Essays on Asset Pricing

Stig R. H. Lundeby

May 10, 2021

Contents

Acknowledgements

First of all, I wish to express my deepest gratitude to my advisor Tommy Stamland for his support and guidance throughout my Ph.D. study. His feedback, enthusiasm and continuous encouragement have been immensely valuable. I greatly appreciate the time and effort he has devoted to making my Ph.D. experience productive, interesting and highly enjoyable. Our many conversations and discussions have been a wealth of inspiration to me.

I am highly grateful to my co-advisor Lars A. Lochstoer for inviting me to spend a year at UCLA. His enthusiasm for research and immense knowledge have helped me become a better researcher. Lars A. Lochstoer and Mikhail Chernov, also at UCLA, co-authors with me on the first paper in this thesis. Working with both of them have been a pleasure and privilege, and I have learned a lot about the whole research process from idea generation to publication.

I would also like to thank the finance faculty and administrative staff at NHH for providing me support during the Ph.D. life. In particular, I am grateful to Espen Eckbo and Karin Thorburn for hiring me as their RA and for their perspectives and encouragement on my own research. I must also thank our head of department, Jøril Meland, for her continuous encouragement and support throughout my Ph.D. life.

I am very thankful to all my Ph.D. colleagues, past and present, for making the Ph.D. experience fun and rewarding. Especially, I would like to thank my friend and colleague Debashis Senapati for all our fun times and interesting conversations.

Lastly, I am greatly indebted to all my friends and family. In particular, my parents, sisters, and grandfather who have stood by me throughout and made me who I am. Words cannot express how lucky I am to have them in my life.

Stig R. H. Lundeby

Bergen, May 10, 2021

Summary

This dissertation consists of three separate papers about the risk-return trade-off in asset markets.

The first paper, co-authored with Mikhail Chernov and Lars Lochstoer, develops a new test for asset pricing models in a multi-horizon setting. The main idea behind the test is that a model should price payoffs correctly at all horizons. We show that the problem of testing whether pricing errors are zero at all horizons can be re-written as a single-horizon problem with appropriately chosen instruments. Thus, our approach effectively generates "new" test assets that are endogenous to the model. We also show formally that our approach can in principle detect most types of conditional misspecification.

The paper carries out an empirical investigation where we apply our test to a set of prominent factor models using the minimal requirement that the model prices its own factors at every horizon. Interestingly, we reject most of the factor models under consideration, indicating that the test has good power properties. Furthermore, the pricing errors on multi-period returns are often large, with annualized pricing errors frequently being similar in magnitude to the average premiums on the factors themselves.

The second paper investigates the implications of a counter-cyclical policy, e.g. monetary or fiscal policy, on the risk-return relationship on a broad stock market index. In particular, I show the presence of such a policy can explain the weak relationship between the volatility and expected returns seen in the literature. The intuition is straightforward. If the policy is expansive in bad states of the economy, it acts as a partial insurance to investors. The value of the insurance increases with risk. At the same time, the insurance is a "negative beta" asset and consequently earns a negative risk premium. The negative effect of the insurance on market risk premium therefore grows when risk increases.

In the time-series this leads to a weakened relationship between conditional volatility and risk premia.

When investors expect the policy to be in place for the foreseeable future, the policy can be viewed as a portfolio of implicit claims where each claim corresponds to the policy in place at a given future date. As expected, the claims to policy in the near future earn negative risk premia. However, the claims to policy further in the future earn large, *positive* risk premia. The reason is that the size of stimulus must in the long run be positively related to the size of the overall economy.

The third paper proposes a distance-metric for the multi-horizon setting analogous to the first metric proposed in Hansen and Jagannathan (1997) for the single-horizon setting. The distance metric can be used to answer the question "Which model is closer to explaining a set of test asset returns in a multi-horizon setting?".

In contrast to the J-statistic derived in Chernov, Lochstoer and Lundeby (2021), the distance metric does not reward variability of the candidate pricing kernel. Thus, if model A has larger pricing errors than model B, the distance metric will be larger for model A. In contrast, the J-statistic for model A might be lower than that of model B if model B is sufficiently variable.

In a simple example economy, I show that the multi-horizon distance metric can be large even in the case that the single-horizon HJ-distance is small or zero. Thus, a model that seemingly does a good job of explaining the risk-return relationship at a given frequency, e.g. monthly, might do a poor job of explaining the risk-return relationship at different frequencies.

Conditional dynamics and the multi-horizon risk-return trade-off [*]

Mikhail Chernov,[†] Lars A. Lochstoer,[‡] and Stig R. H. Lundeby[§]

First Draft: May 12, 2018
This Draft: April 5, 2021

Abstract

We propose testing asset-pricing models using multi-horizon returns (MHR). MHR effectively generate a new set of test assets that are endogenous to the model and that identify a broad set of possible conditional misspecifications. We apply MHR-based testing to prominent linear factor models and show that these models typically do a poor job of pricing longer-horizon returns, with pricing errors that are similar in magnitude to the risk premiums they were designed to explain. We trace the errors to the conditional factor dynamics. Explicitly incorporating factor timing in the models often makes mispricing worse, posing a challenge for future research.

JEL Classification Codes: G12, C51.

Keywords: multi-horizon returns, stochastic discount factor, linear factor models.

---

[*]We thank the editor Ralph Koijen and two referees for important feedback on the manuscript. We are also grateful to Hank Bessembinder, Ian Dew-Becker, Greg Duffee, Valentin Haddad, Serhiy Kozak, Francis Longstaff, Tyler Muir, Christopher Polk, Seth Pruitt, Shri Santosh, Tommy Stamland, and Stan Zin for comments on earlier drafts, as well as participants in the seminars and conferences sponsored by the ASU Winter Finance Conference, BI, David Backus Memorial Conference at Ojai, LAEF, NBER LTAM conference, NBIM, Norwegian School of Economics, the SFS Finance Cavalcade, Stanford GSB, the UBC Winter Finance Conference, and UCLA.

[†]UCLA, NBER, and CEPR; mikhail.chernov@anderson.ucla.edu.

[‡]UCLA; lars.lochstoer@anderson.ucla.edu.

[§] Norwegian School of Economics; stig.lundeby@nhh.no.

1 Introduction

In this paper we propose a new asset-pricing test and apply it to a set of leading linear factor models. Do we need another test, you might ask. Current tests take the set of test assets as given, although it is well-understood that this decision on the part of the researcher is critical for test performance. Over time novel expected return patterns observed in the historical data prompt modifications in the existing models to account for these patterns. As this process unfolds, viewing the test assets as "outside" the model, as the tests assume, becomes tenuous.[1]

We argue that using multi-horizon returns (MHR) offers a useful way to address these issues. Specifically, we show formally that MHR effectively generate a set of test assets that are endogenous to the model at hand and that allow for testing most, if not all, aspects of conditional model misspecification. No matter how much conditioning information a model already accounts for in the construction of the model's factors, the MHR-based test generates new test assets that leverage this information in an endogenous fashion.

Our test is derived using the standard no-arbitrage condition and by formulating models in terms of their implications for the stochastic discount factor (SDF). No-arbitrage implies that the h-period SDF equals the product of the h corresponding single-period SDFs. It is therefore straightforward to derive a model's implication for

[1] A classic test evaluates whether test assets have zero "alpha" (Gibbons, Ross, and Shanken, 1989). Lo and MacKinlay (1990) discuss the effects of data snooping. Lewellen, Nagel, and Shanken (2010) and Daniel and Titman (2012) cover the effects of test asset factor structure.

returns at any horizon. Thus, with MHR we are testing overidentifying restrictions of the model.

In our empirical contribution we show that misspecification of the implicit temporal dynamics in state-of-the-art models of the SDF, as uncovered by MHR, indeed are quantitatively large. Specifically, we consider eight linear factor models: the unconditional CAPM, a two-factor model related to Black, Jensen, and Scholes (1972) (the market factor plus a betting-against-beta factor), the Carhart (1997) four-factor model (the three Fama and French (1993) factors plus momentum), the Fama and French (2015) five-factor model, the Daniel, Mota, Rottke, and Santos (2020) five-factor model, the Stambaugh and Yu (2017) four-factor model, the Hou, Xue, and Zhang (2015) four-factor model, and the Haddad, Kozak, and Santosh (2020) six-factor model. These models which are either workhorse or recent cutting-edge models for empirical risk-return modeling. We test the minimal requirement that a model prices its own factors at multiple return horizons.

As an example of the test results, consider the market factor in the Fama-French model. The h-period gross return to this factor is simply the product of the one-period gross returns from t to $t+h$. The model trivially prices the one-period return to this factor, but quickly generates pricing errors when we consider the model's implications for longer-period returns. At the four-year horizon, the model's annualized pricing error for the market factor is 7% – about the same as the market risk premium itself.

This example is not unique. The average annualized pricing error across all factors and models is 4.5% when tested jointly on horizons of 1, 3, 6, 12, 24, and 48 months. This is about the same magnitude as the average annualized factor risk premiums the models where designed to explain in the first place.

Five out of these eight models are rejected at the 5% level. The Black, Jensen, and Scholes (1972) and the Carhart (1997) models are rejected at the 10% level. The CAPM is not rejected. All the models are rejected at the 5%-level in an alternative test, where we use MHR to a common set of test assets (the five Fama and French (2015) factors) across all the eight models.

The baseline MHR-test rejections imply that the models fail to price their own factors conditionally, indicating that the factors need to be timed in order to span the unconditionally mean-variance efficient (UMVE) portfolio. As a next step, we develop further intuition by illustrating some statistical and economic properties of factors vis-a-vis the implications of the null hypothesis. Under the null, the factors (excess returns on traded portfolios) span the unconditionally mean-variance efficient portfolio. This implies that the conditional expectation of these factors is proportional to their conditional second moment. We use this observation to construct artificial asset returns corresponding to the null.

Armed with these artificial returns, we compare their cumulative autocorrelations and Sharpe ratios with those of actual factors. For the market portfolio, the autocorrelation in the data exhibits little persistence as is the case under the null. Many of the other models, however, have much stronger patterns. For instance, in the Daniel, Mota, Rottke, and Santos (2020) model the autocorrelation in the data is much higher than that under the null and the two are statistically significant different from each other. The Hou, Xue, and Zhang (2015) model exhibits similarly strong but the opposite pattern: the autocorrelation in the data is much lower than that under the null. Thus, the null hypothesis implicitly misspecifies dynamic properties of factors.

Sharpe ratios convey a similar message. In misspecified models the Sharpe ratios under the null and in the data tend to diverge with horizon. If the autocorrelation is higher in the data than under the null, the Sharpe ratios are lower in the data, and vice versa. The differences are economically large. For instance, in the Hou, Xue, and Zhang (2015) model the annualized Sharpe ratios in the data and in the model at the 48-month horizon are 0.75 and 0.48, respectively.

As a final step in our empirical analysis, we evaluate state-of-the-art approaches of factor timing with the ultimate objective to model the UMVE portfolio. Specifically, we consider out-of-sample volatility timing of Moreira and Muir (2017), expected return timing using book-to-market ratio of Haddad, Kozak, and Santosh (2020), and timing on a non-linear function of many stock characteristics of Kozak, Nagel, and Santosh (2020). We apply volatility timing to the CAPM, the Carhart (1997), and the Fama and French (2015) models. Expected return timing is applied to the static version of Haddad, Kozak, and Santosh (2020) model.[2] Lastly, the characteristic-based timing is applied to the UMVE portfolio implied by that approach. We reject all the models using our test, and the MHR pricing errors are of a similar magnitude

[2] Kozak, Nagel, and Santosh (2018) is the antecedent to static Haddad, Kozak, and Santosh (2020). Throughout, we refer to the model as Haddad, Kozak, and Santosh (2020), or HKS, for convenience and because we use the same dataset.

or larger than in the versions of the models without factor timing.

Our rejection of the most advanced models in the literature suggests the extreme challenge of correctly estimating the conditional factor dynamics. In fact, as the number of factors in a model increases, so does the complexity of their conditional dynamics. Simultaneously, those same rejections indicate that our test has good statistical power properties. Taken together our evidence poses a challenge for future research in terms of understanding and estimation of the conditional risk-return trade-off. The MHR-based test that we propose should serve as useful guidance for this endeavor.

Related literature. There are many papers that test conditional versions of factor models. For instance, Boguth, Carlson, Fisher, and Simutin (2011), Ferson and Harvey (1999), Farnsworth, Ferson, Jackson, and Todd (2002), Ghysels (1998), Jagannathan and Wang (1996), Kelly, Pruitt, and Su (2019), Lettau and Ludvigson (2001), Lewellen and Nagel (2006), and Moreira and Muir (2017). Our contribution relative to this literature is to show that MHR in asset pricing tests effectively serve as conditioning variables endogenous to the model and that, empirically, multi-horizon factor returns indeed are informative in terms of uncovering novel conditional dynamics of prominent factor models. In contemporaneous and independent work Haddad, Kozak, and Santosh (2020) and Linnainmaa and Ehsani (2019) use different methods to study factor dynamics with a focus on single-horizon returns.

Our paper makes a connection with a literature that seeks to characterize multi-horizon properties of "zero-coupon" assets, such as bonds, dividends strips, variance swaps, and currencies. Such work includes Backus, Boyarchenko, and Chernov (2018), Belo, Collin-Dufresne, and Goldstein (2015), van Binsbergen, Brandt, and Koijen (2012), Dahlquist and Hasseltoft (2013), Dew-Becker, Giglio, Le, and Rodriguez (2015), Hansen, Heaton, and Li (2008), Koijen, Lustig, and Nieuwerburgh (2017), Lustig, Stathopoulos, and Verdelhan (2013), and Zviadadze (2017).

A related strand of the literature considers multiple frequencies of observations when testing models (e.g., Brennan and Zhang, 2018, Daniel and Marshall, 1997, Jagannathan and Wang, 2007, Kamara, Korajczyk, Lou, and Sadka, 2016, Parker and Julliard, 2005), though none of these consider the implications of a joint test across horizons.

Baba Yara, Boons, and Tamoni (2020) consider the predictive power of characteristics lagged at different horizons. Favero, Melone, and Tamoni (2020) analyze a factor model that also incorporates long-run relationships through cointegration. Bessembinder, Cooper, and Zhang (2020) model and document changes in measures of mutual fund performance at long horizons.

Notation. We use E for expectations and V for variances (a covariance matrix if applied to a vector). A t-subscript on these denotes an expectation or variance conditional on information available at time t, whereas no subscript denotes an unconditional expectation or variance. We use double subscripts for time-series variables, like returns, to explicitly denote the relevant horizon. Thus, a gross return on an investment from time t to time $t+h$ is denoted $R_{t,t+h}$.

2 Linear factor models and multi-horizon returns

This section has three main objectives. First, we offer a unified view of model construction in cross-sectional asset pricing. Second, we highlight difficulties in testing and evaluating progress in improving such models. Third, we introduce a testing approach that is essentially a dual to the dominant testing paradigm in the literature. This new approach is attractive because it allows sidestepping many issues that we describe.

2.1 Factor model construction

A long-standing paradigm in asset pricing is that of the construction of the mean-variance frontier (MVF). Applications include linear beta-pricing models of the cross-section of expected returns, as well as a more general understanding of the properties of the minimum-variance SDF (see, e.g., Cochrane, 2004, Hansen and Richard, 1987). Here, we review key concepts to set the stage and introduce notation for our novel test.

Let $R^{ei}_{t,t+1}$ represent asset i's one-period excess return. Stack excess returns on all assets into an $I_t \times 1$ vector $R^{e}_{t,t+1}$. The unconditional MVE portfolio (UMVE) is then

$$\begin{aligned} R^{U}_{t,t+1} &= k(w^{U}_{t})^{\top} R^{e}_{t,t+1}, \\ w^{U}_{t} &= (1+\theta_t)^{-1} \cdot V_t\left(R^{e}_{t,t+1}\right)^{-1} E_t\left(R^{e}_{t,t+1}\right), \end{aligned} \tag{1}$$

where θ_t is the maximal squared conditional SR:

$$\theta_t = E_t\left(R^e_{t,t+1}\right)^\top V_t\left(R^e_{t,t+1}\right)^{-1} E_t\left(R^e_{t,t+1}\right),$$

and k is a constant governing the leverage of the portfolio. Setting $k = (1 - E((w^U_t)^\top R^e_{t,t+1}))^{-1}$, the SDF

$$M_{t,t+1} \;=\; 1 - (R^U_{t,t+1} - E(R^U_{t,t+1})), \tag{2}$$

prices all excess returns both conditionally and unconditionally: $E_t(M_{t,t+1}R^e_{t,t+1}) = E(M_{t,t+1}R^e_{t,t+1}) = 0$. See Appendix A.1 for a derivation of these relations. Ferson and Siegel (2001) offer an alternative derivation by computing UMVE weights directly.

Finding the UMVE portfolio weights in Equation (1) faces three major hurdles: handling all assets (stocks) is an intractable problem for a variety of reasons, the full information set implicit in the subscript t is not known, and computation of correct conditional mean and variance of returns is not possible without knowing their true distribution at each point in time t. In response to these challenges, the literature evaluates portfolios of stocks (Black, Jensen, and Scholes, 1972) and considers various conditioning variables as explicit proxies for the information set, such as the cross-section of market-to-book ratios (Fama and French, 1993) or the aggregate dividend-price ratio (Fama and French, 1988).

All of these approaches translate to the following form of the UMVE weights:

$$k(w_t^U)^\top = b_t^\top C_t,$$

where C_t is a $K \times N_t$ matrix of stock-level characteristics, and b_t is a $K \times 1$ timing vector. The characteristics define a set of K factors,

$$F_{t,t+1} \equiv C_t R^e_{t,t+1},$$

which, if the model is true, conditionally span the UMVE portfolio: $R^U_{t,t+1} = b_t^\top F_{t,t+1}$. The $K \times 1$ factor timing vector, b_t, optimally combines these factors over time to get to the UMVE portfolio. If b is constant, the factors defined by the characteristics C_t unconditionally span the UMVE portfolio.

Factor timing, as studied in the literature, can be generically represented as

$$b_t = D_0 + D_1 z_t, \tag{3}$$

where D_0 and D_1 are a $K \times 1$ vector and a $K \times L$ matrix of parameters, respectively, while z_t is a $L \times 1$ vector with observable conditioning variables. As one example, if the market dividend-price ratio (dp_t) is used as a conditioning variable for each factor, $z_t = dp_t$, while D_1 is a $K \times 1$ vector(e.g., Ferson and Harvey, 1999, Jagannathan and Wang, 1996, and Lettau and Ludvigson, 2001). As another example, conditioning variables may be factor-specific as well. For instance, Moreira and Muir (2017) advocate using the inverse conditional variance as the timing variable for each factor.

In this case, z_t is a $K \times 1$ vector with factor k's inverse conditional variance in row k, while D_1 is a diagonal $K \times K$ matrix. Similarly, Haddad, Kozak, and Santosh (2020) use factor book-to-market ratios.

Given the factor weights, b_t, the SDF can now be written as:

$$M_{t+1} = 1 - \left(b_t^\top F_{t,t+1} - E(b_t^\top F_{t,t+1})\right). \tag{4}$$

Substituting in for b_t using Equation (3) leads to an SDF with constant loadings on the original factors combined with additional factors that are interactions of the original factors and the variables in z_t.

Theoretically, we know what drives the time-variation in b_t. From Equation (1) for the UMVE portfolio weights, using the factors as the set of base assets, we have:

$$b_t \propto \frac{V_t\left(F_{t,t+1}\right)^{-1} E_t\left(F_{t,t+1}\right)}{1+\theta_t^F}, \tag{5}$$

where θ_t^F is the maximal squared conditional Sharpe ratio possible from investing in the factors. If the frequency of the data is high, $1+\theta_t^F$ is close to 1. Thus, a model where b_t is constant over time implicitly assumes something about the dynamics of $F_{t,t+1}$ in the form of a specific relation between their conditional mean and variance. In fact, an alternative form of this equation is that the conditional factor means must be proportional to the conditional second moment of the factors. See Appendix A.1. If this is not true, the factor model will exhibit conditional mispricing.

In summary, a linear factor model consists of several important ingredients: the initial set of assets, the set of cross-sectional conditioning variables C_t that generates factors that conditionally span the UMVE portfolio, and the time-series conditioning variables z_t that optimally weight these factors over time.

2.2 Testing linear factor models

Consider a version of the SDF in Equation (4) where $b_t = b$ is constant, which implies that the factors span the UMVE portfolio

$$M_{t,t+1} \quad = \quad 1 - b^{\top}(F_{t,t+1} - E(F_{t,t+1})). \tag{6}$$

Researchers focus on the implication from this SDF that any asset's risk premium is linear in the factor risk premiums. The model is therefore commonly tested via the regression

$$\Omega_t R^e_{t,t+1} = \alpha + \beta F_{t,t+1} + \varepsilon_{t+1}, \tag{7}$$

where Ω_t is a $L \times I_t$ matrix of time t portfolio weights, highlighting that test assets typically are trading strategies in the underlying set of base assets. Further, α is a $L \times 1$ vector, β is an $L \times K$ matrix, and ε_{t+1} is an error term with covariance matrix $V(\varepsilon_{t+1})$. If the SDF in Equation (6) is correctly specified, $\alpha = 0$ (see Gibbons, Ross, and Shanken (1989) for the associated test statistic).

The testing challenge is to find a set of test portfolios that are sufficiently informative about a given model. The ideal test asset is the UMVE portfolio, which is unattainable (e.g., Barillas and Shanken, 2017). Thus, in practice, researchers instead search for test portfolios that have two properties: (i) large spread in average returns and (ii) returns that are not spanned by the model factors (e.g., Daniel and Titman, 2012, Lewellen, Nagel, and Shanken, 2010).

Importantly, if a model is rejected in this test, we know how to modify the model so that it prices the test assets in-sample. As explained in MacKinlay (1995), we need to add a factor with portfolio weights proportional to $V^{-1}(\varepsilon_{t+1}) \cdot \alpha$ from regression (7). That is, we tend to use information from the construction of the test assets for the construction of new factors. This insight informs the search for additional characteristics and timing variables, which refine the conditioning information implicit in z_t and C_t. A logical conclusion of this process is the explicit data-mining approach of Kozak, Nagel, and Santosh (2020), which considers all functions of all characteristics used in prior research in the model test and construction.

Thus, "the model" is in practice the union of the factors and the test assets. A question that arises is how to test such a model, where there is a feedback from the in-sample performance of test assets to factor construction (see Lo and MacKinlay (1990) for an early discussion of these issues). In the next section, we show that MHR provide a way to test this broader notion of a model that, unlike existing tests, does not rely on specifying further conditioning information beyond the C_t and z_t already used in the model construction.

2.3 The role of multiple horizons

In this paper we propose using MHR in model tests by making use of the models' no-arbitrage implications for returns across different horizons. It turns out there is a tight connection between correct conditional pricing and unconditional pricing across multiple horizons. In order to describe the connection, we have to extend factor models to multiple horizons.

As is known from extant literature, see e.g., Grossman, Melino, and Shiller (1987), Levhari and Levy (1977), and Longstaff (1989), a factor model does not apply across all horizons. To see this, consider the two-period SDF implied by Equation (6):

$$\begin{aligned} M_{t,t+2} &= M_{t,t+1}M_{t+1,t+2} = (a - b^\top F_{t,t+1})(a - b^\top F_{t+1,t+2}) \\ &= a^2 - ab^\top F_{t,t+1} - ab^\top F_{t+1,t+2} + b^\top F_{t,t+1}F_{t+1,t+2}^\top b, \end{aligned}$$

where $a = 1 + b^\top E(F_{t,t+1})$. This implies that the corresponding regression for the two-period return $R^{ei}_{t,t+2}$ will essentially feature a new set of factors even if the original single-horizon model is correctly specified.

The SDF-based approach is a natural way to translate a regression-based linear factor model for expected returns into its counterpart at any longer horizon h. Denote the one-period gross risk-free rate by $R^f_{t,t+1}$, and the asset's gross return by $R^i_{t,t+1} = R^{ei}_{t,t+1} + R^f_{t,t+1}$. The multi-horizon SDF and returns are simple products of their single-horizon counterparts:

$$
\begin{aligned}
M_{t,t+h} &= \prod_{j=1}^{h} M_{t+j-1,t+j}, \\
R^i_{t,t+h} &= \prod_{j=1}^{h} R^i_{t+j-1,t+j}.
\end{aligned}
$$

Thus, we cast analysis in this paper in terms of SDFs. Switching over to the SDF language means that the focus on the magnitude of α changes to the focus on whether

$$E(M_{t,t+h}R^i_{t,t+h}) = 1. \tag{8}$$

Simply put, in a correctly specified model the present value of any \$1 investment is indeed \$1.

Proposition 1. *Consider the Euler equation (8) at the single and $h-$period horizons.*

1. *Suppose $E_t(M_{t,t+1}R^i_{t,t+1}) = 1$ for any i, then $E(M_{t,t+h}R^i_{t,t+h}) = 1$ for any i, h.*

2. *Suppose $E(M_{t,t+h}R^i_{t,t+h}) = 1$ for any i and h, then for any i, the conditional pricing error, $E_t(M_{t,t+1}R^i_{t,t+1})-1$, is zero-mean and uncorrelated with the lagged "Euler equation errors", $M_{t-h,t}R^i_{t-h,t} - 1$, for any h.*

We present the proof here because it is helpful in developing intuition about the meaning of the Proposition.

Proof.

1. By recursively iterating on the following equation for $h = 1, 2, \ldots,$ we have:

$$\begin{aligned} & E[M_{t-h,t+1}R^i_{t-h,t+1}] = E[M_{t-h,t}R^i_{t-h,t}M_{t,t+1}R^i_{t,t+1}] \\ = \quad & E[M_{t-h,t}R^i_{t-h,t}E_t[M_{t,t+1}R^i_{t,t+1}]] = E[M_{t-h,t}R^i_{t-h,t}] = 1. \end{aligned}$$

2. First note that conditional pricing errors, $E_t(M_{t,t+1}R^i_{t,t+1}) - 1$, have zero mean because $E(M_{t,t+1}R^i_{t,t+1}) - 1 = 0$. Next, consider two-period returns:

$$\begin{aligned} 1 &= E(M_{t-1,t+1} \cdot R^i_{t-1,t+1}) = E(M_{t-1,t}M_{t,t+1} \cdot R^i_{t-1,t}R^i_{t,t+1}) \\ &= E(M_{t-1,t}R^i_{t-1,t}E_t(M_{t,t+1}R^i_{t,t+1})) \\ &= E(M_{t-1,t}R^i_{t-1,t}) \cdot E(M_{t,t+1}R^i_{t,t+1}) + Cov(M_{t-1,t}R^i_{t-1,t}, E_t(M_{t,t+1}R^i_{t,t+1})). \end{aligned}$$

If (8) holds at both one- and two-period horizons, then

$$Cov(M_{t-1,t}R^i_{t-1,t} - 1, E_t(M_{t,t+1}R^i_{t,t+1}) - 1) = 0.$$

A generalization to $(h+1)$-period returns is

$$\sum_{j=1}^{h} Cov(M_{t-j,t}R^i_{t-j,t} - 1, E_t(M_{t,t+1}R^i_{t,t+1}) - 1) = 0. \tag{9}$$

These equations tell us that conditional pricing errors are uncorrelated with Euler equation errors for any h.

The second part of the Proposition forms the basis for the test of asset-pricing models that we propose in this paper. Specifically, we advocate testing if $E(M_{t,t+h}R^i_{t,t+h}) = 1$ jointly for a set of different h and i. In order to appreciate what the rejection of this null tells us, we revisit some elements of the proof.

Express $M_{t,t+h}R_{t,t+h} - 1$ as $\eta_t^{(h)} + \nu_{t,t+h}$, where $\eta_t^{(h)}$ is the pricing error, $\eta_t^{(h)} = E_t(M_{t,t+h}R_{t,t+h}) - 1$, and $\nu_{t,t+h}$ is the innovation, which is uncorrelated with $\eta_t^{(h)}$ by the properties of the conditional expectation. Then, Equation (9) can be re-written as:

$$\sum_{j=1}^{h} Cov(\eta_{t-j}^{(j)} + \nu_{t-j,t}, \eta_t^{(1)}) = 0. \tag{10}$$

In words, rejection of the null implies that either pricing errors are persistent, $Cov(\eta_{t-j}^{(j)}, \eta_t^{(1)}) \neq 0$, or errors are contemporaneously correlated with innovations, $Cov(\nu_{t-j,t}, \eta_t^{(1)}) \neq 0$, or both.

Thus, exploring a model's pricing implications over multiple horizons appears to be a promising avenue. It allows to test for conditional pricing, $E_t(M_{t,t+1}R^i_{t,t+1}) = 1$, using model-implied conditioning information, without the need to specify auxiliary conditioning variables from outside the model. Given that the informational variables z_t and C_t are already implicit in the candidate SDF, $M_{t,t+1}$, MHR-based testing can incorporate whatever conditioning is advocated in the literature.

The next section develops our testing methodology which is applicable to any model that respects the Law of One Price and to any set of test assets. Before we do so, we offer two examples that illustrate the Proposition further.

2.4 Examples

Consider two examples to get a sense of what kind of models can be rejected by the proposed test.

Persistent pricing errors

Suppose that a misspecified one-factor model is

$$\widetilde{M}_{t+1} = 1 - \left(bF_{t,t+1} - E\left[bF_{t,t+1}\right]\right), \quad b = V^{-1}\left(F_{t,t+1}\right) \cdot E\left(F_{t,t+1}\right), \tag{11}$$

where the factor is the excess return to a traded portfolio. The correct model, however, is:

$$M_{t,t+1} = 1 - \left(b_t F_{t,t+1} - E\left[b_t F_{t,t+1}\right]\right)$$

with $b_t = B_0 + B_1 b_{t-1} + u_t$ where u_t is an error term. That is, $F_{t,t+1}$ is only CMVE, not UMVE. The candidate model prices factor returns unconditionally:

$$E\left[\widetilde{M}_{t,t+1} F_{t,t+1}\right] = 0.$$

However, due to the misspecification, we have:

$$E_t\left[\widetilde{M}_{t,t+1} F_{t,t+1}\right] = \left(\frac{1 + bE\left[F_{t,t+1}\right]}{1 + E\left[b_t F_{t,t+1}\right]} b_t - b\right) E_t\left[F_{t,t+1}^2\right].$$

That is, the model does not correctly price the factor conditionally. See Appendix

A.2. The pricing error, $\eta_t^{(1)}$, is persistent since the true b_t is persistent. Thus, $Cov(\eta_{t-j}^{(j)}, \eta_t^{(1)}) \neq 0$. Note that in this case, the timed factor $b_t F_{t,t+1}$ would price assets conditionally and unconditionally.

Correlated pricing errors and innovations

Another example is short-term dependence in returns, as seen in short-term reversal. Again, let the proposed model be as the one in Equation (11). What is different from the previous example is that now the factor returns are i.i.d. Thus, the model prices the factors both conditionally and unconditionally. However, a test asset's returns are not. In particular, consider:

$$R_{t,t+1}^{ie} = \beta_i F_{t,t+1} + \varepsilon_{i,t+1} + \theta \varepsilon_{i,t},$$

where $\varepsilon_{i,t+1}$ is an i.i.d. error term uncorrelated with $F_{t,t+1}$ at all leads and lags. This model represents reversal if $\theta < 0$. The SDF prices $R_{t,t+1}^{ie}$ unconditionally. See Appendix A.3. Thus, a GRS test with $R_{t,t+1}^{e}$ as the test assets would fail to reject this model because all alphas are equal to zero.

However, the model does not correctly price excess returns conditionally:

$$E_t\left[\widetilde{M}_{t,t+1} R_{t,t+1}^{ie}\right] = \theta \varepsilon_{i,t}.$$

See Appendix A.3.

Using the notation in Equation (10), the pricing error $\theta\varepsilon_{i,t} = \eta_t^{(1)}$. Thus, in this example pricing errors are not persistent, so $Cov(\eta_{t-j}^{(j)}, \eta_t^{(1)}) = 0$. However, the one-period pricing error is correlated with the innovation, $Cov(\nu_{t-1,t}, \eta_t^{(1)}) \neq 0$. See Appendix A.3.

To summarize, failure to reject a model using MHR does not imply that the model prices assets conditionally. The model could still have errors with two properties: (i) the errors are not persistent, and (ii) the errors have no contemporaneous correlation with innovations. While, mathematically, it is possible to have a model that is misspecified along these lines, we could not think of any model contemplated in the literature that would match this description. Thus, Proposition 1.2 justifies the use of MHR to test for many important, although formally not all, conditional pricing implications.

3 Testing linear factor models using MHR

3.1 Testing general asset pricing models using MHR

In this section, we develop a GMM-based test using MHR that is applicable to any asset pricing model that satisfies the Law of One Price. As shown in Proposition 1.1, such a model implies Equation (8) for any asset i, which we repeat here for convenience:

$$E(M_{t,t+h}R^i_{t,t+h} - 1) = 0$$

for any asset i and horzion h. This condition can be easily tested jointly for multiple horizons h in a GMM framework.

The proof of the Proposition in section 2.3 demonstrates that these MHR-based moments are equivalent to Equation (9). The Equation implies moment conditions that we ultimately use in our testing. Specifically, for test assets $i = 1, ..., I$

$$f^i_{t+1} = \begin{pmatrix} M_{t,t+1}R^i_{t,t+1} - 1 \\ z^{(h_2)}_{i,t}(M_{t,t+1}R^i_{t,t+1} - 1) \\ \vdots \\ z^{(h_n)}_{i,t}(M_{t,t+1}R^i_{t,t+1} - 1) \end{pmatrix}, \tag{12}$$

where the conditioning variable is

$$z^{(h)}_{i,t} = \sum_{j=1}^{h-1} M_{t-h+j,t}R^i_{t-h+j,t}, \tag{13}$$

n is the number of horizons used in the test, and $\{h_j\}^n_{j=2}$ are the set of horizons used in addition to the single-period horizon. The null hypothesis is $E(f^i_{t+1}) = 0$ for all i, and the test is thus an unconditional test of the conditional properties of the asset pricing model as explained in the second part of the Proposition.

The virtue of the moments in Equation (12) is that the associated residuals are not serially correlated under the null hypothesis. See Appendix A.4. Imposing this additional restriction when estimating the covariance matrix of the moment conditions improves the small-sample properties of the standard errors and test statistics. See Hodrick (1992) for a similar argument in the context of overlapping observations in regressions.[3]

A common approach in the literature is to introduce conditioning information via instrumental variables, such as dividend-price ratios (e.g., Hansen and Singleton, 1982, Hodrick and Zhang, 2001). Denoting such variables by $\tilde{z}_t$, this strategy implies that $\tilde{z}_t(R^i_{t,t+1} - R^f_{t,t+1})$ is just an excess return on another asset. Thus it could be incorporated as a new test asset using the moments outlined in Equation (12). This logic highlights the conceptual difference between the existing and our approaches. While the former relies on exogenously selected conditioning variables, the latter is using those dictated by a given model and set of test assets.

The test falls into the standard GMM framework, where:

$$g(\theta) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} f_t^1(\theta) \\ f_t^2(\theta) \\ \vdots \\ f_t^I(\theta) \end{pmatrix},$$

where θ are the parameters in the SDF to be estimated. The objective function is

[3] Hodrick (1992) emphasizes, in the context of return predictability, that returns could be serially correlated under plausible alternative hypotheses. That prompts him to consider alternative, heteroskedasticity and autocorrelation robust (HAR), standard errors. That consideration is not applicable in our case because the predicted variable is $\eta_t^{(1)}$ in Equation (10). That variable is not serially correlated under an alternative that retains the hypothesis that the model unconditionally prices MHRs. While using HAR standard errors is not incorrect even in this case, imposing the null leads to a more efficient estimate of the test statistic with better small sample properties. For completeness, we report HAR-adjusted results in Appendix A.6.

as usual:

$$\underset{\theta}{\operatorname{argmin}}\, g(\theta)^\top W g(\theta),$$

where W is an $(I \times n) \times (I \times n)$ positive definite weighting matrix (e.g., Hansen and Singleton, 1982). Relevant test statistics and parameter standard errors can be found using the usual GMM toolkit.

3.2 Adopting the general test to linear models

Moment conditions

We slightly re-write the K-factor model in Equation (6) as

$$M_{t,t+1} = 1 - b^\top (F_{t,t+1} - \mu), \tag{14}$$

to emphasize the need to estimate $\mu = E(F_{t,t+1})$. Guaranteeing that this SDF prices the risk-free rate conditionally requires adding auxiliary assumptions that are not explicit in the settings that are traditionally used for testing linear factor models. Because our goal is to assess the original models' performance, we make a slight adjustment to the moment conditions to ensure we do not reject the models based on mispricing of the multi-period risk-free rates, something that they were not designed to match.

Specifically, we note that predicting discounted gross returns, MR^i, as in the covariance condition in Equation (9), is equivalent to predicting discounted excess returns, $M(R^i - R^f)$, if the model prices the risk-free asset. We therefore replace the moment conditions in Equation (12) with the following ones:

$$f^i_{t+1} = \begin{pmatrix} M_{t,t+1}(R^i_{t,t+1} - R^f_{t,t+1}) \\ z^{(h_2)}_{i,t} M_{t,t+1}(R^i_{t,t+1} - R^f_{t,t+1}) \\ \vdots \\ z^{(h_n)}_{i,t} M_{t,t+1}(R^i_{t,t+1} - R^f_{t,t+1}) \end{pmatrix}.$$

The resulting $I \times (n+1)$ GMM moments are:

$$g(b, \mu) = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} F_{t,t+1} - \mu \\ f^1_t(b, \mu) \\ f^2_t(b, \mu) \\ \vdots \\ f^I_t(b, \mu) \end{pmatrix}. \tag{15}$$

Note that the managed portfolio weights $z^{(h)}_{i,t}$ in each f^i are exactly the same as in Equation (13), that is, they still depend on gross returns rather then excess ones.

Test assets

We consider the factors themselves as the set of test assets. A rejection implies that the model does not price assets conditionally, as in the first example of section 2.4. Further, failure to reject does not necessarily imply that the model is well-specified.

A richer cross-section of assets might still reject the model. The reason for this limited choice of test assets is three-fold.

First, the factors in these models are created from mechanical trading strategies to price documented empirical spreads in the cross-section of expected returns. Thus, a natural and minimal requirement for a well-specified model is that the model can price these strategies at any horizon.

Second, it is clear that each model can price the single-horizon excess returns associated with its factors unconditionally. We will in fact estimate b such that the single-horizon returns (SHR) to the factors themselves are priced without error and set μ equal to the factor sample means. We choose the weighting matrix accordingly to ensure these are the only moments used to identify the parameters. That is in line with the standard Black, Jensen, and Scholes (1972) regressions in Equation (7), as the regression imposes the sample mean of the factors in the estimation of α. Thus, any rejection is due to the joint test of the models' pricing of longer-horizon returns.

Third, this choice of test assets implies that there exists an SDF with time-varying loadings b_t as in Equation (4) that does price the factors conditionally and, therefore, prices these factors unconditionally at any horizon per the first part of the Proposition in section 2.3. We discuss this alternative hypothesis in more detail in a later section.

Additional properties associated with linear models

The linear structure of the model allows us to characterize properties of the proposed test more explicitly, which is helpful with interpreting the evidence. Specifically, we can interpret our test as a version of GRS, which corresponds to suitably defined multi-horizon alphas and allows for non-normal and heteroskedastic errors in Equation (7).

In particular, the test asset $k = (i, h)$ is a strategy in factor i, where the time-varying weights are given by lagged h-period discounted returns in factor i, i.e.,

$$R^e_{k,t,t+1} \equiv z^{(h)}_{i,t} F^i_{t,t+1},$$

where the instrument $z^{(h)}_{i,t}$ is endogenous because it depends on the estimated SDF M in Equation (14). Consider the time-series regression

$$R^e_{k,t,t+1} = \alpha_k + \beta_k^\top F_{t,t+1} + \varepsilon_{k,t+1}$$

for each test asset k. In this equation each alpha represents mispricing at some horizon. Collect the alphas into a vector α_z, betas into a matrix β_z, errors into a vector ε_z, and excess returns into a vector R^e_z.

Proposition 2. *Consider the GMM $J-$test corresponding to the moment conditions outlined in Equation (15), with a weighting matrix that sets the factor means equal to the sample return on the factors and one-period pricing errors to zero.*

1. *The test statistic is equal to*

$$
\begin{aligned}
J/T &= E(M_{t,t+1}(R^e_{z,t,t+1} - \beta_z F_{t,t+1}))^\top \times V^{-1}(M_{t,t+1}\varepsilon_{z,t+1}) & (16)\\
&\times \; E(M_{t,t+1}(R^e_{z,t,t+1} - \beta_z F_{t,t+1})) \\
&= \alpha_z^\top \cdot V^{-1}(M_{t,t+1}\varepsilon_{z,t+1}) \cdot \alpha_z. & (17)
\end{aligned}
$$

J is distributed χ^2 with $I \times (n-1)$ degrees of freedom. The $J-$statistic is not affected by the endogeneity of the instruments $z^{(h)}_{i,t}$.

2. *If the squared pricing error ε_z^2 is uncorrelated with the squared SDF, M^2, the test statistic simplifies to*

$$
J/T = \frac{\alpha_z^\top \cdot V^{-1}(\varepsilon_{z,t+1}) \cdot \alpha_z}{1 + \mu_z^\top V^{-1}(F_{t,t+1})\mu_z},
$$

which is identical to the asymptotic GRS test.

See Appendix A.5 for the proof.

As in the GRS test, one can interpret the test statistic as a measure of the mispricing of a portfolio that is orthogonal to the original factors.

Corollary (to Proposition 2)**.** *Consider the portfolio with excess return*

$$
\begin{aligned}
R^{*e}_{t,t+1} &\equiv \alpha_z^\top V^{-1}(M_{t,t+1}\varepsilon_{z,t+1})(R^e_{z,t,t+1} - \beta_z F_{t,t+1}) & (18)\\
&= \alpha_z^\top V^{-1}(M_{t,t+1}\varepsilon_{z,t+1})(\alpha_z + \varepsilon_{z,t+1}).
\end{aligned}
$$

1. *Using this portfolio as the only multihorizon test asset leads to the same $J-$statistic.*

2. *Under the simplifying assumption of Proposition 2.2, the squared SR of this portfolio is equal to the squared maximal Information ratio (IR), $\alpha_z^\top V^{-1}(\varepsilon_{z,t+1})\alpha_z$.*

See Appendix A.5 for the proof. The IR of this portfolio gives a direct measure of the degree of mispricing in a given model. Decomposing this portfolio into the contributions of individual factors characterizes how the factors should be timed to exploit the model misspecification.

Lastly, we discuss whether our test relies on tradeable portfolios. Our initial moment conditions, with their general form given in Equation (8), are tests of whether a candidate SDF prices returns across multiple horizons. Because all the factors we are looking at are tradeable, the longer-run buy-and-hold returns to these factors are tradeable as well. There is therefore no look-ahead bias in the test assets, even though the parameters in the SDF are, as is usually the case, estimated over the full sample.

The interpretation of the MHR moment conditions as managed portfolios in one-period returns is convenient for intuition about the results and helps in formulating moment conditions that are not correlated over time under the null hypothesis. At the same time, this one-period interpretation may suggest that the test is providing specific real-time implementable timing strategies, embedded in the endogenous instruments $z_{i,t}^{(h)}$. These endogenous timing variables are the product of lagged returns and SDFs, where the latter is a function of parameters estimated over the full sample. It is, therefore, interesting to see if the test results survive using managed portfolios that are tradeable in real time. In Appendix A.7, we report results for the case of rolling estimation of the SDF parameters.

4 Evidence

4.1 Models and data

We select our models based on their historical importance, recent advancements, and data availability. Specifically, we include the unconditional CAPM, CAPM combined with the BAB factor (Frazzini and Pedersen, 2014, Black, Jensen, and Scholes, 1972, Novy-Marx and Velikov, 2016), Fama and French five-factor model, FF5, (Fama and French, 2015), a version of the FF5 model with hedged unpriced risks (Daniel, Mota, Rottke, and Santos, 2020), Fama and French three-factor model with momentum (Carhart, 1997, Fama and French, 1993), the four-factor models of Hou, Xue, and Zhang (2015) and Stambaugh and Yu (2017), and the six-factor model of Haddad, Kozak, and Santosh (2020).

The Fama-French five-factor model includes the market factor (MKT), the value factor (HML), the size factor (SMB), the profitability factor (RMW; see also Novy-Marx, 2013), and the investment factor (CMA; see also Cooper, Gulen, and Schill, 2008). These data and the momentum factor MOM (Jegadeesh and Titman, 1993) are provided on Kenneth French's webpage. The returns are monthly and the sample is from July 1963 to June 2017.

The hedged versions of these factors studied by Daniel, Mota, Rottke, and Santos (2020) (DMRS) are available on Kent Daniel's webpage. The sample period is July 1963 to June 2017. The factors studied by Hou, Xue, and Zhang (2015) (HXZ) are MKT, SMB, I/A (investment-to-assets) and ROE (return on equity), and are available on Lu Zhang's website. The sample is from January 1967 to December 2017. Stambaugh and Yu (2017) propose two factors intended to capture stock mispricing, in addition to the existing MKT and SMB factors: PERF and MGMT. We denote this four-factor model as SY. These data are available on Robert Stambaugh's webpage. The sample period for these factors starts January 1963 and ends December 2016. Haddad, Kozak, and Santosh (2020) (HKS) propose, in addition to MKT, factors that are the first five principal components (PC1-5) of fifty anomaly portfolios that are entertained in the literature. Their sample period is January 1974 to December 2017.

Given the recent critique by Novy-Marx and Velikov (2016), we depart from the BAB factor construction of Frazzini and Pedersen (2014). We use the value-weighted beta- and size-sorted portfolios on Kenneth French's webpage as the building blocks for constructing this factor, following Fama and French (2015) and Novy-Marx and Velikov (2016). See Appendix A.8.

Finally, we get the monthly risk-free rate from CRSP and create the real risk-free rate by subtracting realized monthly inflation from the nominal rate. The inflation data are from CRSP as well.

Before we proceed with the discussion of results, we emphasize again that each model is tested using MHR on its own factors. For example, the CAPM is tested using the market return at various horizons, and the FF5 model is tested using the returns to each of its five factors at various horizons. While we test the conditional pricing of the models, we cannot readily compare the results across them because the test assets vary. As a robustness exercise, we consider a common set of test assets across the different models. Specifically, we consider the five factors from the FF5 model. See Appendix A.9.

4.2 MHR pricing errors and model tests

In the tests we use the horizons 3, 6, 12, 24, and 48 months in addition to the one-period (monthly) horizon. Because the evaluated factors are designed as zero-investment long-short portfolios, we construct $R^i = R^f + F^i$ for each factor i when evaluating $z_{i,t}^{(h)}$ in Equation (13).

We start by computing pricing errors for each factor in each model across horizons. The pricing errors should be understood as the net present value of an h-period \$1 buy-and-hold investment in the gross factor return. Since the models are estimated to match one-period returns unconditionally, non-zero net present values are due to mispricing of the conditional factor return. To facilitate comparison we annualize errors so that each reported number reflects the same period irrespective of the horizon. Thus, the pricing error for a factor F^i at horizon h is $12/h \times E(z_{i,t}^{(h)} M_{t,t+1} F_{t,t+1}^i)$. The non-annualized version of these pricing errors are equal to the corresponding elements of α_z appearing in the Corollary to Proposition 2 as our estimation sets $E(M_{t,t+1}) = 1$.

The horizons for reported errors range from 1 to 48 months. Figure 1 displays the pricing errors for the first four factor models. The top left panel shows that the pricing errors of the CAPM are small across horizons, always less than 1% annualized. Thus, for the market model, a constant b coefficient in the SDF works reasonably well for pricing market returns across these horizons.

The top right panel shows the MKT+BAB model. In this case, pricing errors are much larger for both factors. For the BAB factor, the annualized pricing error increases with horizon (in absolute value) to almost 10% per year at the 48-month horizon. That is about twice the average annualized monthly returns on this factor.

The bottom left plot shows the Carhart model (FF3+MOM), where the pricing errors get very large, exceeding 50% p.a. for the 4-year MOM return. The bottom right panel shows the corresponding pricing errors for the FF5 model. Again pricing errors increase in absolute value with horizon. Three of the five factors (MKT, RMW, and CMA) have absolute pricing errors in excess of 5% p.a. at the 4-year horizon.

Panel A of Table 1 gives the p-values of the J-test of these models. The test fails to reject the CAPM. MKT+BAB and FF3+MOM are rejected at the 10% level. Lastly, FF5 is rejected with a $p-$value of 0.02. We calculate the mean absolute pricing error

(MAPE) for each model as the mean of the absolute value of the annualized pricing errors across the factors and horizons. For the CAPM, the MAPE is only 0.7%, for the CAPM+BAB it is 3%, for the Carhart model it is 7.6%, and for the FF5 it is 1.9%.

Figure 2 shows the pricing errors for the remaining four models. The top left panel shows pricing errors for the FF5_{DMRS} model. Its pricing errors exceed 10% p.a. for two factors (their versions of the MKT and SMB factors) and 5% for their version of the CMA factor. The remaining plots show the pricing errors for the SY, HXZ, and HKS models. For these models, pricing errors are even larger, with the largest pricing error exceeding 100% p.a. (PERF in the SY model). As p-values in Panel B of Table 1 indicate, all these models are rejected with high levels of confidence.

Overall, Table 1 shows the average MAPE across all eight models is 4.4%, which is about the same as the annualized factor risk premiums that these models were originally designed to match. We conclude from this that the current benchmark models for risk-adjustment do a poor job accounting for MHRs. For all models but the CAPM, the MHR to the models' own factors have power to detect model misspecification. Economically, requiring that a model is able to price its own factors at various horizons is a minimal requirement of model consistency. Failing this requirement implies that the factor dynamics are not consistent with the constant b assumption.

In robustness checks reported in Appendixes A.6 (HAR standard errors) and A.9 (the same test assets across models) we reject all the eight models. One exception is the CAPM that is not rejected even with HAR standard errors when using MHR to MKT only. The test with out-of-sample instrument estimation in Appendix A.7 rejects FF3+MOM, $FF5_{DMRS}$, SY, and HKS.

4.3 Sharpe and Information ratios

Table 1 also displays the annualized maximal SR, $[E(F)^{\top}V(F)^{-1}E(F)]^{1/2}$, implied by each factor model. A higher SR suggests that the factors are closer to spanning the unconditional MVE portfolio. As is well-known, the SR of the MKT factor is much lower than the maximal SR in more recent multi-factor models. For instance, the SY model has an annualized SR of 1.7 compared to 0.4 for the CAPM.

Finally, Table 1 reports the annualized maximal IR for each model as an alternative economic measure of the mispricing implied by the MHR, as motivated by the Corollary to Proposition 2. The models' IRs are economically large, ranging from 0.61 to 1.14 for the rejected models, similar in magnitude to the model's maximal SR. That is, timing of the model factors, as implied by the MHR-based instruments, yields high ex-post Sharpe ratios of strategies that are orthogonal to the original factors.

We can gain more insight into the sources of mispricing by considering the IRs of individual timing strategies. This is standard practice in the typical cross-sectional tests, where one looks at the individual alphas and SRs of particular portfolios (e.g., value has a positive alpha and high SR while growth has negative alpha and low SR). The difference is that in the case of the MHR test a given portfolio is a horizon-based trading strategy for a particular factor. We focus on IRs instead of alphas so that the magnitudes are easily comparable across horizons.

We select one factor per model for the clarity of exposition by simply picking the factor with the largest average pricing error across horizons: MKT from the CAPM, BAB from MKT+BAB, MOM from FF3+MOM, RMW from FF5, SMB_{DMRS} from FF5_{DMRS}, PERF from SY, ROE from HXZ, and PC2 from HKS. Figures 3 and 4 show the annualized IRs of the factor's timing strategies for each horizon.

The top left plot of Figure 3 shows the IRs of timed versions of the MKT factor in the CAPM. This case is particularly interesting because MKT is the single factor and test asset in this model. Each bar in the chart corresponds to timing based on the instruments for each horizon in our test (3, 6, 12, 24, and 48 months). The IR of the 3-month timing factor is -0.3. As Equation (9) implies, one can thus increase SR by increasing the exposure to MKT when the lagged 2-month risk-adjusted returns, $M_{t-2,t}R_{t-2,t}$, are low. The 6-month timing strategy has a similar IR, while IRs corresponding to longer horizons are negative and decreasing in absolute value. Thus, while the mispricing seems strongest at shorter horizons, all horizons exhibit reversal in risk-adjusted returns.

The test does not reject the CAPM despite the large absolute values of the IRs. To understand why, recall that the joint test in Equation (17) takes into account the correlation between the test assets. If the IRs (alphas) line up with the correlation structure of the test assets, the probability that the pattern of sample alphas are due to a single shock increases and the joint test is less likely to reject.

Specifically, in the case of the CAPM the test actually rejects if we only use the 3-month horizon (the $p-$value is then 0.030). In the general test, the correlations between the returns to the 3-month strategy and the other horizon strategies are positive and declining in the horizon in a similar fashion to how the absolute values of the IRs are declining.[4] That suggests that the alphas (IRs) arise due to a common shock.

As another example, MOM is the factor with the largest average pricing errors in the FF3+MOM model. Its IRs are displayed in the bottom left plot of Figure 3. The term corresponding to the strategy with $h = 3$ months has a positive IR and indicates that one can increase SRs by increasing the exposure to MOM when the lagged 2-month risk-adjusted returns are high, which is opposite to the MKT in the CAPM. However, at the 48-month horizon the IR is negative. This implies that if risk-adjusted momentum did well over the last 47 months, it will likely do poorly the following month.

Besides the sign, IRs differ in magnitude as horizon changes. For instance, the bottom left plot in Figure 4 shows the IRs for the ROE factor in the HXZ model. In this case, there appears to be valuable timing information in the short horizon (2-month) and long-horizon (47-month) instruments, but not in the intermediate horizons. The fact that both short- and long-run discounted returns are important suggests persistent dynamics in b_t are needed at both short- and long-run frequencies.

[4]We use the correlations to get predicted IRs relative to the 3$-$month IR via $\rho_{3,h} \times IR_3$, where h is the other horizons and $\rho_{3,h}$ is the correlation in realized pricing errors between the 3-month timing and the h-month timing. In our sample the predicted IRs of -0.26, -0.21, -0.16, and -0.13 for horizons 6, 12, 24, and 48, respectively. These predicted values are quite close to the estimated IRs reported in the Figure.

Overall, many factors have negative IRs across horizons. That indicates reversal in risk-adjusted returns per Equation (9). These reversal patterns play out at both long and short horizons. The size of the IRs indicates how strong the timing effect is at a particular instrument's horizon. In terms of the implications for the optimal b_t, a reversal in risk-adjusted returns suggests that b_t ought to be lower (higher) after high (low) risk-adjusted returns. Said differently, since b_t governs the conditional risk-return trade-off, these results suggest that the conditional Sharpe ratio of these factors tend to go down when lagged risk-adjusted returns are high. Positive IRs (such as those of PERF from SY, or PC2 from HKS) have the opposite implications as these result from momentum in risk-adjusted returns.

While these plots are informative about the nature of optimal timing activity, a caveat is that they ignore cross-correlations that show up in the test statistic and that are also important for the model rejections. One way to address this additional dimension of the data is to inspect the implied portfolio weights of the rejecting portfolio, $\alpha_z^\top \cdot V^{-1}(M_{t,t+1}\varepsilon_{z,t+1})$. We leave this for future research as in-depth analysis of the nature of timing across these eight models and thirty-one factors is beyond the scope of this paper.

Figure 5 shows each model's MAPE plotted against the respective maximal SRs Interestingly, there is a positive relation. The higher a model's SR the closer it should be to spanning the unconditionally mean-variance efficient portfolio and thus the lower the pricing errors should be. The opposite being the case indicates that the search for high SR models has increased the complexity of the conditional dynamics, consistent with the findings in Haddad, Kozak, and Santosh (2020). Such dynamics have received relatively little attention in the literature as researchers have focused on short-term average returns and Sharpe ratios.

4.4 Factor dynamics and long-horizon investment

The null hypothesis that the SDF in Equation (6) is correctly specified implies that conditional mean of factor returns is proportional to their conditional second moment. See Appendix A.1. That our test rejects the models implies that this requirement does not hold in the data.

In order to gain further intuition about the rejection results, we evaluate both statistical and economic metrics that are relevant for long-horizon investors. We compare the long-term properties of actual asset returns to those of artificial returns that are generated under the null hypothesis. For illustration purposes, we focus on one factor per model that corresponds to the largest pricing error, just like in the previous subsection.

We construct artificial returns in three steps. First, we estimate the conditional second moment of actual returns via the DCC-GARCH(1,1) model. Second, we construct the model-implied conditional mean exploiting the proportionality of the two under the null. Third, we construct artificial return shocks by resampling with replacement the unconditionally de-meaned returns, descaling the resulting series by their sample standard deviation, then rescaling them by the standard deviations from the previously estimated DCC-GARCH(1,1), and, lastly, adding the model-implied conditional mean to the result. The resampling also offers us the means of computing confidence bands under the null hypothesis.

Statistical assessment

One way to illustrate long-term dynamics of returns is to compare cumulative serial correlations of asset returns across multiple horizons. Figures 6 and 7 display such serial correlation for the chosen factor in each model. We also report confidence bands bootstrapped under the null.

The market factor in the CAPM displays almost no departures from the null for horizons up to 20 months, becoming slightly lower afterwards, with the difference peaking at 0.2 at the horizon of 36 months. The subsequent decline in the difference is consistent with a long-run mean-reverting component in market returns.

All the other models display markedly stronger departures from the constant b baseline. To start with the most extreme cases, the BAB factor in MKT+BAB and the SMB_{DMRS} factor FF5_{DMRS} have serial correlations that are steeply and significantly departing from the null, with the difference exceeding 0.6 at 12 to 24 months. For the remaining models the differences between cumulative serial correlations peak between 0.3 and 0.5. The difference is insignificant for the MOM factor in FF3+MOM.

These observations are suggestive of the potential difficulty of spanning UMVE statically. Clearly, there are strong persistent components in the factor return dynamics that are inconsistent with a constant b.

Economic assessment

We offer another perspective on this conclusion by assessing the economic impact of these dynamics for long-horizon investors. The Sharpe ratio of investing in a factor across different investment horizons is affected by the autocorrelation of factor returns, as autocorrelation determines how return variance grows with horizon. We calculate annualized SRs by horizon for the same chosen factors as before, both in the data and as implied by the model. Excess returns at horizon h is the average h-period gross return minus the h-period gross risk-free rate, where both of these are calculated by multiplying together h one-period gross returns. The SR is then the mean excess return divided by the standard deviation of excess returns. We annualize by multiplying with $\sqrt{12/h}$.

Figures 8 and 9 display the SR and their benchmarks as implied by the respective models, accompanied by the confidence bands bootstrapped under the null. In the case of the CAPM, the SR barely changes with horizon, and there is no difference between the data and the null. That is consistent with all the evidence presented on the CAPM heretofore: MHR to MKT do not reject the model, MKT has relatively small MAPE across horizons, and cumulative autocorrelation of MKT returns in the data does not deviate much from that under the null.

