

Demographics and the Term Structure of Stock Market Risk

Carlo A. Favero*

Bocconi University, IGIER & CEPR

Andrea Tamoni†

Bocconi University

This Version: October, 2010

Abstract

This paper examines the consequences for the term structure of stock market risk of the significance of demographics in capturing the time varying mean of the dividend-price ratio and in predicting stock market returns. A potential role for demographic variables has never been considered in the ongoing debate on the slope of the term structure of stock market risk. Intuitive reasoning, formal modeling and empirical evidence show that demographic trends are a slow-moving information variable, that determines the slow moving mean of the dividend-price ratio and has a forecasting power for stock market returns that increases with the horizon. We show that the forward solution of the dynamic dividend growth model augmented with demographics delivers a negative sloping term structure of stock market risk. Direct regressions of returns at different horizon on the relevant predictors are much better suited to capture this feature of the model than VAR based multi-period iterated forecasts. These results are very little affected by parameters' uncertainty, as a parsimonious parameterization is very precisely estimated in the relevant empirical model. Moreover, they are robust to the existence of "imperfect predictors" as forecast of returns in the direct regression approach involve only currently observable variables and no projections of future variables.

KEYWORDS: dynamic dividend growth model, long run returns predictability, stock market risk, demographics, direct regressions, multiperiod iterated forecasts

J.E.L. CLASSIFICATION NUMBERS: G17,C53,E44

*Deutsche Bank Chair in Quantitative Finance and Asset Pricing, Bocconi University, Department of Finance and IGIER, Milan, 20136, Italy. Phone: 39-02-58363306, e-mail: carlo.favero@unibocconi.it. Favero acknowledges financial support from Bocconi University and from MIUR. We thank seminar participants at the 2010 NBER Summer Institute, Working Group on Forecasting & Empirical Methods in Macroeconomics & Finance at 2010 C6 CSEF-IGIER conference in Capri, and EIEF in Rome for helpful comments. Any errors or omissions are the responsibility of the authors.

†Bocconi University, Department of Finance, Milan, 20136, Italy. Phone: +39 02 58360-3362, e-mail: andrea.tamoni@unibocconi.it.

1 Introduction

This paper examines the consequences for the term structure of stock market risk, defined as the per period conditional variance of cumulative returns, of the significance of demographics in capturing the time varying mean of the dividend-price ratio and in predicting stock market returns. To our knowledge demographic variables has never been considered in the ongoing debate on the slope of the term structure of stock market risk. This is a surprising omission as demographic data, by their nature, have a predictive power for stock market returns that increases with the horizon.

The term structure of stock market risk is of special relevance to strategic asset allocation when risk is measured on the basis of a predictive model for returns at different horizons. There are two basic elements determining the shape of risk as the horizon at which returns are defined gets larger. On the one hand, the longer the horizon the more distant is the future to predict and therefore the more uncertainty should matter. On the other hand, if the "information" in fundamentals for stock market returns emerges only in the long-run, while the short-run is dominated by "noise", then the longer is the horizon the less is the "noise" in the variable to be predicted.¹ Demographic trends are by the nature very smooth, they do not contribute to the short-run noise but they are a natural candidate to capture the information that emerges in the long-run.

The empirical evidence on stock market return predictability associated with demographic information is already fairly extensive in the empirical literature (see Geanakoplos, Magill and Quinzii(2004), Ang and Maddaloni(2005), Favero, Gozluklu and Tamoni(2010)), but the term structure of stock market risk is not simply the other side of the coin of predictability of returns.

The strategic asset allocation literature (e.g. Campbell and Viceira (2002), (2005)), is based on a stationary Vector Autoregressive (VAR) specification for predictors and returns that captures time-variation in the investment opportunity set, and constitutes the input into the optimal asset allocation decision of a long-horizon investor. In practice, long-horizon returns are predicted via multi-step ahead projections of a VAR model in which the dividend-price ratio is used as a predictor of returns and no demographic variables are included. Importantly Campbell and Viceira (2005, CV) and more recently Schotman et al. (2008) have shown that whereas absence of predictability entails a flat term structure of risk, predictability per se does not lead to a downward sloping term structure of risk. In fact, risk at the different horizons is determined by three components: i.i.d uncertainty, mean reversion, uncertainty about future predictors. Without predictability the entire term structure is flat at the level determined by i.i.d uncertainty. This is the classical situation where portfolio choice is independent of the investment hori-

¹The use of the terms "noise" and "information" has been inspired by our reading of Chapter 3 of Taleb (2001).

