and yet the model is able to produce this moment with high accuracy. This information is indirectly encoded into the orthogonality conditions though the imposition of the change in the 3-months yield. The last column captures the holding period returns of a strategy that dictates buying a 1-year bond and selling it after 3 quarters. Backus et al. (1989) point to the difficulty of representative agent models to account for both the sign and magnitude of holding period returns in the bond market. Again we note that we did not impose any holding period returns conditions in the estimation procedure and yet the model captures the returns dynamics well. Nevertheless, we should note that by imposing the spread and the change in the short rate conditions, we provide the estimation procedure with enough information about the dynamics of the 1-year and 3-months yields to the extent that the holding period returns are captured accurately by the model.

The top panel in Figure 3.6 presents estimation results over the years $'97-'06$. During this period TIPS bonds were traded in the U.S. and thus provide a good proxy to real yields. The solid line is the average level of the yield curve over this period with 95% confidence bands. The dot-dashed line is the model-implied average yield curve. Note that we only impose two bond market restrictions in the estimation procedure and yet the model can closely imitate the behavior of the entire yield curve (within the confidence bands). The bottom panel depicts the term structure of the volatilities of yields. Clearly, the model can replicate the downward slope due to the mean reversion in the estimated conditional variance process, as discussed earlier. The impression is that the procedure anchors the implied first and second moments of the 1-year yield to its empirical counterpart, but it is still doing a good job in approximating the entire curve.

As discussed earlier, our model can reconcile the difficulty one factor models face when trying to match both the high persistence of yields and the high convexity of the curve. Figure 3.6 shows that the agent prices the yield curve as if shocks to v die out fast.

Table 3.7: Empirical and theoretical bond market moments (with consumption volatility restriction).

Period	T	$\mathcal{Y}_{3m}$		$\mathcal{Y}_{1y}$		$\rho\left(\mathcal{Y}_{3m}\right)$		$\ln\left[\frac{p(1;v_{t+3})}{p(4;v_t)}\right]$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
$Q2.52 - Q4.06$	218	1.531	1.973	2.250	2.659	0.843	0.863	2.465	2.888
		0.263	0.010	0.234	0.007	0.060	0.000	0.274	0.006
$Q1.62 - Q4.06$	180	1.749	2.033	2.241	2.501	0.835	0.881	2.394	2.658
		0.292	0.011	0.279	0.008	0.071	0.000	0.321	0.007
$Q1.72 - Q4.06$	140	1.686	2.106	2.214	2.601	0.830	0.876	2.359	2.766
		0.362	0.012	0.351	0.008	0.075	0.001	0.406	0.007
$Q1.82 - Q4.06$	100	2.190	2.396	2.742	2.926	0.862	0.888	2.932	3.105
		0.366	0.012	0.378	0.009	0.062	0.001	0.443	0.009
$Q1.90 - Q4.06$	68	1.666	1.719	2.015	2.073	0.897	0.892	2.118	2.190
		0.422	0.013	0.371	0.011	0.058	0.001	0.420	0.010
$Q2.52 - Q4.81$	118	0.965	1.565	1.865	2.467	0.804	0.800	1.976	2.768
		0.330	0.020	0.247	0.011	0.096	0.001	0.275	0.008
$Q2.52 - Q4.89$	149	1.469	2.107	2.358	2.934	0.823	0.834	2.586	3.211
		0.328	0.015	0.292	0.009	0.077	0.001	0.348	0.007

The period column represents the time interval of the data that is used to estimate the model. The data is in quarterly frequency with quarterly values. T is the number of quarterly observations used to estimate the model. Columns with the number (1) present the empirical moments. Empirical moments computed with the data and theoretical moments are implied by the estimated model. Columns with the number (2) present the theoretical moments. The theoretical moments were generated using $1,000$ replications of the economy that was calibrated using the estimated parameters over the corresponding period. Robust standard errors are given below each moment. The standard errors were corrected using the Newey-West procedure with 4 lags. The standard errors for the theoretical moments were computed over the $1,000$ replications. All moments, aside from the autocorrelations, are given in % values. $\mathcal{Y}_{3m}$, $\mathcal{Y}_{1y}$, and $\rho\left(\mathcal{Y}_{3m}\right)$ are the real 3 month yield, real 1 year yield and the first order autocorrelation coefficient of the real 3 month yield, respectively. The last column reports real holding period return for buying a one year to maturity bond and selling it after three quarters.

However, Tables 3.7 and 3.8 confirm that the model can still match the persistence of the short rate. Empirically, all yields exhibit the same level of persistence.[38]

[38]The term structure literature usually identifies 3 factors that account well for most of the variation in the yield curve (Litterman and Scheinkman (1991)): level, slope and curvature. The level slope is very persistent and, thus, accounts for most of the observed persistence of yields.

Table 3.8: Empirical and theoretical bond market moments (without consumption volatility restriction).

Period	T	$\mathcal{Y}_{3m}$		$\mathcal{Y}_{1y}$		$\rho\left(\mathcal{Y}_{3m}\right)$		$\ln\left[\frac{p(1;v_{t+3})}{p(4;v_t)}\right]$	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
$Q2.52 - Q4.06$	218	1.531	1.882	2.250	2.525	0.843	0.864	2.465	2.739
		0.263	0.013	0.234	0.009	0.060	0.000	0.274	0.008
$Q1.62 - Q4.06$	180	1.749	1.962	2.241	2.417	0.835	0.874	2.394	2.570
		0.292	0.014	0.279	0.010	0.071	0.001	0.321	0.009
$Q1.72 - Q4.06$	140	1.686	1.935	2.214	2.422	0.830	0.868	2.359	2.585
		0.362	0.015	0.351	0.011	0.075	0.001	0.406	0.010
$Q1.82 - Q4.06$	100	2.190	2.253	2.742	2.777	0.862	0.885	2.932	2.955
		0.366	0.017	0.378	0.013	0.062	0.001	0.443	0.012
$Q1.90 - Q4.06$	68	1.666	1.749	2.015	2.092	0.897	0.891	2.118	2.205
		0.422	0.018	0.371	0.015	0.058	0.001	0.420	0.014
$Q2.52 - Q4.81$	117	0.967	1.559	1.828	2.381	0.806	0.800	1.963	2.650
		0.333	0.027	0.237	0.016	0.096	0.001	0.278	0.013
$Q2.52 - Q4.89$	149	1.469	2.002	2.358	2.798	0.823	0.837	2.586	3.065
		0.328	0.021	0.292	0.014	0.077	0.001	0.348	0.011

The period column represents the time interval of the data that is used to estimate the model. The data is in quarterly frequency with quarterly values. T is the number of quarterly observations used to estimate the model. Columns with the number (1) present the empirical moments. Empirical moments computed with the data and theoretical moments are implied by the estimated model. Columns with the number (2) present the theoretical moments. The theoretical moments were generated using $1,000$ replications of the economy that was calibrated using the estimated parameters over the corresponding period. Robust standard errors are given below each moment. The standard errors were corrected using the Newey-West procedure with 4 lags. The standard errors for the theoretical moments were computed over the $1,000$ replications. All moments, aside from the autocorrelations, are given in % values. $\mathcal{Y}_{3m}$, $\mathcal{Y}_{1y}$, and $\rho\left(\mathcal{Y}_{3m}\right)$ are the real 3 month yield, real 1 year yield and the first order autocorrelation coefficient of the real 3 month yield, respectively. The last column reports real holding period return for buying a one year to maturity bond and selling it after three quarters.

3.5.3. 'Disciplining Fear': Detection Error probabilities

In this section, we undertake the task of interpreting θ. We showed so far that the model can account for different asset pricing facts and puzzles. Nevertheless, we have yet to

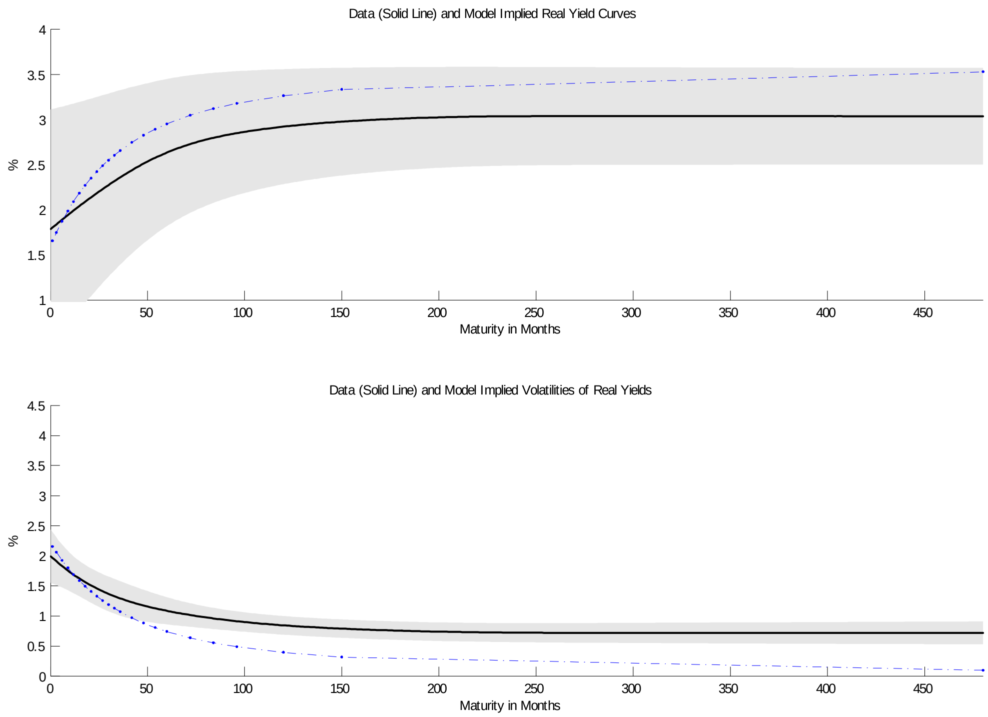

Figure 3.6: Top panel: average real yield curve extracted from the TIPS data from $M1.97 - M12.06$ (solid line) with 95% confidence bands with Newey-West (12 lags) correction. Model implied average yield curve (dot-dashed line). The model is estimated over the same period as the empirical yield curve. Bottom panel: empirical term structure of unconditional volatilities of the TIPS data (solid line). with 95% confidence bands with Newey-West (12 lags) correction. The model is estimated over the same period as the empirical yield curve.

tackle an important question - does the model imply too much uncertainty aversion? Even though we showed that coefficients of relative risk aversion and elasticity of intertemporal substitution of unity are sufficient, we still need to gauge the amount of ambiguity aversion implied by the data. Detection error probabilities (DEP's) are the mechanism through which we can interpret θ, and consequently, assess the amount of ambiguity aversion implied by our estimation.

In order to quantify ambiguity aversion, we ask the following: when the agent examines the (finite amount of) data available to him and has to decide whether the reference or the distorted model generated the data, what is the probability of making a model detection mistake? If the probability is very low, this indicates that the two models are far apart statistically, and that the agent should easily be able to distinguish between them. In this case, one might be led to conclude that the degree of robustness implied by our estimation is unreasonably high. If to the contrary, the DEP is high, then it is reasonable to believe that the agent would find it difficult to determine which model is the true representation of the economy.[39]

Technically, DEP's are a mapping from the space of structural parameters to a probability space, which is inherently more easily interpretable than parameter values. Based on our estimate of the parameter θ, we infer the detection error probabilities from the data. It then allows us to interpret whether the degree of ambiguity aversion in our parameterization seems excessive. Appendix 10 details how to derive the DEP for a given economy using simulations.