The difference between the model-implied and empirical SRs across all other models is expanding with the horizon (HKS is the exception as the difference starts declining after 24 months). Consistent with the autocorrelation plots in Figures 6 and 7, the Sharpe ratios are too low (high) in the data if the factor's autocorrelations are too high (low). Thus, the misspecified dynamics as implied by the constant b models affects inference about long-run investment performance. Just to highlight some statistically significant magnitudes, in the case of FF5 the 48-month SR under the null is 0.2, while it is 0.4 in the data; for SY the numbers are 0.35 and 0.6, respectively; for HXZ they are 0.45 and 0.75; and for HKS they are 0.3 and 0.55 (at 20 months in this case).

The difference between SRs under the null and in the data for other models (MKT+BAB, FF3+MOM, $\mathrm{FF5}_{DMRS}$) are insignificant. Testing SRs of individual factors in a model is less powerful than the MHR test, which weighs all the factors in a given model and does so optimally. The main purpose of this section is to illustrate economic implications of the misspecified factor dynamics associated with the null hypothesis.

4.5 Towards improvement of pricing MHR to factors

The model rejections is a consequence of factor dynamics unaccounted for in the linear SDF specification. In this section, we consider these dynamics. Full accounting for the uncovered role of dynamics and proposing a convincing alternative to each model is beyond the scope of this paper. We have a more modest objective of providing an illustration of what recognizing the dynamic properties of the factors might entail and to suggest a path for future research.

Per part one of Proposition 1, an SDF that prices the factors conditionally also prices MHR to the factors unconditionally. Following the discussion in Sections 2.1

and 2.2, the UMVE portfolio formed by dynamic trading in the factors prices them both conditionally and unconditionally, as in the SDF in Equation (4), where the optimal timing vector b_t is a function of the conditional means and covariance matrix of the factors as in Equation (5).

We now turn to most recent research on the topic of factor timing. Specifically, Haddad, Kozak, and Santosh (2020) and Moreira and Muir (2017) propose factor timing approaches, which are both out-of-sample and rely on estimates of $E_t(F_{t,t+1})$ and $V_t(F_{t,t+1})$ as prescribed by Equation (5). As discussed in Section 2.1, both their approaches can be represented in the form of Equation (3). Moreira and Muir (2017) time volatility via $b_t^i = b^i V_t^{-1}(F_{t,t+1}^i)$, which is estimated using squared realized daily factor returns. HKS use $b_t^i = b^i E_t(F_{t,t+1}^i)$, where out-of-sample conditional expectations are constructed for their factors using each factor's value spread. HKS also contemplate a version with $b_t^i = b^i E_t(F_{t,t+1}^i) V_t^{-1}(F_{t,t+1}^i)$ with V_t estimated in-sample. In all cases, we estimate the constants of proportionality b^i for each factor i by matching the in-sample average returns to the timed factors in the model at hand, analogous to how we estimated the constant b vector in the baseline models.

Under the null of correct conditional pricing of the factors $F_{t,t+1}$, the timed combination of the factors that is the UMVE portfolio $b_t^\top F_{t,t+1}$ should be priced as well. Thus, we include this model-implied timed portfolio into the set of test assets in our MHR-based test, in addition to the original factors. This is an important step because in a model with a time-varying b_t, the UMVE portfolio embodies conditional information not captured by the factors themselves.

Panels A and B in Table 2 reports testing results. Because of the specifics of the methodologies employed in these papers, we cannot test all the models that we have considered heretofore. In the case of Volatility Timing, we report CAPM, FF3+MOM, and FF5. In the case of HKS, we report two versions of the model: out-of-sample timing with E_t only, and hybrid timing with the same E_t and in-sample V_t. All models are rejected with economically large MAPEs for the MHR to the model factors.

For the Volatility Timing case, the SRs of the timing portfolio are higher than those in the constant b case, but the MAPEs are still large. Thus, volatility timing is insufficient for spanning the UMVE correctly. The case of the CAPM is the most striking as volatility timing appears to make matters strictly worse. While the evidence in Moreira and Muir (2017) is clear that such timing improves single-horizon returns, it is better to stay with the original factor if one is interested in MHR. This conclusion is a flipside of our discussion of the baseline test in Table 1: the original factor is rejected by the SHR test, but not by the MHR one.

For the HKS factor timing results, we note that the sample is only the last half of the original sample, due to the estimation of out-of-sample means in HKS. Thus, a direct comparison to the corresponding constant b model in Table 1 cannot be made. That said, the MAPEs are economically large, more than 5% annualized on average per factor. For all of these timing strategies the Information ratios are still large, all above 1.

In Appendix A.10 we introduce our own version of factor timing for the 8 models we considered heretofore. We do so in-sample to simplify the task. We still reject all the models. Overall, the MAPEs and the Information ratios are still large even when considering models that attempt to account for factor timing. While factor timing does tend to increase the maximal SR of the model, this is a one-period metric. Testing the model with MHR, however, brings in moments that are particularly sensitive to persistent misspecification in the conditional factor dynamics. This new perspective indicates that there is substantial work to do in improving estimates of these dynamics.

As we pointed out in section 2.2, yet another alternative to improving a linear factor model is to change the factors themselves rather than their timing. Kozak, Nagel, and Santosh (2020) propose just that by introducing weights on all CRSP stock returns that are non-linear functions of 50 characteristics C_t. The end product is a one-factor model where the factor is an estimate of the UMVE portfolio. They advocate using both what they term sparsity (L^1-norm) and shrinkage (combining L^1- and L^2-norms) penalties in selecting portfolio weights. Because the authors have more characteristics than time periods at the monthly frequency, they opt for estimation of the daily SDF to improve precision.

We evaluate the model using MHR on the UMVE portfolio associated with the proposed SDF. We normalize the volatility of this candidate UMVE portfolio to equal that of the market factor so the MAPE is comparable. We use horizons of 3, 6, 12, 24, and 48 days rather than months to be consistent with the Kozak, Nagel, and Santosh (2020) observation frequency. Panel C of Table 2 reports the results. The model is rejected with MAPEs of 3.4% and 6.6%, indicating that the factor construction even in this case does not pay sufficient attention to factor dynamics.

5 Conclusion

We propose a new unconditional test of conditional asset-pricing models. It is GMM-based and uses multi-horizon returns (MHR) to evaluate the ability of any model to price risky cash flows that accrue at different horizons. We argue that MHR are appealing because they effectively provide a set of test assets that are endogenous to the model being tested and because these assets identify a broad set of conditional model misspecification. Thus, the test does not require any conditioning variables beyond those used in the construction of a model.

Our empirical exercise, involving a number of prominent linear factor models, suggests that our test has statistical power. We reject most of these models, including recently proposed models that explicitly incorporate factor timing in the model construction. The associated pricing errors and Information ratios are economically large.

The reason the models do a poor job pricing longer-horizon returns is that the model-implied conditional properties of risk pricing are strongly at odds with dynamic properties of the factors associated with these models. Because long-run investment entails exposure to conditional return dynamics even in the absence of factor timing, these dynamics show up as large mispricing in longer-run returns.

Many economic applications, such as capital budgeting and consumption-savings decisions, require discounting cash flows accruing at multiple horizons. Our evidence suggests that there is still much to be done to arrive at models that can successfully be applied to such problems. In particular, correct specification of the joint conditional dynamics of pricing factors appears even more quantitatively important than previously emphasized in the literature.

References

Baba Yara, Fahiz, Martijn Boons, and Andrea Tamoni, 2020, New and old characteristic-sorted portfolios: Implications for asset pricing, working paper.

Backus, David K., Nina Boyarchenko, and Mikhail Chernov, 2018, Term structures of asset prices and returns, *Journal of Financial Economics* 129, 1–23.

Barillas, Fancisco, and Jay Shanken, 2017, Which alpha?, *The Review of Financial Studies* 30, 1316–1338.

Belloni, Alexandre, and Victor Chernozhukov, 2013, Least squares after model selection in high-dimensional sparse models, *Bernoulli* 19, 521–547.

Belo, Frederico, Pierre Collin-Dufresne, and Robert Goldstein, 2015, Dividend dynamics and the term structure of dividend strips, *Journal of Finance* 70, 1115–1160.

Bessembinder, Hendrik, Michael J. Cooper, and Feng Zhang, 2020, Mutual fund performance at long horizons, working paper.

Black, Fischer, Michael Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger Publishers Inc.).

Boguth, Oliver, Murray Carlson, Adlai Fisher, and Mikhail Simutin, 2011, Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas, *Journal of Financial Economics* 102, 363–389.

Bollerslev, Tim, 1990, Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *Review of Economics and statistics* 72, 498–505.

Brennan, Michael, and Yuzhao Zhang, 2018, Capital asset pricing with a stochastic horizon, working paper.

Carhart, Mark, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Cochrane, John, 2004, *Asset Pricing* (Princeton University Press: Princeton, NJ) revised edn.

Cooper, Michael, Huseyn Gulen, and Michael Schill, 2008, Asset growth and the cross-section of stock returns, *Journal of Finance* 63, 1609–1651.

Dahlquist, Magnus, and Henrik Hasseltoft, 2013, International bond risk premia, *Journal of International Economics* 90, 17–32.

Daniel, Kent, and David Marshall, 1997, Equity-premium and risk-free-rate puzzles at long horizons, *Macroeconomic Dynamics*.

Daniel, Kent, Lira Mota, Simon Rottke, and Tano Santos, 2020, The cross section of risk and return, *Review of Financial Studies* 33, 1927–1979.

Daniel, Kent, and Sheridan Titman, 2012, Testing factor-model explanations of market anomalies, *Critical Finance Review* 1, 103–139.

Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez, 2015, The price of variance risk, working paper.

Fama, Eugene, and Kenneth French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

——— , 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

——— , 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Farnsworth, Heber, Wayne Ferson, David Jackson, and Steven Todd, 2002, Performance evaluation with stochastic discount factors, *Journal of Business* 75, 473–503.

Favero, Carlo, Alessandro Melone, and Andrea Tamoni, 2020, Factor models with drifting prices, working paper.

Ferson, Wayne, and Andrew Siegel, 2001, The efficient use of conditioning information in portfolios, *Journal of Finance* 56, 967–982.

Ferson, Wayne E., and Campbell R. Harvey, 1999, Conditioning variables and the cross-section of stock returns, *Journal of Finance* 54, 1325–1360.

Frazzini, Andrea, and Lasse Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Ghysels, Eric, 1998, On stable factor structures in the pricing of risk: Do time-varying betas help or hurt?, *Journal of Finance* 53, 549–573.

Gibbons, Michael, Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 59, 1121–1152.

Grossman, Sanford, Angelo Melino, and Robert Shiller, 1987, Estimating the continuous-time consumption-based asset pricing model, *Journal of Business and Economic Statistics* 5, 315–327.

Haddad, Valentin, Serhiy Kozak, and Shri Santosh, 2020, Factor timing, *Review of Financial Studies* 33, 1980–2018.

Hansen, Lars, John Heaton, and Nan Li, 2008, Consumption strikes back? Measuring long-run risk, *Journal of Political Economy* 116, 260 – 302.

Hansen, Lars, and Scott Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269–1286.

Hodrick, Robert, and Xiaoyan Zhang, 2001, Evaluating the specification errors of asset pricing models, *Journal of Financial Economics* 62, 327–376.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *The Review of Financial Studies* 5, 357?386.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *The Review of Financial Studies* 28, 650–705.

Jagannathan, Ravi, and Yong Wang, 2007, Lazy investors, discretionary consumption, and the cross-section of stock returns, *Journal of Finance* 62, 1623–1661.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional capm and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Kamara, Avraham, Robert A. Korajczyk, Xiaoxia Lou, and Ronnie Sadka, 2016, Horizon pricing, *Journal of Financial and Quantitative Analysis* 51, 1769–1793.

Kelly, Bryan, Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and returns, *Journal of Financial Economics* 134, 501–524.

Koijen, Ralph, Hanno Lustig, and Stijn Van Nieuwerburgh, 2017, The cross-section and time series of stock and bond returns, *Journal of Monetary Economics* 88, 50–69.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2018, Interpreting factor models, *Journal of Finance* 73, 1183–1223.

——— , 2020, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

Lettau, Martin, and Sydney Ludvigson, 2001, Resurrecting the (c)capm: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Levhari, David, and Haim Levy, 1977, The capital asset pricing model and the investment horizon, *Review of Economic Studies* 59, 92–104.

Lewellen, Jonathan, and Stefan Nagel, 2006, The conditional capm does not explain asset pricing anomalies, *Journal of Financial Economics* 79, 289–314.

——— , and Jay Shanken, 2010, A skeptical appraisal of asset-pricing tests, *Journal of Financial Economics* 96, 175–194.

Linnainmaa, Juhani, and Sina Ehsani, 2019, Factor momentum and the momentum factor, working paper.

Lo, Andrew, and Craig A. MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *The Review of Financial Studies* 3, 431–467.

Longstaff, Francis, 1989, Temporal aggregation and the continuous-time capital asset pricing model, *Journal of Finance* 44, 871–887.

Lustig, Hanno, Andreas Stathopoulos, and Adrien Verdelhan, 2013, The term structure of currency carry trade risk premia, working paper.

MacKinlay, A. Craig, 1995, Multifactor models do not explain deviations from the capm, *Journal of Financial Economics* 38, 3–28.

Moreira, Alan, and Tyler Muir, 2017, Volatility-managed portfolios, *Journal of Finance* 72, 1611–1644.

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

——— , and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *The Review of Financial Studies* 29, 104–147.

Parker, Jonathan, and Christian Julliard, 2005, Consumption risk and the cross-section of expected returns, *Journal of Political Economy* 113, 185–222.

Stambaugh, Robert, and Yuan Yu, 2017, Mispricing factors, *The Review of Financial Studies* 30, 1270–1315.

van Binsbergen, Jules, Michael Brandt, and Ralph Koijen, 2012, On the timing and pricing of dividends, *American Economic Review* 102, 1596–1618.

Zviadadze, Irina, 2017, Term-structure of consumption risk premia in the cross-section of currency returns, *Journal of Finance* 72, 1529–1566.

A Appendix

A.1 MVE portfolios and the SDF

Portfolios with the maximal time t conditional Sharpe ratio are conditionally mean-variance efficient (CMVE) and given by

$$\begin{aligned} R_{t,t+1}^{C} &= (w_t^C)^\top R_{t,t+1}^e, \\ w_t^C &= k_t^{-1} V_t^{-1}\left(R_{t,t+1}^e\right) E_t\left(R_{t,t+1}^e\right), \end{aligned}$$

where w_t^C is the vector of time t portfolio weights and k_t is any positive constant known at time t, governing the leverage of the portfolio. The CMVE "prices" any combination of these assets in the sense that there exists an SDF, $M_{t,t+1}^*$, derived from the CMVE that prices $R_{t,t+1}^e$ conditionally and unconditionally, $E(M_{t,t+1}^* R_{t,t+1}^e) = E_t(M_{t,t+1}^* R_{t,t+1}^e) = 0$. Specifically,

$$\begin{aligned} M_{t,t+1}^* &= 1 - k_t(R_{t,t+1}^C - E_t(R_{t,t+1}^C)), \\ k_t &= V_t^{-1}\left(R_{t,t+1}^C\right) E_t\left(R_{t,t+1}^C\right). \end{aligned}$$

This SDF is only conditionally linear in the CMVE portfolio, as k_t and $E_t(R_{t,t+1}^C)$ are time-varying, and therefore does not directly imply an unconditional, linear beta-pricing model. To that end, note that if $E_t(M_{t,t+1}^* R_{t,t+1}^e) = 0$, then $E_t(a_t \cdot M_{t,t+1}^* R_{t,t+1}^e) = 0$, if a_t is known at time t. Thus, dividing $M_{t,t+1}^*$ by $1 + k_t E_t(R_{t,t+1}^C)$ we obtain another SDF that prices the same set of assets:

$$\begin{aligned} \widetilde{M}_{t+1} &= 1 - \frac{k_t}{1 + k_t E_t\left(R_{t,t+1}^C\right)} R_{t,t+1}^C \\ &= 1 - \frac{k_t}{1 + E_t\left(R_{t,t+1}^e\right)^\top V_t\left(R_{t,t+1}^e\right)^{-1} E_t\left(R_{t,t+1}^e\right)} R_{t,t+1}^C \\ &= 1 - \delta_t R_{t,t+1}^C. \end{aligned}$$

Next, divide $\widetilde{M}_{t+1}$ by the constant $1 - E(\delta_t R_{t,t+1}^C)$ to get the final version of the SDF that still prices the same set of assets conditionally and unconditionally:

$$M_{t,t+1} = 1 - \left(R_{t,t+1}^U - E(R_{t,t+1}^U)\right),$$

where $R_{t,t+1}^U \equiv \delta_t R_{t,t+1}^C / (1 - E(\delta_t R_{t,t+1}^C))$ is the excess return to timing the CMVE portfolio in a way that renders it *unconditionally* mean variance efficient (UMVE). The latter is true as its returns have a perfectly negative unconditional correlation with $M_{t,t+1}$. Letting $k = (1 - E(\delta_t R_{t,t+1}^C))^{-1}$, the associated portfolio weights are:

$$\begin{aligned} k w_t^U &= k \delta_t w_t^C \\ &= k \frac{V_t^{-1}\left(R_{t,t+1}^e\right) E_t\left(R_{t,t+1}^e\right)}{1 + E_t\left(R_{t,t+1}^e\right)^\top V_t^{-1}\left(R_{t,t+1}^e\right) E_t\left(R_{t,t+1}^e\right)}. \end{aligned}$$

The common model specification in the literature assumes there is a set of K factors, $F_{t,t+1}$, that unconditionally span the UMVE portfolio: $R^U_{t,t+1} = b^\top F_{t,t+1}$. Thus, the SDF can be written

$$M_{t,t+1} = 1 - b^\top \left(F_{t,t+1} - E(F_{t,t+1})\right).$$

Note that this assumption has implications for factor dynamics. In particular, using the above expression for UMVE weights, setting these to a constant vector b, and substituting in the factors for the vector of base assets, we have:

$$\begin{aligned} b &\propto \frac{V_t^{-1}\left(F_{t,t+1}\right) E_t\left(F_{t,t+1}\right)}{1+\theta_t^F} \iff \\ E_t\left(F_{t,t+1}\right) &\propto V_t\left(F_{t,t+1}\right)\left(1+\theta_t^F\right) \times b, \end{aligned}$$

where $\theta_t^F = E_t\left(F_{t,t+1}\right)^\top V_t^{-1}\left(F_{t,t+1}\right) E_t\left(F_{t,t+1}\right)$ is the maximal squared conditional Sharpe ratio possible from combining the factor returns. If the frequency of the data is high, $1+\theta_t^F$ is close to one and the conditional factor means are approximately proportional to their conditional variance. Applying the Sherman-Morrison formula to the above expression, one can show that $E_t\left(F_{t,t+1}\right) \propto E_t\left(F_{t,t+1}F_{t,t+1}^\top\right) \times b$. In fact, this latter result is easiest found by applying the conditional LOOP to the factors themselves:

$$\begin{aligned} E_t(M_{t,t+1}F_{t,t+1}^\top) &= 0 \iff \\ E_t(F_{t,t+1}^\top) &= \frac{b^\top}{1+b^\top E(F_{t,t+1})} E_t(F_{t,t+1}F_{t,t+1}^\top). \end{aligned}$$

In words, if the factors span the UMVE portfolio, the conditional expected factor returns are proportional to the conditional second moment of factor returns. If the frequency of the data is high, the conditional factor means are approximately proportional to the conditional variance.

A.2 Misspecified model with persistent errors

We have:

$$\begin{aligned} E_t\left[\widetilde{M}_{t+1}F_{t+1}\right] &= E_t\left[(1-b(F_{t+1}-E\left[F_{t+1}\right])F_{t+1}\right] \\ &= E_t\left[F_{t+1}\right]\left(1+bE\left[F_{t+1}\right]\right) - bE_t\left[F_{t+1}^2\right]. \end{aligned}$$

The correctly specified model implies that

$$E_t\left[F_{t+1}\right]\left(1+E\left[b_tF_{t+1}\right]\right) - b_tE_t\left[F_{t+1}^2\right] = 0.$$

Thus,

$$E_t\left[F_{t+1}\right] = \frac{b_t}{1+E\left[b_tF_{t+1}\right]} E_t\left[F_{t+1}^2\right],$$

and

$$\begin{aligned} E_t\left[\widetilde{M}_{t+1}F_{t+1}\right] &= E_t\left[F_{t+1}\right]\left(1+bE\left[F_{t+1}\right]\right)-bE_t\left[F_{t+1}^2\right] \\ &= \left(\frac{1+bE[F_{t+1}]}{1+E\left[b_tF_{t+1}\right]}b_t-b\right)E_t\left[F_{t+1}^2\right]. \end{aligned}$$

A.3 Misspecified model with i.i.d. errors

The SDF prices $R_{t,t+1}^{ie}$ unconditionally:

$$\begin{aligned} E\left[\widetilde{M}_{t,t+1}R_{t,t+1}^{ie}\right] &= E\left[\left(1-b\left(F_{t,t+1}-E\left[F_{t,t+1}\right]\right)\right)\left(\beta_iF_{t,t+1}+\varepsilon_{i,t+1}+\theta\varepsilon_{i,t}\right)\right] \\ &= \beta_iE\left[F_{t,t+1}-\frac{E\left[F_{t,t+1}\right]}{Var\left(F_{t,t+1}\right)}\left(F_{t,t+1}-E\left[F_{t,t+1}\right]\right)F_{t,t+1}\right]=0. \end{aligned}$$

However, that is not the case conditionally:

$$\begin{aligned} E_t\left[\widetilde{M}_{t,t+1}R_{t,t+1}^{ie}\right] &= E_t\left[\left(1-b\left(F_{t,t+1}-E\left[F_{t,t+1}\right]\right)\right)\left(\beta_iF_{t,t+1}+\varepsilon_{i,t+1}+\theta\varepsilon_{i,t}\right)\right] \\ &= \beta_iE_t\left[F_{t,t+1}-\frac{E\left[F_{t,t+1}\right]}{Var\left(F_{t,t+1}\right)}\left(F_{t,t+1}-E\left[F_{t,t+1}\right]\right)F_{t,t+1}\right] \\ &\quad +E_t\left[\left(1-b\left(F_{t,t+1}-E\left[F_{t,t+1}\right]\right)\right)\theta\varepsilon_{i,t}\right] \\ &= \theta\varepsilon_{i,t}. \end{aligned}$$

Next, we show that $Cov\left(\nu_{t-1,t},\eta_t^{(1)}\right)\neq 0$. Because

$$\eta_{t-1}^{(1)}+\nu_{t-1,t}=\left(1-b\left(F_{t-1,t}-E\left[F_{t-1,t}\right]\right)\right)\left(\beta_iF_{t-1,t}+\varepsilon_{i,t}+\theta\varepsilon_{i,t-1}\right),$$

we have that:

$$Cov\left(\eta_{t-1}^{(1)}+\nu_{t-1,t},\theta\varepsilon_{i,t}\right) = \theta Var\left(\varepsilon_{i,t}\right).$$

Because $Cov(\eta_{t-1}^{(1)},\theta\varepsilon_{t,t})=Cov(\eta_{t-1}^{(1)},\eta_t^{(1)})=0$, $Cov\left(\nu_{t-1,t},\eta_t^{(1)}\right)=\theta Var\left(\varepsilon_{i,t}\right)$.

A.4 No serial correlation in residuals

That residuals are not autocorrelated follows from Equation (8). For simplicity, consider one horizon, h. We have that

$$E(f_t^i\cdot f_{t+1}^i)=E(f_t^i\cdot E_t(f_{t+1}^i))=E(f_t^i\cdot z_{i,t}^{(h)}E_t(M_{t,t+1}R_{t,t+1}^i-1))=0,$$

because, under the null, $E_t(M_{t,t+1}R_{t,t+1}^i-1)=0$ for all t.

A.5 Derivation of the test statistics for the linear SDF

To avoid clutter, we do not distinguish between populations and sample versions of objects, such as E vs E_T, or V vs V_T, and the various concepts that follow from that. The derivation relies on standard GMM analytics, thus it should not create a confusion.

In this section we will keep the test assets more general, in the sense that we estimate the model to price K factors correctly unconditionally, but allow the other moments to be trading strategies in I excess returns that are assumed priced by the model. As a special case, those K excess returns are the factors themselves.

Let $N = 2K + I(n-1)$, $M_t = 1 + b'\mu - b'F_t$, with μ, b $K \times 1$ vectors. Let $\gamma \equiv [\mu^\top, b^\top]^\top$, and $g(\gamma)$ is given in Equation (15).

Define

$$a(\gamma) \equiv \frac{\partial g^\top(\gamma)}{\partial \gamma} W$$
$$d(\gamma) \equiv \frac{\partial g(\gamma)}{\partial \gamma^\top}$$

where $a(\gamma)$ is an $2K \times N$ matrix, $d(\gamma)$ is an $N \times 2K$ matrix, and W is an $N \times N$ weighting matrix.

The J-test statistic is then

$$J \equiv T g(\gamma)^\top \Big[\Big(I - d(ad)^{-1}a\Big) S \Big(I - d(ad)^{-1}a\Big)^\top\Big]^{-1} g(\gamma)$$

where S is the spectral density matrix of f_t.

We estimate using only the first $2K$ conditions and test on the remaining $N - 2K$ conditions. Therefore, we use the following weighting matrix throughout

$$W = \begin{pmatrix} I_{2K\times 2K} & 0_{2K\times(N-2K)} \\ 0_{(N-2K)\times 2K} & 0_{(N-2K)\times(N-2K)} \end{pmatrix}$$

Then

$$a = \begin{pmatrix} \frac{\partial g^{2K}(\gamma)^\top}{\partial \gamma} & 0_{2K\times(N-2K)} \end{pmatrix}$$
$$d = \begin{pmatrix} \frac{\partial g^{2K}(\gamma)}{\partial \gamma^\top} \\ \\ \frac{\partial g^{N-2K}(\gamma)}{\partial \gamma^\top} \end{pmatrix}$$

where $g^{2K}(\gamma)$ denotes the $2K$ first moment conditions and g^{N-2K} denotes the last $N-2K$ moment conditions. Then

$$ad = \frac{\partial g^{2K}(\gamma)^\top}{\partial \gamma} \frac{\partial g^{2K}(\gamma)}{\partial \gamma^\top}.$$

Furthermore,

$$
\begin{aligned}
\frac{\partial g_{i,h}}{\partial \gamma^\top}(\gamma) &= E\Big(M_{t+1}(\gamma) R^{ie}_{t+1} \frac{\partial z^{(h)}_{i,t}}{\partial \gamma^\top}(\gamma) + z^{(h)}_{i,t}(\gamma) R^{ie}_{t+1} \frac{\partial M_{t+1}}{\partial \gamma^\top}(\gamma)\Big) \\
&= E\Big(z^{(h)}_{i,t}(\gamma) R^{ie}_{t+1} \frac{\partial M_{t+1}}{\partial \gamma^\top}(\gamma)\Big),
\end{aligned}
$$

where the last equality follows from the null. Using the structure of M gives us

$$
\begin{aligned}
\frac{\partial g_{i,h}}{\partial \gamma^\top}(\gamma) &= E\Big(z^{(h)}_{i,t}(\gamma)[R^{ie}_{t+1} b^\top, R^{ie}_{t+1}(\mu - F_{t+1})^\top]\Big) \\
&= \Big(E[z^{(h)}_{i,t}(\gamma) R^{ie}_{t+1}] b^\top, -Cov[z^{(h)}_{i,t}(\gamma) R^{ie}_{t+1}, F^\top_{t+1}]\Big)
\end{aligned}
$$

Stacking the expressions gives us

$$
\begin{aligned}
\frac{\partial g_h}{\partial \gamma^\top}(\gamma) &= \Big(E[z^{(h)}_t(\gamma) \circ R^e_{t+1}] b^\top, -Cov[z^{(h)}_t(\gamma) \circ R^e_{t+1}, F^\top_{t+1}]\Big) \equiv (\mu_{z_h} b^\top, -\Sigma_{z_h,F}), \\
\frac{\partial g}{\partial \gamma^\top}(\gamma) &= \begin{pmatrix} -I_{K\times K} & 0_{K\times K} \\ \mu b^\top & -\Sigma \\ \mu_{z_2} b^\top & -\Sigma_{z_2,F} \\ \vdots & \vdots \\ \mu_{z_n} b^\top & -\Sigma_{z_n,F} \end{pmatrix},
\end{aligned}
$$

where $\mu_{z_h} \equiv \mathbb{E}[z^h_t(\gamma) \circ R^e_{t+1}]$ is a vector of expected excess return and $\Sigma_{z_h,F} \equiv Cov[z^h_t(\gamma) \circ R^e_{t+1}, F^\top_{t+1}]$ is the covariance matrix between $z^h_t(\gamma) \circ R^e_{t+1}$ and F_{t+1} (note that it is generally not symmetric) and μ and Σ denote the expected excess factor returns and covariance matrix, respectively.

We then have

$$
\begin{aligned}
\Big[\frac{\partial g^{2K}}{\partial \gamma^\top}\Big]^{-1} &= \begin{pmatrix} -I_{K\times K} & 0_{K\times K} \\ \mu b^\top & -\Sigma \end{pmatrix}^{-1} = -\begin{pmatrix} I_{K\times K} & 0_{K\times K} \\ \Sigma^{-1}\mu b^\top & \Sigma^{-1} \end{pmatrix}, \\
\Big[\frac{\partial (g^{2K})^\top}{\partial \gamma}\Big]^{-1} &= \begin{pmatrix} -I_{K\times K} & b\mu^\top \\ 0_{K\times K} & -\Sigma \end{pmatrix}^{-1} = -\begin{pmatrix} I_{K\times K} & b\mu^\top \Sigma^{-1} \\ 0_{K\times K} & \Sigma^{-1} \end{pmatrix}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
d(ad)^{-1}a &= d\begin{pmatrix} I_{K\times K} & 0_{K\times K} \\ \Sigma^{-1}\mu b^\top & \Sigma^{-1} \end{pmatrix}\begin{pmatrix} I_{K\times K} & b\mu^\top\Sigma^{-1} \\ 0_{K\times K} & \Sigma^{-1} \end{pmatrix}\begin{pmatrix} -I_{K\times K} & b\mu^\top & 0_{K\times(N-2K)} \\ 0_{K\times K} & -\Sigma & 0_{K\times(N-2K)} \end{pmatrix} \\
&= -d\begin{pmatrix} I_{K\times K} & 0_{K\times K} \\ \Sigma^{-1}\mu b^\top & \Sigma^{-1} \end{pmatrix}\begin{pmatrix} I_{2K\times 2K} & 0_{2K\times(N-2K)} \end{pmatrix} \\
&= -\begin{pmatrix} -I_{K\times K} & 0_{K\times K} \\ \mu b^\top & -\Sigma \\ \mu_{z_2} b^\top & -\Sigma_{z_2,F} \\ \vdots & \vdots \\ \mu_{z_n} b^\top & -\Sigma_{z_n,F} \end{pmatrix}\begin{pmatrix} I_{K\times K} & 0_{K\times K} \\ \Sigma^{-1}\mu b^\top & \Sigma^{-1} \end{pmatrix}\begin{pmatrix} I_{2K\times 2K} & 0_{2K\times(N-2K)} \end{pmatrix} \\
&= \begin{pmatrix} I_{K\times K} & 0_{K\times K} \\ 0_{K\times K} & I_{K\times K} \\ (\Sigma_{z_2,F}\Sigma^{-1}\mu - \mu_{z_2})b^\top & \Sigma_{z_2,F}\Sigma^{-1} \\ \vdots & \vdots \\ (\Sigma_{z_n,F}\Sigma^{-1}\mu - \mu_{z_n})b^\top & \Sigma_{z_n,F}\Sigma^{-1} \end{pmatrix}\begin{pmatrix} I_{2K\times 2K} & 0_{2K\times(N-2K)} \end{pmatrix} \\
&= \begin{pmatrix} I_{K\times K} & 0_{K\times K} & 0_{K\times(N-2K)} \\ 0_{K\times K} & I_{K\times K} & 0_{K\times(N-2K)} \\ (\Sigma_{z_2,F}\Sigma^{-1}\mu - \mu_{z_2})b^\top & \Sigma_{z_2,F}\Sigma^{-1} & 0_{I\times(N-2K)} \\ \vdots & \vdots & \vdots \\ (\Sigma_{z_n,F}\Sigma^{-1}\mu - \mu_{z_n})b^\top & \Sigma_{z_n,F}\Sigma^{-1} & 0_{I\times(N-2K)} \end{pmatrix}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
E(M_{t+1}z_t^{(h)} \circ R_{t+1}^e) &= E(z_t^{(h)} \circ R_{t+1}^e) - Cov(z_t^{(h)} \circ R_{t+1}^e, F_{t+1}^\top)b \\
&= \mu_{z_h} - \Sigma_{z_h,F}\Sigma^{-1}\mu \equiv \alpha_{z_h},
\end{aligned}
$$

where α_{z_h} denotes the vector of pricing errors for horizon h. Furthermore, note that $\beta_{z_h} \equiv \Sigma_{z_h,F}\Sigma^{-1}$ is a matrix of regression coefficients obtained by regressing $z_t^{(h)} \circ R_{t+1}^e$ on F_{t+1} with an intercept. In particular, row i in β_{z_h} are the coefficients from regressing $z_{i,t}^{(h)} R_{t+1}^{ie}$ on F_{t+1}. The intercepts in these regressions equal the pricing error because $E(M) = 1$ and $E(MF) = 0$.

As a result

$$
I - d(ad)^{-1}a = \begin{pmatrix} 0_{K\times K} & 0_{K\times K} & 0_{K\times(N-2K)} \\ 0_{K\times K} & 0_{K\times K} & 0_{K\times(N-2K)} \\ \alpha_{z_2} b^\top & -\beta_{z_2} & \\ \vdots & \vdots & I_{(N-2K)\times(N-2K)} \\ \alpha_{z_n} b^\top & -\beta_{z_n} & \end{pmatrix} = \begin{pmatrix} 0_{2K\times K} & 0_{2K\times K} & 0_{2K\times(N-2K)} \\ \alpha_z b^\top & -\beta_z & I_{(N-2K)\times(N-2K)} \end{pmatrix},
$$

where $\alpha_z \equiv (\alpha_{z_2}^\top, \ldots, \alpha_{z_n}^\top)^\top$ and $\beta_z \equiv (\beta_{z_2}^\top, \ldots, \beta_{z_n}^\top)^\top$.

Partition the spectral density matrix

$$S = \begin{pmatrix} S_{00} & S_{01} & L_{0,z} \\ S_{10} & S_{11} & L_{1,z} \\ L_{z,0} & L_{z,1} & \Gamma \end{pmatrix}$$

where S_{ij} are the parts of S relating solely to the moments we use for estimation, Γ is the part of S relating to the moments used for testing, and $L_{i,j}$ are the parts of S relating to interactions between estimation and test moments.

Then

$$\begin{aligned}
&\Big(I - d(ad)^{-1}a\Big)S\Big(I - d(ad)^{-1}a\Big)^\top = \begin{pmatrix} 0_{2K\times K} & 0_{2K\times K} & 0_{2K\times(N-2K)} \\ \alpha_z b^\top & -\beta_z & I_{(N-2K)\times(N-2K)} \end{pmatrix} \begin{pmatrix} S_{00} & S_{01} & L_{0,z} \\ S_{10} & S_{11} & L_{1,z} \\ L_{z,0} & L_{z,1} & \Gamma \end{pmatrix} \\
&\times \begin{pmatrix} 0_{K\times 2K} & b\alpha_z^\top \\ 0_{K\times 2K} & -\beta_z^\top \\ 0_{(N-2K)\times 2K} & I_{(N-2K)\times(N-2K)} \end{pmatrix} \\
&= \begin{pmatrix} 0_{2K\times K} & 0_{2K\times K} & 0_{2K\times(N-2K)} \\ \alpha_z b^\top S_{00} - \beta_z S_{10} + L_{z,0} & \alpha_z b^\top S_{01} - \beta_z S_{11} + L_{z,1} & \alpha_z b^\top L_{0,z} - \beta_z L_{1,z} + \Gamma \end{pmatrix} \\
&\times \begin{pmatrix} 0_{K\times 2K} & b\alpha_z^\top \\ 0_{K\times 2K} & -\beta_z^\top \\ 0_{(N-2K)\times 2K} & I_{(N-2K)\times(N-2K)} \end{pmatrix} \\
&= \begin{pmatrix} 0_{2K\times 2K} & 0_{2K\times(N-2K)} \\ 0_{(N-2K)\times 2K} & Q_{(N-2K)\times(N-2K)} \end{pmatrix},
\end{aligned}$$

where

$$Q \equiv \Big(\alpha_z b^\top S_{00} - \beta_z S_{10} + L_{z,0}\Big) b\alpha_z^\top - \Big(\alpha_z b^\top S_{01} - \beta_z S_{11} + L_{z,1}\Big)\beta_z^\top + \alpha_z b^\top L_{0,z} - \beta_z L_{1,z} + \Gamma.$$

Under the null, $\alpha_z = 0$, thus we get a simplified Q-matrix

$$Q = \Gamma - \beta_z L_{1,z} - L_{z,1}\beta_z^\top + \beta_z S_{11}\beta_z^\top.$$

Now, suppose we estimate the spectral density matrix S as

$$\begin{aligned}
S &= Cov(f_t, f_t^\top) \\
&= \begin{pmatrix} \Sigma & Cov(F_{t+1}, M_{t+1}F_{t+1}^\top) & Cov(F_{t+1}, M_{t+1}R_{z,t+1}^{e\top}) \\ Cov(M_{t+1}F_{t+1}, F_{t+1}^\top) & Cov(M_{t+1}F_{t+1}, M_{t+1}F_{t+1}^\top) & Cov(M_{t+1}F_{t+1}, M_{t+1}R_{z,t+1}^{e\top}) \\ Cov(M_{t+1}R_{z,t+1}^e, F_{t+1}^\top) & Cov(M_{t+1}R_{z,t+1}^e, M_{t+1}F_{t+1}^\top) & Cov(M_{t+1}R_{z,t+1}^e, M_{t+1}R_{z,t+1}^{e\top}) \end{pmatrix}
\end{aligned}$$

(under the null this is correct for Γ, $L_{1,z}$, $L_{z,1}$ and S_{11}, which are the parts that enter Q). Then

$$\begin{aligned}
Q &= V(MR_z^e) - \beta_z Cov(MF, MR_z^{e^\top}) - Cov(MR_z^e, MF^\top)\beta_z^\top + \beta_z V(MF)\beta_z^\top \\
&= V(M(\alpha_z + \beta_z F + \varepsilon_z)) - Cov(M\beta_z F, M(\alpha_z + \beta_z F + \varepsilon_z)^\top) \\
&- Cov(M(\alpha_z + \beta_z F + \varepsilon_z), M(\beta_z F)^\top) + V(M\beta_z F) \\
&= V(M)\alpha_z\alpha_z^\top + V(M\beta_z F) + V(M\varepsilon_z) + Cov(M\beta_z F, M(\alpha_z + \varepsilon_z)^\top) \\
&+ Cov(M(\alpha_z + \varepsilon_z), M(\beta_z F)^\top) + Cov(M\alpha_z, M\varepsilon_z^\top) \\
&+ Cov(M\varepsilon_z, M\alpha_z^\top) - Cov(M\beta_z F, M(\alpha_z + \varepsilon_z)^\top) \\
&- Cov(M(\alpha_z + \varepsilon_z), M(\beta_z F)^\top) - 2V(M\beta_z F) + V(M\beta_z F) \\
&= V(M)\alpha_z\alpha_z^\top + V(M\varepsilon_z) + Cov(M\alpha_z, M\varepsilon_z^\top) + Cov(M\varepsilon_z, M\alpha_z^\top) \\
&= V(M\varepsilon_z),
\end{aligned}$$

where the last equality again uses the null that $\alpha_z = 0$.

Our J test can therefore be written in any of the following forms:

$$\begin{aligned}
J/T &= E(MR_z^e)^\top V^{-1}(M\varepsilon_z)E(MR_z^e) \\
&= \alpha_z^\top V^{-1}(M\varepsilon_z)\alpha_z \\
&= E(M(R_z^e - \beta_z F))^\top V^{-1}(M\varepsilon_z)E(M(R_z^e - \beta_z F)).
\end{aligned}$$

Moving on to the connection to GRS, we can write

$$\begin{aligned}
V(M\varepsilon_z) &= E(M^2\varepsilon_z\varepsilon_z^\top) - E(M\varepsilon_z)E(M\varepsilon_z^\top) = E(M^2)E(\varepsilon_z\varepsilon_z^\top) + Cov(M^2, \varepsilon_z\varepsilon_z^\top) \\
&= \{E(M)^2 + V(M)\}E(\varepsilon_z\varepsilon_z^\top) + Cov([1 + b^\top\mu - b^\top F]^2, \varepsilon_z\varepsilon_z^\top) \\
&= \{1 + \mu^\top\Sigma^{-1}\mu\}V(\varepsilon_z) - 2(1 + b^\top\mu)Cov(b^\top F, \varepsilon_z\varepsilon_z^\top) + Cov(b^\top FF^\top b, \varepsilon_z\varepsilon_z^\top)
\end{aligned}$$

It is therefore clear that $V(M\varepsilon_z) = \{1 + \mu^\top\Sigma^{-1}\mu\}V(\varepsilon_z)$ if $\varepsilon_z\varepsilon_z^\top$ is uncorrelated with both $b^\top F$ and $b^\top FF^\top b$. For example, that is the case if ε_z is homoscedastic and unpredictable, as in the GRS test. In this case, our J test would simplify to the standard GRS test

$$T\{1 + \mu^\top\Sigma^{-1}\mu\}^{-1}\alpha^\top V(\varepsilon_z)^{-1}\alpha.$$

One natural question is whether we can find a single portfolio that gives us exactly the same J statistic as the original $I(n-1)$ trading strategies. The starting point would be to guess that the portfolio in question is

$$R^{*e} = \alpha_z^\top V^{-1}(M\varepsilon_z)(R_z^e - \beta_z F).$$

The pricing error of this portfolio is

$$\alpha^* \equiv E(MR^{*e}) = \alpha_z^\top V^{-1}(M\varepsilon_z)E(MR_z^e) = \alpha_z^\top V^{-1}(M\varepsilon_z)\alpha_z.$$

Furthermore, $\varepsilon^* = \alpha_z^\top V^{-1}(M\varepsilon_z)\varepsilon_z$. Thus,

$$\begin{aligned} V(M\varepsilon^*) &= V(M\alpha_z^\top V^{-1}(M\varepsilon_z)\varepsilon_z) = \alpha_z^\top V^{-1}(M\varepsilon_z)V(M\varepsilon_z)V^{-1}(M\varepsilon_z)\alpha_z \\ &= \alpha_z^\top V^{-1}(M\varepsilon_z)\alpha_z. \end{aligned}$$

As a result, the $J-$statistic with this single portfolio would be

$$J/T = \frac{E(MR^{e*})^2}{V(M\varepsilon^*)} = \frac{(\alpha_z^\top V^{-1}(M\varepsilon_z)\alpha_z)^2}{\alpha_z^\top V^{-1}(M\varepsilon_z)\alpha_z} = \alpha_z^\top V^{-1}(M\varepsilon_z)\alpha_z,$$

which is the same as the previously derived $J-$statistic.

A.6 HAR standard errors

Our test given in Proposition 2 is derived under the null hypothesis that pricing errors are not predictable, which implies no autocorrelation in the moments. Imposing the null hypothesis this way generates a more efficient test statistic that is better behaved in small samples.

That said, the general GMM formulas can in principle be applied with HAR-adjustment, which accounts for autocorrelation in the moments. Under the null hypothesis errors are not predictable so autocorrelations should be zero and therefore not matter asymptotically. Of course, in a small sample they would add noise to the estimate of the covariance matrix. For completeness, we also report $p-$values from such tests. We include as many lags of potential autocorrelation in the Newey-West procedure as the maximal return horizon (48 months). With the exception of the CAPM, the resulting p-values are uniformly substantially smaller than those we report in our main test in Table A1, now rejecting all of these models at the 1% level.

A.7 Out-of-sample instruments

If the focus is only on one-period returns, one might wonder if the test results survive using managed portfolios that are tradeable in real time. To this end, we estimate the parameters in the SDF using only data up until time t to construct the instrument $z_t^{(i,h)}$. In order to have a reasonable "burn in"-period for all models, we start the test datasets in July 1983.

Table A2 shows that the results for FF3+MOM, FF5$_{DMRS}$, SY, and HKS are highly significant despite the loss of power due to the relatively short samples. IRs continue to be high relative to the SRs of the models' factors.

A.8 Construction of the BAB factor

We construct four value-weighted portfolios: (1) small size, low beta, (2) small size, high beta, (3) big size, low beta, and (4) big size, high beta. The size cutoffs are the 40th and 60th NYSE percentiles. For betas, we use the 20th and 80th NYSE percentiles. Denote these returns as $R_{s\ell}, R_{sh}, R_{b\ell}, R_{bh}$, respectively, where s denotes small size, ℓ denotes low beta, b denotes big size, and h denotes high beta. We also compute the prior beta for each of the four portfolios and shrink towards 1 with a value of 0.5 on the historical estimate. We denote these as $\beta_{s\ell,t}$, $\beta_{b\ell,t}$, $\beta_{sh,t}$, and $\beta_{bh,t}$. We construct these portfolios using the 25 size and market beta sorted portfolio returns, as well as the corresponding market values and 60-month historical betas, given on Kenneth French's webpage.

The factor return is then constructed as follows:

$$
\begin{aligned}
BAB_{t,t+1} \quad &= \quad \frac{1}{\beta_{\ell,t}}\left(\frac{1}{2}R_{s\ell,t,t+1} + \frac{1}{2}R_{b\ell,t,t+1} - R_{f,t,t+1}\right) \\
&- \quad \frac{1}{\beta_{h,t}}\left(\frac{1}{2}R_{sh,t,t+1} + \frac{1}{2}R_{bh,t,t+1} - R_{f,t,t+1}\right),
\end{aligned}
$$

where $\beta_{\ell,t} = \frac{1}{2}\beta_{s\ell,t} + \frac{1}{2}\beta_{b\ell,t}$, and $\beta_{h,t} = \frac{1}{2}\beta_{sh,t} + \frac{1}{2}\beta_{bh,t}$. As a result, the conditional market beta of BAB should be close to zero, as in Frazzini and Pedersen (2014).

A.9 Testing models with MHR to the same assets

In our main test we chose to use MHR to each model's factors as test assets. This allows us to trace a rejection to the dynamics of the model's factors being inconsistent with those implicitly assumed when assuming the factors unconditionally span the UMVE portfolio.

However, as the set of test assets vary across models, we cannot say that one model is "better" than another based on the test. For instance, our MHR test does not reject the unconditional CAPM, but a large literature has soundly rejected this model using various characteristic-sorted portfolios as test assets.

To level the playing field, we ask all models to price MHR to the FF5 factors. We have converged on this set of factors because it is a relatively recent set of factors that are arguably important for the cross-sectional asset pricing. Also, the factors are available across the data spans applicable to all the models that we study.

As in the baseline case, the models are estimated to match its factor single-horizon returns (SHR). Thus, to establish a straw man, we first test if models price SHR to the FF5 factors. One can anticipate that some of our models would be rejected on the basis of SHR alone. After all, that is the reason for Fama and French (2015) to introduce the five-factor structure.

Indeed, as Table A3 demonstrates, all the models on Panel A are rejected on the basis of SHR (the SHR test for FF5 is not applicable because in this case testing coincides with the baseline). In Panel B, which contains more recent models, only HKS is rejected using SHR to the FF5 factors. The MHR test rejects all the models.

A.10 Alternative factor timing

In order to obtain b_t in Equation (5), we explicitly estimate $E_t(F_{t,t+1})$ and $V_t(F_{t,t+1})$ for each model. We emphasize that, because of the illustrative nature of our exercise, the estimation is in-sample. We estimate the conditional monthly variance-covariance matrix of the factor returns using the multivariate CCC-GARCH method of Bollerslev (1990). We estimate conditional mean of each element k of the vector of factors F using a simple regression model that is motivated by the uncovered strong dependencies in factor returns documented in Section 3.5:

$$F^i_{t,t+1} = \beta_{i,0} + \sum_{j=1}^{n} \beta_{i,h_j} x^{(h_j)}_{i,t} + \varepsilon^i_{t+1},$$

where $x^{(h)}_{i,t} = \sum_{j=1}^{h} F^i_{t-j,t-j+1}$. We use the same horizons h as in our GMM tests.

We follow the post LASSO approach of Belloni and Chernozhukov (2013) to estimate the regression above for each factor. That is, we use the LASSO to select strong predictive variables and, because the LASSO yields biased return estimates, we next use OLS with these selected regressors to get conditional expectation $E_t(F^i_{t,t+1})$.

Lastly, because dividing by estimated variance introduces a bias, we rescale each element i of the estimated version of portfolio weights in (5) by a constant, b^i. We mitigate the bias by ensuring, via b^i, for $i = 1, ..., K$ that the unconditional factor SHRs are priced correctly. That is also consistent with our testing strategy in the constant b case of the preceding section. Thereby, this approach connects with factor-timing, b_t, as described in Equation (3), with $D_0 = 0$, diagonal D_1 with element i equal to b^i, and $z_t = b_t$, where the latter are UMVE portfolio weights as defined in Equation (5).

Table A4 presents test results along with pricing errors and Sharpe ratios. The results suggest that, overall, we reject the same models as in the case of constant b. There are some differences. MKT+BAB is marginally rejected when b is constant, while we fail to reject when it is time-varying. FF3+MOMs $p-$value is 0.020, a rejection when b_t is time-varying.

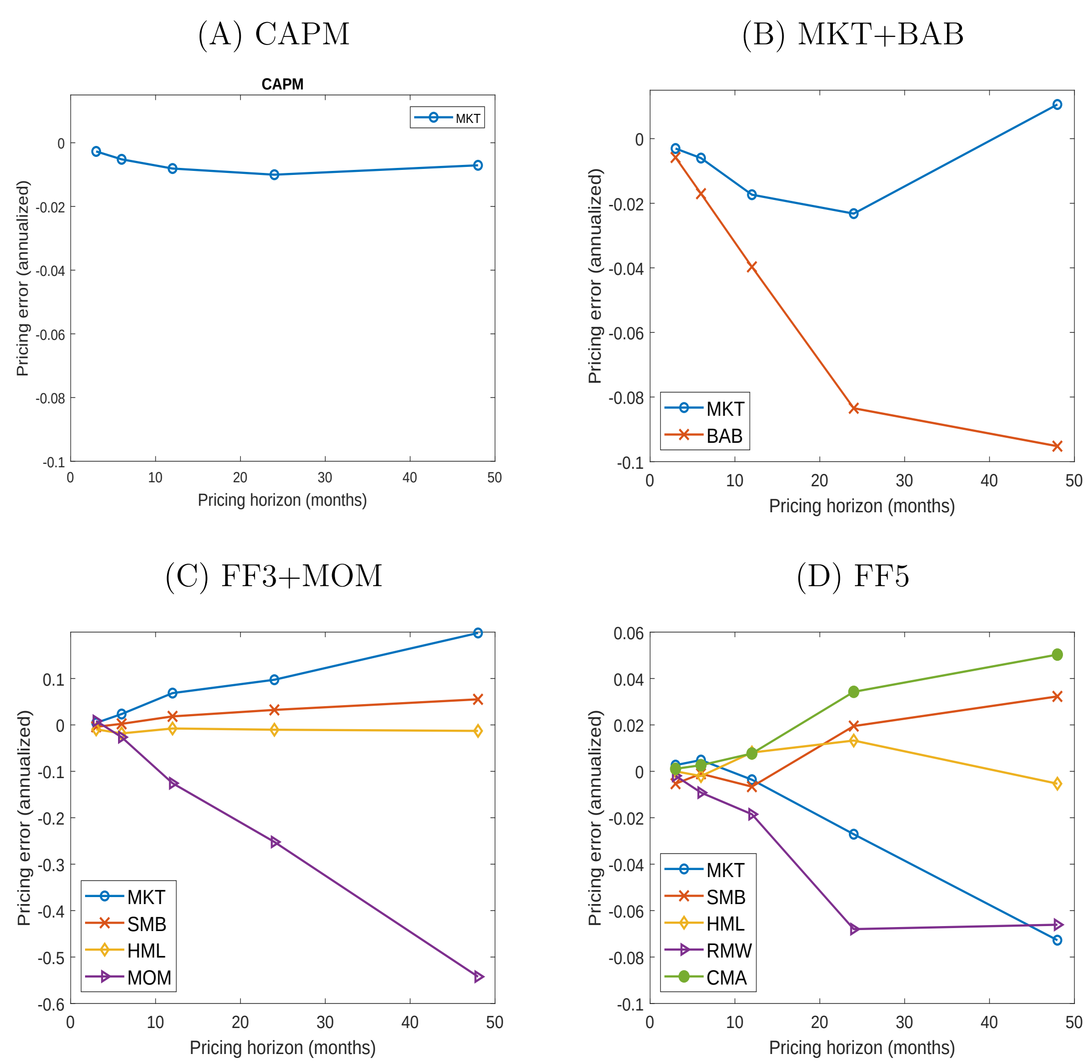
Table A4 shows the Sharpe ratio of the additional test asset, which is constructed as the UMVE portfolios $F^U_{t,t+1}$ implied from each model's estimated conditional factor means and covariance matrices as described in Equation (5). This is the maximal Sharpe ratio, comparable to that given in Table 1 for the constant b models. As one would expect, the Sharpe ratio reported for the models with time-varying b_t are generally higher than those from the constant b versions of the models. SY's maximal Sharpe ratio decreases, indicating poor estimates of the conditional factor dynamics for this model.

Table A4 reports MAPE across the original factors and horizons and is, thereby, comparable to the ones in Table 1. As an example, the constant b MAPE for the $FF5_{DMRS}$ model is 3.4%, while it is 20% for the time-varying b_t case. This occurs despite the maximal Sharpe ratio increasing from 1.59 to 1.83 with factor timing. That is the case more generally. Despite the increase in the maximal Sharpe ratios, the MAPEs across models are in many cases higher than those reported for the constant b case.

While it might seem surprising that pricing errors increase when the model's maximal Sharpe ratio increases, thus presumably getting closer to the true UMVE portfolio, this is a manifestation of the Herculean task of estimating the correct conditional dynamics. Misspecification in the conditional mean and variance processes likely results in persistent errors, which in turn show up in MHR per Equation (9).

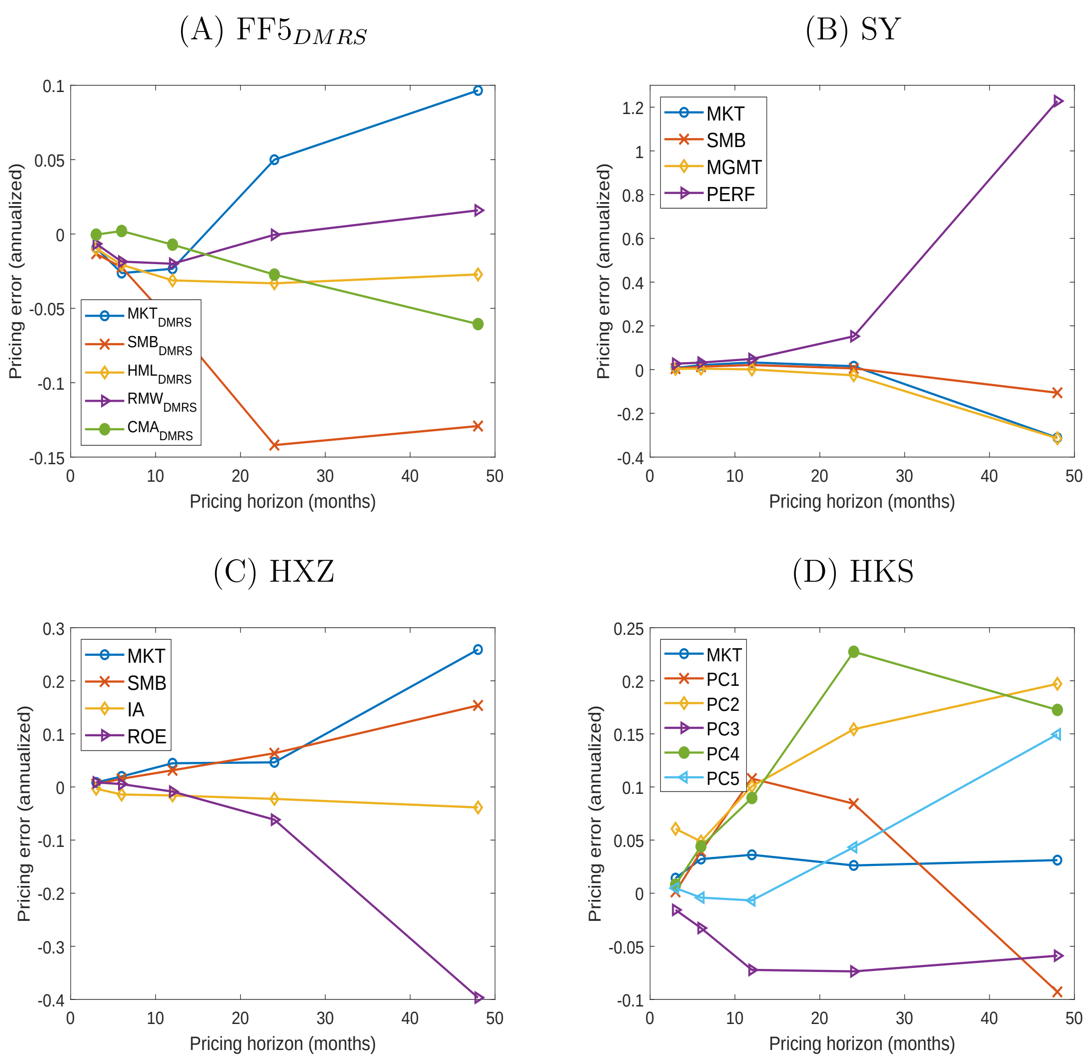
As in Section 4.2, we report the maximal annualized Information ratio for each model by running regressions (7). The left hand side test assets returns are $z_{i,t}^{(h)} \times F_{t,t+1}^{i}$ for factor i and horizon h, where we again add the timing UMVE portfolio as a test asset. In this case, the right-hand side factor is the implied UMVE portfolio. The Information ratios are similar to those of the constant b models, although these are not directly comparable due to the addition of the MHR-based return of UMVE portfolio in each model as a test asset.