zon. In the presence of predictability the mean reversion component is naturally downward sloping with the horizon, while the uncertainty about future predictors increases with the horizon. A negative slope of the term structure of risk then emerges only when the effect of the mean reversion component compensates that of the uncertainty about future predictors. The relative importance of these two effects is only an empirical issue. Overall, the slope of the term structure of risk depends on the significance of predictors in explaining returns, on the contemporaneous correlation between the innovations in the equations for predictors and returns, on the variance of returns and predictors, and on the persistence of the predictors. High negative contemporaneous correlation between innovations in the returns and in the predictors, paired with significant predictability and low persistence in the predictors generate a steeply declining term structure of stock market risk. CV find that empirically a very high negative correlation between innovations compensates the high persistence in the predictors (the dividend-price ratio) to generate a negatively sloping term structure of risk. Pastor and Stambaugh (2008, 2009, PS) extend the framework of CV to consider two additional sources of risk: one reflecting uncertainty around the mean of the process generating returns and one reflecting parameters' uncertainty. Within this extended framework, PS do find a positive slope for the term structure of stock market risk, despite the evidence of predictability.

The contribution of our paper to this debate is twofold.

First, following Geanakoplos, Magill, and Quinzii (2004) and Favero, Gozluklu and Tamoni (2010), we augment the set of predictors of stock market returns by considering the dividend-price ratio together with a demographic variable, MY , the ratio of middle-aged to young population, that captures the slow moving but time-varying mean of the US dividend-price ratio. The quick mean reversion of the dividend-price ratio towards a time varying mean determined by demographics has important implications for the predictability of returns and the slope of the term structure of stock market risk.

Second, we argue that the importance of such variable can be best understood in the framework of the dynamic dividend growth model (Campbell-Shiller (1988)) with a time varying linearization point driven by demographics, and that in this context direct regression rather than multi-step ahead forecasting is the natural way to predict long-horizon returns.

The best way to introduce our work is to refer the reader to Figures 1.1-1.3. Figure 1.1 illustrates, over about one century of US data, the relationship between 1-year real stock market returns and MY . Figure 1.2 relates again demographics and stock market fluctuations, but 20-year real annualized returns are now considered instead of 1-year return. The comovement between demographics and stock market returns is negligible for annual returns and remarkable for 20-year returns. Figure 1.3 rationalizes the positive relation between long-horizons stock market returns and the demographic variable in terms of the negative relation between the dividend-price ratio and MY . If MY captures

the time varying-mean of the dividend price ratio, being negatively related to it², and stock market returns react to correct deviations of the dividend-price from its equilibrium level, then future stock market returns are positively related to MY .

[Insert Figure 1.1-1.3 about here.]

When demographic trends are used to model the slow moving fluctuations in the dividend-price ratio a natural decomposition of this variable into an high volatility “noise” component, and a low-volatility “information” component naturally emerges.

This paper embeds the decomposition in a small “structural” model that allows for an explicit role for demographics in the dynamic dividend growth model. In particular, the dividend-price ratio is made a function of a temporary “noise” component and of a persistent “information” component related to demographics. The term structure of stock market risk is then derived by the simultaneous estimation of a system for stock market returns at different horizons obtained by the forward solution of the model. We show that the forward solution of the dynamic dividend growth model (Campbell-Shiller(1988)) augmented with demographics does naturally progressively eliminate the noise component as the horizon increases. The explicit comparison of our results with the traditional VAR-based methods to derive the term structure of stock market risk shows that the combination of direct regressions methods and the inclusion of the demographic variable in the information set relevant for long-horizon regressions makes the term structure of stock market risk steeply downward sloping.

Importantly, the slope of the term structure of stock market estimated within our proposed framework remains downward sloping even when the two additional sources of risk proposed by PS are considered.

The paper is organized as follows. The first section places our contribution in the literature. In the second section a simple structural model linking demographics and the dynamic dividend growth model is introduced to derive the term structure of stock market risk via direct estimation of a system for returns at different horizons. In the third section, the term structure of stock market risk derived from the forward-looking solution of the model and estimated via direct regressions is compared with the one derived from the backward solution of the VAR adopted by the CV and estimated via the multi-step iterated forecast. We then consider the impact on our results of considering the two additional sources of uncertainty introduced by PS. The last section concludes.

²We shall discuss the sign of this relation and provide an interpretation for it in the next section of the paper.

2 Related Literature

This paper adds to a considerable literature on the relation between the predictability of stock market returns and the term structure of stock market risk.

In describing the “verdict of history” on asset returns on a long-sample (1802-1996) of US historical data J.J.Siegel(1998,pp.32), pointed out that "*...stocks are riskier than fixed-income investment over short-term holding periods. But once the holding period increases to between 15 and 20 years, the standard deviation of average annual returns,..., becomes lower than the standard deviation of average bond and bill returns...*".