The last column in tables 3.2 and 3.3 presents the implied DEP's in each economy. First, it is important to point out that DEPs have to be between 0% and 50% (if both models are the same, then there is a 50% chance of making a mistake when assessing which model is the true one). What we find is that our implied DEPs are definitely not unreasonably small, particularly in the context of a framework where the only source of uncertainty is a single shock. This is, once again, an outcome of the interaction between the two main building blocks of our model - robust decision making and state dependent

[39]For an elaborate discussion of DEP's see, for example, Anderson et al. (2003) and Barillas et al. (2007). For a textbook treatment of robustness and DEP's see chapters 9 and 10 in Hansen and Sargent (2007a).

volatility. Together they imply a high enough market price of risk and uncertainty, and in fact with stochastic volatility the agent does not need to distort the reference model 'too much'. Therefore, the DEPs are sufficiently large.

The lowest DEP is for our benchmark model. This is not surprising for two reasons. First, we use the longest possible sample, making it easier for the agent/econometrician to distinguish between the objective and distorted models. Second, imposing (the very low) consumption growth rate volatility restricts the implied volatility of the SDF severely. Therefore, the model implies a stronger distortion in a way that enables us to achieve the Hansen-Jagannathan bound. When either the number of observations is smaller or we ignore the consumption growth rate volatility, the DEP increases.

Figure 3.7 present two comparative statics exercises on the implied DEPs. The left panel fixes the benchmark model and varies only θ. The right panel introduces variation only in the number of observations available to the econometrician. We see a clear pattern: Higher θ means less robustness. Thus, it becomes harder to statistically distinguish between the reference and the distorted models as the agent distorts less and less. As $\theta \to \infty$ the DEP reaches 0.5. This is not surprising, since $\theta = \infty$ implies that the distortion to the DGP is zero (recall (3.23)) and both models are therefore indistinguishable. On the other hand, a lower value of θ means more robustness and the models become statistically distant from each other (in the relative entropy sense), reflected in a lower DEP. Similarly, more observations reduce the DEP, in line with our earlier intuition.

3.5.3.1. The Evolution of 'Fear'. In this subsection we document the way fear of model misspecification evolved over time in the context of our framework. We constructed Figure 3.8 by estimating our complete model using rolling (overlapping) windows of 20

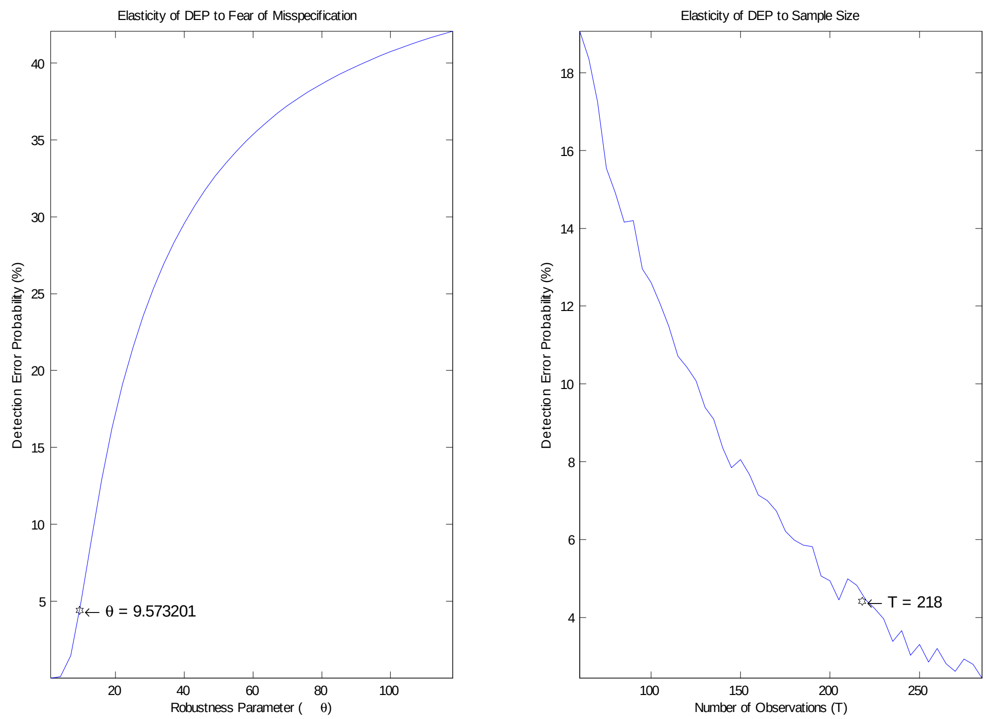

Figure 3.7: Comparative statics on DEP's. Left panel: Fixing all the estimated parameters in the bechmark case with consumption growth volatility as a restriction and over the longest sample $Q2.52 - Q4.06$. Varying robustness in the model (by varying θ on the x-axis) we compute the implied DEP's (y-axis). Right panel: Fixing all the parameters in the benchmark model and varying only the hypothetical number of observations (x-axis) and computing the implied DEP's (y-axis).

years of quarterly data, from the early 1970s to 2007. For any given estimation iteration, we present the point estimate of θ with its corresponding 95% confidence interval and DEP. It is apparent from this figure that θ and DEPs are closely related to each other, with a cross correlation of 0.8113 and Newey-West standard errors with 4 lag correction of 0.0418. Therefore, it strongly confirms the suggestion that we should examine DEPs when trying to understand the level of uncertainty aversion exhibited by economic agents.

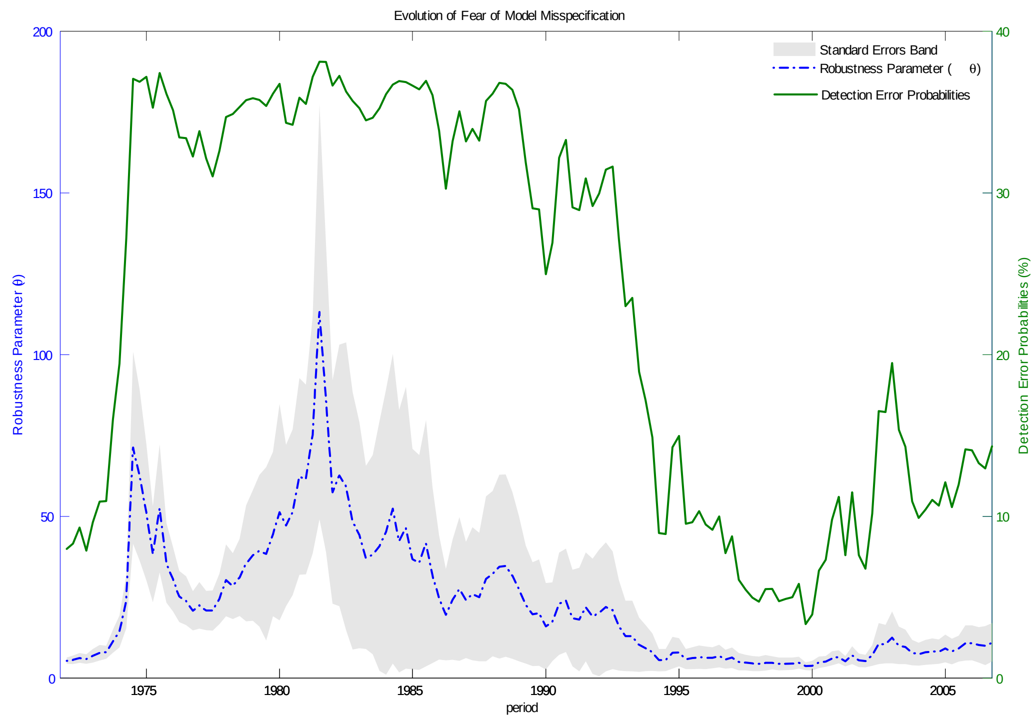

Figure 3.8: Evolution of θ and DEP's over time. We estimate our complete model by using rolling (overlapping) windows of 20 years of quarterly data. For any given estimation iteration, we present the point estimate of θ (dot-dashed line) with its corresponding 95% confidence interval (gray bands). The standard errors are corrected using the Newey-West procedure with 4 lags. Given the estimated parameters in each iteration, we compute the implied DEP (solid line).

On the basis of this exercise it seems that the agent was seeking more robustness in the later period of the sample. An interesting question is to determine whether this evolution could be linked to macroeconomic and financial developments over the same time period, and in particular its link with the discussion about the Great Moderation.[40] Since the

[40]On the one hand, macro volatility, and in particular consumption growth volatility, has steadily declined in the later period of our sample (the Great Moderation). However, market return volatility does not exhibit the same pattern. One possibility is that the smoother consumption growth is interpreted by the estimation procedure as an increased uncertainty aversion which implies the decline in detection error

investigation of any causality is outside the scope of this study, we leave this question for future research.

3.6. Conclusion

We presented an equilibrium dynamic asset pricing model that can account for key regularities in the market for default free bonds, while predicting an equity premium, risk free rate and consumption growth as in the data. We estimated the model and showed that it performs well, even though the structural parameters of risk aversion and elasticity of intertemporal substitution are unitary. The results are driven by the interaction of the robust control decision mechanism and state dependent conditional volatility of consumption growth. We interpreted most of what is usually considered risk premium as a premium for Knightian uncertainty. The agent is being compensated in equilibrium for bearing the possibility of model misspecification. We also showed that modeling robustness can help explain biases in expectations documented in surveys.

We showed that under the assumption of state dependent conditional volatility of consumption growth, not only the market price of risk is stochastic but also the market price of model uncertainty. As part of our research agenda, we are currently investigating a model with heterogenous robust control agents. Such a model can generate both state dependent risk and uncertainty premia even though the conditional volatility of consumption is constant. The channel through which the model generates stochastic market prices

probabilities in the later part of the sample. In other words, it is harder to achieve the HJ bound with smoother consumption. Thus, the estimation procedure compensates for this difficulty by encoding more robustness into the agent's behavior. Consequently, the implied DEPs are higher.

of risk and uncertainty is the trade between the agents and the consequent fluctuations in the agent's relative wealth.[41]

We also suggested that different frequencies in the conditional volatility of consumption growth are potentially important in understanding asset prices. We find it easier to detect high frequency variation in the volatility of consumption growth rate. Also, the full estimation of the model has trouble detecting the lower frequency component. We believe that further investigation of this point is warranted. In addition, an interesting extension would be to consider the link between the evolution of volatility over time and the behavior of asset prices, in the presence of ambiguity aversion. This is directly linked to the recent literature on the Great Moderation in macroeconomics.

We also believe that extending the empirical investigation to a broader asset class can be fruitful. Liu et al. (2005), for example, examine options data in the context of a robust equilibrium with rare events. We believe that one can address different empirical regularities pertaining to the valuation of interest rate sensitive assets with robust considerations. Also, we think that robustness can shed more light on our understanding of exchange rate dynamics, and in particular the failure of uncovered interest rate parity. Finally, our model is a complete characterization of a real economy. One can extend this framework to a nominal one either by assuming an exogenous price level process as in Cox et al. (1985) and Wachter (2001) or by modelling an exogenous money supply process as in Buraschi and Jiltsov (2005) to derive an endogenous price level.

---

[41] Liu et al. (2005) introduce state dependent market price of uncertainty by modeling rare events. Hansen and Sargent (2007b) introduce state dependent market price of uncertainty through the distortion (tilting) of Bayesian model averaging.

References

Abel, A., (2002), "An exploration of the effects of pessimism and doubt on asset returns." *Journal of Economic Dynamics & Control* 26, 1075–1092.

Altig, D., Christiano, L.J., Eichenbaum, M. and J. Linde, 2005, "Firm-Specific Capital, Nominal Rigidities, and the Business Cycle", *NBER Working Paper* 11034.

Amano, R. and Hendry, S., 2003, "Inflation Persistence and Costly Market Share Adjustment: A Preliminary Analysis", BIS Papers No.19, 2003.