Figure 1
Term structure of annualized factor pricing errors I



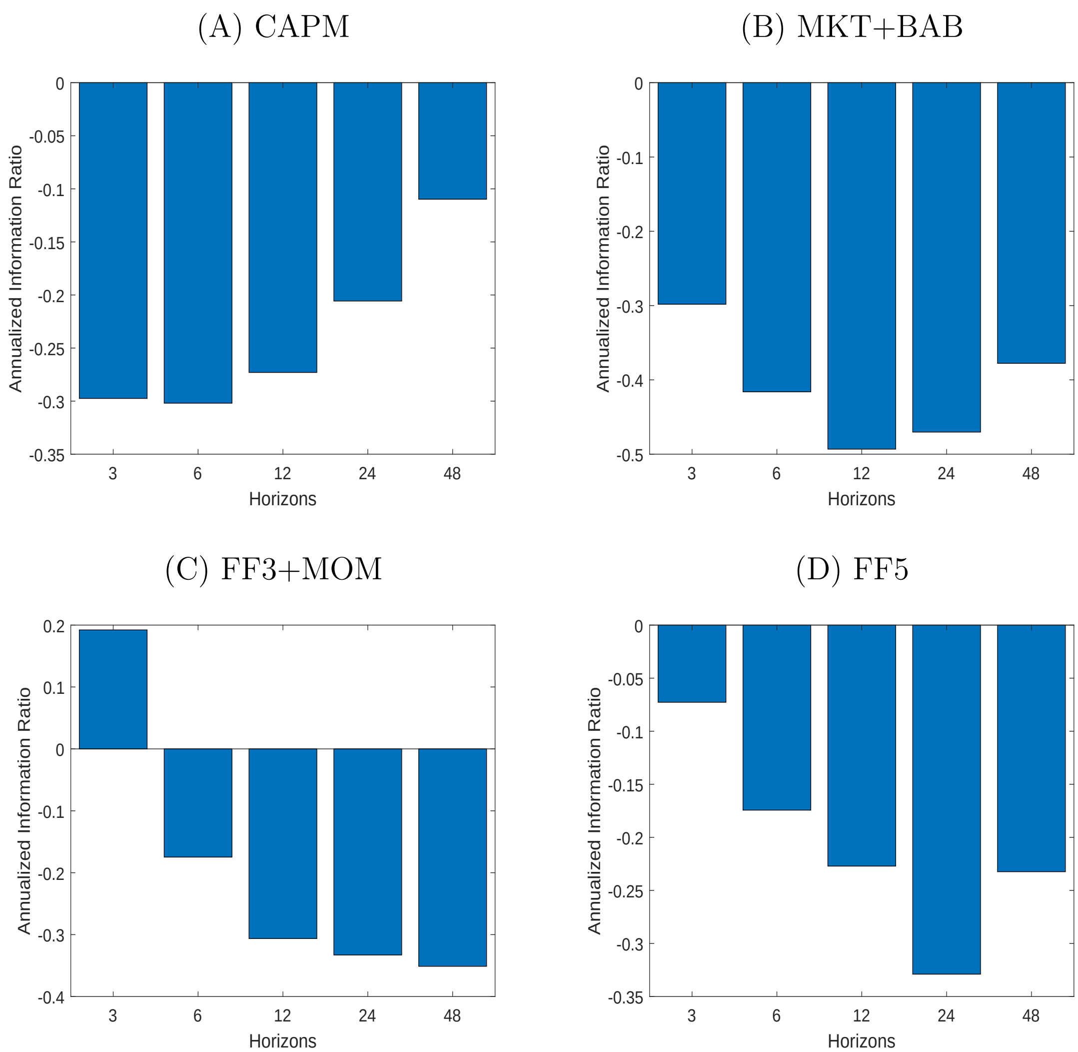
The panels show factor pricing errors for various models at horizons 3, 6, 12, 24, and 48 months. Annualized pricing errors at horizon h are $12/h \times E_T(z_{i,t}^{(h)} M_{t,t+1} F_{t,t+1}^i)$, where E_T denotes the sample average, $z_{i,t}^{(h)}$ is the endogenous conditioning variable for factor i at horizon h described in the main text, and $F_{t,t+1}^i$ is the return to factor i. The population average of a correctly specified model is zero. The sample is monthly, from 1963 to 2017.

Figure 2
Term structure of annualized factor pricing errors II



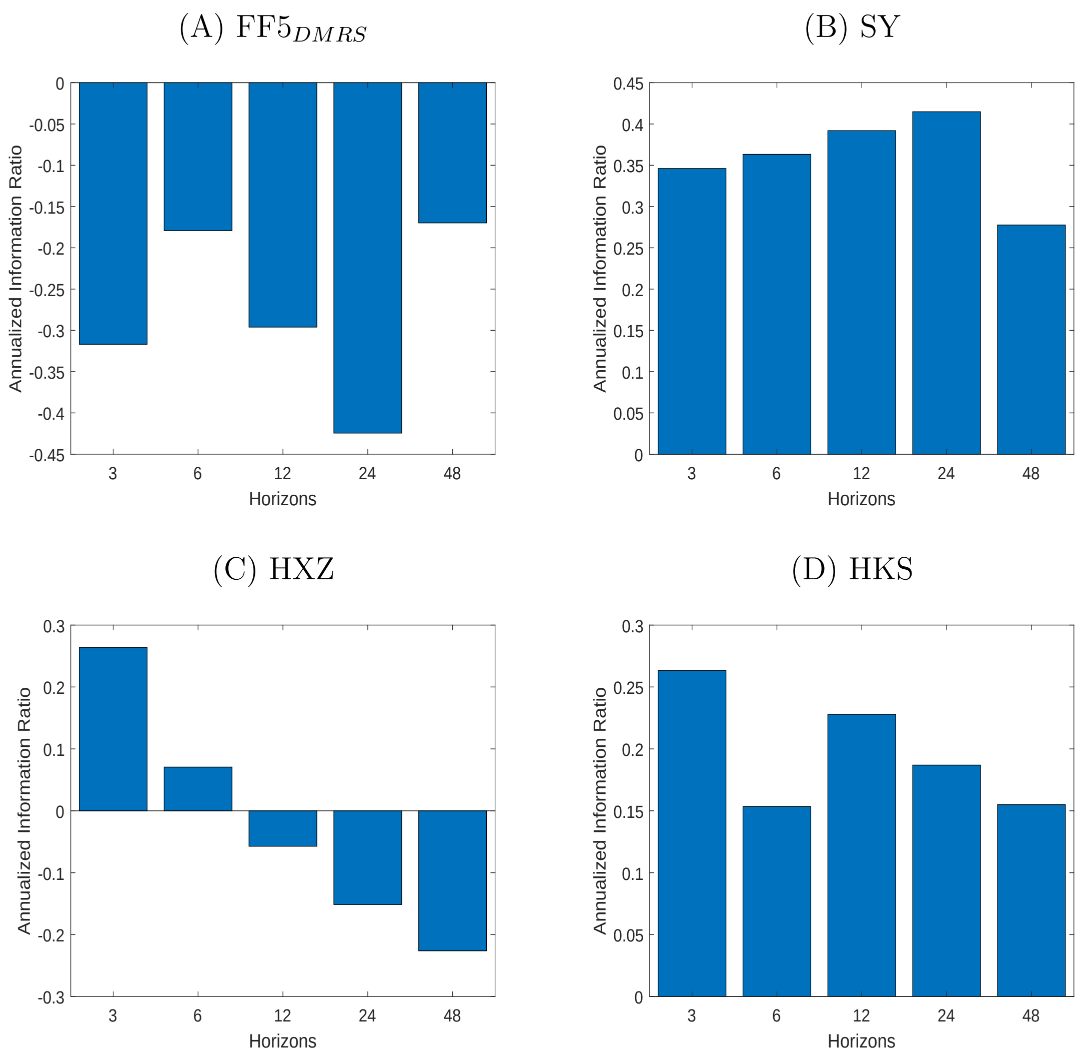
The panels show factor pricing errors for various models at horizons 3, 6, 12, 24, and 48 months. Annualized pricing errors at horizon h are $12/h \times E_T(z_{i,t}^{(h)} M_{t,t+1} F_{t,t+1}^i)$, where E_T denotes the sample average, $z_{i,t}^{(h)}$ is the endogenous conditioning variable for factor i at horizon h described in the main text, and $F_{t,t+1}^i$ is the return to factor i. The population average of a correctly specified model is zero. The sample is monthly, from 1963 to 2017 for $FF5_{DMRS}$, 1963 to 2016 for SY, 1967 to 2017 for HXZ, and 1974 to 2017 for HKS.

Figure 3
Information ratios I



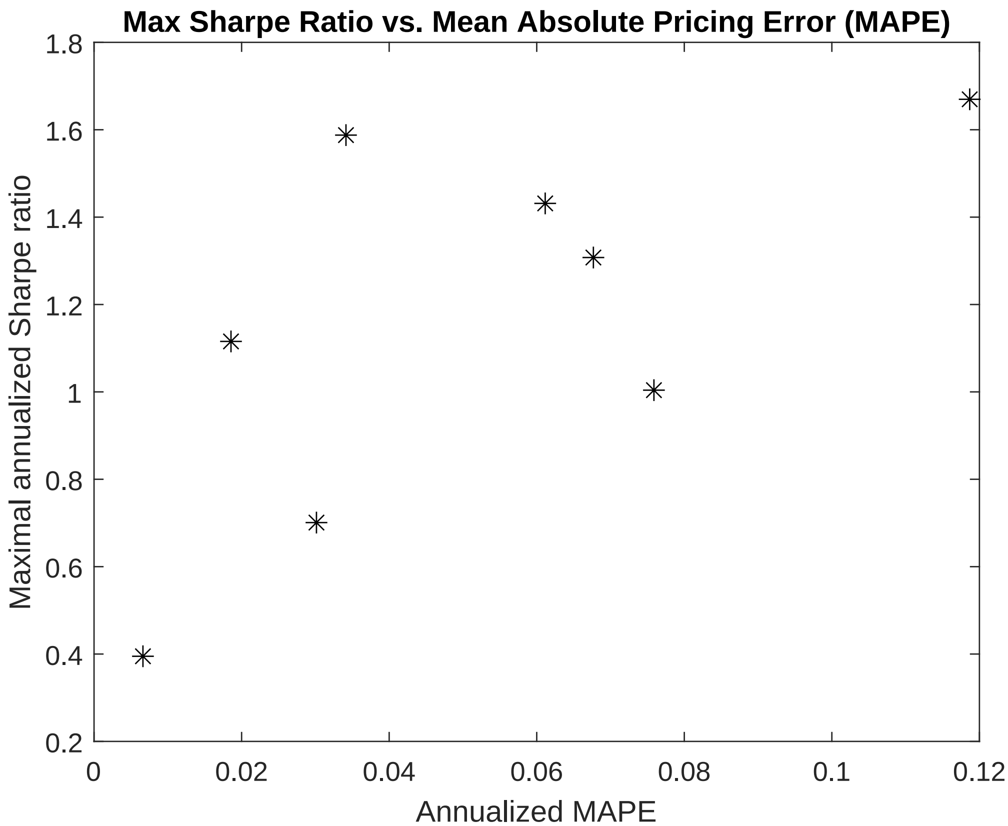
The panels show annualized Information Ratios (IR) for one factor from each model corresponding to timing strategies based on the 3, 6, 12, 24, and 48 month horizon moments. The information ratio is defined as $\sqrt{12}\alpha/\sigma(\varepsilon)$. For each model the chosen factor is the one with the maximal average pricing error (see Figure 1). The chosen factors are MKT (CAPM), BAB (MKT+BAB), MOM (FF3+MOM), and RMW (FF5). The sample is monthly, from 1963 to 2017.

Figure 4
Information ratios II



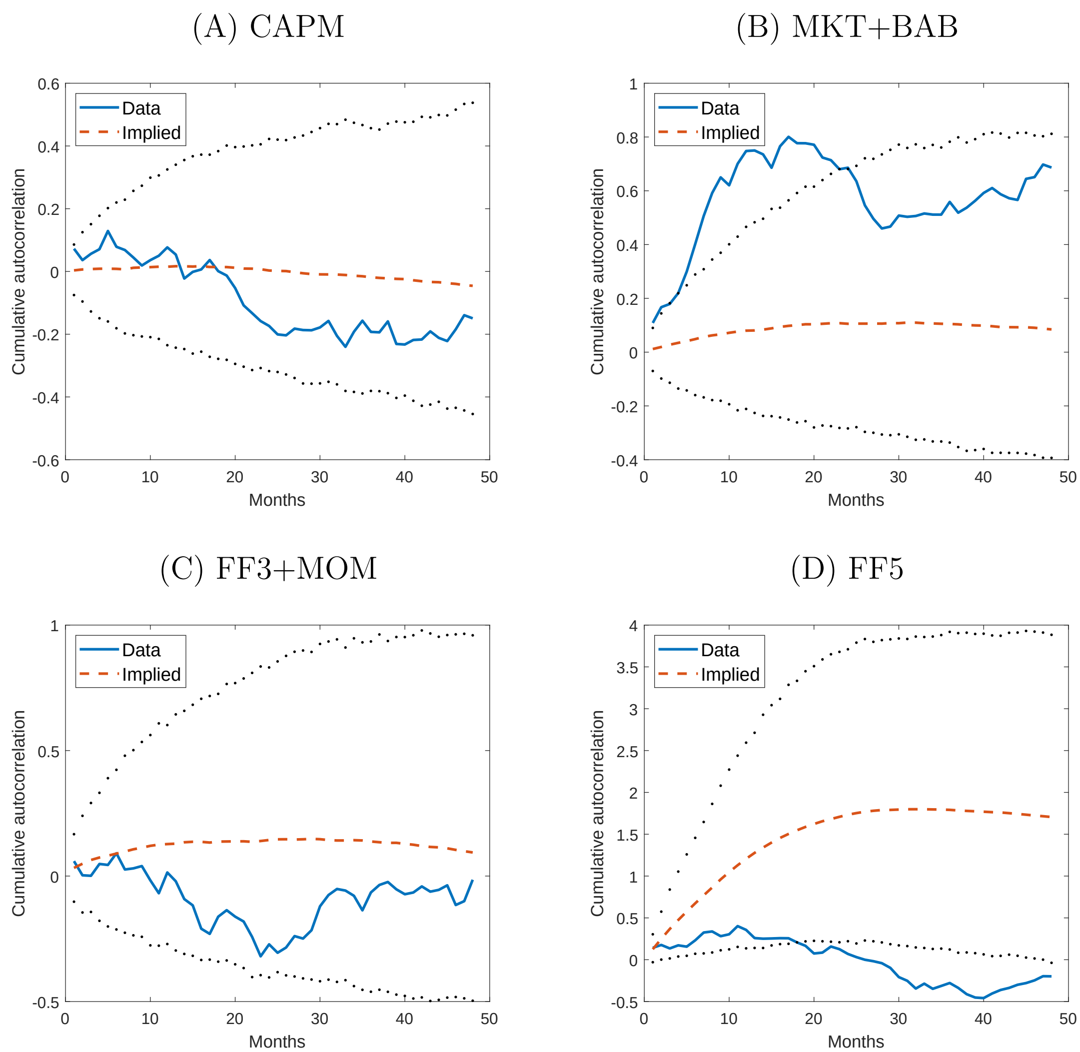
The panels show annualized Information Ratios (IR) for one factor from each model corresponding to timing strategies based on the 3, 6, 12, 24, and 48 month horizon moments. The information ratio is defined as $\sqrt{12}\alpha/\sigma(\varepsilon)$. For each model the chosen factor is the one with the maximal average pricing error (see Figure 2). The chosen factors are SMB (FF5$_{DMRS}$), PERF (SY), ROE (HXZ), and PC2 (HKS). The sample is monthly, from 1963 to 2017 for FF5$_{DMRS}$, 1963 to 2016 for SY, 1967 to 2017 for HXZ, and 1974 to 2017 for HKS.

Figure 5
Max Sharpe ratio of single-horizon factor model vs. multi-horizon pricing errors



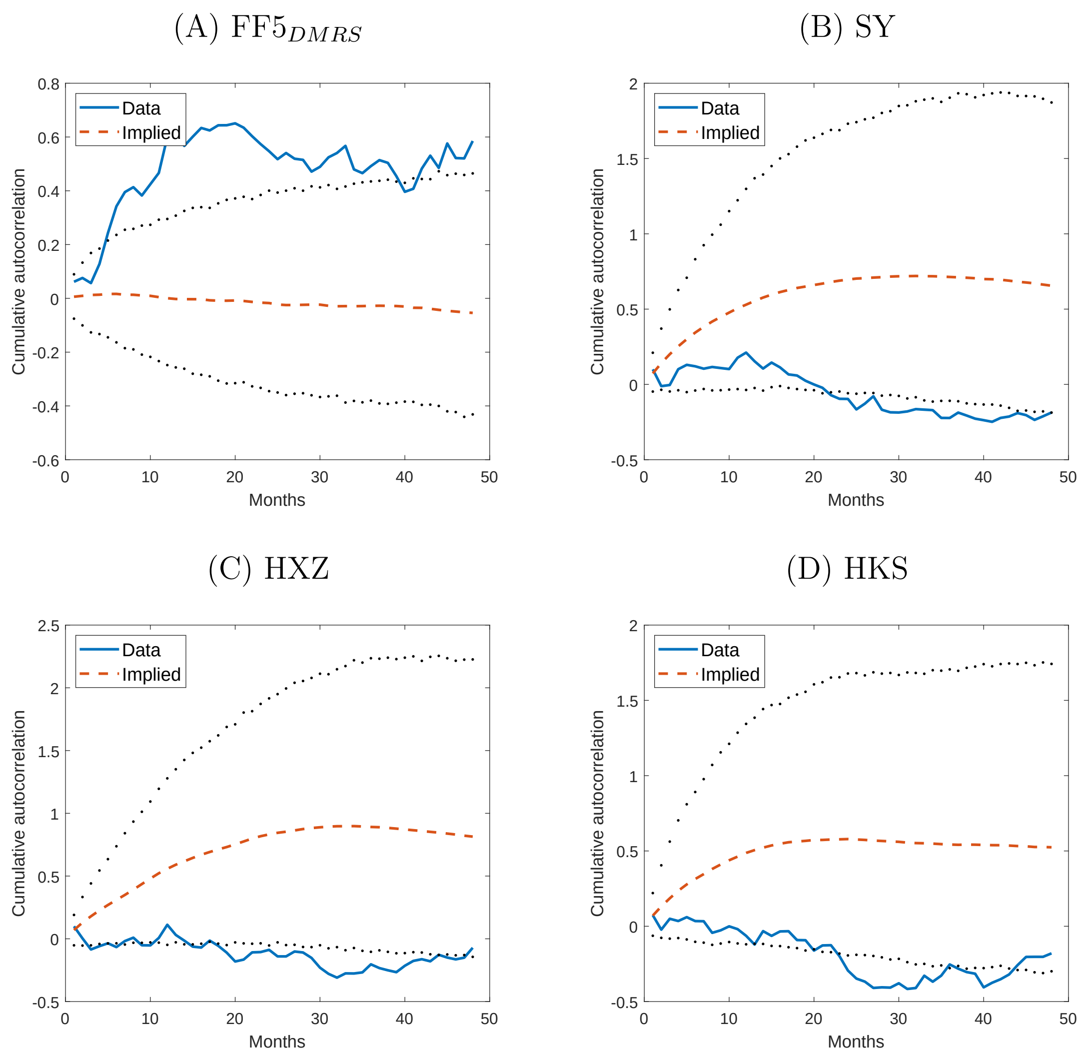
The figure plots the annualized maximal in-sample Sharpe ratio combination of the factors in each model against the annualized mean absolute pricing error (MAPE) of the corresponding model, when the model is estimated using one-period returns and tested on excess factor returns with horizons 1, 3, 6, 12, 24, and 48 months. The sample is monthly, from 1963 to 2017 for all models except SY, which is 1963 to 2016, HXZ, which is 1967 to 2017, and HKS, which is 1974 to 2017.

Figure 6
Cumulative autocorrelations I



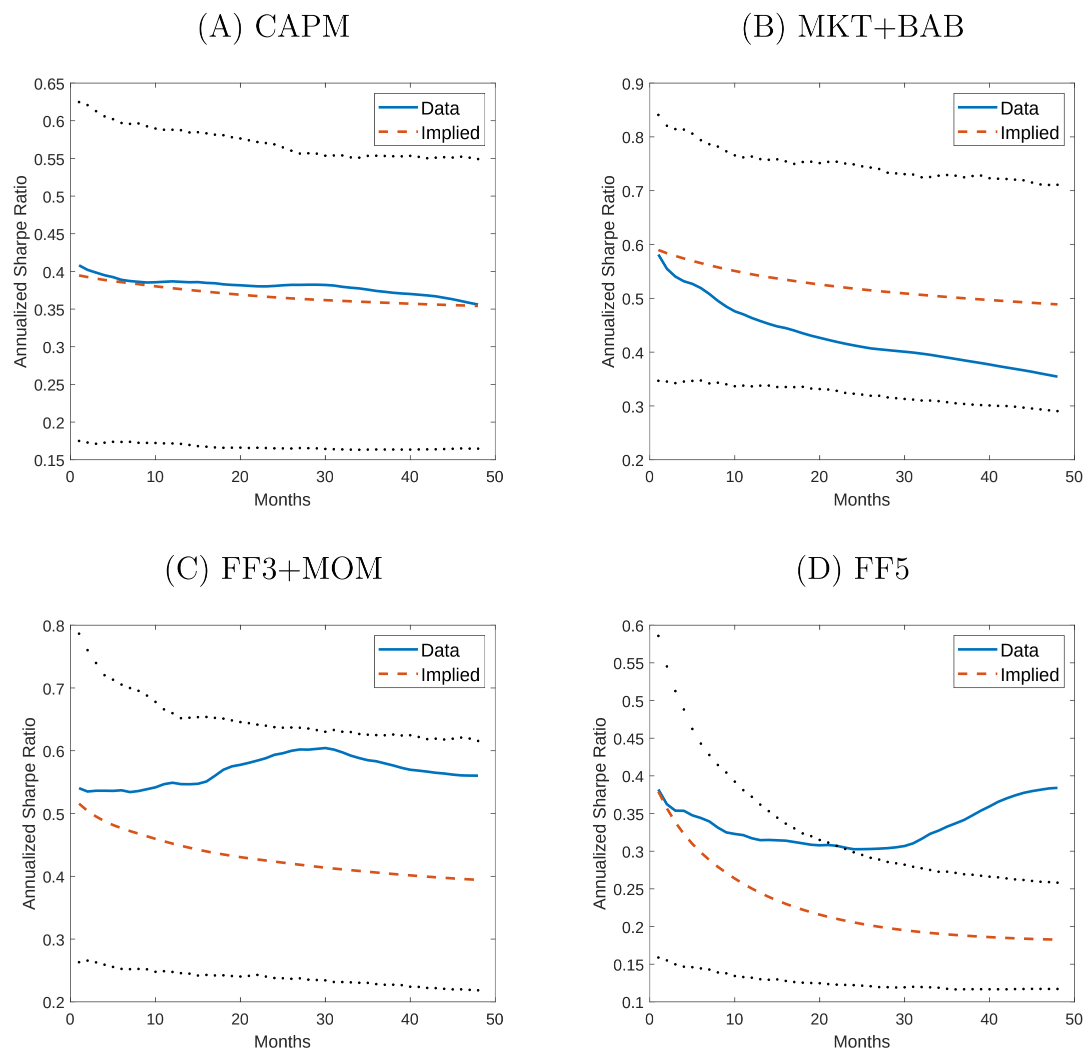
The panels show cumulative autocorrelation coefficients for one factor from each model from the 1- to 48-month horizon. For each model the chosen factor is the one with the maximal average pricing error (see Figure 1). The chosen factors are MKT (CAPM), BAB (MKT+BAB), MOM (FF3+MOM), and RMW (FF5). The red, dashed lines give the cumulative autocorrelation of the factor returns as implied by the model the factor belongs to, and the dotted lines give the associated 90% confidence bands. The sample is monthly, from 1963 to 2017.

Figure 7
Cumulative autocorrelations II



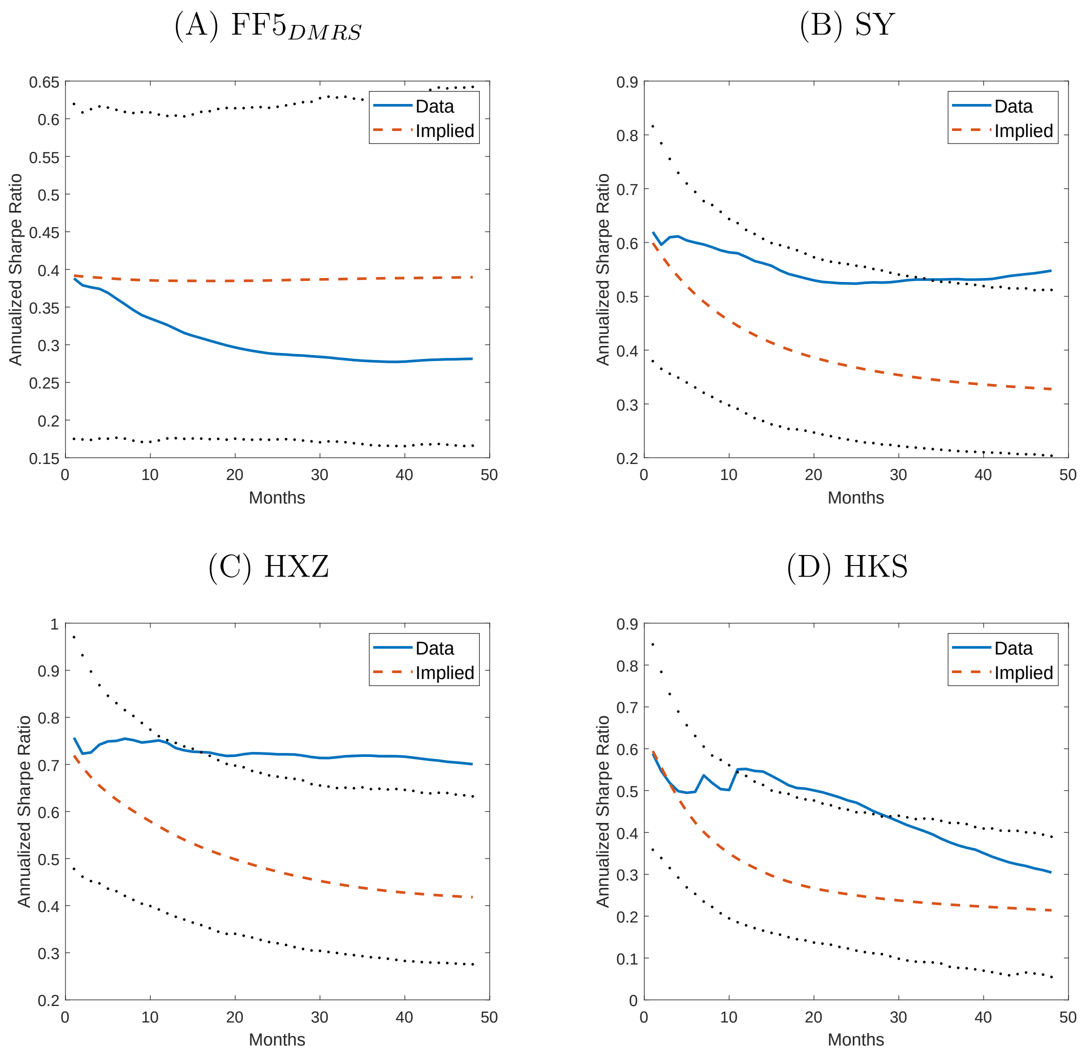
The panels show cumulative autocorrelation coefficients for one factor from each model from the 1- to 48-month horizon. For each model the chosen factor is the one with the maximal average pricing error (see Figure 2). The chosen factors are SMB (FF5$_{DMRS}$), PERF (SY), ROE (HXZ), and PC2 (HKS). The red, dashed lines give the cumulative autocorrelation of the factor returns as implied by the model the factor belongs to, and the dotted lines give the associated 90% confidence bands. The sample is monthly, from 1963 to 2017 for FF5$_{DMRS}$, 1963 to 2016 for SY, 1967 to 2017 for HXZ, and 1974 to 2017 for HKS.

Figure 8
Term structure of Sharpe ratios I



The panels show annualized Sharpe ratios for one factor from each model. For each model the chosen factor is the one with the maximal average pricing error (see Figure 1). The chosen factors are MKT (CAPM), BAB (MKT+BAB), MOM (FF3+MOM), and RMW (FF5). In each case, we add the gross real risk-free rate to the factor return and get h-period returns to this portfolio as $R_{t,t+h} = R_{t,t+1} \times R_{t+1,t+2} \times ... \times R_{t+h-1,t+h}$. The h-period risk-free rate is found in the same way. We then calculate the h-period annualized Sharpe ratio as $\sqrt{12/h} \times E(R_{t,t+h} - R^f_{t,t+h})/V^{1/2}(R_{t,t+h} - R^f_{t,t+h})$ (solid, blue line). The red, dashed lines give the annualized Sharpe ratios as implied by the model the factor belongs to, and the dotted lines give the associated 90% confidence bands. The sample is monthly, from 1963 to 2017.

Figure 9
Term structure of Sharpe ratios II



The panels show annualized Sharpe ratios for one factor from each model. For each model the chosen factor is the one with the maximal average pricing error (see Figure 2). The chosen factors are SMB ($FF5_{DMRS}$), PERF (SY), ROE (HXZ), and PC2 (HKS). We add the gross real risk-free rate to the factor return and get h-period returns to this portfolio as $R_{t,t+h} = R_{t,t+1} \times R_{t+1,t+2} \times \ldots \times R_{t+h-1,t+h}$. The h-period risk-free rate is found similarly. The h-period annualized Sharpe ratio is then $\sqrt{12/h} \times E(R_{t,t+h} - R^f_{t,t+h})/V^{1/2}(R_{t,t+h} - R^f_{t,t+h})$ (solid, blue line). The red, dashed lines give the annualized Sharpe ratios as implied by the model the factor belongs to, and the dotted lines give the associated 90% confidence bands. The sample is monthly, from 1963 to 2017 for $FF5_{DMRS}$, from 1963 to 2016 for SY, 1967 to 2017 for HXZ, from 1974 to 2017 for HKS.

Table 1: MHR tests of linear factor models

Panel A:	**CAPM**	**MKT+BAB**	**FF3+MOM**	**FF5**
$p-$value (GMM)	0.191	0.059	0.073	0.022
MAPE	0.007	0.030	0.076	0.019
Max Sharpe Ratio	0.395	0.701	1.004	1.116
Max Information Ratio	0.380	0.613	0.909	1.025

Panel B:	$\mathbf{FF5}_{DMRS}$	**SY**	**HXZ**	**HKS**
$p-$value (GMM)	0.006	0.024	0.023	0.025
MAPE	0.034	0.119	0.061	0.068
Max Sharpe Ratio	1.588	1.670	1.431	1.308
Max Information Ratio	1.017	0.939	0.907	1.135

The first row of each panel gives the $p-$value from the GMM $J-$test, where the linear factor models are estimated on the one-period factor returns and tested on multi-horizon factor returns. The second row displays the annualized mean absolute price error ($MAPE$) across the test assets. The returns horizons used are 1, 3, 6, 12, 24, and 48 months. The table also reports the sample Sharpe ratio of the in-sample MVE combination of each model's factors, as well as maximal annualized Information ratio implied by the MHR returns. The sample is monthly, from 1963 to 2017 for all models except SY, which is 1963 to 2016, HXZ, which is 1967 to 2017, and HKS, which 1974 to 2017.

Table 2: MHR tests of conditional linear factor models: out-of-sample conditioning

Panel A: **Volatility Timing**	**CAPM**	**FF3+MOM**	**FF5**
$p-$value (GMM)	0.037	0.000	0.000
MAPE	0.014	0.051	0.077
Max Sharpe Ratio	0.416	1.183	1.149
Max Information Ratio	0.642	1.172	1.245

Panel B: **HKS Factor Timing**	E_t **only**	E_t **and** V_t
$p-$value (GMM)	0.040	0.026
MAPE	0.032	0.025
Max Sharpe Ratio	1.242	1.109
Max Information Ratio	1.674	1.709

Panel C: **KNS Stock Timing**	$L^1 - L^2$ **penalty**	L^1 **penalty**
$p-$value (GMM)	0.014	0.000
MAPE	0.034	0.066
Max Sharpe Ratio	3.898	3.028
Max Information Ratio	0.659	0.768

Panels A and B report test statistics from the factor models with time-varying SDF loadings b_t, as opposed to the constant b model tests given in Table 1. The b_t are computed out-of-sample. The exception is HKS, where E_t is computed out-of-sample but the conditional covariance V_t is computed in-sample. The first row of each panel gives the p-value from the GMM J-test. The returns horizons used in the test are 1, 3, 6, 12, 24, and 48 months. The second row gives the mean absolute pricing errors (MAPE) of the model factors across horizons, excluding the pricing errors of the timing portfolio so as to be comparable to the MAPEs in Table 1. The third row gives the sample annualized maximal Sharpe ratio as implied by the model with time-varying b_t, while the fourth row gives the maximal Information ratio implied by all MHR returns. The sample is monthly, from 1963 to 2017 for Volatility Timing, and from 1996 to 2017 for HKS. Panel C shows results from the UMVE construction using non-linear functions of characteristics and machine learning methods by Kozak, Nagel, and Santosh (2020). Here the sample is daily, from 1974 to 2017.

Table A1: MHR tests of linear factor models: HAR standard errors

Panel A:	**CAPM**	**MKT+BAB**	**FF3+MOM**	**FF5**
$p-$value (GMM)	0.693	0.000	0.000	0.000

Panel B:	**FF5**$_{DMRS}$	**SY**	**HXZ**	**HKS**
$p-$value (GMM)	0.000	0.000	0.000	0.000

The first row of each panel gives the $p-$value from the GMM $J-$test, where the linear factor models are estimated on the one-period factor returns and tested on multi-horizon factor returns. Different from Table 1, the $p-$value is calculated via standard GMM formulas with HAR-adjustment, as opposed to fully imposing the null hypothesis as we do in our porposed test. The return horizons used are 1, 3, 6, 12, 24, and 48 months as before. The sample is monthly, from 1963 to 2017 for all models except SY, which is 1963 to 2016, HXZ, which is 1967 to 2017, and HKS, which 1974 to 2017.

Table A2: MHR tests of linear factor models: out-of-sample instruments

Panel A:	**CAPM**	**MKT+BAB**	**FF3+MOM**	**FF5**
$p-$value (GMM)	0.582	0.201	0.044	0.146
MAPE	0.005	0.030	0.102	0.172
Max Sharpe Ratio	0.508	0.781	0.931	1.212
Max Information Ratio	0.366	0.849	1.040	1.180

Panel B:	**FF5**$_{DMRS}$	**SY**	**HXZ**	**HKS**
$p-$value (GMM)	0.000	0.001	0.165	0.041
MAPE	0.190	0.317	0.110	0.172
Max Sharpe Ratio	1.718	1.662	1.372	1.372
Max Information Ratio	1.584	1.257	0.903	1.264

The first row of each panel gives the $p-$value from the GMM $J-$test, where the linear factor models are estimated on the one-period factor returns and tested on multi-horizon factor returns. In this case, the timing weights are estimated in an out-of-sample fashion. The second row displays the annualized mean absolute price error ($MAPE$) across the test assets. The returns horizons used are 1, 3, 6, 12, 24, and 48 months. The table also reports the sample Sharpe ratio of the in-sample MVE combination of each model's factors, as well as maximal annualized Information ratio implied by the MHR returns. The sample is monthly, starting in July 1983 for all models, with the earlier data used as a burn-in period to help with estimation of the out-of-sample instruments.

Table A3: MHR tests of linear factor models: MHR to the same assets

Panel A:	**CAPM**	**MKT+BAB**	**FF3+MOM**	**FF5**
$p-$value (SHR)	0.000	0.000	0.000	NaN
$p-$value (GMM)	0.000	0.000	0.002	0.022
MAPE	0.032	0.026	0.029	0.019
Max Sharpe Ratio	0.395	0.701	1.004	1.116
Max Information Ratio	1.238	1.118	1.079	1.025

Panel B:	$\mathbf{FF5}_{DMRS}$	**SY**	**HXZ**	**HKS**
$p-$value (SHR)	0.562	0.652	0.670	0.010
$p-$value (GMM)	0.015	0.036	0.037	0.005
MAPE	0.037	0.061	0.031	0.023
Max Sharpe Ratio	1.588	1.671	1.429	1.301
Max Information Ratio	1.021	0.958	1.001	1.144

In this table all models are tested using MHR to the FF5 factors. The models are estimated by pricing the one-period returns to the model's own factors without error. The second row of each panel gives the $p-$value from the GMM $J-$test, where the linear factor models are estimated on the one-period factor returns and tested on multi-horizon factor returns. The third row displays the annualized mean absolute price error ($MAPE$) across the test assets. The returns horizons used are 1, 3, 6, 12, 24, and 48 months. The table also reports the sample Sharpe ratio of the in-sample MVE combination of each model's factors, as well as maximal annualized Information ratio implied by the MHR returns. The first row gives the p-values when the additional test assets are only single-period return to the FF5 assets. The sample is monthly, from 1963 to 2017 for all models except SY, which is 1963 to 2016, HXZ, which is 1967 to 2017, and HKS, which 1974 to 2017.

Table A4: MHR tests of conditional linear factor models

Panel A:	**CAPM**	**MKT+BAB**	**FF3+MOM**	**FF5**
$p-$value (GMM)	0.878	0.189	0.020	0.007
MAPE	0.003	0.044	0.088	0.055
Max Sharpe Ratio	0.396	0.919	1.226	1.380
Max Information Ratio	0.347	0.667	0.974	1.093

Panel B:	**FF5**$_{DMRS}$	**SY**	**HXZ**	**HKS**
p-value (GMM)	0.002	0.000	0.003	0.018
MAPE	0.200	0.126	0.201	0.122
Max Sharpe Ratio	1.832	1.604	1.631	1.540
Max Information Ratio	1.140	1.385	1.081	1.261

This table reports test statistics from the factor models with time-varying SDF loadings b_t, as opposed to the constant b model tests given in Table 1. The first row of each panel gives the p-value from the GMM J-test. The returns horizons used in the test are 1, 3, 6, 12, 24, and 48 months. The second row gives the mean absolute pricing errors (MAPE) of the model factors across horizons, excluding the pricing errors of the timing portfolio so as to be comparable to the MAPEs in Table 1. The third row gives the sample annualized Sharpe ratio of the unconditional MVE portfolio as implied by the model with time-varying b_t, as well as maximal annualized Information ratio implied by all MHR returns. The sample is monthly, from 1963 to 2017 for all models except SY, which is 1963 to 2016, HXZ, which is 1967 to 2017, and HKS, which 1974 to 2017.

The Impact of Policy on the Risk-Return Relationship

Stig R. H. Lundeby

May 10, 2021

Abstract

I incorporate counter-cyclical monetary and fiscal policy into an otherwise standard long-run risk model. The policy acts as a partial insurance to investors. As insurances are negative "beta" assets, the policy typically earns a negative risk premium. Furthermore, insurance becomes more valuable in risky times, thereby weakening the risk-return trade-off. I also find that a policy intended to insure the market can at times *increase* risk and risk-premia.

1 Introduction

Negative news about the economy, often signaled or accompanied by stock market turmoil, frequently triggers fiscal and monetary stimulus. The policy response has in recent decades been both swift and strong. Market participants therefore expect substantial policy response to significantly negative stock market returns also going forward.

This paper incorporates fiscal and monetary policy into a standard long-run risk economy (Bansal and Yaron, 2004). The policy response is triggered by poor economic or stock market performance, thus effectively insuring investors. This insurance is a negative beta asset which means it carries a negative risk premium. When overall risk in the economy increases, the insurance becomes more valuable and thus more strongly impacts the stock market. The insurance thereby also gains a stronger impact on expected return on the stock market. Thus, increased risk is attenuated by increased insurance, thereby dampening the increase in expected return in response to increased risk.

> "A rebound fueled by unprecedented fiscal and monetary efforts has pushed global indexes back up, though economies remain fragile and life in many places is on hold amid a slower-than-expected vaccine rollout"
>
> - *Bloomberg, February 12, 2021* [a]
>
> [a]https://www.bloomberg.com/markets/fixed-income

The current pandemic illustrates the point well. Between February 20, 2020 and March 20, 2020, the S&P 500 fell by a little over 30%. By March 26, 2020 S&P 500 had recovered about one third of the initial loss and by April 20, 2020 half the initial loss was recovered. In fact, the S&P 500 had completely recovered by August and, despite second and third waves of infections, closed the year 16% higher than it started. For comparison, the Bureau of Economic Analysis reports that real GDP in the fourth quarter of 2020 was 2.4% smaller than fourth quarter 2019.[1] Over the same time period we saw fiscal and monetary responses of an unprecedented scale. Elgin, Basbug and Yalaman (2020) calculates that as of April 2020 the fiscal stimulus package in the US amounted to 10.50% of GDP. As of January 7, 2021, this figure had risen to 18.22% of GDP. Including the latest stimulus package from the Biden administration, the Washington Post reports that the total US fiscal stimulus amounts to an astonishing 27.09% of GDP.[2]

The US also saw exceptionally accommodating monetary policy. On unscheduled FOMC meetings on March 3, 2020 and later on March 15, 2020 the Federal Reserve cut its target rate by 0.5 and 1 percentage points respectively. Additionally, on the March 15 meeting, the Federal Reserve also announced that it would purchase at least $500 billion of Treasuries and at least $200 billion of agency mortgage-backed securities over the coming months. In fact, the Federal Reserve purchased $3 trillion of assets in the space

[1]Source: https://www.bea.gov/news/2021/gross-domestic-product-fourth-quarter-and-year-2020-second-estimate

[2]The Washington Post on March 10, 2021 (source: https://www.washingtonpost.com/world/2021/03/10/coronavirus-stimulus-international-comparison/)

of three months, almost doubling its balance sheet (Putniņš, 2020). Putniņš (2020) calculates that the aggressive monetary response accounted for one third of the rebound in stock markets since March 2020. A similar conclusion is reached in an IMF report:

> "The disconnect between the performance of stock markets and the real economy ... has become a topic of much interest and debate. After considering several hypotheses for the disconnect, this note concludes that the most compelling is that unprecedented monetary policy actions, while needed to stem the impact of COVID-19, have driven asset prices up"
> - *Igan, Kirti and Peria (2020)*

The policy response during the pandemic is not an isolated event, except perhaps for the scale. The financial crisis of 2007-2009 saw large fiscal and monetary responses both in the US and around the world. Policies ranged from rate cuts and large scale asset purchase programs (QE) by central banks to direct fiscal transfers, tax reductions, and increased government spending. Indeed, counter-cyclical government responses to recessions is nothing new. Ever since the Great Depression and the ensuing stimulus packages, like the New Deal, many governments and central banks have adopted Keynesian ideas by seeking to boost output and employment when economic downturns threaten. Thus, it seems plausible that investors expect some form of government stimulus when times are bad in the overall economy.

Furthermore, there seems to be an increasing amount of evidence for policy not only responding to economy-wide downturns, but also to worsening conditions in asset markets. For instance, Cieslak and Vissing-Jorgensen (2020) use textual analysis of FOMC documents to find that negative stock-market mentions predict federal funds rate cuts in the period since mid-1990s. Rigobon and Sack (2003) estimates that a 5% drop in the stock market is associated with a 50 percentage point increase in the probability of a 25 bp monetary easing. Bernanke and Kuttner (2005) estimates that a 25 bp (unexpected) monetary easing causes the stock market to increase by about one percent. There is an asymmetry in the Federal Reserve's responses to stock market performance. Fed policy responds promptly to falling stock markets while it does not seem to tighten policy in response to rising markets (e.g. Dahiya, Kamrad, Poti and Siddique (2019) and Putniņš (2020)). This asymmetry in policy response has received a lot of attention by practitioners to the point that terms like "the Greenspan put", "the Yellen put", "the Powell put" etc. has become so common that many now just refer to "the Fed put".

> "It's official: there is a Greenspan put option. Yesterday's half a percentage point interest rate cut by the US Federal Reserve may not have been designed explicitly to bail out the stock market. But that is exactly what it is in danger of doing - especially since the cut came between official meetings, thereby heightening its impact"
> - *The Financial Times, January 4, 2001*

Nor is the monetary policy restricted to adjusting policy rates. Since the financial crisis, central banks around the world have increasingly relied on QE to achieve their policy goals. While there have been considerable debate about the effectiveness of QE on the broader economy and the channels through which it works, there is an increasing amount of evidence that QE boosts asset prices (e.g. Barbon and Gianinazzi (2019), Putniņš (2020)). QE comes in many flavors, the Federal Reserve buys Treasuries and mortgage backed securities, while the Bank of Japan buys exchange traded funds (ETFs) to the extent that it has become one of the largest shareholders in several Nikkei index companies (Barbon and Gianinazzi (2019)). Furthermore, the Federal Reserve balance sheet seems to behave in a manner consistent with a "Fed put". Putniņš (2020) estimates that a 10% fall in the stock market is associated with a balance sheet expansion of about 5.5%, while a positive return does not result in a similar balance sheet contraction.

Clearly there is a large degree of heterogeneity in policies and the channels through which they operate. In this paper I wish to capture the insurance aspect. Government policy is therefore modeled as transfers to the market from the non-traded economy, where the rule governing the transfers is allowed to be very general. This allows me to capture two important features: (1) policy is expansive in "bad" times, (2) expansive policy increases market prices all else equal. Since I consider a representative agent, endowment economy, the policy does not affect the pricing kernel. As a consequence, the results in this paper is driven by the *correlation* between market return and the pricing kernel. In particular, when risk is high, the insurance becomes valuable and the correlation between market returns and the pricing kernel decreases in magnitude.

Bruno and Haug (2018) show that leverage causes idiosyncratic return volatility to be correlated with expected return. They arrive at this conclusion by considering equity as a call option on assets. In my paper the government policy can be thought of as put options. The market is then a portfolio of the "fundamental" market and put options, which we expect should behave much like a call. My paper differs from theirs in a few important aspects. First, they take the leverage as given, which translates into a fixed (known) strike price. In my setup, government policy reacts to shocks in the economy. This gives rise to an interesting term structure of discount rates for the policy. Since the payoffs from policies in the near future tend to be negatively related to the economy, they earn a negative risk premium. However, as long as the government stimulus relative to the size of the stock market is stationary, i.e. the size of the stimulus does not persistently grow or fall as a fraction of market capitalization, the payoffs from government policies further in the future must be *positively* related to the economy. As a result, discount rates for policy payoffs far enough into the future must be positive. The policy response is akin to the government writing put options. Since the policy is expected to continue in the future, the policy acts as a portfolio of put options. Some of these options have already been written in response to recent economic developments while a substantial part of the put portfolio reflect expected future policy. This expected future policy is akin to claims issued today that turn into options in the future with strike prices determined at the time of the future issuance. For plausible policies, the strike prices are positively related to economic fundamentals. Thus, the future options have a similar risk exposure as the overall market.

There are several papers that have shown a seeming disconnect between the risk in the stock market as measured by its variance and expected returns, e.g. Moreira and Muir (2017) shows that a timing strategy that takes a smaller (larger) position in the market when the variance is high (low) earns alpha with respect to the market and increases the Sharpe ratio. Similarly, Glosten, Jagannathan and Runkle (1993) find a negative relationship between conditional variance and conditional expected returns. These results are puzzling from a theoretical point of view, as most models imply that the risk premium should be higher in risky times. In my setup, the government policy is able to generate the weak risk-return trade-off observed in these papers.

Lochstoer and Muir (2020) show that the weak risk-return trade-off uncovered in these papers can be explained by investors having biased expectations about volatility. In my paper, investors are fully rational, but the presence of a counter-cyclical policy changes both the market's conditional variance and conditional correlation with the pricing kernel. The variance of market returns therefore becomes *less* connected to priced risk.

2 Model

In this section I present the economy and some theoretical results. The starting point is a long-run risk endowment economy similar to that in e.g. Bansal and Yaron (2004), which I will refer to as the fundamental economy. The aggregate consumption in the economy is the sum of dividends on a stock market index, henceforth the fundamental market, and income from other sources, e.g. human capital. I will then introduce a government policy that acts as a partial insurance to investors in the fundamental market while leaving aggregate consumption unchanged. In other words, the policy payoff at any time is financed by the non-traded economy. This implies the pricing kernel is the same in the insured economy as in the fundamental economy. While it would certainly be interesting to allow for macroeconomic effects of the policy, I show that even in the absence of such effects, government policy can have significant impact on risk-return relationships in financial markets. This is particularly interesting given the debate about the efficacy of QE on stimulating the broader economy.

2.1 Fundamental Economy

The starting point is an economy similar to that in Bansal and Yaron (2004). The investor has Epstein-Zin preferences where the value function is

$$V_t \equiv \max_{C_t,\omega_t} \left\{ C_t^{1-\frac{1}{\psi}} + \beta \mathbb{E}_t (V_{t+1}^{1-\gamma})^{\frac{1-\frac{1}{\psi}}{1-\gamma}} \right\}^{\frac{1}{1-\frac{1}{\psi}}} \tag{1}$$

β is the discount factor of future utility, ψ is the intertemporal elasticity of substitution (IES) and γ is the coefficient of relative risk aversion. The budget constraint is

$$W_{t+1} = (W_t - C_t)\omega_t^\top R_{t+1} \equiv (W_t - C_t) R_{\omega,t+1} \tag{2}$$

ω_t is a vector of time t portfolio weights, R_{t+1} a vector of asset returns and $R_{\omega,t+1}$ is the corresponding portfolio return. We allow for the possibility of non-traded assets (like human capital), in which case the dividend on the non-traded asset is simply the consumption it offers each period. The return on such an asset would only be implicit - the price is set such that the investor would not want to trade it even if he could. Epstein and Zin (1991) shows that the stochastic discount factor in this economy can be written (see appendix A)

$$M_{t+1} = \beta^\theta \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{\theta}{\psi}} R_{c,t+1}^{\theta-1} \tag{3}$$

where $\theta \equiv \frac{1-\gamma}{1-\frac{1}{\psi}}$ and $R_{c,t+1}$ is the implicit return on an asset that pays aggregate consumption as dividends each period. It is worth noting that the CRRA expected utility SDF obtains if $\gamma = \frac{1}{\psi}$. If $\gamma > \frac{1}{\psi}$, we have preference for early resolution of uncertainty. We usually consider $\gamma > 1$ and $\psi \geq 1$, i.e. the investor is more risk-averse and more willing to substitute (risk-free) consumption across time than a log-utility investor. These parameter choices would imply $\theta < 0$.

Suppose the aggregate log consumption and the log dividend on a broad stock portfolio (the fundamental market) evolves as follows

$$\begin{aligned}
\Delta c_{t+1} &= \mu_c + x_{t+1} - \frac{\tau}{2}\sigma_t^2 - \frac{\tau_x}{2}\sigma_{x,t}^2 + \sigma_t \varepsilon_{t+1} \\
\Delta d_{t+1} &= \mu_d + \varrho \Delta c_{t+1} - \frac{\tau_d}{2}\sigma_{d,t}^2 + \sigma_{d,t}\varepsilon_{d,t+1} \\
x_{t+1} &= \rho_x x_t + \sigma_{x,t}\varepsilon_{x,t+1} \\
\sigma_{t+1}^2 &= \varphi \sigma_t^2 + \eta_{t+1} \\
\sigma_{d,t+1}^2 &= \varphi_d \sigma_{d,t}^2 + \eta_{d,t+1} \\
\sigma_{x,t+1}^2 &= \varphi_x \sigma_{x,t}^2 + \eta_{x,t+1}
\end{aligned}$$

where $\rho_x,\ \varphi,\ \varphi_x,\ \varphi_d\ \in (0,1)$. The shocks are assumed independent of each other and across time with

$(\varepsilon_{t+1}, \varepsilon_{x,t+1}, \varepsilon_{d,t+1})' \sim N(0, I)$ and $\eta_{i,t+1} \sim IG\big((1-\varphi_i)\sigma_i^2, \lambda_i\big)$. The variance shocks have means

$$\mathbb{E}_t(\eta_{i,t+1}) = (1-\varphi_i)\sigma_i^2$$

i.e. the variances follow AR(1) processes with means σ_i^2. Throughout, I will refer to the $\varepsilon_{\cdot,t+1}$ shocks as cash-flow shocks, σ_t^2, $\sigma_{x,t}^2$ as systematic risk and $\sigma_{d,t}^2$ as idiosyncratic/unsystematic risk.

This dividend and consumption process is a generalization of Bansal and Yaron (2004) in that we allow current variance processes to directly affect expected consumption and dividend growth. In their set-up, it is assumed that variances does not forecast dividends and consumption in logs, with the consequence that variances must forecast levels positively. In my set-up, setting $\tau_i = 1$ implies levels are not forecasted by variance, setting $\tau_i > 1$ implies levels are negatively forecasted by variance and vice versa for $\tau_i < 1$. There might be some reasonable arguments for why variances should forecast lower growth, e.g. a greater uncertainty might slow down investments in positive NPV risky projects until more information arrives, resulting in a lower growth for dividends or consumption.

Another slight deviation from the typical set-up is that I assume the variance shocks are Inverse Gaussian distributed. The Inverse Gaussian distribution has two attractive features over the more standard assumption that shocks to variance is normal: (1) the Inverse Gaussian distribution has positive support, eliminating the possibility of negative variances, (2) the distribution is positively skewed, which is in line with what we typically observe for variance processes. Additionally, the Inverse Gaussian distribution (like the normal distribution) has a moment generating function, which makes it tractable to find an approximate partial analytical solution to the model using the same log-linearization technique as employed by Bansal and Yaron (2004) and Campbell and Shiller (1988).

To price the cash flows, let us rewrite the discount factor in logs as

$$m_{t+1} = \theta log(\beta) - \frac{\theta}{\psi}\Delta c_{t+1} + (\theta - 1) r_{c,t+1} \tag{4}$$

Note that the log return to an asset i can be written as (see Appendix B)

$$r_{i,t+1} = log\big(e^{z_{i,t} - z_{i,t+1}} + e^{z_{i,t}}\big) + \Delta d_{i,t+1}$$

where $z_{i,t} \equiv log(\frac{D_{i,t}}{P_{i,t}})$. We can approximate the log-return by taking a first-order Taylor approximation of $log\big(e^{z_{i,t} - z_{i,t+1}} + e^{z_{i,t}}\big)$ around $\mathbb{E}(z_{i,t+1}) = \mathbb{E}(z_{i,t}) \equiv z_i$. We then get (see Appendix B)

$$r_{i,t+1} = \kappa_{i,0} + z_{i,t} - \kappa_{i,1} z_{i,t} + \Delta d_{i,t+1}$$

In particular, for the implicit return $r_{c,t+1}$ we have

$$r_{c,t+1} = \kappa_{c,0} + z_{c,t} - \kappa_{c,1} z_{c,t+1} + \Delta c_{t+1} \tag{5}$$

and the log-SDF is

$$m_{t+1} = \theta log(\beta) - \Big(1 - \theta + \frac{\theta}{\psi}\Big)\Delta c_{t+1} + (\theta - 1)(\kappa_{c,0} + z_{c,t} - \kappa_{c,1} z_{c,t+1}) \tag{6}$$

The log consumption-price ratio can be written as an affine function of the state variables x_t, σ_t^2 and $\sigma_{x,t}^2$ as follows [3].

$$z_{c,t} = A_c + A_{c,x} x_t + A_{c,\sigma^2} \sigma_t^2 + A_{c,\sigma_x^2} \sigma_{x,t}^2 \tag{7}$$

Substituting in the assumed processes for the state variables in $t+1$ gives us the following expression for the return and SDF (see Appendix D)

$$r_{c,t+1} = \Gamma_{r_c} + \Gamma_{r_c,x} x_t + \Gamma_{r_c,\sigma^2} \sigma_t^2 + \Gamma_{r_c,\sigma_x^2} \sigma_{x,t}^2 + \Gamma_{r_c,\varepsilon} \sigma_t \varepsilon_{t+1} + \Gamma_{r_c,\varepsilon_x} \sigma_{x,t} \varepsilon_{x,t+1} + \Gamma_{r_c,\eta} \eta_{t+1} + \Gamma_{r_c,\eta_x} \eta_{x,t+1}$$

$$m_{t+1} = \Gamma_m + \Gamma_{m,x} x_t + \Gamma_{m,\sigma^2} \sigma_t^2 + \Gamma_{m,\sigma_x^2} \sigma_{x,t}^2 - \Gamma_{m,\varepsilon} \sigma_t \varepsilon_{t+1} - \Gamma_{m,\varepsilon_x} \sigma_{x,t} \varepsilon_{x,t+1} - \Gamma_{m,\eta} \eta_{t+1} - \Gamma_{m,\eta_x} \eta_{x,t+1}$$

In Appendix D, I show that $\Gamma_{m,\eta}$ and Γ_{m,η_x} are negative for plausible parameter values. Thus, shocks to variance would be positively related to marginal utility. Assets that pay off more in such states, should therefore carry a negative risk premium. A put option is one such asset.

Similarly, we can log-linearize the return to the fundamental market portfolio. In this case, the log dividend-price ratio is also a function of the idiosyncratic volatility since it enters the dividend process directly [4]

$$z_{d,t} = A_d + A_{d,x} x_t + A_{d,\sigma^2} \sigma_t^2 + A_{d,\sigma_x^2} \sigma_{x,t}^2 + A_{d,\sigma_d^2} \sigma_{d,t}^2 \tag{8}$$

And the resulting log-return is given by (see Appendix D)

$$\begin{aligned} r_{d,t+1} = {} & \Gamma_{r_d} + \Gamma_{r_d,x} x_t + \Gamma_{r_d,\sigma^2} \sigma_t^2 + \Gamma_{r_d,\sigma_x^2} \sigma_{x,t}^2 + \Gamma_{r_d,\sigma_d^2} \sigma_{d,t}^2 \\ & + \Gamma_{r_d,\varepsilon} \sigma_t \varepsilon_{t+1} + \Gamma_{r_d,\varepsilon_x} \sigma_{x,t} \varepsilon_{x,t+1} + \Gamma_{r_d,\varepsilon_d} \sigma_{d,t} \varepsilon_{d,t+1} + \Gamma_{r_d,\eta} \eta_{t+1} + \Gamma_{r_d,\eta_x} \eta_{x,t+1} + \Gamma_{r_d,\eta_d} \eta_{d,t+1} \end{aligned} \tag{9}$$

[3] Note that in a representative agent economy, the consumption-price ratio on the consumption portfolio should not depend on the idiosyncratic volatility $\sigma_{d,t}$ as this does not affect current or future consumption. However, the dividend-price ratio on the fundamental market portfolio should depend on $\sigma_{d,t}$ as it affects the expected dividend growth on this portfolio.

[4] Note that the idiosyncratic volatility cannot predict stock market returns in levels. The flip side is that it must predict stock market returns in logs.

Proposition 1. *The conditional (log-) risk premium on the fundamental market is*

$$log\mathbb{E}_t(R_{d,t+1}) - r_{f,t} = \alpha + \gamma\varrho\sigma_t^2 + \frac{\gamma - \frac{\kappa_{c,1}\rho_x}{\psi}}{1-\kappa_{c,1}\rho_x}\frac{\varrho - \frac{\kappa_{d,1}\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x}\sigma_{x,t}^2 \tag{10}$$

Furthermore, the risk premium is increasing in σ_t^2.

Proof. See Appendix D.3. ∎

Corollary 1 (to Proposition 1)**.** $\gamma\psi \geq 1$ *and* $\varrho\psi \geq 1$ *is a sufficient condition for the conditional (log-) risk premium on the fundamental market to be increasing in* $\sigma_{x,t}^2$.