This statement on unconditional second moments has been strengthened by Campbell-Viceira(2002, 2005) who exploited the predictability of returns by estimating VAR models for returns and predictors and by using VAR-based multi-period iterated forecasts to find that the conditional variance of stock return does not grow in proportion with the investment horizon but it grows more slowly. As a consequence the term-structure of stock market risk is downward sloping and the findings by Siegel on the property of the unconditional distribution of stocks returns are extended and strengthened when the conditional distribution of returns is used to measure stock market risk.

However, the downward sloping term-structure of stock market risk, has been recently questioned by Pastor-Stambaugh(2008,2009) who show that, allowing for coefficient uncertainty and imperfect predictors in a Campbell-Viceira type of VAR, makes the conditional variance of stock returns increasing with the horizon as it can even exceed both the unconditional variance and the one period ahead conditional variance.

The VAR-based approach to measure the term structure of stock market risk uses the (log) dividend-price ratio as the predictor for stock market returns at different horizons. This specification has its foundation in the dynamic dividend growth (DDG) model proposed by (Campbell and Shiller 1988). In fact, the DDG model predicts that (log) dividends and prices share a common stochastic trend and that deviations of (log) prices from the common trend in (log) dividends summarize expectations of either stock market returns, or dividend growth or some combination of the two. The empirical investigation of the DDG has established a number of relevant results.

First, the log dividend-price ratio, dp_t , does not have important long-horizon forecasting power for future discounted dividend-growth (Campbell (1991), Cochrane (1991), Campbell, Lo, and Mackinlay (1997), Cochrane (2001) and Cochrane (2008a)). Second, dp_t is a very persistent time-series and forecasts stock market returns and excess returns over horizons of many years (Fama and French (1988), Campbell and Shiller (1988), Cochrane (2001, Ch. 20), and Cochrane (2008a)). Third, the very high persistence of dp_t has led some researchers to question the evidence of its forecasting power for returns, especially at short-horizon. Indeed careful statistical analysis that takes full account of the persistence in dp_t provides little evidence in favour of stock market returns forecasta-

bility (Nelson and Kim (1993); Stambaugh (1999); Ang and Bekaert (2007); Valkanov (2003); Goyal and Welch (2003) and Welch and Goyal (2008)). Structural breaks have also been found in the relation between dp_t and future returns (Neely and Weller (2000) and Weller(2000) and Paye and Timmermann (2006), Rapach and Wohar (2006)). A recent strand of the empirical literature has related the contradictory evidence on the dynamic dividend growth model to the potential weakness of its fundamental hypothesis that log dividend-price ratio is a stationary process (Lettau and Van Nieuwerburgh (2008), LVN henceforth). LVN use a century of US data to show evidence on the breaks in the constant mean \overline{dp} . The potential time-variation of the linearization point creates a link between demographics and the DDG model.

Favero, Gozluklu and Tamoni(2010) have explained the structural breaks found by LVN in the dividend-price process with demographic trends. They point out that theoretical model by Geneakoplos et al.(2004) (GMQ) predicts that a specific demographic variable, MY , the ratio of middle-aged to young population, determines fluctuations in the dividend yield.

GMQ consider an overlapping generation model in which the demographic structure mimics the pattern of live births in the US, that have featured alternating twenty-year periods of boom and busts. They conjecture that the life-cycle portfolio behavior (Bakshi-Chen, 1994), which suggests that agents should borrow when young, invest for retirement when middle-aged, and live off their investment once they are retired, plays an important role in determining equilibrium asset prices. Consumption smoothing by the agents, given the assumed demographic structure requires that when the MY ratio is small (large), there will be excess demand for consumption (saving) by a large cohort of retirees (middle-aged) and for the market to clear, equilibrium prices of financial assets should adjust, i.e. decrease (increase), so that saving (consumption) is encouraged for the middle-aged. As the dividend-price ratio is negatively related to fluctuations in prices, the model predicts a negative relation between this variable and MY . When the GMQ model is taken to the data via the conjecture that fluctuations in MY could capture a slowly evolving mean in dp_t within the dynamic dividend growth model (Favero, Gozluklu and Tamoni (2010)), strong evidence is found in favour of using this variable together with the dividend-price ratio in long-run forecasting regressions for stock market returns. Interestingly, the fluctuations in MY match very well the break-points in the mean of dp_t identified by LVN in the fifties and the nineties. This paper differs from Favero, Gozluklu and Tamoni (2010) in that it concentrates on the implications of the relation between demographics and the dividend-price ratio for the term structure of stock market risk. To our knowledge, this step has never been taken in the available literature. We propose an empirical strategy potentially capable of identifying separately the importance of demographic variables for high-frequency and low-frequency fluctuations in asset prices. Investigations on the interaction between asset prices and demographic variables have traditionally concen-

trated either on high-frequency or low frequency fluctuations but have never considered an empirical framework based on the dynamic dividend growth model, capable of accommodating both of them, with a different role (see Erb, Harvey, and Viskanta (1997), Poterba (2001), Goyal (2004), Ang and Maddaloni (2005) and DellaVigna and Pollet (2005)).