Amirault, D., Kwan, C., and G. Wilkinson, 2006, "Survey of Price-Setting Behaviour of Canadian Companies", *Bank of Canada Working Paper 2006-35.*

Anderson, E., L. P. Hansen, and T. J. Sargent, (2003), "A quartet of semigroups for model specification, robustness, prices of risk, and model detection." *Journal of the European Economic Association* 1 (1):68–123.

Apel, M., R. Friberg and K. Hallsten, 2005, "Microfoundations of Macroeconomic Price Adjustment: Survey Evidence from Swedish Firms", *Journal of Money, Credit and Banking*, Vol. 37, No. 2, 314-338.

Atkeson, A. and A. Burstein, 2005, "Trade Costs, Pricing to Market, and International Relative Prices", mimeo, University of California in Los Angeles.

Aucremanne, L. and M. Druant, 2005, "Price-Setting Behaviour in Belgium: What Can Be Learned from an Ad-Hoc Survey?", *ECB Working Paper No. 448.*

Backus, D., Foresi, S., Mozumdar, A. and Wu, L., (1998), "Predictable changes in yields and forward rates." NBER WP no 6379.

Backus, D., Gregory, A. W. and Zin, S. E., (1989), "Risk premiums in the term structure: Evidence from artificial economies." *Journal of Monetary Economics* 24, 371-99.

Ball, L. and D. Romer, 1990, "Real Rigidities and the Non-Neutrality of Money," *The Review of Economic Studies*, Vol. 57, No. 2, 183-203.

Bansal, R., Katchatrian V., and A. Yaron (2005), "Interpretable asset markets?" *European Economic Review* 49, 531-560.

Bansal, R., Kiku, D., and Yaron, A. (2007), "Risks for the long run: estimation and inference." Working paper, Duke and UPenn.

Bansal, R., and A. Yaron. (2004), "Risks for the long run: a potential resolution of asset pricing puzzles." *Journal of Finance* 59(4):1481–509.

Barillas, F., Hansen, L. P., Sargent, T. J., (2007), "Doubts or variability?" Working paper, University of Chicago and NYU.

Baudry, L., Le Bihan, H., Sevestre, P. and S. Tarrieu, 2004, "Price Rigidity: Evidence from the French CPI Micro-Data", *ECB Working Paper No. 384*.

Bauer, H. H., Hammerschmidt, M. and M. Braehler, 2004, "The Customer Lifetime Value Concept and Its Contribution to Corporate Valuation," *Yearbook of Marketing and Consumer Research*, Vol. 1, Berlin: Duncker & Humblot, 47–67.

Bekaert, G., Hodrick, R. and Marshall, D., (1997), "On biases in tests of the expectations hypothesis of the term structure of interest rates." *Journal of Financial Economics* 44, 309-48.

Besanko, D., Dubé, J.-P., and S. Gupta, 2005, "Own-Brand and Cross-Brand Retail Pass-Through," *Marketing Science*, Vol.24, No. 1, 123-137.

Bils, M., 1987, "The Cyclical Behavior of Marginal Cost and Price", *American Economic Review*, Vol. 77, No 5, 838-855.

Bils, M. and P. Klenow, 2004, "Some Evidence on the Importance of Sticky Prices", *Journal of Political Economy*, Vol. 112, No 5, 947-985.

Blinder, A.S., E.R.D. Canetti, D.E. Lebow and J.B. Rudd , 1998, *Asking about Prices: A New Approach to Understanding Price Stickiness*, New York: Russell Sage.

Bliss, R. (1999), "Fitting term structures to bond prices." Working paper, Federal Reserve Bank of Chicago.

Boivin, J., M. Giannoni and I. Mihov, 2007, "Sticky Prices and Monetary Policy: Evidence from Disaggregated U.S. Data", *NBER Working Paper 12824*.

Borenstein, S., Cameron, A.C., R. Gilbert, 1992, "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?", *NBER Working Paper 4138*.

Brown, S. J. and Dybvig, P. H. (1986), "The empirical implications of the Cox, Ingersoll, and Ross theory of the term structure of interest rates." *Journal of Finance* 41, 617-32.

Brown, R. H. and Schaefer, S. (1994), "The term structure of real interest rates and the Cox, Ingersoll, and Ross Model." *Journal of Financial Economics* 35, 3-42.

Buraschi, A., Jiltsov, A., (2005), "Inflation risk premia and the expectations hypothesis." *Journal of Financial Economics* 75, 429-490.

Burstein, A., Eichenbaum, M., and S. Rebelo, 2005, "Large Devaluations and the Real Exchange Rate", *Journal of Political Economy*, Vol.113, No.4, 742-784.

Burstein, A., M. Eichenbaum and S. Rebelo, 2005, "Modeling Exchange-Rate Passthrough After Large Devaluations," *NBER working Paper 11638.*

Cagetti, M., Hansen, L. P., Sargent, T. and Williams, N. (2002), "Robustness and pricing with uncertain growth." *Review of Financial Studies* 15, 363–404.

Campa, J.M. and L.S. Goldberg, 2002, "Exchange Rate Pass-Through into Import Prices: a Macro or Micro Phenomenon?," *NBER Working Paper 8934.*

Campbell, J. Y., (2000), "Asset pricing at the millennium." *Journal of Finance* 55(4): 1515-1567.

Campbell, J. Y. and Shiller, R., (1991), "Yield spreads and interest rates: A bird's eye view." *Review of Economic Studies* 58, 495-514.

Cason, T.N. and D. Friedman, 2002, "A Laboratory Study of Customer Markets," *Advances in Economic Analysis and Policy*, Vol.2, Issue 1.

Chan, K, Karolyi, A., Longstaff, F. and Sanders, A., (1992), "The volatility of the short-term interest rates: An empirical comparison of alternative models of the term structure of interest rates." *Journal of Finance* 68: 1209-27.

Chapman, D. (1997), "The cyclical properties of consumption growth and the real term structure." *Journal of Monetary Economics* 39, 145-72.

Christiano, L., Eichenbaum, M. and C. Evans, 2005, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy", *Journal of Political Economy*, February, 2005, 113 (1), 1 -45.

Cochrane, J. H. and Piazzesi, M., (2002), "Bond risk premia." *American Economic Review* 95, 138-160.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985), "A theory of the term structure of interest rates." *Econometrica* 53, 385-407.

Croce, M. M., Lettau, M., Ludvigson, S. C. (2006), "Investor information, long-run risk, and the duration of risky cash flows." Working paper, NYU.

Dai, Q. (2003), "Term structure dynamics in a model with stochastic internal habit." Working paper, UNC.

Dai, Q. and Singleton, K. (2000), "Specification analysis of affine term structure models." *Journal of Finance* 55, 1943-78.

Dai, Q. and Singleton, K. (2002), "Expectation puzzles, time-varying risk premia, and affine models of the term structure." *Journal of Financial Economics* 63, 415-41.

Dai, Q. and Singleton, K. (2003), "Term structure modeling in theory and reality." *Review of Financial Studies* 16(3), 631-678.

Devereux, M., 2003, "Price Setting and Exchange Rate Pass-through: Theory and Evidence," *Price Adjustment and Monetary Policy*, Bank of Canada.

Dhyne, E., Alvarez, L.J., Le Bihan, H., Veronese, G., Dias, D., Hoffman, J., Jonker, N., Lünnemann, P., Rumler, F. and J. Vilmunen, 2005, "Price-Setting in the Euro Area: some Stylized Facts from Individual Consumer Price Data", *ECB Working Paper No. 524*.

Dixit, A.K. and J.E. Stiglitz, 1977, "Monopolistic Competition and Optimum Product Diversity", *American Economic Review*, 67, 297-308.

Dominguez, K.M.E. and L.L. Tesar, 2006, "Exchange Rate Exposure", *Journal of International Economics*, January 2006, 68(1): 188-218.

Donaldson, J. B., Johnsen, T. H. and Mehra, R. (1990), "On the term structure of interest rates." *Journal of Economic Dynamics and Control* 14, 571–596.

Dotsey, M. and R. King, 2005, "Implications of State-Dependent Pricing for Dynamic Macroeconomic Models," *Journal of Monetary Economics*, 52(1), 213-242.

Drost, F.C., Nijman, T.E. (1993), "Temporal Aggregation of Garch Processes." Econometrica 61, 909-927.

Duffie, D., Epstein, L., (1992a), "Asset pricing with stochastic differential utility." *Review of Financial Studies* 5, 411–436.

Duffie, D., and Epstein, L., (1992b), "Stochastic differential utility." *Econometrica* LX, 353–394.

Duffie, D. and Kahn, R. (1996), "A yield-factor model of interest rates." *Mathematical Finance* 6, 379-406.

Duffie, D., and Singleton, K., (1993), "Simulated methods estimation of markov models of asset prices." *Econometrica* 61, 929-952.

Ellsberg, D. (1961), "Risk, ambiguity and the savage axioms." *Quarterly Journal of Economics* 25, 643–69.

Epstein, L. G. and Wang, T., (1994), "Intertemporal asset pricing under knightian uncertainty." *Econometrica* 62(2):283–322.

Epstein,L. G. and Schneider, M., (2003), "Recursive multiple priors." *Journal of EconomicTheory* 113:32–50.

Epstein, L. and Zin, S., (1989), "Substitution, risk aversion and the temporal behavior of consumption and asset returns: a theoretical framework." *Econometrica* 57, 937–969.

Erbaş, S. N. and Mirakhor, A. (2007), "The equity premium puzzle, ambiguity aversion, and institutional quality." IMF WP/07/230.

Fabiani, S., Druant, M., Hernando, I., Kwapil, C., Landau, B., Loupias, C., Martins, F., Matha, T.Y., Sabbatini, R., Stahl, H. and A.C.J. Stockman, 2005, "The Pricing Behaviour of Firms in the Euro Area", *ECB Working Paper No. 535.*

Fama, E. F. and Bliss, R.R. (1987), "The information in long-maturity forward rates." *American Economic Review* 77, 680-92.

Froot, K. and P. Klemperer, 1989, "Exchange rate Pass-Through When Market Share Matters," *American Economic Review 79*, 637-654.

Gagliardini, P., Porchia, P. and Trojani, F., (2004), "Ambiguity aversion and the term structure of interest rates." Working paper, University of St.Gallen.

Gagnon, E., 2006, "Price Setting during Low and High Inflation: Evidence from Mexico", mimeo Federal Reserve Board.

Ghanesh, J., M.J. Arnold and K.E. Reynolds, 2000, “Understanding the Customer Base of Service Providers: An Examination of the Differences between Switchers and Stayers”, *Journal of Marketing*, July 2000, 64(3): 65-87.

Gibbons, M. R. and Ramaswamy, K., (1993), “A test of the Cox, Ingersoll and Ross model of the term structure.” *Review of Financial Studies* 6, 619-58.

Gilboa, I. and Schmeidler, D., (1989), "Maxmin expected utility with non-unique prior." *Journal of Mathematical Economics* 18:141–153.

Giordani, P., Söderlind, P., (2006), "Is there evidence of pessimism and doubt in subjective distributions? Implications for the equity premium puzzle." *Journal of Economic Dynamics & Control* 30, 1027-1043.

Glosten, L., Jagannathan, R., Runkle, D. (1993), "On the relation between the expected value and the volatility of the nominal excess return on stocks." *Journal of Finance* 48, 1779-1801.

Golosov, M. and R.E. Lucas, Jr., 2003, “Menu Costs and Phillips Curves”, *NBER Working Paper 10187*.

Gron, A. and D.L. Swenson, 2000, “Cost Pass-through in the U.S. Automobile Market”, *The Review of Economics and Statistics*, 82(2): 316-324.

Hall, S., Walsh, M., and A. Yates, 1997, “How Do U.K. Companies Set Prices?” Bank of England Working Paper No.67.

Hansen, L. P., (1982), “Large sample properties of generalized method of moments estimators.” *Econometrica* 50, 1029-54.