Proof. Recall that $\kappa_{c,1},\ \kappa_{d,1},\ \rho_x \in (0,1)$. The result then follows directly from Proposition 1. ∎

Proposition 1 tells us that the conditional risk premium on the fundamental market is affine in systematic risk and is always increasing in short-run risk σ_t^2. The corollary tells us that if the investor has a preference for late resolution of uncertainty and the fundamental market is sufficiently exposed to consumption shocks, the risk premium is also increasing in long-run risk $\sigma_{x,t}^2$. In particular, the typical parameter choices $\gamma,\ \psi,\ \varrho \geq 1$ satisfies the condition in Corollary 1. Under the condition of Corollary 1 the risk premium is increasing in the risk aversion γ, the intertemporal elasticity of substitution ψ, the persistence of consumption growth ρ_x and the market exposure ϱ to consumption risk. The risk aversion and consumption exposure affects both the compensation for short- and long-run risk, whereas the IES and persistence of consumption growth only affects compensation to long-run risk.

Proposition 2. *Suppose we regress the fundamental market risk premium on the fundamental market return variance as follows*

$$log\mathbb{E}_t(R_{d,t+1}) - r_{f,t} = \phi_0 + \phi_1\mathbb{V}_t(r_{d,t+1}) + u_{t+1} \tag{11}$$

The population regression coefficient is then given by

$$\phi_1 = \frac{\gamma\varrho^3\mathbb{V}(\sigma_t^2) + \frac{\gamma - \frac{\kappa_{c,1}\rho_x}{\psi}}{1-\kappa_{c,1}\rho_x}\left(\frac{\varrho - \frac{\kappa_{d,1}\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x}\right)^3\mathbb{V}(\sigma_{x,t}^2)}{\varrho^4\mathbb{V}(\sigma_t^2) + \left(\frac{\varrho - \frac{\kappa_{d,1}\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x}\right)^4\mathbb{V}(\sigma_{x,t}^2) + \mathbb{V}(\sigma_{d,t}^2)} \tag{12}$$

Furthermore, $\gamma\psi > 1$ *and* $\varrho\psi > 1$ *is a sufficient condition for the regression coefficient to be positive.*

Proof. See Appendix D.3. ∎

Proposition 2 tells us that if the fundamental market risk premium is increasing in long-run risk $\sigma_{x,t}^2$ it is also higher on average when the return variance is high. Importantly, increasing the variance of idiosyncratic risk, lowers the magnitude of the regression coefficient, but does not alter its sign. Additionally, we can see that a higher risk aversion increases the regression coefficient, causing a stronger relationship between fundamental market return variance and risk-premium.

2.2 Insurance policy

Suppose the government introduces a form of insurance policy to the market. In particular, at each time $T_h = t + h$, the government policy implies a payoff

$$X_{T_h} = e^{q_{T_h}(s)}(K_{T_h} - P_{T_h} - D_{T_h})^+ = P_{T_h-1}e^{q_{T_h}(s)}(e^{-g_{T_h}(s)} - e^{r_{d,T_h}})^+ \tag{13}$$

to the market, where $g_{T_h} \equiv log\Big(\frac{P_{T_h-1}}{K_{T_h}}\Big)$. P_{T_h} and D_{T_h} are the price and dividend on the fundamental market at time T_h.

Each of the payoffs can be priced separately and considered as separate assets or claims. Since the payoffs in (13) are reminiscent of put option payoffs, I will often refer to each claim as a put option.

The policy is specified in terms of the pair $(q_{T_h}(s), g_{T_h}(s))$. We can think of $q_{T_h}(s)$ as governing the number of claims the government issues to the market, and $g_{T_h}(s)$ governs the strike price of those claims. A partial insurance would imply $e^{q_{T_h}(s)} \in (0, 1)$. The number s is a constant which I sometimes will refer to as the duration of the options. Both $q_{T_h}(s)$ and $g_{T_h}(s)$ are allowed in general to depend on any shocks and state variables between time $T_h - s$ and T_h, i.e.

$$\begin{aligned} g_{T_h}(s) = \bar{g}_0 + \sum_{j=0}^{s} \Big(&\bar{g}_{x,j}x_{T_h-j} + \bar{g}_{\sigma^2,j}\sigma^2_{T_h-j} + \bar{g}_{\sigma^2_x,j}\sigma^2_{x,T_h-j} + \bar{g}_{\sigma^2_d,j}\sigma^2_{d,T_h-j} + \bar{g}_{\varepsilon,j}\sigma_{T_h-j-1}\varepsilon_{T_h-j} \\ &+ \bar{g}_{\varepsilon_x,j}\sigma_{x,T_h-j-1}\varepsilon_{x,T_h-j} + \bar{g}_{\varepsilon_d,j}\sigma_{d,T_h-j-1}\varepsilon_{d,T_h-j} + \bar{g}_{\eta,j}\eta_{T_h-j} + \bar{g}_{\eta_x,j}\eta_{x,T_h-j} + \bar{g}_{\eta_d,j}\eta_{d,T_h-j}\Big) \end{aligned} \tag{14}$$

$$\begin{aligned} q_{T_h}(s) = \bar{q}_0 + \sum_{j=0}^{s} \Big(&\bar{q}_{x,j}x_{T_h-j} + \bar{q}_{\sigma^2,j}\sigma^2_{T_h-j} + \bar{q}_{\sigma^2_x,j}\sigma^2_{x,T_h-j} + \bar{q}_{\sigma^2_d,j}\sigma^2_{d,T_h-j} + \bar{q}_{\varepsilon,j}\sigma_{T_h-j-1}\varepsilon_{T_h-j} \\ &+ \bar{q}_{\varepsilon_x,j}\sigma_{x,T_h-j-1}\varepsilon_{x,T_h-j} + \bar{q}_{\varepsilon_d,j}\sigma_{d,T_h-j-1}\varepsilon_{d,T_h-j} + \bar{q}_{\eta,j}\eta_{T_h-j} + \bar{q}_{\eta_x,j}\eta_{x,T_h-j} + \bar{q}_{\eta_d,j}\eta_{d,T_h-j}\Big) \end{aligned} \tag{15}$$

Observe that $(q_{T_h}(s), g_{T_h}(s))$ is stationary as long as $s < \infty$. The policy payoff normalized by fundamental market capitalization is therefore stationary as well.

2.2.1 Policy Examples

The specification in (14) and (15) is very general and allows policy to respond to any observable variable in the economy. Table 1 gives four examples of strike prices and the corresponding $\bar{g}$ parameters that implement the policy.

In the first example (second column), strike price is simply proportional to the fundamental market value s periods before maturity. Assuming $q_{T_h}(s) \equiv q$, the policy corresponds to the government every

Table 1: Columns 2-5 are examples of possible strike price specifications contained as special cases of the more general policy in equation (14) and the corresponding parameter values for the $\bar{g}$-coefficients are given in the rows.

	Strike Price K_T			
g	$e^{-g}P_{T-s}$	$e^{-g}\prod_{j=1}^{s} P_{T-j}^{1/s}$	$e^{-g}e^{-z_d+d_{T-s}}$	$e^{-g}e^{-z_{d,T}+d_{T-s}}$
$\bar{g}_0$	$g+(s-1)(\mu_d+\varrho\mu_c)$	$g+\frac{s-1}{2}(\mu_d+\varrho\mu_c)$	$g+(s-1)(\mu_d+\varrho\mu_c)+z_d$	$g+(s-1)(\mu_d+\varrho\mu_c)+A$
$\bar{g}_{x,0}$	0	0	0	A_x
$\bar{g}_{x,1}$	$\varrho - A_x$	$-\frac{s-1}{s}(A_x-\varrho)$	$\varrho - A_x$	$\varrho - A_x$
$\bar{g}_{x,j}$	ϱ	$\frac{A_x}{s}+\frac{s-j}{s}\varrho$	ϱ	ϱ
$\bar{g}_{x,s}$	A_x	$\frac{A_x}{s}$	0	0
$\bar{g}_{\sigma^2,0}$	0	0	0	A_{σ^2}
$\bar{g}_{\sigma^2,1}$	$-A_{\sigma^2}$	$-\frac{s-1}{s}A_{\sigma^2}$	$-A_{\sigma^2}$	$-A_{\sigma^2}$
$\bar{g}_{\sigma^2,j}$	$-\frac{\varrho\tau}{2}$	$\frac{A_{\sigma^2}}{s}-\frac{s+1-j}{s}\frac{\varrho\tau}{2}$	$-\frac{\varrho\tau}{2}$	$-\frac{\varrho\tau}{2}$
$\bar{g}_{\sigma^2,s}$	$A_{\sigma^2}-\frac{\varrho\tau}{2}$	$\frac{1}{s}(A_{\sigma^2}-\frac{\varrho\tau}{2})$	$-\frac{\varrho\tau}{2}$	$-\frac{\varrho\tau}{2}$
$\bar{g}_{\sigma_x^2,0}$	0	0	0	$A_{\sigma_x^2}$
$\bar{g}_{\sigma_x^2,1}$	$-A_{\sigma_x^2}$	$-\frac{s-1}{s}A_{\sigma_x^2}$	$-A_{\sigma_x^2}$	$-A_{\sigma_x^2}$
$\bar{g}_{\sigma_x^2,j}$	$-\frac{\varrho\tau_x}{2}$	$\frac{A_{\sigma_x^2}}{s}-\frac{s+1-j}{s}\frac{\varrho\tau_x}{2}$	$-\frac{\varrho\tau_x}{2}$	$-\frac{\varrho\tau_x}{2}$
$\bar{g}_{\sigma_x^2,s}$	$A_{\sigma_x^2}-\frac{\varrho\tau_x}{2}$	$\frac{1}{s}(A_{\sigma_x^2}-\frac{\varrho\tau_x}{2})$	$-\frac{\varrho\tau_x}{2}$	$-\frac{\varrho\tau_x}{2}$
$\bar{g}_{\sigma_d^2,0}$	0	0	0	$A_{\sigma_d^2}$
$\bar{g}_{\sigma_d^2,1}$	$-A_{\sigma_d^2}$	$-\frac{s-1}{s}A_{\sigma_d^2}$	$-A_{\sigma_d^2}$	$-A_{\sigma_d^2}$
$\bar{g}_{\sigma_d^2,j}$	$-\frac{\varrho\tau_d}{2}$	$\frac{A_{\sigma_d^2}}{s}-\frac{s+1-j}{s}\frac{\tau_d}{2}$	$-\frac{\varrho\tau_d}{2}$	$-\frac{\varrho\tau_d}{2}$
$\bar{g}_{\sigma_d^2,s}$	$A_{\sigma_d^2}-\frac{\varrho\tau_d}{2}$	$\frac{1}{s}(A_{\sigma_d^2}-\frac{\tau_d}{2})$	$-\frac{\varrho\tau_d}{2}$	$-\frac{\varrho\tau_d}{2}$
$\bar{g}_{\varepsilon,0}$	0	0	0	0
$\bar{g}_{\varepsilon,1}$	ϱ	$\frac{s-1}{s}\varrho$	ϱ	ϱ
$\bar{g}_{\varepsilon,j}$	ϱ	$\frac{s-j}{s}\varrho$	ϱ	ϱ
$\bar{g}_{\varepsilon,s}$	0	0	0	0
$\bar{g}_{\varepsilon_x,0}$	0	0	0	0
$\bar{g}_{\varepsilon_x,1}$	0	0	0	0
$\bar{g}_{\varepsilon_x,j}$	0	0	0	0
$\bar{g}_{\varepsilon_x,s}$	0	0	0	0
$\bar{g}_{\varepsilon_d,0}$	0	0	0	0
$\bar{g}_{\varepsilon_d,1}$	1	$\frac{s-1}{s}$	1	1
$\bar{g}_{\varepsilon_d,j}$	1	$\frac{s-j}{s}$	1	1
$\bar{g}_{\varepsilon_d,s}$	0	0	0	0
$\bar{g}_{\eta,j}$	0	0	0	0
$\bar{g}_{\eta_x,j}$	0	0	0	0
$\bar{g}_{\eta_d,j}$	0	0	0	0

period issuing a standard European put-option with a time-to-maturity s. In other words, the government promises the shareholders that if $P_{t+s}+D_{t+s}$ falls sufficiently below the current fundamental market value, the shareholders will be (partly) compensated for the loss.

The second example (the third column in Table 1) offers a smoother policy by setting the strike price proportional to the geometric average fundamental market capitalization between issuance at $T_h - s$ and $T_h - 1$. In this case, the strike price is forward-looking at issuance and falls with the fundamental market. The policy therefore insures more against large, sudden drops in the fundamental price than small, repeated drops that add up to a fall of similar magnitude.

The policies in the fourth and fifth columns have a slightly different flavor. In the former, the strike price on the option expiring at $t + s$ is set proportional to $V_t = e^{-z_d + d_t}$, where z_d is the mean log dividend-price ratio. Note that V_t would be the fundamental market value if the dividend-price ratio at time t was z_d. We can thus think of a policy maker who calculates a pseudo price assuming discount rates and expected dividend-growth is constant and then promises to compensate investors if the actual market price and dividend at $t + s$ is sufficiently below this pseudo price. One rationale for such a policy would be to not insure investors too much at times when discount rates are low (low discount rates imply low $z_{d,t}$ and high P_t).

The final column is similar in spirit. At maturity the policymaker calculates the pseudo price $V_{t+s} = e^{-z_{d,t+s} + d_t}$, which is the fundamental market price at time t had the dividend-price ratio been $z_{d,t} = z_{d,t+s}$. The strike price at maturity will therefore be high if discount rates at maturity are low and vice versa. The policy therefore compensates falls in the stock market due to cash-flow shocks, whereas shocks to discount rates receive less compensation.

Clearly, Table 1 only gives a flavor of the types of policies that can be entertained within my framework. For example, we can design policies to compensate cash-flow or discount rate shocks. Or we could make it explicitly dependent on realized or expected economic growth etc. Furthermore, all the policies in Table 1 can be conditionally scaled based on observable variables in the economy through the appropriate choice of $\bar{q}$. E.g. if the policy aims to mitigate effects arising from uncertainty, $q_t(s)$ could be dependent on $(\sigma_t^2,\ \sigma_{x,t}^2,\ \sigma_{d,t}^2)$ or on $\mathbb{V}_t(r_{d,t+1})$. In summary, the very general nature of my policy specification allows me to capture a lot of the complexity and heterogeneity observed in actual monetary and fiscal responses.

2.2.2 Policy Pricing

The time t value of the insurance payoff occurring at time T_h is

$$P_{h,t}^X \equiv \mathbb{E}_t(M_{t,T_h} X_{T_h}) = P_t \tilde{P}_{h,t}^X \tag{16}$$

$$P_{0,t}^X \equiv X_t \tag{17}$$

In order to get a partial analytical price for the policy claims, the following notation proves useful. Let $\mathcal{F}_t$ denote the full information set at time t generated by $\{\varepsilon_j, \varepsilon_{x,j}, \varepsilon_{d,j}, \eta_j, \eta_{x,j}, \eta_{d,j}\}_{j=-\infty}^t$ and $\mathcal{H}_t$ the partial information set generated by $\{\eta_j, \eta_{x,j}, \eta_{d,j}\}_{j=-\infty}^t$. Note the latter is a coarser (smaller) information set containing only information generated by the shocks to variance. I will use the following notation:

$\mathbb{E}_t(X) \equiv \mathbb{E}(X|\mathcal{F}_t)$ and $\mathbb{E}_t(X|\mathcal{H}_{t+h}) \equiv \mathbb{E}(X|\mathcal{F}_t, \mathcal{H}_{t+h})$, i.e. an expectation with a time subscript refers to the conditional expectation w.r.t. the full information set and use the explicit conditioning when taking conditional expectations w.r.t. the full information set augmented by the realization of (future) variance shocks. Furthermore, for any random variable X let

$$\hat{X}_{T|T-j} \equiv \mathbb{E}_{T-j}(X_T|\mathcal{H}_T)$$
$$\tilde{X}_{T|T-j} \equiv X_T - \hat{X}_{T|T-j}$$

The payoff at time T_h in equation (13) can be written as

$$X_{T_h} = P_{T_h-1}e^{q_{T_h}(s)-g_{T_h}(s)}(1-e^{r_{d,T_h}+g_{T_h}(s)})^+ \tag{18}$$

Note that whether the claim pays off or not only depend on $r_{T_h} + g_{T_h}(s)$. It is useful to represent $r_{d,T_h} + g_{T_h}(s)$ as follows

$$r_{d,T_h} + g_{T_h}(s) = \hat{r}_{d,T_h|t} + \hat{g}_{T_h|t}(s) + \sum_{j=0}^{h-1} K_{\varepsilon,j}\sigma_{T_h-1-j}\varepsilon_{T_h-j} + K_{\varepsilon_x,j}\sigma_{x,T_h-1-j}\varepsilon_{x,T-j} + K_{\varepsilon_d,j}\sigma_{d,T_h-1-j}\varepsilon_{d,T_h-j}$$

where $K_{\cdot,j}$ are constants. Let

$$\begin{aligned} y_{t,T_h} &\equiv log(M_{t,T_h}P_{T_h-1}) - log(P_t) + q_{T_h}(s) - g_{T_h}(s) \\ &= \hat{mp}_{T_h|t} + \hat{q}_{T_h|t}(s) - \hat{g}_{T_h|t}(s) + \sum_{j=0}^{h-1} L_{\varepsilon,j}\sigma_{T_h-1-j}\varepsilon_{T_h-j} + L_{\varepsilon_x,j}\sigma_{x,T_h-1-j}\varepsilon_{x,T_h-j} + L_{\varepsilon_d,j}\sigma_{d,T_h-1-j}\varepsilon_{d,T_h-j} \end{aligned}$$

where $L_{\cdot,j}$ are constants. See Appendix E for derivations and definitions. $\hat{mp}_{T_h|t} \equiv \mathbb{E}_t(log(M_{t,T_h}P_{T_h-1})|\mathcal{H}_{T_h}) - log(P_t)$ is the expected value of $log(M_{t,T_h}P_{T_h-1}) - log(P_t)$ conditional on time t information augmented by the variance shocks up to time T_h. Thus, $\hat{mp}_{T_h|t}$ does not depend on the future cash-flow shocks $\varepsilon_{\cdot,t+j}$ for $j > 0$. Define the following conditional variances and covariance

$$\zeta_{y,t,T_h-1} \equiv \mathbb{V}_t\big(y_{t,T_h}\big|\mathcal{H}_{T_h}\big) = \sum_{j=0}^{h-1} L^2_{\varepsilon,j}\sigma^2_{T_h-1-j} + L^2_{\varepsilon_x,j}\sigma^2_{x,T_h-1-j} + L^2_{\varepsilon_d,j}\sigma^2_{d,T_h-1-j}$$

$$\zeta_{r_d+g,t,T_h-1} \equiv \mathbb{V}_t\big(r_{d,T_h} + g_{T_h}(s)\big|\mathcal{H}_{T_h}\big) = \sum_{j=0}^{h-1} K^2_{\varepsilon,j}\sigma^2_{T_h-1-j} + K^2_{\varepsilon_x,j}\sigma^2_{x,T_h-1-j} + K^2_{\varepsilon_d,j}\sigma^2_{d,T_h-1-j}$$

$$\begin{aligned} \zeta_{y,r_d+g,t,T_h-1} &\equiv Cov_t\big(y_{t,T_h}, r_{d,T_h} + g_{T_h}(s)\big|\mathcal{H}_{T_h}\big) \\ &= \sum_{j=0}^{h-1} L_{\varepsilon,j}K_{\varepsilon,j}\sigma^2_{T_h-1-j} + L_{\varepsilon_x,j}K_{\varepsilon_x,j}\sigma^2_{x,T_h-1-j} + L_{\varepsilon_d,j}K_{\varepsilon_d,j}\sigma^2_{d,T_h-1-j} \end{aligned}$$

The following proposition gives a partial analytical solution to the prices of policy claims

Proposition 3. *Suppose $s < \infty$ and $K_{\cdot,j} \neq 0$ for at least some ε, ε_x, ε_d and some $j = 0, 1, \ldots, h-1$.*

Then the value of the claim that expires h periods in the future is

$$P^X_{h,t} = P_t\mathbb{E}_t\Bigg[e^{\hat{mp}_{T_h|t}+\hat{q}_{T_h|t}-\hat{g}_{T_h|t}(s)+\frac{\zeta_{y,t,T_h-1}}{2}}\Bigg(\Phi\Big(-\frac{\hat{g}_{T_h|t}(s)+\hat{r}_{d,T_h|t}+\zeta_{y,r_d+g,t,T_h-1}}{\sqrt{\zeta_{r_d+g,t,T_h-1}}}\Big)$$
$$-e^{\hat{g}_{T_h|t}(s)+\hat{r}_{d,T_h|t}+\zeta_{y,r_d+g,t,T_h-1}+\frac{\zeta_{r_d+g,t,T_h-1}}{2}}\Phi\Big(-\frac{\hat{g}_{T_h|t}(s)+\hat{r}_{d,T_h|t}+\zeta_{r_d+g,t,T_h-1}}{\sqrt{\zeta_{r_d+g,t,T_h-1}}}-\sqrt{\zeta_{r_d+g,t,T_h-1}}\Big)\Bigg)\Bigg]$$

where $\Phi(.)$ *denotes the standard normal CDF.*

Proof. See Appendix E. ■

Proposition 3 gives us a more efficient and precise way to calculate the claims prices. The condition that at least one $K_{\cdot,j} \neq 0$ requires that whether the claim pays off depends on at least one cash-flow shock.

Corollary 2 (to Proposition 3)**.** *Suppose* $s < \infty$ *and* $K_{\cdot,j} = 0$ *for all* $\varepsilon,\ \varepsilon_x,\ \varepsilon_d$ *and all* $j = 0,\ 1, \ldots, h-1$. *Then the value of the claim that expires h periods in the future is*

$$P^X_{h,t} = P_t\mathbb{E}_t\Bigg[e^{\hat{mp}_{T_h|t}+\hat{q}_{T_h|t}-\hat{g}_{T_h|t}(s)+\frac{\zeta_{y,t,T_h-1}}{2}}\Bigg(1-e^{\hat{g}_{T_h|t}(s)+\hat{r}_{d,T_h|t}}\Bigg)\mathbb{I}_{\hat{g}_{T_h|t}(s)+\hat{r}_{d,T_h|t}\leq 0}\Bigg]$$

where $\mathbb{I}_{\hat{g}_{T_h|t}+\hat{r}_{d,T_h|t}\leq 0}$ *is an indicator function that takes the value* 1 *if* $\hat{g}_{T_h|t}+\hat{r}_{d,T_h|t} \leq 0$ *and* 0 *otherwise.*

Proof. See Appendix E. ■

The corollary provides the partial analytical price for the claim in the case that cash-flow shocks do not influence whether or not the claim pays off. Note that the actual cash-flow at maturity can still depend on cash-flow shocks through $P_{T_h-1}e^{q_{T_h}(s)-g_{T_h}(s)}$. Corollary 2 then gives a way to price such claims without having to simulate the cash-flow shocks.

Part 3 of Theorem 1 shows an interesting and perhaps paradoxical effect of the policy. Although the policy insures the market, implicit claims today corresponding to the market's expectation that further insurance will be performed in the future may *increase* the risk of the market today and thus command a large positive risk premium

Theorem 1. *Assume* $s < \infty$ *and that* $P^X_{h,t}$ *is given by Proposition 3. Then*

1. *For* $h > s$

$$Cov_{t-1}\Big(\sigma_{t-1}\varepsilon_t, log(P^X_{h,t})\Big) = Cov_{t-1}\Big(\sigma_{t-1}\varepsilon_t, log(P_t)\Big) > 0$$

2. *If* $\varrho\psi > 1$, *there exists a number* k *s.t.* $\forall h > k$

$$Cov_{t-1}\Big(\sigma_{x,t-1}\varepsilon_{x,t}, log(P^X_{h,t})\Big) > 0$$

3. If $\varrho\psi > 1$, the limiting covariance

$$\lim_{h\to\infty} Cov_{t-1}\Big(\sigma_{x,t-1}\varepsilon_{x,t}, log(P^X_{h,t})\Big) = \frac{\varrho - \frac{\rho_x}{\psi}}{1-\rho_x}\sigma^2_{x,t-1} > Cov_{t-1}\Big(\sigma_{x,t-1}\varepsilon_{x,t}, log(P_t)\Big)$$

Proof. See Appendix G ∎

Theorem 1 tells us that the policy claims maturing far enough in the future have positive risk premia. For claims expiring $h > s$ periods in the future, the exposure to short-run consumption shocks is the same as that of the fundamental market. To see why this makes sense, consider the payoff at maturity given by (13). Since dividends are random walks (with stochastic drift) the fundamental market price at $T_h - 1$ moves by the same relative amount as P_t in response to a shock ε_t. The requirement $h > s$ implies that the policy cannot respond to ε_t, thus the payoff exposure to ε_t is the same as that of P_{T_h-1}. The last two statements are less straightforward because the policy $(q_{T_h}(s),\ g_{T_h}(s))$ can depend directly on x_{T_h-j} for $j = 0, 1, \ldots, s$ and fundamental market return r_{d,T_h} depend on x_{T_h}. Through the dependence on x_{T_h-j}, they also depend on $\varepsilon_{x,t}$ indirectly. The intuition for the result is that the indirect dependence on $\varepsilon_{x,t}$ decreases exponentially at rate ρ_x^{h-s} with time to maturity, whereas the direct dependence coming from current dividends never die out. Thus, at some point the latter must dominate. These results indicate that the claims maturing far enough into the future, will carry positive risk premiums. Furthermore, the last part of the theorem states that not only will the exposure of $P^X_{h,t}$ to $\varepsilon_{x,t}$ be positive for a sufficiently long time-to-maturity, it can even be greater than the fundamental market exposure to $\varepsilon_{x,t}$. As a consequence, the risk premiums on those claims can actually be *higher* than the risk premium on the fundamental market.

The total market value and dividends at time t can be written

$$P^*_t = P_t + \sum_{h=1}^{\infty} P^X_{h,t} = P_t\Big(1 + \sum_{h=1}^{\infty} \tilde{P}^X_{h,t}\Big)$$
$$D^*_t = D_t + X_t$$

and the total return from t to $t+1$ is

$$R^*_{t+1} = \omega_{0,t}R_{d,t+1} + \sum_{h=1}^{\infty}\omega_{h,t}R^X_{h,t+1}$$

where $\omega_{0,t} \equiv \frac{P_t}{P^*_t}$, $\omega_{h,t} \equiv \frac{P^X_{h,t}}{P^*_t}$ and $R^X_{h,t+1} \equiv \frac{P^X_{h-1,t+1}}{P^X_{h,t}}$.

2.3 Intuition

Theorem 1 part 1 gives us a useful way to classify the claims. The claims maturing $i \le s$ periods from now can be thought of as options that are already issued. Correspondingly, claims that expire $i > s$ periods from now would then be a claim to receive the option $i-s$ periods from now. The market return can then be divided into three parts: (1) the fundamental return R_{t+1}, (2) the return on the portfolio of options already issued $R^X_{I,t+1} \equiv \frac{1}{\sum_{i=1}^{s}\omega_{i,t}}\sum_{i=1}^{s}\omega_{i,t}R^X_{i,t+1}$, and (3) the return on the portfolio of claims to future options $R^X_{U,t+1} \equiv \frac{1}{\sum_{i=s+1}^{\infty}\omega_{i,t}}\sum_{i=s+1}^{\infty}\omega_{i,t}R^X_{i,t+1}$. As shown in section 2.1, the fundamental return earns a positive risk premium unconditionally, and the conditional risk premium is increasing in the fundamental return variance.

For policies that compensates investors in bad times, the return on issued options behaves like that of a standard put option. Firstly, issued put options earn negative risk premiums unconditionally. The reason is that the probability of these options paying off, and the corresponding payoff, both increases when the fundamental market falls. In other words, the expected cash-flow to these options increase when fundamentals worsen. For a given discount rate, the value of the options must therefore increase.[5] Since issued options tend to be negatively correlated with consumption, they earn a negative risk premium. Secondly, the conditional risk premium on issued options should become more negative when the systematic risk rises and less negative when the unsystematic risk is high. The explanation is reminiscent of the finding in Bruno and Haug (2018) that higher systematic (idiosyncratic) risk increases (decreases) the positive correlation with the discount factor.

The claims to future options are less straightforward, but Theorem 1 gives us a general idea. All of these claims are positively correlated with the ε_t consumption shock, which requires compensation. Furthermore, the correlation with $\varepsilon_{x,t}$ also tends to be positive for these claims (possibly with the exception of when ρ_x is very close to 1). We therefore expect the portfolio of future options will earn a positive risk premium unconditionally. This is less puzzling than one might expect for two reasons. First, as long as the stimulus is in some way "proportional" to the market capitalization of the fundamental market over time, the payoffs to claims expiring far into the future must be positively related to dividends, and thereby consumption. These claims therefore carry systematic risk. If this was not the case, the policy payoffs would either end up accounting for the vast majority of cash-flows to shareholders, or an insignificant fraction. Second, if discount rates were negative for all claims, their price would increase with maturity (unless the market capitalization is expected to persistently fall). In particular, as time to maturity goes towards infinity, the corresponding price would too. Consequently, it is very natural to expect discount rates to be positive for long maturities.

Theorem 1 also gives us an idea about what to expect conditionally. Clearly, an increase in σ_t^2 causes next-period risk premium on future options to be higher. Similarly, the second and third part of the theorem indicates that an increase in $\sigma_{x,t}^2$ should cause an increased risk premium as well. Perhaps somewhat paradoxically, the claims to future insurances in some important aspects tend to behave more

[5] Since the options are non-linear in the underlying, assuming discount rates remain the same when the underlying falls is clearly not correct as the "leverage" of the options changes. However, the effect resulting from changing the discount rate should not be large enough to offset the cash-flow effect.

like the fundamental market than an insurance. As a consequence, introducing the policy might lower unconditional risk premium on the market by less than might otherwise be expected.

For the return on the market, R^*, we also have to take into account the portfolio composition effect. Following the above discussion, we can write

$$R^*_{t+1} = \omega_{0,t} R_{d,t+1} + \omega_{I,t} R^X_{I,t+1} + \omega_{U,t} R^X_{U,t+1} \tag{19}$$

$\omega_{0,t}$, $\omega_{I,t}$, $\omega_{U,t}$ are the time t portfolio weights of the fundamental market, the options that are already issued, and future options respectively. R_{t+1}, $R^X_{I,t+1}$, $R^X_{U,t+1}$ are the corresponding returns. When variance is high, the fundamental price tends to be low due to higher discount rates, whereas the value of the insurance is high. We therefore expect $\omega_{0,t}$ to be negatively related to current variance and $\omega_{I,t}$ to be positively related to current variance. The effect of increased variance on $\omega_{U,t}$ is less clear due to two possible opposing effects: (1) higher systematic variance tend to increase the discount rate causing the prices of claims to future options to fall, (2) to the extent that variance is persistent, high current variance predicts variance at issuance to be high as well, thereby raising expected cash-flow and thus current prices. My numerical results indicate that the net effect is that $\omega_{U,t}$ does not change by a lot. The total return is therefore more similar to the issued put return when the risk is high, leading to a weaker, possibly negative, risk-return trade-off.

3 Numerical Results

In this section I report numerical results for the baseline parameters. Under the policy, the government writes 0.05 put options every period with time-to-maturity of 12 months. The strike price is $e^{-0.2} \approx 82\%$ of the geometric average fundamental market capitalization between issuance at time $T_h - 12$ and a month before maturity $T_h - 1$. In other words, at issuance the options have a forward-looking strike price that falls with the fundamental market price. However, the strike price falls less than one-for-one as the fundamental market falls, thus offering an insurance to investors.

3.1 The Risk-Return Relationship

Figure 1 plots annualized mean returns against return variance on the market portfolio. We see that the expected return on the market, the blue line, does not vary much with variance. In fact, it is around 8% both when the variance is in the bottom and top quintile. This is in stark contrast to the fundamental market, the red dashed line, whose expected return increase from about 8.2% when the variance is in the bottom quintile, to about 12% for variance in the top quintile. Mean risk free rate, plotted as purple crosses, remains around 3% for all quintiles of variance. Thus, the policy virtually eliminates the

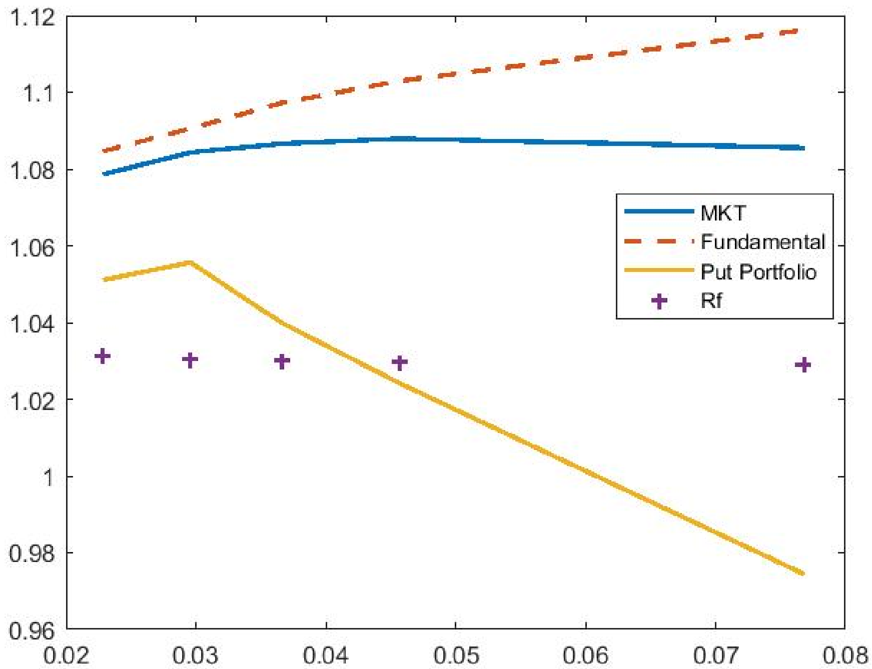


Figure 1: This figure plots average annualized expected returns and risk free rate (y-axis) against annualized market return variance (x-axis). The dashed red line is the fundamental expected return, the blue line is the expected return on the market and the yellow line is the expected return on the portfolio of put options. The +'s represents average risk-free rate.

fundamental relationship between the market's risk and risk premium. In contrast, the risk premium on the fundamental market increases by about 80 percent when return risk goes from the bottom to the top quintile.

It is also interesting to note that despite the presence of the insurance, market variance exhibits substantial spread, from an average annualized variance of 0.02 in the bottom quintile to about 0.08 in the top quintile. For comparison, the fundamental market variance is about 0.02 and 0.09 in the bottom and top quintiles respectively. In other words, the policy significantly weakens the risk-return relationship, without causing a large reduction in variance spreads.

To understand the weakened relationship between risk and return, note that the market return is a weighted average of the fundamental return and policy return. There are therefore two channels that affects how the market risk premium is related to variance. First, for given weights, the market risk premium follow the pattern of its constituent parts, the "direct" channel. Second, the weights used to average over the constituent parts might change with variance. To the extent that expected returns on the fundamental market and the policy differ, we get an additional "composition effect".

From Figure 1, we observe that there is a negative relationship between the expected return on the policy, the yellow line, and the market variance. In the low variance quintile, the expected return on the portfolio of policy claims is about 5%, while it is about -2% in the high variance quintile. Interestingly, the policy carries a *positive* risk premium, about 2%, on average when market risk is low.

The government policy consists of a collection of put options that are already issued and an implicit promise of the issuance of additional put options in the future. In Panel (a) of Figure 2, I plot the average prices of a full unit of all the policy claims conditioning on current market variance. Note that the actual policy consists of 0.05 units of each claim. Thus, to figure out how important each claim is to the overall market capitalization, we must multiply each price by 0.05.

Each point along the x-axis represents a separate claim expiring x periods in the future. The claims expiring less than 12 months in the future (left of the vertical line) are the put options that have already been issued. The claims to the right of the vertical line are the implicit claims to future options. Furthermore, each curve represents a given quintile of market return variance. To make the interpretation easy (and make the prices stationary), the prices have been normalized by the current fundamental market value.

Comparing the curves in Panel (a) of Figure 2, we observe the well-known fact that option values tend to increase with variance. Since all the prices are normalized by fundamental market capitalization, the increase in policy prices implies that the composition of the market is tilted more towards the policy when variance increases. As the expected return on the policy is lower than on the fundamental market for all variance brackets, the "composition" effect further weakens the relationship between risk and return on the market.

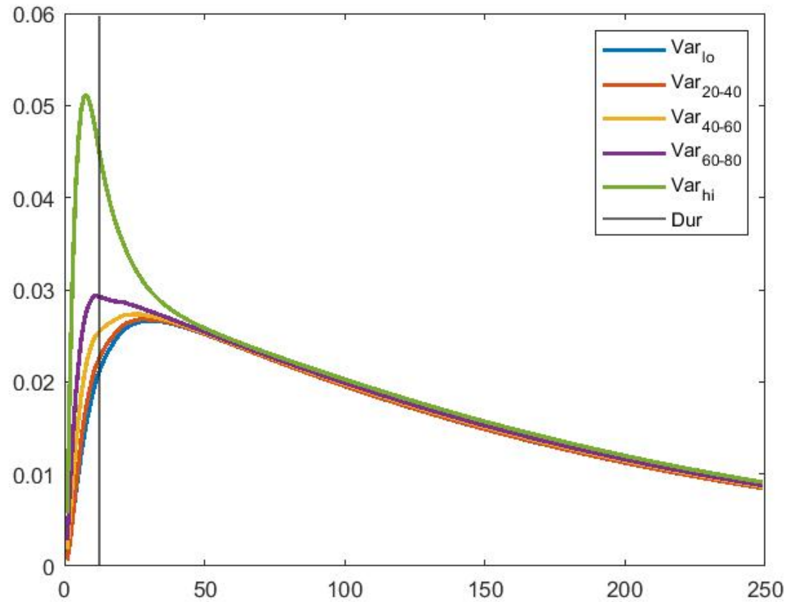
In summary, the weak risk-return trade-off on the market is explained both by the policy return being negatively related to variance (direct effect) and that the weight of the policy is increasing in variance (composition effect).

3.2 Policy Composition and Risk Premium

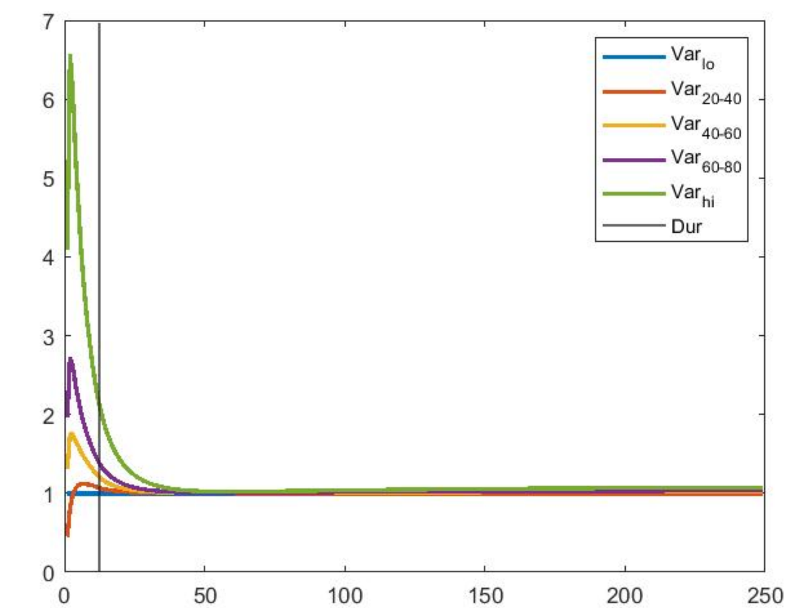
From Panel (a) in Figure 2 we also get the impression that claims have different sensitivities to market variance. In Panel (b), I have therefore plotted the ratio of policy prices from Panel (a) in market variance quintiles 2, 3, 4, and 5 to the corresponding prices in the low quintile. The y-axis therefore represents how many times more valuable a claim is in a given variance quintile compared to what it would be worth if market variance was in the bottom quintile.

We observe that the issued options, left of the vertical line, generally become much more valuable with increasing variance. For instance, when market variance is in the second highest quintile, the issued put options are typically more than twice as valuable as they would have been if market variance was low. Looking at the highest quintile, the ratio is even more extreme - several of the issued options are more than five times as valuable.

The prices of claims to future options, right of the vertical line, are much less responsive to current



(a)



(b)

Figure 2: Panel (a) plots average prices of put options relative to fundamental market value. The average is conditioned on the market return variance, e.g. the blue line represents the average relative put price when the market return variance is in the bottom quintile. The x-axis is the time to maturity for the option. In panel (b), each line in plot (a) is divided by the relative put price when return variance is in the bottom quintile (the blue line).

variance. Thus, rising market risk is associated with short-term policy claims becoming more valuable, while long-term policy claims are unaffected. Increasing market risk therefore essentially causes the policy itself to become more "short-term".

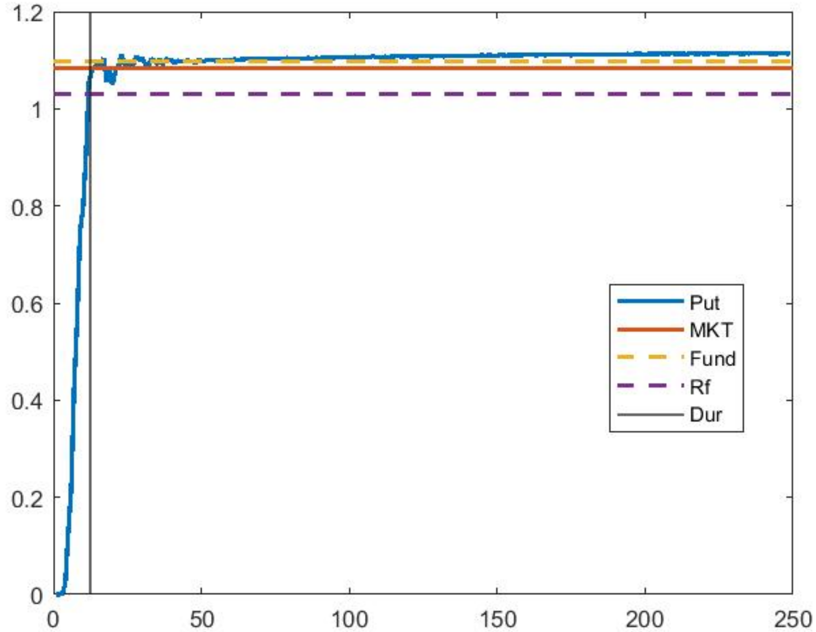
Figure 3 plots annualized expected returns for policy claims with different time-to-maturity (x-axis). The vertical line is located at the duration of the options (12 months). The claims to the left of the vertical line are therefore the options already issued. Similarly, the claims to the right are future options not yet issued. Panel (a) shows the unconditional mean returns, while Panel (b) condition on market variance quintiles.

From Panel (a), we see that the issued options carry large negative risk premiums. The risk-premiums on these options gradually increase with time-to-maturity. The reason for the gradual increase is that the policy uses the forward-looking geometric average fundamental market price as the strike price. This policy implies that the (log) strike on the option expiring in 12 months moves almost one-for-one with next period cash-flow shock, while the responsiveness of near-term strikes are much lower.

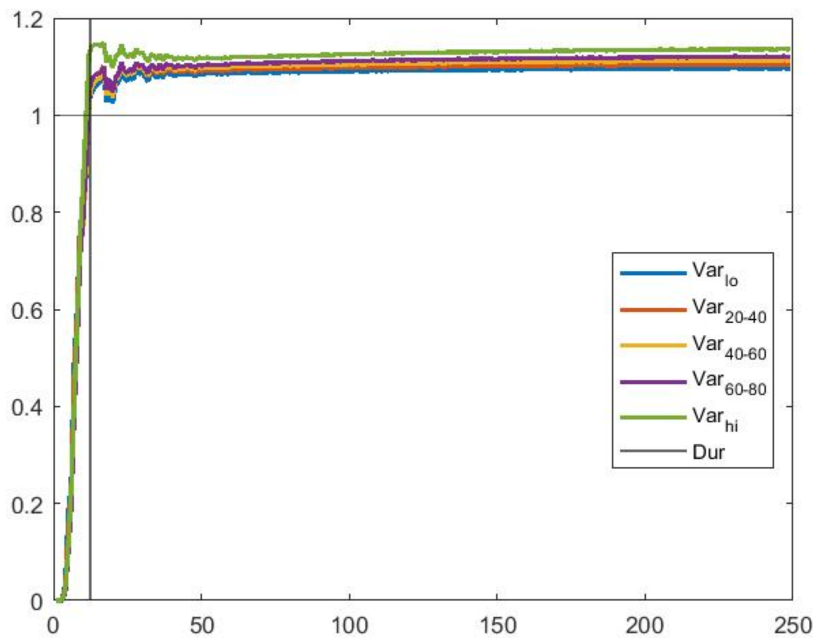
From Theorem 1, we expect the claims expiring in the distant future to have a larger risk premium than the fundamental market. Panel (a) confirms this prediction - claims to future options have a slightly higher unconditional risk premium than the fundamental market. As a consequence, the difference between the expected return on the insured and uninsured market is smaller unconditionally than what might be expected.

From panel (b), we see that the expected returns on claims expiring more than 12 months into the future are increasing in market variance. Although harder to see from Figure 3, my numerical results indicate the opposite holds for the already issued options - expected returns are typically *lower* when risk is high. This result is intuitive because the near term options act as an insurance as they increase in value when the fundamental market falls. The expected return on an insurance should generally be lower when risk increases.

The strike price on options that will be issued in the future will be determined by the future fundamental market level and thus reflect all that happens to that level between now and expiration. Since today's value of future put options increase with increasing strike price, this induces a risk exposure similar to that of the fundamental market. Thus, as the fundamental market demands a higher risk compensation, so does the claims to future options.



(a)



(b)

Figure 3: Panel (a) plots annualized expected returns on the put options (blue line) against the time-to-maturity. Also plotted is the insured market (red line) and the fundamental (uninsured) market (dashed yellow line) and the mean risk free rate (dashed purple line). The vertical line represents the option "duration". Panel (b) shows expected put option returns conditional on market return variance.

3.3 Sorting on Market Variance vs Fundamental Variance

Since the options are on the fundamental market, it is really the fundamental market return variance that affects the option values. Consider therefore the following decomposition of market return variance

$$\begin{aligned}\mathbb{V}_t(R^*_{t+1}) = \omega^2_{0,t}\mathbb{V}_t(R_{d,t+1}) + \omega^2_{I,t}\mathbb{V}_t(R^X_{I,t+1}) + \omega^2_{U,t}\mathbb{V}_t(R^X_{U,t+1}) + 2\omega_{0,t}\omega_{I,t}Cov_t(R_{d,t+1}, R^X_{I,t+1}) \\ + 2\omega_{0,t}\omega_{I,t}Cov_t(R_{d,t+1}, R^X_{U,t+1}) + 2\omega_{I,t}\omega_{U,t}Cov_t(R^X_{I,t+1}, R^X_{U,t+1}) \qquad (20)\end{aligned}$$

where $\omega_{0,t}$, $\omega_{I,t}$, and $\omega_{U,t}$ are the weights of the fundamental market, currently issued options, and claims to future options in the market portfolio and $R_{d,t+1}$, $R^X_{I,t+1}$ and $R^X_{U,t+1}$ are the corresponding (portfolio) returns.

Typically, the variance terms on the right-hand side is positively related to the variance on the fundamental market $\mathbb{V}_t(R_{d,t+1})$. Given the discussion in the previous section and Theorem 1, we expect the covariance between the fundamental market and future claims, $Cov_t(R_{d,t+1}, R^X_{U,t+1})$, to be positive and increasing in $\mathbb{V}_t(R_{d,t+1})$. However, since the issued options behave like insurances, $Cov_t(R_{d,t+1}, R^X_{I,t+1})$ and $Cov_t(R^X_{I,t+1}, R^X_{U,t+1})$ should both be negative, and their magnitudes also grow with $\mathbb{V}_t(R_{d,t+1})$. With partial insurance, and given portfolio weights ω_t, the net effect of increasing fundamental return variance on the market return variance $\mathbb{V}_t(R^*_{t+1})$ is still positive. As a consequence, $\mathbb{V}_t(R_{d,t+1})$ and $\mathbb{V}_t(R^*_{t+1})$ should be highly correlated [6].

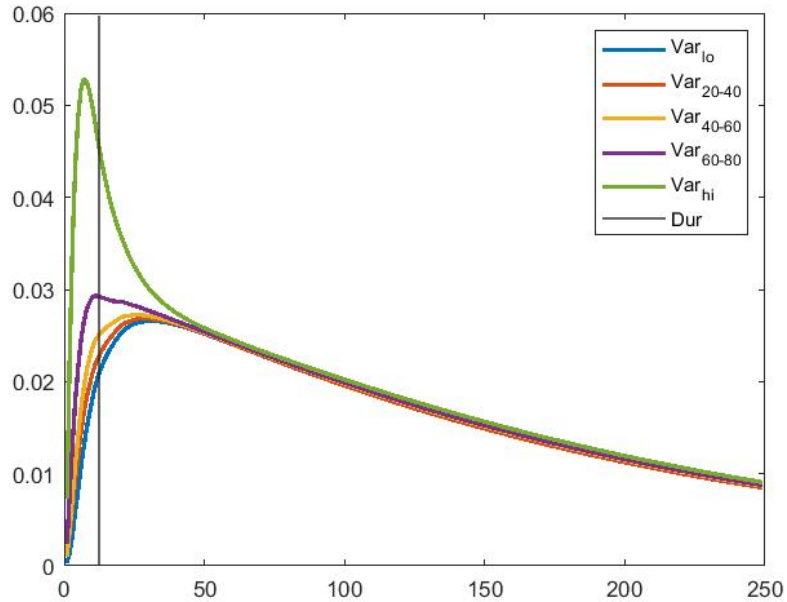
The conditional market return variance is also affected by the portfolio weights ω_t. Our discussion so far indicates that the fundamental market share $\omega_{0,t}$ falls with increasing fundamental return variance while the weight on currently issued options $\omega_{I,t}$ increases. The weight $\omega_{U,t}$ on future claims remains relatively constant.

Note that return variance is not the only determinant of these portfolio weights. Importantly, they are also history dependent. If the fundamental market fell significantly over the last period, $\omega_{I,t}$ is higher this period, i.e. the market is more insured. For a given fundamental return variance, higher $\omega_{I,t}$ lowers $\mathbb{V}_t(R^*_{t+1})$. Market variance is therefore high when fundamental variance is high and issued options are *less* valuable than expected given fundamental variance.

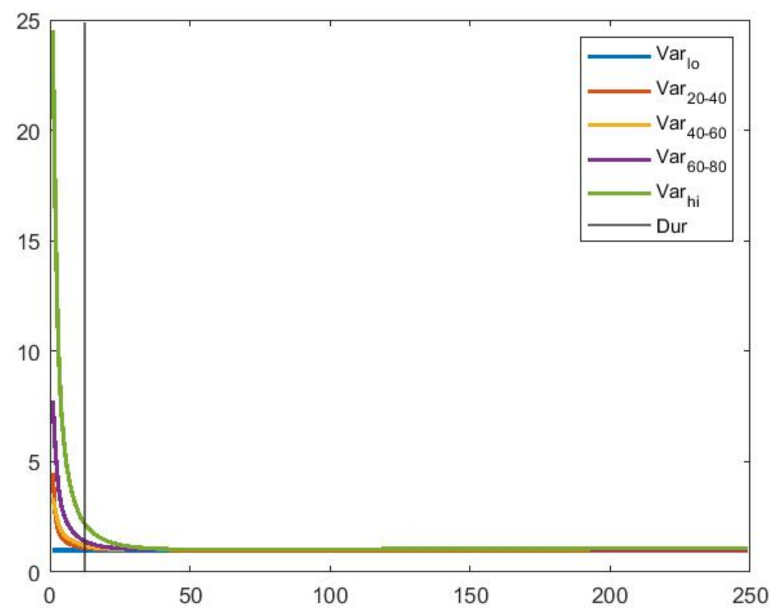
The link between fundamental return variance and market variance could be of particular interest in periods of economic turmoil. These periods are typically associated with a high degree of uncertainty and falling stock markets. In this case, the market variance would initially rise with the increase in uncertainty. However, as the fundamental market falls, the policy becomes more valuable, dampening the market return risk. Thus, the risk on the market might fall without a fall in fundamental uncertainty.

How large the difference between fundamental risk and market risk is, depends on how much the government insures. For instance, if the government increases the number of options it issues every

[6] In unreported numerical results i generally find the two variance series to be highly correlated.



(a)



(b)

Figure 4: Panel (a) plots average prices of put options relative to fundamental market value. The average is conditioned on the fundamental market return variance, e.g. the blue line represents the average relative put price when the market return variance is in the bottom quintile. The x-axis is the time to maturity for the claim. In panel (b), each line in plot (a) is divided by the average relative put price when fundamental return variance is in the bottom quintile (the blue line)

period, the difference in risk will be larger, and the history dependence in the portfolio weights becomes a more important factor.

Figure 4 sorts on fundamental market return variance, but are otherwise identical to Figure 4. We see the patterns are generally similar, but the magnitudes differ. In particular, we see that the ratios, Panel (b), in Figure 4 are more extreme than those in Figure 2. The contrast is particularly stark for the options expiring less than 4 months in the future. For instance, Panel (b) in Figure 2, these options are less valuable in the second variance quintile than in the first quintile. For comparison, in Figure 4 the same options are between 2 and 5 times as valuable in the second quintile. In other words, the fundamental market variance is more strongly related to the prices on currently issued options than the insured market variance is.

Intuitively, there are two (main) determinants of current market variance: (1) the variance of the fundamental market, and (2) the value of issued options. Increasing the fundamental risk tends to increase the market risk, while increasing the value of issued options has the opposite effect. As a result, the correlation between market variance and value of issued options will be *negative* when conditioning on the level of fundamental variance. It is therefore not surprising that issued options are generally more valuable in the high variance quintiles when sorted on the fundamental market variance $\mathbb{V}_t(R_{d,t+1})$ (as in Figure 4) compared to a sort on the market variance $\mathbb{V}_t(R^*_{t+1})$ (Figure 2).

The difference between sorts on fundamental risk and market risk also has implications for plots like Figure 1. If we had plotted expected returns against fundamental return variance, the risk-return trade-off for the market is typically even weaker, while that for the fundamental market is stronger. Furthermore, plotting expected returns against fundamental risk means the risk-return relationship monotonically weakens as we increase the number of options issued every period. However, when plotting against market variance, we can find a non-monotonic relationship. The reason is that with more insurance, sorts on market variance is increasingly a sort on (relatively) low prices for issued options.

3.4 Time-Series of Expected Returns

At first glance, one might expect the flat risk-return trade-off implies that there's very little variation over time in expected market returns. Figure 5 shows that this first glance deceives - the expected market return (blue line) varies significantly over time.

The expected market return is strongly positively correlated with the expected return on the fundamental market (red line), with a correlation coefficient of 0.65. Thus, the risk premium on the market does tend to rise when systematic risk increases, albeit by less than the fundamental market.

Somewhat puzzling, the risk-premium on the market are at times *higher* than that of the fundamental market. The explanation is that the risk-premiums on long-dated claims are higher than that of the fundamental market. Thus, when the current insurance is negligible, the effect of the policy is dominated

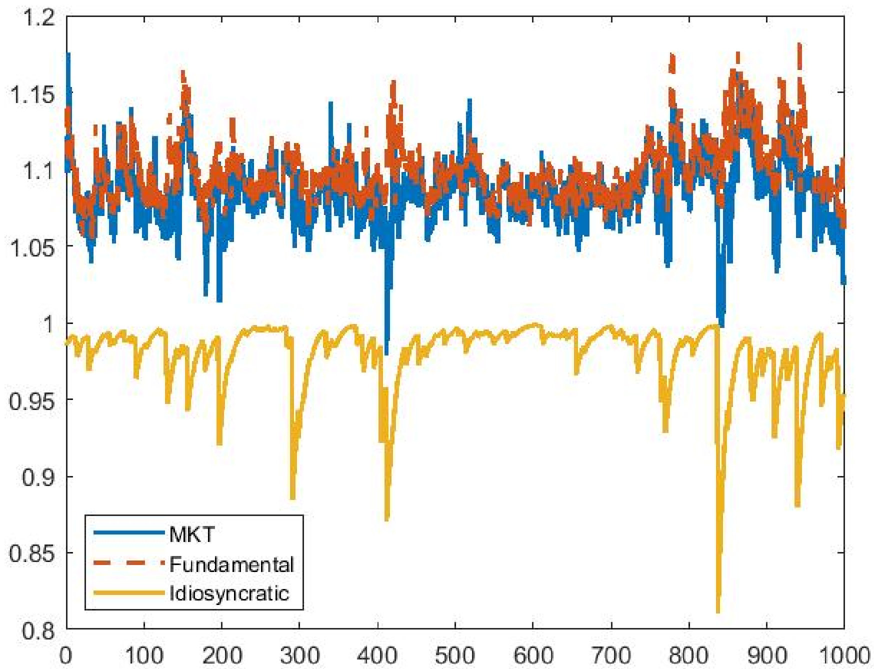


Figure 5: This figure plots a time-series of annualized conditional expected returns on the insured market (blue line) and the fundamental market (red line) calculated from simulations. The purple line is $1 - 12 * \sigma^2_{d,t}$.

by claims to future insurance that actually increases the market risk and risk premium. Under the current policy and parameters, this is not a frequent occurrence however.

We can also see that the expected market return is strongly correlated with the the inverse of idiosyncratic risk, $1 - 12\sigma^2_{d,t}$, represented by the yellow line. In other words, the market return is significantly negatively correlated with idiosyncratic risk, with a correlation coefficient of -0.37. The cause of the negative correlation is reminiscent of the finding in Bruno and Haug (2018). Higher idiosyncratic volatility causes issued options to become more valuable relative to the fundamental market. As a consequence, market returns are more correlated with option returns, and thereby less correlated (in absolute value) with the pricing kernel.

4 Conclusion

Counter-cyclical monetary and fiscal policy can plausibly explain the weak risk-return trade-off uncovered in the literature. Since such policies pays off when times are particularly bad, they offer a form of partial insurance to investors. If negative news are associated with a greater stimulus than the contraction following positive news, the policy is akin to a put option. Put options typically earn negative risk premiums and become more valuable when risk is high, thereby weakening, or possibly even reversing, the risk-return relationship.

If investors expect similar policies to be in place for the foreseeable future, they would rationally price in their expected future impact already today. Interestingly, investors price future policies very differently from current policies. In particular, future policies generally earn risk-premiums comparable, or even higher than, the market itself. The intuition is that in the long run, the stimulus offered by fiscal authorities or central banks must be positively related to the overall size of the economy. Otherwise, the stimulus either becomes irrelevant or the entire economy.

The positive risk premium associated with future policies means that the risk premium on the partly insured market can be close to that of the un-insured market unconditionally, yet behave very differently conditionally. For instance, the policy induces a negative correlation between expected return on the insured market and un-priced cash-flow risk, whereas the expected return on the un-insured market is unrelated to such risk.

