Ambiguity Shifts and the 2007–2008 Financial Crisis*

Nina Boyarchenko

July 15, 2011

Abstract

I analyze the effects of model misspecification on default swap spreads and equity prices for firms that are informationally opaque to the investors. The agents in the economy are misspecification-averse and thus assign higher probabilities to lower utility states. This leads to higher CDS rates, lower equity prices and lower expected times to default. Estimating the model using data on financial institutions, I find that the sudden increase in credit spreads in the summer of 2007 can be partially explained by agents' mistrust of the signals observed in the market. The bailout of Bear Stearns in March 2008 and the liquidation of Lehman Brothers in September 2008 further exacerbated the agents' doubts about signal quality and introduced mistrust about the agents' pricing models, accounting for the further increases in credit spreads after these events.

*Email: nboyarch@chicagobooth.edu. Earlier versions of this paper were circulated under the title "Ambiguity, Information Quality and Credit Risk". I would like to thank my committee Lars P. Hansen (co-chair), Pietro Veronesi (co-chair), Stavros Panageas and John Heaton. I would also like to thank Lubos Pastor, Martin Schneider, Narayana Kocherlakota, three anonymous faculty reviewers at the University of Chicago Booth School of Business, Ariel Zetlin-Jones, Andrei Malenko, Matthew Plosser, Valentin Haddad, participants in the Finance Brownbag, University of Chicago Booth School of Business, Chicago-Minnesota PhD conference, Transatlantic PhD Conference, Summer Meetings of the Econometrics Society, Boston, Society of Economic Dynamics Meetings, Istanbul, European Meetings of the Econometric Society, Barcelona, the Finance Lunch, University of Chicago Booth School of Business and seminar participants at Johns Hopkins University, Carey School of Business, Federal Reserve Board of Governors, Federal Reserve Bank of New York, Ohio State University, Fisher College of Business, the Einaudi Institute for Economics and Finance and Bocconi University for helpful comments on the previous drafts of this paper. All the remaining errors are mine. Research support from the Sanford J. Grossman Fellowship in Honor of Arnold Zellner is gratefully acknowledged; any opinions expressed herein are the author's and not necessarily those of Sanford J. Grossman or Arnold Zellner.

1 Introduction

When making consumption decisions, an investor faces uncertainty about both the relevant underlying state and the data-generating process governing the evolution of the state. While uncertainty about the state is risk that the investor understands and can model, uncertainty about the data-generating process represents agents' pessimism about their ability to identify the correct model. This paper argues that prices of credit securities are sensitive to the investors' preferences toward model uncertainty and that the implied time-variation in the level of model uncertainty is a source of variation in credit spreads that explains the asymmetric response of credit spreads to upturns and downturns in the economy.

I analyze the effects of model misspecification on default swap spreads in secondary markets for the corporate debt of firms that are not perfectly transparent to the investors. The agents in the economy are misspecification-averse and thus mistrust the statistical model of the fundamental value of assets of firms and of the firms' observed earnings process. This mistrust reflects the fact that, when it is difficult for investors to observe firms' assets directly, they are forced to rely on imprecise accounting information. In this situation, investors must draw inference from accounting data and other publicly available information. Investors realize that, although they may be able to pick a model of the fundamental asset value and the accounting signals to best fit the historical data, this may not be the true data-generating model. Under the assumption that investors are misspecification-averse, I derive the asset prices in the economy, explicitly accounting for the implications of imperfect information and model misspecification.

I show several significant implications of model misspecification for the level and variation in the term structure of secondary market default swap spreads. Compared to a model with perfect information, model uncertainty increases the level of the yield curve and the default swap spread curve. Intuitively, in the presence of model misspecification, investors must be compensated for the risk associated with choosing the "wrong" model to describe the evolution of the underlying state. Notice that, as shown in Duffie and Lando (2001), introducing imperfect information to a standard Black and Cox (1976) model has the additional benefit of being able to explain high credit spreads at short maturities.

Next, I show that model uncertainty exacerbates the imperfect information problem faced by the representative investors in the secondary asset markets. In filtering information about the underlying state from the imperfect signals, agents must take into account uncertainty about both the model governing the evolution of the underlying state and the signals about the underlying state. The misspecification-averse agent assigns higher probabilities to lower utility states. Further, how much these probabilities are higher than under the reference model depends on the current conditional probability vector under the reference model.

Model misspecification also impacts the joint probability distribution of the next period's signals and states. In particular, while in states of the economy when no firm defaults, the misspecification averse agent perceives the probability of transitioning to a default state to be higher than under the reference model. Thus, the expected time to default of each firm decreases, increasing default swap spreads. Further, the misspecification averse agent also perceives the transition probability matrix associated with the underlying state to be time varying. The time-variation in the transition probability matrix induces additional time-variation in the expected time to default of each firm and, thus, in default swap spreads.

In this paper, I argue that the increases in CDS spreads observed during the 2007–2008 crisis were due to increases in investors' doubts about the validity of their pricing models and the quality of the signals available to market participants. On August 9, 2007, France's largest bank BNP Paribas announced that it was having difficulties because two of its off-balance-sheet funds had loaded up on securities based on American subprime mortgages. But Paribas was not alone in its troubles: a month before, the German bank IKB announced similar difficulties, and the Paribas announcement was followed the next day by Northern Rock's revelation that it had only had enough reserve cash to last until the end of the month. These and other similar announcements lead to a freeze of the credit markets as banks lost faith in each other's balance sheets. The situation was particularly surprising considering the market conditions shortly before the crisis began. At the beginning of 2007, financial markets were liquidity-unconstrained and credit spreads were at historical lows. Even as late as May 2007, it would have been hard to predict the magnitude of the response that the losses on subprime mortgages had generated. Compared to the total value of financial instruments traded worldwide, the subprime losses were relatively small: even the worst-case estimates put them at around USD 250 billion. [1] Further, for investors familiar with the instruments, the losses were not unexpected. By definition, the subprime mortgages were part of the riskiest segment of the mortgage market, so it was hardly surprising some borrowers would default on the loans. Yet, despite their predictability, the defaults had precipitated the current liquidity crisis that spread between the credit markets.

(Figure 1 about here.)

Using observations of the CDS spreads on financial institutions, I estimate the degree of misspecification aversion of the investors in the secondary debt markets. To evaluate the changes in investors' aversion to misspecification during the crisis, I estimate the misspecification aversion coefficients using three sub-periods– before the start of the crisis in July 2007, from the start of the crisis to the bailout of Bear Stearns in March 2008, and

[1] Source: Caballero and Krishnamurthy (2008b)

from the bailout of Bear Stearns to the liquidation of Lehman Brothers in September 2008 – and find that the three estimates are not statistically significantly different, implying that investors' misspecification preferences did not change during the crisis. In terms of the model, this implies that the observed changes in credit spreads during the financial crisis were due not to changing misspecification attitudes on the part of the investors but rather due to an increase in the amount of misspecification in the economy.

Since the investors' aversion to misspecification did not change significantly during the crisis, I use the pre-crisis estimate of the misspecification aversion coefficient to compute the model-implied time series of CDS spreads, equity prices and the quantity of misspecification. Measuring the amount of misspecification using the expected log likelihood ratio (or relative entropy) between the reference and the worst-case models, I find that the amount of misspecification did in fact increase during the financial crisis. Further, the way that total entropy is decomposed into the contribution from misspecification of the distribution of the future signals and state and the contribution from misspecification of the conditional distribution over the current state changed during the crisis. More specifically, the initial BNP Paribas announcement in August 2007 lead to an increase in the relative entropy due to current period conditional probability misspecification. Intuitively, the BNP Paribas announcement and the subsequent Northern Rock revelations lead to an increase in ambiguity about the quality of the signals available to market participants. The bailout of Bear Stearns and the bailout of Lehman Brothers, on the other hand, lead to an increase in entropy due to misspecification of the distribution of future signals and states. That is, the effective default of these two institutions increased investors' doubts about the quality of the surviving financial institutions. Entropy stabilizes toward the end of the crisis but at a higher level than before the start of the crisis.

To evaluate the quality of the fit of the model, I compare the model-implied CDS spreads and equity prices to the corresponding quantities observed in the data. The model-implied CDS spreads match both the levels and the changes in CDS spreads observed during the crisis, although the performance of the model deteriorates after the liquidation of Lehman Brothers. Further, although the model does not match the overall levels of equity prices and, in fact, is not geared to do so, it does match the changes in equity prices observed during the crisis.

Next, I examine the impact of each source of ambiguity on CDS spreads in greater detail. I reformulate the model to allow investors to have different degrees of aversion to ambiguity about the underlying dynamics and ambiguity about the signal quality. In this formulation, I can alternatively set one or the other degree of ambiguity aversion to zero and thus examine separately the contributions to the behavior of CDS rates from the two sources of ambiguity. I find that, while both sources of ambiguity are important for matching the time series evolution of CDS rates, the relative importance of the contributions from different sources evolves over time. At the start of the crisis, aversion to ambiguity about signal quality plays a greater role while in the later part of the crisis the behavior of CDS rates is influenced more strongly by aversion to ambiguity about the underlying dynamics.

The rest of the paper is organized as follows. I review the related literature in Section 2. In Section 3, I solve a three period example to illustrate the impact of ambiguity aversion on investors' beliefs. I describe the model considered in the paper in Section 4. The results of the estimation of the model are presented in Section 5. Section 6 concludes. Technical details are relegated to the appendix.

2 Literature Review

A rapidly growing literature studies the behavior of asset prices in the presence of ambiguity in dynamic economies. A substantial part of this literature considers investor ambiguity about the data-generating model. Anderson et al. (2003) derive the pricing semigroups associated with robust perturbations of the true state probability law. Trojani and Vanini (2002) use their framework to address the equity premium and the interest rate puzzles, while Leippold et al. (2008) consider also the excess volatility puzzle. Gagliardini et al. (2009) study the term structure implications of adding ambiguity to a production economy. This setting has also been used to study the portfolio behavior of ambiguity-averse investors and the implications for the options markets (see e.g. Trojani and Vanini (2004) and Liu et al. (2005)).

The second strand in the literature, however, assumes that, although the agents in the economy know the "true" data-generating model, they face uncertainty about the quality of the observed signal about an unobservable underlying. Chen and Epstein (2002) study the equity premium and the interest rate puzzles in this set-up, and Epstein and Schneider (2008) consider the implications for the excess volatility puzzle. The portfolio allocation implications of this setting have also been studied extensively in e.g. Uppal and Wang (2003) and Epstein and Miao (2003).

However, none of these papers study the relationship between ambiguity aversion and the term structure of credit spreads. Following Hansen and Sargent (2005, 2007), I introduce model misspecification by considering martingale distortions to the reference model probability law. As Hansen and Sargent (2007) show, the martingale distortion can be factored into distortions of the conditional distribution of the underlying state (signal quality) and the evolution law of the hidden state (asset value dynamics). I assume that the representative investor in the secondary debt market has max-min preferences over consumption paths under possible models.

This paper is also related to the literature on preference-based explanations for credit spreads. Chen (2010) studies two puzzles about corporate debt: the credit spread puzzle – why yield spreads between corporate bonds and treasuries are high and volatile – and the under-leverage puzzle – why firms use debt conservatively despite seemingly large tax benefits and low costs of financial distress. The paper argues that both of these puzzles can be explained by two observations: defaults are more highly concentrated during bad times, when marginal utility is high, and the losses associated with default are higher during such times. Thus, investors demand high risk premia for holding defaultable claims, including corporate bonds and levered firms.

Using similar intuition, Chen et al. (2009) argue that the credit spread puzzle can be explained by the covariation between default rates and market Sharpe ratios. That is, investors must be compensated more for holding credit risk securities because default rates (and, hence, expected losses from default) increase at the same time as market returns are more uncertain. More specifically, the authors investigate the credit spread implications of the Campbell and Cochrane (1999) pricing kernel calibrated to equity returns and aggregate consumption data. Identifying the historical surplus–consumption ratio from aggregate consumption data, the paper finds that the implied level and time-variation of spreads match historical levels well.

3 Three Period Example

In this section, I consider a three period economy consisting of just one firm. In this setting, I can examine analytically the effect of ambiguity aversion on the investors' perceived beliefs. The setting in this section is an extremely simplified version of the model in Section 4 and provides the main intuition for the more general results.

More specifically, consider a firm whose payoffs V each period are given as in the event tree in Fig. 2. At date 0, the firm pays V_0 to the investors for sure. Between dates 0 and 1, the payoff from the firm increases to V_h with probability p_h or decreases to V_l with probability $p_l = 1 - p_h$. Similarly, the probability of the firm payoff increasing between dates 1 and 2 is p_h and the probability of the firm payoff decreasing between dates 1 and 2 is $p_l = 1 - p_h$. Finally, notice that, by assumption, the tree is recombining so that the firm payoff at date 2 if the firm payoff first decreases between dates 0 and 1 and then increases between dates 1 and 2 is the same as the payoff of a firm which first increases and then decreases in value. This simplifying assumption reduces notational complexity without detracting from the main intuition of the model.

(Figure 2 about here.)

In this example, I focus on the investors' problem at date 1. In particular, I assume that, at the intermediate date 1, the investors in the firm do not observe the realized payoff V_1 (which is paid into an interest-free account on their behalf). Instead, the representative investor in the firm observes an imperfect but unbiased signal about the true realization: $y = V_1 + u$. Given a signal y at date 1, the representative investor forms a posterior probability π_h of the firm payoff being in the high state at date 1:

$$\pi_h = \frac{p_h f(y - V_h)}{p_h f(y - V_h) + p_l f(y - V_l)},$$

where $f(\cdot)$ is the probability distribution function of the signal error, u. The corresponding posterior probability of being in the low state at date 1 is then given by:

$$\pi_l = \frac{p_l f(y - V_h)}{p_h f(y - V_h) + p_l f(y - V_l)}.$$

The investors in this firm are risk-neutral but ambiguity-averse. That is, in making investment decisions at date 1, the investors in the firm consider two questions:

1. Given a realization of the firm payoff at date 1, what is the worst-case probability of the firm payoff increasing between dates 1 and 2?
2. Given a signal about the firm payoff at date 1,what is the worst-case probability of the firm payoff being in the high state (i.e. $V_1 = V_h$) at date 1?

The first question captures investors doubts about the validity of the statistical model they have for the evolution of firm payoffs in the future. Notice that, although I consider the imperfect information setting in this example, this concern remains valid even in perfect information settings as, in making investment and consumption decisions, agents make predictions about the future evolution of their cash flows. The second question captures investors doubts about the validity of filter that they use to construct the posterior probabilities at date 1.

More formally, denote by $\tilde{p}_{hh}$ the perceived probability of the firm payoff increasing from V_h at date 1 to V_{hh} at date 2, $\tilde{p}_{lh}$ the perceived probability of the firm payoff increasing from V_l at date 1 to V_{hl} at date 2 and by $\tilde{\pi}_h$ the perceived probability of V_h being the realized firm payoff at date 1. The corresponding probability tree with distorted probabilities is presented in Fig. 3. The ambiguity-averse investor solves:

$$\begin{aligned}\min_{\tilde{\pi},\tilde{p}} \quad & \tilde{\pi}_h \left(V_h + \tilde{p}_{hh}V_{hh} + \tilde{p}_{hl}V_{hl}\right) + \tilde{\pi}_l \left(V_l + \tilde{p}_{lh}V_{lh} + \tilde{p}_{ll}V_{ll}\right) \\ + \quad & \theta_d \left[\tilde{\pi}_h \left(\tilde{p}_{hh} \log \frac{\tilde{p}_{hh}}{p_h} + \tilde{p}_{hl} \log \frac{\tilde{p}_{hl}}{p_l}\right) + \tilde{\pi}_l \left(\tilde{p}_{lh} \log \frac{\tilde{p}_{lh}}{p_h} + \tilde{p}_{ll} \log \frac{\tilde{p}_{ll}}{p_l}\right)\right] \\ + \quad & \theta_s \left(\tilde{\pi}_h \log \frac{\tilde{\pi}_h}{\pi_h} + \tilde{\pi}_l \log \frac{\tilde{\pi}_l}{\pi_l}\right)\end{aligned} \tag{3.1}$$

where the minimization is subject to:

$$\begin{aligned} 1 &= \tilde{\pi}_h + \tilde{\pi}_l \\ 1 &= \tilde{p}_{hh} + \tilde{p}_{hl} \\ 1 &= \tilde{p}_{lh} + \tilde{p}_{ll}. \end{aligned}$$

The minimization problem (3.1) has three components. The first,

$$\tilde{E}[V_1 + V_2|y] = \tilde{\pi}_h \left(V_h + \tilde{p}_{hh} V_{hh} + \tilde{p}_{hl} V_{hl}\right) + \tilde{\pi}_l \left(V_l + \tilde{p}_{lh} V_{lh} + \tilde{p}_{ll} V_{ll}\right),$$

is the usual expected utility component but calculated using the distorted probabilities. The remaining two components measure the entropy between the reference and the misspecified probability laws and discipline the choice of the worst-case likelihood. The entropy due to a misspecification of the underlying dynamics (that is, the entropy associated a different choice of $\tilde{p}_{hh}$ and $\tilde{p}_{lh}$) is:

$$\epsilon^1(\tilde{p}_{hh}, \tilde{p}_{lh}, \tilde{\pi}_h) = \tilde{\pi}_h \left(\tilde{p}_{hh} \log \frac{\tilde{p}_{hh}}{p_h} + \tilde{p}_{hl} \log \frac{\tilde{p}_{hl}}{p_l}\right) + \tilde{\pi}_l \left(\tilde{p}_{lh} \log \frac{\tilde{p}_{lh}}{p_h} + \tilde{p}_{ll} \log \frac{\tilde{p}_{ll}}{p_l}\right),$$

and the entropy due to a misspecification of the signal quality (that is, the entropy associated with a different choice of $\tilde{\pi}_h$) is:

$$\epsilon^2(\tilde{\pi}_h) = \tilde{\pi}_h \log \frac{\tilde{\pi}_h}{\pi_h} + \tilde{\pi}_l \log \frac{\tilde{\pi}_l}{\pi_l}.$$

θ_d^{-1} and θ_s^{-1} measure the degree of investors' ambiguity aversion toward ambiguity about the underlying dynamics and ambiguity about the signal quality, respectively. For $\theta_d^{-1} = 0$ and $\theta_s^{-1} = 0$, the investors' problem reduces to that of the ambiguity-neutral investor. As either θ_d^{-1} or θ_s^{-1} increases, the degree of ambiguity aversion of the representative agent increases and the agent considers a large set of possible models when making investment decisions.

(Figure 3 about here.)

Solving the first order conditions, we obtain:

$$\frac{\tilde{p}_{hh}}{p_h} = \frac{\exp\left(-\theta_d^{-1} V_{hh}\right)}{\mathbb{E}\left[\exp(-\theta_d^{-1} V_2) \middle| V_1 = V_h\right]} \tag{3.2}$$

$$\frac{\tilde{p}_{lh}}{p_h} = \frac{\exp\left(-\theta_d^{-1} V_{lh}\right)}{\mathbb{E}\left[\exp(-\theta_d^{-1} V_2) \middle| V_1 = V_l\right]} \tag{3.3}$$

$$\frac{\tilde{\pi}_h}{\pi_h} = \frac{\exp\left(-\theta_s^{-1} V_h + \frac{\theta_d}{\theta_s} \log \mathbb{E}\left[\exp\left(-\theta_d^{-1} V_2\right)\middle| V_1 = V_h\right]\right)}{\mathbb{E}\left[\exp\left(-\theta_s^{-1} V_1 + \frac{\theta_d}{\theta_s} \log \mathbb{E}\left[\exp\left(-\theta_d^{-1} V_2\right)\middle| V_1\right]\right)\middle| y\right]} \tag{3.4}$$

Notice first that, in general, $\tilde{p}_{hh} \neq \tilde{p}_{lh}$ so that, under the misspecified model, the probability of the firm payoff increasing between date 1 and date 2 is state-contingent. Next, we see that the ambiguity-averse agent tilts the probability distribution toward the lower utility states using an exponential tilt. Thus, the distorted probability of the firm payoff increasing between dates 1 and 2 is lower than the reference model probability and decreases as the agent becomes more averse to ambiguity about the underlying dynamics. Similarly, the distorted posterior probability of the firm payoff being high at date 1 is lower than the reference model probability and decreases as the agent becomes more averse to ambiguity about the signal quality. Notice also that the tilt in the posterior probabilities incorporates the investors' distorted beliefs about the probability of the firm payoff increasing between dates 1 and 2. Finally, (3.2)-(3.4) imply that the minimized objective of the ambiguity-averse investor is given by:

$$J = -\theta_s \log \mathbb{E}\left[\exp\left(-\theta_s^{-1} V_1 + \frac{\theta_d}{\theta_s} \log \mathbb{E}\left[\exp(-\theta_d^{-1} V_2) \middle| V_1\right]\right) \middle| y\right]. \tag{3.5}$$

Thus, in calculating expected utility, the ambiguity-averse investor applies an exponential tilt away from the higher utility states.

There is no independent study of the magnitude of ambiguity aversion in the literature. One way of interpreting the degree of ambiguity aversion is through a thought experiment related to the Ellsberg Paradox (Ellsberg (1961)). Suppose that there are two urns. The participants in the experiment know that there are 50 black and 50 white balls in urn 1. Urn 2 also contains 100 balls but, unlike urn 1, either contains 100 white balls or 100 black balls. The participants are asked to pick a ball from an urn and are given a prize if they pick a black ball. Experimental evidence suggests that participants prefer to bet on urn 1 (see e.g. Camerer (1999) and Halevy (2007)). The Ellsberg Paradox is that, if the participants are asked to pick a white ball, they still prefer to bet on urn 1. This effect cannot be explained by the standard expected utility model with any beliefs or any risk aversion level. This behavior is, however, consistent with ambiguity aversion as, for an ambiguity-averse agent picking a black ball, the worst case beliefs imply that urn 2 contains only white balls while, for the agent picking a white ball, the corresponding worst case belief is that urn 2 contains only black balls. Thus, ambiguity aversion and risk aversion have distinct behavioral meanings.

More formally, denote by w the wealth of a risk-neutral, ambiguity-averse participant and by d the prize money. Since the distribution of balls in urn 1 is known for sure to be $(1/2, 1/2)$, the participant's certainty equivalent from picking a ball from urn 1 is:

$$\frac{1}{2}(w+d) + \frac{1}{2}w. \tag{3.6}$$

The distribution of balls in urn 2 is, on the other hand, unknown and the participant evaluates the gamble using a version of the recursion (3.5):

$$-\theta \log\left[\frac{1}{2}\exp\left(-\frac{w+d}{\theta}\right)+\frac{1}{2}\exp\left(-\frac{w}{\theta}\right)\right]. \tag{3.7}$$

Notice that, since the exponential function is more concave than a linear function, the expression in (3.6) is larger than that in (3.7), so that the agent prefers to bet on urn 1 rather than urn 2. The difference between the certainty equivalents in (3.6) and (3.7) is a measure of the ambiguity premium. Fig. 4 plots the ambiguity premium as a function of the ambiguity aversion parameter, θ, for a prize-wealth ratio of 1%. Increasing the prize-wealth ratio raises the ambiguity premium. Camerer (1999) reports that the ambiguity premium is typically on the order of 10 – 20 percent of the expected value of a bet in the Ellsberg-style experiments. Thus, except for extreme levels of the ambiguity aversion coefficient, the ambiguity premium seems small and reasonable.

(Figure 4 about here.)

Return now to the three period tree in Fig. 2. Fig. 5 plots the distorted probability of the firm payoff increasing between date 1 and date 2 as a function of the reference probability of the same event for two different levels of investors' ambiguity aversion to misspecification of the underlying dynamics, θ_d^{-1}. As the degree of ambiguity aversion increases, the probability becomes more distorted. Furthermore, the distortion is non-linear in the reference model probability, p_h, and is larger for intermediate values of p_h. Intuitively, if the agent knows that the firm payoff increases between dates 1 and 2 for sure, then he has no misspecification doubts and, thus, the implied misspecified probability coincides with the reference model probability. Similarly, Fig. 6 plots the distorted posterior probability of the high state at date 1 as a function of the reference probability of the same event for two different levels of investors' ambiguity aversion to misspecification of the signal quality, θ_s^{-1}. Similarly to the probability of an increase, as the degree of ambiguity aversion increases, the distortion increases and is larger for intermediate levels of the reference model probability, π_h.

(Figure 5 about here.)

(Figure 6 about here.)

The interaction between the two sources of ambiguity is illustrated in Fig. 7, which plots the distorted probability of default as a function of the reference model probability for different combinations of θ_d^{-1} and θ_2^{-1}. Notice first that, even for low levels of the reference model probability of default, the distorted probability of default can be quite high: for the case $\theta_d^{-1} = \theta_s^{-1} = 0.5$, a reference model default probability of 25 bps translates into a perceived 65% probability for the ambiguity-averse agent. Notice further that, although the distortion increases when the degree of ambiguity aversion to either source of ambiguity increases, the relation between the reference and the distorted probability is different depending on whether the agent is more averse to ambiguity about the underlying dynamics or the signal quality. More specifically, for the case when the agent is more averse to ambiguity about the underlying dynamics ($\theta_d^{-1} = 0.5$, $\theta_s^{-1} = 0.1$), the distorted probability of default increases over the whole range of the reference probability. On the other hand, for the case when the agent is more averse to ambiguity about the signal quality ($\theta_d^{-1} = 0.1$, $\theta_s^{-1} = 0.5$), the distorted probability of default levels off. That is, while the distorted probability of default for the agent more averse to ambiguity about underlying dynamics increases faster than the reference model probability, the distorted probability of default for the agent more averse to ambiguity about the signal quality increases slower than the reference probability.

(Figure 7 about here.)

Consider finally the two contributions, ϵ^1 and ϵ^2, plotted in Fig. 8. The left panel plots the entropy due to misspecification of the underlying dynamics, ϵ^1, as a function of the probability of the firm payoff increasing between dates 1 and 2 for two different levels of the ambiguity aversion to misspecification of the underlying dynamics, while the right panel plots the entropy due to misspecification of the signal quality as a function of the reference model posterior probability of the high state at date 1 for two different levels of the ambiguity aversion to misspecification of the signal quality. As the degree of ambiguity aversion increases to the corresponding source of ambiguity, the entropy increases. Furthermore, entropy is non-linear in the corresponding probability, larger for intermediate levels of the probability and is zero at the extremes of the distribution. Thus, for example, the entropy due to misspecification of the underlying dynamics is larger when the probability of the firm payoff increasing between dates 1 and 2 is close to 1/2. Intuitively, if the agent knows that the firm payoff increases between dates 1 and 2 for sure, then he has no misspecification doubts and, thus, the entropy between the misspecified and the reference models is zero.

(Figure 8 about here.)

4 Model

In this section, I present the economy considered in this paper. I begin by describing the reference model used for pricing credit securities and then proceed to the misspecification problem faced by the representative agent in the economy.