3 The Dynamic Dividend Growth Model, Demographics and the Term Structure of Stock Market Risk.

The objective of this section is to propose a new model to measure the term structure of stock market risk.

Consider the continuously compounded stock market return from time t to time $t + 1$, \mathbf{r}_{t+1} . Define $\boldsymbol{\mu}_t$, the conditional expected log return given information up to time t , as follows:

$$\mathbf{r}_{t+1} = \boldsymbol{\mu}_t + \mathbf{u}_{t+1}$$

where \mathbf{u}_{t+1} is the unexpected log return. Define the k -period cumulative return from period $t + 1$ through period $t + k$, as follows:

$$\mathbf{r}_{t,t+k} = \sum_{i=1}^k \mathbf{r}_{t+i}$$

The term structure of risk is defined as the conditional variance of cumulative returns, given the investor's information set, scaled by the investment horizon

$$\Sigma_r(k) \equiv \frac{1}{k} \text{Var}(\mathbf{r}_{t,t+k} \mid D_t) \tag{1}$$

where $D_t \equiv \sigma\{z_k : k \leq t\}$ consists of the full histories of returns as well as predictors that investors use in forecasting returns.

In the light of the results of the empirical investigations on the DDG model and on the evidence of the relation between demographics and the dividend price ratio dp_t , we consider the following small "structural" model³:

³The model representation omits constants that have been included in the estimated version.

$$\Delta d_{t+1} = \varepsilon_{1,t+1} \quad (2)$$

$$dp_{t+1} = \varphi_{22}dp_t + \varphi_{23}MY_{t+1} + \varepsilon_{2,t+1} \quad (3)$$

$$r_{t+1}^s = \Delta d_{t+1} - \rho [dp_{t+1} - \bar{dp}_{t+1}] + [dp_t - \bar{dp}_t] + \varepsilon_{3,t+1} \quad (4)$$

$$\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix} \sim \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, & \sigma_1^2 & 0 & 0 \\ & 0 & \sigma_2^2 & 0 \\ & 0 & 0 & \sigma_3^2 \end{bmatrix}$$

Equation (2) specifies the process for the dividend growth as a white noise, where we label $\varepsilon_{1,t+1}$ the innovation to real dividend growth $\varepsilon_{1,t+1}$. This simple parameterization is fully consistent with the evidence of very little predictability of dividend growth⁴.

Equation (3) specifies the process for the dividend-price ratio as fluctuating around a time-varying mean determined by the age structure of the population, MY , a slowly evolving highly predictable variable (the Bureau of Census makes available through its web page projections of this variable up to 2050). Such a modification is justified by the theoretical model of Geanakoplos, Magill, and Quinzii (2004) and by the empirical evidence provided in Favero, Gozluklu, and Tamoni (2010). MY constitutes the information component of the dividend price ratio and there is no uncertainty attached to it: we take it as an exogenous variable whose path for the relevant future is known. However, the dividend-price is also affected by some short-term idiosyncratic noise $\varepsilon_{2,t+1}$. If $|\varphi_{22}| < 1$, then the dividend-price is mean reverting toward a long-run trend determined by the information variable and the effect of the noise shock on the process is only temporary. In fact, our empirical results will show that the speed of mean reversion of the dividend-price ratio toward its long-run mean determined by demographic trends is much higher than that of the dividend-price process itself. Importantly, stability analysis conducted via the Quandt-Andrews test (see Andrews, 1993) for unknown breakpoints confirms the evidence of instability discussed in LVN for the parameters of a simple autoregressive process for dp_{t+1} , while the null of no-break cannot be rejected when the autoregressive model is augmented with MY_{t+1} ⁵. This reduced persistence in the financial ratio is important for two reasons. First, it makes inference less problematic, as there is little doubt on the stationarity of dividend-price ratio around a demographic trend. Second, the reduced half-life of the shocks to the short-term idiosyncratic noise in

⁴Following the approach of Lacerda and Santa-Clara (2010) we also tried a specification where agents forecast the dividend growth rate from the average of past dividend growth rates, i.e. $E_t[\Delta d_{t+k}] = \bar{g}_t$. Results are unaffected and available upon requests.