A Appendix A - Recursive Utility SDF - Epstein and Zin (1991)

The value function is given by

$$V_t \equiv \max_{C_t,\omega_t} \Big\{ C_t^{1-\frac{1}{\psi}} + \beta \mathbb{E}_t (V_{t+1}^{1-\gamma})^{\frac{1-\frac{1}{\psi}}{1-\gamma}} \Big\}^{\frac{1}{1-\frac{1}{\psi}}} \tag{21}$$

where β is the discount factor of future utility, ψ is the inter-temporal rate of substitution and γ is the coefficient of relative risk aversion. Furthermore, the budget constraint is

$$W_{t+1} = (W_t - C_t)\omega_t^\top R_{t+1} \equiv (W_t - C_t) R_{\omega,t+1} \tag{22}$$

The derivative of the right-hand side of (1) w.r.t. consumption is

$$\begin{aligned} \frac{\partial}{\partial C_t} : \frac{V_t^{\frac{1}{\psi}}}{1-\frac{1}{\psi}} \Big\{ \big(1 - \frac{1}{\psi}\big) C_t^{-\frac{1}{\psi}} + \frac{1-\frac{1}{\psi}}{1-\gamma} \beta \mathbb{E}_t (V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} (1-\gamma) \mathbb{E}_t \Big(V_{t+1}^{-\gamma} \frac{\partial V_{t+1}}{\partial W_{t+1}} \frac{\partial W_{t+1}}{\partial C_t} \Big) \Big\} \\ = V_t^{\frac{1}{\psi}} \Big\{ C_t^{-\frac{1}{\psi}} + \beta \mathbb{E}_t (V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} \mathbb{E}_t \Big(V_{t+1}^{-\gamma} \frac{\partial V_{t+1}}{\partial W_{t+1}} \frac{\partial W_{t+1}}{\partial C_t} \Big) \Big\} \end{aligned} \tag{23}$$

Note that the budget constraint implies

$$\frac{\partial W_{t+1}}{\partial C_t} = -R_{\omega,t+1} \tag{24}$$

The first-order condition for consumption therefore implies

$$C_t^{-\frac{1}{\psi}} = \beta \mathbb{E}_t (V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} \mathbb{E}_t \Big(V_{t+1}^{-\gamma} \frac{\partial V_{t+1}}{\partial W_{t+1}} R_{\omega,t+1} \Big) \tag{25}$$

Let us now consider the derivative of the value function w.r.t. current wealth

$$\begin{aligned} \frac{\partial V_t}{\partial W_t} &= \frac{V_t^{\frac{1}{\psi}}}{1-\frac{1}{\psi}} \Big\{ \big(1 - \frac{1}{\psi}\big) C_t^{-\frac{1}{\psi}} \frac{\partial C_t}{\partial W_t} + \frac{1-\frac{1}{\psi}}{1-\gamma} \beta \mathbb{E}_t (V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} (1-\gamma) \mathbb{E}_t \Big(V_{t+1}^{-\gamma} \frac{\partial V_{t+1}}{\partial W_{t+1}} \Big[\frac{\partial W_{t+1}}{\partial W_t} + \frac{\partial W_{t+1}}{\partial C_t} \frac{\partial C_t}{\partial W_t} \Big] \Big) \Big\} \\ &= V_t^{\frac{1}{\psi}} \Big\{ \Big[C_t^{-\frac{1}{\psi}} + \beta \mathbb{E}_t (V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} \mathbb{E}_t \Big(V_{t+1}^{-\gamma} \frac{\partial V_{t+1}}{\partial W_{t+1}} \frac{\partial W_{t+1}}{\partial C_t} \Big) \Big] \frac{\partial C_t}{\partial W_t} \\ &+ \beta \mathbb{E}_t (V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} \mathbb{E}_t \Big(V_{t+1}^{-\gamma} \frac{\partial V_{t+1}}{\partial W_{t+1}} \frac{\partial W_{t+1}}{\partial W_t} \Big) \Big\} \\ &= \beta \mathbb{E}_t (V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} \mathbb{E}_t \Big(V_{t+1}^{-\gamma} \frac{\partial V_{t+1}}{\partial W_{t+1}} \frac{\partial W_{t+1}}{\partial W_t} \Big) V_t^{\frac{1}{\psi}} \end{aligned} \tag{26}$$

where the last equality follows from the first-order condition for consumption. Since the budget constraint implies $\frac{\partial W_{t+1}}{\partial W_t} = R_{\omega,t+1}$, we get

$$\begin{aligned}\frac{\partial V_t}{\partial W_t} &= \beta \mathbb{E}_t(V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} \mathbb{E}_t\Big(V_{t+1}^{-\gamma} \frac{\partial V_{t+1}}{\partial W_{t+1}} R_{\omega,t+1}\Big) V_t^{\frac{1}{\psi-1}} \\ &= C_t^{-\frac{1}{\psi}} V_t^{\frac{1}{\psi}} \end{aligned} \tag{27}$$

Substituting the last expression into (25) yields

$$\begin{aligned} 1 &= \beta \mathbb{E}_t(V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} \mathbb{E}_t\Big(V_{t+1}^{\frac{1}{\psi}-\gamma} \Big[\frac{C_{t+1}}{C_t}\Big]^{-\frac{1}{\psi}} R_{\omega,t+1}\Big) \\ &= \beta \mathbb{E}_t\left(\left[\frac{V_{t+1}}{\mathbb{E}_t(V_{t+1}^{1-\gamma})^{\frac{1}{1-\gamma}}}\right]^{\frac{1}{\psi}-\gamma} \Big[\frac{C_{t+1}}{C_t}\Big]^{-\frac{1}{\psi}} R_{\omega,t+1}\right) \end{aligned} \tag{28}$$

Thus,

$$M_{t+1} = \beta \left[\frac{V_{t+1}}{\mathbb{E}_t(V_{t+1}^{1-\gamma})^{\frac{1}{1-\gamma}}}\right]^{\frac{1}{\psi}-\gamma} \Big[\frac{C_{t+1}}{C_t}\Big]^{-\frac{1}{\psi}} \tag{29}$$

is a valid SDF for the portfolio return $R_{\omega,t+1}$. Note that this SDF reduces to the standard CRRA expected utility SDF when $\gamma = \frac{1}{\psi}$. When the risk aversion is different from the inverse intertemporal rate of substitution, the SDF has an additional factor which is a function of the value function itself. The denominator in the brackets involving the value function, is the time t certainty equivalent of the value function. To see that this SDF is indeed a valid SDF for all assets in the economy, consider the first-order condition for the optimal portfolio weights ω_t. Let us rewrite the budget constraint as

$$W_{t+1} = (W_t - C_t)(R_{0,t+1} + \bar{\omega}_t^\top R_{t+1}^e) \tag{30}$$

and take the derivative of the right hand side of (21) w.r.t. $\bar{\omega}_t$ and set it equal 0

$$\begin{aligned} &\frac{\partial}{\partial \bar{\omega}_t} : \frac{V_t^{\frac{1}{\psi}}}{1-\frac{1}{\psi}} \Big\{ \frac{1-\frac{1}{\psi}}{1-\gamma} \beta \mathbb{E}_t(V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} (1-\gamma) \mathbb{E}_t\Big(V_{t+1}^{-\gamma} \frac{\partial V_{t+1}}{\partial W_{t+1}} \frac{\partial W_{t+1}}{\partial \bar{\omega}_t}\Big)\Big\} \\ &= V_t^{\frac{1}{\psi}} \Big\{\beta \mathbb{E}_t(V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} \mathbb{E}_t\Big(V_{t+1}^{-\gamma} \frac{\partial V_{t+1}}{\partial W_{t+1}} (W_t - C_t) R_{t+1}^e\Big)\Big\} = 0 \\ &\Updownarrow \\ &0 = \beta \mathbb{E}_t(V_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} \mathbb{E}_t\Big(V_{t+1}^{\frac{1}{\psi}-\gamma} C_{t+1}^{-\frac{1}{\psi}} R_{t+1}^e\Big) \\ &\Updownarrow \\ &0 = \mathbb{E}_t\Big(M_{t+1} R_{t+1}^e\Big) \end{aligned} \tag{31}$$

By the homogeneity of (21), the value function must be linear in wealth each time t. Thus, we can write

$$V_t = \phi_t W_t \tag{32}$$

where ϕ_t does not depend on wealth. Thus, $\frac{\partial V_t}{\partial W_t} = \phi_t$. Using this in (25) we get

$$\begin{aligned}
C_t^{-\frac{1}{\psi}} &= \beta \mathbb{E}_t(\phi_{t+1}^{1-\gamma} W_{t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} \mathbb{E}_t\Big(\phi_{t+1}^{-\gamma} W_{t+1}^{-\gamma} \phi_{t+1} R_{\omega,t+1}\Big) \\
&= \beta \mathbb{E}_t(\phi_{t+1}^{1-\gamma} (W_t - C_t)^{1-\gamma} R_{\omega,t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} \mathbb{E}_t\Big(\phi_{t+1}^{1-\gamma} (W_t - C_t)^{-\gamma} R_{\omega,t+1}^{1-\gamma}\Big) \\
&= \beta (W_t - C_t)^{\gamma-\frac{1}{\psi}} \mathbb{E}_t(\phi_{t+1}^{1-\gamma} R_{\omega,t+1}^{1-\gamma})^{\frac{\gamma-\frac{1}{\psi}}{1-\gamma}} (W_t - C_t)^{-\gamma} \mathbb{E}_t\Big(\phi_{t+1}^{1-\gamma} R_{\omega,t+1}^{1-\gamma}\Big) \\
&= \beta (W_t - C_t)^{-\frac{1}{\psi}} \mathbb{E}_t(\phi_{t+1}^{1-\gamma} R_{\omega,t+1}^{1-\gamma})^{\frac{1-\frac{1}{\psi}}{1-\gamma}} && (33) \\
&\equiv \beta (W_t - C_t)^{-\frac{1}{\psi}} \xi_t^{1-\frac{1}{\psi}} && (34)
\end{aligned}$$

where $\xi_t \equiv \mathbb{E}_t(\phi_{t+1}^{1-\gamma} R_{\omega,t+1}^{1-\gamma})^{\frac{1}{1-\gamma}}$. Rearranging yields

$$\xi_t^{1-\frac{1}{\psi}} = \beta^{-1} \Big(\frac{C_t}{W_t - C_t}\Big)^{-\frac{1}{\psi}} \tag{35}$$

$$\Updownarrow$$

$$\xi_t = \beta^{\frac{\psi}{1-\psi}} \Big(\frac{C_t}{W_t - C_t}\Big)^{\frac{1}{1-\psi}} \tag{36}$$

Note that

$$\mathbb{E}_t\Big(V_{t+1}^{1-\gamma}\Big)^{\frac{1-\frac{1}{\psi}}{1-\gamma}} = (W_t - C_t)^{1-\frac{1}{\psi}} \xi_t^{1-\frac{1}{\psi}} \tag{37}$$

Substitute (37) into (21) to obtain

$$\begin{aligned}
\phi_t W_t &= \Big\{C_t^{1-\frac{1}{\psi}} + \beta (W_t - C_t)^{1-\frac{1}{\psi}} \xi_t^{1-\frac{1}{\psi}}\Big\}^{\frac{1}{1-\frac{1}{\psi}}} \\
&= \Big\{C_t^{1-\frac{1}{\psi}} + \beta (W_t - C_t)^{1-\frac{1}{\psi}} \beta^{-1} \Big(\frac{C_t}{W_t - C_t}\Big)^{-\frac{1}{\psi}}\Big\}^{\frac{1}{1-\frac{1}{\psi}}} \\
&= \Big\{C_t^{1-\frac{1}{\psi}} + (W_t - C_t) C_t^{-\frac{1}{\psi}}\Big\}^{\frac{1}{1-\frac{1}{\psi}}} \\
&= \Big\{\Big(C_t + (W_t - C_t)\Big) C_t^{-\frac{1}{\psi}}\Big\}^{\frac{1}{1-\frac{1}{\psi}}} \\
&= W_t^{\frac{1}{1-\frac{1}{\psi}}} C_t^{\frac{1}{1-\psi}} && (38)
\end{aligned}$$

$$\Updownarrow$$

$$\phi_t = \Big(\frac{C_t}{W_t}\Big)^{\frac{1}{1-\psi}} \tag{39}$$

To arrive at an expression for the SDF, use (36), (37), and (38) in (29) to obtain

$$\begin{aligned}
M_{t+1} &= \beta\left[\frac{\phi_{t+1}W_{t+1}}{(W_t - C_t)\xi_t}\right]^{\frac{1}{\psi}-\gamma}\left[\frac{C_{t+1}}{C_t}\right]^{-\frac{1}{\psi}} \\
&= \beta\left[\frac{W_{t+1}^{\frac{1}{1-\frac{1}{\psi}}}C_{t+1}^{\frac{1}{1-\psi}}}{(W_t - C_t)\beta^{\frac{\psi}{1-\psi}}\left(\frac{C_t}{W_t - C_t}\right)^{\frac{1}{1-\psi}}}\right]^{\frac{1}{\psi}-\gamma}\left[\frac{C_{t+1}}{C_t}\right]^{-\frac{1}{\psi}} \\
&= \beta^{\frac{1-\gamma}{1-\frac{1}{\psi}}}\left[\frac{W_{t+1}}{W_t - C_t}\right]^{\frac{\frac{1}{\psi}-\gamma}{1-\frac{1}{\psi}}}\left[\frac{C_{t+1}}{C_t}\right]^{\frac{1-\gamma}{1-\psi}} \\
&= \beta^{\theta}\left[\frac{W_{t+1}}{W_t - C_t}\right]^{\theta-1}\left[\frac{C_{t+1}}{C_t}\right]^{-\frac{\theta}{\psi}} \\
&= \beta^{\theta}R_{\omega,t+1}^{\theta-1}\left[\frac{C_{t+1}}{C_t}\right]^{-\frac{\theta}{\psi}} \qquad (40)
\end{aligned}$$

where $\theta \equiv \frac{1-\gamma}{1-\frac{1}{\psi}}$.

B Appendix B - Log-linearization

Let $z_t \equiv log\Big(\frac{D_t}{P_t}\Big)$ be the log dividend-price ratio. Following Campbell and Shiller (1988), the log return on any asset can be written

$$
\begin{aligned}
r_{t+1} &= log\Big(\frac{P_{t+1} + D_{t+1}}{P_t}\Big) = log\Big(\Big[\frac{P_{t+1}}{D_{t+1}} + 1\Big]\frac{D_t}{P_t}\frac{D_{t+1}}{D_t}\Big) \\
&= log\Big(\frac{P_{t+1}}{D_{t+1}}\frac{D_t}{P_t} + \frac{D_t}{P_t}\Big) + log\Big(\frac{D_{t+1}}{D_t}\Big) \\
&= log\Bigg(exp\Big\{log\Big[\frac{P_{t+1}}{D_{t+1}}\Big] + log\Big[\frac{D_t}{P_t}\Big]\Big\} + exp\Big\{log\Big[\frac{D_t}{P_t}\Big]\Big\}\Bigg) + log\Big(\frac{D_{t+1}}{D_t}\Big) \\
&= log\big(e^{z_t - z_{t+1}} + e^{z_t}\big) + \Delta d_{t+1} \qquad (41)
\end{aligned}
$$

Consider a first order Taylor approximation to the function $h(z_t, z_{t+1}) \equiv log\big(e^{z_t - z_{t+1}} + e^{z_t}\big)$ around $z \equiv \mathbb{E}(z_t)$

$$
\begin{aligned}
h(z_t, z_{t+1}) &\approx h(z, z) + \frac{\partial h}{\partial z_t}(z, z)(z_t - z) + \frac{\partial h}{\partial z_{t+1}}(z, z)(z_{t+1} - z) \\
&= log(1 + e^z) + (z_t - z) - \frac{1}{1 + e^z}(z_{t+1} - z) \\
&\equiv \kappa_0 + z_t - \kappa_1 z_{t+1} \qquad (42)
\end{aligned}
$$

where $\kappa_0 \equiv log(1 + e^z) - \frac{e^z}{1+e^z} z$ and $\kappa_1 \equiv \frac{1}{1+e^z}$. The log return can therefore be approximated as

$$
r_{t+1} \approx \kappa_0 + z_t - \kappa_1 z_{t+1} + \Delta d_{t+1} \qquad (43)
$$

C Appendix C - MGF Inverse Gaussian

The moment generating function (MGF) for $\eta_i \sim IG((1-\varphi_i)\sigma_i^2, \lambda_i)$ is

$$\mathbb{M}_{\eta_i}(s) \equiv \mathbb{E}(e^{s\eta_i}) = exp\Big\{\frac{\lambda_i}{(1-\varphi_i)\sigma_i^2}\Big(1-\sqrt{1-\frac{2(1-\varphi_i)^2\sigma_i^4 s}{\lambda_i}}\Big)\Big\} \tag{44}$$

D Appendix D - Model Solution with log-linearization

D.1 SDF and Consumption Claim

Recall that the assumed exogenous processes guarding the evolution of the economy is

$$\begin{aligned}
\Delta c_{t+1} &= \mu_c + x_{t+1} - \frac{\tau}{2}\sigma_t^2 - \frac{\tau_x}{2}\sigma_{x,t}^2 + \sigma_t\varepsilon_{t+1}, & \varepsilon_{t+1} &\sim N(0,1)\\
\Delta d_{t+1} &= \mu_d + \varrho\Delta c_{t+1} - \frac{\tau_d}{2}\sigma_{d,t}^2 + \sigma_{d,t}\varepsilon_{d,t+1}, & \varepsilon_{d,t+1} &\sim N(0,1)\\
x_{t+1} &= \rho_x x_t + \sigma_{x,t}\varepsilon_{x,t+1}, & \varepsilon_{x,t+1} &\sim N(0,1)\\
\sigma_{t+1}^2 &= \varphi\sigma_t^2 + \eta_{t+1}, & \eta_{t+1} &\sim IG((1-\varphi)\sigma^2, \lambda)\\
\sigma_{d,t+1}^2 &= \varphi_d\sigma_{d,t}^2 + \eta_{d,t+1}, & \eta_{d,t+1} &\sim IG((1-\varphi_d)\sigma_d^2, \lambda_d)\\
\sigma_{x,t+1}^2 &= \varphi_x\sigma_{x,t}^2 + \eta_{x,t+1}, & \eta_{x,t+1} &\sim IG((1-\varphi_x)\sigma_x^2, \lambda_x)
\end{aligned}$$

and the log pricing kernel can be expressed

$$m_{t+1} = \theta log(\beta) - \frac{\theta}{\psi}\Delta c_{t+1} + (\theta-1)r_{c,t+1} \tag{45}$$

Furthermore, the return on the consumption portfolio can be approximated as

$$r_{c,t+1} = \kappa_{c,0} + z_{c,t} - \kappa_{c,1}z_{c,t+1} + \Delta c_{t+1} \tag{46}$$

Let us guess that

$$z_{c,t} = A_c + A_{c,x}x_t + A_{c,\sigma^2}\sigma_t^2 + A_{c,\sigma_x^2}\sigma_{x,t}^2 \tag{47}$$

Then

$$
\begin{aligned}
m_{t,t+1} &= \theta log(\beta) - \frac{\theta}{\psi}\Delta c_{t+1} + (\theta-1)(\kappa_{c,0} + z_{c,t} - \kappa_{c,1} z_{c,t+1} + \Delta c_{t+1}) \\
&= \theta log(\beta) + (\theta-1)\kappa_{c,0} - \Big(1 + \frac{\theta}{\psi} - \theta\Big)\Delta c_{t+1} + (\theta-1) z_{c,t} - (\theta-1)\kappa_{c,1} z_{c,t+1} \\
&= \theta log(\beta) + (\theta-1)\kappa_{c,0} - \gamma\Big(\mu_c + \rho_x x_t + \sigma_{x,t}\varepsilon_{x,t+1} - \frac{\tau}{2}\sigma_t^2 - \frac{\tau_x}{2}\sigma_{x,t}^2 + \sigma_t \varepsilon_{t+1}\Big) \\
&+ (\theta-1)\big(A_c + A_{c,x} x_t + A_{c,\sigma^2}\sigma_t^2 + A_{c,\sigma_x^2}\sigma_{x,t}^2\big) \\
&- (\theta-1)\kappa_{c,1}\Big(A_c + A_{c,x}(\rho_x x_t + \sigma_{x,t}\varepsilon_{x,t+1}) + A_{c,\sigma^2}(\varphi\sigma_t^2 + \eta_{t+1}) + A_{c,\sigma_x^2}(\varphi_x \sigma_{x,t}^2 + \eta_{x,t+1})\Big) \\
&= \Big(\theta log(\beta) + (\theta-1)\kappa_{c,0} - \gamma\mu_c + (\theta-1)(1-\kappa_{c,1})A_c\Big) \\
&+ \Big((\theta-1)(1-\kappa_{c,1}\rho_x)A_{c,x} - \gamma\rho_x\Big) x_t \\
&+ \Big((\theta-1)(1-\kappa_{c,1}\varphi)A_{c,\sigma^2} + \frac{\tau\gamma}{2}\Big)\sigma_t^2 \\
&+ \Big((\theta-1)(1-\kappa_{c,1}\varphi_x)A_{c,\sigma_x^2} + \frac{\tau_x\gamma}{2}\Big)\sigma_{x,t}^2 \\
&- \gamma\sigma_t\varepsilon_{t+1} - \Big(\gamma + (\theta-1)\kappa_{c,1}A_{c,x}\Big)\sigma_{x,t}\varepsilon_{x,t+1} - (\theta-1)\kappa_{c,1}\Big(A_{c,\sigma^2}\eta_{t+1} + A_{c,\sigma_x^2}\eta_{x,t+1}\Big) \\
&\equiv \Gamma_m + \Gamma_{m,x} x_t + \Gamma_{m,\sigma^2}\sigma_t^2 + \Gamma_{m,\sigma_x^2}\sigma_{x,t}^2 - \Gamma_{m,\varepsilon}\sigma_t\varepsilon_{t+1} - \Gamma_{m,\varepsilon_x}\sigma_{x,t}\varepsilon_{x,t+1} - \Gamma_{m,\eta}\eta_{t+1} - \Gamma_{m,\eta_x}\eta_{x,t+1}
\end{aligned}
$$

where the third equality used $1 + \frac{\theta}{\psi} - \theta = \gamma$. Similarly, the return on the consumption portfolio can be written

$$
\begin{aligned}
r_{c,t+1} &= \kappa_{c,0} + z_{c,t} - \kappa_{c,1} z_{c,t+1} + \Delta c_{t+1} \\
&= \Big(\kappa_{c,0} + (1-\kappa_{c,1})A_c + \mu_c\Big) + \Big(\rho_x + (1-\kappa_{c,1}\rho_x)A_{c,x}\Big)x_t + \Big((1-\kappa_{c,1}\varphi)A_{c,\sigma^2} - \frac{\tau}{2}\Big)\sigma_t^2 \\
&+ \Big((1-\kappa_{c,1}\varphi_x)A_{c,\sigma_x^2} - \frac{\tau_x}{2}\Big)\sigma_{x,t}^2 + \sigma_t\varepsilon_{t+1} + (1-\kappa_{c,1}A_{c,x})\sigma_{x,t}\varepsilon_{x,t+1} - \kappa_{c,1}(A_{c,\sigma^2}\eta_{t+1} + A_{c,\sigma_x^2}\eta_{x,t+1}) \\
&\equiv \Gamma_{r_c} + \Gamma_{r_c,x} x_t + \Gamma_{r_c,\sigma^2}\sigma_t^2 + \Gamma_{r_c,\sigma_x^2}\sigma_{x,t}^2 + \Gamma_{r_c,\varepsilon}\sigma_t\varepsilon_{t+1} + \Gamma_{r_c,\varepsilon_x}\sigma_{x,t}\varepsilon_{x,t+1} + \Gamma_{r_c,\eta}\eta_{t+1} + \Gamma_{r_c,\eta_x}\eta_{x,t+1}
\end{aligned}
$$

Where

$$
\begin{aligned}
\Gamma_m &\equiv \theta log(\beta) + (\theta - 1)\kappa_{c,0} - \gamma\mu_c + (\theta - 1)(1 - \kappa_{c,1})A_c \\
\Gamma_{m,x} &\equiv (\theta - 1)(1 - \kappa_{c,1}\rho_x)A_{c,x} - \gamma\rho_x \\
\Gamma_{m,\sigma^2} &\equiv (\theta - 1)(1 - \kappa_{c,1}\varphi)A_{c,\sigma^2} + \frac{\tau\gamma}{2} \\
\Gamma_{m,\sigma_x^2} &\equiv (\theta - 1)(1 - \kappa_{c,1}\varphi_x)A_{c,\sigma_x^2} + \frac{\tau_x\gamma}{2} \\
\Gamma_{m,\varepsilon} &\equiv \gamma \\
\Gamma_{m,\varepsilon_x} &\equiv \gamma + (\theta - 1)\kappa_{c,1}A_{c,x} \\
\Gamma_{m,\eta} &\equiv (\theta - 1)\kappa_{c,1}A_{c,\sigma^2} \\
\Gamma_{m,\eta_x} &\equiv (\theta - 1)\kappa_{c,1}A_{c,\sigma_x^2}
\end{aligned}
$$

and

$$
\begin{aligned}
\Gamma_{r_c} &\equiv \kappa_{c,0} + (1 - \kappa_{c,1})A_c + \mu_c \\
\Gamma_{r_c,x} &\equiv \rho_x + (1 - \kappa_{c,1}\rho_x)A_{c,x} \\
\Gamma_{r_c,\sigma^2} &\equiv (1 - \kappa_{c,1}\varphi)A_{c,\sigma^2} - \frac{\tau}{2} \\
\Gamma_{r_c,\sigma_x^2} &\equiv (1 - \kappa_{c,1}\varphi_x)A_{c,\sigma_x^2} - \frac{\tau_x}{2} \\
\Gamma_{r_c,\varepsilon} &\equiv 1 \\
\Gamma_{r_c,\varepsilon_x} &\equiv 1 - \kappa_{c,1}A_{c,x} \\
\Gamma_{r_c,\eta} &\equiv -\kappa_{c,1}A_{c,\sigma^2} \\
\Gamma_{r_c,\eta_x} &\equiv -\kappa_{c,1}A_{c,\sigma_x^2}
\end{aligned}
$$

It is therefore clear that

$$
\begin{aligned}
m_{t+1} + r_{c,t+1} = {} & \Gamma_{mr_c} + \Gamma_{mr_c,x}x_t + \Gamma_{mr_c,\sigma^2}\sigma_t^2 + \Gamma_{mr_c,\sigma_x^2}\sigma_{x,t}^2 \\
& + \Gamma_{mr_c,\varepsilon}\sigma_t\varepsilon_{t+1} + \Gamma_{mr_c,\varepsilon_x}\sigma_{x,t}\varepsilon_{x,t+1} + \Gamma_{mr_c,\eta}\eta_{t+1} + \Gamma_{mr_c,\eta_x}\eta_{x,t+1}
\end{aligned}
$$

with

$$\begin{aligned}\Gamma_{mr_c} &\equiv \Gamma_{r_c} + \Gamma_m \\ &= \theta log(\beta) - (\gamma-1)\mu_c + \theta\kappa_{c,0} + \theta(1-\kappa_{c,1})A_c\end{aligned}$$

$$\begin{aligned}\Gamma_{mr_c,x} &\equiv \Gamma_{r_c,x} + \Gamma_{m,x} \\ &= \theta(1-\kappa_{c,1}\rho_x)A_{c,x} - (\gamma-1)\rho_x\end{aligned}$$

$$\begin{aligned}\Gamma_{mr_c,\sigma^2} &\equiv \Gamma_{r_c,\sigma^2} + \Gamma_{m,\sigma^2} \\ &= \theta(1-\kappa_{c,1}\varphi)A_{c,\sigma^2} + \frac{\tau}{2}\Big(\frac{\theta}{\psi} - \theta\Big)\end{aligned}$$

$$\begin{aligned}\Gamma_{mr_c,\sigma_x^2} &\equiv \Gamma_{r_c,\sigma_x^2} + \Gamma_{m,\sigma_x^2} \\ &= \theta(1-\kappa_{c,1}\varphi_x)A_{c,\sigma_x^2} + \frac{\tau_x}{2}\Big(\frac{\theta}{\psi} - \theta\Big)\end{aligned}$$

$$\begin{aligned}\Gamma_{mr_c,\varepsilon} &\equiv \Gamma_{r_c,\varepsilon} - \Gamma_{m,\varepsilon} \\ &= -(\gamma-1)\end{aligned}$$

$$\begin{aligned}\Gamma_{mr_c,\varepsilon_x} &\equiv \Gamma_{r_c,\varepsilon_x} - \Gamma_{m,\varepsilon_x} \\ &= -\theta\kappa_{c,1}A_{c,x} - (\gamma-1)\end{aligned}$$

$$\begin{aligned}\Gamma_{mr_c,\eta} &\equiv \Gamma_{r_c,\eta} - \Gamma_{m,\eta} \\ &= -\theta\kappa_{c,1}A_{c,\sigma^2}\end{aligned}$$

$$\begin{aligned}\Gamma_{mr_c,\eta_x} &\equiv \Gamma_{r_c,\eta_x} - \Gamma_{m,\eta_x} \\ &= -\theta\kappa_{c,1}A_{c,\sigma_x^2}\end{aligned}$$

Note that all these coefficients equals 0 if $\gamma = 1$. $m_{t+1} + r_{c,t+1}$ is not conditionally normal as seen from time t. It is however, conditionally normal based on the information set that contains both time t

information and the realization of $\eta_{t+1},\ \eta_{x,t+1}$. Thus

$$
\begin{aligned}
\mathbb{E}_t(e^{m_{t+1}+r_{c,t+1}}) &= \mathbb{E}_t\big(\mathbb{E}_t(e^{m_{t+1}+r_{c,t+1}}|\eta_{t+1},\ \eta_{x,t+1})\big) \\
&= \mathbb{E}_t\Big(e^{\Gamma_{mr_c}+\Gamma_{mr_c,x}x_t+(\Gamma_{mr_c,\sigma^2}+\frac{\Gamma^2_{mr_c,\varepsilon}}{2})\sigma_t^2+(\Gamma_{mr_c,\sigma_x^2}+\frac{\Gamma^2_{mr_c,\varepsilon_x}}{2})\sigma^2_{x,t}+\Gamma_{mr_c,\eta}\eta_{t+1}+\Gamma_{mr_c,\eta_x}\eta_{x,t+1}}\Big) \\
&\equiv B_{mr_c}e^{\Gamma_{mr_c,x}x_t+(\Gamma_{mr_c,\sigma^2}+\frac{\Gamma^2_{mr_c,\varepsilon}}{2})\sigma_t^2+(\Gamma_{mr_c,\sigma_x^2}+\frac{\Gamma^2_{mr_c,\varepsilon_x}}{2})\sigma^2_{x,t}}
\end{aligned}
$$

where $B_{mr_c} \equiv e^{\Gamma_{mr_c}}\mathbb{M}_\eta(\Gamma_{mr_c,\eta})\mathbb{M}_{\eta_x}(\Gamma_{mr_c,\eta_x})$ is a constant (due to $\eta_{\cdot,t+1}$ i.i.d. across time). The pricing condition

$$
\begin{aligned}
1 &= \mathbb{E}_t(e^{m_{t+1}+r_{c,t+1}}) \Leftrightarrow \\
0 &= log(B_{mr_c}) + \Gamma_{mr_c,x}x_t + \Big(\Gamma_{mr_c,\sigma^2}+\frac{\Gamma^2_{mr_c,\varepsilon}}{2}\Big)\sigma_t^2 + \Big(\Gamma_{mr_c,\sigma_x^2}+\frac{\Gamma^2_{mr_c,\varepsilon_x}}{2}\Big)\sigma^2_{x,t}
\end{aligned} \tag{48}
$$

This expression can only be satisfied for all t if

$$
\begin{aligned}
0 &= log(B_{mr_c}) \\
0 &= \Gamma_{mr_c,x} \\
0 &= \Gamma_{mr_c,\sigma^2} + \frac{\Gamma^2_{mr_c,\varepsilon}}{2} \\
0 &= \Gamma_{mr_c,\sigma_x^2} + \frac{\Gamma^2_{mr_c,\varepsilon_x}}{2}
\end{aligned} \tag{49}
$$

We therefore get

$$
\begin{aligned}
0 &= \theta(1-\kappa_{c,1}\rho_x)A_{c,x} - (\gamma-1)\rho_x \Leftrightarrow \\
A_{c,x} &= -\frac{1-\frac{1}{\psi}}{1-\kappa_{c,1}\rho_x}\rho_x
\end{aligned} \tag{50}
$$

$$
\begin{aligned}
0 &= \theta(1-\kappa_{c,1}\varphi)A_{c,\sigma^2} + \frac{\tau}{2}(\gamma-1) + \frac{(\gamma-1)^2}{2} \Leftrightarrow \\
A_{c,\sigma^2} &= \frac{1}{2}\Big(\tau-1+\gamma\Big)\frac{1-\frac{1}{\psi}}{1-\kappa_{c,1}\varphi}
\end{aligned}
$$

$$
\begin{aligned}
0 &= \theta(1-\kappa_{c,1}\varphi_x)A_{c,\sigma_x^2} + \frac{\tau_x}{2}(\gamma-1) + \frac{1}{2}\Big(\gamma-1+\theta\kappa_{c,1}A_{c,x}\Big)^2 \Leftrightarrow \\
A_{c,\sigma_x^2} &= \frac{1}{2}\Big(\tau_x + \Big(\frac{1}{1-\kappa_{c,1}\rho_x}\Big)^2(\gamma-1)\Big)\frac{1-\frac{1}{\psi}}{1-\kappa_{c,1}\varphi_x}
\end{aligned} \tag{51}
$$

Note that all these expressions would be the same if the variances were conditionally normal instead of Inverse Gaussian (the only expression that would be different is that for B_{mr_c}). A_c is determined by the condition $log(B_{mr_c}) = 0$

$$\begin{aligned} 0 = logB_{mr_c} &\Leftrightarrow \Gamma_{mr_c} = -log\mathbb{M}_\eta(\Gamma_{mr_c,\eta}) - log\mathbb{M}_{\eta_x}(\Gamma_{mr_c,\eta_x}) \Leftrightarrow \\ A_c &= -\frac{\theta log\beta - (\gamma-1)\mu_c + \theta\kappa_{c,0} + log\mathbb{M}_\eta(\Gamma_{mr_c,\eta}) + log\mathbb{M}_{\eta_x}(\Gamma_{mr_c,\eta_x})}{\theta(1-\kappa_{c,1})} \\ &= -\frac{log\beta + \kappa_{c,0} + \left(1-\frac{1}{\psi}\right)\mu_c + \left(log\mathbb{M}_\eta(\Gamma_{mr_c,\eta}) + log\mathbb{M}_{\eta_x}(\Gamma_{mr_c,\eta_x})\right)\frac{1-\frac{1}{\psi}}{1-\gamma}}{1-\kappa_{c,1}} \end{aligned} \tag{52}$$

which depends on the distribution of η_i because its moment generating function enters the expression. Note that if $\psi = 1$, the expression simplifies to $A_c = -\frac{log\beta+\kappa_{c,0}}{1-\kappa_{c,1}}$ and if $\gamma = 1$, we have $A_c = -\frac{log\beta+\kappa_{c,0}+\left(1-\frac{1}{\psi}\right)\mu_c}{1-\kappa_{c,1}}$ since $\Gamma_{mr_c,\eta} = \Gamma_{mr_c,\eta_x} = 0$. Furthermore, the constants $\kappa_{c,0},\ \kappa_{c,1}$ are functions of the endogenous expected dividend-price ratio z_c, which is determined by solving the following non-linear equation for z_c

$$0 = A_c(z_c) + A_{\sigma^2}(z_c)\sigma^2 + A_{\sigma_x^2}(z_c)\sigma_x^2 - z_c \tag{53}$$

The sign on $A_{c,x}$ is determined by ψ. If $\psi > 1$, $A_{c,x} < 0$. Thus, the log dividend-price ratio is decreasing in x_t. The sign on A_{c,σ^2} is slightly more complicated. If $\psi > 1$, the sign is given by $\tau - 1 + \gamma$. In particular, when $\tau = 1$, i.e. σ_t^2 does not predict consumption growth in levels, $A_{c,\sigma^2} > 0$. If $\tau = 0$, i.e. σ_t^2 does not predict log consumption growth, $A_{c,\sigma^2} > 0$ if $\gamma > 1$. As a result, the log dividend-price ratio is increasing in σ_t^2 for all $\gamma \geq 1$ as long as $\tau \geq 0$. Similarly, the sign on A_{c,σ_x^2} is positive if $\psi > 1$, $\gamma > 1$ and $\tau_x \geq 0$. A dividend-price ratio that is increasing in risk makes intuitive sense, as higher risk causes prices to fall for a given expected cash-flow.

To gain some understanding as to what is considered bad states of the world by this model, it is helpful to revisit the expressions for $\Gamma_{m,i}$.

$$\begin{aligned}
\Gamma_{m,x} &= -\frac{\frac{1}{\psi}-\gamma}{1-\frac{1}{\psi}}(1-\kappa_{c,1}\rho_x)\frac{1-\frac{1}{\psi}}{1-\kappa_{c,1}\rho_x}\rho_x - \gamma\rho_x \\
&= -\frac{1}{\psi}\rho_x && (54)\\
\Gamma_{m,\sigma^2} &= \frac{\frac{1}{\psi}-\gamma}{1-\frac{1}{\psi}}(1-\kappa_{c,1}\varphi)\frac{1}{2}\Big(\tau-1+\gamma\Big)\frac{1-\frac{1}{\psi}}{1-\kappa_{c,1}\varphi}+\frac{\tau}{2}\gamma \\
&= \frac{1}{2}\Big(\frac{\tau}{\psi}+\Big[\frac{1}{\psi}-\gamma\Big](\gamma-1)\Big) && (55)\\
\Gamma_{m,\sigma_x^2} &= \frac{\frac{1}{\psi}-\gamma}{1-\frac{1}{\psi}}(1-\kappa_{c,1}\varphi_x)\frac{1}{2}\Big(\tau_x+\Big(\frac{1}{1-\kappa_{c,1}\rho_x}\Big)^2(\gamma-1)\Big)\frac{1-\frac{1}{\psi}}{1-\kappa_{c,1}\varphi_x}+\frac{\tau_x}{2}\gamma \\
&= \frac{1}{2}\Big(\frac{\tau_x}{\psi}+\Big[\frac{1}{\psi}-\gamma\Big]\frac{\gamma-1}{(1-\kappa_{c,1}\rho_x)^2}\Big) && (56)\\
\Gamma_{m,\varepsilon} &= \gamma && (57)\\
\Gamma_{m,\varepsilon_x} &= \gamma+\frac{\frac{1}{\psi}-\gamma}{1-\frac{1}{\psi}}\kappa_{c,1}\frac{\frac{1}{\psi}-1}{1-\kappa_{c,1}\rho_x}\rho_x \\
&= \frac{\gamma-\frac{\kappa_{c,1}\rho_x}{\psi}}{1-\kappa_{c,1}\rho_x} && (58)\\
\Gamma_{m,\eta} &= \frac{\frac{1}{\psi}-\gamma}{1-\frac{1}{\psi}}\kappa_{c,1}\frac{1}{2}\Big(\tau-1+\gamma\Big)\frac{1-\frac{1}{\psi}}{1-\kappa_{c,1}\varphi} \\
&= -\frac{1}{2}(\tau-1+\gamma)\kappa_{c,1}\frac{\gamma-\frac{1}{\psi}}{1-\kappa_{c,1}\varphi} && (59)\\
\Gamma_{m,\eta_x} &= \frac{\frac{1}{\psi}-\gamma}{1-\frac{1}{\psi}}\kappa_{c,1}\frac{1}{2}\Big(\tau_x+\Big(\frac{1}{1-\kappa_{c,1}\rho_x}\Big)^2(\gamma-1)\Big)\frac{1-\frac{1}{\psi}}{1-\kappa_{c,1}\varphi_x} \\
&= -\frac{1}{2}\Big(\tau_x+\frac{\gamma-1}{(1-\kappa_{c,1}\rho_x)^2}\Big)\kappa_{c,1}\frac{\gamma-\frac{1}{\psi}}{1-\kappa_{c,1}\varphi_x} && (60)
\end{aligned}$$

Suppose $\gamma > 1$, $\psi \geq 1$ and $\tau,\ \tau_x = 0$. Then $\Gamma_{m,\eta},\ \Gamma_{m,\eta_x} < 0$, which implies m_{t+1} (marginal utility) is increasing in shocks to variance. In this case, assets that pay off more when shocks to variance are positive, will tend to carry a lower (negative) risk premium. Similarly, $\Gamma_{m,\varepsilon},\ \Gamma_{m,\varepsilon_x} > 0$, and as a result m_{t+1} (marginal utility) is decreasing in shocks to short and long run consumption (as expected).

D.2 Fundamental Stock Market Portfolio

Now, let us turn our attention to the fundamental market portfolio. Use the log linearization technique for the return on the market to get

$$r_{d,t+1} \approx \kappa_{d,0} + z_{d,t} - \kappa_{d,1}z_{d,t+1} + \Delta d_{d,t+1} \qquad (61)$$

and assume $z_{d,t} = A_d + A_{d,x}x_t + A_{d,\sigma^2}\sigma_t^2 + A_{d,\sigma_x^2}\sigma_{x,t}^2 + A_{d,\sigma_d^2}\sigma_{d,t}^2$. We then have

$$
\begin{aligned}
r_{d,t+1} &= \kappa_{d,0} + A_d + A_{d,x}x_t + A_{d,\sigma^2}\sigma_t^2 + A_{d,\sigma_x^2}\sigma_{x,t}^2 + A_{d,\sigma_d^2}\sigma_{d,t}^2 \\
&- \kappa_{d,1}\big(A_d + A_{d,x}x_{t+1} + A_{d,\sigma^2}\sigma_{t+1}^2 + A_{d,\sigma_x^2}\sigma_{x,t+1}^2 + A_{d,\sigma_d^2}\sigma_{d,t+1}^2\big) \\
&+ \mu_d + \varrho\big(\mu_c + x_{t+1} - \frac{\tau}{2}\sigma_t^2 - \frac{\tau_x}{2}\sigma_{x,t}^2 + \sigma_t\varepsilon_{t+1}\big) - \frac{\tau_d}{2}\sigma_{d,t}^2 + \sigma_{d,t}\varepsilon_{d,t+1} \\
&= \kappa_{d,0} + A_d + A_{d,x}x_t + A_{d,\sigma^2}\sigma_t^2 + A_{d,\sigma_x^2}\sigma_{x,t}^2 + A_{d,\sigma_d^2}\sigma_{d,t}^2 \\
&- \kappa_{d,1}\Big(A_d + A_{d,x}(\rho_x x_t + \sigma_{x,t}\varepsilon_{x,t+1}) + A_{d,\sigma^2}(\varphi\sigma_t^2 + \eta_{t+1}) + A_{d,\sigma_x^2}(\varphi_x\sigma_{x,t}^2 + \eta_{x,t+1}) \\
&+ A_{d,\sigma_d^2}(\varphi_d\sigma_{d,t}^2 + \eta_{d,t+1})\Big) \\
&+ \mu_d + \varrho\Big(\mu_c + (\rho_x x_t + \sigma_{x,t}\varepsilon_{x,t+1}) - \frac{\tau}{2}\sigma_t^2 - \frac{\tau_x}{2}\sigma_{x,t}^2 + \sigma_t\varepsilon_{t+1}\Big) - \frac{\tau_d}{2}\sigma_{d,t}^2 + \sigma_{d,t}\varepsilon_{d,t+1} \\
&= \Big(\kappa_{d,0} + (1-\kappa_{d,1})A_d + \mu_d + \varrho\mu_c\Big) \\
&+ \Big((1-\kappa_{d,1}\rho_x)A_{d,x} + \varrho\rho_x\Big)x_t \\
&+ \Big((1-\kappa_{d,1}\varphi)A_{d,\sigma^2} - \varrho\frac{\tau}{2}\Big)\sigma_t^2 \\
&+ \Big((1-\kappa_{d,1}\varphi_x)A_{d,\sigma_x^2} - \varrho\frac{\tau_x}{2}\Big)\sigma_{x,t}^2 \\
&+ \Big((1-\kappa_{d,1}\varphi_d)A_{d,\sigma_d^2} - \frac{\tau_d}{2}\Big)\sigma_{d,t}^2 \\
&+ \varrho\sigma_t\varepsilon_{t+1} + (\varrho - \kappa_{d,1}A_{d,x})\sigma_{x,t}\varepsilon_{x,t+1} + \sigma_{d,t}\varepsilon_{d,t+1} - \kappa_{d,1}A_{d,\sigma^2}\eta_{t+1} - \kappa_{d,1}A_{d,\sigma_x^2}\eta_{x,t+1} - \kappa_{d,1}A_{d,\sigma_d^2}\eta_{d,t+1} \\
&\equiv \Gamma_{r_d} + \Gamma_{r_d,x}x_t + \Gamma_{r_d,\sigma^2}\sigma_t^2 + \Gamma_{r_d,\sigma_x}\sigma_{x,t}^2 + \Gamma_{r_d,\sigma_d^2}\sigma_{d,t}^2 \\
&+ \Gamma_{r_d,\varepsilon}\sigma_t\varepsilon_{t+1} + \Gamma_{r_d,\varepsilon_x}\sigma_{x,t}\varepsilon_{x,t+1} + \Gamma_{r_d,\varepsilon_d}\sigma_{d,t}\varepsilon_{d,t+1} + \Gamma_{r_d,\eta}\eta_{t+1} + \Gamma_{r_d,\eta_x}\eta_{x,t+1} + \Gamma_{r_d,\eta_d}\eta_{d,t+1} \quad (62)
\end{aligned}
$$

where

$$\Gamma_{r_d} \equiv \kappa_{d,0} + (1-\kappa_{d,1})A_d + \mu_d + \varrho\mu_c \tag{63}$$

$$\Gamma_{r_d,x} \equiv (1-\kappa_{d,1}\rho_x)A_{d,x} + \varrho\rho_x \tag{64}$$

$$\Gamma_{r_d,\sigma^2} \equiv (1-\kappa_{d,1}\varphi)A_{d,\sigma^2} - \frac{\varrho\tau}{2} \tag{65}$$

$$\Gamma_{r_d,\sigma_x^2} \equiv (1-\kappa_{d,1}\varphi_x)A_{d,\sigma_x^2} - \frac{\varrho\tau_x}{2} \tag{66}$$

$$\Gamma_{r_d,\sigma_d^2} \equiv (1-\kappa_{d,1}\varphi_d)A_{d,\sigma_d^2} - \frac{\tau_d}{2} \tag{67}$$

$$\Gamma_{r_d,\varepsilon} \equiv \varrho \tag{68}$$

$$\Gamma_{r_d,\varepsilon_x} \equiv \varrho - \kappa_{d,1}A_{d,x} \tag{69}$$

$$\Gamma_{r_d,\varepsilon_d} \equiv 1 \tag{70}$$

$$\Gamma_{r_d,\eta} \equiv -\kappa_{d,1}A_{d,\sigma^2} \tag{71}$$

$$\Gamma_{r_d,\eta_x} \equiv -\kappa_{d,1}A_{d,\sigma_x^2} \tag{72}$$

$$\Gamma_{r_d,\eta_d} \equiv -\kappa_{d,1}A_{d,\sigma_d^2} \tag{73}$$

As a result,

$$\begin{aligned} m_{t+1} + r_{d,t+1} &= (\Gamma_m + \Gamma_{r_d}) + (\Gamma_{m,x} + \Gamma_{r_d,x})x_t + (\Gamma_{m,\sigma^2} + \Gamma_{r_d,\sigma^2})\sigma_t^2 + (\Gamma_{m,\sigma_x^2} + \Gamma_{r_d,\sigma_x^2})\sigma_{x,t}^2 + \Gamma_{r_d,\sigma_d^2}\sigma_{d,t}^2 \\ &+ (\Gamma_{r_d,\varepsilon} - \Gamma_{m,\varepsilon})\sigma_t\varepsilon_{t+1} + (\Gamma_{r_d,\varepsilon_x} - \Gamma_{m,\varepsilon_x})\sigma_{x,t}\varepsilon_{x,t+1} + \Gamma_{r_d,\varepsilon_d}\sigma_{d,t}\varepsilon_{d,t+1} + (\Gamma_{r_d,\eta} - \Gamma_{m,\eta})\eta_{t+1} \\ &+ (\Gamma_{r_d,\eta_x} - \Gamma_{m,\eta_x})\eta_{x,t+1} + \Gamma_{r_d,\eta_d}\eta_{d,t+1} \\ &\equiv \Gamma_{mr_d} + \Gamma_{mr_d,x}x_t + \Gamma_{mr_d,\sigma^2}\sigma_t^2 + \Gamma_{mr_d,\sigma_x^2}\sigma_{x,t}^2 + \Gamma_{r_d,\sigma_d^2}\sigma_{d,t}^2 + \Gamma_{mr_d,\varepsilon}\sigma_t\varepsilon_{t+1} \\ &+ \Gamma_{mr_d,\varepsilon_x}\sigma_{x,t}\varepsilon_{x,t+1} + \Gamma_{r_d,\varepsilon_d}\sigma_{d,t}\varepsilon_{d,t+1} + \Gamma_{mr_d,\eta}\eta_{t+1} + \Gamma_{mr_d,\eta_x}\eta_{x,t+1} + \Gamma_{r_d,\eta_d}\eta_{d,t+1} \end{aligned} \tag{74}$$

It is clear from the pricing condition that

$$0 = log(B_{mr_d}) + \Gamma_{mr_d,x}x_t + (\Gamma_{mr_d,\sigma^2} + \frac{\Gamma_{mr_d,\varepsilon}^2}{2})\sigma_t^2 + (\Gamma_{mr_d,\sigma_x^2} + \frac{\Gamma_{mr_d,\varepsilon_x}^2}{2})\sigma_{x,t}^2 + (\Gamma_{r_d,\sigma_d^2} + \frac{\Gamma_{r_d,\varepsilon_d}^2}{2})\sigma_{d,t}^2 \tag{75}$$

where $B_{mr_d} \equiv e^{\Gamma_{mr_d}}\mathbb{M}_\eta(\Gamma_{mr_d,\eta})\mathbb{M}_{\eta_x}(\Gamma_{mr_d,\eta_x})\mathbb{M}_{\eta_d}(\Gamma_{r_d,\eta_d})$. In order for the pricing condition to hold for all realizations of the state variables, the coefficients multiplying the state variables must each be equal to 0. Thus

$$
\begin{aligned}
0 = \Gamma_{mr_d,x} &\Leftrightarrow \Gamma_{r_d,x} = -\Gamma_{m,x} \Leftrightarrow \\
A_{d,x} &= -\frac{\Gamma_{m,x} + \varrho\rho_x}{1-\kappa_{d,1}\rho_x}
\end{aligned} \tag{76}
$$

$$
\begin{aligned}
0 = \Gamma_{mr_d,\sigma^2} + \frac{\Gamma^2_{mr_d,\varepsilon}}{2} &\Leftrightarrow \Gamma_{r_d,\sigma^2} = -\Gamma_{m,\sigma^2} - \frac{\Gamma^2_{mr_d,\varepsilon}}{2} \Leftrightarrow \\
A_{d,\sigma^2} &= -\frac{\Gamma_{m,\sigma^2}}{1-\kappa_{d,1}\varphi} - \frac{1}{2}\frac{(\varrho-\gamma)^2 - \varrho\tau}{1-\kappa_{d,1}\varphi}
\end{aligned} \tag{77}
$$

$$
\begin{aligned}
0 = \Gamma_{mr_d,\sigma_x^2} + \frac{\Gamma^2_{mr_d,\varepsilon_x}}{2} &\Leftrightarrow \Gamma_{r_d,\sigma_x^2} = -\Gamma_{m,\sigma_x^2} - \frac{\Gamma^2_{mr_d,\varepsilon_x}}{2} \Leftrightarrow \\
A_{d,\sigma_x^2} &= -\frac{\Gamma_{m,\sigma_x^2}}{1-\kappa_{d,1}\varphi_x} - \frac{1}{2}\frac{\left(\varrho - \kappa_{d,1}A_{d,x} - \frac{\gamma - \frac{\kappa_{c,1}\rho_x}{\psi}}{1-\kappa_{c,1}\rho_x}\right)^2 - \varrho\tau_x}{1-\kappa_{d,1}\varphi_x}
\end{aligned} \tag{78}
$$

$$
\begin{aligned}
0 = \Gamma_{r_d,\sigma_d^2} + \frac{\Gamma^2_{r_d,\varepsilon_d}}{2} &\Leftrightarrow \\
A_{d,\sigma_d^2} &= \frac{1}{2}\frac{\tau_d - 1}{1-\kappa_{d,1}\varphi_d}
\end{aligned} \tag{79}
$$

Notice that the condition for A_{d,σ_d^2} states that expected return (not log-returns) on the market portfolio is not predicted by the idiosyncratic variance. Furthermore, if $\tau_d = 1$, the dividend-price ratio is not a function of the idiosyncratic variance. This makes sense as dividend growth (not log-dividend growth) is not predictable by the idiosyncratic variance in this case. Since expected return is also unrelated to the idiosyncratic variance, the dividend-price ratio must be unrelated as well. If $\tau_d > 1$, future dividends are expected to fall, resulting in a fall in stock market price and an increasing dividend-price ratio.