4.1 Reference model

As the reference model, I consider a modified version of the Black and Cox (1976) economy. Consider a (sector of the) economy consisting of I firms, indexed by $i = 1, \ldots, I$ and denote by $A_{it} = e^{a_{it}}$ the fundamental value of the assets of firm i at date $t = 1, 2, \ldots$. To fix ideas, assume that there are $n_y = 12$ data periods in a year, so that each period corresponds to a month. I assume that the log-asset value of each firm can be decomposed into the sum of two components:

$$a_{it} = z_{it} + \rho_i z_{ct}, \tag{4.1}$$

where z_{it} is an idiosyncratic shock to the asset value of firm i, z_{ct} is an aggregate shock to the asset values of all the firms in the sector and ρ_i is the loading of firm i on the aggregate component. Denote by $z_t = [z_{1t}, \ldots, z_{It}, z_{ct}]'$ the vector of the components of asset values at date t. I assume that the vector z_t evolves according to an N-state Markov chain, with possible values $\xi_1, \ldots, \xi_N$ and the transition probability matrix Λ defined as:

$$\{\Lambda\}_{jk} \equiv \lambda_{jk} = \mathbb{P}\left(z_{t+1} = \xi_k \middle| \, z_t = \xi_j \right). \tag{4.2}$$

There are two types of agents in the economy: managers and investors. All the day-to-day operations of the firm are delegated to the respective manager. I assume that there are no agency problems between a firm's managers and the equity holders of the firm, so that the managers act in the best interest of the equity holders. Further, similarly to Duffie and Lando (2001), I assume that the managers are better informed about the firm they manage than the participants in the public markets and, in particular, that the managers of the firm observe perfectly the evolution of the fundamental value of the firm's assets. To prevent information spill-over, I assume that managers are precluded from trading in the public assets markets.

In this paper, I abstract from modeling the operational decisions of the firm managers and, in particular, from modeling the optimal capital structure, dividend payment and default decisions faced by the managers. As in Leland (1994) and Duffie and Lando (2001), I assume that each firm i issues perpetual debt with face value D_i. This debt is serviced by a constant coupon rate C_i. While firm i is in operation, it generates a constant fraction δ_i of assets as cash-flows which accrue, minus the coupon payments, as equity in the firm.

The managers decide on behalf of the equity holders when to default. As in Black and Cox (1976), I abstract from modeling the liquidation decision faced by the managers and assume instead that a firm defaults automatically whenever the fundamental value of the firm's assets reaches the lowest possible value implied by the Markov chain $\{z_t\}_{t=1}^{+\infty}$. In particular, denote by ξ_{ji} the i^{th} element of the asset values vector in state j, ξ_j. Let $i^* = argmin_{j=1,\ldots,N} \xi_{ji} + \rho_i \xi_{jc}$ be the state index at which firm i achieves its lowest possible value and by $a_{B_i} = \xi_{i^*}$ the corresponding state. Then the (stochastic) default date τ_i of firm i is the first hitting time of the state a_{B_i}: $\tau_i = \inf \{t : z_t = a_{B_i}\}$. In economic terms, the exogenous default rule can be interpreted as a debt covenant. The firm is liquidated at the present value of the discontinued cash flows, with the proceeds distributed among the firm's primary debt holders and the equity holders receiving 0. For simplicity, I assume that each firm has a single default state and that firms do not default simultaneously. Notice that, since managers observe perfectly the asset value evolution of the firm under their management, there is no uncertainty about the firm being liquidated upon hitting its default boundary. Finally, denote by $a_B = \bigcup_{i=1}^{I} a_{B_i}$ the union of the default states of all the firms and by a_B^c its complement, which is the set of states where none of the firms default.

Consider now the participants in the public markets. Similarly to Duffie and Lando

(2001), I assume that the representative investor does not observe the true evolution of asset values in the sector and receives instead imperfect, unbiased signals about the fundamental value of the assets of each firm, $\hat{A}_{it} = e^{y_{it}}$, and the aggregate component of asset values in the sector, $\hat{A}_{ct} = e^{y_{ct}}$. More specifically, assume that $y_{it} = a_{it} + u_{it}$ and $y_{ct} = z_{ct} + u_{ct}$ where the signal errors $u_t = [u_{1t}, \ldots, u_{It}, u_{ct}]'$ are serially uncorrelated and normally distributed, independent of the true realization of z_t: $u_t \sim N(\overline{u}, \Sigma_u)$. Here, $\overline{u}$ is the mean signal error and Σ_u^{-1} the signal quality. At each date t, the representative agent also observe whether any of the default states have been reached and any of the firms have been liquidated. Thus, the information set of the representative agent at date t is:

$$\mathcal{G}_t = \sigma \left\{ y_s,\ \mathbf{1}_{z_s \in a_B^c} :\ s = 1, \ldots, t \right\},$$

where $y_t = [y_{1t}, \ldots, y_{It}, y_{ct}]'$ is the full signal vector at date t.

Denote by p_{jt} the probability, conditional on the date t information set of the representative investors, of the vector z being in state j at date t:

$$p_{jt} = \mathbb{P}\left(z_t = \xi_j | \, \mathcal{G}_t \right).$$

For mathematical reasons, it is easier to formulate the updating rule in terms of unnormalized probabilities $\vec{\pi}_t$, which are related to the proper probabilities by:

$$p_{jt} = \frac{\pi_{jt}}{\sum_{j=1}^{N} \pi_{jt}}, \qquad j = 1, \ldots, N.$$

Let $\pi_{j0} = \mathbb{P}(z_1 = \xi_j)$ be the prior probability. Then the following result holds.

Lemma 4.1. *(Wonham Filter)*

Assume that the transition probability matrix, Λ, and the prior distribution π_{j0} are *known. Then the date 1 update to the unnormalized probability vector is given by:*

$$\pi_{j1} = \mathbf{1}_{\xi_j \notin a_B} \pi_{j0} f(y_1 - \xi_j), \qquad j = 1, \ldots, N, \tag{4.3}$$

where $f(\cdot)$ *is the density function of the observation errors,* u*. For* $t > 1$*, the "predict" step of the update to the unnormalized probability vector is given by:*

$$\tilde{\pi}_{t+1} = diag(\mathbf{f}(y_{t+1}))\Lambda' \vec{\pi}_t, \tag{4.4}$$

where $\mathbf{f}(y) = [f(y - \xi_1), \ldots, f(y - \xi_N)]'$ *and* $diag(\cdot)$ *creates a diagonal matrix from the vector* $\cdot$*. Then, conditional on no firm defaulting in period* $t+1$*, the updated unnormalized probability vector is given by:*

$$\vec{\pi}_{t+1} = diag(\mathbf{1}_{a_B^c})\tilde{\pi}_{t+1},$$

where $\mathbf{1}_{a_B^c} = \left[\mathbf{1}_{\xi_1 \in a_B^c}, \ldots, \mathbf{1}_{\xi_N \in a_B^c}\right]'$.

Proof. See e.g. Frey and Schmidt (2009). □

Finally, consider the utility of the representative agent. I assume the representative agent is risk-neutral and, thus, holds all the claims to the firm's asset value. Thus, the date t expected present value of the utility of the representative agent is given by:

$$J_t = \mathbb{E}\left[\sum_{s=0}^{+\infty} \beta^s \sum_{i=1}^{I} \delta_i A_{i,t+s} \middle| \mathcal{G}_t\right], \tag{4.5}$$

where β is the subjective discount factor. For the discussion below, it is useful to represent the expected present value of utility in recursive form:

$$J_t = \mathbb{E}\left[\sum_{i=1}^{I} \delta_i A_{it} + \beta \mathbf{1}_{z_{t+1} \in a_B^c} J_{t+1} + \beta \sum_{i=1}^{I} \mathbf{1}_{z_{t+1} = a_{B_i}} J^{\tau_i} \middle| \mathcal{G}_t \right], \tag{4.6}$$

where J^{τ_i} is the value function of the representative agent in case firm i defaults.

4.2 Asset prices

In this paper, I consider three types of claims to the assets of the firm: a claim to the firm's equity, a zero-coupon, risky bond and a credit default swap (CDS) written on the bond. Recall that equity in firm i accrues as a constant fraction δ_i of the fundamental asset value and that the equity holders receive 0 in case of the firm being liquidated. Thus, the date t price of a claim to equity of firm i, V_{it}, is given by:

$$V_{it} = \mathbb{E}\left[\sum_{s=0}^{\tau_i} \beta^s (\delta_i A_{i,t+s} - C_i) \middle| \mathcal{G}_t \right],$$

where τ_i is the (stochastic) default date of the firm i. Notice that the equity price satisfies the Euler equation:

$$V_{it} = \mathbb{E}\left[\delta_i A_{it} - C_i + \beta \mathbf{1}_{\tau_i > t+1} V_{i,t+1} \middle| \mathcal{G}_t \right]. \tag{4.7}$$

Consider now the default swaps written on the primary debt of firm i. With a given maturity T, a default swap is an exchange of an annuity stream at a constant coupon rate until maturity or default, whichever is first, in return for a payment of X at default, if default is before T, where X is the difference between the face value and the recovery value on the stipulated underlying bond. A default swap can thus be thought of as a default insurance contract for bond holders that expires at a given date T, and makes up the difference between face and recovery values in the event of default.

I assume, as typical in practice, that the default swap annuity payments are made semiannually, and that the default swaps maturity date T is a coupon date. As in Duffie and Lando (2001), I take the underlying bond for the default swap on firm i to be the consol bond issued by firm i. Recall that, in case of default, the debt holders receive the present value of the discontinued cash flows. Thus, the payment X_i per unit of primary debt if firm i defaults before the swap maturity date T is given by:

$$X_i = 1 - \frac{\delta_i (I - \beta\Lambda)^{-1}_{i*} e^{\xi}}{(1-\beta) D_i}.$$

The at-market default swap spread is the annualized coupon rate $c_i(t,T)$ that makes the default swap sell at date t for a market value of 0. Thus, with $T = t + 6n$ for a given positive integer n,[2] the CDS spread is given by:

$$c_i(t,T) = \frac{2X_i \mathbb{E}\left[\beta^{\tau_i - t}\mathbf{1}_{\tau_i < T} \middle| \mathcal{G}_t\right]}{\sum_{s=1}^{n} \beta^{6s} \mathbb{E}\left[\mathbf{1}_{\tau_i \geq t+6s} \middle| \mathcal{G}_t\right]}. \tag{4.8}$$

default swap spreads are a standard for price quotation and credit information in bond markets. In this setting, they have the additional virtue of providing implicitly the term structure of credit spreads for par floating-rate bonds of the same credit quality as the underlying consol bond, in terms of default time and recovery at default. Denote by $B_i(t,T)$ the date t price of a zero-coupon bond with maturity date T on the debt of firm i. The implicit discount curve is then given by:

$$\begin{aligned} B_i(t, t+6) &= \frac{1}{1 + c_i(t, t+6)} \\ B_i(t, t+6s) &= \frac{1 - \frac{c_i(t,t+6s)}{2} \sum_{j=1}^{s-1} B_i(t, t+6j)}{1 + c_i(t, t+6s)/2}. \end{aligned}$$

[2]Recall that there are 12 data periods in a year

4.3 Model misspecification

This paper studies asset prices in a setting where the representative agent makes decision rules robust to possible misspecifications of asset value and accounting signals models. In reality, the correct specification assumption of the reference model is overly restrictive. It implies that, even though the participants in the public markets only observe imperfect signals about the evolution of the fundamental asset value, they can still correctly identify the parametric model that governs the relevant dynamics. More realistically, I assume that the representative investor in the firm fears misspecification of the probability law generated by the model above and believes instead that the signals are related to the true asset value realizations by a family of likelihoods.

As in Hansen and Sargent (1995), Hansen et al. (1999), Tallarini Jr. (2000) and Anderson et al. (2003), I model preferences of the representative agent in the presence of model misspecification using the recursion:

$$J_t = -\theta_s \log \mathbb{E}\left[\exp\left[-\frac{U(z_t)+\mathcal{R}_t(\beta J_{t+1};\theta_d)}{\theta_s}\right]\middle|\,\mathcal{G}_t\right], \tag{4.9}$$

where:

$$\mathcal{R}_t(\beta J_{t+1};\theta_d) \equiv -\theta_d \log \mathbb{E}\left[\exp\left(-\frac{\beta J_{t+1}}{\theta_d}\right)\middle|\,\mathcal{F}_t\right].$$

The tilted recursion (4.9) replaces the standard utility recursion (4.6), incorporating the representative agent's misspecification doubts in two steps. First, the tilted continuation function $\mathcal{R}_t$ makes an additional risk adjustment to the continuation value function of the representative agent, accounting for misspecification fears about the fundamental asset value evolution dynamics. Second, the tilted expectations over the current period utility adjusts for misspecification fears about the filtered probability distribution over the current state. As emphasized by Hansen and Sargent (1995), the log-exp specification of the recursion links risk-sensitive control theory and a more general recursive utility specification of Epstein and Zin (1989). The degree of the representative agent's aversion to misspecification of the underlying dynamics is quantified by θ_d^{-1} (where d stands for dynamics) and the degree of the representative agent's aversion to misspecification of the filter distribution is quantified by θ_s^{-1} (where s stands for signals). When $\theta_s^{-1} = \theta_d^{-1} = 0$, the risk-sensitive recursion (4.9) reverts to the usual utility recursion under Von Neumann-Morgenstern form of state additivity. For values of θ_d^{-1} greater than zero, the recursion (4.9) implies an increased aversion to risk associated with the time evolution of the hidden state *vis a vis* the Von Neumann-Morgenstern specification. Similarly, for values of θ_s^{-1} greater than zero, the recursion implies an increased aversion to risk associated with the unobservability of the true state *vis a vis* the Von Neumann-Morgenstern specification. Maenhout (2004) links the degree of misspecification to the value function itself, so that the agent becomes more misspecification-averse as the present value of her utility increases. Finally, notice that, since the distorted continuation value function $\mathcal{R}_t$ conditions on the full information set $\mathcal{F}_t$ (and, hence, on the true realization of the hidden state), the investors in this economy have the option of focusing their attention on misspecification of the joint probability distribution of future signals and state.

To understand better the recursion (4.9), consider the following static optimization problem:

$$\min_{m \geq 0;\ \mathbb{E}[m]=1} \mathbb{E}\left[mV\right] + \theta \mathbb{E}\left[m \log m\right]. \tag{4.10}$$

The random variable m is the likelihood ratio between the reference model and an alternative model. m implies a distorted expectation operator: $\tilde{\mathbb{E}}\left[V\right] = \mathbb{E}\left[mV\right]$. The optimization problem (4.10) then minimizes the expected value of the payoff V under alternative models but is penalized in utility terms for deviations from the reference model (parametrized by $m = 1$). The term $\mathbb{E}[m \log m]$ measures the discrepancy in relative entropy terms between the reference model and an alternative model. As noted in Jacobson (1973), the relative entropy $\mathbb{E}[m \log m]$ is the expected log-likelihood between the reference and the misspecified models. Thus, the parameter θ can be interpreted as a penalization parameter for large deviations away from the reference model. The problem (4.10) can thus be interpreted as a robust way of alternating probability measures. The minimizing choice of m, the so-called worst-case model, is given by:

$$m^* = \frac{\exp\left(-\frac{1}{\theta}V\right)}{\mathbb{E}\left[\exp\left(-\frac{1}{\theta}V\right)\right]}$$

and the outcome of the minimization problem by:

$$-\theta \log \mathbb{E}\left[\exp\left(-\frac{1}{\theta}V\right)\right].$$

Thus, in choosing between alternative models, the representative agent tilts the probability toward bad (in terms of payoffs) states.

Turn now back to the recursion (4.9). Hansen and Sargent (2007) show that, corresponding to the tilted continuation function $\mathcal{R}_t$ is the worst-case likelihood ratio:

$$\phi_t(z_{t+1}, y_{t+1}) = \frac{\exp\left(-\frac{\beta J_{t+1}}{\theta_d}\right)}{\mathbb{E}\left[\exp\left(-\frac{\beta J_{t+1}}{\theta_d}\right) \middle| \mathcal{F}_t\right]}. \tag{4.11}$$

ϕ_t captures the difference between the evolution of future signals and states under the misspecified model and under the reference model. In particular, ϕ_t is the date t probability distortion to the joint distribution of next period's signals and state. Relative to the reference model distribution, ϕ_t tilts the joint distribution toward lower continuation value states, decreasing the expected future value of the continuation utility. Similarly, corresponding to the recursion (4.9), is the worst-case likelihood ratio:

$$\psi_t(z) = \frac{\exp\left(-\frac{U(z)+\mathcal{R}_t(\beta J_{t+1};\theta_d)}{\theta_s}\right)}{\mathbb{E}\left[\exp\left(-\frac{U(z)+\mathcal{R}_t(\beta J_{t+1};\theta_d)}{\theta_s}\right)\middle|\,\mathcal{G}_t\right]} \tag{4.12}$$

between the conditional distribution over the state at date t under the misspecfied and reference models. ψ_t tilts the conditional distribution toward lower utility states, decreasing the expected value of utility at the current date.

An alternative interpretation of the recursion (4.9) is in terms of the smooth robustness preferences of Klibanoff et al. (2005, 2009) and recursive preferences of Epstein and Zin (1989). In the smooth ambiguity preference setting, the representative agent does not choose the "worst-case model" and instead assigns a preference ordering to the alternative models. In particular, let u be the agent's utility over realizations of consumption, μ index different models, f the agent's utility function over different models and π the belief vector over the different models. Then, an agent with the smooth robustness preferences evaluates consumption according to:

$$f^{-1}\left(\mathbb{E}_\pi\left[f\left(\mathbb{E}\left[u|\,\mu\right]\right)\right]\right).$$

Compare this to the recursion (4.9). Notice first that the tilted continuation utility, $\mathcal{R}(\beta J_{t+1};\theta_d)$, corresponds to the continuation utility for an agent with Epstein and Zin (1989) preferences, so corresponding to the inner expectation conditional on a "model" in smooth utility preferences, the worst-case utility model evaluates future consumption using Epstein and Zin (1989) preferences conditional on the current state. The recursion (4.9) then uses an exponential utility function to rank continuation values and current period utilities for different realizations of the state.

To quantify the amount of distortion in the economy, Hansen and Sargent (2007) introduce measures of the conditional relative entropy between the reference and the distorted models. In particular, under the full information setting, the conditional relative entropy between the reference and the worst case model of the future state and signals evolution is defined as:

$$\epsilon_t^1(\phi_{t+1}, \xi_j) = \sum_{k=1}^{N} \int \tau(\xi_k, y_{t+1}|\xi_j)\phi(\xi_k, y_{t+1}) \log \phi(\xi_k, y_{t+1}) dy_{t+1}$$

and the conditional relative entropy between the reference and the worst case model of the current state by:

$$\epsilon_t^2(\psi_t) = \sum_{j=1}^{N} p_j \psi_j \log \psi_j.$$

The total conditional relative entropy between the reference model and the worst case model at date t is then given by:

$$\epsilon_t = \epsilon_t^2(\psi_t) + \tilde{\mathbb{E}}\left[\epsilon_t^1(\phi_{t+1}, z_t)\middle|\, \mathcal{G}_t\right] \equiv \epsilon_t^2(\psi_t) + \hat{\epsilon}_t^1(\phi_{t+1}). \tag{4.13}$$

In general, the recursion (4.9) does not have a closed-form solution. Instead, I look for a first order approximation to the representative agent's value function around the point $\theta_d^{-1} = \theta_s^{-1} = 0$, which corresponds to the solution under the reference model. Notice that the approximation I construct here is different in its nature from the small noise approximations constructed in Campi and James (1996) and Anderson et al. (2010) as, instead of approximating the reference model value function around the deterministic steady state, I approximate around the value function corresponding to the zero signal precision case.[3] The following result holds:

[3]When the signal precision approaches 0, the agent does not update the conditional probability distribution and, hence, the stationary distribution can be used as the prior distribution.

Lemma 4.2. *The first order approximation to the value function around the point* $\theta_d^{-1} = \theta_s^{-1} = 0$ *is given by;*

$$J(\pi; \theta_d^{-1}, \theta_s^{-1}) \approx J_0(\pi) + \theta_d^{-1} J_{\theta_d^{-1}}(\pi) + \epsilon J_\epsilon(\pi), \tag{4.14}$$

where $\epsilon = \theta_s^{-1}/\theta_d^{-1} - 1$*. Here,* J_0 *is the value function of the ambiguity-neutral agent,* $J_{\theta_d^{-1}}$ *measures the change in the value function as the overall level of ambiguity increases and* J_ϵ *measures the change in the value function as the agent becomes more averse to ambiguity about the signal quality than ambiguity about the underlying dynamics. In the non-default states of the economy, the first order approximations to* J_0*,* $J_{\theta_d^{-1}}$*, and* J_ϵ *in terms of log-deviations,* $\hat{\pi}$*, from the stationary distribution of the underlying Markov chain are given, respectively, by:*

$$J_0(\pi) \approx \gamma_{00} + \gamma_{01}' diag(\bar{\pi})\hat{\pi} \tag{4.15}$$

$$J_{\theta_d^{-1}}(\pi) \approx \gamma_{10} + \gamma_{11}' diag(\bar{\pi})\hat{\pi} \tag{4.16}$$

$$J_\epsilon(\pi) \approx \gamma_{\epsilon 0} + \gamma_{\epsilon 1}' diag(\bar{\pi})\hat{\pi}, \tag{4.17}$$

where the coefficients γ_{00}*,* γ_{01}*,* γ_{10}*,* γ_{11}*,* $\gamma_{\epsilon 0}$ *and* $\gamma_{\epsilon 1}$ *solve the system (C.2)-(C.10). The first order approximation to the implied distortion to the conditional joint distribution of next period's signals and state is then given by:*

$$\begin{aligned} \phi_t(z_{t+1}, y_{t+1}) \approx\ & 1 - \theta_d^{-1}\beta\left(J_0(\pi_{t+1}) - \mathbb{E}\left[\left. J_0(\pi_{t+1})\right| \mathcal{F}_t\right]\right) \\ & - \left(\theta_s^{-1} - \theta_d^{-1}\right)\beta\left(J_\epsilon(\pi_{t+1}) - \mathbb{E}\left[\left. J_\epsilon(\pi_{t+1})\right| \mathcal{F}_t\right]\right), \end{aligned} \tag{4.18}$$

and the first order approximation to the implied distortion to the conditional distribution *of the current state by:*

$$\psi_t(z_t) \approx 1 - \left(\theta_d^{-1} - \theta_s^{-1}\right) \beta \left(\mathbb{E}\left[J_\epsilon(\pi_{t+1}) \middle| \mathcal{F}_t\right] - \mathbb{E}\left[J_\epsilon(\pi_{t+1}) \middle| \mathcal{G}_t\right]\right) \tag{4.19}$$
$$- \theta_s^{-1} \left(U(z_t) + \beta \mathbb{E}\left[J_0(\pi_{t+1}) \middle| \mathcal{F}_t\right] - \mathbb{E}\left[U(z_t) + \beta J_0(\pi_{t+1}) \middle| \mathcal{G}_t\right]\right).$$

Proof. See Appendix C.1. □

Notice that J_0 is the value function of the representative agent under the reference model. The vector γ_{01} captures the first-order dependence of the reference model value function on the conditional distribution of the hidden state. From (C.3) we can see that the right hand side of the equation determining $\gamma_{01,j}$ is positive when the agent's utility in state j is higher than the stationary probability distribution weighted average of utility in different states. Thus, $\gamma_{01,j}$ is positive for states that have higher utility and negative for lower utility states. Intuitively, the expected present value of the representative agent's utility should be higher when the probability of the economy being in a good state is higher and lower when the probability of being in a bad state is higher. $J_{\theta_d^{-1}}$ measures the change in the value function as the overall level of ambiguity increases. Since the ambiguity-averse agent is solving a minimization problem in determining optimal consumption and portfolio choices, $J_{\theta_d^{-1}}$ is negative. Thus, as the agent becomes more ambiguity averse overall, his value function decreases. J_ϵ, on the other hand, measures the change to the value function of the ambiguity-averse agent as the agent becomes more averse to ambiguity about signal quality than ambiguity about the underlying dynamics. Thus, keeping the level of aversion to ambiguity about the underlying dynamics, J_ϵ is also negative as the agent imposes a smaller entropy cost on choosing more distorted (toward lower utility states) posterior distributions. Notice that, since both $J_{\theta_d^{-1}}$ and J_ϵ are time-varying, the agent's perceived risk attitudes change depending on the conditional distribution of the hidden state, with the vectors γ_{11} and $\gamma_{\epsilon 1}$ describing the loadings on the individual components of the probability vector.

Consider now the first order approximation to the worst-case distortion to the conditional joint distribution of next period's signals and states. Notice first that, when the representative agent has equal degree of aversion to both sources of ambiguity, so that $\theta_s = \theta_d$, the first order approximation depends only on the value function of the ambiguity-neutral agent, J_0. Recall that J_ϵ is the derivative of the value function of the ambiguity-averse agent in the direction of greater aversion to ambiguity about the signal quality. Since the value function of the ambiguity-averse agent is lower than that of the ambiguity-neutral agent, the distortion to the conditional joint distribution of next period's signals and state is smaller when the agent in more averse to ambiguity about signal quality than ambiguity about the underlying dynamics. That is, if the agent is more averse to ambiguity about signal quality than ambiguity about the underlying dynamics (so that $\theta_d^{-1} < \theta_s^{-1}$) , the aversion to ambiguity about signal quality mitigates the impact of ambiguity about the underlying dynamics on the perceived transition probabilities. On the other hand, if the agent is more averse to ambiguity about the underlying dynamics (so that $\theta_d^{-1} > \theta_s^{-1}$) , the aversion to ambiguity about signal quality exacerbates the impact of ambiguity about the underlying dynamics on the perceived transition probabilities. Intuitively, if the agent is more averse to ambiguity about signal quality, then he will not apply as a large of a distortion to the transition probabilities since misspecification of the underlying dynamics is a smaller concern in his mind. If, however, the agent is more averse to ambiguity about the underlying dynamics, then misspecification of the transition probabilities is a primary concern which only worsens due to misspecification concerns about the posterior distribution over the possible realizations of the current state.

Similarly, when $\theta_s = \theta_d$, the first order approximation to the worst-case distortion to the posterior distribution of the current state also depends only on the value function of the ambiguity-neutral agent. If the agent is more averse to ambiguity about the signal quality than about the underlying dynamics (so that $\theta_d^{-1} < \theta_s^{-1}$), the aversion to ambiguity about the underlying dynamics exacerbates the impact of the ambiguity about the signal quality. If, on the other hand, the agent is more averse to ambiguity about the underlying dynamics (so that $\theta_d^{-1} > \theta_s^{-1}$), the aversion to ambiguity about the underlying dynamics mitigates the impact of the ambiguity about the signal quality. Intuitively, if the representative agent is more averse to ambiguity about the underlying dynamics, he will not apply as large a distortion to the current period posterior probabilities since misspecification of the filter dynamics is a smaller concern in his mind. If, however, the agent is more averse to ambiguity about the signal quality, then misspecification of the current period posterior probabilities is a primary concern which only worsens due to misspecification concerns about the transition probabilities.

Substituting for J_0 and J_ϵ in (4.18)-(4.19), we can express:

$$\begin{aligned}
\phi_t(z^* = \xi_k, y^*|z = \xi_j) &= 1 + \theta_d^{-1}\left(\varphi_{01,jk} + \varphi'_{\pi 1,jk} diag(\overline{\pi})\hat{\pi} + \varphi'_{y1,jk}\log f(y^* - \xi_k)\right) \\
&\quad + \left(\theta_s^{-1} - \theta_d^{-1}\right)\left(\varphi_{0\epsilon,jk} + \varphi'_{\pi\epsilon,jk} diag(\overline{\pi})\hat{\pi} + \varphi'_{y\epsilon,jk}\log f(y^* - \xi_k)\right) \\
\psi_t(\xi_j) &= 1 + \theta_s^{-1}\left(\zeta_{01,j} + \zeta'_{11,j} diag(\overline{\pi})\hat{\pi}\right) + \left(\theta_d^{-1} - \theta_s^{-1}\right)\left(\zeta_{0\epsilon,j} + \zeta'_{1\epsilon,j} diag(\overline{\pi})\hat{\pi}\right)
\end{aligned}$$

where the expansion coefficients are given by (C.11)-(C.20). Integrating over the future state, the implied transition probability matrix is:

$$\begin{aligned}
\tilde{\lambda}_{jk} &\equiv \tilde{\mathbb{P}}\left(z_{t+1} = \xi_k \middle| z_t = \xi_j\right) \\
&\approx \lambda_{jk}\left[1 + \theta_d^{-1}\left(\varphi_{01,jk} + \varphi'_{\pi 1,jk} diag(\overline{\pi})\hat{\pi} + \varphi'_{y1,jk}\Delta_k^1\right)\right] \\
&\quad + \lambda_{jk}\left(\theta_s^{-1} - \theta_d^{-1}\right)\left(\varphi_{0\epsilon,jk} + \varphi'_{\pi\epsilon,jk} diag(\overline{\pi})\hat{\pi} + \varphi'_{y\epsilon,jk}\Delta_k^1\right).
\end{aligned} \tag{4.20}$$

Notice that this implies that, unlike the reference model, the transition probability matrix under the misspecified model is time-dependent. This effect introduces additional time variation into asset prices above that implied by the time evolution of fundamental asset values under the reference model.