⁵The Quandt-Andrews test for unknown breakpoints (with a trimming of 10 per cent of the observations) takes a Maximum Wald statistic of 20.06 in 1954 with a tail probability of 0.001 for the parameters in the autoregressive process for dp_t . When the same test is applied to equation (??) the Maximum Wald statistic takes a value 11.68 with a tail probability of 0.076 and the null of parameters stability cannot be rejected.

dividend-price ratios has a direct impact on the term structure of stock market risk.

Equation (4) is a version of the extension of the Campbell-Shiller log linear approximation of real returns proposed by Lettau&VanNieuwerburgh (2008) to allow for time-varying steady-state growth rates and returns.

In Appendix B of their paper, Lettau&VanNieuwerburgh (2008) derive the following log-linear approximation of returns⁶:

$$(r_{t+1}^s - \bar{r}_{t+1}) = (\Delta d_{t+1} - \bar{\Delta d}_{t+1}) - \rho_t [dp_{t+1} - \bar{dp}_{t+1}] + [dp_t - \bar{dp}_t] + \Delta \bar{dp}_{t+1} \quad (5)$$

We obtain our equation (4) from (5) by assuming that of the three processes for returns, dividend growth and the dividend-price ratio only the last one is persistent ($\bar{r}_t = \bar{\Delta d}_t = const$), that the time-varying mean of the dividend-price ratio is very slowly evolving, i.e. $\Delta \bar{dp}_{t+1} \approx 0$ and that the linearization parameter is constant, $\rho_t = \rho$.⁷ We introduce an error term $\varepsilon_{3,t+1}$ to capture the effect of our approximation.

We report in Figure 2.1 the three endogenous time series in our small structural model.

Insert Figure 2.1 here

The graphical evidence is suggestive that the speed of mean reversion towards a constant mean of the dividend-price ratio is very different from that of annual real returns and annual real dividend growth. We model this feature of the data by introducing a time-varying mean for the dividend-price ratio, driven by demographics. The figure makes clear that without this step it would be very hard to reconcile the time-series properties of dp_t with those of r_t^s and Δd_{t+j} . We believe that this feature of the data, rather overlooked in the literature, is at the heart of the controversy on the predictive power of the dividend-price ratio for stock market returns.

By solving eq. (4) forward we obtain:

$$\begin{aligned} \sum_{j=1}^m \rho^{j-1} (r_{t+j}^s) &= [dp_t - \bar{dp}_t] + \sum_{j=1}^m \rho^{j-1} (\Delta d_{t+j}) - \rho^m [dp_{t+m} - \bar{dp}_{t+m}] \\ &+ \sum_{j=1}^m \rho^{j-1} (\varepsilon_{1,t+j} + \varepsilon_{3,t+j}) \end{aligned} \quad (6)$$

Eq. (6) clearly shows that deviations of the dividend/price ratio from its equilibrium value at time t have a predictive power for m -period ahead stock market returns (and/or

⁶ $r_{t+1}^s = \ln \left(\frac{P_{t+1} + D_{t+1}}{P_t} \right)$,

$dp_t = \ln \left(\frac{D_t}{P_t} \right)$

⁷Rytchkov (2008) estimates a system of equation similar to ours and study how sensitive ML parameters are to variation in this parameter. He concludes that there is almost no sensitivity to the choice of ρ (see Table 1 in his paper).

dividend growth) that increases with the horizon, as the larger is m the smaller is the effect of future noise in the dividend-price ratio $[dp_{t+m} - \overline{dp}_{t+m}]$. However, this term cannot be ignored in the computation of the term structure of stock market risk that considers typically horizons from 1-year onwards. To bring (6) to the data, an observable counterpart of the time varying linearization value for the dividend-price must be considered. Consistently with (3), we assume that the relevant linearization value for computing returns from time t to time $t + m$ is the conditional expectation of the dividend-yield for time $t + m$, given the information available at time t . We then have

$$\begin{aligned} \sum_{j=1}^m \rho^{j-1} (r_{t+j}^s) &= dp_t - \left[\varphi_{22}^m dp_t + \sum_{j=1}^m \varphi_{22}^{j-1} \varphi_{23} MY_{t+m+1-j} \right] + u_{t+m} \quad (7) \\ &= (1 - \varphi_{22}^m) dp_t - \sum_{j=1}^m \varphi_{22}^{j-1} \varphi_{23} MY_{t+m+1-j} + u_{t+m} \\ u_{t+m} &= \sum_{j=1}^m \rho^{j-1} (\varepsilon_{1,t+j} + \varepsilon_{3,t+j}) - \rho^m \sum_{j=1}^m \varphi_{22}^{j-1} \varepsilon_{2,t+m+1-j} \end{aligned}$$