Finally, we get A_d from

$$
\begin{aligned}
0 = log(B_{mr_d}) &\Leftrightarrow \Gamma_{mr_d} = -log\mathbb{M}_{\eta}(\Gamma_{mr_d,\eta}) - log\mathbb{M}_{\eta_x}(\Gamma_{mr_d,\eta_x}) - log\mathbb{M}_{\eta_d}(\Gamma_{r_d,\eta_d}) \Leftrightarrow \\
\Gamma_{r_d} &= -\Gamma_m - log\mathbb{M}_{\eta}(\Gamma_{mr_d,\eta}) - log\mathbb{M}_{\eta_x}(\Gamma_{mr_d,\eta_x}) - log\mathbb{M}_{\eta_d}(\Gamma_{r_d,\eta_d}) \Leftrightarrow \\
A_d &= -\frac{\Gamma_m + \kappa_{d,0} + \mu_d + \varrho\mu_c + log\mathbb{M}_{\eta}(\Gamma_{mr_d,\eta}) + log\mathbb{M}_{\eta_x}(\Gamma_{mr_d,\eta_x}) + log\mathbb{M}_{\eta_d}(\Gamma_{r_d,\eta_d})}{1-\kappa_{d,1}}
\end{aligned} \tag{80}
$$

As in the previous section, $\kappa_{d,0}$, $\kappa_{d,1}$ are functions of the endogenous expected dividend-price ratio z_d, which is the solution to the following non-linear equation

$$
0 = A_d(z_d) + A_{d,\sigma^2}(z_d)\sigma^2 + A_{d,\sigma_x^2}(z_d)\sigma_x^2 + A_{d,\sigma_d^2}(z_d)\sigma_d^2 - z_d \tag{81}
$$

D.3 Risk free rate and conditional expected fundamental market return

The risk-free rate is given by

$$R_{f,t} = \mathbb{E}_t(e^{m_{t+1}})^{-1} \Leftrightarrow r_{f,t} = -log(\mathbb{E}_t(e^{m_{t+1}})) \tag{82}$$

We have

$$\mathbb{E}_t(e^{m_{t+1}}) = B_m e^{\Gamma_{m,x}x_t + (\Gamma_{m,\sigma^2} + \frac{\Gamma^2_{m,\varepsilon}}{2})\sigma_t^2 + (\Gamma_{m,\sigma_x^2} + \frac{\Gamma^2_{m,\varepsilon_x}}{2})\sigma^2_{x,t}}$$

where

$$\begin{aligned} B_m &\equiv e^{\Gamma_m}\mathbb{M}_\eta(-\Gamma_{m,\eta})\mathbb{M}_{\eta_x}(-\Gamma_{m,\eta_x}) \\ &= e^{\Gamma_m} exp\Big\{\frac{\lambda}{(1-\varphi)\sigma^2}\Big(1 - \sqrt{1 + \frac{2(1-\varphi)^2\sigma^4\Gamma_{m,\eta}}{\lambda}}\Big)\Big\} exp\Big\{\frac{\lambda_x}{(1-\varphi_x)\sigma_x^2}\Big(1 - \sqrt{1 + \frac{2(1-\varphi_x)^2\sigma_x^4\Gamma_{m,\eta_x}}{\lambda_x}}\Big)\Big\} \end{aligned} \tag{83}$$

The log risk free rate is therefore

$$\begin{aligned} r_{f,t} &= -log(B_m) - \Gamma_{m,x}x_t - \Big(\Gamma_{m,\sigma^2} + \frac{\Gamma^2_{m,\varepsilon}}{2}\Big)\sigma_t^2 - \Big(\Gamma_{m,\sigma_x^2} + \frac{\Gamma^2_{m,\varepsilon_x}}{2}\Big)\sigma^2_{x,t} \\ &= -log(B_m) + \frac{1}{\psi}\rho_x x_t - \frac{1}{2}\Big(\frac{\tau}{\psi} + \Big[\frac{1}{\psi} - \gamma\Big](\gamma - 1) + \gamma^2\Big)\sigma_t^2 \\ &- \frac{1}{2}\Big(\frac{\tau_x}{\psi} + \Big[\frac{1}{\psi} - \gamma\Big]\frac{\gamma - 1}{(1-\kappa_{c,1}\varphi_x)^2} + \Big[\frac{\gamma - \frac{\kappa_{c,1}\rho_x}{\psi}}{1 - \kappa_{c,1}\rho_x}\Big]^2\Big)\sigma^2_{x,t} \end{aligned} \tag{84}$$

The risk free rate is thus increasing in expected consumption growth through x_t, $-\frac{\tau-1}{2}\sigma_t^2$ and $-\frac{\tau_x-1}{2}\sigma^2_{x,t}$. Note that if $\tau_i < 1$, there is a positive effect of increasing $\sigma_{i,t}$ on the expected consumption growth, which would in turn increase the interest rate. The total effect on the interest rate from an increase in variance depends both on the expected consumption growth effect and the precautionary savings effect - higher variance implies more precautionary savings, resulting in lower interest rates. If $\tau_i \geq 1$, i.e. expected consumption growth decreases in variance, the interest rate falls with higher variance.

$$\mathbb{E}_t(R_{d,t+1}) = \mathbb{E}_t(e^{r_{d,t+1}}) = B_{r_d} e^{\Gamma_{r_d,x}x_t + (\Gamma_{r_d,\sigma^2} + \frac{\Gamma^2_{r_d,\varepsilon}}{2})\sigma_t^2 + (\Gamma_{r_d,\sigma_x^2} + \frac{\Gamma^2_{r_d,\varepsilon_x}}{2})\sigma^2_{x,t}} \tag{85}$$

where

$$
\begin{aligned}
B_{r_d} &\equiv e^{\Gamma_{r_d}} \mathbb{M}_{\eta}(\Gamma_{r_d,\eta}) \mathbb{M}_{\eta_x}(\Gamma_{r_d,\eta_x}) \mathbb{M}_{\eta_d}(\Gamma_{r_d,\eta_d}) \\
&= e^{\Gamma_{r_d}} exp\Big\{\frac{\lambda}{(1-\varphi)\sigma^2}\Big(1-\sqrt{1-\frac{2(1-\varphi)^2\sigma^4\Gamma_{r_d,\eta}}{\lambda}}\Big)\Big\} exp\Big\{\frac{\lambda_x}{(1-\varphi_x)\sigma_x^2}\Big(1-\sqrt{1-\frac{2(1-\varphi_x)^2\sigma_x^4\Gamma_{r_d,\eta_x}}{\lambda_x}}\Big)\Big\} \\
&\times exp\Big\{\frac{\lambda_d}{(1-\varphi_d)\sigma_d^2}\Big(1-\sqrt{1-\frac{2(1-\varphi_d)^2\sigma_d^4\Gamma_{r_d,\eta_d}}{\lambda_d}}\Big)\Big\} \\
&= e^{-\Gamma_m} exp\Big\{\frac{\lambda}{(1-\varphi)\sigma^2}\Big(\sqrt{1-\frac{2(1-\varphi)^2\sigma^4\Gamma_{mr_d,\eta}}{\lambda}}-\sqrt{1-\frac{2(1-\varphi)^2\sigma^4\Gamma_{r_d,\eta}}{\lambda}}\Big)\Big\} \\
&\times exp\Big\{\frac{\lambda_x}{(1-\varphi_x)\sigma_x^2}\Big(\sqrt{1-\frac{2(1-\varphi_x)^2\sigma_x^4\Gamma_{mr_d,\eta_x}}{\lambda_x}}-\sqrt{1-\frac{2(1-\varphi_x)^2\sigma_x^4\Gamma_{r_d,\eta_x}}{\lambda_x}}\Big)\Big\}
\end{aligned} \tag{86}
$$

where the last equality follows from 80. As expected, the expected return on the fundamental market portfolio does not depend on idiosyncratic risk. The log risk premium can therefore be written

$$
\begin{aligned}
log\mathbb{E}_t(R_{d,t+1}) - r_{f,t} &= log(B_{r_d}) + log(B_m) + (\Gamma_{m,x}+\Gamma_{r_d,x})x_t + \Big(\Gamma_{m,\sigma^2}+\Gamma_{r_d,\sigma^2}+\frac{\Gamma_{m,\varepsilon}^2+\Gamma_{r_d,\varepsilon}^2}{2}\Big)\sigma_t^2 \\
&+ \Big(\Gamma_{m,\sigma_x^2}+\Gamma_{r_d,\sigma_x^2}+\frac{\Gamma_{m,\varepsilon_x}^2+\Gamma_{r_d,\varepsilon_x}^2}{2}\Big)\sigma_{x,t}^2 \\
&= log(B_{r_d}) + log(B_m) + \Gamma_{mr_d,x}x_t + \Big(\Gamma_{mr_d,\sigma^2}+\frac{\Gamma_{mr_d,\varepsilon}^2+2\Gamma_{m,\varepsilon}\Gamma_{r_d,\varepsilon}}{2}\Big)\sigma_t^2 \\
&+ \Big(\Gamma_{mr_d,\sigma_x^2}+\frac{\Gamma_{mr_d,\varepsilon_x}^2+2\Gamma_{m,\varepsilon_x}\Gamma_{r_d,\varepsilon_x}}{2}\Big)\sigma_{x,t}^2 \\
&= log(B_{r_d}) + log(B_m) + \Gamma_{m,\varepsilon}\Gamma_{r_d,\varepsilon}\sigma_t^2 + \Gamma_{m,\varepsilon_x}\Gamma_{r_d,\varepsilon_x}\sigma_{x,t}^2 \\
&= \alpha + \gamma\varrho\sigma_t^2 + \frac{\gamma-\frac{\kappa_{c,1}\rho_x}{\psi}}{1-\kappa_{c,1}\rho_x}(\varrho-\kappa_{d,1}A_{d,x})\sigma_{x,t}^2
\end{aligned} \tag{87}
$$

where

$$
\begin{aligned}
\alpha &\equiv log(B_{r_d}) + log(B_m) \\
&= \frac{\lambda}{(1-\varphi)\sigma^2}\left(1+\sqrt{1-\frac{2(1-\varphi)^2\sigma^4\Gamma_{mr_d,\eta}}{\lambda}}-\sqrt{1-\frac{2(1-\varphi)^2\sigma^4\Gamma_{r_d,\eta}}{\lambda}}-\sqrt{1+\frac{2(1-\varphi)^2\sigma^4\Gamma_{m,\eta}}{\lambda}}\right) \\
&+ \frac{\lambda_x}{(1-\varphi_x)\sigma_x^2}\left(1+\sqrt{1-\frac{2(1-\varphi_x)^2\sigma_x^4\Gamma_{mr_d,\eta_x}}{\lambda_x}}-\sqrt{1-\frac{2(1-\varphi_x)^2\sigma_x^4\Gamma_{r_d,\eta_x}}{\lambda_x}}-\sqrt{1+\frac{2(1-\varphi_x)^2\sigma_x^4\Gamma_{m,\eta_x}}{\lambda_x}}\right)
\end{aligned} \tag{88}
$$

Using the expression for $\Gamma_{m,x}$, we can write $A_{d,x}$ as

$$
A_{d,x} = -\frac{\Gamma_{m,x}+\varrho\rho_x}{1-\kappa_{d,1}\rho_x} = -\frac{-\frac{1}{\psi}\rho_x+\varrho\rho_x}{1-\kappa_{d,1}\rho_x} = -\frac{\varrho-\frac{1}{\psi}}{1-\kappa_{d,1}\rho_x}\rho_x \tag{89}
$$

$$
\tag{90}
$$

The expression for the log risk premium then becomes

$$log\mathbb{E}_t(R_{d,t+1}) - r_{f,t} = \alpha + \gamma\varrho\sigma_t^2 + \frac{\gamma - \frac{\kappa_{c,1}\rho_x}{\psi}}{1-\kappa_{c,1}\rho_x}\frac{\varrho - \frac{\kappa_{d,1}\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x}\sigma_{x,t}^2 \tag{91}$$

Consider the conditional variance of log returns

$$\begin{aligned}
\mathbb{V}_t(r_{d,t+1}) &= \delta_{r_d}^2 + \Gamma_{r,\varepsilon}^2\sigma_t^2 + \Gamma_{r,\varepsilon_x}^2\sigma_{x,t}^2 + \Gamma_{r,\varepsilon_d}^2\sigma_{d,t}^2 \\
&= \delta_{r_d}^2 + \varrho^2\sigma_t^2 + (\varrho - \kappa_{d,1}A_{d,x})^2\sigma_{x,t}^2 + \sigma_{d,t}^2 \\
&= \delta_{r_d}^2 + \varrho^2\sigma_t^2 + \Big(\frac{\varrho - \frac{\kappa_{d,1}\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x}\Big)^2\sigma_{x,t}^2 + \sigma_{d,t}^2 \qquad (92)
\end{aligned}$$

Regressing $log\mathbb{E}_t(R_{t+1}) - r_{f,t}$ on $\mathbb{V}_t(r_{d,t+1})$

$$log\mathbb{E}_t(R_{d,t+1}) - r_{f,t} = \phi_0 + \phi_1\mathbb{V}_t(r_{d,t+1}) + u_{t+1} \tag{93}$$

where

$$\begin{aligned}
\phi_1 &= \frac{Cov\Big(log(\mathbb{E}_t(R_{d,t+1})) - r_{f,t}, \mathbb{V}_t(r_{d,t+1})\Big)}{\mathbb{V}\Big(\mathbb{V}_t(r_{d,t+1})\Big)} \\
&= \frac{Cov\Big(\gamma\varrho\sigma_t^2 + \frac{\gamma-\frac{\kappa_{c,1}\rho_x}{\psi}}{1-\kappa_{c,1}\rho_x}\frac{\varrho-\frac{\kappa_{d,1}\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x}\sigma_{x,t}^2, \varrho^2\sigma_t^2 + \Big(\frac{\varrho-\frac{\kappa_{d,1}\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x}\Big)^2\sigma_{x,t}^2\Big)}{\mathbb{V}\Big(\varrho^2\sigma_t^2 + \Big[\frac{\varrho-\frac{\kappa_{d,1}\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x}\Big]^2\sigma_{x,t}^2 + \sigma_{d,t}^2\Big)} \\
&= \frac{\gamma\varrho^3\mathbb{V}(\sigma_t^2) + \frac{\gamma-\frac{\kappa_{c,1}\rho_x}{\psi}}{1-\kappa_{c,1}\rho_x}\Big(\frac{\varrho-\frac{\kappa_{d,1}\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x}\Big)^3\mathbb{V}(\sigma_{x,t}^2)}{\varrho^4\mathbb{V}(\sigma_t^2) + \Big(\frac{\varrho-\frac{\kappa_{d,1}\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x}\Big)^4\mathbb{V}(\sigma_{x,t}^2) + \mathbb{V}(\sigma_{d,t}^2)} \qquad (94)
\end{aligned}$$

We see that this regression coefficient is clearly positive if $\varrho,\ \gamma,\ \psi > 1$ as $\kappa_{c,1},\ \kappa_{d,1},\ \rho_x \in (0,1)$. Note that the regression coefficient will be positive even if some of the parameters do not satisfy the restrictions given here, as the first term in the numerator of ϕ_1 is clearly positive. We also see that increasing $\mathbb{V}(\sigma_{d,t}^2)$, the variance of the idosyncratic variance, lowers the magnitude of the regression coefficient, but does not alter its sign.

E Appendix E - Policy Pricing

In this appendix, we will find a partial analytical price for the governemnt policy. This will allow us to solve the rest via simulation without too much difficulty. It will prove useful to define the objects $\mathcal{F}_t$ as the information set at time t generated by $\{\varepsilon_s, \varepsilon_{x,s}, \varepsilon_{d,s}, \eta_s, \eta_{x,s}, \eta_{d,s}\}_{s=-\infty}^{t}$ and $\mathcal{H}_t$ as the information set generated by $\{\eta_s, \eta_{x,s}, \eta_{d,s}\}_{s=-\infty}^{t}$. Note the latter is a coarser (smaller) information set containing only information generated by the shocks to variance. I will use the following notation: $\mathbb{E}_t(X) \equiv \mathbb{E}(X|\mathcal{F}_t)$ and $\mathbb{E}_t(X|\mathcal{H}_s) \equiv \mathbb{E}(X|\mathcal{F}_t, \mathcal{H}_s)$, i.e. an expectation with a time subscript refers to the conditional expectation w.r.t. the full information set and use the explicit conditioning when taking conditional expectations w.r.t. the full information set augmented by the realization of variance shocks. Similarly, a function of the full information set at time t can be denoted as $f(\mathcal{F}_t)$, a function only of the variance shocks $f(\mathcal{H}_t)$, and a function of the ε_i shocks up to time t and variance shocks up to time T will be denoted as $f(\mathcal{F}_t, \mathcal{H}_T)$.

The following definitions will prove useful. For any random variable X let

$$\hat{X}_{T|T-s} \equiv \mathbb{E}_{T-s}(X_T|\mathcal{H}_T) \tag{95}$$

$$\tilde{X}_{T|T-s} \equiv X_T - \hat{X}_{T,T-s} \tag{96}$$

I will furthermore use the notation $\hat{X}$ for variables that are related to projections on the information set $\mathcal{H}$ and $\tilde{X}$ for variables that are related to the shock around this projection.

To begin with let the payout at time T to an option be given by

$$\begin{aligned} X_T &= e^{q_T(s)}(K_T(s) - P_T - D_T)^+ = e^{q_{T_i}(s)}(K_T(s) - P_{T-1}R_T)^+ \\ &= P_{T-1}e^{q_T(s)}(e^{-g_T(s)} - R_T)^+ \end{aligned} \tag{97}$$

I.e. this is a put option whose strike price is known at time T, but possibly dependent on things that happen between $T-s$ and T where $s \geq 0$. The (possibly stochastic) number $g_T(s)$ is defined as

$$g_T(s) \equiv log\Big(\frac{P_{T-1}}{K_T(s)}\Big) \tag{98}$$

which is known at time T.

E.1 Policy Representation

We will assume the following general structure for $(q_T(s),\ g_T(s))$

$$
\begin{aligned}
q_T(s) = \bar{q}_0 + \sum_{j=0}^{s} \Big(& \bar{q}_{x,j} x_{T-j} + \bar{q}_{\sigma^2,j} \sigma^2_{T-j} + \bar{q}_{\sigma^2_x,j} \sigma^2_{x,T-j} + \bar{q}_{\sigma^2_d,j} \sigma^2_{d,T-j} + \bar{q}_{\varepsilon,j} \sigma_{T-j-1} \varepsilon_{T-j} \\
& + \bar{q}_{\varepsilon_x,j} \sigma_{x,T-j-1} \varepsilon_{x,T-j} + \bar{q}_{\varepsilon_d,j} \sigma_{d,T-j-1} \varepsilon_{d,T-j} + \bar{q}_{\eta,j} \eta_{T-j} + \bar{q}_{\eta_x,j} \eta_{x,T-j} + \bar{q}_{\eta_d,j} \eta_{d,T-j} \Big) \qquad (99)
\end{aligned}
$$

$$
\begin{aligned}
g_T(s) = \bar{g}_0 + \sum_{j=0}^{s} \Big(& \bar{g}_{x,j} x_{T-j} + \bar{g}_{\sigma^2,j} \sigma^2_{T-j} + \bar{g}_{\sigma^2_x,j} \sigma^2_{x,T-j} + \bar{g}_{\sigma^2_d,j} \sigma^2_{d,T-j} + \bar{g}_{\varepsilon,j} \sigma_{T-j-1} \varepsilon_{T-j} \\
& + \bar{g}_{\varepsilon_x,j} \sigma_{x,T-j-1} \varepsilon_{x,T-j} + \bar{g}_{\varepsilon_d,j} \sigma_{d,T-j-1} \varepsilon_{d,T-j} + \bar{g}_{\eta,j} \eta_{T-j} + \bar{g}_{\eta_x,j} \eta_{x,T-j} + \bar{g}_{\eta_d,j} \eta_{d,T-j} \Big) \qquad (100)
\end{aligned}
$$

Let $h \equiv T - t$ and $\bar{q}_{\cdot,j} = \bar{g}_{\cdot,j} = 0\ \forall j > s$. If $h \leq s$ we can divide $(q_T(s),\ g_T(s))$ into history and future. First, $g_T(s)$

$$
\begin{aligned}
g_T(s) = \bar{g}_0 + \sum_{j=0}^{h-1} \Big(& \bar{g}_{x,j} x_{T-j} + \bar{g}_{\sigma^2,j} \sigma^2_{T-j} + \bar{g}_{\sigma^2_x,j} \sigma^2_{x,T-j} + \bar{g}_{\sigma^2_d,j} \sigma^2_{d,T-j} + \bar{g}_{\varepsilon,j} \sigma_{T-j-1} \varepsilon_{T-j} \\
& + \bar{g}_{\varepsilon_x,j} \sigma_{x,T-j-1} \varepsilon_{x,T-j} + \bar{g}_{\varepsilon_d,j} \sigma_{d,T-j-1} \varepsilon_{d,T-j} + \bar{g}_{\eta,j} \eta_{T-j} + \bar{g}_{\eta_x,j} \eta_{x,T-j} + \bar{g}_{\eta_d,j} \eta_{d,T-j} \Big) \\
& + \sum_{j=h}^{s} \Big(\bar{g}_{x,j} x_{T-j} + \bar{g}_{\sigma^2,j} \sigma^2_{T-j} + \bar{g}_{\sigma^2_x,j} \sigma^2_{x,T-j} + \bar{g}_{\sigma^2_d,j} \sigma^2_{d,T-j} + \bar{g}_{\varepsilon,j} \sigma_{T-j-1} \varepsilon_{T-j} \\
& + \bar{g}_{\varepsilon_x,j} \sigma_{x,T-j-1} \varepsilon_{x,T-j} + \bar{g}_{\varepsilon_d,j} \sigma_{d,T-j-1} \varepsilon_{d,T-j} + \bar{g}_{\eta,j} \eta_{T-j} + \bar{g}_{\eta_x,j} \eta_{x,T-j} + \bar{g}_{\eta_d,j} \eta_{d,T-j} \Big) \\
& = \bar{g}_0 + W_{g,h,t} + \sum_{j=0}^{h-1} \Big(\bar{g}_{x,j} x_{T-j} + \bar{g}_{\sigma^2,j} \sigma^2_{T-j} + \bar{g}_{\sigma^2_x,j} \sigma^2_{x,T-j} + \bar{g}_{\sigma^2_d,j} \sigma^2_{d,T-j} + \bar{g}_{\varepsilon,j} \sigma_{T-j-1} \varepsilon_{T-j} \\
& + \bar{g}_{\varepsilon_x,j} \sigma_{x,T-j-1} \varepsilon_{x,T-j} + \bar{g}_{\varepsilon_d,j} \sigma_{d,T-j-1} \varepsilon_{d,T-j} + \bar{g}_{\eta,j} \eta_{T-j} + \bar{g}_{\eta_x,j} \eta_{x,T-j} + \bar{g}_{\eta_d,j} \eta_{d,T-j} \Big) \qquad (101)
\end{aligned}
$$

where

$$
\begin{aligned}
W_{g,h,t} \equiv \sum_{j=h}^{s} \Big(& \bar{g}_{x,j} x_{T-j} + \bar{g}_{\sigma^2,j} \sigma^2_{T-j} + \bar{g}_{\sigma^2_x,j} \sigma^2_{x,T-j} + \bar{g}_{\sigma^2_d,j} \sigma^2_{d,T-j} + \bar{g}_{\varepsilon,j} \sigma_{T-j-1} \varepsilon_{T-j} \\
& + \bar{g}_{\varepsilon_x,j} \sigma_{x,T-j-1} \varepsilon_{x,T-j} + \bar{g}_{\varepsilon_d,j} \sigma_{d,T-j-1} \varepsilon_{d,T-j} + \bar{g}_{\eta,j} \eta_{T-j} + \bar{g}_{\eta_x,j} \eta_{x,T-j} + \bar{g}_{\eta_d,j} \eta_{d,T-j} \Big) \qquad (102)
\end{aligned}
$$

Use the convention that $W_{g,h,t} = 0$ when $h > s$. We can then rewrite $g_T(s)$ for general h as follows

$$
\begin{aligned}
g_T(s) = g_0 + W_{g,h,t} + g_{x,h-1} x_t + g_{\sigma^2,h-1} \sigma^2_t + g_{\sigma^2_x,h-1} \sigma^2_{x,t} + g_{\sigma^2_d,h-1} \sigma^2_{d,t} + \sum_{j=0}^{h-1} \Big(& g_{\varepsilon,j} \sigma_{T-j-1} \varepsilon_{T-j} \\
+ g_{\varepsilon_x,j} \sigma_{x,T-j-1} \varepsilon_{x,T-j} + g_{\varepsilon_d,j} \sigma_{d,T-j-1} \varepsilon_{d,T-j} + g_{\eta,j} \eta_{T-j} + g_{\eta_x,j} \eta_{x,T-j} + g_{\eta_d,j} \eta_{d,T-j} \Big) & \qquad (103)
\end{aligned}
$$

where

$$
\begin{aligned}
g_0 &\equiv \bar{g}_0 \\
g_{x,j} &\equiv \sum_{l=0}^{j} \bar{g}_{x,l}\rho_x^{j-l} \\
g_{\sigma^2,j} &\equiv \sum_{l=0}^{j} \bar{g}_{\sigma^2,l}\varphi^{j-l} \\
g_{\sigma_x^2,j} &\equiv \sum_{l=0}^{j} \bar{g}_{\sigma_x^2,l}\varphi_x^{j-l} \\
g_{\sigma_d^2,j} &\equiv \sum_{l=0}^{j} \bar{g}_{\sigma_d^2,l}\varphi_d^{j-l} \\
g_{\varepsilon,j} &\equiv \bar{g}_{\varepsilon,j} \\
g_{\varepsilon_x,j} &\equiv g_{x,j} + \bar{g}_{\varepsilon_x,j} \\
g_{\varepsilon_d,j} &\equiv \bar{g}_{\varepsilon_d,j} \\
g_{\eta,j} &\equiv g_{\sigma^2,j} + \bar{g}_{\eta,j} \\
g_{\eta_x,j} &\equiv g_{\sigma^2,j} + \bar{g}_{\eta_x,j} \\
g_{\eta_d,j} &\equiv g_{\sigma^2,j} + \bar{g}_{\eta_d,j}
\end{aligned}
$$

We can then divide $g_T(s)$ into the following two parts

$$
\begin{aligned}
\hat{g}_{T|t}(s) = g_0 &+ W_{g,h,t} + g_{x,h-1}x_t + g_{\sigma^2,h-1}\sigma_t^2 + g_{\sigma_x^2,h-1}\sigma_{x,t}^2 + g_{\sigma_d^2,h-1}\sigma_{d,t}^2 \\
&+ \sum_{j=0}^{h-1}\Big(g_{\eta,j}\eta_{T-j} + g_{\eta_x,j}\eta_{x,T-j} + g_{\eta_d,j}\eta_{d,T-j}\Big)
\end{aligned} \tag{104}
$$

and

$$
\tilde{g}_{T|t}(s) = \sum_{j=0}^{h-1}\Big(g_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j} + g_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j} + g_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j}\Big) \tag{105}
$$

Note that we can $q_T(s)$ in exactly the same way. Thus,

$$
\begin{aligned}
q_T(s) = q_0 + W_{q,h,t} + q_{x,h-1}x_t + q_{\sigma^2,h-1}\sigma_t^2 + q_{\sigma_x^2,h-1}\sigma_{x,t}^2 + q_{\sigma_d^2,h-1}\sigma_{d,t}^2 + \sum_{j=0}^{h-1}\Big(q_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j} \\
+ q_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j} + q_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j} + q_{\eta,j}\eta_{T-j} + q_{\eta_x,j}\eta_{x,T-j} + q_{\eta_d,j}\eta_{d,T-j}\Big)
\end{aligned} \tag{106}
$$

where

$$W_{q,h,t} \equiv \sum_{j=h}^{s} \Big(\bar{q}_{x,j} x_{T-j} + \bar{q}_{\sigma^2,j} \sigma^2_{T-j} + \bar{q}_{\sigma^2_x,j} \sigma^2_{x,T-j} + \bar{q}_{\sigma^2_d,j} \sigma^2_{d,T-j} + \bar{q}_{\varepsilon,j} \sigma_{T-j-1} \varepsilon_{T-j}$$

$$+ \bar{q}_{\varepsilon_x,j} \sigma_{x,T-j-1} \varepsilon_{x,T-j} + \bar{q}_{\varepsilon_d,j} \sigma_{d,T-j-1} \varepsilon_{d,T-j} + \bar{q}_{\eta,j} \eta_{T-j} + \bar{q}_{\eta_x,j} \eta_{x,T-j} + \bar{q}_{\eta_d,j} \eta_{d,T-j} \Big) \qquad (107)$$

and

$$\begin{aligned}
q_0 &\equiv \bar{g}_0 \\
q_{x,j} &\equiv \sum_{l=0}^{j} \bar{g}_{x,l} \rho_x^{j-l} \\
q_{\sigma^2,j} &\equiv \sum_{l=0}^{j} \bar{g}_{\sigma^2,l} \varphi^{j-l} \\
q_{\sigma^2_x,j} &\equiv \sum_{l=0}^{j} \bar{g}_{\sigma^2_x,l} \varphi_x^{j-l} \\
q_{\sigma^2_d,j} &\equiv \sum_{l=0}^{j} \bar{g}_{\sigma^2_d,l} \varphi_d^{j-l} \\
q_{\varepsilon,j} &\equiv \bar{q}_{\varepsilon,j} \\
q_{\varepsilon_x,j} &\equiv q_{x,j} + \bar{q}_{\varepsilon_x,j} \\
q_{\varepsilon_d,j} &\equiv \bar{q}_{\varepsilon_d,j} \\
q_{\eta,j} &\equiv q_{\sigma^2,j} + \bar{q}_{\eta,j} \\
q_{\eta_x,j} &\equiv q_{\sigma^2_x,j} + \bar{q}_{\eta_x,j} \\
q_{\eta_d,j} &\equiv q_{\sigma^2_d,j} + \bar{q}_{\eta_d,j}
\end{aligned}$$

We can also divide $q_T(s)$ into two parts

$$\hat{q}_{T|t}(s) = q_0 + W_{q,h,t} + g_{q,h-1} x_t + q_{\sigma^2,h-1} \sigma^2_t + q_{\sigma^2_x,h-1} \sigma^2_{x,t} + q_{\sigma^2_d,h-1} \sigma^2_{d,t}$$

$$+ \sum_{j=0}^{h-1} \Big(q_{\eta,j} \eta_{T-j} + q_{\eta_x,j} \eta_{x,T-j} + q_{\eta_d,j} \eta_{d,T-j} \Big) \qquad (108)$$

and

$$\tilde{q}_{T|t}(s) = \sum_{j=0}^{h-1} \Big(q_{\varepsilon,j} \sigma_{T-j-1} \varepsilon_{T-j} + q_{\varepsilon_x,j} \sigma_{x,T-j-1} \varepsilon_{x,T-j} + q_{\varepsilon_d,j} \sigma_{d,T-j-1} \varepsilon_{d,T-j} \Big) \qquad (109)$$

E.2 Intermediary Expressions

Let

$$
\begin{aligned}
y_{t,T} &\equiv \sum_{j=1}^{h} m_{t+j} + p_{T-1} - p_t + q_T(s) - g_T(s) \\
&= \hat{m}p_{T|t} + \hat{q}_{T|t}(s) - \hat{g}_{T|t}(s) + \sum_{j=0}^{h-1} L_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j} + L_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j} + L_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j}
\end{aligned}
\tag{110}
$$

where

$$
\begin{aligned}
\hat{m}p_{T|t} &\equiv \mathbb{E}_t\Big(\sum_{j=1}^{\Delta_T} m_{t+j} + p_{T-1} - p_t \Big| \mathcal{H}_T\Big) \\
L_{\varepsilon,0} &\equiv q_{\varepsilon,0} - g_{\varepsilon,0} - \Gamma_{m,\varepsilon} \\
L_{\varepsilon_x,0} &\equiv q_{\varepsilon_x,0} - g_{\varepsilon_x,0} - \Gamma_{m,\varepsilon_x} \\
L_{\varepsilon_d,0} &\equiv q_{\varepsilon_d,0} - g_{\varepsilon_d,0} \\
L_{\varepsilon,j} &\equiv \varrho + q_{\varepsilon,j} - g_{\varepsilon,j} - \Gamma_{m,\varepsilon} \\
L_{\varepsilon_x,j} &\equiv (\varrho + \Gamma_{m,x})\frac{1-\rho_x^j}{1-\rho_x} - A_x\rho_x^{j-1} + q_{\varepsilon_x,j} - g_{\varepsilon_x,j} - \Gamma_{m,\varepsilon_x} \\
L_{\varepsilon_d,j} &\equiv 1 + q_{\varepsilon_d,j} - g_{\varepsilon_d,j} - \Gamma_{m,\varepsilon_d}
\end{aligned}
$$

Furthermore, let

$$
g_T(s) + r_{d,T} = \hat{g}_{T|t}(s) + \hat{r}_{d,T|t} + \sum_{j=0}^{h-1} K_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j} + K_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j} + K_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j}
\tag{111}
$$

where

$$
\begin{aligned}
K_{\varepsilon,0} &\equiv \Gamma_{r_d,\varepsilon} + g_{\varepsilon,0} \\
K_{\varepsilon_x,0} &\equiv \Gamma_{r_d,\varepsilon_x} + g_{\varepsilon_x,0} \\
K_{\varepsilon_d,0} &\equiv \Gamma_{r_d,\varepsilon_d} + g_{\varepsilon_d,0} \\
K_{\varepsilon,j} &\equiv g_{\varepsilon,j} \\
K_{\varepsilon_x,j} &\equiv g_{\varepsilon_x,j} + \Gamma_{r_d,x}\rho_x^{j-1} \\
K_{\varepsilon_d,j} &\equiv g_{\varepsilon_d,j}
\end{aligned}
$$

We can then write the claim payoff multiplied with the h-period pricing kernel as follows

$$
\begin{aligned}
M_{t,T}X_T &= M_{t,T}P_{T-1}e^{q_T(s)-g_T(s)}\Big(1-e^{g_T(s)+r_{d,T}}\Big)^+ = P_t e^{y_{t,T}}\Big(1-e^{g_T(s)+r_{d,T}}\Big)^+ \\
&= P_t e^{\hat{mp}_{T|t}+\hat{q}_{T|t}(s)-\hat{g}_{T|t}(s)+\sum_{j=0}^{h-1}L_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j}+L_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j}+L_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j}} \\
&\times\Big(1-e^{\hat{g}_{T|t}(s)+\hat{r}_{d,T|t}+\sum_{j=0}^{h-1}K_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j}+K_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j}+K_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j}}\Big)^+ \qquad (112)
\end{aligned}
$$

E.3 Proof of Proposition 3

Proof. The time t price of the claim that expires at time $T=t+h$ is

$$P_t^X=\mathbb{E}_t(M_{t,T}X_T)=\mathbb{E}_t\Big(\mathbb{E}_t(M_{t,T}X_T|\mathcal{H}_T)\Big) \qquad (113)$$

Assume at least one of $K_{\varepsilon,l}$, $K_{\varepsilon_x,l}$, or $K_{\varepsilon_d,l}$ is different from 0 for some $l<h$. Denote one such $K_{\cdot,l}$ as $K_{\varepsilon_i,j}$. Assume first that $K_{\varepsilon_i,j}>0$. Using equation (112), we see that the policy pays off iff

$$
\begin{aligned}
0 &\geq \hat{r}_{d,T|t}+\hat{g}_{T|t}(s)+\sum_{j=0}^{h-1}K_{\varepsilon,j}\sigma_{T-1-j}\varepsilon_{T-j}+K_{\varepsilon_x,j}\sigma_{x,T-1-j}\varepsilon_{x,T-j}+K_{\varepsilon_d,j}\sigma_{d,T-1-j}\varepsilon_{d,T-j} \\
\varepsilon_{i,T-l} &\leq -\frac{\hat{r}_{d,T|t}+\hat{g}_{T|t}(s)+\sum_{j=0}^{h-1}\bar{K}_{\varepsilon,j}\sigma_{T-1-j}\varepsilon_{T-j}+\bar{K}_{\varepsilon_x,j}\sigma_{x,T-1-j}\varepsilon_{x,T-j}+\bar{K}_{\varepsilon_d,j}\sigma_{d,T-1-j}\varepsilon_{d,T-j}}{K_{\varepsilon_i,l}\sigma_{i,T-1-j}}\equiv a_T
\end{aligned}
\qquad (114)
$$

with $\bar{K}_{\varepsilon_k,j}\equiv K_{\varepsilon_k,j}$ except for $k=i$ and $j=l$ where $\bar{K}_{\varepsilon_i,l}\equiv 0$. Similarly, if $K_{\varepsilon_i,l}<0$ we have

$$\varepsilon_{i,T-l}\geq a_T \qquad (115)$$

Note that a_T is *not* a function of $\varepsilon_{i,T-l}$ in either case.

Using equation (112) and the conditions in (114) and (115) we get

$$
\begin{aligned}
\mathbb{E}_t(M_{t,T}X_T|\mathcal{H}_T) &= P_t\mathbb{E}_t\Big[e^{\hat{mp}_{T|t}+\hat{q}_{T|t}(s)-\hat{g}_{T|t}(s)+\sum_{j=0}^{h-1}L_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j}+L_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j}+L_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j}} \\
&\times\Big(1-e^{\hat{g}_{T|t}(s)+\hat{r}_{d,T|t}+\sum_{j=0}^{h-1}K_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j}+K_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j}+K_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j}}\Big)\mathbb{I}_{\varepsilon_{i,T-l}\leq a_T}\Big|\mathcal{H}_T\Big]
\end{aligned}
\qquad (116)
$$

$$
\begin{aligned}
\mathbb{E}_t(M_{t,T}X_T|\mathcal{H}_T) &= P_t\mathbb{E}_t\Big[e^{\hat{mp}_{T|t}+\hat{q}_{T|t}(s)-\hat{g}_{T|t}(s)+\sum_{j=0}^{h-1}L_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j}+L_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j}+L_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j}} \\
&\times\Big(1-e^{\hat{g}_{T|t}(s)+\hat{r}_{d,T|t}+\sum_{j=0}^{h-1}K_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j}+K_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j}+K_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j}}\Big)\mathbb{I}_{\varepsilon_{i,T-l}\geq a_T}\Big|\mathcal{H}_T\Big]
\end{aligned}
\qquad (117)
$$

Using the integration results in equations (134) and (136) in Appendix F gives us

$$\mathbb{E}_t(M_{t,T}X_T|\mathcal{H}_T) = P_t\Bigg[e^{\hat{mp}_{T|t}+\hat{q}_{T|t}(s)-\hat{g}_{T|t}(s)+\frac{1}{2}\sum_{j=0}^{h-1}L_{\varepsilon,j}^2\sigma_{T-j-1}^2+L_{\varepsilon_x,j}^2\sigma_{x,T-j-1}^2+L_{\varepsilon_d,j}^2\sigma_{d,T-j-1}^2}$$
$$\Phi\Bigg(-\frac{\hat{g}_{T|t}(s)+\hat{r}_{d,T|t}+\sum_{j=0}^{h-1}L_{\varepsilon,j}K_{\varepsilon,j}\sigma_{T-j-1}^2+L_{\varepsilon_x,j}K_{\varepsilon_x,j}\sigma_{x,T-j-1}^2+L_{\varepsilon_d,j}K_{\varepsilon_d,j}\sigma_{d,T-j-1}^2}{\sqrt{\sum_{j=0}^{h-1}K_{\varepsilon,j}^2\sigma_{T-j-1}^2+K_{\varepsilon_x,j}^2\sigma_{x,T-j-1}^2+K_{\varepsilon_d,j}^2\sigma_{d,T-j-1}^2}}\Bigg)$$
$$-e^{\hat{mp}_{T|t}+\hat{q}_{T|t}(s)+\hat{r}_{d,T|t}+\frac{1}{2}\sum_{j=0}^{h-1}(L_{\varepsilon,j}+K_{\varepsilon,j})^2\sigma_{T-j-1}^2+(L_{\varepsilon_x,j}+K_{\varepsilon_x,j})^2\sigma_{x,T-j-1}^2+(L_{\varepsilon_d,j}+K_{\varepsilon_d,j})^2\sigma_{d,T-j-1}^2}$$
$$\Phi\Bigg(-\frac{\hat{g}_{T|t}(s)+\hat{r}_{d,T|t}+\sum_{j=0}^{h-1}(L_{\varepsilon,j}+K_{\varepsilon,j})K_{\varepsilon,j}\sigma_{T-j-1}^2+(L_{\varepsilon_x,j}+K_{\varepsilon_x,j})K_{\varepsilon_x,j}\sigma_{x,T-j-1}^2+(L_{\varepsilon_d,j}+K_{\varepsilon_d,j})K_{\varepsilon_d,j}\sigma_{d,T-j-1}^2}{\sqrt{\sum_{j=0}^{h-1}K_{\varepsilon,j}^2\sigma_{T-j-1}^2+K_{\varepsilon_x,j}^2\sigma_{x,T-j-1}^2+K_{\varepsilon_d,j}^2\sigma_{d,T-j-1}^2}}\Bigg)\Bigg] \tag{118}$$

Defining

$$\zeta_{y,t,T-1} \equiv \mathbb{V}_t\big(y_{t,T}\big|\mathcal{H}_T\big) = \sum_{j=0}^{h-1}L_{\varepsilon,j}^2\sigma_{T-1-j}^2+L_{\varepsilon_x,j}^2\sigma_{x,T-1-j}^2+L_{\varepsilon_d,j}^2\sigma_{d,T-1-j}^2 \tag{119}$$

$$\zeta_{r_d+g,t,T-1} \equiv \mathbb{V}_t\big(r_{d,T}+g_T(s)\big|\mathcal{H}_T\big) = \sum_{j=0}^{h-1}K_{\varepsilon,j}^2\sigma_{T-1-j}^2+K_{\varepsilon_x,j}^2\sigma_{x,T-1-j}^2+K_{\varepsilon_d,j}^2\sigma_{d,T-1-j}^2 \tag{120}$$

$$\begin{aligned}\zeta_{y,r_d+g,t,T-1} &\equiv Cov_t\big(y_{t,T},r_{d,T}+g_T(s)\big|\mathcal{H}_T\big)\\ &= \sum_{j=0}^{h-1}L_{\varepsilon,j}K_{\varepsilon,j}\sigma_{T-1-j}^2+L_{\varepsilon_x,j}K_{\varepsilon_x,j}\sigma_{x,T-1-j}^2+L_{\varepsilon_d,j}K_{\varepsilon_d,j}\sigma_{d,T-1-j}^2\end{aligned} \tag{121}$$

and taking the expectation conditional on $\mathcal{F}_t$ gives us the result

$$P_{h,t}^X = P_t\mathbb{E}_t\Bigg[e^{\hat{mp}_{T|t}+\hat{q}_{T|t}(s)-\hat{g}_{T|t}(s)+\frac{\zeta_{y,t,T-1}}{2}}\Bigg(\Phi\Big(-\frac{\hat{g}_{T-1|t}(s)+\hat{r}_{d,T|t}+\zeta_{y,r_d+g,t,T-1}}{\sqrt{\zeta_{r_d+g,t,T-1}}}\Big)$$
$$-e^{\hat{g}_{T|t}(s)+\hat{r}_{d,T|t}+\zeta_{y,r_d+g,t,T-1}+\frac{\zeta_{r_d+g,t,T-1}}{2}}\Phi\Big(-\frac{\hat{g}_{T|t}(s)+\hat{r}_{d,T|t}+\zeta_{r_d+g,t,T-1}}{\sqrt{\zeta_{r_d+g,t,T-1}}}-\sqrt{\zeta_{r_d+g,t,T-1}}\Big)\Bigg)\Bigg] \tag{122}$$

■

E.4 Proof of Corollary 2

Proof. Assume $K_{\varepsilon,l} = K_{\varepsilon_x,l} = K_{\varepsilon_d,l} = 0$ for all $l < h$. Using equation (112), we see that the policy pays off iff

$$0 \geq \hat{r}_{d,T|t} + \hat{g}_{T|t}(s) \tag{123}$$

Using equations (112) and (123), the time t price of the claim that expires at time $T = t + h$ is

$$\begin{aligned}
P_t^X &= \mathbb{E}_t(M_{t,T}X_T) = \mathbb{E}_t\Big(\mathbb{E}_t(M_{t,T}X_T|\mathcal{H}_T)\Big) \\
&= P_t\mathbb{E}_t\Bigg[\mathbb{E}_t\Bigg(e^{\hat{mp}_{T|t}+\hat{q}_{T|t}(s)-\hat{g}_{T|t}(s)+\sum_{j=0}^{h-1}L_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j}+L_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j}+L_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j}} \\
&\times\Big(1-e^{\hat{g}_{T|t}(s)+\hat{r}_{d,T|t}}\Big)\mathbb{I}_{0\geq\hat{r}_{d,T|t}+\hat{g}_{T|t}(s)}\Bigg|\mathcal{H}_T\Bigg)\Bigg] \\
&= P_t\mathbb{E}_t\Bigg[\mathbb{E}_t\Bigg(e^{\hat{mp}_{T|t}+\hat{q}_{T|t}(s)-\hat{g}_{T|t}(s)+\sum_{j=0}^{h-1}L_{\varepsilon,j}\sigma_{T-j-1}\varepsilon_{T-j}+L_{\varepsilon_x,j}\sigma_{x,T-j-1}\varepsilon_{x,T-j}+L_{\varepsilon_d,j}\sigma_{d,T-j-1}\varepsilon_{d,T-j}}\Bigg|\mathcal{H}_T\Bigg) \\
&\times\Big(1-e^{\hat{g}_{T|t}(s)+\hat{r}_{d,T|t}}\Big)\mathbb{I}_{0\geq\hat{r}_{d,T|t}+\hat{g}_{T|t}(s)}\Bigg] \\
&= P_t\mathbb{E}_t\Bigg[e^{\hat{mp}_{T|t}+\hat{q}_{T|t}(s)-\hat{g}_{T|t}(s)+\frac{1}{2}\sum_{j=0}^{h-1}L_{\varepsilon,j}^2\sigma_{T-j-1}^2+L_{\varepsilon_x,j}^2\sigma_{x,T-j-1}^2+L_{\varepsilon_d,j}^2\sigma_{d,T-j-1}^2} \\
&\times\Big(1-e^{\hat{g}_{T|t}(s)+\hat{r}_{d,T|t}}\Big)\mathbb{I}_{0\geq\hat{r}_{d,T|t}+\hat{g}_{T|t}(s)}\Bigg] \qquad (124)
\end{aligned}$$

■

F Integration results

The following integration result will be useful

$$\int_{-\infty}^{\infty}\Phi(a+B\varepsilon)\phi(\varepsilon)d\varepsilon = \Phi\Big(\frac{a}{\sqrt{1+B^2}}\Big) \qquad (125)$$

As a result

$$\begin{aligned}
\int_{-\infty}^{\infty}e^{c+L\varepsilon}\Phi(a+K\varepsilon)\phi(\varepsilon)d\varepsilon &= e^c\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{L\varepsilon-\frac{\varepsilon^2}{2}}\Phi(a+B\varepsilon)d\varepsilon \\
&= e^{c+\frac{L^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(\varepsilon-L)^2}{2}}\Phi(a+B\varepsilon)d\varepsilon \\
&= e^{c+\frac{L^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\Phi\big(a+B(y+L)\big)dy \\
&= e^{c+\frac{L^2}{2}}\int_{-\infty}^{\infty}\Phi\big((a+BL)+By\big)\phi(y)dy \\
&= e^{c+\frac{L^2}{2}}\Phi\Big(\frac{a+BL}{\sqrt{1+B^2}}\Big) \qquad (126)
\end{aligned}$$

Let

$$c_n \equiv c+\sum_{j=1}^{n}L_j\varepsilon_j \qquad (127)$$

$$a_n \equiv a+\sum_{j=1}^{n}B_j\varepsilon_j \qquad (128)$$

We can then generalize to multiple integrals

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} e^{c+\sum_{j=1}^{n}L_j\varepsilon_j}\Phi\Big(a+\sum_{j=1}^{n}B_j\varepsilon_j\Big)\phi(\varepsilon_n)\dots\phi(\varepsilon_1)d\varepsilon_n\dots d\varepsilon_1\\
&\quad=\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{c_{n-1}+L_n\varepsilon_n}\Phi\Big(a_{n-1}+B_n\varepsilon_n\Big)\phi(\varepsilon_n)d\varepsilon_n\right)\phi(\varepsilon_{n-1})\dots\phi(\varepsilon_1)d\varepsilon_{n-1}\dots d\varepsilon_1\\
&\quad=\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} e^{c_{n-1}+\frac{L_n^2}{2}}\Phi\Big(\frac{a_{n-1}+B_nL_n}{\sqrt{1+K_n^2}}\Big)\phi(\varepsilon_{n-1})\dots\phi(\varepsilon_1)d\varepsilon_{n-1}\dots d\varepsilon_1\\
&\quad=e^{\frac{L_n^2}{2}}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\\
&\quad\left(\int_{-\infty}^{\infty} e^{c_{n-2}+L_{n-1}\varepsilon_{n-1}}\Phi\Big(\frac{a_{n-2}+B_nL_n}{\sqrt{1+B_n^2}}+\frac{B_{n-1}}{\sqrt{1+B_n^2}}\varepsilon_{n-1}\Big)\phi(\varepsilon_{n-1})d\varepsilon_{n-1}\right)\phi(\varepsilon_{n-2})\dots\phi(\varepsilon_1)d\varepsilon_{n-2}\dots d\varepsilon_1\\
&\quad=e^{\frac{L_n^2}{2}}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} e^{c_{n-2}+\frac{L_{n-1}^2}{2}}\Phi\left(\frac{\frac{a_{n-2}+B_nL_n+B_{n-1}L_{n-1}}{\sqrt{1+B_n^2}}}{\sqrt{1+\Big(\frac{B_{n-1}}{\sqrt{1+B_n^2}}\Big)^2}}\right)\phi(\varepsilon_{n-2})\dots\phi(\varepsilon_1)d\varepsilon_{n-2}\dots d\varepsilon_1\\
&\quad=e^{\frac{L_n^2+L_{n-1}^2}{2}}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} e^{c_{n-2}}\Phi\left(\frac{a_{n-2}+B_nL_n+B_{n-1}L_{n-1}}{\sqrt{1+B_n^2+B_{n-1}^2}}\right)\phi(\varepsilon_{n-2})\dots\phi(\varepsilon_1)d\varepsilon_{n-2}\dots d\varepsilon_1\\
&\quad\vdots\\
&\quad=e^{c+\frac{1}{2}\sum_{j=1}^{n}L_j^2}\Phi\left(\frac{a+\sum_{j=1}^{n}B_jL_j}{\sqrt{1+\sum_{j=1}^{n}B_j^2}}\right)
\end{aligned}
\tag{129}
$$

The final result we will need in order to solve for the option price is

$$
\begin{aligned}
\int_{-\infty}^{a} e^{c+L\varepsilon}\phi(\varepsilon)d\varepsilon&=e^{c}\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{L\varepsilon-\frac{\varepsilon^2}{2}}d\varepsilon=e^{c+\frac{L^2}{2}}\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{\frac{-(\varepsilon-L)^2}{2}}d\varepsilon=e^{c+\frac{L^2}{2}}\int_{-\infty}^{a-L}\frac{1}{\sqrt{2\pi}}e^{\frac{-y^2}{2}}dy\\
&=e^{c+\frac{L^2}{2}}\Phi(a-L)
\end{aligned}
\tag{130}
$$

Using the last two results together, gives us

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\int_{-\infty}^{a_n} e^{c+\sum_{j=1}^{n+1}L_j\varepsilon_j}\phi(\varepsilon_{n+1})\phi(\varepsilon_n)\dots\phi(\varepsilon_1)d\varepsilon_{n+1}d\varepsilon_n\dots d\varepsilon_1\\
&\quad=\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} e^{\sum_{j=1}^{n}L_j\varepsilon_j}\left(\int_{-\infty}^{a_n} e^{c+L_{n+1}\varepsilon_{n+1}}\phi(\varepsilon_{n+1})d\varepsilon_{n+1}\right)\phi(\varepsilon_n)\dots\phi(\varepsilon_1)d\varepsilon_n\dots d\varepsilon_1\\
&\quad=\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} e^{\sum_{j=1}^{n}L_j\varepsilon_j}e^{c+\frac{L_{n+1}^2}{2}}\Phi\Big((a-L_{n+1})+\sum_{j=1}^{n}B_j\varepsilon_j\Big)\phi(\varepsilon_n)\dots\phi(\varepsilon_1)d\varepsilon_n\dots d\varepsilon_1\\
&\quad=e^{c+\frac{1}{2}\sum_{j=1}^{n+1}L_j^2}\Phi\left(\frac{a-L_{n+1}+\sum_{j=1}^{n}B_jL_j}{\sqrt{1+\sum_{j=1}^{n}B_j^2}}\right)
\end{aligned}
\tag{131}
$$

Note that $e^{c+\frac{1}{2}\sum_{j=1}^{n+1}L_j^2} = \mathbb{E}(e^{c+\sum_{j=1}^{n+1}L_j\varepsilon_j})$. Let

$$B_j \equiv -\frac{K_j}{K_{n+1}} \tag{132}$$

$$a \equiv -\frac{\hat{a}}{K_{n+1}} \tag{133}$$

Assume $K_{n+1} > 0$. Then

$$\begin{aligned}
&\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\int_{-\infty}^{a_n} e^{c+\sum_{j=1}^{n+1}L_j\varepsilon_j}\phi(\varepsilon_{n+1})\phi(\varepsilon_n)\dots\phi(\varepsilon_1)d\varepsilon_{n+1}d\varepsilon_n\dots d\varepsilon_1 \\
&\quad= e^{c+\frac{1}{2}\sum_{j=1}^{n+1}L_j^2}\Phi\left(-\frac{\frac{\hat{a}}{K_{n+1}}+L_{n+1}+\sum_{j=1}^{n}\frac{K_j}{K_{n+1}}L_j}{\sqrt{1+\sum_{j=1}^{n}\left(\frac{K_j}{K_{n+1}}\right)^2}}\right) \\
&\quad= e^{c+\frac{1}{2}\sum_{j=1}^{n+1}L_j^2}\Phi\left(-\frac{\hat{a}+K_{n+1}L_{n+1}+\sum_{j=1}^{n}K_jL_j}{\sqrt{K_{n+1}^2+\sum_{j=1}^{n}K_j^2}}\right) \\
&\quad= e^{c+\frac{1}{2}\sum_{j=1}^{n+1}L_j^2}\Phi\left(-\frac{\hat{a}+\sum_{j=1}^{n+1}K_jL_j}{\sqrt{\sum_{j=1}^{n+1}K_j^2}}\right)
\end{aligned} \tag{134}$$

Suppose we instead have

$$\begin{aligned}
&\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\int_{a_n}^{\infty} e^{c+\sum_{j=1}^{n+1}L_j\varepsilon_j}\phi(\varepsilon_{n+1})\phi(\varepsilon_n)\dots\phi(\varepsilon_1)d\varepsilon_{n+1}d\varepsilon_n\dots d\varepsilon_1 \\
&\quad= \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} e^{c+\frac{L_{n+1}^2}{2}+\sum_{j=1}^{n}L_j\varepsilon_j}\left(\int_{(a_n-L_{n+1})}^{\infty}\phi(\varepsilon_{n+1})d\varepsilon_{n+1}\right)\phi(\varepsilon_n)\dots\phi(\varepsilon_1)d\varepsilon_n\dots d\varepsilon_1 \\
&\quad= \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} e^{c+\frac{L_{n+1}^2}{2}+\sum_{j=1}^{n}L_j\varepsilon_j}\Phi\Big(-(a-L_{n+1})-\sum_{j=1}^{n}B_j\varepsilon_j\Big)\phi(\varepsilon_n)\dots\phi(\varepsilon_1)d\varepsilon_n\dots d\varepsilon_1 \\
&\quad= e^{c+\frac{1}{2}\sum_{j=1}^{n+1}L_j^2}\Phi\left(-\frac{a-L_{n+1}+\sum_{j=1}^{n}B_jL_j}{\sqrt{1+\sum_{j=1}^{n}B_j^2}}\right)
\end{aligned} \tag{135}$$

Assume $K_{n+1} < 0$. Then

$$\begin{aligned}
&\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\int_{a_n}^{\infty} e^{c+\sum_{j=1}^{n+1}L_j\varepsilon_j}\phi(\varepsilon_{n+1})\phi(\varepsilon_n)\dots\phi(\varepsilon_1)d\varepsilon_{n+1}d\varepsilon_n\dots d\varepsilon_1 \\
&\quad= e^{c+\frac{1}{2}\sum_{j=1}^{n+1}L_j^2}\Phi\left(-\frac{-\frac{\hat{a}}{K_{n+1}}-L_{n+1}-\sum_{j=1}^{n}\frac{K_j}{K_{n+1}}L_j}{\sqrt{1+\sum_{j=1}^{n}\left(\frac{K_j}{K_{n+1}}\right)^2}}\right) \\
&\quad= e^{c+\frac{1}{2}\sum_{j=1}^{n+1}L_j^2}\Phi\left(|K_{n+1}|\frac{\frac{\hat{a}}{K_{n+1}}+L_{n+1}+\sum_{j=1}^{n}\frac{K_j}{K_{n+1}}L_j}{\sqrt{K_{n+1}^2+\sum_{j=1}^{n}K_j^2}}\right) \\
&\quad= e^{c+\frac{1}{2}\sum_{j=1}^{n+1}L_j^2}\Phi\left(-\frac{\hat{a}+L_{n+1}K_{n+1}+\sum_{j=1}^{n}K_jL_j}{\sqrt{K_{n+1}^2+\sum_{j=1}^{n}K_j^2}}\right) \\
&\quad= e^{c+\frac{1}{2}\sum_{j=1}^{n+1}L_j^2}\Phi\left(-\frac{a+\sum_{j=1}^{n+1}K_jL_j}{\sqrt{\sum_{j=1}^{n+1}K_j^2}}\right)
\end{aligned} \tag{136}$$

which is the same result as in (134).

G Proof of Theorem 1

Proof. of Theorem 1.

Assumption: $s < \infty$.

Let $T_h = t + h$. From Proposition 3, the ratio of the claim price to the fundamental market price is

$$\tilde{P}^X_{h,t} = \mathbb{E}_t\Bigg[e^{\hat{mp}_{T_h|t}+\hat{q}_{T_h|t}-\hat{g}_{T_h|t}(s)+\frac{\zeta_{y,t,T_h-1}}{2}}\Bigg(\Phi\Big(-\frac{\hat{g}_{T_h-1|t}(s)+\hat{r}_{d,T_h|t}+\zeta_{y,r_d+g,t,T_h-1}}{\sqrt{\zeta_{r_d+g,t,T_h-1}}}\Big)$$
$$-e^{\hat{g}_{T_h|t}(s)+\hat{r}_{d,T_h|t}+\zeta_{y,r_d+g,t,T_h-1}+\frac{\zeta_{r_d+g,t,T_h-1}}{2}}\Phi\Big(-\frac{\hat{g}_{T_h|t}(s)+\hat{r}_{d,T_h|t}+\zeta_{y,r_d+g,t,T_h-1}}{\sqrt{\zeta_{r_d+g,t,T_h-1}}}-\sqrt{\zeta_{r_d+g,t,T_h-1}}\Big)\Bigg)\Bigg]$$

and

$$P^X_{h,t} = P_t \tilde{P}^X_{h,t}$$

is the actual claim price. It is useful to write the log fundamental market price as

$$\begin{aligned}
p_t &= -z_{d,t} + d_t = -z_{d,t} + d_{t-1} + \Delta d_t \\
&\equiv \varrho\sigma_t\varepsilon_t + (\varrho - A_{d,x})\sigma_{x,t-1}\varepsilon_{x,t} + \xi_{p,t} \\
&\equiv \varrho\sigma_t\varepsilon_t + \theta_p\sigma_{x,t-1}\varepsilon_{x,t} + \xi_{p,t}
\end{aligned} \tag{137}$$

where $z_{d,t}$ is the log dividend price ratio on the fundamental market, d_t is the log dividend and ξ_t is independent of ε_t and $\varepsilon_{x,t}$. Then note

$$\begin{aligned}
\hat{mp}_{T_h|t} &= \mathbb{E}_t\Big(\sum_{j=1}^{h} m_{t+j} + p_{t+h-1}\Big|\mathcal{H}_{t+h}\Big) - p_t \\
&= \mathbb{E}_t\Big(\sum_{j=1}^{h} m_{t+j} - z_{d,t+h-1} + d_t + \sum_{j=1}^{h-1}\Delta d_{t+j}\Big|\mathcal{H}_{t+h}\Big) + z_{d,t} - d_t \\
&= \mathbb{E}_t\Big(\sum_{j=1}^{h} m_{t+j} - z_{d,t+h-1} + \sum_{j=1}^{h-1}\Delta d_{t+j}\Big|\mathcal{H}_{t+h}\Big) + z_{d,t}
\end{aligned} \tag{138}$$

is independent of ε_t, but depends on x_t (and thereby $\varepsilon_{x,t}$) directly through $z_{d,t}$ and indirectly through the terms inside the expectation due to the persistence of x. We can write $\hat{mp}_{T_h|t}$ as follows

$$
\begin{aligned}
\hat{mp}_{T_h|t} &= \Big((1-\rho_x^{h-1})A_{d,x} + \Gamma_{m,x}\sum_{j=1}^{h}\rho_x^{j-1} + \varrho\sum_{j=1}^{h-1}\rho_x^{j}\Big)\sigma_{x,t-1}\varepsilon_{x,t} + \xi_{\hat{mp},t} \\
&= \Big((1-\rho_x^{h-1})A_{d,x} + \Gamma_{m,x}\frac{1-\rho_x^h}{1-\rho_x} + \varrho\frac{\rho_x-\rho_x^h}{1-\rho_x}\Big)\sigma_{x,t-1}\varepsilon_{x,t} + \xi_{\hat{mp},t} \\
&\equiv \theta_{\hat{mp},h}\sigma_{x,t-1}\varepsilon_{x,t} + \xi_{\hat{mp},t} \qquad (139)
\end{aligned}
$$

where $\xi_{\hat{mp},t}$ is independent of ε_t and $\varepsilon_{x,t}$.

Statement 1.

Let $h > s$. First, note that $P_{i,t}^X$ only depend on ε_t through P_t, as $\zeta_{\cdot,t,T_h-1}$ and $\hat{r}_{d,T_h|t}$ always independent of ε_t and $h > s$ implies $\hat{q}_{T_h}(s)$ and $\hat{g}_{T_h}(s)$ do not depend on ε_t. Thus

$$
\begin{aligned}
Cov_{t-1}\Big(\sigma_{t-1}\varepsilon_t, log(P_{h,t}^X)\Big) &= Cov_{t-1}\Big(\sigma_{t-1}\varepsilon_t, log(P_t) + log(\tilde{P}_{h,t}^X)\Big) = Cov_{t-1}\Big(\sigma_{t-1}\varepsilon_t, log(P_t)\Big) \\
&= Cov_{t-1}\Big(\sigma_{t-1}\varepsilon_t, \varrho\sigma_{t-1}\varepsilon_t + \xi_t\Big) \\
&= \varrho\sigma_{t-1}^2 > 0 \qquad (140)
\end{aligned}
$$

Statement 2.