Hansen, L. P. and Jagannathan, R., (1991), "Implications of security market data for models of dynamic economies." *Journal of Political Economy* 99:225–262.

Hansen, L. P. and Sargent, T. J., (2007a), "Robustness." Princeton, New Jersey: Princeton University Press.

Hansen, L.P. and Sargent, T. J., (2007b), "Fragile beliefs and the price of model uncertainty." Working paper, NYU and Univ. of Chicago.

Hansen, L. P., Sargent, T. J., Turmuhambetova, G. A. and Williams, N. (2006), "Robust control, min-max expected utility, and model misspecification." *Journal of Economic Theory* 128, 45–90.

Hansen, L. P., Sargent, T. and Tallarini, T. (1999), "Robust permanent income and pricing." *Review of Economic Studies* 66, 873–907.

Ireland, P.N., 1998, "Customer Flows, Countercyclical Markups, and the Output Effects of Technology Shocks", *Journal of Macroeconomics*, 20(4), 649-664.

Keaveney, S.M., 1995, "Customer Switching Behavior in Service Industries: An Explanatory Study", *Journal of Marketing*, April 1995, 59(2): 71-82.

Kim, M., Kliger, D., and B. Vale, 2001, "Estimating Switching Costs and Oligopolistic Behavior", Wharton School for Financial Institutions Center, WP 01-13.

Klemperer, P., 1995, "Competition when Consumers Have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade", *The Review of Economic Studies*, 62:4, 515-539.

Klenow, P.J. and J.L. Willis, 2006, "Real Rigidities and Nominal Price Changes," Kansas City Fed, Working Paper 06-03.

Kwapil, C., Baumgartner, J. and J. Scharley, 2005, "The Price-Setting Behavior of Austrian Firms: Some Survey Evidence", *ECB Working Paper No. 464*.

Litterman, R. and Scheinkman, J. (1991), "Common factors affecting bond returns." *Journal of Fixed Income* 1, 54-61.

Liu, J., Pan, J. and Wang, T. (2005), "An equilibrium model of rare-event premia and its implication for option smirks." *Review of Financial Studies* 18(1), 131–64.

Lkhagvasuren, D., 2006, "Big Locational Differences in Unemployment Despite High Labor Mobility," mimeo, Northwestern University, 2006.

Longstaff, F. and Schwartz, E., (1992), "Interest rate volatility and the term structure: A two-factor general equilibrium model." *Journal of Finance* 47, 1259-82.

Lucas, Jr., R. E., (1978). "Asset prices in an exchange economy." *Econometrica* 46, 1426–14.

Ludwig, A. and Zimper, A. (2006), "Rational expectations and ambiguity: A comment on Abel (2002)." *Economics Bulletin*, Vol. 4, No. 2 pp. 1-15.

Maccheroni, F., Marrinacchi, M., Rustichini, A, (2006) "Ambiguity aversion, robustness, and the variational representation of preferences." *Econometrica* 74, 1447-1498.

Maenhout, P. (2004), "Robust portfolio rules and asset pricing." *Review of Financial Studies* 17, 951–983.

Mair, J., 2005, "How the Appreciation of the Canadian Dollar Has Affected Canadian Firms: Evidence from the Bank of Canada," *Bank of Canada Review*, Autumn 2005.

McCulloch, J. H. and Kwon, H. C. (1993), "U.S. term structure data, 1947-1991." Working paper, Ohio State University.

Mehra, R. and Prescott, E., (1985), "The equity premium: a puzzle." *Journal of Monetary Economics* 15, 145–61.

Miao, J., (2004), "A note on consumption and savings under knightian uncertainty." *Annals of Economics and Finance* 5, 299-301.

Nelson, D., (1991), "Conditional heteroskedasticity in asset returns: A new approach." *Econometrica* 59, 347-370.

Neumark, D. and S.A. Sharpe, 1992, "Market Structure and the Nature of Price Rigidity: Evidence from the Market for Consumer Deposits," *The Quarterly Journal of Economics*, 107:2, 657-680.

Okun, A. ,1981, *Prices and Quantities: A Macroeconomic Analysis*, Blackwell:Oxford.

Peltzman, S., 2000, "Prices Rise Faster than They Fall", *Journal of Political Economy*, 108:3, 466-502.

Phelps, E.S. and S. Winter, 1970, "Optimal Price Policy under Atomistic Competition," *Microeconomic Foundations of Employment and Inflation Theory* (Norton).

Piazzesi, M. (2004), "Affine term structure models", Y. Aït-Sahalia and L. P. Hansen (eds.), *Handbook of Financial Econometrics*, Amsterdam: North Holland.

Ravn, M., S. Schmitt-Grohé, and M. Uribe, 2006,"Deep Habits," *Review of Economic Studies* 73, 2006, 195-218.

Renner, E. and J.-R. Tyran, 2004, "Price Rigidity in Customer Markets", *Journal of Economic Behavior & Organization*, Vol. 55, 575-593.

Rotemberg, J., 2005, "Customer Anger at Price Increases, Time Variation in the Frequency of Price Changes and Monetary Policy", *Journal of Monetary Economics* 52, 829-852.

Rotemberg, J., J., and M. Woodford , 1997, “An Optimization-Based Econometric Model for the Evaluation of Monetary Policy,” *NBER Macroeconomics Annual* 12: 297-346

Rotemberg, J., J., and M. Woodford , 1999, “The Cyclical Behavior of Prices and Costs,” in J.B. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*, 1B.

Routledge, B. R., and Zin, S. E. (2001), "Model uncertainty and liquidity." Working paper, NBER.

Skiadas, C. (2003), "Robust control and recursive utility." *Finance and Stochastics* 7, 475–489.

Söderlind, P., (2006), "C-CAPM without ex-post data.". WP 344, *NCCR FINRISK*

Stiglitz, J. E., 1979, “Equilibrium in Product Markets with Imperfect Information”, *American Economic Review*, 69, 339-345.

Stock, J. H. and Watson, M. W., (2003), "Has the business cycles changed? evidence and explanations." in "Monetary policy and uncertainty: adapting to a changing economy." pp. 9-56. Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming.

Strzalecki, T. (2007), "Axiomatic foundations of multiplier preferences." Working paper, Northwestern University.

Uppal, R., Wang, T., (2003), "Model misspecification and underdiversification." *Journal of Finance* 58, 2465–2486.

Wachter, J., (2006), “A consumption-based model of the term structure of interest rates.” *Journal of Financial Economics* 79, 365–399.

Weil, P., (1989), "The equity premium puzzle and the risk free rate puzzle." *Journal of Monetary Economics* 24, 401–421.

Weil, P., (1990), "Nonexpected Utility in Macroeconomics." *Quarterly Journal of Economics* 105 (1):29–42.

Whitelaw, R., (1994), "Time variation and covariations in the expectation and volatility of stock market returns." *Journal of Finance* 49, 515–541.

Wiederholt, M., and B. Mackowiak, 2006, “Optimal Sticky Prices and Rational Inattention,” mimeo Northwestern University.

Woglom, G., 1982, "Underemployment Equilibrium with Rational Expectations", *Quarterly Journal of Economics*, 97, 89- 107.

Zbaracki, M.J., Ritson, M., Levy, D., Dutta, S., and M. Bergen, 2004, "Managerial and Customer Costs of Price Adjustment: Direct Evidence from Industrial Markets", *The Review of Economics and Statistics*, 86(2), 514-533.

Appendix

1. Switching Rule

In this section we show the details behind the derivaton of the object $\partial\widehat{\delta}_{it}(k)/\partial p_{it}(k)$. The threshold switching cost $\widehat{\delta}_{it}(k)$ is implicitly defined by equating the value of remaining with the current supplier, V_0, to the value of switching, V_1:

$$V_0\left[\left\{p_i^t\right\}, p_i^t(k)\right] = V_1\left[\left\{p_i^t\right\}, \widehat{\delta}_{it}(k)\right].$$

Even if different threshold consumers have different aggregates $\tilde{p}_t^j$ and $\tilde{c}_t^j$ out of equilibrium, those variables do not affect the marginal decisions to stay or switch for a particular seller. Clearly, this choice is only a function of the switching cost, the price charged by the home seller and the distribution of prices from other sellers. Therefore, we focus on a typical threshold costumer and drop the j subscripts to be concise.

By defining the function $G\left[\left\{p_i^t\right\}, p_i^t(k), \widehat{\delta}_{it}(k)\right] = V_0 - V_1 = 0$, it is easy to verify that at the symmetric equilibrium where $p_i^t(l) = p_i^t$, $\forall l$, we have that $\widehat{\delta}_{it} = 1$ and the following conditions are satisfied:

$$G\left[\left\{p_i^t\right\}, p_i^t, \widehat{\delta}_{it} = 1\right] = 0$$

$$\frac{\partial G}{\partial\widehat{\delta}_{it}}\left[\left\{p_i^t\right\}, p_i^t, \widehat{\delta}_{it} = 1\right] \neq 0.$$

We therefore know that the implicit function theorem applies around the symmetric equilibrium, which is all we need for our purpose.

$$\frac{\partial \widehat{\delta}_{it}(k)}{\partial p_{it}(k)}\left[\left\{p_i^t\right\}, p_i^t, \widehat{\delta}_{it}=1\right]=-\frac{\frac{\partial G}{\partial p_{it}(k)}\left[\left\{p_i^t\right\}, p_i^t, \widehat{\delta}_{it}=1\right]}{\frac{\partial G}{\partial \widehat{\delta}_{it}(k)}\left[\left\{p_i^t\right\}, p_i^t, \widehat{\delta}_{it}=1\right]}$$

V_0 depends only on $p_{it}(k)$, and based on the result of the Section 2.3.3, we know that the continuation value is not a function of $p_{it}(k)$. Denote the optimal demand of the typical threshold consumer by c_{0it} and c_{1it} in case he is either staying or switching, respectively:

$$\begin{aligned}
\partial G/\partial p_{it}(k) &= \frac{\partial U(c_{0it})}{\partial c_{0it}}\frac{\partial c_{0it}}{\partial p_{it}(k)} \\
&= (\tilde{c}_t)^{\frac{1}{\gamma}-\sigma}(\tilde{c}_{0it})^{-\frac{1}{\gamma}}\left(-\frac{\gamma c_{0it}}{p_{it}(k)}\right) \\
&= -\gamma \tilde{c}_t^{-\sigma}\frac{c_{0it}}{\tilde{p}_t}
\end{aligned}$$

where we use our earlier result that $(\tilde{c}_t)^{\frac{1}{\gamma}}(c_{0it})^{-\frac{1}{\gamma}} = p_{it}(k)/\tilde{p}_t$. Also, only V_1 depends on $\widehat{\delta}_{it}(k)$, and because the switching cost are i.i.d., the derivative with respect to the continuation value drops out. Hence:

$$\begin{aligned}
\partial G/\partial \widehat{\delta}_{it}(k) &= -\partial V_1/\partial \widehat{\delta}_{it}(k) \\
&= -\int_0^1 \left[\frac{\partial U(c_{1it}(l))}{\partial \widehat{\delta}_{it}(k)} + \frac{\partial U(c_{1it}(l))}{\partial c_{1it}(l)}\frac{\partial c_{1it}(l)}{\partial \widehat{\delta}_{it}(k)}\right] dl \\
&= \int_0^1 \gamma(\tilde{c}_t)^{-\sigma} c_{1it}(l)\frac{1}{\widehat{\delta}_{it}(k)}\frac{p_{it}(l)}{\tilde{p}_t} dl
\end{aligned}$$

where we use $(\tilde{c}_t)^{\frac{1}{\gamma}} (\tilde{c}_{1it}(l))^{-\frac{1}{\gamma}} \widehat{\delta}_{it}(k)^{\frac{1}{\gamma}-1} = p_{it}(l)/\tilde{p}_t$. We know that:

$$\frac{c_{0it}}{c_{1it}(l)} = \widehat{\delta}_{it}(k)^{\gamma-1}\left(\frac{p_{it}(k)}{p_{it}(l)}\right)^{-\gamma}$$

therefore, we obtain:

$$\begin{aligned}\frac{\partial\widehat{\delta}_{it}(k)}{\partial p_{it}(k)} &= \frac{\gamma\tilde{c}_t^{-\sigma}c_{0it}}{\int\limits_0^1 \gamma(\tilde{c}_t)^{-\sigma}c_{1it}(l)\frac{1}{\widehat{\delta}_{it}(k)}p_{it}(l)\,dl} \\ &= \widehat{\delta}_{it}(k)\left[\int\limits_0^1 \frac{c_{1it}(l)}{c_{0it}}p_{it}(l)\,dl\right]^{-1}\end{aligned}$$

$$\frac{\partial\widehat{\delta}_{it}(k)}{\partial p_{it}(k)} = \left(\widehat{\delta}_{it}(k)\right)^{\gamma}\left[\int\limits_0^1 \left(\frac{p_{it}(k)}{p_{it}(l)}\right)^{\gamma} p_{it}(l)\,dl\right]^{-1}$$

which, evaluated at the symmetric equilibrium, simplifies to:

$$\frac{\partial\widehat{\delta}_{it}(k)}{\partial p_{it}(k)}\left[\left\{p_i^t\right\}, p_i^t, \widehat{\delta}_{it} = 1\right] = \frac{1}{p_{it}}.$$