4.4 Asset prices under the misspecified model

Since the representative agent evaluates expectations under the worst-case measure when making consumption decisions, the Euler equation holds under the worst-case likelihood and assets can be priced using the worst-case Euler equation. In particular, under the misspecified model, the date t price of a claim to the equity of firm i satisfies:

$$V_{it} = \tilde{\mathbb{E}}\left[\delta_i A_{it} - C_i + \beta \mathbf{1}_{\tau_i > t+1} V_{i,t+1} \middle| \mathcal{G}_t\right]. \tag{4.21}$$

As with the value function, consider a first order expansion of the equity price around the reference model equity price. That is, I look for a first order approximation to the solution of the worst-case Euler equation (4.21) in the form:

$$V_i(\pi_t; \theta_d^{-1}, \theta_s^{-1}) \approx V_{i0}(\pi_t) + \theta_d^{-1} V_{i\theta_d^{-1}}(\pi_t) + \epsilon V_{i\epsilon}(\pi_t). \tag{4.22}$$

The following result holds:

Lemma 4.3. *The first order approximations to V_{i0}, $V_{i\theta_d^{-1}}$, and $V_{i\epsilon}$ in terms of log-deviations, $\hat{\pi}$, from the stationary distribution of the underlying Markov chain, are given,* *respectively, by:*

$$V_{i0}(\pi_t) = \nu_{i,00} + \nu_{i,01}' diag(\overline{\pi})\hat{\pi} + O_2(\hat{\pi}) \tag{4.23}$$

$$V_{i\theta_d^{-1}}(\pi_t) = \nu_{i,10} + \nu_{i,11}' diag(\overline{\pi})\hat{\pi} + O_2(\hat{\pi}) \tag{4.24}$$

$$V_{i\epsilon}(\pi_t) = \nu_{i,\epsilon 0} + \nu_{i,\epsilon 1}' diag(\overline{\pi})\hat{\pi} + O_2(\hat{\pi}), \tag{4.25}$$

where the coefficients $\nu_{i,00}$, $\nu_{i,01}$, $\nu_{i,10}$, $\nu_{i,11}$, $\nu_{i,\epsilon 0}$ and $\nu_{i,\epsilon 1}$ solve the system of linear equations in Appendix C.2.

Proof. See Appendix C.2. □

Consider now the CDS spreads under the misspecified model. Recall that the spread on a default swap with maturity $T = t + 6n$ on the primary debt of firm i is given by:

$$c_i(t,T) = \frac{2X\tilde{\mathbb{E}}\left[\beta^{\tau_i - t}\mathbf{1}_{\tau_i<T}\middle|\,\mathcal{G}_t\right]}{\sum_{k=1}^{n}\beta^{6s}\tilde{\mathbb{E}}\left[\mathbf{1}_{\tau_i\geq t+6s}\middle|\,\mathcal{G}_t\right]}. \tag{4.26}$$

The misspecification concerns of the representative agent influence the CDS calculations in two ways. First, the conditional probability of the current state is distorted using $\psi(\xi_j)$, with the lower utility states receiving a higher probability. Second, the transition probability matrix Λ is replaced with the time-dependent distorted probability matrix $\tilde{\Lambda}$. Introduce the following notation:

$$\begin{aligned}
\Upsilon^i(\pi,t,T) &= \tilde{\mathbb{E}}\left[\beta^{\tau_i - t}\mathbf{1}_{\tau_i<T}\middle|\,\mathcal{G}_t\right] \\
\Psi^i(\pi,t,T) &= \tilde{\mathbb{E}}\left[\mathbf{1}_{\tau_i\geq T}\middle|\,\mathcal{G}_t\right],
\end{aligned}$$

so that:

$$c_i(t,T) = \frac{2X_i\Upsilon^i(\pi,t,T)}{\sum_{k=1}^{n}\beta^{6s}\Psi^i(\pi,t,t+6s)}.$$

By definition:

$$\begin{aligned}
\Upsilon^{i}(\pi,t,T) &= \sum_{j=1}^{N} p_{jt}\psi(\xi_j)\sum_{s=1}^{T-t}\beta^{s}\tilde{\mathbb{E}}\left[\mathbf{1}_{\tau_i=t+s}\middle|\, z_t=\xi_j\right] \\
&= \sum_{j=1}^{N} p_{jt}\psi(\xi_j)\sum_{s=1}^{T-t}\beta^{s}\tilde{\mathbb{P}}\left(\tau_i=t+s\middle|\, z_t=\xi_j\right) \\
\Psi^{i}(\pi,t,T) &= \sum_{j=1}^{N} p_{jt}\psi(\xi_j)\left(1-\sum_{s=1}^{T-t}\tilde{\mathbb{E}}\left[\mathbf{1}_{\tau_i=t+s}\middle|\, z_t=\xi_j\right]\right) \\
&= \sum_{j=1}^{N} p_{jt}\psi(\xi_j)\left(1-\sum_{s=1}^{T-t}\tilde{\mathbb{P}}\left(\tau_i=t+s\middle|\, z_t=\xi_j\right)\right).
\end{aligned}$$

Let $\tilde{q}^{n}_{ij} = \tilde{\mathbb{P}}\left(z_{t+n}=\xi_j,\ z_{t+n-1}\neq\xi_j,\ldots,z_{t+1}\neq\xi_j\middle|\, z_t=\xi_i\right)$ be the probability that $t+n$ is the first hitting time of state j conditional on being in state i at date t, so that $\tilde{\mathbb{P}}\left(\tau_i=t+n\middle|\, z_t=\xi_j\right) = \tilde{q}^{n}_{ji^*}$. Introduce also the matrix of the first hitting time probabilities: $\tilde{Q}^{n} = \left\{\tilde{q}^{n}_{ij}\right\}_{i,j=1}^{N}$ and let $\tilde{Q}^{n}_{0} = diag(\tilde{Q}^{n})$. Then, from Seneta (1981), $\tilde{Q}^{n}$ is given recursively by:

$$\tilde{Q}^{n} = \tilde{\Lambda}\tilde{Q}^{n}_{0}.$$

Notice that, since $\tilde{\Lambda}$ is time-dependent, the first hitting time distribution under the misspecified model is also time-dependent.

Consider now the contagion effect of the default of firm i at date τ_i on the expected time to default of the surviving firms. I will use the following property of Markov chains:

Lemma 4.4. *Let $\{X_t\}_{t\geq 0}$ be a Markov chain on the probability space $(\Omega,\mathcal{F},\mathbb{P})$ with transition matrix P and state space I. Define the hitting time of a subset A of I to be the random variable $\tau_A:\ \Omega\rightarrow\{0,1,2,\ldots\}$ such that $\tau_A(\omega)=\inf\{t\geq 0: X_t(\omega)\in A\}$. Denote by $k_A(i)$ the mean time taken for $(X_t)_{t\geq 0}$ to reach A after starting from state i:*

$$k_A(i) = \mathbb{E}\left[\tau_A\middle|\, X_0=i\right].$$

Let Q denote the matrix obtained by deleting the rows and columns corresponding to the set A from P. Then:

$$k_A(i) = \sum_{j \notin A} (I - Q)^{-1}_{ij} \tag{4.27}$$

Proof. See e.g. Seneta (1981). □

Similarly to Frey and Schmidt (2010), I define the contagion effect as the change in the expected time to default of firm j at the default time τ_i of firm i, which is given by:

$$\hat{k}_{a_{B_j}}(\tau_i) - \hat{k}_{a_{B_j}}(\tau_i - 1) \equiv \tilde{\mathbb{E}}\left[k_{a_{B_j}} \middle| \mathcal{G}_{\tau_i} \right] - \tilde{\mathbb{E}}\left[k_{a_{B_j}} \middle| \mathcal{G}_{\tau_i - 1} \right]. \tag{4.28}$$

Notice that there are two opposing effects of observing one of the firms default. On the one hand, it reveals to the representative agent the current state of the economy, thus reducing the misspecification concerns faced by the agent. On the other hand, the conditional probability of the other firms defaulting next period increases as the agent knows that the aggregate component of the fundamental value is in its lowest state.

5 The 2007–2008 Financial Crisis

In this section, I estimate the model in Section 4 using data on financial institutions. Although the returns on the equity of financial institutions accounts for a small portion of the overall level of consumption in the economy, these institutions were at the forefront of the 2007 financial crisis and, thus, to understand the asset price movements during the crisis, it is important to understand the movements in the prices of claims on these institutions. I begin by estimating the parameters of the reference model using the observations of book equity of financial institutions as firm-specific signals and the Case-Shiller 10 Cities Housing (CS10) Index as the signal about the common component of the asset values. Although book equity of financial institutions make for imperfect signals, both because by it's nature the series is backward-looking and because of the infrequency of observation, using observations of market equity is also fraught with difficulty. The problem of inverting market prices to obtain time series observations of risk-neutral probabilities is a complex problem in general. For the model considered in this paper, the matter is complicated by the fact that all market prices reflect the worst-case distortions imposed by the preferences of the representative agent. Thus, if market prices were to be used as firm-specific signals, then estimation could no longer be decomposed into two parts – estimating independently the parameters of the reference model and then using these parameters as inputs to estimate the ambiguity preference parameters, further complicating the estimation procedure. While I believe that observations of market equity for the institutions concerned provide interesting insights into the nature of risk and uncertainty in the market, in the current paper I prefer to use these time series to provide an outside test for the validity of the estimated parameters. I choose observations of the CS10 Index as the aggregate signal to capture the exposure of the financial institutions to risks associated with the national housing market. Notice from Table I that the financial institutions considered have higher correlations with the CS10 Index than with stock market indices, such as the S&P 500 index.

(Table I about here.)

Next, I use historical observations of CDS spreads for the financial institutions prior to the start of the crisis to obtain an estimate of θ_s and θ_d – the parameters governing the investors' aversion to misspecification of the filter distribution and underlying dynamics, respectively. Using these estimates, I compute the implied relative entropy between the reference and the worst-case models and decompose the entropy calculation into the contributions from misspecification of the signal model and misspecification of the fundamental value of assets model for the whole time series. Next, I compare the model-implied CDS rates and equity prices to the observed time series. Finally, I examine the differential effect of the two sources of misspecification in the model on asset prices.

5.1 Estimating the reference model

To estimate the reference model, I use historical observations of the firm-specific signals and the aggregate signal. Below, I provide the outline of the estimation procedure. The details of the estimation are provided in Appendix A.

As observations of firm-specific signals, I use balance sheet data from COMPUSTAT. In particular, I use observations of book equity as the accounting signals. As observations of the aggregate signals, I take the time series of the Case-Shiller 10 Cities index. Notice that, while balance sheet data are observed at a quarterly frequency only, observations of the CS10 Index are available at a monthly frequency. The procedure described in Appendix A accounts explicitly for this dual frequency of observations. Notice that, to estimate the reference model, I only use observations up to Q2 2007 to avoid introducing measurement error by including observations of the accounting signal which reflect markdowns taken since the start of the crisis, as well as the increased ambiguity discount in credit derivatives held on the balance sheets of these institutions.

Recall that the reference model is described by the parameters:

- Λ, $\{\xi_j\}_{j=1}^N$: transition probability matrix and states of the fundamental asset values
- $\{\rho_i\}_{i=1}^I$: firm-specific loadings on the common component
- Σ_u: signal error covariance matrix
- $\{D_i, C_i\}_{i=1}^I$: level of perpetual debt and coupon payments

- $\{\delta_i\}_{i=1}^{I}$: fraction of assets generated as cash-flows.

I begin the estimation by identifying the face value of the perpetual bond issued by firm i, D_i, with the last pre-crisis (Q2 2007) observation of the firm's value of long-term debt; notice that, since the model-implied debt has infinite maturity, long-term debt is a better measure than total debt as it excludes short-term liabilities. The coupon payment, C_i, is then chosen to make the level of debt D_i optimal. Following the model assumption that each firm generates a constant fraction δ_i of assets as cash-flows, I identify δ_i as the time-series average of the total earnings as a fraction of total assets. The rest of the parameters are estimated using the Gibbs sampling procedure of Appendix A.

The estimated reference model parameters are presented in Tables II – III and the filtered time series of the expected fundamental value of each firm's assets under the reference model is plotted in Fig. 9. Notice first that, while changes in the filtered fundamental asset values mimic the observed changes in the corresponding book values, the level of the fundamental asset values are lower than that of the book values. Notice also that the states of the economy are highly persistent. The probability of the firm-specific component of asset values staying in the same state next period is around 54% and the probability of the aggregate component of asset values staying in the same state next period is around 99%. Notice also that, although book values have a high correlation with the Case Shiller 10 Index, the estimated firm-specific loadings on the aggregate component of asset values is lower, ranging from 27% for JP Morgan and less than 1% for Bear Stearns. Finally, notice that the signal errors have low cross-correlations of at most 7% and a higher variance, ranging from 73% for the Case Shiller 10 Index to 15% for Goldman Sachs.

(Table II about here.)

(Table III about here.)

(Figure 9 about here.)

5.2 Estimating the misspecification preference parameter

Consider now estimating the degree of aversion toward misspecification of the underlying dynamics, θ_d^{-1}, and misspecification of the filter distribution, θ_s^{-1}. Rewrite the CDS equation (4.26) as:

$$0 = 2X_i \Upsilon^i (\pi, t, T) - c_i(t, T) \sum_{k=1}^{n} \beta^{6s} \Psi^i (\pi, t, t + 6s) .$$

Recall that the expectation functions Υ^i and Ψ^i are calculated using the worst-case likelihood and, thus, depend on investors' preferences toward ambiguity. Thus, to estimate θ_d^{-1} and θ_s^{-1}, I assume that the CDS rates are observed with a measurement error. In particular, assume that the observation equation is given by:

$$0 = 2X_i \Upsilon^i (\pi, t, T) - \hat{c}_i(t, T) \sum_{k=1}^{n} \beta^{6s} \Psi^i (\pi, t, t + 6s) + \eta_{iT,t},$$

where $\hat{c}_i(t, T)$ are the observed CDS rates and the vector of maturity-specific measurement errors $\eta_{it} = [\eta_{i1,t}, \ldots, \eta_{iT,t}]$ is normally distributed and i.i.d. across time and firms: $\eta_i \sim N(0, \Sigma_\eta)$. Taking the point estimates from the Gibbs sampling procedure as estimates of the reference model parameters, I make draws of the vector $\theta^{-1} = \left[\theta_d^{-1}\ \theta_s^{-1}\right]'$ using a Random Walk Metropolis algorithm with a flat prior. The accept/reject probability for the draws of θ^{-1} is the ratio of the likelihood of the CDS rates for all firm, at all available data points and for all available maturities.

To evaluate whether the investors' attitudes toward ambiguity change during the crisis, I conduct three different estimations of the parameters θ_d^{-1} and θ_s^{-1} using different data sub-periods: the pre-crisis period, the period from the start of the crisis to the bailout of Bear Stearns and the period from the bailout of Bear Stearns to the liquidation of Lehman Brothers. The results of these estimations are presented in Table IV. Notice that the estimation suggests that the investors in this market exhibit a higher degree of aversion to ambiguity about the filter distribution than to ambiguity about the underlying dynamics. Notice further that the three different periods do not yield significantly different estimates of θ_d^{-1} and θ_s^{-1}, suggesting that the investors attitudes toward model misspecification do not change significantly during the crisis period. Instead, the observed changes in CDS rates are due to the amount of ambiguity faced by agents in the markets and to shifts in what kind of misspecification the agents are more concerned about.

(Table IV about here.)

Consider now the model-implied time series evolution of credit spreads. Since the estimate of θ^{-1} does not change during the crisis, I use the pre-crisis estimates of θ_d^{-1} and θ_s^{-1} to compute the model-implied CDS spreads. Table V presents the observed 5 year CDS spreads for the financial institutions at five dates of interest – before the start of the crisis, July 2007, at the start of the crisis in August 2007, after the bailout of Bear Stearns in March 2008, after the liquidation of Lehman Brothers in September 2008 and after the introduction of TARP in October 2008 – together with the model-implied CDS rates at these dates. Notice first that the model-implied CDS rates follow the observed pattern of increasing during the financial crisis. Further, for most institutions, the implied CDS rates match the levels of CDS spreads over time, although the performance of the estimated model worsens after the liquidation of Lehman Brothers.

(Table V about here.)

Recall that the estimates of the ambiguity aversion parameters θ_d^{-1} and θ_s^{-1} do not change significantly over the crisis. What then generates the time series behavior of CDS rates that we observed during the crisis? In this paper, I argue that the observed rapid increases in CDS spreads are driven by changes in the amount of total ambiguity faced by investors in the market, as well as changes in how this total amount of ambiguity is decomposed into the amount of ambiguity about the underlying dynamics and the amount of ambiguity about the filter distribution. In Fig. 10, I plot the time series evolution of ambiguity, as measured by the conditional relative entropy between the reference and the worst-case likelihoods[4], during the crisis. Notice that at the beginning of the crisis (after the BNP Paribas announcement in August 2007), only ambiguity about the filter distribution increased. Intuitively, at the beginning of the crisis, although investors observed the negative reports released by financial institutions, since no defaults of major institutions occurred, they could interpret these new signals as informing about the quality of information and not about the quality of the assets on the balance sheets of these institutions. Contrast this with the changes to the amount of ambiguity after the bailout of Bear Stearns in March 2008 and after the liquidation of Lehman Brothers in September 2008. In these cases, it is the amount of ambiguity about the underlying dynamics that increases. Unlike the start of the crisis, investors now observed default by major financial institutions, which increased their uncertainty about the quality of assets held by financial institutions. Notice also that, although the amount of total entropy decreased a little after the introduction of TARP in October 2008, it was still higher than before the start of the crisis, which is consistent with the Caballero and Krishnamurthy (2008a) intuition that ambiguity increased during the crisis.

(Figure 10 about here.)

[4]See the definition of conditional relative entropy in eq. (4.13)

To further understand the time series evolution of asset prices during the crisis, consider the time series evolution of expected time to default of the financial institutions. Table VI presents the expected time to default for the financial institutions at five dates of interest – before the start of the crisis, July 2007, at the start of the crisis in August 2007, after the bailout of Bear Stearns in March 2008, after the liquidation of Lehman Brothers in September 2008 and after the introduction of TARP in October 2008 – together with the percentage change in the time to default relative to the previous month. The initial BNP Paribas announcement in August 2007 only lead to a decrease in the expected time to default for only Bank of America, Goldman Sachs and Morgan Stanley. This coincides with the intuition that investors interpreted the announcements by financial institutions at the beginning of the crisis as increasing the amount of ambiguity about the filter distribution, and not necessarily indicative of the quality of assets held by the financial institutions. More interesting is the fact that the bailout of Bear Stearns in March 2008 lead to a decrease in the expected time to default for only some firms. In contrast, the liquidation of Lehman Brothers in September 2008 decreased the expected time to default for all institutions except Goldman Sachs and Morgan Stanley. Thus, while the bailout of Bear Stearns induced contagion effects for some of the financial institutions, the effect was not as widespread as that induced by the liquidation of Lehman Brothers. Intuitively, while the bailout of Bear Stearns increased the ambiguity about the underlying dynamics, the fact that the institution was bailed out and not liquidated implied a guarantee for the other institutions and thus mitigated the contagion effects. The liquidation of Lehman Brothers, on the other hand, removed this guarantee and thus the liquidation had more systematic effects.

(Table VI about here.)

Compare this to the evolution of expected times to default under the reference model, presented in Panel B of Table VI. Notice first that, under the reference model, the expected time to default is longer at all dates than under the misspecified model. Intuitively, the misspecification-averse agent perceives the probability of default next period to be greater than under the reference model, decreasing the expected time to default. Next, consider the time evolution of the expected times to default. The initial BNP Paribas announcement leads to a slight (0.03%) decrease in the expected time to default for each of the institutions. As under the misspecified model, the decrease in the expected time to default is much greater after the bailout of Bear Stearns. Notice, however, that the decrease under the reference model is greater than under the misspecified model. Intuitively, since the misspecification-averse agent already has more pessimistic views of the future, observing the bailout of Bear Stearns did not have as a large of an impact on her beliefs as it did on the beliefs under the reference model. Similarly, the decrease under the reference model after the liquidations of Lehman Brothers is larger than under the misspecified model. After introduction of TARP in October 2008, however, the increase in the expected time to default under the reference model is much smaller than under the misspecified model.

Finally, consider the implied time series evolution of equity prices during the crisis. Fig. 11 plots the observed evolution of equity prices together with the model-implied evolution. Although we cannot hope to match the level of equity prices since the firm earnings model in the paper is extremely simplistic, the model should be able to match the observed movements in equity prices. Comparing the model-implied evolution to the true evolution of equity prices, we see that the model-implied equity prices lag the observed equity prices. This is not surprising since the signals used to construct the time series evolution of conditional probabilities are backward-looking; for example, the Case-Shiller 10 Index is constructed using observations over the previous three months. The model is able to capture the overall downward trend of equity prices during the crisis and especially well the sharp drop in equity prices after the bailout of Bear Stearns and after the liquidation of Lehman Brothers.

(Figure 11 about here.)

5.3 Sources of ambiguity and CDS rates

I will now investigate the impact of ambiguity aversion on CDS rates and the expected times to default of the financial institutions examined in Section 5 in greater detail. In order to evaluate the disparate effect of the two sources of ambiguity on market quantities, I will examine two cases:

1. Investors are averse to ambiguity about the underlying dynamics but not to ambiguity about the signal quality ($\theta_s^{-1} = 0$).

2. Investors are averse to ambiguity about the signal quality but not to ambiguity about the underlying dynamics ($\theta_d^{-1} = 0$).

In this exercise, I assume that the value of the non-zero degree of ambiguity aversion is the value estimated in Section 5 using the pre-crisis data (see Table IV). This allows me to examine the counterfactual implication of having only one source of ambiguity for market prices. While it is possible to estimate the full model with different degrees of aversion to different sources of ambiguity using a version of the Metropolis-Hastings algorithm described in the previous section, it would be less illuminating about the differential impact of the two sources of ambiguity and I leave that exercise for future research.

Consider now the CDS rates implied for different sources of ambiguity. Table VII presents the percent deviation between the model-implied and the observed five year CDS rates for different financial institutions at five dates of interest – before the start of the crisis in July 2007, at the start of the crisis in August 2007, after the bailout of Bear Stearns in March 2008, after the liquidation of Lehman Brothers in September 2008 and after the introduction of TARP in October 2008 – for three different specifications of ambiguity preference. In particular, the column "Main" refers to the case in the main body of the paper, with the representative agent averse to ambiguity about both the underlying dynamics and the filter distribution; the column θ_s^{-1} refers to the case when the representative agent is averse to ambiguity about the underlying dynamics only; the column $\theta_d^{-1} = 0$ refers to the case when the representative agent is averse to ambiguity about the filter distribution only. Notice first that, the model with aversion to both sources of ambiguity fits the data better overall than either the model with aversion to only the underlying dynamics or the model with aversion to only the filter distribution. Thus, both sources of ambiguity are important in explaining the behavior of CDS rates during the crisis. Notice also that neither the model with aversion to only ambiguity about the underlying dynamics outperforms consistently the model with aversion to only ambiguity about filter distribution nor *vice versa*. This is consistent with the intuition provided by the time series of the different contributions to entropy in Fig. 10. More specifically, at the start of the crisis (July – August, 2007), the model with aversion to only ambiguity about the filter distribution outperforms the model with aversion to only ambiguity about the underlying dynamics, implying that, during this period, ambiguity about the filter distribution had a stronger impact on CDS rates. This relationship reversed as the crisis progressed, with the model with aversion to only ambiguity about the underlying dynamics outperforming the model with aversion to only ambiguity about the filter distribution. Thus, during the later part of the crisis, ambiguity about the underlying dynamics had a stronger impact on CDS spreads.

(Table VII about here.)

6 Conclusion

In this paper, I consider the implications of model misspecification for default swap spreads. Using an incomplete information version of the Black and Cox (1976) model of credit spreads as the reference model, I find that introducing misspecification concerns exacerbates the imperfect information problem faced by the representative agent. This leads to an increased level of default swap spreads overall and greater sensitivity of CDS spreads to bad news. The misspecification-averse agent perceives the probability of default next period to be higher than under the reference model, increasing CDS spreads and decreasing expected time to default. Observing a bad signal not only increases the conditional probability of being in a low-payoff state in the current period but also increased the perceived probability of default in the next period.

To investigate whether the model can produce reasonable magnitudes of CDS spreads, I estimated the parameters of the reference model using observations of the book value of equity of several financial institutions as firm-specific signals and of Case Shiller 10 Index as observations of aggregate signals. The misspecification preference parameters θ_d^{-1} and θ_s^{-1} were then estimated using observations of CDS spreads for the financial institutions over time. The results of the estimation procedure suggest that, while agents' preference toward model misspecification did not change during the crisis, the amount of entropy in the economy and how that entropy is decomposed into the contributions from misspecification of the joint distribution of next period's signals and state and misspecification of the conditional probability distribution of the current state did change. In particular, the initial BNP Paribas announcement in August 2007 lead to an increase in ambiguity about the filter distribution. The bailout of Bear Stearns in March 2008 and the liquidation of Lehman Brothers in September 2008 on the other hand lead to an increase in the amount of ambiguity about the underlying dynamics.

Examining the implied time-series evolution of equity prices, I find that, while the model is able to match the overall movements of the equity prices, the model-implied equity prices lag the observed equity prices. A possible avenue of future research is to estimate the model under the assumption that the representative agent in the economy observes more information than the econometrician estimating the model. While allowing for equity prices to adjust quicker to news in the market, this would also allow us to estimate the model at a higher frequency and, thus, extract more information from the observed CDS spreads.

I turn next to examining the disparate effect of the two sources of ambiguity aversion on market prices. I find that, while both sources of ambiguity play an important role in determining CDS spreads, aversion to different sources of ambiguity plays an important role during different times. In particular, the model with aversion to only ambiguity about the filter distribution is better able to explain the evolution of CDS spreads at the start of the crisis, while aversion to only ambiguity about the underlying dynamics performs better during the later part of the crisis. Notice that this corresponds to the intuition delivered by looking at the time series evolution of the two contributions to entropy.

Notice that, while the model described in this paper is geared toward explaining the observed increases in CDS spreads, similar intuition could be used to explain observed changes to prices of collateralized debt obligations (CDOs) and other complex securities. In fact, since arguably CDOs have a more complicated underlying structure than default swaps, model misspecification concerns would be even more relevant in pricing these securities. A formal treatment of this problem, however, is left for future research.

References

J.H. Albert and S. Chib. Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts. *Journal of Business and Economic Statistics*, 11(1):1–15, 1993.

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

E. W. Anderson, L. P. Hansen, and T. J. Sargent. Small noise methods for risk-sensitive/robust economies. Working Paper, 2010.