Assumption: $\varrho\psi > 1$

Let $h > s$. The second part of the theorem is a little more cumbersome as $P_{h,t}^X$ depend on ε_t through both P_t and $\tilde{P}_{h,t}^X$ because $\varepsilon_{x,t}$ affects the state variable x_t. Note that we can write the claim price as

$$
\begin{aligned}
P_{h,t}^X = e^{(\theta_p+\theta_{\hat{mp},h}+\theta_{q,h}-\theta_{g,h})\sigma_{x,t-1}\varepsilon_{x,t}} \Bigg(&\mathbb{E}_t\Big[e^{u_{1,h,T_h}}\Phi\Big(-\frac{(\theta_{g,h}+\theta_{r,h})\sigma_{x,t-1}\varepsilon_{x,t}+v_{h,T_h}}{\sqrt{\zeta_{r_d+g,t,T_h-1}}}\Big)\Big] \\
- e^{(\theta_{g,h}+\theta_{r,h})\sigma_{x,t-1}\varepsilon_{x,t}} &\mathbb{E}_t\Big[e^{u_{2,h,T_h}}\Phi\Big(-\frac{(\theta_{g,h}+\theta_{r,h})\sigma_{x,t-1}\varepsilon_{x,t}+v_{h,T_h}}{\sqrt{\zeta_{r_d+g,t,T_h-1}}} - \sqrt{\zeta_{r_d+g,t,T_h-1}}\Big)\Big]\Bigg) \qquad (141) \\
\equiv e^{(\theta_p+\theta_{\hat{mp},h}+\theta_{q,h}-\theta_{g,h})\sigma_{x,t-1}\varepsilon_{x,t}} V_{h,t} \qquad (142)
\end{aligned}
$$

where u and v are independent of $\varepsilon_{x,t}$. Note that $V_{h,t}$ is a smooth function $\varepsilon_{x,t}$. For $h \geq s$, the coefficients on $\sigma_{x,t-1}\varepsilon_{x,t}$ satisfies

$$\theta_{q,h+1} = \rho_x\theta_{q,h} \qquad (143)$$

$$\theta_{g,h+1} = \rho_x\theta_{g,h} \qquad (144)$$

$$\theta_{r,h+1} = \rho_x\theta_{r,h} \qquad (145)$$

The two first recursions follows from the policy specification - $(q_{T_h}(s),\ g_{T_h}(s))$ can at most directly depend on x_{T_h-s} and ε_{x,T_h-s}. If the policy specification gives $(\bar{q}_{x,j},\ \bar{g}_{x,j}) \neq 0$ for some $j = 0, 1, \ldots, s$,

$(q_{T_h}(s),\ g_{T_h}(s))$ will depend on x_t indirectly through $x_{T_h-j} = \rho_x^{h-j}x_t + e_{x,T_h-j}$. Thus, for $h > s$, any policy dependence on x_t, and thereby $\sigma_{x,t-1}\varepsilon_{x,t}$, must decrease exponentially with h. The final recursion follows directly from $r_{d,T_h} = \Gamma_{r_d,x}x_{T_h-1} + e_{r,T_h}$, where e_{r,T_h} is independent of ε_{x,T_h-j} $\forall j = 1,\ldots\infty$. Thus, r_{T_h} depends indirectly on x_t through the persistence of x_t.

We can also write

$$\begin{aligned}\theta_{x,h} \equiv \theta_p + \theta_{\hat{mp},h} &= \varrho - A_{d,x} + (1-\rho_x^{h-1})A_{d,x} + \Gamma_{m,x}\frac{1-\rho_x^h}{1-\rho_x} + \varrho\frac{\rho_x - \rho_x^h}{1-\rho_x} \\ &= \left(\varrho - \frac{\rho_x}{\psi}\right)\frac{1-\rho_x^h}{1-\rho_x} + \frac{\varrho - \frac{1}{\psi}}{1-\kappa_{d,1}\rho_x}\rho_x^h > 0 \end{aligned} \tag{146}$$

where I have used $\Gamma_{m,x} = -\frac{1}{\psi}\rho_x$ and $A_{d,x} = -\frac{\varrho-\frac{1}{\psi}}{1-\kappa_{d,1}\rho_x}\rho_x$. It is clear that

$$\theta_{x,h+1} > \theta_{x,h} \tag{147}$$

$$\theta_{x,h} \to \frac{\varrho - \frac{\rho_x}{\psi}}{1-\rho_x} \quad \text{as} \quad h \to \infty \tag{148}$$

$$|\theta_{y,h+1}| = \rho_x|\theta_{y,h}| \qquad \text{for} \quad y = q,\ g,\ r \tag{149}$$

$$\theta_{y,h} \to 0 \qquad\qquad \text{as} \quad h \to \infty \tag{150}$$

The covariance of interest can then be written

$$\begin{aligned} Cov_{t-1}\Big(\sigma_{x,t-1}\varepsilon_{x,t},\ log(P^X_{h,t})\Big) &= Cov_{t-1}\Big(\sigma_{x,t-1}\varepsilon_{x,t},\ \theta_{x,h}\sigma_{x,t-1}\varepsilon_{x,t} + (\theta_{q,h} - \theta_{g,h})\sigma_{x,t-1}\varepsilon_{x,t} + log(V_{h,t})\Big) \\ &= \theta_{x,h}\sigma^2_{x,t-1} + (\theta_{q,h} - \theta_{g,h})\sigma^2_{x,t-1} + Cov_{t-1}\Big(\sigma_{x,t-1}\varepsilon_{x,t},\ log(V_{h,t})\Big) \end{aligned} \tag{151}$$

The first term is clearly positive and increasing in h. The last two terms could be either positive or negative depending on the particular policy specification, but these terms can be made arbitrarily small in magnitude by increasing h. Thus, there exists some number k s.t. for all $h > k$

$$Cov_{t-1}\Big(\sigma_{x,t-1}\varepsilon_{x,t},\ log(P^X_{h,t})\Big) > 0 \tag{152}$$

Statement 3. From equation (151) we have

$$Cov_{t-1}\Big(\sigma_{x,t-1}\varepsilon_{x,t},\ log(P^X_{h,t})\Big) \to \frac{\varrho - \frac{\rho_x}{\psi}}{1-\rho_x}\sigma^2_{x,t-1} \quad \text{as} \quad h \to \infty \tag{153}$$

Furthermore,

$$\begin{aligned} Cov_{t-1}\Big(\sigma_{x,t-1}\varepsilon_{x,t},\ log(P_t)\Big) &= Cov_{t-1}\Big(\sigma_{x,t-1}\varepsilon_{x,t},\ -A_x x_t + \varrho\sigma_{x,t-1}\varepsilon_{x,t}\Big) = (\varrho - A_x)\sigma^2_{x,t-1} \\ &= \frac{[1 + (1-\kappa_{d,1})\rho_x]\varrho - \frac{\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x}\sigma^2_{x,t-1} \end{aligned} \tag{154}$$

Note that

$$
\begin{aligned}
&\frac{\varrho - \frac{\rho_x}{\psi}}{1-\rho_x} > \frac{[1+(1-\kappa_{d,1})\rho_x]\varrho - \frac{\rho_x}{\psi}}{1-\kappa_{d,1}\rho_x} \Leftrightarrow (1-\kappa_{d,1}\rho_x)\frac{\varrho - \frac{\rho_x}{\psi}}{1-\rho_x} > [1+(1-\kappa_{d,1})\rho_x]\varrho - \frac{\rho_x}{\psi} \Leftrightarrow \\
&\frac{\varrho - \frac{\rho_x}{\psi}}{1-\rho_x} + \frac{\rho_x}{\psi} - (1+\rho_x)\varrho > \kappa_{d,1}\rho_x\Big(\frac{\varrho - \frac{\rho_x}{\psi}}{1-\rho_x} - \varrho\Big) \Leftrightarrow \\
&\varrho - \frac{\rho_x}{\psi} + \frac{\rho_x}{\psi}(1-\rho_x) - (1+\rho_x)(1-\rho_x)\varrho > \Big(\varrho - \frac{1}{\psi}\Big)\rho_x^2\kappa_{d,1} \Leftrightarrow \\
&\Big(\varrho - \frac{1}{\psi}\Big)\rho_x^2 > \Big(\varrho - \frac{1}{\psi}\Big)\rho_x^2\kappa_{d,1} \Leftrightarrow \qquad (155) \\
&\kappa_{d,1} < 1 \qquad (156)
\end{aligned}
$$

Thus, the covariance $Cov_{t-1}(\sigma_{x,t-1}\varepsilon_{x,t},\ log(P^X_{h,t})) > Cov_{t-1}(\sigma_{x,t-1}\varepsilon_{x,t},\ log(P_t))$ as $h \to \infty$ ■

References

Bansal, R., and A. Yaron. 2004. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance* 59: 1481–1509.

Barbon, A., and V. Gianinazzi. 2019. Quantitative easing and equity prices: Evidence from the ETF program of the Bank of Japan. *The Review of Asset Pricing Studies* 9: 210–255.

Bernanke, B.S., and K.N. Kuttner. 2005. What explains the stock market's reaction to Federal Reserve policy? *Journal of Finance* 60: 1221–1257.

Bruno, G., and J. Haug. 2018. Expected equity returns should correlate with idiosyncratic risk. Working Paper, Norwegian School of Economics.

Campbell, J.Y., and R.J. Shiller. 1988. Stock prices, earnings, and expected dividends. *Journal of Finance* 43: 661–676.

Cieslak, A., and A. Vissing-Jorgensen. 2020. The economics of the Fed put. NBER Working Paper.

Dahiyay, S., B. Kamradz, V. Potix, and A. Siddique. 2019. The Greenspan put. Working Paper.

Elgin, C., G. Basbug, and A. Yalaman. 2020. Economic policy responses to a pandemic: Developing the Covid-19 economic stimulus index. CEPR Press.

Epstein, L.G., and S. E. Zin. 1991. Risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy* 99: 263–286.

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

Lochstoer, L., and T. Muir. 2020. Volatility expectations and returns. Working Paper, UCLA.

Moreira, A., and T. Muir. 2017. Volatility-managed portfolios. *Journal of Finance* 72: 1611–1644.

Putniņš, T.J. 2020. From Free markets to Fed markets: How unconventional monetary policy impacts equity markets. Working Paper, University of Technology Sydney; Stockholm School of Economics, Riga.

Rigobon, R., and B. Sack. 2003. Measuring the reaction of monetary policy to the stock market. *The Quarterly Journal of Economics* 118: 639–669.

Swanson, E. T. 2015. Measuring the effects of unconventional monetary policy on asset prices. NBER Working Paper.

Multi-Horizon HJ-Distance

Stig R. H. Lundeby

May 10, 2021

Abstract

I present a multi-period generalization of the standard HJ-distance. My distance metric uses the fact that a pricing kernel should price payoffs at any horizon. Intuitively, the multi-period distance is small if pricing errors are small for all horizons under consideration. I show that the multi-horizon perspective can detect significant model mis-specification where an analogous 1-period HJ-distance might conclude errors are small.

1 Introduction

Most investment and savings decisions take a multi-horizon perspective with in- and out-flows at several dates. A useful asset pricing model should therefore be able to price relevant payoffs at multiple horizons. Comparing models only at a single frequency, e.g. monthly, might therefore lead us astray. Yet, most empirical work only tests model implications for a given horizon at a time. In this paper, I develop a multi-horizon distance metric for model comparisons that tests model implications at several horizons jointly. A low multi-horizon distance indicates that the model performs well at all horizons under consideration.

As in Chernov et.al. (2021), I view payoffs at different horizons as separate test assets. Furthermore, each horizon can be thought of as a separate collection of states in an expanded state space. Viewing horizons as states allows me to employ the same machinery as in Hansen and Jagannathan (1997).

The multi-horizon metric is therefore a natural generalization of the metric developed in Hansen and Jagannathan (1997) to a multi-period setting. In fact, it turns out that my distance metric is a weighted average of HJ-distances for each horizon under consideration. As a consequence, the multi-period distance metric is zero if and only if the HJ-distance is zero at all horizons.

As in Hansen and Jagannathan (1997), the metric developed in this paper does not reward model variability. In typical GMM J-tests, the “numerator” is the model pricing errors and the “denominator” is essentially the variance of pricing errors. As a consequence, we might fail to reject a model either because the pricing errors are low, or because the pricing kernel is very variable. In my metric, the “denominator” is a block-diagonal matrix, where each block is the second moment matrix of test assets at a given horizon. The second-moment matrix of test asset payoffs is model-independent. Thus, high pricing kernel variability does not in itself lower the distance.

The model-independent “denominator” also means that the metric is ideal for model comparisons. If model (A) has a higher distance than model (B), model (A) has larger pricing errors than model (B). The distance metric is therefore a direct measure of how large the errors in the model-implied Euler equations are in economic terms, whereas the J-test can only conclude whether those errors are statistically significant.

Pricing errors for a given model can in principle be due to pricing “levels” of payoffs wrong or because the model mis-prices risk, i.e. excess returns. In some applications we might wish to distinguish between these channels of mis-specification. I therefore decompose the distance metric into an excess return distance and a level distance and show that the two components are orthogonal. In other word, the total distance squared can be written as the sum of the component distances squared.

It turns out that the distance metric is closely related to the (standardized) expected utility gain for a quadratic utility investor from exploiting the model-mis-specification. In particular, if we consider only the pricing of excess returns, the utility gain of the investor is proportional to the squared excess return distance metric. As a result, a larger excess return distance implies a greater utility gain from exploiting the mis-pricing of excess returns.

In a simple example economy, I show that the multi-period distance metric can reveal severe model mis-specification in a model even in the case where the standard 1-period HJ-distance is 0. Furthermore, the rankings of models often change when going from the single period metric to the multi-period metric, indicating that the multi-horizon view indeed adds valuable information.

The rest of the paper is organized as follows. Section 2 presents useful terminology and definitions that will be used throughout the paper. Section 3 defines the distance metric and presents some of its general properties. In section 4, I investigate the relationship between the distance metric and the utility loss experienced by an investor using a mis-specified model. Section 5 illustrates the use of the metric in a simple example economy.

2 Pricing kernels and payoff spaces

We are interested in finding a measure of how close a model y is to the "truth". Previous papers, most notably Hansen and Jagannathan (1997) have already developed such a distance metric for the single period case. However, as Chernov et.al (2021) shows, there is additional information in considering multi-horizon returns jointly. Incorporating multiple horizons often lead to rejections of models that are generally successful at a single horizon. Given that all models are at best approximations to the truth, we might be more interested in knowing which model is closest to explaining the data. It therefore makes sense to consider distance metrics that take into account model implications across horizons.

2.1 Candidate Model

Let y denote a candidate model for the pricing kernel in an economy. The model has some specified functional form at the 1-period horizon. For instance, in the unconditional CAPM, $y_{t,t+1} = a - bR^e_{m,t,t+1}$, where a and b are positive constants and $R^e_{m,t,t+1}$ is the 1-period excess market return. Another example is the standard CRRA expected utility kernel $y_{t,t+1} = \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}$, where C_t is the investor's consumption at date t, γ is his risk aversion and β measures his patience. The null-hypothesis is that y prices all returns[1] conditionally

$$\mathbb{E}_t(y_{t,t+1}R_{i,t,t+1}) = 1, \qquad \forall i \tag{1}$$

[1] Note that if the model prices all returns, it also prices all cash-flows assuming the Law of One Price holds.

Since the econometrician does not have access to the full information set of the investor at time t, the standard approach to testing (1) is by taking unconditional expectation

$$\mathbb{E}(y_{t,t+1}R_{i,t,t+1}) = 1, \qquad \forall i \tag{2}$$

However, as noted in Chernov et.al (2021), the Law of One Price has the additional implication that

$$\mathbb{E}_t\Big(y_{t,t+h}R_{i,t,t+h}\Big) = 1, \qquad \forall i \tag{3}$$

where $R_{i,t,t+h}$ is the h-period return on asset i and $y_{t,t+h} \equiv \prod_{j=0}^{h-1} y_{t+j,t+j+1}$ is the h-period pricing kernel implied by y. If the model is well-specified, (3) should hold for all horizons h. To simplify the notation, let $y^{(h)}$ denote the h-period discount factor and $R^{(h)}$ a vector of h-period returns. I will omit the time-subscripts unless needed for clarity.

In general, we might be interested in pricing other payoffs than returns. It is therefore useful to define pricing functionals for y at each horizon as follows

$$\pi^{(h)}_{t,y}\big(p^{(h)}\big) \equiv \mathbb{E}_t\big(y^{(h)}p^{(h)}\big) \qquad \forall p^{(h)} \tag{4}$$

$\pi^{(h)}_{t,y}(p^{(h)})$ is the hypothetical time t price y would assign to a (stochastic) payoff $p^{(h)}$ at $t+h$. Note that if y is misspecified, these hypothetical prices will generally differ from actual prices. Since we do not have access to the information embedded in $\mathbb{E}_t$, we do not observe the hypothetical prices. However, if we assume that payoffs and the pricing kernel are stationary, we can get an estimate of the average price assigned to a payoff

$$\pi^{(h)}_{y}\big(p^{(h)}\big) \equiv \mathbb{E}\big(y^{(h)}p^{(h)}\big) \qquad \forall p^{(h)} \tag{5}$$

The general idea of the multi-period distance metric presented in this paper, is to compare these average hypothetical prices to average observable market prices. We say that the model is well-specified if $\pi^{(h)}_{y}(p^{(h)})$ is close to average market prices for "all" payoffs and "all" horizons.

2.2 Payoff space

In order to be more precise about what "all" payoffs mean, we have to define a payoff space. When looking at a single horizon, a natural way to define the payoff space, $\mathcal{P}^{(1)}$, is to choose an $N_1 \times 1$ vector of basis returns $R^{(1)}_b$ and let the payoff-space be all linear combinations of these basis returns. Formally

$$\mathcal{P}^{(1)} \equiv \{c^\top R^{(1)}_b \quad : \quad c \in \mathbb{R}^{N_1}\} \tag{6}$$

Importantly, we accommodate conditioning information by allowing the basis returns to also contain specified trading strategies. The choice of using a vector of returns as basis is not essential, but it has the benefit that all payoffs will be stationary as long as returns are stationary. Furthermore, the market prices for returns are particularly easy as they equal 1 at all times.

Given the single-period payoff space, it seems natural in the multi-period setting to consider any payoffs attainable by creating portfolios of the basis returns $R?b^{(1)}$ and simply re-balance the portfolios every period from t to $t+h$, i.e.

$$\hat{\mathcal{P}}^{(h)} \equiv \left\{ c \cdot \left(\prod_{j=1}^{h} \omega^\top R_{b,t+j}^{(1)} \right) \quad : \quad c \in \mathbb{R}^1, \ \omega^\top \mathbf{1} = 1 \right\} \tag{7}$$

However, this space is too "large" due to the non-linear (multiplicative) way ω enters. As a consequence, (7) generally does not have a finite basis. To see this, consider a simple example where $R^{(1)}$ consists of the risk free rate and market return. $\hat{\mathcal{P}}^{(h)}$ would then clearly contain the h-period risk free rate and market return. However, it would also contain the h-period return on a portfolio investing half in each, rebalancing every period for h periods. This h-period portfolio return will not be spanned by the corresponding market and risk-free rate. In fact, there are infinitely many such portfolios. Thus, there is no finite set of basis returns that would span $\hat{\mathcal{P}}^{(h)}$ even in this simple 2-asset case.

As in the single-period setting, it is necessary for empirical implementation to limit attention to payoff spaces spanned by finitely many assets/strategies. I therefore take the simpler approach of specifying the basis returns for each horizon separately and let the payoff space consist of all linear combinations of basis returns. Formally, for each h, let $R_b^{(h)}$ denote an $N_h \times 1$ vector of basis returns. Then

$$\mathcal{P}^{(h)} \equiv \{ c^\top R_b^{(h)} \quad : \quad c \in \mathbb{R}^{N_h} \} \tag{8}$$

Note that we don't require $N_h = N_1$. Based on the discussion about $\hat{\mathcal{P}}^{(h)}$, we may want $N_h \geq N_1$ in order to get multi-horizon payoff spaces that get closer to spanning $\hat{\mathcal{P}}^{(h)}$. The following lemma states that if we choose basis returns at horizon h that are static strategies in returns in $\mathcal{P}^{(1)}$, then $\mathcal{P}^{(h)}$ is a subspace of $\hat{\mathcal{P}}^{(h)}$.

Lemma 1. *Suppose a basis for $\mathcal{P}^{(h)}$ is $R_b^{(h)}$. If each row i in $R_b^{(h)}$ can be written $R_{b,i,t,t+h}^{(h)} = \prod_{j=1}^{h} \omega_i^\top R_{t+j}$, where R is an $N \times 1$ vector of returns s.t. $R_j \in \mathcal{P}^{(1)}$ $\forall j = 1, \ldots, N$, and ω_i is a corresponding vector of portfolio weights. Then $\mathcal{P}^{(h)} \subset \hat{\mathcal{P}}^{(h)}$.*

Proof. Since all payoffs in $\mathcal{P}^{(h)}$ can be written as linear combinations of $R_b^{(h)}$, all we need to show is that each row of $R_b^{(h)}$ is in $\hat{\mathcal{P}}^{(h)}$. To show this, note $R_j \in \mathcal{P}^{(1)}$ implies that $R_j = \varphi_j^\top R_b^{(1)}$ for some vector of portfolio weights φ_j, where $R_b^{(1)}$ is a vector of basis returns for $\mathcal{P}^{(1)}$. By construction, $\hat{\mathcal{P}}^{(h)}$ contains every element that can be written as $\prod_{j=1}^{h} \omega^\top R_{t+j}^{(1)}$. The result follows immediately. ∎

2.2.1 Excess return space

It will prove useful to define the excess return space, $\mathcal{P}^{e(h)}$ associated with $\mathcal{P}^{(h)}$

$$\mathcal{P}^{e(h)} \equiv \left\{c \cdot \varphi^\top R_b^{(h)} \qquad : \qquad c \in \mathbb{R} \text{ and } \varphi^\top \mathbf{1} = 0\right\} \tag{9}$$

The excess return space is therefore all zero cost payoffs that can be generated from $\mathcal{P}^{(h)}$. The requirement that φ sums to 0, implies that we can represent the excess return space in a slightly different way. Without loss of generality, let $\bar{R}_b^{(h)}$ denote the last $N_h - 1$ rows of the vector of basis returns $R_b^{(h)}$ for $\mathcal{P}^{(h)}$. We can then generate a basis vector of excess returns, $R_b^{e(h)}$, as follows

$$R_b^{e(h)} \equiv \bar{R}_b^{(h)} - R_{b,1}^{(h)} \mathbf{1}$$

The excess return space is then

$$\mathcal{P}^{e(h)} = \left\{c^\top R_b^{e(h)} \qquad : \qquad c \in \mathbb{R}^{N_h - 1}\right\} \tag{10}$$

2.3 Admissible pricing kernels

Let $\pi_t^{(h)}(p)$ denote the observed time t price on an h-period ahead payoff and $\pi^{(h)}(p)$ its average. I assume the pricing functionals $\pi_t^{(h)}$ and $\pi^{(h)}$ are linear. For instance, the price of a payoff $p = c^\top R^{(h)}$ is $\pi_t^{(h)}(p) = c^\top \pi_t^{(h)}\left(R^{(h)}\right) = c^\top \mathbf{1}$. We say that an h-period pricing kernel is admissible if it assigns the correct *average* price to all payoffs in a payoff space. In particular,

$$\hat{\mathcal{M}}^{(h)} \equiv \{m \qquad : \quad \mathbb{E}(mp) = \pi^{(h)}(p) \text{ for every } p \in \hat{\mathcal{P}}^{(h)}\} \tag{11}$$

$$\mathcal{M}^{(h)} \equiv \{m \qquad : \quad \mathbb{E}(mp) = \pi^{(h)}(p) \text{ for every } p \in \mathcal{P}^{(h)}\} \tag{12}$$

$\hat{\mathcal{M}}^{(h)}$ and $\mathcal{M}^{(h)}$ denotes the sets of valid pricing kernels for payoff space $\hat{\mathcal{P}}^{(h)}$ and $\mathcal{P}^{(h)}$ respectively. Note that if m prices all payoffs in $\mathcal{P}^{(h)}$, it also prices all excess returns in $\mathcal{P}^{e(h)}$.

If the condition in Lemma 1 is satisfied, then the following lemma shows that $\hat{\mathcal{M}}^{(h)}$ is a subset of $\mathcal{M}^{(h)}$.

Lemma 2. *If $\mathcal{P}^{(h)} \subset \hat{\mathcal{P}}^{(h)}$ then $\hat{\mathcal{M}}^{(h)} \subset \mathcal{M}^{(h)}$.*

Proof. Need to show that an arbitrary $m \in \hat{\mathcal{M}}^{(h)}$ is also in $\mathcal{M}^{(h)}$. By definition, $m \in \hat{\mathcal{M}}^{(h)} \Leftrightarrow \mathbb{E}(mp) = \pi^{(h)}(p)$ for every $p \in \hat{\mathcal{P}}^{(h)}$. By Lemma 1 $\mathcal{P}^{(h)} \subset \hat{\mathcal{P}}^{(h)} \Leftrightarrow (p \in \hat{\mathcal{P}}^{(h)} \Rightarrow p \in \mathcal{P}^{(h)})$. Thus, $m \in \hat{\mathcal{M}}^{(h)} \Rightarrow \mathbb{E}(mp) = \pi^{(h)}(p)$ for every $p \in \mathcal{P}^{(h)} \Leftrightarrow m \in \mathcal{M}^{(h)}$. ■

Notice that definitions (11) and (12) only requires that admissible pricing kernels price payoffs at a given horizon h. Another natural way to define the sets of admissible pricing kernels would be to require that admissible discount factors prices all payoffs at "all" horizons. Let $\mathcal{H} \equiv \{1, h_2, \ldots, h_H\}$ denote the collection of horizons we are interested in. I assume that all horizons are natural numbers. Then the following sets of discount factors prices all payoffs correctly for every horizon in $\mathcal{H}$.

$$
\begin{aligned}
\hat{\bar{\mathcal{M}}}^{(1)} \equiv \Big\{ m^{(1)} \quad &: \quad \mathbb{E}(m^{(h)} p^{(h)}) = \pi^{(h)}(p^{(h)}) \text{ for every } p^{(h)} \in \hat{\mathcal{P}}^{(h)} \text{ and every } \ h \in \mathcal{H}, \\
&\quad \text{where } m^{(h)}_{t,t+h} \equiv \prod_{j=1}^{h} m^{(1)}_{t+j}. \Big\} \qquad (13) \\
\bar{\mathcal{M}}^{(1)} \equiv \Big\{ m^{(1)} \quad &: \quad \mathbb{E}(m^{(h)} p^{(h)}) = \pi^{(h)}(p^{(h)}) \text{ for every } p^{(h)} \in \mathcal{P}^{(h)} \text{ and every } \ h \in \mathcal{H}, \\
&\quad \text{where } m^{(h)}_{t,t+h} \equiv \prod_{j=1}^{h} m^{(1)}_{t+j}. \Big\} \qquad (14)
\end{aligned}
$$

In general it is often difficult to characterize a single element in either of these sets. However, Lemma 3 tells us that the sets in (13) and (14) are smaller than the sets in (11) and (12).

Lemma 3. *Consider arbitrary $\hat{m}^{(1)} \in \hat{\bar{\mathcal{M}}}^{(1)}$ and $m^{(1)} \in \bar{\mathcal{M}}^{(1)}$. Define $\hat{m}^{(h)}_{t,t+h} \equiv \prod_{j=1}^{h} \hat{m}^{(1)}_{t+j}$ and $m^{(h)}_{t,t+h} \equiv \prod_{j=1}^{h} m^{(1)}_{t+j}$ for every $h \in \mathcal{H}$. Then*

1. *$\hat{m}^{(h)} \in \hat{\mathcal{M}}^{(h)}$ for every $h \in \mathcal{H}$*
2. *$m^{(h)} \in \mathcal{M}^{(h)}$ for every $h \in \mathcal{H}$*

Proof. The results follow directly from the definitions of $\hat{\bar{\mathcal{M}}}^{(1)}$ and $\bar{\mathcal{M}}^{(1)}$. ■

2.4 Viewing dates as states

In order to define the distance metric, it is useful to think of each horizon as separate (collections of) states. To that end, let Ω^P denote the physical state-space and define the pseudo state-space

$$\Omega \equiv \mathcal{H} \times \Omega^P = \{(h, \omega) \qquad : \qquad h \in \mathcal{H} \text{ and } \omega \in \Omega^P\}$$

where $\times$ denotes the Cartesian product. For any h-horizon payoff $p^{(h)}$, we can define a payoff p on Ω as follows

$$p(j, \omega) \equiv \begin{cases} p^{(h)}(\omega) & \text{if } j = h \\ 0 & \text{otherwise} \end{cases} \qquad (15)$$

and in particular, the h-horizon vector of basis returns becomes

$$\tilde{R}^{(h)}_b(j, \omega) \equiv \begin{cases} R^{(h)}_b(\omega) & \text{if } j = h \\ 0 & \text{otherwise} \end{cases}$$

In words, $\tilde{R}_b^{(h)}$ pays of $R_b^{(h)}$ at horizon h and 0 at all other horizons. We can then collect basis returns for all horizons into a single vector $\tilde{R}_b \equiv (\tilde{R}_b^{(1)\top}, \tilde{R}_b^{(h_2)\top}, \ldots, \tilde{R}_b^{(h_H)\top})^\top$, which has $N = \sum_{h\in\mathcal{H}} N_h$ rows. From $\tilde{R}_b$, we can generate all the payoffs in $\mathcal{P}^{(1)}$, ..., $\mathcal{P}^{(h_H)}$. Note that payoffs p on Ω derived from $p^{(h)} \in \mathcal{P}^{(h)}$, can be represented as linear combinations of $\tilde{R}_b$. Thus, $\tilde{R}_b$ is a basis for our multi-horizon payoff space

$$\mathcal{P} \equiv \{c^\top \tilde{R}_b \quad : \quad c \in \mathbb{R}^N\} \tag{16}$$

The multi-horizon excess return space can be defined the same way. That is,

$$\tilde{R}_b^{e(h)}(j,\omega) \equiv \begin{cases} R_b^{e(h)}(\omega) & \text{if } j = h \\ 0 & \text{otherwise} \end{cases}$$

and $\tilde{R}_b^e \equiv (\tilde{R}_b^{e(1)\top}, \tilde{R}_b^{e(h_2)\top}, \ldots, \tilde{R}_b^{e(h_H)\top})^\top$. The multi-horizon excess return space is then

$$\mathcal{P}^e \equiv \{c^\top \tilde{R}_b^e \quad : \quad c \in \mathbb{R}^{N-H}\} \tag{17}$$

The sets of admissible pricing kernels defined on Ω becomes

$$\mathcal{M} \equiv \{m \quad : \quad m(h,\cdot) \in \mathcal{M}^{(h)} \ \ \forall\, h \in \mathcal{H}\} \tag{18}$$

$$\hat{\mathcal{M}} \equiv \{m \quad : \quad m(h,\cdot) \in \hat{\mathcal{M}}^{(h)} \ \ \forall\, h \in \mathcal{H}\} \tag{19}$$

$$\bar{\mathcal{M}} \equiv \left\{m \quad : \quad m(h,\omega) = m^{(h)}(\omega) \ \ \forall\, h \in \mathcal{H}, \text{ where } m_{t,t+h}^{(h)} = \prod_{j=1}^{h} m_{t+j}^{(1)} \quad \text{and } m^{(1)} \in \bar{\mathcal{M}}^{(1)}\right\} \tag{20}$$

The first two sets simply require that an admissible pricing kernel must be admissible at every horizon. The third set places the additional requirement that an admissible pricing kernel m must be multiplicative, i.e. m in states associated with horizon $h > 1$ must be the product of m in states associated with the 1-period horizon.

We can extend the physical probability measure $\mathbb{P}$ to Ω as follows

$$\mathbb{Q}(h \times A) = \theta_h \mathbb{P}(A), \qquad \forall h \in \mathcal{H} \text{ and } \forall A \subset \Omega^P$$

where $\theta_h > 0$ and $\sum_{h\in\mathcal{H}} \theta_h = 1$. Let $\hat{\mathbb{E}}_{\theta,\mathcal{H}}$ denote the expectation operator implied by the measure $\mathbb{Q}$. For any random variable X defined on Ω, we then have

$$\hat{\mathbb{E}}_{\theta,\mathcal{H}}(X) = \sum_{h\in\mathcal{H}} \theta_h \mathbb{E}(X(h,\omega)) \tag{21}$$

where $\mathbb{E}$ denotes the physical expectation operator. The expectation operator $\hat{\mathbb{E}}_{\theta,\mathcal{H}}$ naturally leads us to consider the norm

$$||x||_{\theta,\mathcal{H}} \equiv \hat{\mathbb{E}}_{\theta,\mathcal{H}}(x^2)^{\frac{1}{2}} \tag{22}$$

Clearly, the norm depends on both the choice of "probabilities" θ, and on what horizons $\mathcal{H}$ we consider. Although it might seem strange to assign a probability to a horizon (which hopefully occurs with probability 1), what I am actually doing is to introduce a weighting scheme so that the norm does not mechanically grow by adding more horizons. One natural choice for θ would be to just equal-weight the horizons. However, section 4 gives a utility-based justification for deviating from equal weighting.

3 Multi-horizon distance metric

Having defined a norm, we can define multi-period analogues to the HJ-distance. For a candidate discount factor y, let

$$\delta_{\theta,\mathcal{H},y} \equiv \min_{m\in\mathcal{M}} ||y-m||_{\theta,\mathcal{H}} \tag{23}$$

$$\hat{\delta}_{\theta,\mathcal{H},y} \equiv \min_{m\in\hat{\mathcal{M}}} ||y-m||_{\theta,\mathcal{H}} \tag{24}$$

$$\bar{\delta}_{\theta,\mathcal{H},y} \equiv \min_{m\in\bar{\mathcal{M}}} ||y-m||_{\theta,\mathcal{H}} \tag{25}$$

The first metric measures the distance between y and a set of pricing kernels that price all payoffs in $\mathcal{P}^{(h)}$ for every horizon. The second metric requires admissible pricing kernels to price any multi-period payoff that can be generated by creating portfolios of 1-period returns. For the first two metrics, there is no requirement that m is multiplicative at different horizons. The third metric is the distance between y and the closest pricing kernel that is multiplicative and prices payoffs at all horizons. In the rest of the paper, I will refer to the first metric given in (23) as the distance metric.

The following proposition shows that the squared multi-period distance metric is a weighted average of squared standard HJ-distances at each horizon

Proposition 1. *Let $\delta_{\theta,\mathcal{H},y}$ be given by (23), then $\delta^2_{\theta,\mathcal{H},y} = \sum_{h\in\mathcal{H}} \theta_h \delta^2_{y,h}$ where $\delta_{y,h}$ is the standard HJ-distance for horizon h and payoff-space $\mathcal{P}^{(h)}$.*

Proof. The minimization problem in (23) is equivalent to

$$
\begin{aligned}
\delta^2_{\theta,\mathcal{H},y} &= \min_{m\in\mathcal{M}} ||y-m||^2_{\theta,\mathcal{H}} = \min_{m\in\mathcal{M}} \sum_{h\in\mathcal{H}} \theta_h \mathbb{E}\big((y^{(h)}-m^{(h)})^2\big) \\
&= \sum_{h\in\mathcal{H}} \theta_h \min_{m^{(h)}\in\mathcal{M}^{(h)}} \mathbb{E}\big((y^{(h)}-m^{(h)})^2\big) \\
&= \sum_{h\in\mathcal{H}} \theta_h \delta^2_{y,h} \qquad (26)
\end{aligned}
$$

The second equality uses the definitions of the norm (22) and the pseudo-expectation (21). ∎

An interesting question is whether we can say anything about $\hat{\delta}_{\theta,\mathcal{H},y}$ and $\bar{\delta}_{\theta,\mathcal{H},y}$. The following theorem shows that $\delta_{\theta,\mathcal{H},y}$ is a lower bound for these alternative metrics.

Theorem 1. *Let $\hat{\delta}_{\theta,\mathcal{H},y}$ and $\bar{\delta}_{\theta,\mathcal{H},y}$ be defined as in (24) and (25) respectively. Then*

1. $\bar{\delta}_{\theta,\mathcal{H},y} \geq \delta_{\theta,\mathcal{H},y}$
2. *If the basis returns for $\mathcal{P}^{(h)}$ are h-period portfolio returns in the basis assets of $\mathcal{P}^{(1)}$ for every $h \in \mathcal{H}$, then $\hat{\delta}_{\theta,\mathcal{H},y} \geq \delta_{\theta,\mathcal{H},y}$*

Proof. Lemma 3 implies 1. Lemma 1 and 2 implies 2. ∎

Using the proposed metric allows us to use the same machinery as in Hansen and Jagannathan (1997). The remaining part of this sub-section therefore follows Hansen and Jagannathan (1997) closely. Let

$$\hat{\pi}_y(p) \equiv \sum_{h\in\mathcal{H}} \theta_h \pi^{(h)}_y(p^{(h)}) \qquad (27)$$

$$\hat{\pi}(p) \equiv \sum_{h\in\mathcal{H}} \theta_h \pi^{(h)}(p^{(h)}) \qquad (28)$$

where $\pi^{(h)}_y(p^{(h)})$ is the price assigned by model y to the payoff $p^{(h)}$, and $\pi^{(h)}(p^{(h)})$ is the actual price of payoff $p^{(h)}$.[2] Thus, $\hat{\pi}(p)$ is the actual price of a cash-flow that pays $\theta_h p^{(h)}$ at every horizon $h \in \mathcal{H}$, and $\hat{\pi}_y(p)$ is the corresponding price assigned by model y. From the definition of the pseudo-expectation, we can write (27) and (28) as

$$\hat{\pi}_y(p) = \hat{\mathbb{E}}_{\theta,\mathcal{H}}(yp) \qquad \forall p \in \mathcal{P}$$

$$\hat{\pi}(p) = \hat{\mathbb{E}}_{\theta,\mathcal{H}}(mp) \qquad \forall m \in \mathcal{M} \text{ and } \forall p \in \mathcal{P}$$

We can also define a pricing error functional as

$$\tilde{\pi}_y(p) \equiv \hat{\pi}_y(p) - \hat{\pi}(p) \qquad \forall p \in \mathcal{P} \qquad (29)$$

[2]Note that the structure of $\mathcal{P}$ implies that the price of a payoff equals its average. $\mathcal{P}$ only includes constant linear combinations of returns. Returns have constant price 1. Thus, the price on a payoff is a constant times 1.

In particular, applying the pricing error functional to an h-period return gives us $\tilde{\pi}_y(\tilde{R}^{(h)}) = \theta_h \Big(\pi_y^{(h)}(R^{(h)}) - \pi^{(h)}(R^{(h)})\Big)$. By the Riesz representation theorem, there exists a unique $\tilde{p} \in \mathcal{P}$ s.t.

$$\hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{p}p) = \tilde{\pi}_y(p) \qquad\qquad \forall p \in \mathcal{P} \tag{30}$$

I will therefore refer to $\tilde{p}$ as the mis-pricing payoff.

It follows that

$$\hat{\mathbb{E}}_{\theta,\mathcal{H}}\big((y - m - \tilde{p})p\big) = 0 \qquad\qquad \forall m \in \mathcal{M} \;\text{ and }\; \forall p \in \mathcal{P}$$

In other words, $y - m - \tilde{p}$ is orthogonal to $\mathcal{P}$ for any $m \in \mathcal{M}$, which leads us to consider

$$y - m = \tilde{p} + \varepsilon_m$$

where $\hat{\mathbb{E}}_{\theta,\mathcal{H}}(\varepsilon_m p) = 0$ for any payoff $p \in \mathcal{P}$. The mis-pricing payoff $\tilde{p}$ is therefore a least-squares projection of $y-m$ on $\mathcal{P}$. The representation allows us to arrive at the following lower bound for the distance between y and any admissible pricing kernel m

$$\begin{aligned}||y - m||^2_{\theta,\mathcal{H}} &= \hat{\mathbb{E}}_{\theta,\mathcal{H}}\big((y - m)^2\big) = \hat{\mathbb{E}}_{\theta,\mathcal{H}}\big((\tilde{p} + \varepsilon_m)^2\big) = \hat{\mathbb{E}}_{\theta,\mathcal{H}}\big(\tilde{p}^2 + 2\tilde{p}\varepsilon_m + \varepsilon_m^2\big) \\ &= \hat{\mathbb{E}}_{\theta,\mathcal{H}}\big(\tilde{p}^2 + \varepsilon_m^2\big) \geq \hat{\mathbb{E}}_{\theta,\mathcal{H}}\big(\tilde{p}^2\big) = ||\tilde{p}||^2_{\theta,\mathcal{H}}\end{aligned} \tag{31}$$

where the last equality uses $\hat{\mathbb{E}}_{\theta,\mathcal{H}}(\varepsilon_m p) = 0$ for any $p \in \mathcal{P}$ and that $\tilde{p} \in \mathcal{P}$. Since the inequality in (31) applies for any $m \in \mathcal{M}$, it also places a lower bound on the distance metric

$$\delta_{\theta,\mathcal{H},y} \equiv \min_{m \in \mathcal{M}} ||y - m||_{\theta,\mathcal{H}} \geq ||\tilde{p}||_{\theta,\mathcal{H}} \tag{32}$$

Furthermore, since $(y - \tilde{p}) \in \mathcal{M}$, we also get an upper bound on the distance metric

$$\delta_{\theta,\mathcal{H},y} \equiv \min_{m \in \mathcal{M}} ||y - m||_{\theta,\mathcal{H}} \leq ||y - (y - \tilde{p})||_{\theta,\mathcal{H}} = ||\tilde{p}||_{\theta,\mathcal{H}} \tag{33}$$

Using both (32) and (33) leads us to conclude

$$\delta_{\theta,\mathcal{H},y} = ||\tilde{p}||_{\theta,\mathcal{H}}$$

It is also worth noting that $\delta_{\theta,\mathcal{H},y}$ has the interpretation as the maximum mis-pricing per unit-norm payoff since by the Cauchy-Schwarz inequality

$$|\hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{p}p)|^2 \leq \hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{p}^2)\hat{\mathbb{E}}_{\theta,\mathcal{H}}(p^2) = ||\tilde{p}||^2_{\theta,\mathcal{H}} \cdot ||p||^2_{\theta,\mathcal{H}}$$

Using (30), then taking the square-root on both sides, and rearranging gives us

$$\frac{|\tilde{\pi}_y(p)|}{||p||_{\theta,\mathcal{H}}} \leq ||\tilde{p}||_{\theta,\mathcal{H}} \tag{34}$$

The weak inequality in (34) is satisfied with equality for $\tilde{p}$.

$$\max_{p\in\mathcal{P},\ ||p||_{\theta,\mathcal{H}}=1} |\tilde{\pi}_y(p)| = ||\tilde{p}||_{\theta,\mathcal{H}}$$

In other words, $\tilde{p}$ is also the maximally mis-priced payoff.

3.1 Representing $\tilde{p}$ and $\delta_{\theta,\mathcal{H},y}$

Since $\tilde{R}_b$ is a basis for $\mathcal{P}$, any payoff $p \in \mathcal{P}$ can be written $c^\top \tilde{R}_b$. Consider the vector of pseudo pricing errors on the base assets

$$\tilde{\alpha} \equiv \tilde{\pi}_y(\tilde{R}_b)$$

We wish to represent $\tilde{p} = \tilde{c}^\top \tilde{R}_b$. Since by the definition of $\tilde{p}$, $\tilde{\alpha} = \hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{p}\tilde{R}_b)$, we get

$$\tilde{c} = \hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{R}_b\tilde{R}_b^\top)^{-1}\tilde{\alpha} = \begin{pmatrix} \mathbb{E}\big(R_b^{(1)}R_b^{(1)\top}\big)^{-1}\alpha^{(1)} \\ \mathbb{E}\big(R_b^{(h_2)}R_b^{(h_2)\top}\big)^{-1}\alpha^{(h_2)} \\ \vdots \\ \mathbb{E}\big(R_b^{(h_H)}R_b^{(h_H)\top}\big)^{-1}\alpha^{(h_H)} \end{pmatrix}$$

where $\alpha^{(h)} \equiv \mathbb{E}(y^{(h)}R_b^{(h)}) - 1$ are vectors of actual pricing errors. Importantly, $\tilde{c}$, and by extension $\tilde{p}$, does not depend on the pseudo-probabilities θ. In fact, $\tilde{p}$ can be thought of as an asset that pays single-horizon mis-pricing payoffs $\tilde{p}^{(h)}$ at every horizon h

$$\begin{aligned} \tilde{p}^{(h)} &= \alpha^{(h)\top}\mathbb{E}\big(R_b^{(h)}R_b^{(h)\top}\big)^{-1}R_b^{(h)} \\ \tilde{p} &= \sum_{h\in\mathcal{H}} \tilde{p}^{(h)} \end{aligned}$$

The multi-period distance metric can be computed as

$$\begin{aligned} \delta_{\theta,\mathcal{H},y} = ||\tilde{p}||_{\theta,\mathcal{H}} &= \sqrt{\tilde{\alpha}^\top\hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{R}_b\tilde{R}_b^\top)^{-1}\tilde{\alpha}} \\ &= \sqrt{\sum_{h\in\mathcal{H}} \theta_h \alpha^{(h)\top}\mathbb{E}\big(R_b^{(h)}R_b^{(h)\top}\big)^{-1}\alpha^{(h)}} \end{aligned}$$

3.2 Decomposing the distance metric

The distance metric $\delta_{\theta,\mathcal{H},y}$ does not distinguish between mis-pricing "levels" of returns and risk. For instance, if all returns have identical mis-pricing, the candidate model y prices excess returns (risk) correctly. On the other hand, if the cross-sectional weighted average mis-pricing of returns is 0, there is no mis-pricing in levels.

To investigate the two channels of mis-specification, note that the (pseudo) pricing errors associated with h-period excess returns is

$$\tilde{\alpha}^{e(h)} \equiv \hat{\mathbb{E}}_{\theta,\mathcal{H}}(y\tilde{R}_b^{e(h)}) = \tilde{\alpha}_{-1}^{(h)} - \tilde{\alpha}_1^{(h)}\mathbf{1}$$

where $\tilde{\alpha}_{-1}^{(h)}$ is the last $N_h - 1$ rows of the vector of pricing errors associated with the h-period basis returns, and $\tilde{\alpha}_1^{(h)}$ is the first row of the same vector. Stack the excess return pricing errors for every horizon into a single vector $\tilde{\alpha}^e \equiv (\tilde{\alpha}^{e(1)\top}, \ldots, \tilde{\alpha}^{e(h_H)\top})^\top$.

Clearly, there exists an excess return $\tilde{p}^e \in \mathcal{P}^e$ that gives the same pricing errors as y for all excess returns

$$\hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{p}^e p^e) = \tilde{\pi}_y(p^e), \qquad\qquad \forall\, p^e \in \mathcal{P}^e$$

To show this, note that all we need to show is that there exists an excess return $\tilde{p}^e = c^\top \tilde{R}_b^e$ such that

$$\tilde{\alpha}^e = \hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{p}^e \tilde{R}_b^e)$$

It is easy to see that

$$\tilde{p}^e = \tilde{\alpha}^{e\top}\hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{R}_b^e\tilde{R}_b^{e\top})^{-1}\tilde{R}^e \tag{35}$$

does the job. Furthermore, $m^e \equiv y - \tilde{p}^e$ is an admissible pricing kernel for the excess returns. We will use this fact to define an excess return distance metric

$$\delta_{\theta,\mathcal{H},y}^e \equiv ||y - m^e||_{\theta,\mathcal{H}} = ||\tilde{p}^e||_{\theta,\mathcal{H}} = \sqrt{\tilde{\alpha}^{e\top}\hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{R}_b^e\tilde{R}_b^{e\top})^{-1}\tilde{\alpha}^e} \tag{36}$$

In general, m^e will not price other payoffs in $\mathcal{P}$. However, all returns at a given horizon h will have identical pricing errors when using m^e. We can therefore think about the distance between m^e and $m^* \equiv y - \tilde{p}$ as the level distance. Let

$$\tilde{p}^L \equiv \tilde{p} - \tilde{p}^e \tag{37}$$

Then, the level distance can be defined as

$$\delta^L_{\theta,\mathcal{H},y} \equiv ||m^e - m^*||_{\theta,\mathcal{H}} = ||\tilde{p}^L||_{\theta,\mathcal{H}} \tag{38}$$

The following proposition states that the decomposition of $\tilde{p}$ into $\tilde{p}^e$ and $\tilde{p}^L$ is orthogonal

Proposition 2. *Let $\tilde{p}$, $\tilde{p}^e$, and $\tilde{p}^L$ be given by equations (30), (35), and (37) respectively. Then,*

1. *The level payoff $\tilde{p}^L$ is orthogonal to all excess returns. In particular,*

$$\hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{p}^L\tilde{p}^e) = 0$$

2. *The level and excess return distances are orthogonal*

$$\delta^2_{\theta,\mathcal{H},y} = (\delta^L_{\theta,\mathcal{H},y})^2 + (\delta^e_{\theta,\mathcal{H},y})^2$$

Proof. For the first statement, note that $\tilde{p}^L + \tilde{p}^e = \tilde{p}$, and that $m^* = y - \tilde{p}$ prices all payoffs in $\mathcal{P}$. In particular, m^* prices all excess returns. Thus,

$$\hat{\mathbb{E}}_{\theta,\mathcal{H}}(m^*p^e) = 0 \qquad\qquad \forall\, p^e \in \mathcal{P}^e$$

By construction, $m^e = y - \tilde{p}^e$ also prices all excess returns. Thus,

$$0 = \hat{\mathbb{E}}_{\theta,\mathcal{H}}([m^e - m^*]p^e) = \hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{p}^L p^e) \qquad\qquad \forall p^e \in \mathcal{P}^e$$

Since $\tilde{p}^e \in \mathcal{P}^e$, the result follows.

For the second statement we have

$$\begin{aligned}\delta^2_{\theta,\mathcal{H},y} &= \hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{p}^2) = \hat{\mathbb{E}}_{\theta,\mathcal{H}}([\tilde{p}^L + \tilde{p}^e]^2) = \hat{\mathbb{E}}_{\theta,\mathcal{H}}([\tilde{p}^L]^2) + \hat{\mathbb{E}}_{\theta,\mathcal{H}}([\tilde{p}^e]^2) + 2\hat{\mathbb{E}}_{\theta,\mathcal{H}}(\tilde{p}^L\tilde{p}^e)\\ &= (\delta^L_{\theta,\mathcal{H},y})^2 + (\delta^e_{\theta,\mathcal{H},y})^2\end{aligned}$$

where the last equality uses equations (38) and (36) and that $\tilde{p}^L$ is orthogonal to $\tilde{p}^e$. ■

4 Quadratic utility and multi-horizon distance metric

In this section we are interested in getting an economic assessment of the error made by an investor who uses a pricing model y for his investment decisions. To keep things simple, consider a quadratic utility investor who maximizes expected utility

$$\max_{c_0,\{c_h\}} \; u(c_0) + \mathbb{E} \sum_{h\in\mathcal{H}} \beta^h u(c_h) \tag{39}$$
$$u(c_h) = -\frac{(c_h^b - c_h)^2}{2}$$

where c_h^b denotes the potentially stochastic consumption bliss point. The maximization is subject to the constraint that the date 0 value of his consumption stream equals his wealth. The utility function in (39) is normalized to have $u''(c) = -1$. Thus, any utility change can be thought of as utility change normalized by the absolute value of the second derivative. Utility changes normalized in this way have the benefit that they are invariant to positive affine (preference preserving) transformations of u.

The investor's intertemporal marginal rate of substitution (IMRS) is

$$M_{0,h}^I(c_0, c_h) = \beta^h \frac{c_h^b - c_h}{c_0^b - c_0} \tag{40}$$

Risk averse investors who maximize expected utility wish to smooth consumption across states and dates. Proposition 3 tells us that there is a close connection between the norm defined by (22) and the utility maximization problem in (39). In fact, given optimal date 0 consumption, the optimal future consumption plan minimize the norm of the investor's IMRS/pricing kernel

Proposition 3. *Let* $S \equiv \sum_{h\in\mathcal{H}} \beta^{-h}$ *and* $\theta_h \equiv \frac{\beta^{-h}}{S}$. *Then the optimization problem in (39) can be written*

$$\max_{c_0} \left\{ u(c_0) - \frac{1}{2}(c_0^b - c_0)^2 S \min_{\tilde{c}} ||M^I(c_0,\tilde{c})||^2_{\theta,\mathcal{H}} \right\}$$

where $\tilde{c}$ *takes the value* c_h *in* h*-states and* M^I *is given by (40).*

Proof. Note that we can substitute equation (40) into (39) to obtain

$$\begin{aligned}
&\max_{c_0,\{c_h\}} u(c_0) - \frac{1}{2}(c_0^b - c_0)^2 \; \mathbb{E} \sum_{h\in\mathcal{H}} \beta^{-h} M_{0,h}^I(c_0,c_h)^2 \\
&= \max_{c_0,\{c_h\}} u(c_0) - \frac{1}{2}(c_0^b - c_0)^2 S \; \mathbb{E} \sum_{h\in\mathcal{H}} \frac{\beta^{-h}}{S} M_{0,h}^I(c_0,c_h)^2 \\
&= \max_{c_0,\{c_h\}} u(c_0) - \frac{1}{2}(c_0^b - c_0)^2 S \; \mathbb{E} \sum_{h\in\mathcal{H}} \theta_h M_{0,h}^I(c_0,c_h)^2 \\
&= \max_{c_0,\{\tilde{c}\}} u(c_0) - \frac{1}{2}(c_0^b - c_0)^2 S \; \hat{\mathbb{E}}_{\theta,\mathcal{H}} \Big(M^I(c_0,\tilde{c})^2 \Big)
\end{aligned}$$

where the last equality follows from the definition of $\hat{\mathbb{E}}_{\theta,\mathcal{H}}$. The result then follows immediately from the definition of the norm. ∎

In order to assess a candidate model, suppose the investor initially believes y is a valid discount factor for every h in $\mathcal{H}$. In the current section, the implication is that the investor should seek to equate his IMRS to y at all horizons and in all states, i.e.

$$M^I_{0,h}(c_0, c_h) = y^{(h)}_{0,h} \quad \forall h \in \mathcal{H}$$

which implies the following consumption stream

$$c^y_h = c^b_h - \beta^{-h}(c^b_0 - c^y_0)y^{(h)}_{0,h} \tag{41}$$

assuming c^y attainable. In what follows, I will assume $c^b_h \in \mathcal{P}^{(h)}$ and $y^{(h)} \in \mathcal{P}^{(h)}$ for all $h \in \mathcal{H}$, which is a sufficient condition for c^y to be attainable. The expected utility from using model y is then

$$U_y \equiv u(c^y_0) - \frac{1}{2}(c^b_0 - c^y_0)^2 S\ ||y||^2_{\theta,\mathcal{H}}$$

If the candidate SDF y is indeed an admissible pricing kernel, c^y is his optimal consumption plan and U_y the maximum attainable utility.

4.1 Return Mis-pricing

Suppose the investor learns that the candidate pricing kernel unconditionally mis-prices a set of (basis) returns that he can trade

$$\mathbb{E}\big(y^{(h)}R^{(h)}_b\big) = \mathbf{1} + \alpha^{(h)} \tag{42}$$

where α_h denotes the vector of pricing errors. It then follows that there exists a set of trades at date zero that only requires knowledge about the unconditional distribution of returns and the candidate SDF y to improve his expected utility unconditionally.

Proposition 4. *There exists an admissible pricing kernel m^* s.t. the consumption stream implied by m^* is attainable, and*

1. *The utility gain normalized by $|u''|$ for a quadratic utility investor from using the pricing kernel m^* instead of the mis-specified model y is*

$$\frac{U_{m^*} - U_y}{-u''(c_0)} = \frac{1}{2}a^{-2}S\Big(\delta^2_{\theta,\mathcal{H},y} - \frac{S}{1 + S||m^*||^2}\hat{\pi}(\tilde{p})^2\Big)$$

where $a \equiv \frac{-u''(c^y_0)}{u'(c^y_0)}$ is the coefficient of absolute risk aversion evaluated at his initial consumption under model y.

2. *The pricing kernel m^* minimizes the distance to y*

$$m^* = \arg\min_{m \in \mathcal{M}} ||y - m||$$

Proof. Part 1. Let $m^* = y - \tilde{p}$. By construction of $\tilde{p}$, $m^* \in \mathcal{M}$. Furthermore, since $y \in \mathcal{P}$ and $\tilde{p} \in \mathcal{P}$, we have $m^* \in \mathcal{P}$ as well. It then follows that the consumption stream

$$\begin{aligned} c_h^{m^*} &= c_h^b - (c_0^b - c_0^y - \Delta c_0)\beta^{-h}\left(y^{(h)} - \tilde{p}^{(h)}\right) \\ &= c_h^y - \Delta c_0 \beta^{-h} y^{(h)} + (c_0^b - c_0^y - \Delta c_0)\beta^{-h}\tilde{p}^{(h)} \end{aligned}$$

is attainable. The cost of the change in future consumption is

$$\pi(c^{m*} - c^y) = \Delta c_0 S\hat{\pi}(y - \tilde{p}) + (c_0^b - c_0^y)S\hat{\pi}(\tilde{p})$$

In order to satisfy the investor's budget constraint, the date 0 consumption must adjust s.t. $\Delta c_0 + \pi(c^{m*} - c^y) = 0$. Thus

$$\Delta c_0 = -\frac{(c_0^b - c_0^y)S\hat{\pi}(\tilde{p})}{1 + S\hat{\pi}(y - \tilde{p})}$$

The resulting utility (normalized by $|u''|$) is

$$\begin{aligned} U_{m^*} &= -\frac{(c_0^b - c_0^y - \Delta c_0)^2}{2}\left(1 + S\,||y - \tilde{p}||^2\right) \\ &= -\frac{(c_0^b - c_0^y)^2}{2}\left(1 + \frac{S\hat{\pi}(\tilde{p})}{1 + S\hat{\pi}(y - \tilde{p})}\right)^2\left(1 + S\,||y - \tilde{p}||^2\right) \end{aligned}$$

Taking the difference $U_{m^*} - U_y$ yields the result.

The proof of the second part is trivial. We have

$$||y - m^*|| = ||y - (y - \tilde{p})|| = ||\tilde{p}|| = \delta_{\theta,\mathcal{H},y}$$

By definition, $\delta_{\theta,\mathcal{H},y}$ is the minimum distance between y and any admissible pricing kernel. ■

Proposition 4 tells us that the normalized utility gain is linear and increasing in the multi-period distance metric. Furthermore, the utility gain is lower if the investor is more risk averse. This makes sense as a more risk averse investor generally would take smaller positions to offset the mis-specification.

We also see that the utility gain is related to $\hat{\pi}(\tilde{p})$, which is the price of the payoff that pays $\theta_h \tilde{p}^{(h)}$ at every horizon $h \in \mathcal{H}$. This term comes about due to inter-temporal shifts in consumption. In particular, as returns are mis-priced, date 0 consumption is not optimal either. The extent to which date 0 consumption needs to change depends on the price of $\tilde{p}$. In the special case that $\tilde{p}$ is an excess return so that $\hat{\pi}(\tilde{p}) = 0$, date 0 consumption remains the same. Note that in tat special case, going from model y to m^* only shifts consumption across states and not inter-temporally. Furthermore, the normalized utility gain becomes

$$\frac{U_{m^*} - U_y}{-u''(c_0)} = \frac{1}{2} a^{-2} S \delta^2_{\theta,\mathcal{H},y}$$

The normalized utility gain does not have a familiar economic interpretation. The following corollary therefore provides an equivalent increase in date 0 consumption to gauge the importance of model misspecification.