2. Price Index

The derivation of $\tilde{p}_t$ is straightforward if we impose symmetry across goods i. In this section, we instead show under which conditions the aggregate price index is not household specific when we assume only symmetry within each sector. The following has to hold by definition:

$$\begin{aligned}\tilde{p}_t^j \tilde{c}_t^j &= \int_0^1 p_{it} c_{it}^j di = \int_0^1 p_{it} \left[\left(\delta_{it}^j\right)^{s_{it}^j} \right]^{1-\gamma} \left(\frac{p_{it}}{\tilde{p}_t^j} \right)^{-\gamma} \tilde{c}_t^j di \\ \tilde{p}_t^j &= \left[\int_0^1 \left(\left(\delta_{it}^j\right)^{s_{it}^j} p_{it} \right)^{1-\gamma} di \right]^{\frac{1}{1-\gamma}}.\end{aligned}$$

We start by dividing the mass of sectors into N supersectors. A supersector n is composed of a continuum of *identical* sectors which all have the same price p_{nt}. For simplicity, we consider supersectors of similar size $1/N$ and normalize the mass of goods within each sector to 1. Therefore, we can now rewrite the price index as:

$$\tilde{p}_t^j = \left[N^{-1} \sum_{n=1}^N p_{nt}^{1-\gamma} \int_0^1 \left(\left(\delta_{it}^j\right)^{s_{it}^j} \right)^{1-\gamma} di \right]^{\frac{1}{1-\gamma}}.$$

Now consider a particular supersector n. For any household j, the distribution of switching costs across goods within that supersector is the same:

$$\begin{aligned}\int_0^1 \left(\left(\delta_{it}^j\right)^{s_{it}^j} \right)^{1-\gamma} di &= \int_{i:s_{it}^j=0} dF(\delta) + \int_{i::s_{it}^j=1} \delta^{1-\gamma} dF(\delta) \\ &= 1 - F(1) + \int_{\underline{\delta}}^1 \delta^{1-\gamma} dF(\delta).\end{aligned}$$

That the last term takes into account the weighting introduced by the switching costs in those instances when the consumer decides to switch. This implies that the aggregate price index is not household specific anymore, and we can write it as:[42]

$$\tilde{p}_t = \left[N^{-1} \sum_{n=1}^{N} p_{nt}^{1-\gamma} \left(1 - F(1) + \int_{\underline{\delta}}^{1} \delta^{1-\gamma} dF(\delta) \right) \right]^{\frac{1}{1-\gamma}}$$

$$\tilde{p}_t = \overline{p}_t \left(1 - F(1) + \int_{\underline{\delta}}^{1} \delta^{1-\gamma} dF(\delta) \right)^{\frac{1}{1-\gamma}}$$

where

$$\overline{p}_t = \left[N^{-1} \sum_{n=1}^{N} p_{nt}^{1-\gamma} \right]^{\frac{1}{1-\gamma}} .$$

We assume that N is large enough so that any movement in one supersector has no impact on the price index $\overline{p}_t$, and hence on the aggregate index $\tilde{p}_t$. Finally, if we impose symmetry across goods such that $p_{it} = p_t$, we obtain:

$$\tilde{p}_t = p_t \left(1 - F(1) + \int_{\underline{\delta}}^{1} \delta^{1-\gamma} dF(\delta) \right)^{\frac{1}{1-\gamma}} .$$

In the Dixit-Stiglitz framework where $\delta_{it} = 1, \forall i, t$, we would have that $\tilde{p}_t = p_t$.

[42]We can similarly show that the consumption index, $\tilde{c}_t^j$, is also not household specific.

3. Equilibrium in the Static Case

We prove analytically that even under a model with customer flows, the unique symmetric equilibrium is one where the optimal pricing strategy is to fully pass-through marginal cost shocks to prices.[43]

Consider a sector i which is initially in steady state at $t-1$ with a marginal cost of 1. We denote the new exogenous marginal cost at time t by $mc_{it} = \kappa$, and by $p_{it} = \kappa' p_i$ the new price charged by all competitors in sector i, where p_i is the steady state value of the price. For example, in the case where marginal cost goes up ($\kappa > 1$), pass-through is incomplete if $\kappa' < \kappa$ and complete if $\kappa' = \kappa$.

Since $\beta = 0$, the problem of the firm is purely static, and we can easily express the impact on a firm's profit if it decides to deviate by changing its price:

$$\frac{\partial \Pi_{it}(k)}{\partial p_{it}(k)} = \frac{\partial c_{it}(k)}{\partial p_{it}(k)} \left[p_{it}(k) - mc_{it}\right] + \frac{\partial \left[p_{it}(k) - mc_{it}\right]}{\partial p_{it}(k)} c_{it}(k).$$

We evaluate this expression around the current state where all firms charge the same price. Hence, we set $p_{it}(k) = p_{it}$:

$$\begin{aligned}
\left.\frac{\partial \Pi_{it}(k)}{\partial p_{it}(k)}\right|_{\frac{p_{it}(k)}{p_{it}}=1} &= A(1)\left(\frac{p_{it}}{\tilde{p}}\right)^{-\gamma} \tilde{c} - \left(\frac{p_{it}}{\tilde{p}_t}\right)^{-\gamma-1} \frac{\tilde{c}}{\tilde{p}} \left[\gamma A(1) + f(1)\right]\left[p_{it} - mc_{it}\right] \\
&= \left(\frac{p_{it}}{\tilde{p}}\right)^{-\gamma} \tilde{c} \left[A(1) - \frac{1}{p_{it}}\left[\gamma A(1) + f(1)\right]\left[p_{it} - mc_{it}\right]\right].
\end{aligned}$$

[43]We have shown that such an equilibrium exists when we derive (2.24).

Next, we use our initial definitions that $mc_{it} = \kappa$ and $p_{it} = \kappa' p_i$, and introduce $\varkappa = \frac{\kappa}{\kappa'}$:

$$\begin{aligned} \left.\frac{\partial \Pi_{it}(k)}{\partial p_{it}(k)}\right|_{\frac{p_{it}(k)}{p_{it}}=1} &= \left(\frac{\kappa' p_i}{\tilde{p}}\right)^{-\gamma} \tilde{c}\left[A(1) - \frac{1}{\kappa' p_i}\left[\gamma A(1) + f(1)\right]\left[\kappa' p_i - \kappa\right]\right] \\ &= \left(\frac{\kappa' p_i}{\tilde{p}}\right)^{-\gamma} \tilde{c}\left[A(1) - \left[\gamma A(1) + f(1)\right]\left[1 - \frac{\varkappa}{p_i}\right]\right]. \end{aligned}$$

Finally, we use the expression for the steady state price in our model,

$$p_i = \frac{\gamma A(1) + f(1)}{(\gamma - 1) A(1) + f(1)}$$

and simplify to obtain:

$$\left.\frac{\partial \Pi_{it}(k)}{\partial p_{it}(k)}\right|_{\frac{p_{it}(k)}{p_{it}}=1} = \left(\frac{\kappa' p_i}{\tilde{p}}\right)^{-\gamma} \tilde{c}\left[(\gamma - 1) A(1) + f(1)\right](\varkappa - 1). \tag{.33}$$

We can now analyze this expression to prove that there is a unique, stable symmetric equilibrium under the static case. For example, consider that following an increase of the marginal cost at time t from 1 to κ, all sellers are setting $p_{it} = \kappa' p_i$ where $\kappa' < \kappa$. In this case, there is incomplete pass-through, as the rise in the price is proportionately less than the increase in the marginal cost (in other words, the markup falls). The expression in (.33) tells us that since $\varkappa = \frac{\kappa}{\kappa'} > 1$, $\frac{\partial \Pi_{it}(k)}{\partial p_{it}(k)} > 0$ and therefore a seller k has an incentive to deviate and increase its price. Conversely, if $\kappa' > \kappa$ (firms overshoot following the rise in marginal cost), then our derivation above shows that any producer can raise its profits by lowering its price. The same analysis can be applied to a fall in the marginal cost ($\kappa < 1$).

Most importantly, (.33) shows that the unique symmetric equilibrium is where $\varkappa = 1$, that is where cost pass-through is complete: only in this particular case is there no incentive for sellers to deviate. In any other instances where firms behave symmetrically, but do not practice complete pass-through, there is an incentive to price in order to maintain a constant markup. This confirms that if firms face real rigidities in the context of a non-dynamic profit-maximization problem, the unique and stable symmetric equilibrium is one in which producers fully pass-through cost shocks.

4. Equilibrium in a Dynamic Setting

We now study the incentive of a seller k to deviate from a symmetric equilibrium where all firms fully pass-through the marginal cost shock into their prices. Here we focus on the case of a purely transitory shock ($\rho_z = 0$) because it is analytically tractable.

Recall that the discounted sum of profits $\widehat{\Pi}_{i0}(k)$ of the (i,k) seller is given by (2.12). Its derivative with respect to price at $t = 0$ around the symmetric equilibium is:

$$\left.\frac{\partial \widehat{\Pi}_{i0}(k)}{\partial p_{i0}(k)}\right|_{\left\{\frac{p_{it}(k)}{p_{it}}=1\right\}} = \mu_0 c_{i0} + \sum \beta^t \mu_t \left.\frac{\partial c_{it}(k)}{\partial p_{io}(k)}\right|_{\frac{p_{it}(k)}{p_{it}}=1} \left[p_{it} - mc_{it}\right].$$

Since we are starting from a complete pass-through equilibrium, we denote the new exogenous marginal cost at time 0 by $mc_{i0} = \kappa$, and the new price charged by all sellers in sector i as $p_{i0} = \kappa p_i$, where p_i is the steady state value of the price. Since we use the nominal wage w as the numeraire and normalize it to 1, $mc_{i0} = \kappa$ can also be interpreted as $z_{i0} = 1/\kappa$. For $t > 0$, we assume that $mc_{it} = 1$ (temporary shock), and since the model is purely forward-looking in equilibrium, it can be shown that $p_{it} = p_i$, that is the model is back to steady state starting from period $t = 1$.