Note that the relevance of the noisy component $\varepsilon_{2,t+m+1-j}$ of the dividend-price ratio in the distribution of m -period returns decreases with the horizon: as the horizon gets longer the mean-reversion of the dividend-yield process around the information variable makes the informative content of this variable dominant. The speed at which the effect of the noise is dampened depends on the speed of mean reversion of the dividend process and on the discount parameter ρ . However, even for values of ρ close to unity, the mean reversion in dividend-prices is sufficient to cause a dampening over the horizon of the effect of the noise $\varepsilon_{2,t}$. The second and the third component of the noise in m -period returns are the uncertainty in the dividend process and the real returns that die out much more slowly than the effect of the noise $\varepsilon_{2,t+j}$ and they become persistent when ρ approaches the unit value. Eq. (7) implies that the fit of direct predictive regressions projecting returns at different horizon on the information available at time t should improve with the horizon. It also predicts that the residuals of such predictive regressions have a moving-average component that should be taken care of in estimation. This is a well-known result (see for example, Valkanov (2003)). Interestingly, the model also predicts that the coefficient on the dividend-yield in the projections of long-horizon returns on this variable should be increasing with the horizon.

We measure the term structure of stock market risk by estimating the following “structural” system of eleven equations⁸:

⁸Our “structural” estimation is similar to that by Van Binsbergen and Koijen (2009) with two main differences: equations at all relevant horizons are simultaneously estimated and all variables included in the model are observable.

$$\begin{aligned} \frac{1}{\sqrt{m}} \sum_{j=1}^m (r_{t+j}^s) &= \delta_{0,m} + \frac{1}{\sqrt{m}} (1 - \varphi_{22}^m) dp_t - \frac{\varphi_{23}}{\sqrt{m}} \left(\sum_{j=1}^m \varphi_{22}^{j-1} MY_{t+m+1-j} \right) + u_{t+m} \quad (8) \\ m &= 1, \dots, 10 \\ dp_{t+1} &= \varphi_{20} + \varphi_{22} dp_t + \varphi_{23} MY_{t+1} + \varepsilon_{2,t+1} \end{aligned}$$

The specification of (8) slightly differs from the model in that we use as a dependent variable the unweighted annualized period-returns ($\rho = 1$). This is because the objective of our exercise is to compare the term structure of stock market risk obtained by direct regression and by iterative multi-step iterated VAR based forecasts. To assess the potential cost of the several approximations we have also estimated an unrestricted version of (8) to perform a test of the validity of the relevant restrictions:

$$\begin{aligned} \frac{1}{\sqrt{m}} \sum_{j=1}^m (r_{t+j}^s) &= \delta_{0,m} + \frac{\delta_{1m}}{\sqrt{m}} dp_t + \frac{\delta_{2m}}{\sqrt{m}} \left(\sum_{j=1}^m \varphi_{22}^{j-1} MY_{t+m+1-j} \right) + u_{t+m} \quad (9) \\ m &= 1, \dots, 10 \\ dp_{t+1} &= \varphi_{20} + \varphi_{22} dp_t + \varphi_{23} MY_{t+1} + \varepsilon_{2,t+1} \end{aligned}$$

Note that (8) and (9) are both specified with $\frac{1}{\sqrt{m}} \sum_{j=1}^m (r_{t+j}^s)$ as the dependent variable to obtain directly the conditional annualized standard error of returns from the standard error of the regression.

We estimate the model on a dataset of annual observations for the period 1910-2009. The data are from Welch and Goyal (2008)⁹, who provide detailed descriptions of the data and their sources. Stock returns are measured as continuously compounded returns on the S&P 500 index, including dividends. To compute real returns we calculate inflation rate from the CPI (all urban consumers). The predictor for the equity premium is the dividend-price ratio, computed as the difference between the log of dividends paid on the S&P 500 index and log of stock prices (S&P 500 index), where dividends are measured using a one-year moving sum.

The results of the estimation are reported in Table 1. The GMM method allows to estimate the restricted model parameters φ_{22} and φ_{23} by taking into account the MA(h-1) structure of the error terms in computing standard errors. We also report estimates of the standard error that uses the SURE method proposed by Pesaran, Pick and Timmermann(2010) to account for serial correlation in the residuals of the multi-period direct forecasting model¹⁰. The Table shows an highly significant effect of MY

⁹The data are available at www.bus.emory.edu/AGoyal/Research.html.