F. Black and J.C. Cox. Valuing corporate securities: Some effects of bond indenture provisions. *Journal of Finance*, 31(2):351–367, 1976.

R. J. Caballero and A. Krishnamurthy. Collective risk management in a flight to quality episode. *Journal of Finance*, 63:2195–2230, 2008a.

R. J. Caballero and A. Krishnamurthy. Musical chairs: a comment on the credit crisis. *Banque de France, Financial Stability Review – special issue on liquidity*, 11, 2008b.

C.F. Camerer. Ambiguity − aversion and non − additive probability: Experimental evidence, models and applications. In L. Luini, editor, *Uncertain Decisions: Bridging Theory and Experiments*, pages 53–80. Kluwer Academic Publisher, 1999.

J. Campbell and J. Cochrane. By force of habit: A consumption-based explanation of aggregate stock-market behavior. *Journal of Political Economy*, 107:205–251, 1999.

M.C. Campi and M.R. James. Nonlinear discrete-time risk-sensitive optimal control. *International Journal of Robust and Nonlinear Control*, 6:1–19, 1996.

H. Chen. Macroeconomic conditions and the puzzles of credit spreads and capital structure. *Journal of Finance*, 65, 2010.

L. Chen, P. Collin-Dufresne, and R. Goldstein. On the relation between the credit spread puzzle and the equity premium puzzle. *Review of Financial Studies*, 22:3367–3409, 2009.

Z. Chen and L. G. Epstein. Ambiguity, risk and asset returns in continuous time. *Econometrica*, 70:1403–1443, 2002.

D. Duffie and D. Lando. Term structures of credit spreads with incomplete accounting information. *Econometrica*, 69:633–664, 2001.

D. Ellsberg. Risk, ambiguity and the savage axiom. *Quarterly Journal of Economics*, 75: 643–669, 1961.

L. G. Epstein and J. Miao. A two-person dynamic equilibrium under ambiguity. *Journal of Economic Dynamics and Control*, 27:1253–1288, 2003.

L. G. Epstein and M. Schneider. Ambiguity, information quality and asset pricing. *Journal of Finance*, LXIII(1):197–228, 2008.

L. G. Epstein and S.E. Zin. Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57:937–969, 1989.

R. Frey and T. Schmidt. Pricing corporate securities under noisy asset information. *Mathematical Finance*, 19(3):403–421, 2009.

R. Frey and T. Schmidt. Filtering and incomplete information in credit risk. In D. Brigo, T. Bielecki, and F. Patras, editors, *Credit Risk Frontiers-Suprime Chrisis, Pricing and Hedging*. John Wiley and Sons, 2010.

S. Fruhwirth-Schnatter. Markov chain monte carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96: 194–209, 2001.

S. Fruhwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. Springer-Verlag, 2006.

P. Gagliardini, P. Porchia, and F. Trojani. Ambiguity aversion and the term structure of interest rates. *Review of Financial Studies*, 22:4157–4188, 2009.

Y. Halevy. Ellsberg revisited: An experimental study. *Econometrica*, 75:503–536, 2007.

L.P. Hansen and T.J. Sargent. Discounted linear exponential quadratic gaussian control. *IEEE Trans. Autom. Control*, 40:968–971, 1995.

L.P. Hansen and T.J. Sargent. Robust estimation and control under commitment. *Journal of Economic Theory*, 124(2):258–301, 2005.

L.P. Hansen and T.J. Sargent. Recursive robust estimation and control without commitment. *Journal of Economic Theory*, 136:1–27, 2007.

L.P. Hansen, T.J. Sargent, and T.D. Tallarini Jr. Robust permanent income and pricing. *Review of Economic Studies*, 66:873–907, 1999.

D.H. Jacobson. Optimal stochastics linear systems with exponential performance criteria and their relation to deterministic differential games. *IEEE Transactions on Automatic Control*, 18:124–131, 1973.

P. Klibanoff, M. Marinacci, and S. Mukerji. A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892, 2005.

P. Klibanoff, M. Marinacci, and S. Mukerji. Recursive smooth ambiguity preferences. *Journal of Economic Theory*, 144:930–976, 2009.

M. Leippold, F. Trojani, and P. Vanini. Learning and asset prices under ambiguous information. *Review of Financial Studies*, 21:2565–2597, 2008.

H.E. Leland. Corporate debt value, bond covenants and optimal capital structure. *Journal of Finance*, 49(4):1213–1252, 1994.

J. Liu, J. Pan, and T. Wang. An equilibrium model of rare-event premia and its implication for option smirks. *Review of Financial Studies*, 18:131–164, 2005.

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

R.E. McCulloch and R.S. Tsay. Statistical analysis via markov switching models. *Journal of Time Series Analysis*, 15(5):523–539, 1994.

E. Seneta. *Non-negative matrices and Markov Chains*. Springer, 1981.

T.D. Tallarini Jr. Risk-sensitive real business cycles. *Journal of Monetary Economics*, 45: 507–532, 2000.

F. Trojani and P. Vanini. A note on robustness in merton's model of intertemporal consumption and portfolio choice. *Journal of Economic Dynamics and Control*, 26(3): 423–435, 2002.

F. Trojani and P. Vanini. Robustness and ambiguity aversion in general equilibria. *The Review of Finance*, 2:423–435, 2004.

R. Uppal and T. Wang. Model misspecification and under-diversification. *Journal of Finance*, LIX:2465–2486, 2003.

A Reference Model Estimation

I estimate the time series parameters of the reference model using a Gibbs sampling procedure. Recall the Gibbs sampling allows to sequentially make parameter draws from conditional posteriors. Because the model can be broken down into as many conditional posteriors as needed, it is possible to fully estimate the reference model specified in Section 4.

Recall that for the model of Section 4.1, the observations in the economy are given by:

$$\begin{aligned} y_{it} &= \xi_{i,s_t} + \rho_i \xi_{I+1,s_t} + u_{it}, \qquad i = 1, \ldots, I \\ y_{ct} &= \xi_{I+1,s_t} + u_{ct}, \end{aligned}$$

where s_t is the indicator of the state at date t. In reality, observations of book value occur at a quarterly frequency while observations of the aggregate signal occur at a monthly frequency. Denote by n_y the number of periods in between observations of firm-specific signals and by t_y the number of available observations of the firm-specific signals.

To reduce the number of parameters to be estimated, I impose additional restrictions on the model. In particular, I assume that the vector of the firm-specific components $z_{ft} \equiv [z_{1t}, \ldots, z_{It}]$ of the hidden state vector z evolves independently of the aggregate component, z_{ct}. That is, I assume that z_{ft} and z_{ct} evolve as two independent Markov chains. The vector of firm-specific components z_f evolves as an n_f-state Markov chain, with values $\zeta_{f1}, \ldots, \zeta_{f,n_f}$ and transition probability matrix Ω_f defined by:

$$\{\Omega_f\}_{jk} \equiv \omega_{f,ik} = \mathbb{P}\left(z_{f,t+1} = \zeta_{fk} \middle| \, z_{ft} = \zeta_{fj} \right).$$

Similarly, the aggregate component z_{ct} evolves as an n_c Markov chain, with values $\zeta_{c1}, \ldots, \zeta_{cn_c}$ and transition matrix Ω_c defined by:

$$\{\Omega_c\}_{jk} \equiv \omega_{ci,jk} = \mathbb{P}\left(z_{c,t+1} = \zeta_{ck} \middle| \, z_{ct} = \zeta_{cj} \right).$$

I impose also the assumption that the signal errors of the aggregate signal are uncorrelated with the signal errors of the firm-specific signals but allow for the errors of the firm-specific signals to be cross-sectionally correlated. That is, I partition the signal covariance matrix into:

$$\Sigma_u = \begin{bmatrix} \Sigma_{uf} & \vec{0}_{I,1} \\ \vec{0}_{1,I} & \Sigma_{uc} \end{bmatrix},$$

where Σ_{uf} is the covariance matrix of the firm-specific signals and Σ_{uc} is the variance of the aggregate signal. Notice that this formulation allows me to estimate the firm value at default directly from the signal observations: since z_{ft} and z_{ct} evolve as two independent Markov chains, it is possible to recover the lowest value of the firm-specific component and the lowest value of the common component without observing default.

Denote by Θ the full set of parameters to be estimated:

$$\Theta = \left\{\Omega_f, \Omega_c, \zeta_{f1}, \ldots, \zeta_{fn_f}, \zeta_{c1}, \ldots, \zeta_{cn_c}, \rho_1, \ldots, \rho_I, \Sigma_{uf}, \Sigma_{uc}, \{s_{ft}\}_{t=1}^{t_y}, \{s_{ct}\}_{t=1}^{T}\right\}$$

and by Θ_{-A} the set of all parameters except A: $\Theta_{-A} = \Theta \backslash A$. With the above assumptions, the main steps in the Gibbs procedure are as follows:

Step 1. Conditional on a draw of $\Theta_{-\Omega_f}$, make a draw of Ω_f

Step 2. Conditional on a draw of $\Theta_{-\Omega_c}$, make a draw of Ω_c

Step 3. Conditional on a draw of $\Theta_{-\{\zeta_{ci}\}_{i=1}^{n_c}}$, make a draw of $\zeta_{c1}, \ldots, \zeta_{c,n_c}$

Step 4. Conditional on a draw of $\Theta_{-\left\{\zeta_{fi}\right\}_{i=1}^{n_f}}$, make a draw of $\zeta_{f1}, \ldots, \zeta_{f,n_f}$

Step 5. Conditional on a draw of $\Theta_{-\{\rho_i\}_{i=1}^{I}}$, make a draw of $\rho_1, \ldots, \rho_I$

Step 6. Conditional on a draw of $\Theta_{-\Sigma_{uf}}$, make a draw of Σ_{uf}

Step 7. Conditional on a draw of $\Theta_{-\Sigma_{uc}}$, make a draw of Σ_{uc}

Step 8. Conditional on a draw of $\Theta_{-\{s_{ct}\}_{t=1}^{T}}$, make a draw of $\{s_{ct}\}_{t=1}^{T}$

Step 9. Conditional on a draw of $\Theta_{-\left\{s_{ft}\right\}_{t=1}^{t_y}}$, make a draw of $\{s_{ft}\}_{t=1}^{t_y}$

Step 10. Permute the state indicators

For the conditional posteriors below, I rely on Gibbs sampling results for regime-switching models. The initial application of MCMC estimation methods to regime-switching models in due to Albert and Chib (1993), who estimate an autoregressive model with Markov jumps following a two-state Markov process. McCulloch and Tsay (1994) extend this to situations where the regime-switching model includes non-regime-specific (common) components. For the most part (and in the algorithm below), practical MCMC estimation uses the principle of data augmentation and treats the indicators of the state of the latent Markov chain as missing data. Treating observations of the latent Markov chain as missing data allows for the use of conjugate priors in estimating the parameters of the model. For a more exhaustive discussion of the use of MCMC methods for estimating the parameters of Markov chains, see Fruhwirth-Schnatter (2006).

A.1 Conditional on a draw of $\Theta_{-\Omega_f}$, make a draw of Ω_f

Denote by $\tilde{\Omega}_f$ the n_y-periods-ahead transition probability matrix of the firm-specific value vector z_{ft}: $\tilde{\Omega}_f = \Omega_f^{n_y}$. Since obsevations of the firm-specific signals occur only every n_y periods and the aggregate signals are not informative about the firm-specific state, I make draws of $\tilde{\Omega}_f$ and then infer the corresponding draw of Ω_f. Generalizing the results of Albert and Chib (1993) and McCulloch and Tsay (1994) to the multiple state case, the conjugate prior for the j^{th} row of $\tilde{\Omega}_f$ is :

$$\tilde{\omega}_{f,j} \sim Dir\left(\alpha_{j1}^f, \ldots, \alpha_{j,n_f}^f\right),$$

where Dir denotes the Dirichlet distribution[5]. The posterior is then given by:

$$\tilde{\omega}_{f,j} \sim Dir\left(\alpha_{j1}^f + n_{j1}^f, \ldots, \alpha_{j,n_f}^f + n_{jn_f}^f\right),$$

where n_{jk}^f is the number of times the Markov chain z_f transitions from state j to state k in the current draw of $\{s_{ft}\}_{t=1}^{t_y}$. Once a draw of $\tilde{\Omega}_f$ is made, the corresponding draw of the original transition probability matrix is computed as $\Omega_f = \tilde{\Omega}_f^{\frac{1}{n_y}}$.

A.2 Conditional on a draw of $\Theta_{-\Omega_c}$, make a draw of Ω_c

Similarly to $\tilde{\Omega}_f$, the conjugate prior for the j^{th} row of Ω_c is the Dirichlet distribution:

$$\omega_{c,j} \sim Dir\left(\alpha_{j1}^c, \ldots, \alpha_{j,n_c}^c\right),$$

and the posterior is given by:

$$\omega_{c,j} \sim Dir\left(\alpha_{j1}^c + n_{j1}^c, \ldots, \alpha_{j,n_c}^c + n_{j,n_c}^c\right),$$

where n_{jk}^c is the number of times the Markov chain z_c transitions from state j to state k in the current draw of $\{s_{ct}\}_{t=1}^{T}$.

A.3 Conditional on a draw of $\Theta_{-\{\zeta_{ci}\}_{i=1}^{n_c}}$, make a draw of $\zeta_{c1}, \ldots, \zeta_{c,n_c}$

To derive the conditional posterior of ζ_{ci}, $i = 1, \ldots, n_c$, notice that the firm-specific signals contain information about the common component of the fundamental asset values. In particular, notice that the likelihood function of the signals is given by:

$$\mathcal{L}\left(y|\,\Theta\right) \;\; \propto \;\; \exp\left\{-\frac{1}{2}\sum_{\tau=1}^{t_y} \hat{y}_{f,\tau\Delta_y}' \Sigma_{uf}^{-1} \hat{y}_{f,\tau\Delta_y} - \frac{1}{2}\sum_{t=1}^{T} \frac{\left(y_{ct} - \zeta_{c,s_{ct}} - \overline{u}_c\right)^2}{\Sigma_{uc}}\right\},$$

where $\hat{y}_{f,\tau\Delta_y} = y_{f,\tau\Delta_y} - \zeta_{f,s_{f,\tau\Delta_y}} - \rho\zeta_{c,s_{c,\tau\Delta_y}} - \overline{u}_f$. Let $i_1 < i_2 < \ldots < i_{n_j}$ denote all the time indices such that $s_{ci_k} = j$ and let $y_{c,j} = \left(y_{ci_1}, \ldots, y_{ci_{n_j}}\right)'$. The conjugate prior is

[5] Recall that the Dirichlet distribution generalizes the beta distribution to the multinomial case

given by:

$$\zeta_{cj} \sim N\left(\zeta_{c,0j},\ \Sigma_{uc} A_{c,0j}^{-1}\right)$$

and the conditional prior by:

$$\zeta_{cj} \sim N\left(\bar{\zeta}_{cj},\ \Sigma_{uc} \bar{A}_{cj}^{-1}\right),$$

where

$$\begin{aligned}
\bar{A}_{cj} &= n_{y_j} \Sigma_{uc} \rho' \Sigma_{uf}^{-1} \rho + n_j + A_{c,0j} \\
\bar{\zeta}_{cj} &= \bar{A}_{cj}^{-1} \left[\Sigma_{uc} \rho' \Sigma_{uf}^{-1} \sum_{k=1}^{n_{y_j}} (y_{f,\tau_k \Delta_y} - \zeta_{f,s_{1\tau_k \Delta_y}} - \overline{u}_f) + \sum_{k=1}^{n_j} (y_{c,i_k} - \overline{u}_c) + A_{c,0j} \zeta_{c,0j} \right]
\end{aligned}$$

A.4 Conditional on a draw of $\Theta_{-\{\zeta_{fi}\}_{i=1}^{n_f}}$, make a draw of $\zeta_{f1}, \ldots, \zeta_{f,n_f}$

Denote: $\tilde{y}_{ft} = y_{ft} - \rho \zeta_{cs_{ct}} - \bar{u}$. Let $j_1 < j_2 < \ldots < j_{n_j}$ denote all the time indices such that $s_{fj_k} = j$ and let $\tilde{y}_{fj} = \left(\tilde{y}_{f,j_1}, \ldots, \tilde{y}_{f,j_{n_j}}\right]'$. Then the conjugate prior distribution is:

$$\zeta_{fj} \sim N\left(\zeta_{f,0j},\ \Sigma_{uf} A_{f,0j}^{-1}\right)$$

and the conditional posterior by:

$$\zeta_{fj} \sim N\left(\bar{\zeta}_{fj},\ \Sigma_{uf} \bar{A}_{fj}^{-1}\right),$$

where:

$$\begin{aligned}
\bar{A}_{fj} &= n_j + A_{f,j0} \\
\bar{\zeta}_{fj} &= \bar{A}_{fj}^{-1} \left[\sum_{k=1}^{n_j} \tilde{y}_{f,j_k} + A_{f,j0} \zeta_{f,j0} \right]
\end{aligned}$$

A.5 Conditional on a draw of $\Theta_{-\{\rho_i\}_{i=1}^{I}, \Sigma_{uf}}$, make a draw of $\rho_1, \ldots, \rho_I$ and

Define: $\tilde{y}_{ft} = y_{ft} - \overline{u}_f - \zeta_{fs_{ft}}$. The prior distribution for the vector ρ is then

$$\rho \sim N\left(\beta_0, \Sigma_{uf} A_{0\rho}^{-1}\right)$$

and the conjugate posterior is given by:

$$\rho \sim N\left(\beta_*, \Sigma_{uf} A_{*,\rho}^{-1}\right),$$

where:

$$\begin{aligned} A_{*,\rho} &= \sum_{t=1}^{t_y} \zeta_{c,s_{ct}}^2 + A_{0\rho} \\ \beta_* &= A_{*,\rho}^{-1} \left(\sum_{t=1}^{t_y} \tilde{y}_{ft} \zeta_{c,s_{ct}} + A_{0\rho}\beta_0 \right). \end{aligned}$$

The conjugate prior for Σ_{uf} is given by:

$$\Sigma_{uf} \sim IW(\nu, V)$$

and the posterior is given by:

$$\Sigma_{uf} \sim IW(\nu + t_y, V + S),$$

where

$$S = \sum_{t=1}^{t_y} \left(\tilde{y}_{f,t\Delta_y} - \rho\zeta_{c,s_{c,t\Delta_y}} \right)' \left(\tilde{y}_{f,t\Delta_y} - \rho\zeta_{c,s_{c,t\Delta_y}} \right) + (\beta_* - \beta_0)' A_{0\rho} (\beta_* - \beta_0).$$

A.6 Conditional on a draw of $\Theta_{-\Sigma_{uc}}$, make a draw of Σ_{uc}

In this case, the conjugate prior is an inverse χ^2-distribution:

$$\Sigma_{uc} \sim \frac{\nu_c \bar{s}_c^2}{\chi^2(\nu_c)}$$

and the conditional posterior by:

$$\Sigma_{uc} \sim \frac{\nu_c \bar{s}_c^2 + TE_c}{\chi^2(\nu_c + T)},$$

where:

$$TE_c = \sum_{t=1}^{T} (y_{ct} - \zeta_{cs_{ct}} - \overline{u}_c)^2.$$

A.7 Conditional on a draw of $\Theta_{-\{s_{ct}\}_{t=1}^T}$, make a draw of $\{s_{ct}\}_{t=1}^T$

Denote by S_{ct} the history of observations of the aggregate regime up to date t, $S_{c,-t}$ the full history of the aggregate regime except at date t. The conditional posterior is then given by:

$$\mathbb{P}\left(s_{ct}\middle|\{y_{c\tau}\}_{\tau=1}^{T}, S_{c,-t}\right) \propto \mathbb{P}(s_{ct}|s_{c,t-1})\mathbb{P}(s_{c,t+1}|s_{ct})\exp\left\{-\frac{1}{2}\frac{(y_{ct}-\zeta_{c,s_{ct}})^2}{\Sigma_{uc}}\right\} \qquad t \neq nn_y, n \in \mathbb{N}$$

$$\mathbb{P}\left(s_{ct}\middle|\{y_{c\tau}\}_{\tau=1}^{T}, S_{c,-t}\right) \propto \mathbb{P}(s_{ct}|s_{c,t-1})\mathbb{P}(s_{c,t+1}|s_{ct})\exp\left\{-\frac{1}{2}\frac{(y_{ct}-\zeta_{c,s_{ct}})^2}{\Sigma_{uc}}\right\}$$
$$\times \exp\left\{-\frac{1}{2}\left(y_{ft}-\zeta_{f,s_{ft}}-\rho\zeta_{cs_{ct}}\right)'\Sigma_{uf}^{-1}\left(y_{ft}-\zeta_{f,s_{ft}}-\rho\zeta_{cs_{ct}}\right)\right\} \qquad t = nn_y, n \in \mathbb{N}$$

A.8 Conditional on a draw of $\Theta_{-\{s_{ft}\}_{t=1}^{t_y}}$, make a draw of $\{s_{ft}\}_{t=1}^{t_y}$

Denote S_{ft} the history of observations of the firm-specific regime up to date t, $S_{f,-t}$ the full history of observations of the firm-specific regime except at date t. Then, the conditional posterior for s_{ft} is given by:

$$\mathbb{P}\left(s_{ft}\middle|\{y_{f\tau}\}_{\tau=1}^{T}, S_{f,-t}\right) \propto \mathbb{P}(s_{ft}|s_{f,t-1})\mathbb{P}(s_{f,t+1}|s_{ft})$$
$$\times \exp\left\{-\frac{1}{2}\left(y_{ft}-\zeta_{f,s_{ft}}-\rho\zeta_{cs_{ct}}\right)'\Sigma_{uf}^{-1}\left(y_{ft}-\zeta_{f,s_{ft}}-\rho\zeta_{cs_{ct}}\right)\right\}.$$

A.9 Permute the state indicators

As discussed in Fruhwirth-Schnatter (2001), the behavior of the sampler described above is somewhat unpredictable, and the sampler might be trapped at one modal region of the Markov mixture posterior distribution or may jump occasionally between different model regions causing label switching. A simple but efficient solution to obtain a sampler that explores the full Markov mixture posterior distribution is suggested in Fruhwirth-Schnatter (2006). Each draw from the Gibbs sampler is concluded by selecting randomly one of $n_f!$ possible permutations of the current labeling of the firm-specific states and one $n_c!$ possible permutations of the current labeling of the aggregate states. This permutation is then applied to the transition probability matrices Ω_f and Ω_c, the state-specific parameters ζ_f and ζ_c and the state indicators $s_f^{t_y}$ and s_c^T.

B Model Misspecification

In this section, I describe the derivation of the risk-sensitive recursion (4.9) and the distortions (4.11)-(4.12) to the filtering distributions. I rely on the results of Hansen and Sargent (2005), Hansen and Sargent (2007) to formulate the model misspecification problem faced by the representative investor.

Let M_t be a non-negative $\mathcal{F}_t$-measurable random variable, with $\mathbb{E}[M_t] = 1$. Using M_t as a Radon-Nikodym derivative generates a distorted probability measure that is absolutely continuous with respect to the probability measure over $\mathcal{F}_t$ generated by the model (4.1). Under the distorted measure, the expectation of a bounded $\mathcal{F}_t$-measurable random variable W_t is $\tilde{\mathbb{E}}\left[W_t\right] = \mathbb{E}\left[M_t W_t\right]$.

To construct the implied (distorted) conditional density, Hansen and Sargent (2007) factor the martingale M_t into one-step-ahead random variables. More specifically, for a non-negative martingale $\{M_t\}_{t\geq 0}$ form:

$$m_{t+1} = \begin{cases} \frac{M_{t+1}}{M_t} & \text{if } M_t > 0 \\ 1 & \text{if } M_t = 0 \end{cases}$$

Then $M_{t+1} = m_{t+1} M_t$ and, for any $t \geq 0$, the martingale M_t can be represented as:

$$M_t = M_0 \prod_{j=1}^{t} m_j,$$

where the random variable M_0 has unconditional expectation equal to unity. Notice that, by construction, m_{t+1} has date t conditional expectation equal to unity. Thus, for a bounded $\mathcal{F}_{t+1}$-measurable random variable W_{t+1}, the distorted conditional expectation implied by the martingale $\{M_t\}_{t\geq 0}$ is constructed as:

$$\tilde{\mathbb{E}}\left[W_{t+1} \middle| \mathcal{F}_t\right] \equiv \frac{\mathbb{E}\left[M_{t+1} W_{t+1} \middle| \mathcal{F}_t\right]}{\mathbb{E}\left[M_{t+1} \middle| \mathcal{F}_t\right]} = \frac{\mathbb{E}\left[M_{t+1} W_{t+1} \middle| \mathcal{F}_t\right]}{M_t} = \mathbb{E}\left[m_{t+1} W_{t+1} \middle| \mathcal{F}_t\right],$$

provided that $M_t > 0$. Let $\mathcal{M}_t$ be the space of all non-negative $\mathcal{F}_t$-measurable random variables m_t for which $\mathbb{E}\left[m_t \middle| \mathcal{F}_{t-1}\right] = 1$. The elements of $\mathcal{M}_{t+1}$ represent all possible distortions of the conditional distribution over $\mathcal{F}_{t+1}$ given $\mathcal{F}_t$; that is, each $m_{t+1} \in \mathcal{M}_{t+1}$ represents a possible distortion to the underlying asset value dynamics. The amount of distortion introduced by m_{t+1} is measured each period using the conditional relative entropy between the reference and distorted models:

$$\epsilon_t^1(m_{t+1}) = \mathbb{E}\left[m_{t+1} \log m_{t+1} \middle| \mathcal{F}_t\right]. \tag{B.1}$$

To introduce distortion to the signal model, consider factoring the martingale M_t in a different way. More specifically, introduce the $\mathcal{G}_t$-measurable random variable $\hat{M}_t = \mathbb{E}\left[M_t \middle| \mathcal{G}_t\right]$ and define:

$$h_t = \begin{cases} \frac{M_t}{\hat{M}_t} & \text{if } \hat{M}_t > 0 \\ 1 & \text{if } \hat{M}_t = 0 \end{cases}$$

The $\mathcal{G}_t$-measurable random variable $\hat{M}_t$ implies a likelihood ratio for the partial information set $\mathcal{G}_t$ while the $\mathcal{F}_t$-measurable random variable h_t represents distortions to the probability distribution over $\mathcal{F}_t$ given $\mathcal{G}_t$. Define $\mathcal{H}_t$ to be the space of all non-negative $\mathcal{F}_t$-measurable random variables h_t for which $\mathbb{E}\left[h_t \middle| \mathcal{G}_t\right] = 1$. Similarly to (B.1), the amount of distortion induced by h_t is measured as:

$$\epsilon_t^2(h_t) = \mathbb{E}\left[h_t \log h_t | \, \mathcal{G}_t\right]. \tag{B.2}$$

To solve for the worst-case likelihood, introduce an entropy penalization parameter $\theta > 0$ which captures the beliefs of the representative agent about the amount of misspecification in the economy: as θ increases, the set of admissible alternative models decreases, with the limiting case $\theta = \infty$ corresponding to only the reference model being admissible. Begin by considering the full-information case. Corresponding to each $\mathcal{F}_{t+1}$-measurable random variable m_{t+1} is a relative density $\phi(z^*, y^*)$. The minimizing agent solves:

$$\min_{\phi \geq 0} \sum_{j=1}^{N} \int \left[W(\xi_j, \pi^*, y^*) + \theta \log \phi(\xi_j, y^*)\right] \phi(\xi_j, y^*) \tau(\xi_j, y^* | z, y) dy^*$$

subject to:

$$\sum_{j=1}^{N} \int \phi(\xi_j, y^*) \tau(\xi_j, y^* | z, y) dy^* = 1,$$

where $*$ denote next period values and $\tau(z^*, y^*|z, y)$ is the joint transition probability:

$$\tau(z^*, y^*|z, y) = |2\pi\Sigma_u|^{-\frac{1}{2}} \lambda_{zz^*} \exp\left[-\frac{1}{2}(y^* - z^* - \overline{u})'\Sigma_u^{-1}(y^* - z^* - \overline{u})\right]. \tag{B.3}$$