Based on the demand function (2.11), we find that:

$$\left.\frac{\partial c_{i0}(k)}{\partial p_{io}(k)}\right|_{\frac{p_{i0}(k)}{p_{i0}}=1} = -\left(\frac{p_{i0}}{\tilde{p}_0}\right)^{-\gamma-1} \frac{\tilde{c}_0}{\tilde{p}_0}\left[\gamma A(1) + f(1)\right].$$

The derivative of future consumption with respect to $p_{i0}(k)$ identifies an effect only through the extensive margin. This is because a change in price today will impact the market share in the future, but not the per-customer level of consumption (the intensive margin):

$$\left.\frac{\partial c_{it}(k)}{\partial p_{io}(k)}\right|_{\frac{p_{it}(k)}{p_{it}}=1} = -\frac{f(1)}{p_{i0}} A(1) \left(\frac{p_{it}}{\tilde{p}_t}\right)^{-\gamma} \tilde{c}_t.$$

Recall that $\tilde{p} = p_i A(1)^{\frac{1}{1-\gamma}}$, which yields

$$\left.\frac{\partial c_{it}(k)}{\partial p_{io}(k)}\right|_{\frac{p_{it}(k)}{p_{it}}=1} = -\frac{f(1)}{p_{i0}} A(1)^{\frac{1}{1-\gamma}} \tilde{c}.$$

We plug this into our initial expression and use $mc_{i0} = \kappa$ and $p_{i0} = \kappa p_i$ to get:

$$\begin{aligned}\left.\frac{\partial \widehat{\Pi}_{i0}(k)}{\partial p_{i0}(k)}\right|_{\left\{\frac{p_{it}(k)}{p_{it}}=1\right\}} &= \left(\frac{\kappa p_i}{\tilde{p}_0}\right)^{-\gamma} \mu_0 \tilde{c}_0 \left[A(1) - \frac{1}{kp_i}\left[\gamma A(1) + f(1)\right]\left[\kappa p_i - \kappa\right]\right] \\ &\quad - \sum \beta^t \mu_t \frac{f(1)}{kp_i} A(1)^{\frac{1}{1-\gamma}} \tilde{c}\left[p_i - 1\right].\end{aligned}$$

Next, we use the steady state expression for the price p_i and re-arrange in order to simplify the equation:

$$\left.\frac{\partial \widehat{\Pi}_{i0}(k)}{\partial p_{i0}(k)}\right|_{\left\{\frac{p_{it}(k)}{p_{it}}=1\right\}} = \left[\frac{\frac{\beta f(1) A(1)}{1-\beta}}{\gamma + \frac{\beta f(1)}{1-\beta} + \frac{f(1)}{A(1)}}\right]\left[\left(\frac{\kappa p_i}{\tilde{p}_0}\right)^{-\gamma} \frac{\tilde{c}_0^{1-\sigma}}{\tilde{p}_0} - \frac{\tilde{c}^{1-\sigma}}{\kappa p_i A(1)}\right],$$

where the last line uses the definition of the aggregate price index $\tilde{p}$.

We now analyze some specific cases which we consider in the text. First, notice that if firms and households do not care about the future ($\beta = 0$), the derivative of the profit function with respect to the price around a full pass-through equilibrium is simply equal to 0. In other words, when the agents are not forward-looking, full pass-through is a sustainable symmetric equilibrium, in line with our previous results. This is also true if $f(1) = 0$, that is, if the market share is price inelastic.

Second, if the shock at $t = 0$ is sector specific, the price index $\tilde{p}_0$ and the aggregate consumption remain constant as sectors are atomistic. By setting $\tilde{p}_0 = p_i A(1)^{\frac{1}{1-\gamma}}$ and $\tilde{c}_0 = \tilde{c}$ we obtain:

$$\left.\frac{\partial \widehat{\Pi}_{i0}(k)}{\partial p_{i0}(k)}\right|_{\left\{\frac{p_{it}(k)}{p_{it}}=1\right\}} = \left[\frac{\frac{\beta f(1) A(1)}{1-\beta}}{\gamma + \frac{\beta f(1)}{1-\beta} + \frac{f(1)}{A(1)}}\right] \frac{\tilde{c}^{1-\sigma}}{\kappa p_i A(1)} \left[\frac{1}{\kappa^{\gamma-1}} - 1\right].$$

In this case, there is an incentive to deviate for any non-zero shock to the marginal cost ($\kappa \neq 1$). For example, if the marginal cost increases in period 0 ($\kappa > 1$), the term in the last bracket becomes negative, indicating that a seller has an incentive to deviate from the full pass-through equilibrium by lowering its price. Therefore, under the scenario of a sector-specific shock, the symmetric equilibrium will be one where firms do not fully pass-through changes in their marginal cost.

Finally, we consider the case of an economy-wide shock hitting all sectors simultaneously. We know from our previous results that the aggregate price index can be replaced by $\tilde{p}_0 = p_{i0} A(1)^{\frac{1}{1-\gamma}} = \kappa p_i A(1)^{\frac{1}{1-\gamma}}$:

$$\left.\frac{\partial \widehat{\Pi}_{i0}(k)}{\partial p_{i0}(k)}\right|_{\left\{\frac{p_{it}(k)}{p_{it}}=1\right\}} = \left[\frac{\frac{\beta f(1) A(1)}{1-\beta}}{\gamma + \frac{\beta f(1)}{1-\beta} + \frac{f(1)}{A(1)}}\right] \frac{1}{\kappa p_i A(1)} \left[\tilde{c}_0^{1-\sigma} - \tilde{c}^{1-\sigma}\right].$$

The implications for the symmetric equilibrium are clear. Only in the case of log utility ($\sigma = 1$) there is no incentive to deviate from a full pass-through equilibrium. However, when $\sigma > 1$ there is a tendency to overshoot in the most likely case that aggregate consumption is a positive function of the productivity level. In our simulations, this effect proves to be very small for any reasonable value of σ.

5. Markup and Value of Extensive Margin

Recall our definitions for the variables v_{it} and $\bar{v}_{it}$:

$$v_{it} = \beta E_t \lambda_{it+1} A\left(1\right) \left(\frac{p_{it+1}}{\tilde{p}}\right)^{-\gamma} \tilde{c} + \beta E_t v_{it+1} \tag{.34}$$

$$\bar{v}_{it} = \lambda_{it} \left(\frac{p_{it}}{\tilde{p}}\right)^{-\gamma} \tilde{c} + v_{it}. \tag{.35}$$

Plugging (.34) into (.35), we get

$$\bar{v}_{it} = \lambda_{it} \left(\frac{p_{it}}{\tilde{p}}\right)^{-\gamma} \tilde{c} + \beta E_t \lambda_{it+1} A\left(1\right) \left(\frac{p_{it+1}}{\tilde{p}}\right)^{-\gamma} \tilde{c} + \beta E_t v_{it+1}. \tag{.36}$$

Next, we lead (.35) by one period and plug it into (.36). After rearranging we obtain:

$$\lambda_{it} \left(\frac{p_{it}}{\tilde{p}}\right)^{-\gamma} \tilde{c} + \beta E_t \lambda_{it+1} \left[A\left(1\right) - 1\right] \left(\frac{p_{it+1}}{\tilde{p}}\right)^{-\gamma} \tilde{c} = \bar{v}_{it} - \beta E_t \bar{v}_{it+1}. \tag{.37}$$

We define the gross markup as $mk_{it} = p_{it}/mc_{it}$ where $mc_{it} = w_t/z_{it}$. This implies:

$$\lambda_{it} = mc_{it} \left[mk_{it} - 1\right].$$

Plugging into (.37) and applying a first-order Taylor expansion around the steady-state yields:

$$\widehat{\Phi}_{it} + \beta \left[A\left(1\right) - 1\right] E_t \widehat{\Phi}_{it+1} \approx \bar{v}_i \left[\widehat{\bar{v}}_{it} - \beta E_t \widehat{\bar{v}}_{it+1}\right] \tag{.38}$$

where

$$\widehat{\Phi}_{it} = \left(\frac{p_i}{\tilde{p}}\right)^{-\gamma} \tilde{c}\left[mk_i\widehat{mk}_{it} + [mk_i - 1]\,\widehat{mc}_{it} - \gamma\,[mk_i - 1]\,\widehat{p}_{it}\right]. \tag{.39}$$

All the hatted variables indicate percentage deviations from steady state. Next, we simplify (.38). The second term on the left hand side of (.38) is multiplied by $[A(1) - 1]$, which is very small for almost all distributional assumptions. We therefore drop it as it is dwarfed by the other terms. Similarly, the last two terms in (.39) are multiplied by $[mk_i - 1]$. As the steady state markup is very low in our benchmark, any movements in $\widehat{mc}_{it}$ and $\widehat{p}_{it}$ will be dwarfed by fluctuations in $\widehat{mk}_{it}$.[44] Therefore, after setting $\tilde{p} = p_i$ and $\tilde{c} = 1$ to simplify the exposition without any loss of generality, we obtain the following approximate relation:

$$\widehat{mk}_{it} \approx -\frac{\bar{v}_i}{mk_i}\left[\beta E_t\widehat{\bar{v}}_{it+1} - \widehat{\bar{v}}_{it}\right].$$

[44]This is not true for all parameterizations. For example, when $\beta = 0$, we know that $mk_i = \frac{\gamma}{\gamma-1}$, which is significantly larger than 1.

6. Proof of Lemma 1

We will prove the lemma in the context of the two period example. Given the class of Normal distributions, we need to calculate the relative entropy between the reference measure $\mathbb{P}$ and an arbitrary Normal distribution $\mathbb{Q} \sim N\left(\mu_{\mathbb{Q}}, \sigma^2_{\mathbb{Q}}\right)$. Recall that the relative entropy between two distributions is defined as

$$\mathcal{R}(\mathbb{Q}) \equiv \int \ln\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right) d\mathbb{Q}.$$

We first calculate the integrand

$$\ln\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right) = \ln\frac{\sigma_{\mathbb{P}}}{\sigma_{\mathbb{Q}}} + \left[-\frac{\left(x-\mu_{\mathbb{Q}}\right)^2}{2\sigma^2_{\mathbb{Q}}} + \frac{\left(x-\mu_{\mathbb{P}}\right)^2}{2\sigma^2_{\mathbb{P}}}\right].$$

Then, we take expectations with respect to $\mathbb{Q}$

$$\begin{aligned}
\int \ln\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right) d\mathbb{Q} &= \ln\frac{\sigma_{\mathbb{P}}}{\sigma_{\mathbb{Q}}} + \mathbb{E}^{\mathbb{Q}}\left[-\frac{\left(x-\mu_{\mathbb{Q}}\right)^2}{2\sigma^2_{\mathbb{Q}}} + \frac{\left(x-\mu_{\mathbb{P}}\right)^2}{2\sigma^2_{\mathbb{P}}}\right] \\
&= \ln\frac{\sigma_{\mathbb{P}}}{\sigma_{\mathbb{Q}}} - \frac{1}{2} + \mathbb{E}^{\mathbb{Q}}\frac{\left[x-\mu_{\mathbb{P}}+\left(\mu_{\mathbb{Q}}-\mu_{\mathbb{Q}}\right)\right]^2}{2\sigma^2_{\mathbb{P}}} \\
&= \ln\frac{\sigma_{\mathbb{P}}}{\sigma_{\mathbb{Q}}} - \frac{1}{2} + \frac{\sigma^2_{\mathbb{Q}} + \left(\mu_{\mathbb{Q}}-\mu_{\mathbb{P}}\right)^2}{2\sigma^2_{\mathbb{P}}}.
\end{aligned}$$

In the log-utility case, we need to minimize (3.1) over the (potentially) distorted mean and variance. First, we take the first order condition with respect to $\mu_{\mathbb{Q}}$

$$\begin{aligned}
& \underbrace{1}_{\text{Benefit}} + \underbrace{\theta\frac{\mu_{\mathbb{Q}}-\mu_{\mathbb{P}}}{\sigma^2_{\mathbb{P}}}}_{\text{Cost}} = 0 \\
\Longrightarrow \quad & \mu_{\mathbb{Q}} = \mu_{\mathbb{P}} - \frac{\sigma^2_{\mathbb{P}}}{\theta}.
\end{aligned}$$