¹⁰See the Appendix on Robustness for a full description of the results.

both in the equation for dp_t and in all ten predictive regressions. The performance of the restricted model, that estimates only two parameters in addition to eleven constants, is very similar in term of adjusted R^2 and standard error of the equations to that of the unrestricted model that estimates twenty more parameters and the restrictions are not rejected by the relevant chi-square test.

[Insert Table 1 about here.]

The estimates of the parameters φ_{22} and φ_{23} show that demographics are strongly significant in explaining the dividend-price ratio and that the dividend-price ratio is clearly mean reverting around a mean determined by MY. Figure 2.2 brings more evidence on this issue by reporting dp_t along with the time-varying linearization point used in the model and the breaks identified by LVN.

[Insert Figure 2.2 about here.]

Finally, the term structure of stock market risk described by the estimation of the structural system of direct regression is steeply downward sloping as it can be read directly off the standard errors of regressions reported in Table 1.

4 Direct regression versus VAR multistep-ahead forecasts

Our empirical results on the term structure of stock market risk differ rather importantly from those derived in the literature based on VAR models. To illustrate the point, we consider first a simple representation of the VAR adopted by CV by estimating a model for continuously compounded total stock market returns, r_t^s , and the log dividend price, dp_t :

$$\begin{aligned} (z_t - E_z) &= \Phi_1 (z_{t-1} - E_z) + \nu_t \\ \nu_t &\sim \mathcal{N}(0, \Sigma_\nu) \end{aligned}$$

where

$$\begin{aligned} z_t &= \begin{bmatrix} r_t^s \\ dp_t \end{bmatrix}, E_z = \begin{bmatrix} E_{r^s} \\ E_{d-p} \end{bmatrix} \\ \Phi_1 &= \begin{bmatrix} 0 & \varphi_{1,2} \\ 0 & \varphi_{2,2} \end{bmatrix} \\ \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} &\sim \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix}, & \begin{matrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{matrix} \end{bmatrix} \end{aligned}$$

The bivariate model for returns and the predictor features a restricted dynamics such that only the lagged predictor is significant to determine current returns ($\varphi_{1,1} = 0$) and the predictor is itself a strongly exogenous variable ($\varphi_{2,1} = 0$).

Given the VAR representation and the assumption of constant Σ_ν

$$\begin{aligned} Var_t [(z_{t+1} + \dots + z_{t+k}) | D_t] &= \Sigma_\nu + (I + \Phi_1)\Sigma_\nu(I + \Phi_1)' + \\ &(I + \Phi_1 + \Phi_1^2)\Sigma_\nu(I + \Phi_1 + \Phi_1^2)' + \dots \\ &+(I + \Phi_1 + \dots + \Phi_1^{k-1})\Sigma_\nu(I + \Phi_1 + \dots + \Phi_1^{k-1})' \end{aligned} \quad (10)$$

from which we can derive:

$$\begin{aligned} \Sigma_r(k) &= \frac{1}{k} \sum_{i=0}^{k-1} D_i \Sigma D_i' \\ D_i &= I + \Phi_1 \Xi_{i-1} \quad i > 0 \\ \Xi_i &= \Xi_{i-1} + \Phi_1^i \quad i > 0 \\ D_0 &\equiv I, \quad \Xi_0 \equiv I \end{aligned}$$

Note that, under the chosen specification of the matrix Φ_1 we can write the generic term $D_i \Sigma D_i'$, as follows:

$$\begin{aligned} D_i \Sigma D_i' &= \left(\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{12}' & M_{22} \end{array} \right) \quad (11) \\ M_{11} &= \Sigma_{1,1} + \Phi_{1,2} \Xi_{i-1}^{(22)} \Sigma'_{1,2} + \Sigma_{1,2} \Xi_{i-1}^{(22)'} \Phi'_{1,2} + \Phi_{1,2} \Xi_{i-1}^{(22)} \Sigma_{2,2} \Xi_{i-1}^{(22)'} \Phi'_{1,2} \\ M_{12}' &= \Xi_i^{(22)} \Sigma'_{1,2} + \Xi_i^{(22)} \Sigma_{2,2} \Xi_{i-1}^{(22)'} \Phi'_{1,2} \\ M_{22} &= \Xi_i^{(22)} \Sigma_{2,2} \Xi_i^{(22)'} \end{aligned}$$

where we have used the fact that

$$\begin{aligned} \Xi_i &= \sum_{j=0}^i \Phi_1^j \\ &= \left(\begin{array}{c|c} 0 & \phi_{1,2} \sum_{j=0}^{i-1} \phi_{2,2}^j \\ \hline 0 & \sum_{j=0}^i \phi_{2,2}^j \end{array} \right) \end{aligned}$$

and

$$D_i = I + \Phi_1 \Xi_{i-1} \\ = \left(\begin{array}{c|c} I & \phi_{1,2} \sum_{j=0}^{i-1} \phi_{2,2}^j \\ \hline 0 & \sum_{j=0}^i \phi_{2,2}^j \end{array} \right)$$