The solution to the minimization problem implies a transformation T^1 that maps the value function that depends on next period's state (ξ_j, π^*, y^*) into a risk-adjusted value function that depends on the current state (z, π, y):

$$T^1(W|\theta) = -\theta \log \sum_{j=1}^{N} \int \exp\left(-\frac{W(\xi_j, \pi^*, y^*)}{\theta}\right) \tau(\xi_j, y^*|z, y) dy^*. \tag{B.4}$$

The minimizing choice of ϕ is given by:

$$\phi_t(z^*, y^*) = \frac{\exp\left(-\frac{W(z^*, \pi^*, y^*)}{\theta}\right)}{\mathbb{E}\left[\exp\left(-\frac{W(z^*, \pi^*, y^*)}{\theta}\right) \Big| \, \mathcal{F}_t\right]}.$$

Similarly, corresponding to each $\mathcal{G}_t$-measurable random variable h_t is a relative density $\psi(z)$, with the worst-case distortion given as the solution to:

$$\min_{\psi \geq 0} \sum_{j=1}^{N} \left[\hat{W}(\pi, \xi_j) + \theta \log \psi(\xi_j)\right] \psi(\xi_j) p_j$$

subject to:

$$\sum_{j=1}^{N} \psi(\xi_j) p_j = 1.$$

This implies another operator:

$$T^2(\hat{W}|\theta)(\pi) = -\theta \log \sum_{j=1}^{N} \exp\left(-\frac{\hat{W}(\pi, \xi_j)}{\theta} \right) \psi(\xi_j) p_j. \tag{B.5}$$

The corresponding minimizing choice of ψ is given by:

$$\psi_t(z) = \frac{\exp\left(-\frac{\hat{W}(\pi,z)}{\theta}\right)}{\mathbb{E}\left[\exp\left(-\frac{\hat{W}(\pi,z)}{\theta}\right)\middle| \mathcal{G}_t\right]}.$$

C Proofs

C.1 Proof of Lemma 4.2

To obtain a first order approximation to the value function around the point $\theta_d^{-1} = \theta_s^{-1} = 0$, we need to take a second order expansion of the risk-sensitive recursion. In particular, introduce $\theta_s^{-1}/\theta_d^{-1} = 1 + \epsilon$ and rewrite the recursion (4.9) as:

$$\exp\left(-\theta_d^{-1}(1+\epsilon)J_t\right) = \mathbb{E}\left[\exp\left\{-\theta_d^{-1}(1+\epsilon)U(z_t) + (1+\epsilon)\log \mathbb{E}\left[\exp\left(-\beta\theta_d^{-1}J_{t+1}\right)\middle|\right]\right\}\middle| \mathcal{G}_t\right].$$

Then, taking a second order expansion of the above recursion, we have:

$$\begin{aligned}
&\exp\left(-\theta_d^{-1}(1+\epsilon)J_t\right) \approx 1 - J(p;0,0)\theta_d^{-1} + \frac{\theta_d^{-2}}{2}\left(J(p;0,0)^2 - 2\frac{\partial J(p;0,0)}{\partial \theta_d^{-1}}\right) \\
&- \epsilon\left(J(p;0,0) + \frac{\partial J(p;0,0)}{\partial \epsilon}\right) \\
&\exp\left\{-\theta_d^{-1}(1+\epsilon)U(z_t) + (1+\epsilon)\log \mathbb{E}\left[\exp\left(-\beta\theta_d^{-1}J_{t+1}\right)\middle|\right]\right\} \approx 1 - \theta_d^{-1}\left(U(z_t) + \beta\mathbb{E}\left[J(p;0,0)|\,\mathcal{F}_t\right]\right) \\
&+ \frac{\theta_d^{-2}}{2}\left\{\left(U(z_t) + \beta\mathbb{E}\left[J(p;0,0)|\,\mathcal{F}_t\right]\right)^2 - 2\beta\mathbb{E}\left[\frac{\partial J(p;0,0)}{\partial\theta_d^{-1}}\middle| \mathcal{F}_t\right] - \beta^2\mathbb{E}\left[J(p;0,0)|\,\mathcal{F}_t\right]^2\right\} \\
&- \theta_d^{-1}\epsilon\left\{U(z_t) + \beta\mathbb{E}\left[J(p;0,0) + \frac{\partial J(p;0,0)}{\partial\epsilon}\middle| \mathcal{F}_t\right]\right\}.
\end{aligned}$$

Denote:

$$J_0 = J(p;0,0) \qquad J_{\theta_d^{-1}} = \frac{\partial J(p;0,0)}{\partial \theta_d^{-1}} \qquad J_\epsilon = \frac{\partial J(p;0,0)}{\partial \epsilon},$$

so that:

$$
\begin{aligned}
&\exp\left(-\theta_d^{-1}(1+\epsilon)J_t\right) \approx 1 - J_0\theta_d^{-1} + \frac{\theta_d^{-2}}{2}\left(J_0^2 - 2J_{\theta_d^{-1}}\right) - \epsilon\left(J_0 + J_\epsilon\right) \\
&\exp\left\{-\theta_d^{-1}(1+\epsilon)U(z_t) + (1+\epsilon)\log\mathbb{E}\left[\exp\left(-\beta\theta_d^{-1}J_{t+1}\right)\right]\right\} \approx 1 - \theta_d^{-1}\left(U(z_t) + \beta\mathbb{E}\left[J_0(t+1)|\,\mathcal{F}_t\right]\right) \\
&+ \frac{\theta_d^{-2}}{2}\left\{\left(U(z_t) + \beta\mathbb{E}\left[J_0(t+1)|\,\mathcal{F}_t\right]\right)^2 - 2\beta\mathbb{E}\left[J_{\theta_d^{-1}}(t+1)\middle|\,\mathcal{F}_t\right] - \beta^2\mathbb{E}\left[J_0(t+1)|\,\mathcal{F}_t\right]^2\right\} \\
&- \theta_d^{-1}\epsilon\left\{U(z_t) + \beta\mathbb{E}\left[J_0(t+1) + J_\epsilon(t+1)|\,\mathcal{F}_t\right]\right\}
\end{aligned}
$$

Substituting into the risk-sensitive recursion and equating coefficients on powers of θ_d^{-1} and ϵ, we obtain the following system of equations for J_0, $J_{\theta_d^{-1}}$ and J_ϵ:

$$
\begin{aligned}
J_0(t) &= \mathbb{E}\left[U(z_t) + \beta J_0(t+1)|\,\mathcal{G}_t\right] \\
2J_{\theta_d^{-1}}(t) &= J_0(t)^2 + \mathbb{E}\left[2\beta J_{\theta_d^{-1}}(t+1) + \beta^2\mathbb{E}\left[J_0(t+1)|\,\mathcal{F}_t\right]^2 - \left(U(z_t) + \beta\mathbb{E}\left[J_0(t+1)|\,\mathcal{F}_t\right]\right)^2\middle|\,\mathcal{G}_t\right] \\
J_\epsilon(t) &= -J_0(t) + \mathbb{E}\left[U(z_t) + \beta\left(J_0(t+1) + J_\epsilon(t+1)\right)|\,\mathcal{G}_t\right].
\end{aligned}
$$

To solve the above system, notice that, since the unnormalized probability vector π is proportional to the conditional probability vector p, we can use π as the state variable as long as we recognize that J_0 and J_1 are homogeneous of degree 0 in π, so that $J_0(\alpha\pi) = J_0(\pi)$, $J_{\theta_d^{-1}}(\alpha\pi) = J_{\theta_d^{-1}}(\pi)$ and $J_\epsilon(\alpha\pi) = J_\epsilon(\pi)$ $\forall\alpha\in\mathbb{R}$. I look for a first order approximation to the solution in terms of log deviations from the steady state conditional distribution. In particular, denote by $\overline{\pi}$ the stationary distribution of the Markov chain $\{z_t\}_{t\geq 0}$: $\overline{\pi} = \Lambda'\overline{\pi}$. Notice that $\overline{\pi}$ is the steady state conditional distribution in the limiting case of arbitrarily uninformative signals where $\Sigma_u^{-1} = 0$. Introduce $\hat{\pi}$ to be the vector of log deviations from the stationary distribution:

$$
\hat{\pi}_{jt} = \begin{cases} \log\tilde{\pi}_{jt} - \log\overline{\pi}_j; & j \in a_B^c \\ 0 & j \in a_B \end{cases}, \tag{C.1}
$$

where $\tilde{\pi}_{jt} = f(y_t - \xi_j)\sum_{k=1}^N \lambda_{kj}\pi_{k,t-1}$ is the unnormalized probability vector before conditioning on observations of default at date t, and approximate:

$$
\begin{aligned}
J_0(\pi) &\approx J_0(\overline{\pi}) + \frac{\partial J_0(\overline{\pi})'}{\partial \pi} diag(\overline{\pi})\hat{\pi} \\
J_0(\pi)^2 &\approx J_0(\overline{\pi})^2 + 2J_0(\overline{\pi})\frac{\partial J_0(\overline{\pi})'}{\partial \pi} diag(\overline{\pi})\hat{\pi} \\
J_{\theta_d^{-1}}(\pi) &\approx J_{\theta_d^{-1}}(\overline{\pi}) + \frac{\partial J_{\theta_d^{-1}}(\overline{\pi})'}{\partial \pi} diag(\overline{\pi})\hat{\pi} \\
J_\epsilon(\pi) &\approx J_\epsilon(\overline{\pi}) + \frac{\partial J_\epsilon(\overline{\pi})'}{\partial \pi} diag(\overline{\pi})\hat{\pi} \\
(U(z) + \beta \mathbb{E}\left[J_0(\pi_{t+1}) | \mathcal{F}_t \right]) &\approx (U(z) + \beta J_0(\overline{\pi}))^2 + 2\left(U(z) + \beta J_0(\overline{\pi})\right)\beta \frac{\partial J_0(\overline{\pi})'}{\partial \pi} diag(\overline{\pi})\hat{\pi} \\
\mathbb{E}\left[J_0(\pi_{t+1}) | \mathcal{F}_t \right]^2 &\approx J_0(\overline{\pi})^2 + 2J_0(\overline{\pi})\frac{J_0(\overline{\pi})'}{\partial \pi} diag(\overline{\pi})\hat{\pi}.
\end{aligned}
$$

For simplicity, denote: $\gamma_{00} = J_0(\overline{\pi})$, $\gamma_{01} = \frac{\partial J_0(\overline{\pi})}{\partial \pi}$, $\gamma_{10} = J_{\theta_d^{-1}}(\overline{\pi})$, $\gamma_{11} = \frac{\partial J_{\theta_d^{-1}}(\overline{\pi})}{\partial \pi}$, $\gamma_{\epsilon 0} = J_\epsilon(\overline{\pi})$, and $\gamma_{\epsilon 1} = \frac{\partial J_\epsilon(\overline{\pi})}{\partial \pi}$, so that:

$$
\begin{aligned}
J_0(\pi) &\approx \gamma_{00} + \gamma_{01}' diag(\overline{\pi})\hat{\pi} \\
J_0(\pi)^2 &\approx \gamma_{00}^2 + 2\gamma_{00}\gamma_{01}' diag(\overline{\pi})\hat{\pi} \\
J_{\theta_d^{-1}}(\pi) &\approx \gamma_{10} + \gamma_{11}' diag(\overline{\pi})\hat{\pi} \\
J_\epsilon &\approx \gamma_{\epsilon 0} + \gamma_{\epsilon 1}' diag(\overline{\pi})\hat{\pi} \\
(U(z) + \beta \mathbb{E}\left[J_0(\pi_{t+1}) | \mathcal{F}_t \right]) &\approx (U(z) + \beta\gamma_{00})^2 + 2\left(U(z) + \beta\gamma_{00}\right)\beta\gamma_{01}' diag(\overline{\pi})\hat{\pi} \\
\mathbb{E}\left[J_0(\pi_{t+1}) | \mathcal{F}_t \right]^2 &\approx \gamma_{00}^2 + 2\gamma_{00}\gamma_{01}' diag(\overline{\pi})\hat{\pi}.
\end{aligned}
$$

Notice that the restrictions $J_0(\alpha\pi) = J_0(\pi)$, $J_{\theta_d^{-1}}(\alpha\pi) = J_{\theta_d^{-1}}(\pi)$ and $J_\epsilon(\alpha\pi) = J_\epsilon(\pi)$ imply that:

$$
\begin{aligned}
\sum_{j \in a_B^c} \gamma_{01,j}\overline{\pi}_j &= 0 \\
\sum_{j \in a_B^c} \gamma_{11,j}\overline{\pi}_j &= 0 \\
\sum_{j \in a_B^c} \gamma_{\epsilon 1,j}\overline{\pi}_j &= 0.
\end{aligned}
$$

Consider now the updating rule for log deviations from the steady state. Recall that unnormalized probabilities are updated according to:

$$\pi_i^* = f(y^* - \xi_j) \sum_{k=1}^{N} \lambda_{kj} \pi_k,$$

or, equivalently,

$$\frac{\pi_j^*}{\overline{\pi}_j} = f(y^* - \xi_j) \sum_{k \in a_B^c} \lambda_{kj} \frac{\overline{\pi}_k}{\overline{\pi}_j} \frac{\pi_k}{\overline{\pi}_k}.$$

Taking logs of both sides, we obtain the following updating rule for log deviations from the stationary distribution:

$$\hat{\pi}_j^* = \log f(y^* - \xi_j) + \log \left(\sum_{k \in a_B^c} \lambda_{kj} \frac{\overline{\pi}_k}{\overline{\pi}_j} \exp(\hat{\pi}_k) \right).$$

Approximating once again around $\hat{\pi} = \overrightarrow{0}$, we obtain:

$$\hat{\pi}_j^* \approx \log f(y^* - \xi_j) + \log \left(\sum_{k \in a_B^c} \lambda_{kj} \frac{\overline{\pi}_k}{\overline{\pi}_j} \right) + \left(\sum_{k \in a_B^c} \lambda_{kj} \overline{\pi}_k \right)^{-1} \sum_{k \in a_B^c} \lambda_{kj} \overline{\pi}_k \hat{\pi}_k.$$

Denote: $\mathcal{L}_{0j} = \log \left(\sum_{k \in a_B^c} \lambda_{kj} \frac{\overline{\pi}_k}{\overline{\pi}_j} \right)$, $\mathcal{L}_{1,jk} = \left(\sum_{k \in a_B^c} \lambda_{kj} \overline{\pi}_k \right)^{-1} \lambda_{kj}$, so that the first order approximation to the evolution equation is given by:

$$\hat{\pi}_j^* \approx \log f(y^* - \xi_k) + \mathcal{L}_{0j} + \mathcal{L}'_{1j} diag(\overline{\pi}) \hat{\pi}.$$

Notice also that, after observing the default state of firm i, i^*, the predicted vector of conditional distribution is given by:

$$\hat{\pi}_j^* = \log f(y^* - \xi_j) + \log \lambda_{i^*j} - \log \overline{\pi}_j, \qquad \forall j \in a_B^c.$$

Similarly, approximate:

$$p_j = \frac{\pi_j}{\mathbf{1}'_N \pi} \approx \overline{\pi}_j (1 + \hat{\pi}_j - \overline{\pi}' \hat{\pi}).$$

Substituting into the above system and equating coefficients, we obtain the following system:

$$
\begin{aligned}
\gamma_{00} &= \sum_{j\in a_B^c} \overline{\pi}_j \left\{ U(\xi_j) + \beta \sum_{k\in a_B^c} \lambda_{jk} \left(\gamma_{00} + \gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})[\Delta_k^1 + \mathcal{L}_0] \right) \right\} && \text{(C.2)} \\
&\quad + \beta \sum_{j\in a_B^c} \overline{\pi}_j \sum_{i=1}^{I} \lambda_{ji^*} J_0^{\tau_i} \\
\gamma_{01,j} &= U(\xi_j) + \beta \sum_{k\in a_B^c} \lambda_{jk} \left(\gamma_{00} + \gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})[\Delta_k^1 + \mathcal{L}_0] \right) && \text{(C.3)} \\
&\quad + \beta \sum_{i=1}^{I} \lambda_{ji^*} J_0^{\tau_i} - \gamma_{00} + \beta \gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \mathcal{L}_{1j}' \left(\sum_{l,k\in a_B^c} \overline{\pi}_l \lambda_{lk} \right) \\
J_0^{\tau_i} &= \sum_{k\in a_B^c} \lambda_{i^*k} \left(U(\xi_{i^*}) + \beta \left[\gamma_{00} + \gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \left(\Delta_k^1 + \mathcal{L}_{0i^*}^d \right) \right] \right) && \text{(C.4)} \\
&\quad + \sum_{j=1}^{I} \lambda_{i^*j^*} \left[U(\xi_{i^*}) + \beta J_0^{\tau_j} \right]
\end{aligned}
$$

$$
\begin{aligned}
2\gamma_{10} &= \gamma_{00}^2 + 2 \sum_{j,k\in a_B^c} \overline{\pi}_j \lambda_{jk} \beta \left[\gamma_{10} + \gamma_{11}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \left(\Delta_k^1 + \mathcal{L}_0 \right) \right] && \text{(C.5)} \\
&+ \sum_{j,k\in a_B^c} \overline{\pi}_j \lambda_{jk} \beta^2 \left[\gamma_{00}^2 + 2\gamma_{00} \gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \left(\Delta_k^1 + \mathcal{L}_0 \right) \right] \\
&- \sum_{j,k\in a_B^c} \overline{\pi}_j \lambda_{jk} \left(U(\xi_j) + \beta \gamma_{00} \right)^2 \\
&- 2\beta \sum_{j,k\in a_B^c} \overline{\pi}_j \lambda_{jk} \left[U(\xi_j) + \beta \gamma_{00} \right] \gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \left(\Delta_k^1 + \mathcal{L}_0 \right) \\
&\quad + \sum_{j\in a_B^C} \sum_{i=1}^{I} \overline{\pi}_j \lambda_{ji^*} \left[2\beta J_{\theta_d^{-1}}^{\tau_i} + \beta^2 \left(J_0^{\tau_i} \right)^2 - \left(U(\xi_j) + \beta J_0^{\tau_i} \right)^2 \right].
\end{aligned}
$$

$$
\begin{aligned}
2\gamma_{11,j} &= 2\gamma_{00}\gamma_{01,j} + \gamma_{00}^2 - 2\gamma_{10} + \sum_{i=1}^{I} \lambda_{ji^*}\left[2\beta J^{\tau_i}_{\theta_d^{-1}} + \beta^2\left(J_0^{\tau_i}\right)^2 - \left(U(\xi_j) + \beta J_0^{\tau_i}\right)^2\right] \qquad \text{(C.6)}\\
&+ 2\sum_{k\in a_B^c} \lambda_{jk}\beta\left[\gamma_{10} + \gamma_{11}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\left(\Delta_k^1 + \mathcal{L}_0\right)\right]\\
&+ \sum_{k\in a_B^c} \lambda_{jk}\beta^2\left[\gamma_{00}^2 + 2\gamma_{00}\gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\left(\Delta_k^1 + \mathcal{L}_0\right)\right]\\
&- \sum_{k\in a_B^c} \lambda_{jk}\left(U(\xi_j) + \beta\gamma_{00}\right)^2\\
&- 2\beta\sum_{k\in a_B^c} \lambda_{jk}\left[U(\xi_j) + \beta\gamma_{00}\right]\gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\left(\Delta_k^1 + \mathcal{L}_0\right)\\
&+ 2\beta\gamma_{11}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_{1j}' \sum_{l,k\in a_B^c} \overline{\pi}_l\lambda_{lk}\\
&- 2\beta\gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_{1j}' \sum_{l,k\in a_B^c} U(\xi_l)\overline{\pi}_l\lambda_{lk}
\end{aligned}
$$

$$
\begin{aligned}
2J^{\tau_i}_{\theta_d^{-1}} &= \left(J_0^{\tau_i}\right)^2 + 2\sum_{k\in a_B^c} \lambda_{i^*k}\beta\left[\gamma_{10} + \gamma_{11}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\left(\Delta_k^1 + \mathcal{L}_{0i}^d\right)\right] \qquad \text{(C.7)}\\
&+ \sum_{k\in a_B^c} \lambda_{i^*k}\beta^2\left[\gamma_{00}^2 + 2\gamma_{00}\gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\left(\Delta_k^1 + \mathcal{L}_{0i}^d\right)\right]\\
&- \sum_{k\in a_B^c} \lambda_{i^*k}\left(U(\xi_{i^*}) + \beta\gamma_{00}\right)^2\\
&- \beta\sum_{k\in a_B^c} \lambda_{i^*k}\left[U(\xi_{i^*}) + \beta\gamma_{00}\right]\gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\left(\Delta_k^1 + \mathcal{L}_{0i}^d\right)\\
&+ \sum_{j=1}^{I} \lambda_{i^*j^*}\left[2\beta J^{\tau_j}_{\theta_d^{-1}} + \beta^2\left(J_0^{\tau_j}\right)^2 - \left(U(\xi_{i^*}) + \beta J_0^{\tau_j}\right)^2\right]
\end{aligned}
$$

$$
\begin{aligned}
\gamma_{\epsilon 0} &= -\gamma_{00} + \sum_{j \in a_B^c} \sum_{i=1}^{I} \overline{\pi}_j \lambda_{ji^*} \left(U(\xi_j) + \beta J_0^{\tau_i} + \beta J_\epsilon^{\tau_i} \right) && \text{(C.8)} \\
&+ \sum_{j,k \in a_B^c} \overline{\pi}_j \lambda_{jk} \left[U(\xi_j) + \beta \left(\gamma_{00} + \gamma_{\epsilon 0} \right) \right] \\
&+ \beta \sum_{j,k \in a_B^c} \overline{\pi}_j \lambda_{jk} \left(\gamma_{01} + \gamma_{\epsilon 1} \right)' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \left(\Delta_k^1 + \mathcal{L}_0 \right) \\
\gamma_{\epsilon 1, j} &= -\gamma_{01,j} - \gamma_{00} - \gamma_{\epsilon 0} + \beta \left(\gamma_{01} + \gamma_{\epsilon 1} \right)' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \mathcal{L}'_{1j} \sum_{l,k \in a_B^c} \overline{\pi}_l \lambda_{lk} && \text{(C.9)} \\
&+ \sum_{k \in a_B^c} \lambda_{jk} \left[U(\xi_j) + \beta \left(\gamma_{00} + \gamma_{\epsilon 0} \right) \right] \\
&+ \beta \sum_{k \in a_B^c} \lambda_{jk} \left(\gamma_{01} + \gamma_{\epsilon 1} \right)' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \left(\Delta_k^1 + \mathcal{L}_0 \right) \\
&+ \sum_{i=1}^{I} \lambda_{ji^*} \left(U(\xi_j) + \beta J_0^{\tau_i} + \beta J_\epsilon^{\tau_i} \right) \\
J_\epsilon^{\tau_i} &= J_0^{\tau_i} + \sum_{j=1}^{I} \lambda_{i^* j^*} \left(U(\xi_{j^*}) + \beta J_0^{\tau_j} + \beta J_\epsilon^{\tau_j} \right) && \text{(C.10)} \\
&+ \sum_{k \in a_B^c} \lambda_{i^* k} \left[U(\xi_{i^*}) + \beta \left(\gamma_{00} + \gamma_{\epsilon 0} \right) \right] \\
&+ \beta \sum_{k \in a_B^c} \lambda_{i^* k} \left(\gamma_{01} + \gamma_{\epsilon 1} \right)' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \left(\Delta_k^1 + \mathcal{L}_{0i}^d \right),
\end{aligned}
$$

where Δ_k^1 is a constant vector given by:

$$
\Delta_{kj}^1 = -\frac{1}{2} - \frac{1}{2} (\xi_j - \xi_k)' \Sigma_u^{-1} (\xi_j - \xi_k).
$$

Consider now the distortion to the conditional joint distribution of next period's signals and state. Recall that, in terms of the value function, this is given by:

$$
\phi_t(z^*, y^*) = \frac{\exp\left(-\frac{\beta J(\pi^*; \theta_d^{-1})}{\theta} \right)}{\mathbb{E}\left[\exp\left(-\frac{\beta J(\pi^*; \theta_d^{-1})}{\theta_d} \right) \middle| \mathcal{F}_t \right]}.
$$

Taking a first order approximation around the point $\theta_d^{-1} = \theta_s^{-1} = 0$, we obtain:

$$
\begin{aligned}
\phi_t\left(z_{t+1}, y_{t+1} \right) \approx\ & 1 - \theta_d^{-1} \beta \left(J_0(\pi_{t+1}) - \mathbb{E}\left[J_0(\pi_{t+1}) \middle| \mathcal{F}_t \right] \right) \\
& - \left(\theta_s^{-1} - \theta_d^{-1} \right) \beta \left(J_\epsilon(\pi_{t+1}) - \mathbb{E}\left[J_\epsilon(\pi_{t+1}) \middle| \mathcal{F}_t \right] \right).
\end{aligned}
$$