Therefore, the mean distortion is additive and equals to $-\frac{\sigma_{\mathbb{P}}^2}{\theta}$. Also, the first order condition with respect to the variance distortion reveals that the robust control agent chooses not to distort the variance

$$\begin{aligned} &\underbrace{\theta\left[-\sigma_{\mathbb{Q}}^{-1}+\sigma_{\mathbb{Q}}\sigma_{\mathbb{P}}^{-2}\right]}_{\text{Cost}}=0 \\ \Longrightarrow\quad &\sigma_{\mathbb{Q}}=\sigma_{\mathbb{P}} \end{aligned}$$

We next show that when the agent has power utility with risk aversion coefficient $\gamma\neq 1$ he chooses to distort both the mean and variance of the reference distribution. However, the variance distortion has a particular structure. We let $u(C)=C^{1-\gamma}/(1-\gamma),\gamma\neq 1$ and take first order condition with respect to the distorted mean

$$\underbrace{\alpha^{1-\gamma}\exp\left[(1-\gamma)\mu_{\mathbb{Q}}+(1-\gamma)^2\sigma_{\mathbb{Q}}^2/2\right]}_{\text{Benefit}}+\underbrace{\theta\frac{\mu_{\mathbb{Q}}-\mu_{\mathbb{P}}}{\sigma_{\mathbb{P}}^2}}_{\text{Cost}}=0. \tag{.40}$$

And the first order condition with respect to the variance distortion is given by

$$\underbrace{\alpha^{1-\gamma}(1-\gamma)\sigma_{\mathbb{Q}}\exp\left[(1-\gamma)\mu_{\mathbb{Q}}+(1-\gamma)^2\sigma_{\mathbb{Q}}^2/2\right]}_{\text{Benefit}}+\underbrace{\theta\left[-\sigma_{\mathbb{Q}}^{-1}+\sigma_{\mathbb{Q}}\sigma_{\mathbb{P}}^{-2}\right]}_{\text{Cost}}=0. \tag{.41}$$

These two equations can be solved numerically to obtain the optimal distortions. However, we can show that the variance distortion is linked to the mean distortion in a particular way. Divide (.41) by (.40) and rearrange to isolate for the distorted variance

$$\sigma_{\mathbb{Q}}^2=\frac{\sigma_{\mathbb{P}}^2}{1-(\gamma-1)\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)}.$$

Assuming $\gamma > 1$ and the distorted mean $\mu_{\mathbb{Q}} < \mu_{\mathbb{P}}$ then $\sigma^2_{\mathbb{Q}} > \sigma^2_{\mathbb{P}}$ and $\partial\sigma^2_{\mathbb{Q}} / \partial \left(\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\right) > 0$. This completes the proof.

7. Optimal Policies and Equilibrium

In this appendix we provide some additional details on the derivation of the optimal policies of the agent and the solution of the value function in equilibrium. With a slight abuse of notation, we write the HJB equation as

$$0 = \left[\log C_t + \frac{\theta}{2} h_t^2\right] dt + \mathbb{E}_t^{\mathbb{Q}} dJ - \rho J dt. \tag{.42}$$

We posit the following guess for the agent's value function

$$J\left(W_t, v_t\right) = \frac{\log W_t}{\rho} + \delta_0 + \delta_1 v_t. \tag{.43}$$

Applying Ito's lemma to (.43) and omitting time subscripts for convenience we get

$$\begin{aligned}
dJ &= J_W dW + J_v dv + \frac{1}{2} J_{WW} \left[dW\right]^2 \\
&= \left[J_W \mu_W^{\mathbb{Q}} + J_v \left(a_0 + a_1 v + \sigma_v h \sqrt{v}\right) + \frac{1}{2} J_{WW} \sigma_W^2\right] dt + \left[J_W \sigma_W + J_v \sqrt{v} \sigma_v\right] dB^{\mathbb{Q}} \\
&= \left[\frac{1}{\rho W}\left(rW + \alpha W\left(\mu_R - r\right) - C + \alpha W \sigma_R h\right) + \delta_1 \left(a_0 + a_1 v + \sigma_v h \sqrt{v}\right) - \frac{\alpha^2 \sigma_R^2}{2\rho}\right] dt \\
&\quad + \left[\frac{\alpha \sigma_R}{\rho} + \delta_1 \sqrt{v} \sigma_v\right] dB^{\mathbb{Q}}.
\end{aligned} \tag{.44}$$

Next, for the minimization problem we take the derivative of the value function with respect to h and obtain

$$\begin{aligned}
&\underbrace{\theta h}_{\text{Cost}} + \underbrace{\frac{\alpha \sigma_R}{\rho} + \delta_1 \sigma_v \sqrt{v}}_{\text{Benefit}} = 0 \\
\Longrightarrow \quad & h = -\frac{1}{\theta}\left(\frac{\alpha \sigma_R}{\rho} + \delta_1 \sigma_v \sqrt{v}\right).
\end{aligned}$$

Recall that in equilibrium the agent holds the entire claim on the output process, and thus $\alpha = 1$. This yields expression (3.23).

Deriving α requires taking first order conditions in the maximization problem which shows up only in the drift of dJ. Also, deriving the consumption policy yields the usual envelope type condition $u'(C) = J_W$.

We now solve for the parameters δ_0 and δ_1. First, plug in (.43) and (.44) into (.42) and use the optimal policies for h, α and C and the equilibrium risk free rate and market return

$$\begin{aligned}
0 &= \log\rho + \log W + \frac{v}{2\theta}\left(\frac{1}{\rho} + \delta_1\sigma_v\right)^2 - \rho\left(\frac{1}{\rho}\log W + \delta_0 + \delta_1 v\right) \\
&\quad + \frac{1}{\rho W}\left(rW + \alpha W\left(\mu_R - r\right) - C + \alpha W\sigma_R h\right) + \delta_1\left(a_0 + a_1 v + \sigma_v\sqrt{v}h\right) - \frac{\alpha^2\sigma_R^2}{2\rho}, \\
0 &= \log\rho + \frac{v}{2\theta}\left(\frac{1}{\rho^2} + \frac{2\delta_1\sigma_v}{\rho} + \delta_1^2\sigma_v^2\right) - \rho\delta_0 - \rho\delta_1 v \\
&\quad + \frac{1}{\rho}\left[\mu - \frac{v}{\theta}\left(\frac{1}{\rho} + \delta_1\sigma_v\right)\right] + \delta_1\left(a_0 + a_1 v - \sigma_v\frac{v}{\theta}\left(\frac{1}{\rho} + \delta_1\sigma_v\right)\right) - \frac{v}{2\rho}.
\end{aligned}$$

Collecting coefficients for v

$$\frac{1}{2\theta}\left(\frac{1}{\rho^2} + \frac{2\delta_1\sigma_v}{\rho} + \delta_1^2\sigma_v^2\right) - \rho\delta_1 - \frac{1}{\rho\theta}\left(\frac{1}{\rho} + \delta_1\sigma_v\right) + \delta_1 a_1 - \delta_1\frac{\sigma_v}{\theta}\left(\frac{1}{\rho} + \delta_1\sigma_v\right) - \frac{1}{2\rho} = 0.$$

Rearranging to get a quadratic in δ_1

$$\begin{aligned}
\left(\frac{\sigma_v^2}{2\theta} - \frac{\sigma_v^2}{\theta}\right)\delta_1^2 + \left(\frac{\sigma_v}{\theta\rho} + a_1 - \rho - \frac{\sigma_v}{\theta\rho} - \frac{\sigma_v}{\theta\rho}\right)\delta_1 + \left(\frac{1}{2\theta\rho^2} - \frac{1}{2\rho} - \frac{1}{\theta\rho^2}\right) &= 0, \\
\frac{\sigma_v^2}{2\theta}\delta_1^2 - \left(a_1 - \rho - \frac{\sigma_v}{\theta\rho}\right)\delta_1 + \left(\frac{1}{2\rho} + \frac{1}{2\theta\rho^2}\right) &= 0.
\end{aligned}$$

Solve the quadratic equation

$$\begin{aligned} A &= \frac{\sigma_v^2}{2\theta}, \\ B &= -\left(a_1 - \rho - \frac{\sigma_v}{\theta\rho}\right), \\ C &= \frac{1}{2\rho} + \frac{1}{2\theta\rho^2}. \end{aligned}$$

To prove that it is indeed an equilibrium check that

$$\begin{aligned} B^2 - 4AC &= \left(a_1 - \rho - \frac{\sigma_v}{\theta\rho}\right)^2 - 4\frac{\sigma_v^2}{2\theta}\left(\frac{1}{2\rho} + \frac{1}{2\theta\rho^2}\right) \\ &= \left(a_1 - \rho - \frac{\sigma_v}{\theta\rho}\right)^2 - \frac{\sigma_v^2}{\rho\theta}\left(1 + \frac{1}{\theta\rho}\right) \\ &= \left(a_1 - \rho\right)^2 - 2\left(a_1 - \rho\right)\frac{\sigma_v}{\theta\rho} - \frac{\sigma_y^2}{\theta\rho} \\ &\geq \left(a_1 - \rho\right)^2 - 2\left(a_1 - \rho\right)\frac{\sigma_v}{\theta\rho} - \frac{\sigma_y^2}{\theta^2\rho^2} \\ &= \left(a_1 - \rho - \frac{\sigma_v}{\theta\rho}\right)^2 \\ &\geq 0. \end{aligned}$$

The first inequality follows from the fact that $\frac{\sigma_v^2}{\theta\rho} \geq 0$ and $0 \leq \theta\rho < 1$.

With a solution for δ_1, we find δ_0 by collecting the constant terms

$$\begin{aligned} \log\rho - \rho\delta_0 + \frac{\mu}{\rho} + \delta_1 a_0 = 0 \\ \Longrightarrow \quad \delta_0 = \frac{1}{\rho}\left(\log\rho + \frac{\mu}{\rho} + \delta_1 a_0\right). \end{aligned}$$

8. Pricing the Term Structure

In this appendix we give a more detailed derivation of the bond price. We use the partial differential equation approach which is very common when pricing fixed income securities.