Corollary 1. *If* $1 \geq S\delta^2_{\theta,\mathcal{H},y} - \frac{S}{1+S||y-\tilde{p}||^2_{\theta,\mathcal{H}}} S\hat{\pi}(\tilde{p})^2$*, there exists an equivalent increase in date 0 consumption s.t.* $u(c_0^y + ce) + \mathbb{E}\sum_{h\in\mathcal{H}} \beta^h u(c_h^y) = U_{m^*}$*. The equivalent consumption increase is given by*

$$\frac{ce}{c_0^y} = \frac{1 - \sqrt{1 - S\delta^2_{\theta,\mathcal{H},y} + \frac{S}{1+S||y-\tilde{p}||^2_{\theta,\mathcal{H}}} S\hat{\pi}(\tilde{p})^2}}{\gamma}$$

where $\gamma \equiv \frac{-u''(c_0^y)c_0^y}{u'(c_0^y)}$ *is the coefficient of relative risk aversion.*

The condition that $1 \geq S\delta^2_{\theta,\mathcal{H},y} - \frac{S}{1+S||y-\tilde{p}||^2_{\theta,\mathcal{H}}} S\hat{\pi}(\tilde{p})^2$ is due to utility derived from date 0 consumption is bounded above by $u(c_0^b)$. Consuming more beyond the bliss point lowers utility. Thus, there exists an equivalent consumption increase at date 0 only if the misspecification of y is not so large that the consumer cannot realize all the potential utility gain at a single date. Due to the blisspoints, the consumer would need several dates to absorb the gain if it is large. As S grows with the number of horizons, the condition is less likely to be satisfied if the number of horizons is large.

4.2 Excess Return Mis-pricing

Suppose the investor instead considers the pricing of excess returns. In particular, let $\bar{R}_b^{(h)}$ denote the last $N_h - 1$ rows of $R_b^{(h)}$ and $R_{b,1}^{(h)}$ be the first row. We can then generate a vector of excess returns $R_b^{e,(h)} = \bar{R}_b^{(h)} - R_{b,1}^{(h)}\mathbf{1}$. The prices assigned to these excess returns by $y^{(h)}$ is then

$$\mathbb{E}\big(y^{(h)} R_b^{e,(h)}\big) = \bar{\alpha}^{(h)} - \alpha_1^{(h)}\mathbf{1} \equiv \alpha^{e(h)} \tag{43}$$

It then follows that there exists a set of trades with price 0 at date 0 that improves the investor's expected utility. Proposition 5 states that the normalized utility gain from these trades are proportional to the multi-period excess return distance metric

Proposition 5. *Consider a quadratic utility investor who consumes at dates* 0 *and* $\mathcal{H}$*. Assume his IMRS,* M^I*, equals y at all* $h \in \mathcal{H}$*. If* $\mathbb{E}(y^{(h)} R^{e,(h)}) \neq 0$ *for some* $h \in \mathcal{H}$*, then:*

1. *There exists an excess return pricing kernel* m^e *s.t. the consumption stream implied by* m^e *is attainable. Furthermore, the trades required to attain the implied consumption stream costs 0 at*

date 0.

2. *The distance between y and m^e is $\delta^e_{\theta,\mathcal{H},y}$.*

3. *The utility gain normalized by $|u''|$ for a quadratic utility investor from using the excess return pricing kernel m^e instead of the mis-specified model y is*

$$\frac{U_{m^e} - U_y}{-u''(c_0)} = \frac{1}{2} a^{-2} S (\delta^e_{\theta,\mathcal{H},y})^2$$

where $S \equiv \sum_{h \in \mathcal{H}} \beta^{-h}$, $\theta_h \equiv \frac{\beta^{-h}}{S}$, and $a \equiv -\frac{u''(c_0^y)}{u'(c_0^y)}$ is the coefficient of absolute risk aversion evaluated at c_0^y.

4. *If $1 \geq S\ (\delta^e_{\theta,\mathcal{H},y})^2$, there exists an equivalent increase in date 0 consumption, ce_0, that gives the same utility as going from the model implied consumption c_h^y to c_h^{e*} for all $h \in \mathcal{H}$. ce_0 is given by*

$$\frac{ce_0}{c_0^y} = \frac{1 - \sqrt{1 - S\ (\delta^e_{\theta,\mathcal{H},y})^2}}{\gamma}$$

where $\gamma \equiv -\frac{c_0^y u''(c_0^y)}{u'(c_0^y)}$ is the coefficient of relative risk aversion evaluated at c_0^y.

Proof. Part 1. An admissible pricing kernel for excess returns is $m^e = y - \tilde{p}^e$, where $\tilde{p}^e \in \mathcal{P}^e$. The implied consumption stream is

$$\begin{aligned} c_h^{m^e} &= c_h^b - \beta^{-h}(c_0^b - c_0^y)(y^{(h)} - \tilde{p}^{e(h)}) \\ &= c_h^y + (c_0^b - c_0^y) S \theta_h \tilde{p}^{e(h)} \end{aligned}$$

Since $\tilde{p}^{e(h)}$ is an excess return, it is tradeable and costs 0 at date 0. As a consequence, the investor's budget constraint is satisfied without adjusting date 0 consumption.

Part 2. The distance between y and $m^e = y - \tilde{p}^e$ is

$$||y - m^e||_{\theta,\mathcal{H}} = ||\tilde{p}^e||_{\theta,\mathcal{H}} = \delta^e_{\theta,\mathcal{H},y}$$

Part 3. The expected utility from using m^e is

$$\begin{aligned} U_{m^e} &\equiv u(c_0^y) - \frac{1}{2}(c_0^b - c_0^y)^2 S\ ||m^e||^2_{\theta,\mathcal{H}} \\ &= u(c_0^y) - \frac{1}{2}(c_0^b - c_0^y)^2 S\ ||y - \tilde{p}^e||^2_{\theta,\mathcal{H}} \\ &= u(c_0^y) - \frac{1}{2}(c_0^b - c_0^y)^2 S\ \left(||y||^2_{\theta,\mathcal{H}} + ||\tilde{p}^e||^2_{\theta,\mathcal{H}} - 2\hat{\mathbb{E}}_{\theta,\mathcal{H}}(y\tilde{p}^e)\right) \\ &= u(c_0^y) - \frac{1}{2}(c_0^b - c_0^y)^2 S\ \left(||y||^2_{\theta,\mathcal{H}} - ||\tilde{p}^e||^2_{\theta,\mathcal{H}}\right) \end{aligned}$$

Taking the difference between U_{m^e} and U_y yields the result.

Part 4. The increase in date 0 consumption ce_0, keeping future consumption at c_h^y, that yields the same utility gain is (if it exists) given by

$$\frac{u(c_0^y + ce_0) - u(c_0^y)}{-u''(c_0^y)} = \frac{1}{2}a^{-2}S(\delta_{\theta,\mathcal{H},y}^e)^2 \Leftrightarrow$$
$$0 = \frac{ce_0^2}{2} - a^{-1}\ ce_0 + \frac{1}{2}a^{-2}S(\delta_{\theta,\mathcal{H},y}^e)^2$$

The solutions are given by

$$ce_0 = a^{-1} \pm \sqrt{a^{-2} - a^{-2}S(\delta_{\theta,\mathcal{H},y}^e)^2} = \frac{1 \pm \sqrt{1 - S(\delta_{\theta,\mathcal{H},y}^e)^2}}{a}$$

Note that the negative root is the solution we are interested in as the positive root gives a consumption that overshoots the bliss point. The equivalent date 0 consumption is well-defined if $1 \geq S(\delta_{\theta,\mathcal{H},y}^e)^2$, which is the stated assumption. Dividing both sides by c_0^y and using $\gamma = a \cdot c_0^y$ gives the result. ■

Proposition 5 tells us that the normalized utility gain from exploiting the mis-pricing of risk is proportional to the excess return distance metric. Normalizing the utility gain by $-u''$ gives us a measure of utility change that is invariant preference-preserving transformations of u. Furthermore, the theorem tells us that as long as the model mis-specification is not too large, there exists an increase in date 0 consumption, keeping all future consumption fixed, that delivers the same utility gain as exploiting the mis-pricing. The requirement that the mis-specification not be too large, is caused by the peculiar form of the utility - increasing date 0 consumption beyond the bliss point does not deliver extra utility.

Corollary 2. *Iff* $1 > S\ (\delta_{\theta,\mathcal{H},y}^e)^2$*, then:*

1. $\frac{ce_0}{c_0^y}$ *is strictly increasing in the distance metric* $\delta_{\theta,\mathcal{H},y}^e$
2. $\frac{ce_0}{c_0^y}$ *is strictly decreasing in the coefficient of relative risk aversion* γ

Proof. For part 1, note that $S \equiv \sum_{h\in\mathcal{H}} \beta^{-h} > 0$. The result then follows immediately from Proposition 5. For the second part, note $1 > S\ (\delta_{\theta,\mathcal{H},y}^e)^2$ and $S > 0$ implies $\sqrt{1 - S\ (\delta_{\theta,\mathcal{H},y}^e)^2} \in (0,1)$ and thus $1 - \sqrt{1 - S\ (\delta_{\theta,\mathcal{H},y}^e)^2} > 0$. Since the numerator does not depend on γ, the result follows immediately. ■

5 A model economy

In this section I will assume the base assets for $\mathcal{P}^{(h)}$ are the h period returns on N assets. Let r_{t+1} denote an $N \times 1$ vector of log-returns on the base assets and $r_{f,t}$ be the risk-free rate

$$r_{t+1} = r_{f,t} + \mu_t^e - \frac{1}{2}diag(\sigma_{\nu,t}\sigma_{\nu,t}^\top) - \frac{1}{2}diag(\sigma_{\varepsilon,t}\sigma_{\varepsilon,t}^\top) + \sigma_{\nu,t}\nu_{t+1} + \sigma_{\varepsilon,t}\varepsilon_{t+1} \tag{44}$$

where $diag(X)$ denotes a matrix whose diagonal entries equals that of X and the off-diagonal entries are zero. I assume that ν and ε are independent standard normal random variables. Furthermore, I assume that ν and ε are $K \times 1$ and $N \times 1$ vectors respectively. Let $r_{i,t+1}$, $\sigma_{i,\nu,t}$ and $\sigma_{i,\varepsilon,t}$ denote the i-th rows of the corresponding vector/matrix. Then $diag(\sigma_{\nu,t}\sigma_{\nu,t}^{\top})_{i,i} = \sigma_{i,\nu,t}\sigma_{i,\nu,t}^{\top}$ and $diag(\sigma_{\varepsilon,t}\sigma_{\varepsilon,t}^{\top})_{i,i} = \sigma_{i,\varepsilon,t}\sigma_{i,\varepsilon,t}^{\top}$ are the conditional variances of $r_{i,t+1}$ associated with the ν and ε shocks respectively.

Suppose market prices are determined by the true log pricing kernel given by

$$m_{t+1} = -r_{f,t} - \frac{\lambda_t^{\top}\lambda_t}{2} - \lambda_t^{\top}\nu_{t+1}$$

where λ_t is a $K \times 1$ vector of risk-prices. Note that m_{t+1} prices the risk-free rate by construction. Since it also prices risky returns, we have that

$$\mu_t^e = \sigma_{\nu,t}\lambda_t \tag{45}$$

Letting $R^{(h)}$ and $M^{(h)}$ denote the h-period returns and SDF, it follows from the law of iterated expectations that

$$\mathbf{1} = \mathbb{E}_t(M^h R^{(h)}), \qquad \forall h$$

Now suppose we have a candidate discount factor of the following form

$$y_{t+1} = -r_{f,t} - \frac{\xi_{\nu,t}^{\top}\xi_{\nu,t}}{2} - \frac{\xi_{\varepsilon,t}^{\top}\xi_{\varepsilon,t}}{2} - \xi_{\nu,t}^{\top}\nu_{t+1} - \xi_{\varepsilon,t}^{\top}\varepsilon_{t+1} \tag{46}$$

which also prices the risk-free rate by construction. However, y_{t+1} is potentially mis-specified along two dimensions. First, it might have the wrong risk-prices for the priced shocks ν, i.e. $\xi_{\nu,t} \neq \lambda_t$. Second, it might falsely assign a non-zero price of risk to the ε shocks, i.e. $\xi_{\varepsilon,t} \neq 0$. A special case of the candidate discount factor is $y_{t+1} = a_t - \omega^{\top} r_{t+1}$.

Let us first consider the conditional price assigned by y to 1-period returns.

$$\begin{aligned} q_{y,t}(R^{(1)}) &\equiv \mathbb{E}_t(e^{y_{t+1}+r_{t+1}}) = e^{\mu_t^e - \sigma_{\nu,t}\xi_{\nu,t} - \sigma_{\varepsilon,t}\xi_{\varepsilon,t}} \\ &= e^{\sigma_{\nu,t}(\lambda_t - \xi_{\nu,t}) - \sigma_{\varepsilon,t}\xi_{\varepsilon,t}} \end{aligned} \tag{47}$$

Clearly, $q_{y,t}(R_{t+1})$ is in general not equal 1.

In order to price multi-horizon returns and take unconditional expectations, we must impose more structure. I will therefore consider a simple example economy, with three different candidate models y.

5.1 Example Economy 1

To keep things simple, let $\sigma_{\nu,t} = \sigma_\nu$, $\sigma_{\varepsilon,t} = \sigma_\varepsilon$ and $r_{f,t} = r_f$. Let the risk prices follow an AR(1) with independent standard normal shocks, i.e.

$$\lambda_{k,t+1} = (1-\varphi_k)\lambda_k + \varphi_k \lambda_{k,t} + \varrho_k \eta_{k,t+1}$$

where $\varphi_k \in (-1,1)$. Without loss of generality, we assume $\lambda_k > 0$.

Some further notation will prove useful. Throughout, I will for a matrix X, take $X(k)$ to mean the k-th column of X. Furthermore, X^k raises every element of X to the power of k. A fraction with an $L \times M$ matrix A in the numerator and an $L \times 1$ vector in the denominator divides each row of A by the corresponding row of v. Similarly, $\circ$ denotes element-wise multiplication. If v is an $L \times 1$ vector and A an $L \times M$ matrix, $(v \circ A)_{i,j} \equiv (A \circ v)_{i,j} \equiv A_{i,j} v_i$, $i = 1, \ldots, L$, $j = 1, \ldots, M$. Similarly, if v is an $1 \times M$ vector and A an $L \times M$ matrix, $(v \circ A)_{i,j} \equiv (A \circ v)_{i,j} \equiv A_{i,j} v_j$, $i = 1, \ldots, L$, $j = 1, \ldots, M$. Finally, for two $L \times M$ matrices A and B, $(A \circ B)_{i,j} \equiv (B \circ A)_{i,j} \equiv A_{i,j} B_{i,j}$, $i = 1, \ldots, L$, $j = 1, \ldots, M$. The covariance matrix of λ_t can then be written

$$\Sigma_{\lambda,k,k} = diag\Big(\frac{\varrho^2}{1-\varphi^2}\Big)$$

The conditional and unconditional second moment matrices of 1-period returns is

$$\begin{aligned}
\mathbb{E}_t(R^{(1)} R^{(1)\top}) &= exp\Big\{2r_f + \sigma_\nu \lambda_t \mathbf{1}^\top + \mathbf{1}\lambda_t^\top \sigma_\nu^\top + 2\sigma_\nu \sigma_\nu^\top + 2\sigma_\varepsilon \sigma_\varepsilon^\top\Big\} \\
\mathbb{E}(R^{(1)} R^{(1)\top}) &= exp\Big\{2r_f + \sigma_\nu \lambda \mathbf{1}^\top + \mathbf{1}\lambda^\top \sigma_\nu^\top + 2\sigma_\nu \sigma_\nu^\top + 2\sigma_\varepsilon \sigma_\varepsilon^\top + 2\sigma_\nu \Sigma_\lambda \sigma_\nu^\top\Big\} \\
&\equiv A \circ exp\{2\sigma_\nu \Sigma_\lambda \sigma_\nu^\top\}
\end{aligned}$$

In the multi-period case, the conditional second moment is

$$\mathbb{E}_t(R^{(h)} R^{(h)\top}) = A^h \circ exp\Big\{2\sigma_\nu \Sigma_\eta^{(h)} \sigma_\nu^\top + \sigma_\nu \Big(\frac{1-\varphi^h}{1-\varphi} \circ (\lambda_t - \lambda)\Big)\mathbf{1}^\top + \mathbf{1}\Big(\frac{1-\varphi^h}{1-\varphi} \circ (\lambda_t - \lambda)\Big)^\top \sigma_\nu^\top\Big\}$$

where $\Sigma_\eta^{(h)}$ is the conditional covariance matrix of $\sum_{i=0}^{h-1} \lambda_{t+i}$, which can be expressed analytically as follows

$$\Sigma_\eta^{(h)} \equiv diag\left(\Big(\frac{\varrho}{1-\varphi}\Big)^2 \circ \Big(h - 1 - 2\frac{\varphi - \varphi^h}{1-\varphi} + \frac{\varphi^2 - \varphi^{2h}}{1-\varphi^2}\Big)\right)$$

Note that $\Sigma_\eta^{(1)} = \mathbf{0}$. Thus

$$\mathbb{E}(R_{t,t+h} R_{t,t+h}^\top) = A^h \circ exp\Big\{2\sigma_\nu \Sigma_\eta^{(h)} \sigma_\nu^\top + 2\sigma_\nu \Sigma_\lambda^{(h)} \sigma_\nu^\top\Big\} \tag{48}$$

$\Sigma_{\lambda}^{(h)}$ is the covariance matrix of $\mathbb{E}_t \sum_{i=0}^{h-1} \lambda_{t+i}$, which equals

$$\Sigma_{\lambda}^{(h)} \equiv \Big(\frac{1-\varphi^h}{1-\varphi}\Big)^2 \circ \Sigma_{\lambda}$$

Thus, letting $h=1$, $\Sigma_{\lambda}^{(1)} = \Sigma_{\lambda}$ as expected. Let us now turn our attention towards the candidate pricing kernels. Note that $\Sigma_{\lambda}^{(h)} + \Sigma_{\eta}^{(h)} = \mathbb{V}(\sum_{i=0}^{h-1} \lambda_{t+i})$ is the unconditional covariance matrix of $\sum_{i=0}^{h-1} \lambda_{t+i}$.

Candidate Model 1

Let sub-script v_{-K} denote the first $K-1$ rows of v. Consider the model

$$y_{1,t+1} = -r_{f,t} - \frac{\lambda_{-K,t}^{\top}\lambda_{-K,t}}{2} - \lambda_{-K,t}^{\top}\nu_{-K,t+1} \tag{49}$$

which prices risks associated with the $K-1$ first ν-shocks correctly, but falsely claims that ν_K-shocks carries zero risk-price. Note that in this case $\xi_{\varepsilon,t}=0$. Using (49) in (47) gives

$$q_{y_1,t}(R^{(1)}) = e^{\sigma_\nu(K)\lambda_{K,t}}$$

Clearly, the extent of conditional mis-pricing depends on how costly it is to bear ν_K risk and how exposed the assets are to this risk, i.e. how "big" $\sigma_\nu(K)$ is.

Since $\lambda_{K,t} \sim N(\lambda_K, \frac{\varrho_K^2}{1-\varphi_K^2})$ the average price assigned by y_1 to 1-period returns is

$$q_{y_1}(R^{(1)}) = e^{\sigma_\nu(K)\lambda_K + \frac{1}{2}\sigma_\nu(K)^2\Sigma_{\lambda,K,K}}$$

For the average prices assigned by y_1, it also matters how variable the K-th risk price is - a more variable risk-price implies that y_1 assigns a higher price to assets with a non-zero exposure. As expected, y_1 considers any asset with positive loading on ν_K undervalued. Or equivalently, y_1 overvalues assets exposed to ν_K. The intuition is straightforward - investors require compensation for bearing ν_K risk, but y_1 assumes no such compensation is required and therefore value exposed assets higher.

The conditional price assigned to multi-horizon returns by y_1 is

$$q_{y_1,t}(R^{(h)}) = exp\Big\{h\sigma_\nu(K)\lambda_K + \sigma_\nu(K)\frac{1-\varphi_K^h}{1-\varphi_K}(\lambda_{K,t}-\lambda_K) + \frac{1}{2}\sigma_\nu(K)^2\Sigma_{\eta,K,K}^{(h)}\Big\} \tag{50}$$

The terms in (50) have straightforward interpretations. First, $\sigma_\nu(2)\lambda_2$ is the average log-risk premium coming from compensation to ν_K risk. The average compensation scales up linearly with horizon. The second term, $\sigma_\nu(K)(\lambda_{K,t}-\lambda_K)$ is the extra compensation coming from the current risk price being higher or lower than average. Since the risk price revert back towards its mean in the long run, the term scales up less than linearly with horizon. The final term is the conditional variance of $\sum_{s=0}^{h-1}\sigma_\nu(2)\lambda_{2,t+s}$.

Taking the unconditional expectation of (50) gives us the following average price

$$q_{y_1}(R^{(h)}) = exp\Big\{h\sigma_\nu(K)\lambda_K + \frac{1}{2}\sigma_\nu(K)^2\Sigma^{(h)}_{\lambda,K,K} + \frac{1}{2}\sigma_\nu(K)^2\Sigma^{(h)}_{\eta,K,K}\Big\} \tag{51}$$

Candidate Model 2

The second candidate model correctly identifies all the priced shocks, but mistakenly assumes the corresponding risk-prices are constant, i.e.

$$y_{2,t+1} = -r_f - \frac{\xi_\nu^\top \xi_\nu}{2} - \xi_\nu^\top \nu_{t+1}$$

y_2 therefore prices 1-period returns as follows

$$q_{y_2,t}(R^{(1)}) = e^{\sigma_\nu(\lambda_t - \xi_\nu)}$$
$$q_{y_2}(R^{(1)}) = e^{\sigma_\nu(\lambda - \xi_\nu) + \frac{1}{2}\sigma_\nu^2\Sigma_\lambda \mathbf{1}}$$

If $\xi_\nu = \lambda$, i.e. the assumed risk price equals the mean risk price, assets with non-zero exposure to ν will be overvalued by y_2. Alternatively, setting ξ_ν sufficiently larger than λ will cause the returns to be undervalued. In the special case $K = N$, i.e. there are the same number of priced and un-priced shocks, y_2 could price the 1-period return perfectly (on average) if $\xi_\nu = \lambda + \frac{1}{2}\sigma_\nu^{-1}\sigma_\nu^2\Sigma_\lambda$. However, even in this case, the multi-period returns would be mis-priced

$$q_{y_2,t}(R^{(h)}) = exp\Big\{h\sigma_\nu(\lambda - \xi_\nu) + \sigma_\nu\Big(\frac{1-\varphi^h}{1-\varphi} \circ (\lambda_t - \lambda)\Big) + \frac{1}{2}\sigma_\nu^2\Sigma_\eta^{(h)}\mathbf{1}\Big\}$$
$$q_{y_2}(R^{(h)}) = exp\Big\{h\sigma_\nu(\lambda - \xi_\nu) + \frac{1}{2}\sigma_\nu^2\Big(\Sigma_\lambda^{(h)} + \Sigma_\eta^{(h)}\Big)\mathbf{1}\Big\} \tag{52}$$

we see that even if the 1-period return is undervalued by y_2, the multi-period return may be overvalued if the price of risk is sufficiently persistent and variable.

Candidate Model 3

The final candidate model I will consider is

$$y_{3,t+1} = -r_f - \frac{\xi_\varepsilon^\top \xi_\varepsilon}{2} - \xi_\varepsilon^\top \varepsilon_{t+1} \tag{53}$$

In other words, the candidate model wrongly assumes ε is priced, while failing to recognize that ν is priced. Despite appearances, we could easily fail to reject this model if we only considered 1-period returns. To see why, note that the prices assigned by y_3 is now

$$
\begin{aligned}
q_{y_3,t}(R^{(1)}) &= e^{\sigma_\nu \lambda_t - \sigma_\varepsilon \xi_\varepsilon} \\
q_{y_3}(R^{(1)}) &= e^{\sigma_\nu \lambda + \frac{1}{2}\sigma_\nu^2 \Sigma_\lambda \mathbf{1} - \sigma_\varepsilon \xi_\varepsilon}
\end{aligned} \tag{54}
$$

Thus, assuming σ_ε is of full rank, we can price 1-period returns perfectly on average if we choose

$$
\xi_\varepsilon = \sigma_\varepsilon^{-1}\Big(\sigma_\nu \lambda + \frac{1}{2}\sigma_\nu^2 \Sigma_\lambda \mathbf{1}\Big) \tag{55}
$$

Rewriting (55) as

$$
\xi_\varepsilon = (\sigma_\varepsilon^\top \sigma_\varepsilon)^{-1}\sigma_\varepsilon^\top \Big(\sigma_\nu \lambda + \frac{1}{2}\sigma_\nu^2 \Sigma_\lambda \mathbf{1}\Big)
$$

we see that ξ_ε are regression coefficients resulting from regressing log expected 1-period excess returns $log\mathbb{E}\Big(\frac{R}{R_f}\Big) = \sigma_\nu \lambda + \frac{1}{2}\sigma_\nu^2 \Sigma_\lambda \mathbf{1}$ on the columns of σ_ε. Since we assumed σ_ε was invertible, the columns are linearly independent. The regression therefore has as many independent explanatory variables as observations of the dependent variable, resulting in a perfect fit. Using the risk-prices in (55) gives us the following 1-period prices

$$
\begin{aligned}
q_{y_3,t}(R^{(1)}) &= e^{\sigma_\nu(\lambda_t - \lambda) - \frac{1}{2}\sigma_\nu^2 \Sigma_\lambda \mathbf{1}} \\
q_{y_3}(R^{(1)}) &= \mathbf{1}
\end{aligned}
$$

With general ξ_ε, the multi-period prices becomes

$$
\begin{aligned}
q_{y_3,t}(R^{(h)}) &= exp\Big\{h(\sigma_\nu \lambda - \sigma_\varepsilon \xi_\varepsilon) + \sigma_\nu\Big(\frac{1-\varphi^h}{1-\varphi} \circ (\lambda_t - \lambda)\Big) + \frac{1}{2}\sigma_\nu^2 \Sigma_\eta^{(h)} \mathbf{1}\Big\} \\
q_{y_3}(R^{(h)}) &= exp\Big\{h(\sigma_\nu \lambda - \sigma_\varepsilon \xi_\varepsilon) + \frac{1}{2}\sigma_\nu^2\Big(\Sigma_\lambda^{(h)} + \Sigma_\eta^{(h)}\Big)\mathbf{1}\Big\}
\end{aligned} \tag{56}
$$

which in the special case of ξ_ε given by (55) becomes

$$
\begin{aligned}
q_{y_3,t}(R_{t,t+h}) &= exp\Big\{-\frac{h}{2}\sigma_\nu^2 \Sigma_\lambda \mathbf{1} + \sigma_\nu\Big(\frac{1-\varphi^h}{1-\varphi} \circ (\lambda_t - \lambda)\Big) + \frac{1}{2}\sigma_\nu^2 \Sigma_\eta^{(h)} \mathbf{1}\Big\} \\
q_{y_3}(R_{t,t+h}) &= exp\Big\{-\frac{h}{2}\sigma_\nu^2 \Sigma_\lambda \mathbf{1} + \frac{1}{2}\sigma_\nu^2\Big(\Sigma_\lambda^{(h)} + \Sigma_\eta^{(h)}\Big)\mathbf{1}\Big\}
\end{aligned}
$$

As can be seen from (56), estimating the candidate pricing kernel to have zero pricing error for 1-period returns might lead to high pricing errors at longer horizons. In fact, (56) suggests that

$$
\hat{\xi}_\varepsilon = \sigma_\varepsilon^{-1}\Big[\sigma_\nu \lambda + \frac{1}{2h}\sigma_\nu^2\Big(\Sigma_\lambda^{(h)} + \Sigma_\eta^{(h)}\Big)\mathbf{1}\Big] \tag{57}
$$

is an equally sensible choice for ξ_ε as this causes h-period returns to be priced correctly on average. The two are different to the extent that Σ_λ is different from $\frac{\Sigma_\lambda^{(h)}+\Sigma_\eta^{(h)}}{h}$. Clearly, $\hat{\xi}_\varepsilon$ can then be interpreted as the coefficients resulting from regressing $\sigma_\nu\lambda + \frac{1}{2h}\sigma_\nu^2\Big(\Sigma_\lambda^{(h)} + \Sigma_\eta^{(h)}\Big)\mathbf{1}$ on σ_ε. Since $log\mathbb{E}\Big(\frac{R^{(h)}}{R_f}\Big)^{\frac{1}{h}} = \sigma_\nu\lambda + \frac{1}{2h}\sigma_\nu^2\Big(\Sigma_\lambda^{(h)} + \Sigma_\eta^{(h)}\Big)\mathbf{1}$, it can be interpreted as the average log expected excess return on $R^{(h)}$ measured per period. Thus, if the expected h-period returns are higher than expected 1-period returns (per period), $\hat{\xi}_\varepsilon$ will tend to be bigger in magnitude than ξ_ε

There is a striking similarity between the expressions in (56) and (52). In fact, for a given set of test assets, we can choose ξ_ε in y_3 s.t. y_2 and y_3 have identical pricing implications for all horizons. However, the two SDFs would price new assets very differently. To see this, note that we can think of $\tilde{\varepsilon_{t+1}} \equiv \xi_\varepsilon^\top\varepsilon_{t+1}$ as a single factor. A new asset would have some exposure to $\tilde{\varepsilon}$ that typically does not line up with its exposure to ν. In general, we would therefore expect the external validity of y_2 to be greater than that of y_3.

5.2 Example Economy 2

The equivalence of pricing implications for y_2 and y_3 is caused by the simplifying assumption that $\sigma_{\nu,t}$ and $\sigma_{\varepsilon,t}$ are constant. It might therefore be of some interest to investigate the pricing implications if we relaxed those assumptions. Let us therefore consider an identical economy and candidate models as in sub-section 5.1 except that

$$\sigma_{\varepsilon,t+1} = \sigma_\varepsilon z_t$$

$$z_{t+1} = 1 + b(z_t - 1) + du_{t+1}$$

where u_{t+1} is a standard normal shock and $b \in (-1, 1)$ denotes the persistence of z_t. In other words, conditional volatility for every asset-shock pair i, j is the corresponding mean volatility scaled by z_t. While this might not be the most realistic volatility process, the main point is to illustrate the effect of time-varying volatility on pricing.

The second moment matrix of h-period returns is in this case (see Appendix B for details)

$$\mathbb{E}(R_{t,t+h}R_{t,t+h}^\top) = \hat{A}^h \circ exp\Big\{2\sigma_\nu\Sigma_\eta^{(h)}\sigma_\nu^\top + 2\sigma_\nu\Sigma_\lambda^{(h)}\sigma_\nu^\top + 2\Sigma_\varepsilon^{(h)}\Big\} \tag{58}$$

where

$$\hat{A} \equiv exp\Big\{2r_f + \sigma_\nu\lambda\mathbf{1}^\top + \mathbf{1}\lambda^\top\sigma_\nu^\top + 2\sigma_\nu\sigma_\nu^\top\Big\}$$

The prices assigned by candidate models y_1 and y_2 are the same as in the previous section. For y_3, the 1-period prices are now

$$
\begin{aligned}
q_{y_3,t}(R^{(1)}) &= e^{\sigma_\nu \lambda_t - \sigma_\varepsilon \xi_\varepsilon z_t} \\
q_{y_3}(R^{(1)}) &= e^{\sigma_\nu \lambda + \frac{1}{2}\sigma_\nu^2 \Sigma_\lambda \mathbf{1} - \sigma_\varepsilon \xi_\varepsilon + \frac{1}{2}(\sigma_\varepsilon \xi_\varepsilon)^2 \sigma_z^2}
\end{aligned}
$$

where $\sigma_z^2 \equiv \frac{d^2}{1-b^2}$ is the variance of z_t. The corresponding multi-period prices are

$$
\begin{aligned}
q_{y_3,t}(R^{(h)}) = exp\Big\{&h(\sigma_\nu \lambda - \sigma_\varepsilon \xi_\varepsilon) + \sigma_\nu \Big(\frac{1-\varphi^h}{1-\varphi} \circ (\lambda_t - \lambda)\Big) + \frac{1}{2}\sigma_\nu^2 \Sigma_\eta^{(h)} \mathbf{1} \\
&- \sigma_\varepsilon \xi_\varepsilon \frac{1-b^h}{1-b}(z_t - 1) + \frac{1}{2}(\sigma_\varepsilon \xi_\varepsilon)^2 \sigma_u^{(h)2} \Big\} \\
q_{y_3}(R^{(h)}) = exp\Big\{&h(\sigma_\nu \lambda - \sigma_\varepsilon \xi_\varepsilon) + \frac{1}{2}\sigma_\nu^2 \Big(\Sigma_\lambda^{(h)} + \Sigma_\eta^{(h)}\Big)\mathbf{1} + \frac{1}{2}(\sigma_\varepsilon \xi_\varepsilon)^2 \Big(\sigma_z^{(h)2} + \sigma_u^{(h)2}\Big)\Big\}
\end{aligned} \tag{59}
$$

where

$$
\begin{aligned}
\sigma_u^{(h)} &\equiv \sqrt{h - 1 - 2\frac{b-b^h}{1-b} + \frac{b^2 - b^{2h}}{1-b^2}}\frac{d}{1-b} \\
\sigma_z^{(h)} &\equiv \frac{1-b^h}{1-b}\sigma_z
\end{aligned}
$$

We can interpret $\sigma_u^{(h)2}$ as the conditional variance of $\sum_{i=0}^{h-1} z_{t+i}$ and $\sigma_z^{(h)2}$ as the variance of $\mathbb{E}_t \sum_{i=0}^{h-1} z_{t+i}$. $\sigma_z^{(h)2} + \sigma_u^{(h)2}$ is then the unconditional variance of $\sum_{i=0}^{h-1} z_{t+i}$. From (59) we see that stochastic volatility for the unpriced shocks introduces another source for mis-pricing. For this reason, the pricing implications of y_2 and y_3 are no longer equivalent. The size of the additional mis-pricing depends on both how volatile and persistent idiosyncratic volatility is.

5.3 Numerical Results

In this section I report numerical results for for the multi-period distance metrics associated with example economies 1 and 2. I assume the true log pricing kernel m is a two-factor model, and that factor 1 with a mean risk price of $\lambda_1 = \frac{1}{\sqrt{12}}$, compared to $\lambda_2 = \frac{0.05}{\sqrt{12}}$, is by far the most important on average. Furthermore, the risk-prices are assumed to be fairly volatile - the standard deviation of $\lambda_{1,t}$ is equal to the mean λ_1, and the standard deviation of $\lambda_{2,t}$ is equal to 10 times its mean λ_2. Since the volatility of factor 2 risk price is very high, it is an important pricing factor at some points in time. The number of test assets at every horizon is 50. Each asset's loading on each risk source is drawn independently from a uniform distribution.

The first candidate model captures the main source of priced risk perfectly, but it overlooks a source of priced risk that is unimportant on average. Candidate model 2 on the other hand, includes both risk-sources, but fails to capture the volatility of the prices of risk. In candidate model 2, we have to choose the model-implied price of risk parameters ξ_ν. In this section, I assume $\xi_\nu = \lambda$, i.e. the model-implied risk-price is correct on average. From an intuitive standpoint, the types of mis-specification captured in these models seem very plausible in empirical models.

The third candidate model can be thought of as a data-mining exercise - we have N free parameters and N assets. I consider two versions of this model. The first chooses parameters, assuming constant idiosyncratic volatility, to price 1-period returns unconditionally and the second prices 12-period returns unconditionally. Note that these models are completely uncorrelated with the true log-pricing kernel m. However, by design the covariance between asset returns and these candidate models (in levels) line up perfectly with the covariance of asset returns and e^m at the relevant horizon. As a consequence, these models would typically perform badly if we attempted to price a return that is not in the span of the original N assets.

Figure 1 plots the standard HJ-distance for each candidate model against horizon. Panel (a) gives the results for Example Economy 1 with constant idiosyncratic volatility, while Panel (b) shows the results for Example Economy 2 where idiosyncratic volatility is time-varying.

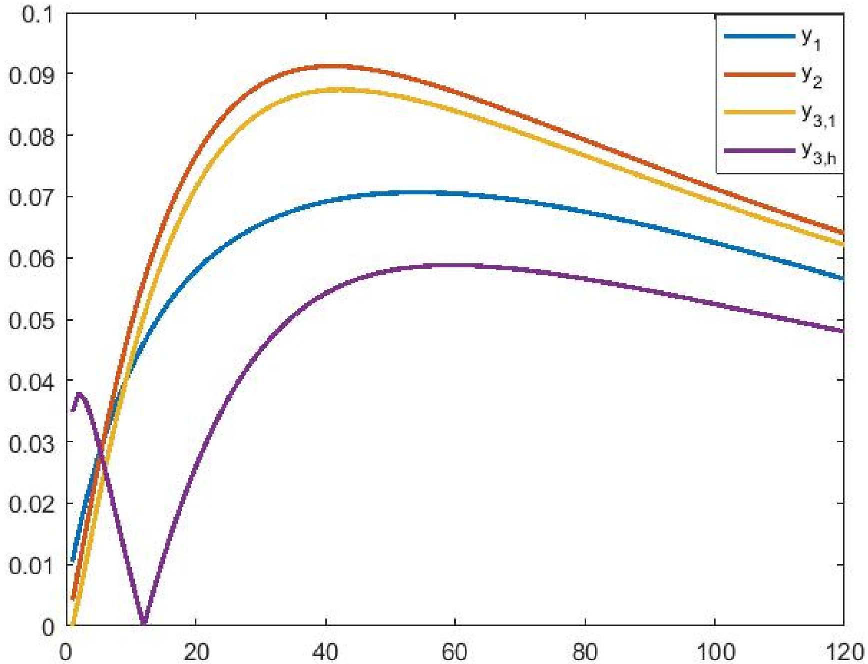
From Panel (a) we see that all models, except the version of model 3 that prices 12 month returns, perform well when looking at 1-period returns. In this case, the HJ-distance is around 1% for candidate model 1 or below for the two other models. It is also interesting to note that if we where to rank models on the basis of 1-period HJ-distance, version 1 of model 3 would be ranked best, followed by model 2, and then model 1. Given that model 3 is the least related to the actual pricing kernel among the candidates, this might be somewhat undesirable.

The version of candidate model 3 that prices 1-period returns unconditionally, gives a 1-period HJ-distance of 0 as expected. However, the HJ-distance grows quickly with horizon peaking right below 0.09 at the 40 month horizon. In other words, the maximally mis-priced payoff has a pricing error of 9% of its norm.

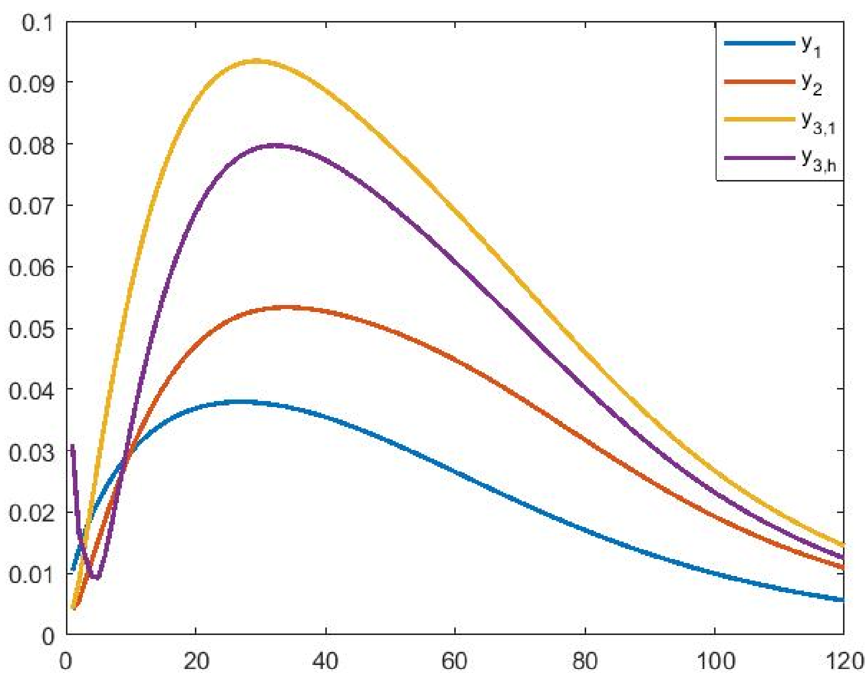
Candidate model 2 does slightly worse than version 1 of model 3 at every horizon. It is therefore clear that the multi-horizon distance metric cannot reverse the ranking between model 2 and 3 in this case.

In the case of candidate model 1, the picture is slightly different. It starts off being ranked below model 2 and version 1 of model 3, but the HJ-distance grows more slowly with horizon causing it to be ranked above both of these models for horizons exceeding 12 months. As a consequence, the multi-period distance metric will rank model 1 above models 2 and version 1 of 3 provided enough weight is placed on longer horizons.

The picture is quite different for the version of the third candidate model that prices 12-month returns. The HJ-distance at the 1 month horizon is quite large at about 0.035, which falls to 0 at the 12 month horizon as expected. For longer horizons, the distance metric grows, peaking right below 0.06 for the 50-60 month horizons. Interestingly, the HJ-distances for this model is below the three other candidates for all horizons exceeding 6 months.



(a)



(b)

Figure 1: This figure plots the HJ-distance for each model vs investment horizon. Panel (a) assumes the constant idiosyncratic volatility economy of subsection 5.1, whereas Panel (b) shows the results for the stochastic idiosyncratic volatility economy of subsection 5.2.

The overall picture from Panel (a) is that the two versions of model 3 performs better than might be expected given that they are uncorrelated with the true pricing kernel. However, as already noted, there is a certain equivalence between candidate models 2 and 3 when volatility is constant. The reason is that both models assume constant risk prices. With constant volatilities, one can always choose the N risk-price parameters in model 3 to match the risk-premiums implied by model 2. With stochastic idiosyncratic volatility, the implied risk premium from model 3 is time-varying, whereas that of model 2 is constant. As a consequence, the equivalence breaks down.

From Panel (b), we see that idiosyncratic volatility changes the rankings and HJ-distance of models quite drastically, particularly at longer horizons. At the 1-period horizon, the metrics are almost identical to those in Panel (a). However, at horizons exceeding 15 months both versions of model 3 performs significantly worse than models 1 and 2. Furthermore, model 1 is now the best performing according to the HJ-distance for all horizons greater than 15 months.

It is worth noting that stochastic idiosyncratic volatility generally *lowers* HJ-distances for model 1 and 2. The reason is straightforward. Idiosyncratic volatility does not affect the pricing errors of basis returns for model 1 and 2. However, it does affect the second-moment matrix of returns, which is in the denominator of the HJ-distance. Positively auto-correlated volatility cause the second moment matrix to grow more with horizon than in the constant volatility case, causing the HJ-distance to fall.

For the two versions of model 3, the pricing errors are generally larger when idiosyncratic volatility is stochastic. Thus, the effect on the HJ-distance is ambiguous and generally depend on what horizon we are looking at.

6 Conclusion

In this paper I propose a distance metric to compare candidate pricing kernels in a multi-horizon setting. The metric measures the distance between a misspecified candidate pricing kernel and the "closest" kernels that prices the test assets correctly. Viewing horizons as collections of states makes the problem analogous to that in Hansen and Jagannathan (1997). As such, my metric is a natural generalization of the standard HJ-distance to a multi-period setting. Indeed, the distance metric can be written as a weighted average of standard HJ-distances across the horizons under consideration. Thus, the multi-period distance is low only if the HJ-distance is low at every horizon.

The multi-period distance metric is closely related to the utility loss of a quadratic utility investor from using a misspecified model. In particular, if the model mis-prices excess returns, the utility gain from exploiting the misspecification is proportional to the excess return distance squared.

A Moment Generating Function

Let v be an $H \times 1$ vector of normally distributed random variables with 0 mean and covariance matrix Γ. Furthermore, let τ_1 denote an $H \times 1$ vector and τ_2 denote an $H \times H$ matrix. Assume

$$\det(I - 2\tau_2\Gamma) > 0$$

and define

$$A \equiv \Gamma\Big(I - 2\tau_2\Gamma\Big)^{-1} \tag{60}$$

The following moment generating function will prove useful

$$\begin{aligned}
\mathbb{E}\Big(e^{\tau_1^\top v + v^\top \tau_2 v}\Big) &= (2\pi)^{-\frac{H}{2}} \det(\Gamma)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{\tau_1^\top v + v^\top \tau_2 v - \frac{1}{2} v^\top \Gamma^{-1} v} dv \\
&= e^{\frac{1}{2}\tau_1^\top A \tau_1} (2\pi)^{-\frac{H}{2}} \det(\Gamma)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(v - A\tau_1)^\top A^{-1}(v - A\tau_1)} dv \\
&= e^{\frac{1}{2}\tau_1^\top A \tau_1} (2\pi)^{-\frac{H}{2}} \det(\Gamma)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} x^\top I^{-1} x} \det(A)^{\frac{1}{2}} dx \\
&= e^{\frac{1}{2}\tau_1^\top A \tau_1} \det(\Gamma)^{-\frac{1}{2}} \det(A)^{\frac{1}{2}} \int_{-\infty}^{\infty} (2\pi)^{-\frac{H}{2}} \det(I)^{-\frac{1}{2}} e^{-\frac{1}{2} x^\top I^{-1} x} dx \\
&= e^{\frac{1}{2}\tau_1^\top A \tau_1} \det(A^{-1}\Gamma)^{-\frac{1}{2}}
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{E}\Big(e^{\tau_1^\top v + v^\top \tau_2 v}\Big) &= e^{\frac{1}{2}\tau_1^\top A \tau_1} \det(A^{-1}\Gamma)^{-\frac{1}{2}} \\
&= \det(I - 2\tau_2\Gamma)^{-\frac{1}{2}} e^{\frac{1}{2}\tau_1^\top (I - 2\tau_2\Gamma)^{-1}\Gamma\tau_1}
\end{aligned} \tag{61}$$

B Example Economy 2

Let H denote the maximum horizon. Recall the assumed idiosyncratic volatility process

$$\begin{aligned}
\sigma_{\varepsilon,t} &= \sigma_\varepsilon z_t \\
\Sigma_{\varepsilon,t} \equiv \sigma_{\varepsilon,t}\sigma_{\varepsilon,t}^\top &= \sigma_\varepsilon \sigma_\varepsilon^\top z_t^2 \equiv \Sigma_\varepsilon z_t^2
\end{aligned}$$

where z_t is a stochastic scalar following an AR(1)

$$z_{t+1} = z + b\tilde{z}_t + du_{t+1}$$

where u_{t+1} is i.i.d. standard normal.

Then

$$z_{t+h} = z + b^h \tilde{z}_t + \sum_{j=1}^{h} b^{h-j} du_{t+j}$$

Thus, letting $\hat{z}_{t,h} \equiv z + b^h \tilde{z}_t$

$$\begin{aligned} z_{t+h}^2 &= \hat{z}_{t,h}^2 + 2\hat{z}_{t,h} \sum_{j=1}^{h} b^{h-j} du_{t+j} + \Big(\sum_{j=1}^{h} b^{h-j} du_{t+j} \Big)^2 \\ &= \hat{z}_{t,h}^2 + 2\hat{z}_{t,h} v_s + v_h^2 \end{aligned}$$

where

$$v_h \equiv \sum_{j=1}^{h} b^{h-j} du_{t+j}$$

Note that v_h is normally distributed with variance $\kappa_h^2 \equiv \sum_{j=1}^{h} b^{2(h-j)} d^2 = \frac{b^2 - b^{2h}}{1-b^2} d^2$. Let $v \equiv (v_1, \ldots, v_{H-1})'$. The covariance matrix Γ of v is then given by

$$\begin{aligned} \Gamma_{l,k} &\equiv Cov(v_l, v_k) = Cov\Big(\sum_{j=1}^{l} b^{l-j} du_{t+j}, \sum_{j=1}^{k} b^{k-j} du_{t+j} \Big) \\ &= b^{l-k} Cov\Big(\sum_{j=1}^{k} b^{k-j} du_{t+j}, \sum_{j=1}^{k} b^{k-j} du_{t+j} \Big) \\ &= b^{l-k} \kappa_k^2, \qquad\qquad l \geq k \\ \Gamma_{l,k} &\equiv Cov(v_l, v_k) = b^{k-l} \kappa_l^2, \qquad\qquad l \leq k \end{aligned}$$

Finally, note that

$$\sum_{l=0}^{h-1} \hat{z}_{t,l}^2 = \sum_{l=0}^{h-1} \Big(z^2 + 2z\tilde{z}_t b^l + \tilde{z}_t^2 b^{2l} \Big) = hz^2 + 2\frac{1-b^h}{1-b} z\tilde{z}_t + \frac{1-b^{2h}}{1-b^2} \tilde{z}_t^2$$

We are interested in

$$B_t^{(h)} \equiv \mathbb{E}_t exp\Big\{ \sum_{l=0}^{h-1} 2\Sigma_{\varepsilon,t} \Big\} = \mathbb{E}_t exp\Big\{ \sum_{l=0}^{h-1} 2\Sigma_\varepsilon z_{t+l}^2 \Big\} = \mathbb{E}_t exp\Big\{ 2\Sigma_\varepsilon \sum_{l=0}^{h-1} \Big(\hat{z}_{t,l}^2 + 2\hat{z}_{t,l} v_l + v_l^2 \Big) \Big\}$$

where we use the notation that $v_0 = 0$. It is useful to define

$$\begin{aligned} \tau_{1,t,l}^{i,j} &= 4\hat{z}_{t,l} \Sigma_{\varepsilon,i,j}, \qquad\qquad l = 1, \ldots, H-1 \\ \tau_{2,l,l}^{i,j} &= 2\Sigma_{\varepsilon,i,j}, \qquad\qquad l = 1, \ldots, H-1 \\ \tau_{2,l,k}^{i,j} &= 0, \qquad\qquad \forall\ l \neq k \end{aligned}$$

We can then find the i,j-th element of $B^{(h)}$ by applying the moment generating function in (61), using only the first $h-1$ elements of $\tau_{1,t}^{i,j}$ (note the dependence on time t information through $\hat{z}_{t,l}$) and the first $h-1$ columns and rows of $\tau_2^{i,j}$, which I will denote as $\tau_{1,t}^{i,j(h)}$ and $\tau_2^{i,j(h)}$ respectively. Similarly, the covariance matrix is the first $h-1$ columns and rows of Γ, which I will denote by $\Gamma^{(h)}$. We can then write

$$
\begin{aligned}
B_{t,i,j}^{(h)} &= exp\Big\{2\Sigma_{\varepsilon,i,j}\sum_{l=0}^{h-1}\hat{z}_{t,l}^2\Big\} \times \mathbb{E}_t exp\Big\{\tau_{1,t}^{i,j(h)\top}v^{(h)} + v^{(h)\top}\tau_2^{i,j(h)}v^{(h)}\Big\} \\
&= exp\Big\{2\Sigma_{\varepsilon,i,j}\Big(hz^2 + 2\frac{1-b^h}{1-b}z\tilde{z}_t + \frac{1-b^{2h}}{1-b^2}\tilde{z}_t^2\Big)\Big\} \times \det(I - 2\tau_2^{i,j(h)}\Gamma^{(h)})^{-\frac{1}{2}} e^{\frac{1}{2}\tau_{1,t}^{i,j(h)\top}A^{i,j(h)}\tau_{1,t}^{i,j(h)}} \\
&= \det(I - 2\tau_2^{i,j(h)}\Gamma^{(h)})^{-\frac{1}{2}} \times exp\Big\{2\Sigma_{\varepsilon,i,j}\Big(hz^2 + 2\frac{1-b^h}{1-b}z\tilde{z}_t + \frac{1-b^{2h}}{1-b^2}\tilde{z}_t^2\Big) \\
&+ \frac{1}{2}\sum_{l=1}^{h-1}\sum_{k=1}^{h-1}4(z+b^l\tilde{z}_t)\Sigma_{\varepsilon,i,j}A_{l,k}^{i,j(h)}4(z+b^k\tilde{z}_t)\Sigma_{\varepsilon,i,j}\Big\} \\
&= \det(I - 2\tau_2^{i,j(h)}\Gamma^{(h)})^{-\frac{1}{2}} \times exp\Big\{2\Sigma_{\varepsilon,i,j}\Big(hz^2 + 2\frac{1-b^h}{1-b}z\tilde{z}_t + \frac{1-b^{2h}}{1-b^2}\tilde{z}_t^2\Big) \\
&+ 8\Sigma_{\varepsilon,i,j}^2\sum_{l=1}^{h-1}\sum_{k=1}^{h-1}(z+b^l\tilde{z}_t)(z+b^k\tilde{z}_t)A_{l,k}^{i,j(h)}\Big\} \\
&= \det(I - 2\tau_2^{i,j(h)}\Gamma^{(h)})^{-\frac{1}{2}} \times exp\Big\{2\Sigma_{\varepsilon,i,j}\Big(h + 4\Sigma_{\varepsilon,i,j}\sum_{l=1}^{h-1}\sum_{k=1}^{h-1}A_{l,k}^{i,j(h)}\Big)z^2 \\
&+ 2\Sigma_{\varepsilon,i,j}\Big(2\frac{1-b^h}{1-b}z\tilde{z}_t + \frac{1-b^{2h}}{1-b^2}\tilde{z}_t^2\Big) \\
&+ 8\Sigma_{\varepsilon,i,j}^2\sum_{l=1}^{h-1}\sum_{k=1}^{h-1}(b^l+b^k)A_{l,k}^{i,j(h)}z\tilde{z}_t + 8\Sigma_{\varepsilon,i,j}^2\sum_{l=1}^{h-1}\sum_{k=1}^{h-1}b^{l+k}A_{l,k}^{i,j(h)}\tilde{z}_t^2\Big\}
\end{aligned}
$$

It is therefore clear that

$$
\begin{aligned}
B_{t,i,j}^{(h)} &= \det(I - 2\tau_2^{i,j(h)}\Gamma^{(h)})^{-\frac{1}{2}} \times exp\Big\{2\Sigma_{\varepsilon,i,j}\Big(h + 4\Sigma_{\varepsilon,i,j}\mathbf{1}^\top A^{i,j(h)}\mathbf{1}\Big)z^2 \\
&+ 4\Sigma_{\varepsilon,i,j}\Big(\frac{1-b^h}{1-b} + 2\Sigma_{\varepsilon,i,j}\sum_{l=1}^{h-1}\sum_{k=1}^{h-1}(b^l+b^k)A_{l,k}^{i,j(h)}\Big)z\tilde{z}_t \\
&+ 2\Sigma_{\varepsilon,i,j}\Big(\frac{1-b^{2h}}{1-b^2} + 4\Sigma_{\varepsilon,i,j}\sum_{l=1}^{h-1}\sum_{k=1}^{h-1}b^{l+k}A_{l,k}^{i,j(h)}\Big)\tilde{z}_t^2\Big\}
\end{aligned}
\tag{62}
$$

Note that $A^{i,j(h)}$ is a matrix of constants given by equation (60). Furthermore, $\tilde{z}$ is normally distributed unconditionally, with mean 0 and variance $\sigma_z^2 = \frac{d^2}{1-b^2}$. We can therefore obtain the unconditional expectation of $B_t^{(h)}$ by applying the moment generating function in (61) to (62), letting

$$
\begin{aligned}
\hat{\tau}_1^{i,j(h)} &= 4\Sigma_{\varepsilon,i,j}\Big(\frac{1-b^h}{1-b} + 2\Sigma_{\varepsilon,i,j}\sum_{l=1}^{h-1}\sum_{k=1}^{h-1}(b^l+b^k)A_{l,k}^{i,j(h)}\Big)z \\
\hat{\tau}_2^{i,j(h)} &= 2\Sigma_{\varepsilon,i,j}\Big(\frac{1-b^{2h}}{1-b^2} + 4\Sigma_{\varepsilon,i,j}\sum_{l=1}^{h-1}\sum_{k=1}^{h-1}b^{l+k}A_{l,k}^{i,j(h)}\Big) \\
\hat{\Gamma} &= \frac{d^2}{1-b^2}
\end{aligned}
$$

which are all constants. Therefore

$$
\begin{aligned}
B_{i,j}^{(h)} &\equiv \mathbb{E}(B_{t,i,j}^{(h)}) = \det(I - 2\tau_2^{i,j(h)}\Gamma^{(h)})^{-\frac{1}{2}} \times exp\Big\{2\Sigma_{\varepsilon,i,j}\Big(h + 4\Sigma_{\varepsilon,i,j}\mathbf{1}^\top A^{i,j(h)}\mathbf{1}\Big)z^2\Big\} \times \mathbb{E}e^{\hat{\tau}_1^{i,j(h)}\tilde{z}_t + \hat{\tau}_2^{i,j(h)}\tilde{z}_t^2} \\
&= \det(I - 2\tau_2^{i,j(h)}\Gamma^{(h)})^{-\frac{1}{2}}(1 - 2\hat{\tau}_2^{i,j(h)}\hat{\Gamma})^{-\frac{1}{2}} \times exp\left\{2\Sigma_{\varepsilon,i,j}\Big(h + 4\Sigma_{\varepsilon,i,j}\mathbf{1}^\top A^{i,j(h)}\mathbf{1}\Big)z^2 + \frac{(\hat{\tau}_1^{i,j(h)})^2\hat{\Gamma}}{2(1 - 2\hat{\tau}_2^{i,j(h)}\hat{\Gamma})}\right\}
\end{aligned} \tag{63}
$$

Let us define

$$
\begin{aligned}
\Sigma_{\varepsilon,i,j}^{(h)} \equiv \Sigma_{\varepsilon,i,j}\Big(h + 4\Sigma_{\varepsilon,i,j}\mathbf{1}^\top A^{i,j(h)}\mathbf{1}\Big)z^2 + \frac{(\hat{\tau}_1^{i,j(h)})^2\hat{\Gamma}}{4(1 - 2\hat{\tau}_2^{i,j(h)}\hat{\Gamma})} \\
- \frac{1}{4}\log\Big(\det(I - 2\tau_2^{i,j(h)}\Gamma^{(h)})\Big) - \frac{1}{4}\log\big(1 - 2\hat{\tau}_2^{i,j(h)}\hat{\Gamma}\big)
\end{aligned} \tag{64}
$$

Note that for $h = 1$, (64) simplifies to

$$
\Sigma_{\varepsilon,i,j}^{(1)} = \Sigma_{\varepsilon,i,j}z^2 + \frac{4\Sigma_{\varepsilon,i,j}^2 z^2\hat{\Gamma}}{1 - 4\Sigma_{\varepsilon,i,j}\hat{\Gamma}} - \frac{1}{4}\log\big(1 - 4\Sigma_{\varepsilon,i,j}\hat{\Gamma}\big) \tag{65}
$$

The first term is the average covariance matrix of ε. The last two terms accounts for the stochasticity of $\Sigma_{\varepsilon,t}$. If $\Sigma_{\varepsilon,i,j}$ is "small", we can approximate (65) as

$$
\Sigma_{\varepsilon,i,j}^{(1)} \approx \Sigma_{\varepsilon,i,j}(z^2 + \hat{\Gamma})
$$

Recall that $\hat{\Gamma}$ is the variance of z_t. Thus, the covariance is "bigger" than the average covariance matrix by the amount $\Sigma_\varepsilon\hat{\Gamma}$. Note that to the extent that the approximation is accurate, each term of the covariance matrix is adjusted by the same factor. This does not generally hold at longer horizons.

Finally, we can write

$$
B^h = exp\{2\Sigma_\varepsilon^{(h)}\}
$$

References

Hansen, L.P., and R. Jagannathan. 1997. Assessing specification errors in stochastic discount factor models. *Journal of Finance* 52: 557–590.

Chernov, M., L. Lochstoer, and S.R.L. Lundeby. 2021. Conditional dynamics and the multi-horizon risk-return trade-off. Working Paper.