Substituting for J_0 and J_ϵ, we obtain:

$$\begin{aligned}\phi(z^*, y^*|z=\xi_j) = 1 &+ \theta_d^{-1}\left(\varphi_{01,j} + \varphi'_{\pi 1,j} diag(\overline{\pi})\hat{\pi} + \varphi'_{y1,j}\log f(y^*)\right)\\ &+ (\theta_s^{-1} - \theta_d^{-1})\left(\varphi_{0\epsilon,j} + \varphi'_{\pi\epsilon,j} diag(\overline{\pi})\hat{\pi} + \varphi'_{y\epsilon,j}\log f(y^*)\right),\end{aligned}$$

where:

$$\varphi_{01,jk} = \begin{cases} \beta\left(\sum_{i=1}^{I}\lambda_{ji^*}J_0^{\tau_i} - \gamma_{00} - \gamma'_{01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_0\right) & \\ +\beta\sum_{l\in a_B^c}\lambda_{jl}\left[\gamma_{00} + \gamma'_{01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\Delta_k^1 + \mathcal{L}_0)\right] & j,k\in a_B^c \\ \beta\left(\sum_{i=1}^{I}\lambda_{ji^*}J_0^{\tau_i} - J_0^{\tau_k}\right) + & \\ +\beta\sum_{l\in a_B^c}\lambda_{jl}\left[\gamma_{00} + \gamma'_{01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\Delta_k^1 + \mathcal{L}_0)\right] & j\in a_B^c, k\in a_B \\ \beta\left(\sum_{i=1}^{I}\lambda_{j^*i^*}J_0^{\tau_i} - \gamma_{00} - \gamma'_{01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_{0j}^d\right) & \\ +\beta\sum_{l\in a_B^c}\lambda_{j^*l}\left[\gamma_{00} + \gamma'_{01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_{0j}^d\right] & j\in a_B, k\in a_B^c \\ \beta\left(\sum_{i=1}^{I}\lambda_{j^*i^*}J_0^{\tau_i} - \gamma_{00} - \gamma'_{01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_{0j}^d\right) & \\ +\beta\sum_{l\in a_B^c}\lambda_{j^*l}\left[\gamma_{00} + \gamma'_{01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_{0j}^d\right] & j,k\in a_B \end{cases} \tag{C.11}$$

$$\varphi_{\pi 1,jk} = \begin{cases} \beta\gamma'_{01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_1\left(\sum_{l\in a_B^c}\lambda_{jl} - 1\right) & j,k\in a_B^c \\ \beta\mathcal{L}'_1 diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\gamma_{01}\sum_{k\in a_B^c}\lambda_{jk} & j\in a_B^c, k\in a_B \\ 0 & j\in a_B \end{cases} \tag{C.12}$$

$$\varphi_{y1,jk} = \begin{cases} -\beta\gamma'_{01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) & j,k\in a_B^c \\ -\beta diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\gamma_{01} & j\in a_B, k\in a_B^c \\ 0 & k\in a_B \end{cases} \tag{C.13}$$

$$\varphi_{0\epsilon,jk} = \begin{cases} \beta\left(\sum_{i=1}^{I}\lambda_{ji^*}J_\epsilon^{\tau_i} - \gamma_{\epsilon 0} - \gamma'_{\epsilon 1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_0\right) & \\ +\beta\sum_{l\in a_B^c}\lambda_{jl}\left[\gamma_{\epsilon 0} + \gamma'_{\epsilon 1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\Delta_k^1 + \mathcal{L}_0)\right] & j,k\in a_B^c \\ \beta\left(\sum_{i=1}^{I}\lambda_{ji^*}J_\epsilon^{\tau_i} - J_\epsilon^{\tau_k}\right) + & \\ +\beta\sum_{l\in a_B^c}\lambda_{jl}\left[\gamma_{\epsilon 0} + \gamma'_{\epsilon 1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\Delta_k^1 + \mathcal{L}_0)\right] & j\in a_B^c, k\in a_B \\ \beta\left(\sum_{i=1}^{I}\lambda_{j^*i^*}J_\epsilon^{\tau_i} - \gamma_{\epsilon 0} - \gamma'_{\epsilon 1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_{0j}^d\right) & \\ +\beta\sum_{l\in a_B^c}\lambda_{j^*l}\left[\gamma_{\epsilon 0} + \gamma'_{\epsilon 1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_{0j}^d\right] & j\in a_B, k\in a_B^c \\ \beta\left(\sum_{i=1}^{I}\lambda_{j^*i^*}J_\epsilon^{\tau_i} - \gamma_{\epsilon 0} - \gamma'_{\epsilon 1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_{0j}^d\right) & \\ +\beta\sum_{l\in a_B^c}\lambda_{j^*l}\left[\gamma_{\epsilon 0} + \gamma'_{\epsilon 1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_{0j}^d\right] & j,k\in a_B \end{cases} \tag{C.14}$$

$$\varphi_{\pi\epsilon,jk} = \begin{cases} \beta\gamma_{\epsilon 1}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_1 \left(\sum_{l\in a_B^c}\lambda_{jl} - 1\right) & j,k \in a_B^c \\ \beta\mathcal{L}_1' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\gamma_{\epsilon 1}\sum_{k\in a_B^c}\lambda_{jk} & j\in a_B^c, k\in a_B \\ 0 & j \in a_B \end{cases} \tag{C.15}$$

$$\varphi_{y\epsilon,jk} = \begin{cases} -\beta\gamma_{\epsilon 1}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) & j,k \in a_B^c \\ -\beta diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\gamma_{\epsilon 1} & j\in a_B, k\in a_B^c \\ 0 & k \in a_B \end{cases} \tag{C.16}$$

Turn now to the distortion to the current period's conditional probability vector. Recall that this is given by:

$$\psi_t(z) = \frac{\exp\left(-\frac{U(z)+\mathcal{R}_t(\beta J(p^*);\theta_d)}{\theta_s}\right)}{\mathbb{E}\left[\exp\left(-\frac{U(z)+\mathcal{R}_t(\beta J(p^*);\theta_d)}{\theta_s}\right)\Big|\,\mathcal{G}_t\right]}.$$

Approximating once again around the point $\theta_d^{-1} = \theta_s^{-1} = 0$, we obtain:

$$\begin{aligned}\psi_t(z_t) \approx 1 &- \left(\theta_d^{-1} - \theta_s^{-1}\right)\beta\left(\mathbb{E}\left[J_\epsilon(\pi_{t+1})|\,\mathcal{F}_t\right] - \mathbb{E}\left[J_\epsilon(\pi_{t+1})|\,\mathcal{G}_t\right]\right) \\ &- \theta_s^{-1}\left(U(z_t) + \beta\mathbb{E}\left[J_0(\pi_{t+1})|\,\mathcal{F}_t\right] - \mathbb{E}\left[U(z_t) + \beta J_0(\pi_{t+1})|\,\mathcal{G}_t\right]\right).\end{aligned}$$

Recall that, in terms of log-deviations from the steady state, $p_l \approx \overline{\pi}_l\left(1 + \hat{\pi}_l - \overline{\pi}'\hat{\pi}\right)$. We can represent:

$$\psi(\xi_j) = 1 + \theta_s^{-1}\left(\zeta_{j,01} + \zeta_{j,11}' diag(\overline{\pi})\hat{\pi}\right) + \left(\theta_d^{-1} - \theta_s^{-1}\right)\left(\zeta_{j,0\epsilon} + \zeta_{j,1\epsilon}' diag(\overline{\pi})\hat{\pi}\right),$$

where:

$$\begin{aligned}\zeta_{j,01} = &\sum_{l\in a_B^c}\overline{\pi}_l\left\{U(\xi_l) + \beta\sum_{k\in a_B^c}\lambda_{lk}\left[\gamma_{00} + \gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\Delta_k^1 + \mathcal{L}_0)\right]\right\} \\ &+ \beta\sum_{l\in a_B^c}\overline{\pi}_l\sum_{i=1}^{I}\lambda_{li^*}J_0^{\tau_k} - U(\xi_l) \\ &- \beta\sum_{k\in a_B^c}\lambda_{lk}\left[\gamma_{00} + \gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\Delta_k^1 + \mathcal{L}_0)\right] - \beta\sum_{i=1}^{I}\lambda_{li^*}J_0^{\tau_k}\end{aligned} \tag{C.17}$$

$$\zeta_{j,11,k} = -\zeta_{k,00} + \beta\gamma_{01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\mathcal{L}_{1k}'\left(\sum_{l,m\in a_B^c}\overline{\pi}_l\lambda_{lm} - \sum_{l\in a_B^c}\lambda_{kl}\right) \tag{C.18}$$

$$\zeta_{j,0\epsilon} = \beta \sum_{l,k\in a_B^c} \overline{\pi}_l \lambda_{lk} \left[\gamma_{\epsilon 0} + \gamma_{\epsilon 1}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\Delta_k^1 + \mathcal{L}_0)\right] + \beta \sum_{l\in a_B^c} \overline{\pi}_l \sum_{i=1}^{I} \lambda_{li^*} J_\epsilon^{\tau_k} \quad \text{(C.19)}$$

$$- \beta \sum_{k\in a_B^c} \lambda_{lk} \left[\gamma_{\epsilon 0} + \gamma_{\epsilon 1}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\Delta_k^1 + \mathcal{L}_0)\right] - \beta \sum_{i=1}^{I} \lambda_{li^*} J_\epsilon^{\tau_k}$$

$$\zeta_{j,1\epsilon,k} = -\zeta_{k,0\epsilon} + \beta \gamma_{\epsilon 1}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \mathcal{L}_{1k}' \left(\sum_{l,m\in a_B^c} \overline{\pi}_l \lambda_{lm} - \sum_{l\in a_B^c} \lambda_{kl} \right). \quad \text{(C.20)}$$

C.2 Proof of Lemma 4.3

Notice first that, substituting the first order expansion to the value of equity of firm i into the worst-case Euler equation and equating coefficients on the powers of θ^{-1}, we obtain the following system for V_{i0} and V_{i1}:

$$V_{i0}(\pi) = -C_i + \sum_{j\in a_B^c} p_j \left[\delta_i A_{ij} + \beta \sum_{k\in a_B^c} \lambda_{jk} \left|2\pi\Sigma_u\right|^{-\frac{1}{2}} \int V_{i0}(\pi^*) df(y^* - \xi_k) + \beta \sum_{k\neq i} \lambda_{jk^*} V_{i0}^{\tau_k} \right]$$

$$V_{i0}^{\tau_j} = -C_i + \delta_i A_{ij^*} + \beta \left[\sum_{k\in a_B^c} \lambda_{j^*k} \left|2\pi\Sigma_u\right|^{-\frac{1}{2}} \int V_{i0}(\pi^*) df(y^* - \xi_k) + \sum_{k\neq i} \lambda_{j^*k^*} V_{i0}^{\tau_k} \right]$$

$$\begin{aligned}
V_{i\theta_d^{-1}} = & \sum_{j\in a_B^c} p_j \left[\zeta_{01,j} + \zeta_{11,j}' diag(\overline{\pi})\hat{\pi}\right] \left[\delta_i A_{ij} + \beta \sum_{k\neq i} \lambda_{jk^*} V_{i0}^{\tau_k} \right] \\
& + \beta \sum_{j,k\in a_B^c} p_j \left[\zeta_{01,j} + \zeta_{11,j}' diag(\overline{\pi})\hat{\pi}\right] \lambda_{jk} \left|2\pi\Sigma_u\right|^{-\frac{1}{2}} \int V_{i0}(\pi^*) df(y^* - \xi_k) \\
& + \beta \sum_{j,k\in a_B^c} p_j \lambda_{jk} \left|2\pi\Sigma_u\right|^{-\frac{1}{2}} \int V_{i\theta_d^{-1}}(\pi^*) df(y^* - \xi_k) \\
& + \beta \sum_{j,k\in a_B^c} p_j \lambda_{jk} \left|2\pi\Sigma_u\right|^{-\frac{1}{2}} \int V_{i0}(\pi^*) \left(\varphi_{01,jk} + \varphi_{\pi 1,jk}' diag(\overline{\pi})\hat{\pi} + \varphi_{y1,jk} \log f(y^* - \xi_k)\right) df(y^* - \xi_k) \\
& + \beta \sum_{j\in a_B^c} p_j \sum_{k\neq i} \lambda_{jk^*} \left[V_{i\theta_d^{-1}}^{\tau_k} + V_{i0}^{\tau_k} \left(\varphi_{01,jk^*} + \varphi_{\pi 1,jk}' diag(\overline{\pi})\hat{\pi}\right) \right]
\end{aligned}$$

$$V^{\tau_j}_{i\theta_d^{-1}} = \beta \sum_{k \in a_B^c} \lambda_{j^*k} \left|2\pi\Sigma_u\right|^{-\frac{1}{2}} \int \left[V_{i\theta_d^{-1}}(\pi^*) + V_{i0}(\pi^*)\left(\varphi_{01,j^*k} + \varphi_{y1,j^*k} \log f(y^* - \xi_k)\right)\right] df(y^* - \xi_k)$$
$$+ \beta \sum_{k \neq i} \lambda_{j^*k^*} \left[V^{\tau_k}_{i\theta_d^{-1}} + V^{\tau_k}_{i0} \varphi_{01,j^*k^*}\right]$$

$$V_{i\epsilon} = \beta \sum_{j \in a_B^c} p_j \left[\sum_{k \in a_B^c} \lambda_{jk} \left|2\pi\Sigma_u\right|^{-\frac{1}{2}} \int V_{i\epsilon}(\pi^*) df(y^* - \xi_k) + \sum_{k \neq i} \lambda_{jk^*} V^{\tau_k}_{i\epsilon}\right]$$
$$V^{\tau_j}_{i\epsilon} = \beta \sum_{k \in a_B^c} \lambda_{j^*k} \left|2\pi\Sigma_u\right|^{-\frac{1}{2}} \int V_{i\epsilon}(\pi^*) df(y^* - \xi_k) + \beta \sum_{k \neq i} \lambda_{j^*k^*} V^{\tau_k}_{i\epsilon},$$

where A_{ij} is the fundamental value of asset of firm i in state j. Substituting the approximations (4.23)-(4.25) and equating coefficients, we obtain:

$$\nu_{i,00} = -C_i + \sum_{j \in a_B^c} \overline{\pi}_j \left(\delta_i A_{ij} + \beta \sum_{k \in a_B^c} \lambda_{jk} \left[\nu_{i,00} + \nu'_{i,01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\mathcal{L}_0 + \Delta^1_k)\right] + \beta \sum_{k \neq i} \lambda_{jk^*} V^{\tau_k}_{i0}\right)$$

$$\nu_{i,01,j} = \delta_i A_{ij} + \beta \sum_{k \in a_B^c} \lambda_{jk} \left[\nu_{i,00} + \nu'_{i,01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\mathcal{L}_0 + \Delta^1_k)\right] + \beta \sum_{k \neq i} \lambda_{jk^*} V^{\tau_k}_{i0}$$
$$- \nu_{i,00} + \beta \sum_{k \neq i} \lambda_{jk^*} V^{\tau_k}_{i0} + \beta \nu'_{i,01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \mathcal{L}'_{1j} \sum_{l,k \in a_B^c} \overline{\pi}_l \lambda_{lk}$$

$$V^{\tau_j}_{i0} = -C_i + \delta_i A_{ij^*} + \beta \sum_{k \neq i} \lambda_{j^*k^*} V^{\tau_k}_{i0} + \beta \sum_{k \in a_B^c} \lambda_{j^*k} \left[\nu_{i,00} + \nu'_{i,01} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\left(\Delta^1_k + \mathcal{L}^d_{0j}\right)\right]$$

$$\nu_{i,\epsilon 0} = \beta \sum_{j \in a_B^c} \overline{\pi}_j \left(\sum_{k \in a_B^c} \lambda_{jk} \left[\nu_{i,\epsilon 0} + \nu'_{i,\epsilon 1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\mathcal{L}_0 + \Delta^1_k)\right] + \sum_{k \neq i} \lambda_{jk^*} V^{\tau_k}_{i\epsilon}\right)$$

$$\nu_{i,01,j} = \beta \sum_{k \in a_B^c} \lambda_{jk} \left[\nu_{i,\epsilon 0} + \nu'_{i,\epsilon 1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\mathcal{L}_0 + \Delta^1_k)\right] + \beta \sum_{k \neq i} \lambda_{jk^*} V^{\tau_k}_{i\epsilon}$$
$$- \nu_{i,\epsilon 0} + \beta \sum_{k \neq i} \lambda_{jk^*} V^{\tau_k}_{i\epsilon} + \beta \nu'_{i,\epsilon 1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \mathcal{L}'_{1j} \sum_{l,k \in a_B^c} \overline{\pi}_l \lambda_{lk}$$

$$V_{i\epsilon}^{\tau_j} = \beta \sum_{k \neq i} \lambda_{j^*k^*} V_{i\epsilon}^{\tau_k} + \beta \sum_{k \in a_B^c} \lambda_{j^*k} \left[\nu_{i,\epsilon 0} + \nu_{i,\epsilon 1}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \left(\Delta_k^1 + \mathcal{L}_{0j}^d \right) \right]$$

$$\begin{aligned}
\nu_{i,10} = & \sum_{j \in a_B^c} \overline{\pi}_j \zeta_{j0} \left\{ \delta_i A_{ij} + \beta \sum_{k \neq i} \lambda_{jk^*} V_{i0}^{\tau_k} + \beta \sum_{k \in a_B^c} \lambda_{jk} \left[\nu_{i,00} + \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) (\mathcal{L}_0 + \Delta_k^1) \right] \right\} \\
& + \beta \sum_{j,k \in a_B^c} \overline{\pi}_j \lambda_{jk} \left[\nu_{i,10} + \nu_{i,11}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) (\mathcal{L}_0 + \Delta_k^1) \right] \\
& + \beta \sum_{j,k \in a_B^c} \overline{\pi}_j \lambda_{jk} \left[\nu_{i,00} + \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) (\mathcal{L}_0 + \Delta_k^1) \right] \varphi_{01,jk} \\
& + \beta \sum_{j,k \in a_B^c} \overline{\pi}_j \lambda_{jk} \left[\nu_{i,00} \varphi_{y1,jk}' \Delta_k^1 + \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \Delta_k^2 \varphi_{y1,jk} \right] \\
& + \beta \sum_{j \in a_B^c} \overline{\pi}_j \sum_{k \neq i} \lambda_{jk^*} \left[V_{i0}^{\tau_k} \varphi_{01,jk^*} + V_{i\theta_d^{-1}}^{\tau_k} \right]
\end{aligned}$$

$$
\begin{aligned}
\nu_{i,11,j} = \zeta_{j0} &\left\{ \delta_i A_{ij} + \beta \sum_{k \neq i} \lambda_{jk^*} V_{i0}^{\tau_k} + \beta \sum_{k \in a_B^c} \lambda_{jk} \left[\nu_{i,00} + \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\mathcal{L}_0 + \Delta_k^1) \right] \right\} \\
&+ \beta \sum_{k \in a_B^c} \lambda_{jk} \left[\nu_{i,10} + \nu_{i,11}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\mathcal{L}_0 + \Delta_k^1) \right] \\
&+ \beta \sum_{k \in a_B^c} \lambda_{jk} \left[\nu_{i,00} + \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\mathcal{L}_0 + \Delta_k^1) \right] \varphi_{01,jk} - \nu_{i,10} \\
&+ \beta \sum_{k \in a_B^c} \lambda_{jk} \left[\nu_{i,00} \varphi_{y1,jk}' \Delta_k^1 + \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \Delta_k^2 \varphi_{y1,jk} \right] + \beta \sum_{k \neq i} \lambda_{jk^*} \left[V_{i0}^{\tau_k} \varphi_{01,jk^*} + V_{i\theta_d^{-1}}^{\tau_k} \right] \\
&+ \sum_{l \in a_B^c} \zeta_{l1,j} \left[\delta_i A_{il} + \beta \sum_{k \neq i} \lambda_{lk^*} V_{i0}^{\tau_k} + \beta \sum_{k \in a_B^c} \lambda_{lk} \left(\nu_{i,00} + \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\mathcal{L}_0 + \Delta_k^1) \right) \right] \\
&+ \beta \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \mathcal{L}_{1j}' \sum_{l,k \in a_B^c} \overline{\pi}_l \zeta_{l0} \lambda_{lk} + \beta \nu_{i,11}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \mathcal{L}_{1j}' \sum_{l,k \in a_B^c} \overline{\pi}_l \zeta_{l0} \lambda_{lk} \\
&+ \beta \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \mathcal{L}_{1j}' \sum_{l,k \in a_B^c} \overline{\pi}_l \lambda_{lk} \left[\varphi_{01,lk} + \varphi_{y1,lk}' \Delta_k^1 \right] + \beta \sum_{l \in a_B^c} \sum_{k \neq i} \overline{\pi}_l \lambda_{lk^*} V_{i0}^{\tau_k} \varphi_{\pi 1,lk^*,j} \\
&+ \beta \sum_{l,k \in a_B^c} \overline{\pi}_l \lambda_{lk} \left[\nu_{i,00} + \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\mathcal{L}_0 + \Delta_k^1) \right] \varphi_{\pi 1,lk,j} \\
V_{i\theta_d^{-1}}^{\tau_j} = \beta &\sum_{k \in a_B^c} \lambda_{j^*k} \left[\nu_{i,00} + \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\mathcal{L}_{0j}^d + \Delta_k^1) \right] \varphi_{01,j^*k} \\
&+ \beta \sum_{k \in a_B^c} \lambda_{j^*k} \left[\nu_{i,00} \varphi_{y1,j^*k}' \Delta_k^1 + \nu_{i,01}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \Delta_k^2 \varphi_{y1,j^*k} \right] \\
&+ \beta \sum_{k \in a_B^c} \lambda_{j^*k} \left[\nu_{i,10} + \nu_{i,11}' diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})(\mathcal{L}_{0j}^d + \Delta_k^1) \right] + \beta \sum_{k \neq i} \lambda_{j^*k^*} \left(V_{i0}^{\tau_k} \phi_1(\xi_{k^*} | \xi_{j^*}) + V_{i\theta_d^{-1}}^{\tau_k} \right)
\end{aligned}
$$

where Δ_k^2 is a constant matrix given by:

$$
\begin{aligned}
\Delta_{k,jl}^2 &\equiv |2\pi\Sigma_u|^{-\frac{1}{2}} \int \left(\frac{1}{2} (y^* - \xi_j - \overline{u})' \Sigma_u^{-1} (y^* - \xi_j - \overline{u}) \right) \left(\frac{1}{2} (y^* - \xi_j - \overline{u})' \Sigma_u^{-1} (y^* - \xi_j - \overline{u}) \right) df(y^* - \xi_k) \\
&= \frac{1}{4} \left[(\xi_k - \xi_j)' \Sigma_u^{-1} (\xi_k - \xi_j) \right] \left[(\xi_k - \xi_l)' \Sigma_u^{-1} (\xi_k - \xi_l) \right] + \frac{1}{4} (\xi_k - \xi_j)' \Sigma_u^{-1} (\xi_k - \xi_j) \\
&+ \frac{1}{4} (\xi_k - \xi_l)' \Sigma_u^{-1} (\xi_k - \xi_l) + (\xi_k - \xi_l)' \Sigma_u^{-1} (\xi_k - \xi_j) + \frac{5}{4}.
\end{aligned}
$$

D Risk aversion benchmark

In this section, I investigate the performance of a power utility model in explaining the time series evolution of credit spreads and equity prices. Instead of solving the portfolio allocation problem of the representative risk-averse agent, I take the stochastic discount factor as given. In particular, let $S_{t,t+s}$ be the stochastic discount factor between dates t and $t+s$. In an economy where the representative agent evaluates consumption paths using a power utility function:

$$u(\mathcal{K}_t) = \frac{\mathcal{K}_t^{1-\gamma}}{1-\gamma},$$

where $\mathcal{K}_t \equiv \mathcal{K}(z_t) = \sum_{i=1}^{I} \delta_i A_i(z_t)$ is the level of consumption at date t and $\gamma > 0$ is the degree of risk aversion, the stochastic discount factor is given by:

$$S_{t,t+s} = \beta^s \frac{\mathcal{K}_{t+s}^{-\gamma}}{\mathcal{K}_t^{-\gamma}}. \tag{D.1}$$

Consider first the CDS spread on a swap with maturity $T = t+6n$ on the consol bond of firm i. In the economy with the risk-averse representative agent, this is given by:

$$c_i(t,T) = \frac{2X_i \mathbb{E}\left[\beta^{\tau_i - t}\mathcal{K}_{\tau_i}^{-\gamma}\mathbf{1}_{\tau_i<T}\middle|\,\mathcal{G}_t\right]}{\sum_{s=1}^{n}\beta^{6s}\mathbb{E}\left[\mathcal{K}_{t+6s}^{-\gamma}\mathbf{1}_{\tau_i\geq t+6s}\middle|\,\mathcal{G}_t\right]}.$$

Notice that:

$$\begin{aligned}
\mathbb{E}\left[\beta^{\tau_i - t}\mathcal{K}_{\tau_i}^{-\gamma}\mathbf{1}_{\tau_i<T}\middle|\,\mathcal{G}_t\right] &= \sum_{s=1}^{T-t}\beta^s \mathbb{E}\left[\mathbf{1}_{\tau_i=t+s}\mathcal{K}_{t+s}^{-\gamma}\middle|\,\mathcal{G}_t\right] \\
\mathbb{E}\left[\mathcal{K}_{t+6s}^{-\gamma}\mathbf{1}_{\tau_i\geq t+6s}\middle|\,\mathcal{G}_t\right] &= \mathbb{E}\left[\mathcal{K}_{t+6s}^{-\gamma}\middle|\,\mathcal{G}_t\right] - \sum_{k=1}^{6s}\mathbb{E}\left[\mathcal{K}_{t+6s}^{-\gamma}\mathbf{1}_{\tau_i=t+k}\middle|\,\mathcal{G}_t\right].
\end{aligned}$$

Let $q_{ij}^{(n)} = \mathbb{P}\left(z_{t+n}=\xi_j,\ z_{t+n-1}\neq\xi_j,\ldots,\ z_{t+1}\neq\xi_j\middle|\, z_t=\xi_i\right)$ be the probability that $t+n$ is the first hitting time of ξ_j conditional on being in state i at date t. Then:

$$\begin{aligned}
\mathbb{E}\left[\mathcal{K}_{t+s}^{-\gamma}\mathbf{1}_{\tau_i=t+s}\middle|\,\mathcal{G}_t\right] &= \sum_{j=1}^{N} p_{jt} q_{ji^*}^{(s)}\mathcal{K}(a_{B_i})^{-\gamma} \\
\mathbb{E}\left[\mathcal{K}_{t+s}^{-\gamma}\middle|\,\mathcal{G}_t\right] &= \sum_{j\in a_B^c} p_{jt}\sum_{k=1}^{N}\left\{\Lambda^s\right\}_{jk}\mathcal{K}(\xi_k)^{-\gamma} \\
\mathbb{E}\left[\mathcal{K}_{t+s_1}^{-\gamma}\mathbf{1}_{\tau_i=t+s_2}\middle|\,\mathcal{G}_t\right] &= \sum_{j\in a_B^c} p_{jt} q_{ji^*}\sum_{k=1}^{N}\left\{\Lambda^{s_1-s_2}\right\}_{jk}\mathcal{K}(\xi_k)^{-\gamma}.
\end{aligned}$$

Table VIII presents the 5 year CDS spreads on the financial institutions for different levels of risk aversion at four of interest – before the start of the crisis, July 2007, at the start of the crisis in August 2007, after the bailout of Bear Stearns in March 2008 and after the liquidation of Lehman Brothers in September 2008 – together with the observations of the 5 year CDS spreads at these dates. Notice first that while the model-implied credit spreads do increase during the crisis, the magnitude of the model-implied spreads remains much smaller than that of the observed CDS spreads. The only exception is JP Morgan, the model-implied spreads for which exceed the observed ones during the crisis. Notice also, that the dependence on the degree of risk aversion is not monotone. In particular, although the spreads increase initially as the degree of risk aversion increases, further increases in risk aversion decrease the model-implied CDS spreads.

(Table VIII about here.)