Applying Ito's lemma to (3.26) we derive the dynamics of the bond price with arbitrary maturity

$$\frac{dp_t}{p_t} = -\left[\beta_0'(\tau) + \beta_1'(\tau)v_t\right]dt + \beta_1(\tau)\,dv_t + \frac{1}{2}\beta_1^2(\tau)\left[dv_t\right]^2. \tag{.45}$$

Next, plug (3.5), (3.24), (3.25) and (.45) into (3.27) to get

$$\begin{aligned} 0 &= -\left[\beta_0'(\tau) + \beta_1'(\tau)v_t\right] + \beta_1(\tau)\left[a_0 + \left(a_1 + \sigma_v(1-\phi)\right)v_t\right] + \frac{1}{2}\beta_1^2(\tau)\,\sigma_v^2 v_t \\ &\quad -(\rho + b_0) + \phi v_t - \beta_1(\tau)\,\sigma_v v_t. \end{aligned}$$

Collecting the coefficients of v and the free coefficients we get two simple ordinary differential equations. The first is a Riccati equation with constant coefficients

$$\beta_1'(\tau) = \frac{1}{2}\sigma_v^2\beta_1^2(\tau) + (a_1 - \phi\sigma_v)\,\beta_1(\tau) + \phi,$$

and the second becomes trivial after we solve for β_1

$$\beta_0'(\tau) = \beta_1(\tau)\,a_0 - (\rho + b_0),$$

with the boundary conditions

$$\beta_0(0) = \beta_1(0) = 0.$$

let $\bar{\beta}_1$ be a particular (constant) solution. In that case $\bar{\beta}_1' = 0$, and $\bar{\beta}_1$ is given by

$$\bar{\beta}_1 = \frac{-\left(a_1 - \phi\sigma_v\right) - \sqrt{\left(a_1 - \phi\sigma_v\right)^2 - 2\sigma_v^2\phi}}{\sigma_v^2}.$$

Let $\beta_1 = \bar{\beta}_1 + \frac{1}{z}$. Then

$$\begin{aligned}
\left[\bar{\beta}_1\left(\tau\right) + \frac{1}{z\left(\tau\right)}\right]' &= \frac{1}{2}\sigma_v^2\left[\bar{\beta}_1 + \frac{1}{z\left(\tau\right)}\right]^2 + \left(a_1 - \phi\sigma_v\right)\left[\bar{\beta}_1 + \frac{1}{z\left(\tau\right)}\right] + \phi, \\
-\frac{z'\left(\tau\right)}{z^2\left(\tau\right)} &= \frac{1}{2}\sigma_v^2\left[\frac{2\bar{\beta}_1}{z\left(\tau\right)} + \frac{1}{z^2\left(\tau\right)}\right] + \left(a_1 - \phi\sigma_v\right)\frac{1}{z\left(\tau\right)} \\
&\Longrightarrow z'\left(\tau\right) + \left[\sigma_v^2\bar{\beta}_1 + \left(a_1 - \phi\sigma_v\right)\right]z\left(\tau\right) + \frac{1}{2}\sigma_v^2 = 0.
\end{aligned}$$

The solution is derived by simple integration. The boundary condition on z is determined through the boundary condition on β_1. Since $\beta_1\left(0\right) = 0$ we have that $\bar{\beta}_1 + \frac{1}{z(0)} = 0 \Longrightarrow z\left(0\right) = -\frac{1}{\bar{\beta}_1}$. Define

$$\begin{aligned}
\Xi &\equiv \sigma_v^2\bar{\beta}_1 + \left(a_1 - \phi\sigma_v\right) \\
&= -\sqrt{\left(a_1 - \phi\sigma_v\right)^2 - 2\sigma_v^2\phi}.
\end{aligned}$$

Then,

$$\begin{aligned}
z\left(\tau\right)e^{\Xi\tau} &= -\frac{1}{2}\sigma_v^2\int e^{\Xi s}ds + const. \\
&= -\frac{1}{2}\sigma_v^2\left[\frac{e^{\Xi s}}{\Xi}\right|_0^\tau + const. \\
&= -\frac{1}{2}\sigma_v^2\left[\frac{e^{\Xi\tau} - 1}{\Xi}\right] + const..
\end{aligned}$$

Taking into account the normalizing constant, we get

$$
\begin{aligned}
z(\tau) &= -\frac{1}{2}\sigma_v^2\left[\frac{1-e^{-\Xi\tau}}{\Xi}\right]-\frac{e^{-\Xi\tau}}{\bar{\beta}_1} \\
&= -\frac{\sigma_v^2}{2\Xi}+\left[\frac{\sigma_v^2}{2\Xi}-\frac{1}{\bar{\beta}_1}\right]e^{-\Xi\tau} \\
&= \zeta_0+\zeta_1 e^{-\Xi\tau}.
\end{aligned}
$$

Finally, we need to back-out $\beta_0(\tau)$ with the boundary condition $\beta_0(0)=0$

$$
\begin{aligned}
\beta_0(\tau) &= a_0\int \beta_1(s)\,ds-(\rho+b_0)\,\tau+const. \\
&= a_0\int\left[\bar{\beta}_1+\frac{1}{z(s)}\right]ds-(\rho+b_0)\,\tau+const. \\
&= a_0\int\frac{ds}{z(s)}-\left(\rho+b_0-a_0\bar{\beta}_1\right)\tau+const. \\
&= a_0\left[\frac{s}{\zeta_0}+\frac{1}{\Xi\zeta_0}\ln\left|\zeta_0+\zeta_1 e^{-\Xi s}\right|\right|_0^\tau-\left(\rho+b_0-a_0\bar{\beta}_1\right)\tau+const. \\
&= a_0\left[\frac{\tau}{\zeta_0}+\frac{1}{\Xi\zeta_0}\ln\left|\frac{\zeta_0+\zeta_1 e^{-\Xi\tau}}{\zeta_0+\zeta_1}\right|\right]-\left(\rho+b_0-a_0\bar{\beta}_1\right)\tau+const. \\
&= \frac{a_0}{\Xi\zeta_0}\ln\left|\frac{\zeta_0+\zeta_1 e^{-\Xi\tau}}{\zeta_0+\zeta_1}\right|-\left(\rho+b_0-a_0\bar{\beta}_1-\frac{a_0}{\zeta_0}\right)\tau.
\end{aligned}
$$

Given the bond pricing rule, one can easily price the forward yield curve. Let $f(\tau;v_t)$ be the *instantaneous* forward rate contracted at time t for delivery at time $t+\tau$ (i.e., instantaneous borrowing or lending at time $t+\tau$). Then,

$$
\begin{aligned}
f(\tau;v_t) &= -\frac{p_\tau(\tau;v_t)}{p(\tau;v_t)} \\
&= -\left[\beta_0'(\tau)+\beta_1'(\tau)\,v_t\right],
\end{aligned}
$$

where p_τ is the derivative of p with respect to maturity τ.

Similarly, given the prices of all default-free zero-coupon bonds, we can price any arbitrary forward contract. Let $F(\tau, s; v_t)$ be the forward rate (price) contracted at time t for delivery at time $t+\tau$ with maturity $t+s$, where $s \geq \tau$. Then,

$$\begin{aligned} F(\tau, s; v_t) &\equiv \frac{\ln p(\tau; v_t) - \ln p(s; v_t)}{s-\tau} \\ &= \frac{1}{s-\tau} \times \left[-\tau \mathcal{Y}(\tau; v_t) + s\mathcal{Y}(s; v_t)\right] \\ &= \frac{1}{s-\tau} \left\{\left[\beta_0(\tau) - \beta_0(s)\right] + \left[\beta_1(\tau) - \beta_1(s)\right] v_t\right\}. \end{aligned}$$

Note that $\lim_{s \downarrow \tau} F(\tau, s; v_t) = f(\tau; v_t)$ and $\lim_{\tau, s \uparrow \infty} F(\tau, s; v_t) = r_t$.

Using forward rates, one can conduct regression analysis as in Fama and Bliss (1987) and Backus et al. (1998) to verify the failure of the expectation hypothesis (return predictability).

9. Data

Unless otherwise stated, all data are quarterly from $Q2.1952 - Q4.2006$.

- McCulloch-Kwon-Bliss data set: nominal prices and yields of zero coupon bonds - see McCulloch and Kwon (1993) and Bliss (1999). In the estimation exercises we use only the 3 month and 1 year nominal yields at the quarterly frequency to create the real counterparts. The data we use spans the period $Q2.52 - Q4.96$
- Treasury Inflation-Protected Securities (TIPS) data from McCulloch: real yields from $M1.97 - M12.06$. Although the data is available at higher frequencies, we use only observations at the quarterly frequency
- Quarterly market index (NYSE/AMEX/NASDAQ) including distributions from CRSP
- Quarterly CPI (all items), SA, from the BLS (see FREDII data source maintained by the federal reserve bank of St. Louis for full description)
- Semi-annual inflation expectations from the Livingston survey (maintained by the federal reserve bank of Philadelphia) - period $H1.52 - H1.81$. From $Q3.81$ quarterly inflation expectations data from the Survey of Professional Forecasters (SPF) becomes availabe
- Quarterly inflation expectations from the SPF maintained by the federal reserve bank of Philadelphia. The sample period covers $Q3.81 - Q4.06$
- Quarterly real Personal Consumption Expenditures (PCE): *services* and *non-durables* from the BEA, SA
- Quarterly real Personal Consumption Expenditures PCE: *imputed services of durables* from the Federal Board of Governors

- Civilian Noninstitutional Population series from the BLS
- Monthly real dividends obtained from Robert Shiller's website over the period $M1.52 - M12.06$ (http://www.econ.yale.edu/~shiller/data.htm). This data set was used in the GARCH-GJR exercise

Since we use only real data in the estimation, we convert nominal prices to real ones using the price level data. For the short rate (3 months) we use a 3 year moving average of realized inflation to construct a 3 month ahead expected inflation measure. For the 1 year yield we use both the Livingston and SPF survey data to construct a quarterly series of expected inflation. The SPF is sampled at quarterly frequency but it is available only in the latter part of the sample. We interpolate the semi-annual Livingston data to construct quarterly data using piecewise cubic Hermite interpolation.

10. Computing Detection Error Probabilities

In this appendix we shortly discuss how we compute DEP's. The discussion is based on chapter 9 in Hansen and Sargent (2007a). The econometrician observes $\left\{\frac{\Delta C_{t+1}}{C_t}\right\}_{t=1}^{T}$ and construct the log-likelihood ratio of the distorted model relative to the objective model

$$\ell^T = \sum_{t=1}^{T} \log \frac{f\left(\frac{\Delta C_{t+1}}{C_t}|\theta < \infty\right)}{f\left(\frac{\Delta C_{t+1}}{C_t}|\theta = \infty\right)}.$$

The distorted model is denoted with $f\left(\cdot|\theta < \infty\right)$ and the reference model is denoted with $f\left(\cdot|\theta = \infty\right)$. The distorted model is selected when $\ell^T > 0$ and the objective model is selected otherwise.

There are two types of detection errors:

(1) Choosing the distorted model when actually the reference model generated the data

$$\mathbb{P}\left(\ell^T > 0|\theta = \infty\right) = \mathbb{E}\left(\mathbf{1}_{\{\ell^T > 0\}}|\theta = \infty\right).$$

(2) Choosing the reference model when actually the distorted model generated the data

$$\begin{aligned}
\mathbb{P}\left(\ell^T < 0|\theta < \infty\right) &= \mathbb{E}\left(\mathbf{1}_{\{\ell^T < 0\}}|\theta < \infty\right) \\
&= \mathbb{E}\left(\exp\left(\ell^T\right)\mathbf{1}_{\{\ell^T < 0\}}|\theta = \infty\right).
\end{aligned}$$

Therefore, the *average error (denoted $\wp$) with a prior of equiprobable models* is

$$
\begin{aligned}
\wp &= \frac{1}{2}\left[\mathbb{P}\left(\ell^{T} > 0|\theta = \infty\right) + \mathbb{P}\left(\ell^{T} < 0|\theta < \infty\right)\right] \\
&= \frac{1}{2}\mathbb{E}\left\{\min\left[\exp\left(\ell^{T}\right), 1\right]|\theta = \infty\right\}. \qquad (.46)
\end{aligned}
$$

We can write an (approximate) transition likelihood ratio as

$$
\begin{aligned}
\frac{f\left(\frac{\Delta C_{t+1}}{C_t}|\theta < \infty\right)}{f\left(\frac{\Delta C_{t+1}}{C_t}|\theta = \infty\right)} &= \exp\left[-\frac{1}{2} \times \frac{\left(\frac{\Delta C_{t+1}}{C_t} - \mu - h_t\sqrt{v_t}\right)^2 - \left(\frac{\Delta C_{t+1}}{C_t} - \mu\right)^2}{v_t}\right] \\
&= \exp\left[-\frac{1}{2} \times \frac{-2\left(\frac{\Delta C_{t+1}}{C_t} - \mu\right)(1-\phi)\, v_t + (1-\phi)^2 v_t^2}{v_t}\right] \\
&= \exp\left[\left(\frac{\Delta C_{t+1}}{C_t} - \mu\right)(1-\phi) - \frac{(1-\phi)^2 v_t}{2}\right].
\end{aligned}
$$

We simulate the economy $5,000$ times using the point estimates of the parameters and construct a likelihood ratio for each economy. Using (.46) we can immediately derive $\wp$.