Consider now the date t price of a claim to the equity of firm i. Using the stochastic discount factor of the risk-averse agent, the equity price satisfies the Euler equation:

$$V_{it} = \mathbb{E}\left[\delta_i A_{it} - C_i + S_{t,t+1}\mathbf{1}_{\tau_i > t+1} V_{i,t+1} \middle| \mathcal{G}_t\right]. \tag{D.2}$$

Similarly to the case with model misspecification, I look for a first order approximation to the equity price in terms of log deviations from the steady state:

$$V_{ir}(\pi_t) = \nu_{ir,0} + \nu'_{ir,1} diag(\overline{\pi})\hat{\pi}_t + O_2(\hat{\pi}_t).$$

Substituting into the Euler equation and equation coefficients, we obtain:

$$\begin{aligned}
\nu_{ir,0} &= \sum_{j\in a_B^c} \overline{\pi}_j \left\{ \delta_i A_{ij} - C_i + \beta \sum_{k\in a_B^c} \frac{\mathcal{K}(\xi_k)^{-\gamma}}{\mathcal{K}(\xi_j)^{-\gamma}} \lambda_{jk} \left[\nu_{ir,0} + \nu'_{ir,1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}\left(\mathcal{L}_0 + \Delta_k^1\right)\right] \right\} \\
&\quad + \beta \sum_{j\in a_B^c} \overline{\pi}_j \sum_{k\neq i} \lambda_{jk^*} \frac{\mathcal{K}(\xi_{k^*})^{-\gamma}}{\mathcal{K}(\xi_j)^{-\gamma}} \\
\nu_{ir,1j} &= \delta_i A_{ij} - C_i + \beta \sum_{k\in a_B^c} \frac{\mathcal{K}(\xi_k)^{-\gamma}}{\mathcal{K}(\xi_j)^{-\gamma}} \lambda_{jk} \left[\nu_{ir,0} + \nu'_{ir,1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}\left(\mathcal{L}_0 + \Delta_k^1\right)\right] \\
&\quad + \beta \sum_{k\neq i} \lambda_{jk^*} \frac{\mathcal{K}(\xi_{k^*})^{-\gamma}}{\mathcal{K}(\xi_j)^{-\gamma}} - \nu_{ir,0} + \beta \nu'_{ir,1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c}) \mathcal{L}'_{1j} \sum_{l,k\in a_B^c} \overline{\pi}_l \lambda_{lk} \frac{\mathcal{K}(\xi_k)^{-\gamma}}{\mathcal{K}(\xi_l)^{-\gamma}}.
\end{aligned}$$

Similarly, at the time of default of firm j, the equity price of firm i solves:

$$\begin{aligned}
V_{ir}^{\tau_j} &= \delta_i A_{ij^*} - C_i + \beta \sum_{k\in a_B^c} \frac{\mathcal{K}(\xi_k)^{-\gamma}}{\mathcal{K}(\xi_{j^*})^{-\gamma}} \lambda_{j^*k} \left[\nu_{ir,0} + \nu'_{ir,1} diag(\overline{\pi}) diag(\mathbf{1}_{a_B^c})\left(\mathcal{L}_0 + \mathcal{L}_1 diag(\overline{\pi})\hat{\pi} + \Delta_k^1\right)\right] \\
&\quad + \beta \sum_{k\neq i} \lambda_{j^*k^*} \frac{\mathcal{K}(\xi_{k^*})^{-\gamma}}{\mathcal{K}(\xi_{j^*})^{-\gamma}} V_{ir}^{\tau_k}.
\end{aligned}$$

	Mean	St. Dev.	Correlations										
			BAC	BSC	Citi	GS	JPM	MER	MS	WB	WFC	CS10	SP500
BAC	56038.09	65351.99	100.00									76.99	61.27
BSC	4508.48	3687.57	98.08	100.00								95.85	83.32
Citi	53210.59	51820.79	91.39	56.58	100.00							91.01	75.66
GS	32395.77	18845.19	96.01	-20.99	93.12	100.00						48.25	-10.57
JPM	51611.38	57647.31	79.25	33.03	74.92	78.50	100.00					67.35	48.97
MER	15995.35	13145.67	88.73	73.51	97.45	92.10	71.71	100.00				95.81	76.41
MS	16683.24	14747.40	95.92	43.47	97.14	96.66	76.59	78.11	100.00			83.31	73.77
WB	21541.91	22530.87	96.66	92.99	93.97	96.84	71.40	95.30	93.99	100.00		93.78	75.48
WFC	23943.94	28543.76	96.35	19.57	88.91	94.29	73.87	63.31	94.59	31.24	100.00	68.15	55.87

Table I: Sample moments of and empirical correlations between the book equity of financial institutions and aggregate market indicators. CS10 is the Case-Shiller 10 index; SP500 is the Standard and Poor's 500 index. Data Source: Compustat, CRSP

	BAC	BSC	Citi	GS	JPM	MER	MS	WB	WFC
δ	3.67	3.41	3.97	4.46	9.67	3.39	4.69	3.13	3.13
D	1398608	381237.992	2093112	891128	1338831	1034133	1160482	647525	492564
X	47.61	70.20	67.91	98.77	75.05	66.96	65.11	45.07	46.35
C	530.96	120.85	767.83	669.21	496.40	305.24	404.80	225.59	202.27

Table II: Parameters estimated outside the MCMC procedure of Appendix A. δ, the fraction of assets generated as payoffs, and X, the payment in case of default, are reported in percentage terms; D is the face value of the consol bond, taken to be the last available observation of long-term debt; C is the monthly coupon payment on the consol bond.

ζ_f

BAC	BSC	Citi	GS	JPM	MER	MS	WB	WFC
7.40	7.98	9.66	5.54	10.00	9.21	9.11	9.37	9.13
9.83	6.92	9.58	5.29	9.99	9.10	9.00	9.28	9.02
9.75	7.93	6.91	5.19	9.92	9.08	8.98	9.27	8.99
9.74	7.91	9.55	0.13	9.91	9.04	8.97	9.25	8.95
9.86	7.99	9.71	5.57	8.00	9.01	8.91	9.17	8.86
9.98	8.07	9.90	5.99	10.11	8.02	8.87	9.16	8.85
9.89	7.99	9.73	5.66	10.03	9.12	6.82	9.14	8.85
9.73	7.91	9.54	5.19	9.91	8.98	8.88	7.49	8.84
9.75	7.92	9.59	5.21	9.95	9.01	8.90	9.19	7.05
9.73	7.88	9.52	5.05	9.90	8.99	8.87	9.14	8.84
9.74	7.94	9.58	5.28	9.91	9.04	8.91	9.16	8.88
9.86	8.01	9.73	5.66	10.02	9.12	8.99	9.27	8.98
9.69	7.89	9.49	5.10	9.90	8.98	8.85	9.10	8.80
9.71	7.89	9.54	5.16	9.88	9.02	8.87	9.13	8.85
9.71	7.91	9.55	5.14	9.90	9.01	8.87	9.13	8.83
9.78	7.94	9.60	5.37	9.95	9.03	8.93	9.18	8.89
9.75	7.92	9.57	5.30	9.93	9.03	8.92	9.18	8.87
9.83	7.97	9.67	5.47	10.00	9.09	8.98	9.23	8.96

ζ_c

4.06	4.82

ρ

BAC	BSC	Citi	GS	JPM	MER	MS	WB	WFC
0.20	0.00	0.09	0.01	0.27	0.09	0.09	0.06	0.01

Ω_c

99.32	0.68
0.54	99.46

Ω_f

54.90	2.74	2.63	2.56	2.71	2.68	2.60	2.58	2.66	2.70	2.53	2.63	2.73	2.64	2.63	2.63	2.70	2.75
2.58	55.04	2.61	2.68	2.69	2.73	2.66	2.66	2.65	2.66	2.59	2.57	2.62	2.64	2.69	2.59	2.74	2.61
2.63	2.65	55.12	2.65	2.60	2.59	2.72	2.66	2.71	2.68	2.67	2.65	2.59	2.69	2.57	2.57	2.63	2.62
2.63	2.65	2.61	55.08	2.63	2.75	2.53	2.63	2.68	2.62	2.65	2.65	2.63	2.69	2.67	2.62	2.69	2.62
2.62	2.64	2.57	2.73	54.93	2.60	2.73	2.63	2.70	2.62	2.66	2.59	2.59	2.73	2.66	2.62	2.70	2.68
2.68	2.64	2.67	2.62	2.63	54.99	2.80	2.59	2.59	2.64	2.59	2.65	2.55	2.77	2.68	2.57	2.70	2.62
2.79	2.66	2.64	2.67	2.67	2.61	54.87	2.68	2.70	2.58	2.67	2.67	2.63	2.69	2.66	2.59	2.63	2.60
2.65	2.54	2.76	2.65	2.75	2.76	2.47	54.97	2.69	2.63	2.72	2.65	2.54	2.61	2.72	2.61	2.65	2.62
2.70	2.67	2.60	2.64	2.84	2.67	2.65	2.66	54.98	2.59	2.52	2.58	2.73	2.64	2.67	2.73	2.56	2.57
2.73	2.58	2.65	2.70	2.64	2.63	2.74	2.62	2.67	54.76	2.69	2.56	2.69	2.62	2.74	2.67	2.58	2.73
2.70	2.58	2.71	2.64	2.58	2.65	2.68	2.62	2.51	2.78	55.04	2.71	2.61	2.57	2.62	2.68	2.66	2.65
2.63	2.65	2.58	2.68	2.62	2.57	2.61	2.62	2.57	2.66	2.75	55.23	2.62	2.56	2.66	2.69	2.74	2.57
2.64	2.66	2.64	2.57	2.57	2.75	2.69	2.66	2.64	2.67	2.60	2.57	55.08	2.53	2.64	2.76	2.64	2.69
2.58	2.59	2.63	2.60	2.64	2.69	2.65	2.61	2.65	2.63	2.61	2.58	2.68	54.96	2.60	2.73	2.79	2.81
2.61	2.66	2.67	2.66	2.60	2.59	2.67	2.65	2.75	2.67	2.75	2.66	2.60	2.60	54.82	2.71	2.68	2.63
2.74	2.67	2.53	2.65	2.65	2.61	2.66	2.73	2.72	2.61	2.62	2.72	2.62	2.76	2.70	54.82	2.51	2.65
2.67	2.65	2.61	2.58	2.70	2.66	2.68	2.71	2.68	2.61	2.65	2.60	2.70	2.59	2.74	2.70	54.87	2.58
2.59	2.64	2.66	2.63	2.52	2.64	2.67	2.69	2.65	2.64	2.66	2.76	2.62	2.72	2.61	2.70	2.71	54.88

Σ_u

	BAC	BSC	Citi	GS	JPM	MER	MS	WB	WFC	CS10
BAC	31.92	3.85	8.56	-0.05	7.85	3.71	5.99	6.69	5.76	0
BSC	3.85	25.47	3.85	0.01	4.02	2.34	3.14	3.63	2.81	0
Citi	8.56	3.85	32.76	-0.04	7.49	3.80	6.21	6.67	6.23	0
GS	-0.05	0.01	-0.04	14.45	-0.03	0.01	-0.06	-0.02	-0.03	0
JPM	7.85	4.02	7.49	-0.03	35.26	4.10	5.10	7.12	5.23	0
MER	3.71	2.34	3.80	0.01	4.10	25.42	2.96	3.63	3.08	0
MS	5.99	3.14	6.21	-0.06	5.10	2.96	27.16	4.76	4.32	0
WB	6.69	3.63	6.67	-0.02	7.12	3.63	4.76	29.20	4.88	0
WFC	5.76	2.81	6.23	-0.03	5.23	3.08	4.32	4.88	27.54	0
CS10	0	0	0	0	0	0	0	0	0	72.93

Table III: Reference model parameters estimated using the MCMC procedure of Appendix A. The transition probability matrices Ω_f and Ω_c as well as the covariance matrix Σ_u are reported in percentage terms. The parameters are estimated using 10000 draws from the Gibbs sampler, with a 1000 draw burn-in period.

	Pre-crisis	Pre- Bear Stearns	Pre- Lehman Brothers
θ_d^{-1}	0.23	0.21	0.29
	(0.017)	(0.028)	(0.018)
θ_s^{-1}	0.64	0.65	0.69
	(0.021)	(0.029)	(0.019)
MSE	2.63	4.04	8.51

Table IV: Estimates of investors' degree of aversion to ambiguity about the underlying dynamics, θ_d^{-1}, and degree of aversion to the filter distribution, θ_s^{-1}. Half-width of the 95% confidence intervals are reported in parentheses. Mean Squared Error (MSE) is the average squared error between the observed CDS rates and the model-implied CDS rates for all institutions, maturities, and observations. Draws are made using the Metropolis-Hastings procedure of Section 5.2.

	BAC	BSC	Citi	GS	JPM	MER	MS	WB	WFC
Jul 31 07	36.2	161.7	37.2	81.2	55	74.7	75.2	39.5	35.9
	29.6	51.3	34.9	71.2	54.7	73.8	73.2	39.3	38.7
Aug 31 07	39.7	135.7	45.5	68.8	45.4	71.7	68.8	39.4	35
	22.1	128.7	41.4	36.5	43.2	72.2	64.7	32.3	32.2
Mar 31 08	86.8	122.7	138.2	115	87.5	195.8	153.9	142.8	80.8
	65.1	157.0	128.62	133.9	85.3	109.7	93.4	83.2	90.9
Sep 30 08	170	143.3	301.7	452.5	143.8	410.8	1022	385.8	170
	102.7	144.5	486.4	482.8	152.6	380.03	317.3	386.5	138.1
Oct 31 08	133.1	120.2	197.6	313.3	119.9	216.2	413.3	121	97.2
	133.3	161.3	178.8	380.7	146.6	203.5	416.26	131.3	93.1

Table V: Observed 5 year CDS rates and model-implied 5 year CDS rates at five different dates. The reference model parameters are estimated using the Gibbs sampling procedure of Appendix A; the degree of aversion to ambiguity about the underlying dynamics, θ_d^{-1}, and the degree of aversion to ambiguity about the filter distribution, θ_s^{-1}, are estimated using the Metropolis-Hastings algorithm of Section 5.2.

Panel A: Misspecified Model

	BAC	BSC	Citi	GS	JPM	MER	MS	WB	WFC
Jul 31 07	5.15	9.54	10.04	13.32	11.65	10.80	13.34	11.02	10.96
	100.43	-18.79	-25.16	256.87	-9.95	-32.18	47.10	-26.28	-28.05
Aug 31 07	2.61	11.74	13.42	3.77	13.01	16.03	9.11	15.03	15.32
	-49.44	23.04	33.73	-71.71	11.64	48.39	-31.73	36.44	39.81
Mar 31 08	4.81	0.00	16.22	5.12	11.64	13.42	9.08	12.83	13.01
	5.11	-100.00	19.85	37.19	-9.44	-15.05	0.53	-13.45	-13.90
Sep 30 08	4.27	–	9.55	10.31	11.50	10.47	13.77	10.74	10.66
	20.39		-42.75	155.77	-0.53	-23.03	59.41	-16.98	-18.83
Oct 31 08	5.00	–	13.86	4.55	10.44	11.81	8.25	11.34	11.49
	17.20		45.09	-55.83	-9.24	12.77	-40.09	5.61	7.72

Panel B: Reference Model

	BAC	BSC	Citi	GS	JPM	MER	MS	WB	WFC
Jul 31 07	16.47	16.47	16.47	16.47	16.47	16.47	16.47	16.47	16.47
	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Aug 31 07	16.46	16.46	16.46	16.46	16.46	16.46	16.46	16.46	16.46
	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03
Mar 31 08	11.23	0.00	11.26	11.23	11.21	11.19	11.21	11.22	11.21
	-31.74	-100.00	-31.63	-31.79	-31.88	-31.98	-31.90	-31.81	-31.87
Sep 30 08	11.23	–	11.25	11.25	11.26	11.22	11.21	11.22	11.22
	-28.05		-27.97	-27.99	-27.92	-28.12	-28.18	-28.13	-28.15
Oct 31 08	11.32	–	11.35	11.34	11.35	11.32	11.31	11.32	11.31
	0.81		0.85	0.84	0.79	0.84	0.85	0.87	0.85

Table VI: Expected time to default and the percentage change in the expected time to default relative to previous month for different financial institutions. Panel A: expected time to default perceived by the misspecification-averse agent. Panel B: expected time to default under the reference model. The reference model parameters are estimated using the Gibbs sampling procedure of Appendix A; the degree of aversion to ambiguity about the underlying dynamics, θ_d^{-1}, and the degree of aversion to ambiguity about the filter distribution, θ_s^{-1}, are estimated using the Metropolis-Hastings algorithm of Section 5.2.

July 31, 2007

	Main	$\theta_s^{-1} = 0$	$\theta_d^{-1} = 0$
BAC	-5.53	-0.82	-14.20
BSC	-71.79	-53.66	-5.90
Citi	31.10	-7.75	3.55
GS	1.19	0.59	1.88
JPM	-3.42	15.32	3.09
MER	-6.05	4.52	0.35
MS	-1.41	-0.15	0.59
WB	-2.12	0.98	-1.18
WFC	3.01	-3.67	0.95

August 31, 2007

	Main	$\theta_s^{-1} = 0$	$\theta_d^{-1} = 0$
BAC	-17.38	-2.52	-44.33
BSC	-61.64	-48.49	-5.16
Citi	-77.8	20.15	-9.01
GS	-29.22	-14.97	-46.95
JPM	5.51	-22.82	-4.85
MER	-12.13	9.07	0.7
MS	14.53	1.45	-5.96
WB	-31.98	15.23	-18.02
WFC	-25.14	31.14	-8

March 31, 2008

	Main	$\theta_s^{-1} = 0$	$\theta_d^{-1} = 0$
BAC	1.22	-22.47	-25
BSC	53.14	31.05	27.95
Citi	-6.98	-44.97	-6.93
GS	18.7	6.7	16.43
JPM	-28.23	-53.49	-2.51
MER	-45.49	-68.65	-43.97
MS	-42.18	-67.52	-39.31
WB	-45.67	-70.22	-41.74
WFC	6.84	-40.22	12.5

September 30, 2008

	Main	$\theta_s^{-1} = 0$	$\theta_d^{-1} = 0$
BAC	12.59	2.24	-39.59
BSC	-0.28	15.35	0.84

Citi	61.37	44.14	61.22
GS	-66.27	-22.98	6.7
JPM	21.32	-6.95	6.12
MER	-7.28	15.86	-7.49
MS	-65.75	-63.05	-68.95
WB	1.56	18.53	0.18
WFC	-18.53	-14.18	-18.76

October 31, 2008

	Main	$\theta_s^{-1} = 0$	$\theta_d^{-1} = 0$
BAC	-0.69	-0.10	-1.77
BSC	-108.75	-81.28	-8.93
Citi	-92.31	23.02	-10.54
GS	0.37	0.18	0.59
JPM	-4.69	21.02	4.25
MER	-11.10	8.29	0.64
MS	-1.53	-0.16	0.64
WB	-49.26	22.82	-27.47
WFC	-28.55	34.80	-9.03

Table VII: Percent deviation between the observed 5 year CDS rates and model-implied 5 year CDS rates for different ambiguity aversion specifications at five different dates. "Main" refers to the setting in the main body of the paper, with the representative agent averse to ambiguity about both the underlying dynamics and the filter distribution; θ_s^{-1} refers to the case when the representative agent is averse to ambiguity about the underlying dynamics only; $\theta_d^{-1} = 0$ refers to the case when the representative agent is averse to ambiguity about the filter distribution only. The reference model parameters are estimated using the Gibbs sampling procedure of Appendix A; the degree of aversion to ambiguity about the underlying dynamics, θ_d^{-1}, and the degree of aversion to ambiguity about the filter distribution, θ_s^{-1}, are estimated using the Metropolis-Hastings algorithm of Section 5.2.

July 31 2007

	Data	$\gamma = 0$	0.5	1	2	3	4	5
BAC	36.2	1.3	1.4	1.6	1.7	1.7	1.5	1.2
BSC	161.7	1.6	1.6	1.6	1.6	1.4	1.0	0.7
Citi	37.2	1.4	1.5	1.5	1.6	1.5	1.2	0.9
GS	81.2	1.4	1.4	1.3	1.2	1.0	0.8	0.5
JPM	55	1.2	1.8	2.7	5.7	10.8	17.9	25.9
MER	74.7	1.4	1.5	1.6	1.7	1.6	1.4	1.0
MS	75.2	1.4	1.5	1.6	1.8	1.7	1.5	1.1
WB	39.5	1.4	1.5	1.6	1.7	1.7	1.5	1.1
WFC	35.9	1.4	1.5	1.6	1.7	1.7	1.4	1.1

August 31 2007

	Data	$\gamma = 0$	0.5	1	2	3	4	5
BAC	39.7	1.9	2.1	2.2	2.5	2.5	2.2	1.7
BSC	135.7	2.1	2.1	2.1	2.0	1.7	1.3	0.9
Citi	45.5	2.0	2.1	2.2	2.2	2.1	1.7	1.2
GS	68.8	2.0	2.0	1.9	1.8	1.5	1.1	0.7
JPM	45.4	1.8	2.6	3.9	8.2	15.6	25.9	37.5
MER	71.7	2.0	2.1	2.2	2.4	2.3	2.0	1.5
MS	68.8	2.0	2.1	2.3	2.5	2.5	2.2	1.6
WB	39.4	2.0	2.1	2.3	2.5	2.5	2.1	1.6
WFC	35	2.0	2.1	2.3	2.4	2.4	2.0	1.5

March 31 2008

	Data	$\gamma = 0$	0.5	1	2	3	4	5
BAC	86.8	19.0	19.4	19.6	19.2	17.3	13.9	9.9
BSC	122.7	861.4	795.8	729.6	591.7	445.4	301.7	181.3
Citi	138.2	19.5	19.3	18.9	17.4	14.7	11.1	7.4
GS	115	20.2	18.8	17.4	14.2	10.7	7.2	4.2
JPM	87.5	18.6	26.2	36.5	67.9	116.0	176.8	237.7
MER	195.8	20.4	20.6	20.6	19.6	17.2	13.4	9.3
MS	153.9	19.9	20.2	20.4	19.8	17.7	14.1	9.9
WB	142.8	19.8	20.1	20.2	19.5	17.3	13.7	9.5
WFC	80.8	19.7	19.9	20.0	19.1	16.8	13.2	9.2

September 30 2008

	Data	$\gamma = 0$	0.5	1	2	3	4	5
BAC	170	19.2	19.6	19.8	19.5	17.5	14.1	10.0
BSC	143.3	19.8	18.8	17.7	14.9	11.6	8.1	4.9
Citi	301.7	19.8	19.6	19.2	17.7	15.0	11.3	7.5
GS	452.5	19.7	18.4	17.0	13.9	10.5	7.0	4.1
JPM	143.8	17.5	24.6	34.3	63.9	109.2	166.4	223.8

MER	410.8	19.7	19.9	19.8	18.9	16.6	13.0	9.0
MS	1022	20.0	20.3	20.5	19.9	17.8	14.2	9.9
WB	385.8	20.1	20.4	20.5	19.7	17.5	13.9	9.7
WFC	170	19.7	20.0	20.0	19.2	16.9	13.3	9.2

Table VIII: Observed five year CDS and implied CDS for different levels of the risk aversion coefficient, γ, at four different dates. $\gamma = 0$ corresponds to the case of risk-neutral investors, which is the reference model case in this paper. The reference model parameters are estimated using the Gibbs sampling procedure of Appendix A.

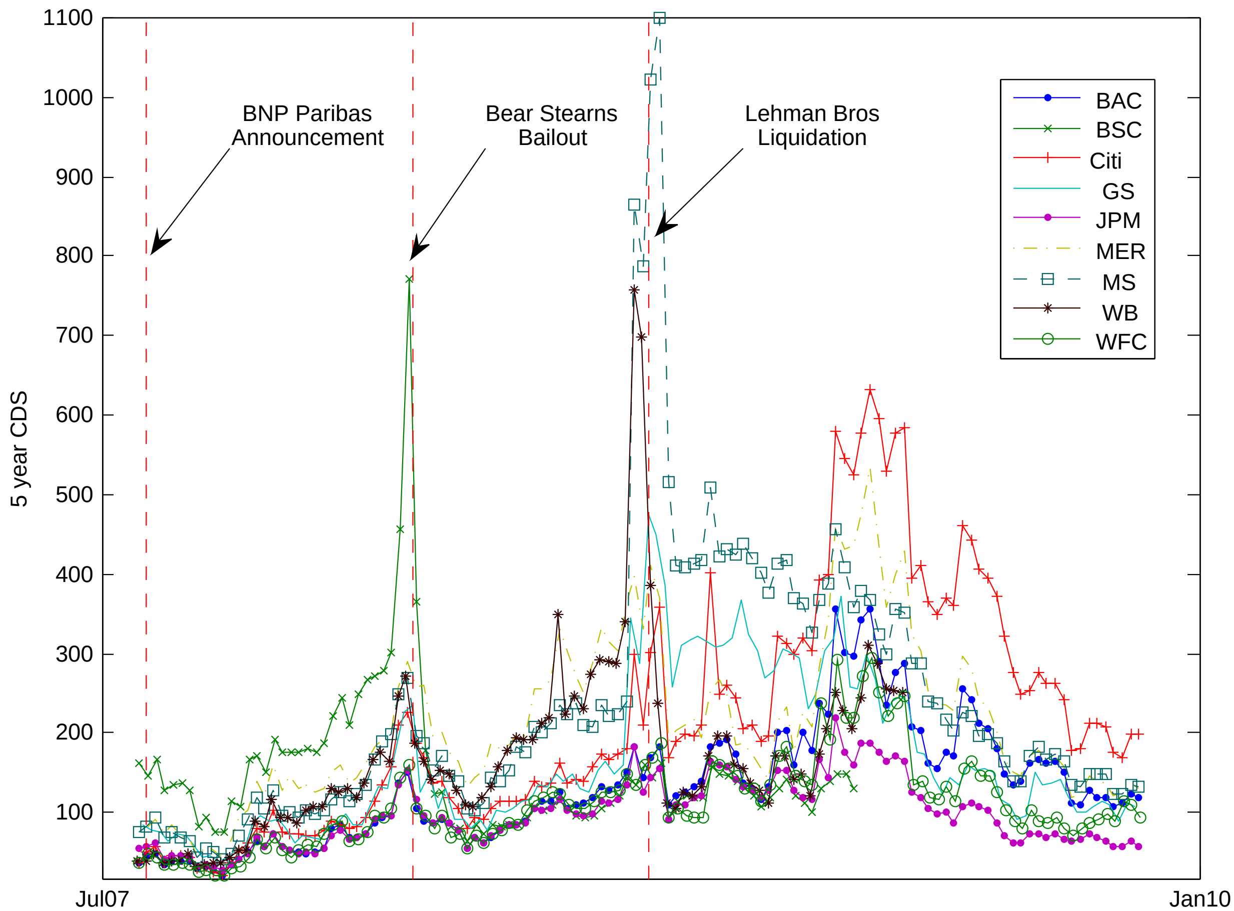

Figure 1: Evolution of the five year CDS spreads for financial institutions over the course of the crisis. Data source: Datastream

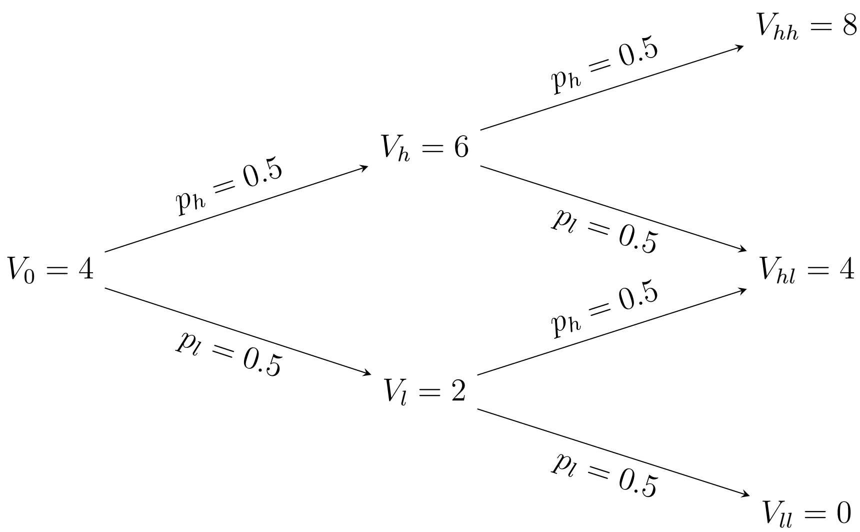

Figure 2: Event tree for the three period economy in Section 3.

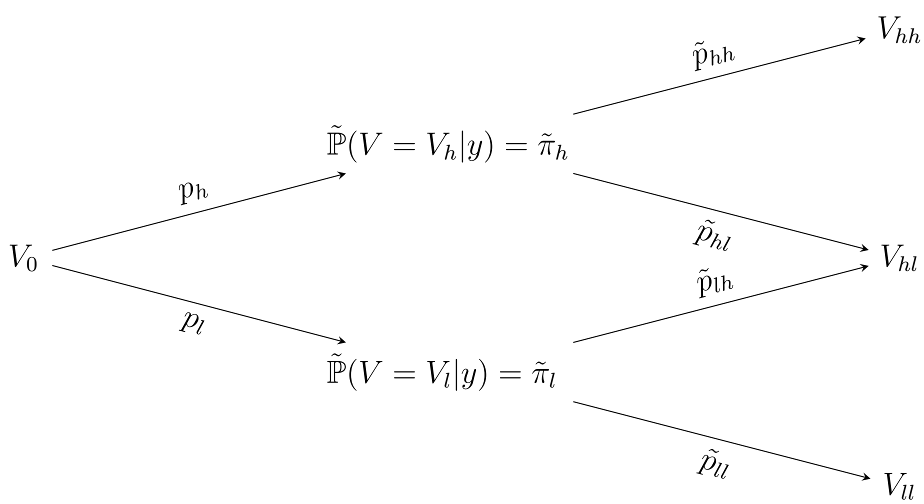

Figure 3: Misspecified probability tree for the three period economy in Section 3.

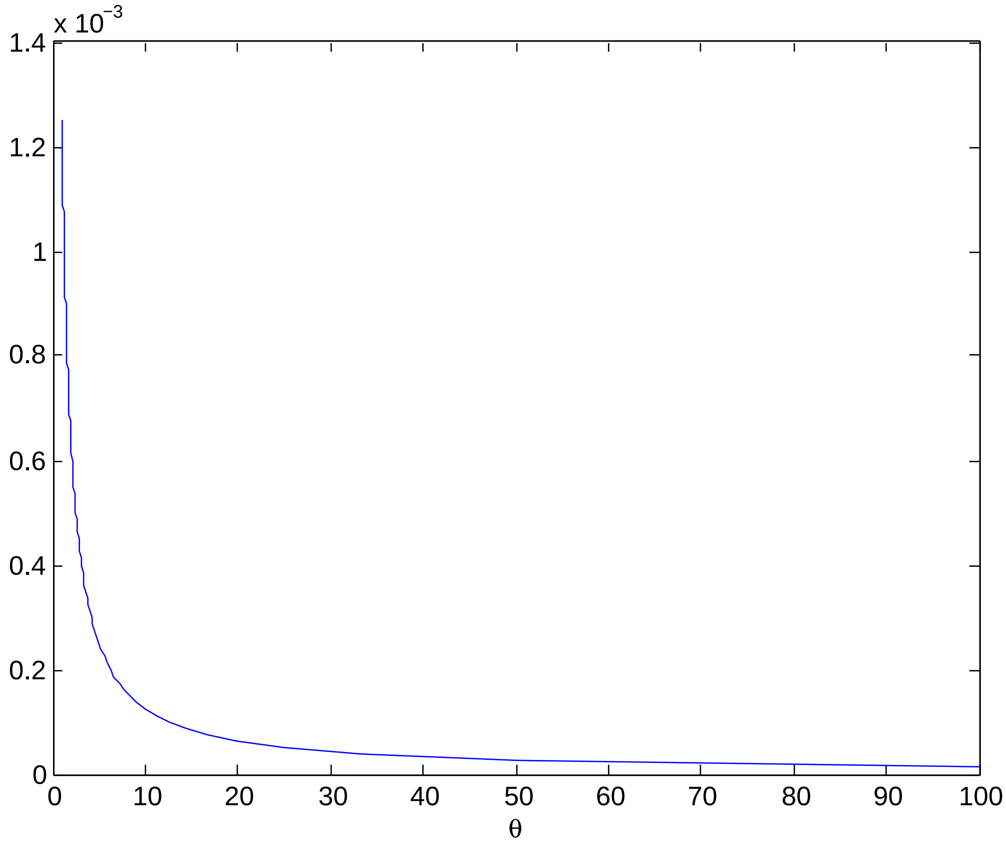

Figure 4: Ambiguity premium as a function of ambiguity aversion in an Ellsberg-style experiment with the prize-wealth ratio equal to 1 percent.

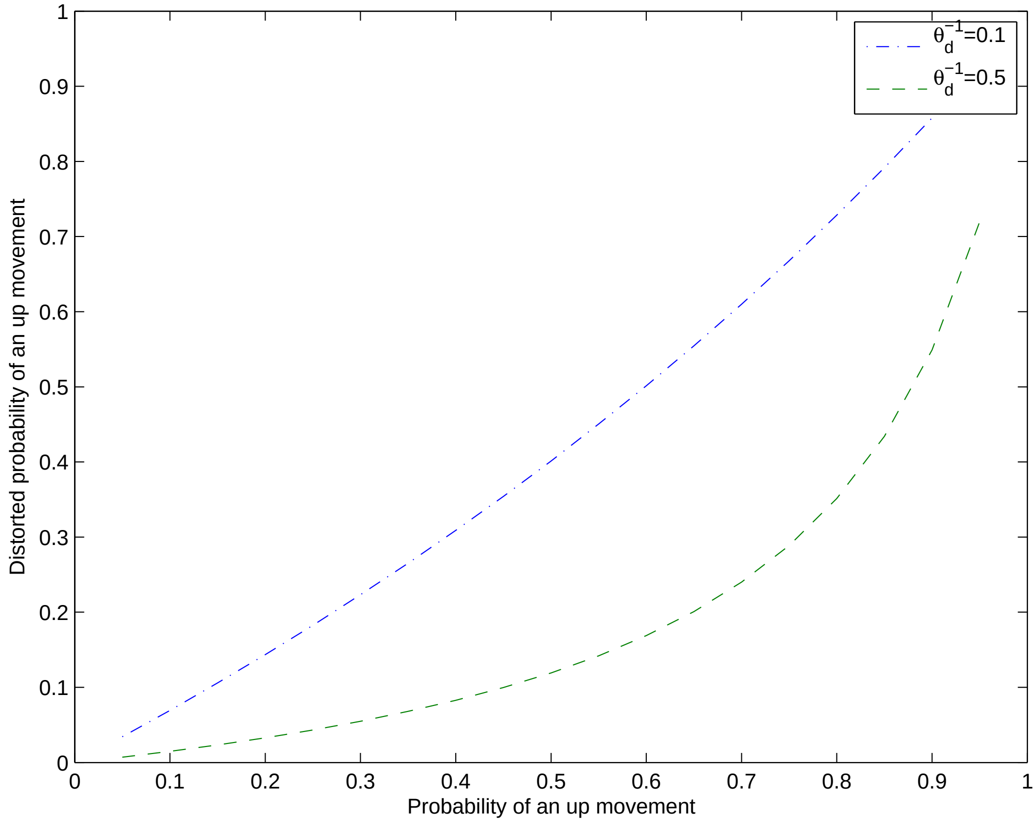

Figure 5: Distorted probability of an up movement for various levels of ambiguity aversion.

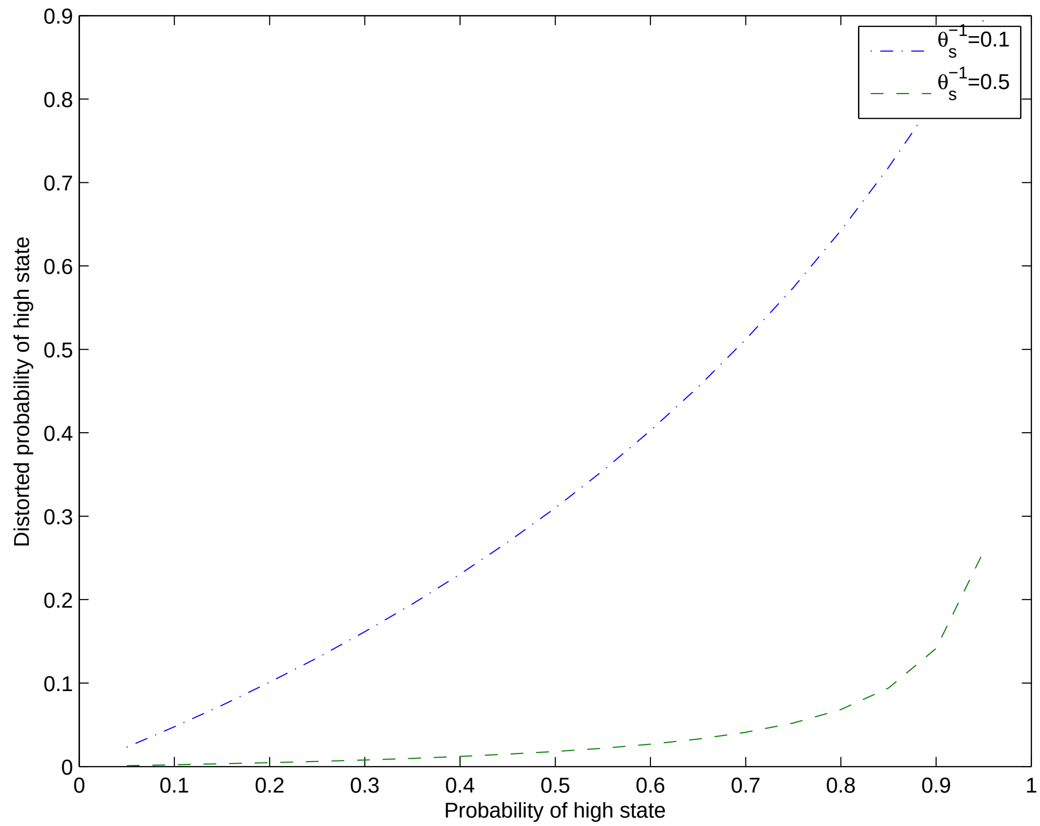

Figure 6: Distorted probability of the high state for various levels of ambiguity aversion.

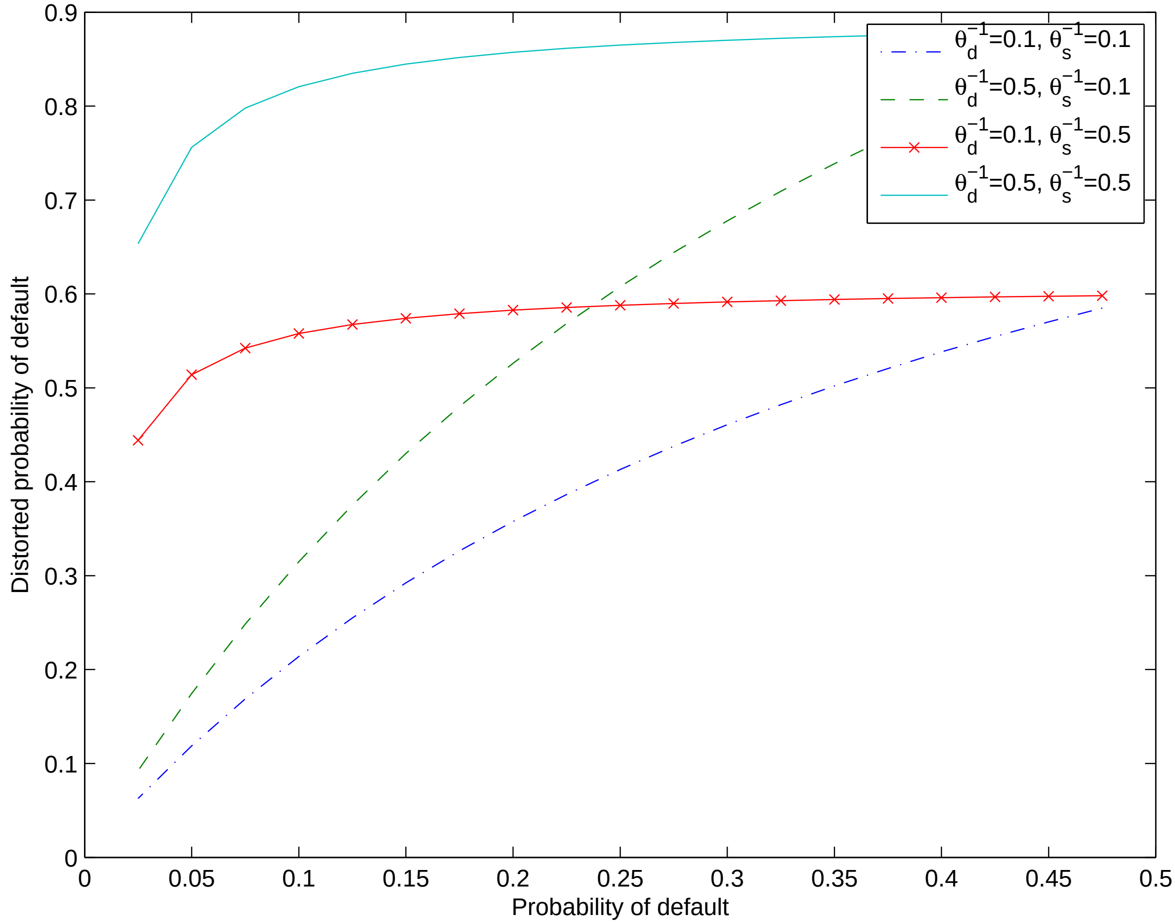

Figure 7: Distorted probability of default for various levels of ambiguity aversion.

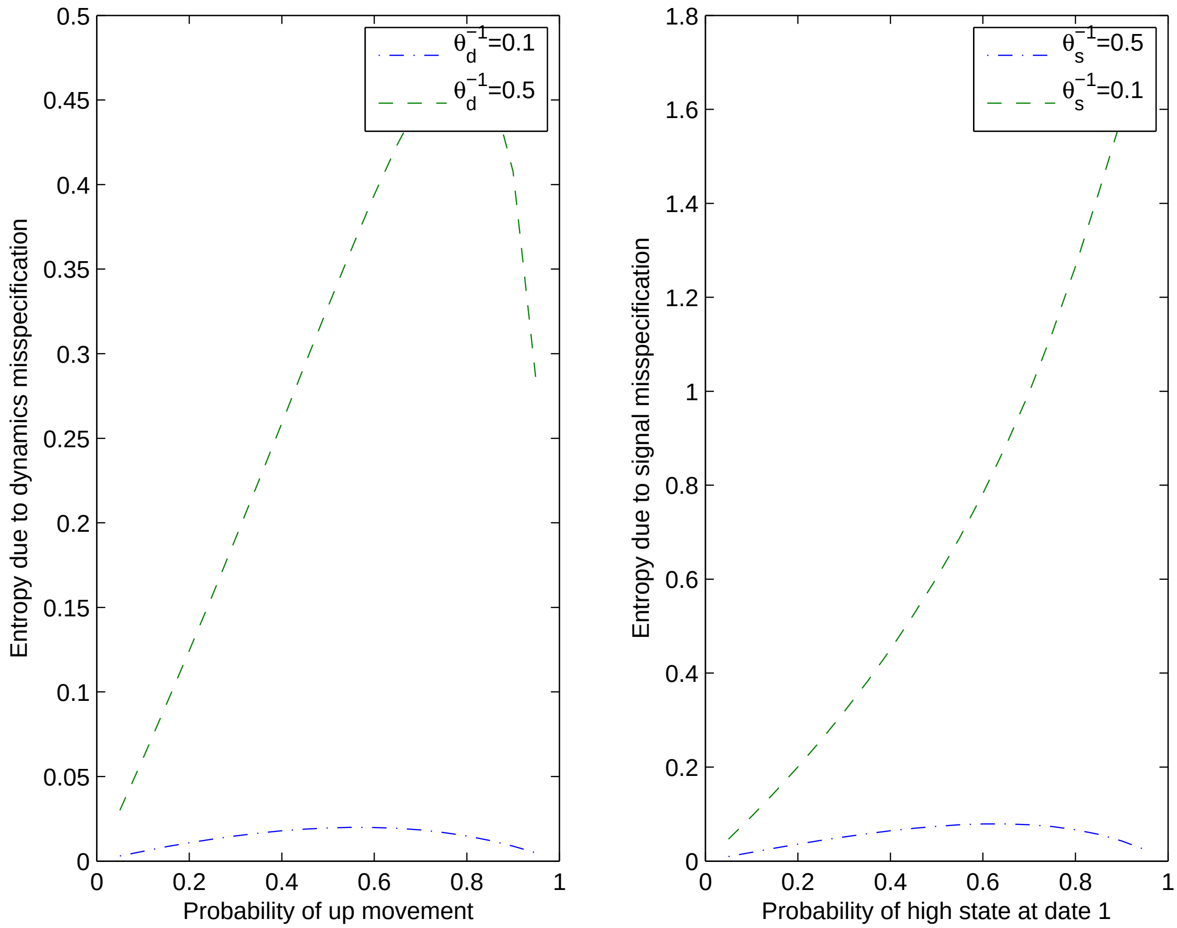

Figure 8: Contribution to entropy for various levels of ambiguity aversion.

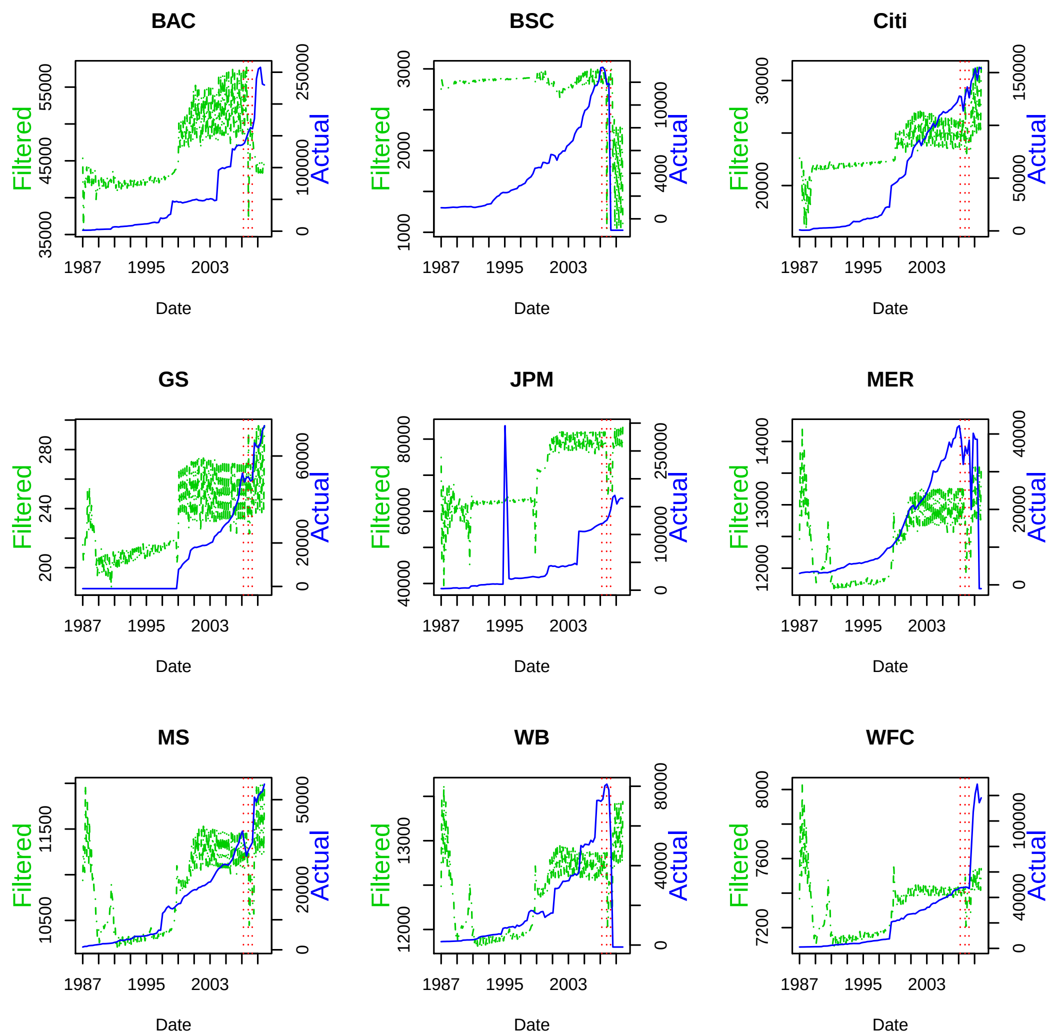

Figure 9: Filtered estimates of fundamental asset value evolution (left-hand scale) and the observed book value evolution (right-hand scale) for different financial institutions. Parameters are estimated using the Gibbs Sampling procedure of Appendix A with 10000 draws and 1000 draw burn-in. The different states are weighted using the reference model conditional probabilities.

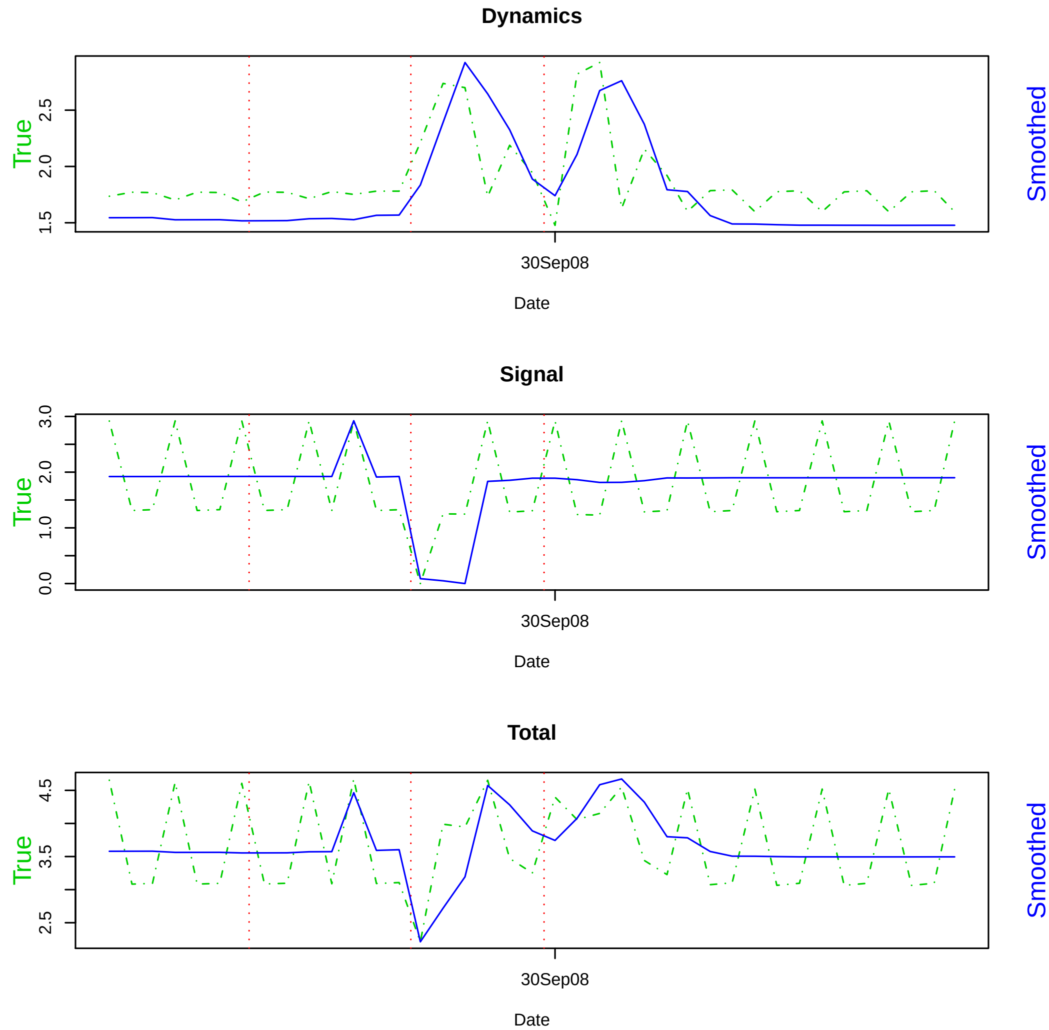

Figure 10: Components of relative entropy between the reference and misspecified models over time. Upper panel: entropy due to misspecification of the joint signals and states dynamics; central panel: entropy due to misspecification of the current period conditional probability; lower panel: total entropy. The right hand scale in each panel is the three month moving average of the corresponding entropy measure. The degree of aversion to ambiguity about the underlying dynamics, θ_d^{-1}, and the degree of aversion to ambiguity about the filter distribution, θ_s^{-1}, are estimated using the Metropolis-Hastings algorithm of Section 5.2.

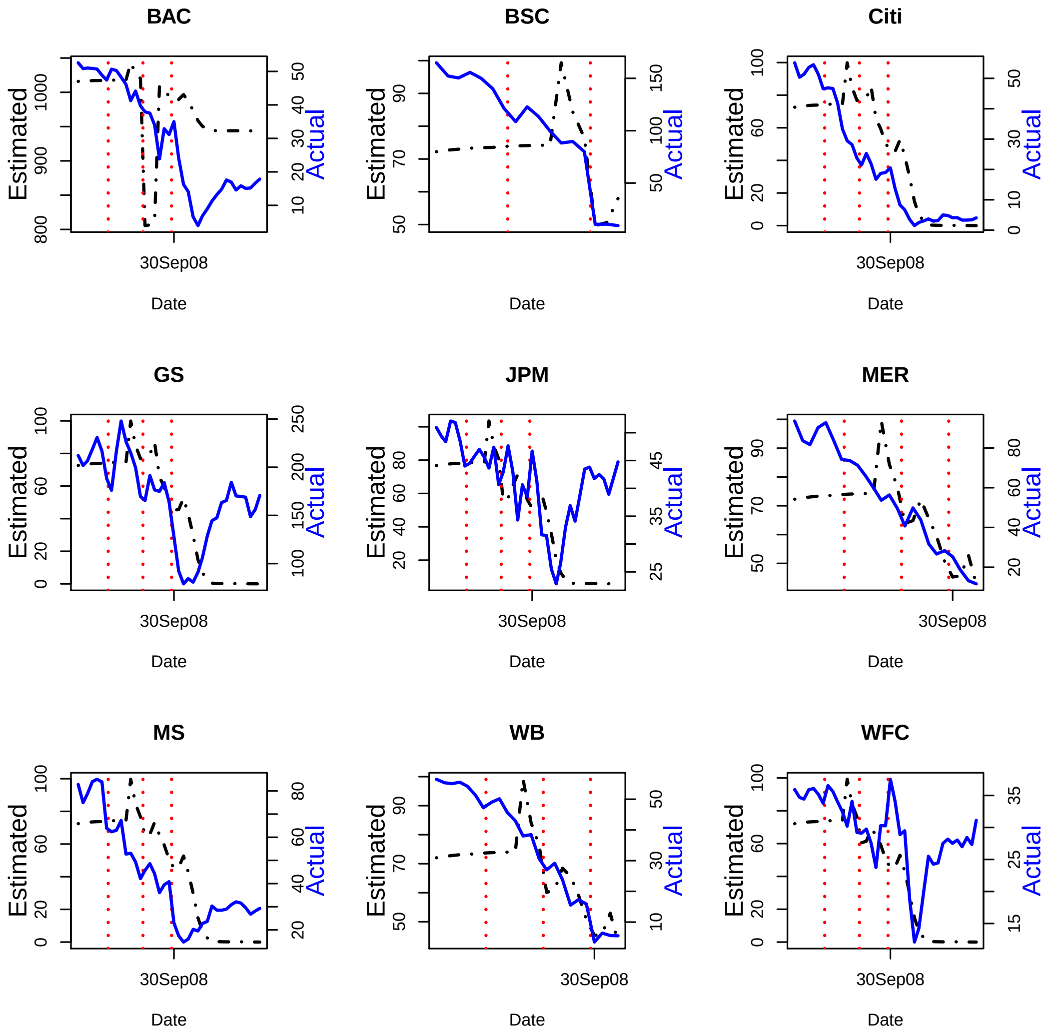

Figure 11: Value of equity under the misspecified model (left-hand scales) and the observed value of equity (right-hand scale) of the different financial institutions over time. The degree of aversion to ambiguity about the underlying dynamics, θ_d^{-1}, and the degree of aversion to ambiguity about the filter distribution, θ_s^{-1}, are estimated using the Metropolis-Hastings algorithm of Section 5.2.