Eq. (11) implies that, in our simple bivariate example, the term structure of stock market risk takes the form

$$\sigma_r^2(k) = \sigma_1^2 + 2\varphi_{1,2}\sigma_{1,2}\psi_1(k) + \varphi_{1,2}^2\sigma_{2,2}^2\psi_2(k) \quad (12)$$

where

$$\psi_1(k) = \frac{1}{k} \sum_{l=0}^{k-2} \sum_{i=0}^l \varphi_{2,2}^i \quad k > 1 \\ \psi_2(k) = \frac{1}{k} \sum_{l=0}^{k-2} \left(\sum_{i=0}^l \varphi_{2,2}^i \right)^2 \quad k > 1 \\ \psi_1(1) = \psi_2(1) = 0$$

The total stock market risk can be decomposed in three components: i.i.d uncertainty, mean reversion, uncertainty about future predictors. Without predictability ($\varphi_{1,2} = 0$) the entire term structure is flat at the level σ_1^2 . This is the classical situation where portfolio choice is independent of the investment horizon. The possible downward slope of the term structure of risk depends on the second term, and it is therefore crucially affected by predictability and a negative correlation between the innovations in dividend price ratio and in stock market returns ($\sigma_{1,2}$), the third term is always positive and increasing with the horizon when the autoregressive coefficient in the dividend yield process is positive. Overall, the slope of the term structure of risk depends on the significance of the dividend-price in explaining returns, on the contemporaneous correlation between the innovations in the equations for the dividend-price and returns, on the variance of returns and the dividend-price, and on the persistence of the dividend-price.

[Insert Table 2 about here.]

Table 2 summarizes the results of the estimation of the system. The estimation results confirm the noisy nature of 1-year stock market returns and the high persistence of the dividend-price ratio. The covariance structure of the innovations is such that the unexpected log excess stock returns are highly negatively correlated with the innovations in the log dividend price ratio. Figure 3 plots the term structure of risk resulting from the estimation of the restricted VAR and its decomposition. The evidence of a downward

sloping curve with risk halving from the one-year to the thirty year horizon replicates the results in Campbell and Viceira (2002), based on the estimation of a larger model including bond and stock excess returns, the nominal and real risk free rate together with the dividend-yield and the yield spread as predictors.

[Insert Figure 3 about here.]

The slope of the term structure of the stock market risk estimated within a traditional VAR features a much smaller decline than that obtained by direct regression of our structural model with demographics. In fact, risk at one year horizon is estimated at about the same level in both models, as the one-period ahead standard deviation is of about 20 per cent . However, from the one-year horizon onward the term structure based on the structural model with demographics declines much more steeply to reach an annualized standard deviation of about 11 per cent at the 10-year horizon against an estimate fifty per cent larger of about 16 per cent from the VAR model.

4.1 Understanding the difference

Why are our results so different from those based on traditional VAR analysis? There are two reasons: the inclusion of MY_t in the investor's information set relevant for long-horizon and the derivation of the term structure of risk via direct regression.

To assess the first effect we estimate the standard Campbell-Viceira VAR augmented with the exogenous MY variable. Results are reported in Table 4.

[Insert Table 4 about here.]

The comparison of Tables 2 and 4 illustrates clearly that the inclusion of MY in the Campbell-Viceira VAR model would cause a downward shift and a steepening of the term structure of stock market risk as it reduces the persistence of dp_t with respect to the traditional specification omitting demographics (the parameter $\varphi_{2,2}$ of the own lag of dp_t is reduced from .89 to .73 when MY_t is included in the VAR specification) and it also reduces its conditional and unconditional variance (the one-step ahead volatility σ_2 goes from .22 to .21). Interestingly, the correlation between the residuals of the equation for one-year returns and the equation for the dividend- price ratio is very little affected by the inclusion of MY_t in the specification for the dynamics of the predictor (as it goes from -.86 to -.85).

As for the second effect, the use of the direct regression rather than iterative multi-step ahead forecast from the VAR has a less immediate implication that is however empirically and theoretically relevant. To illustrate this point we re-write, consistently with equation (7), the term structure of stock market risk implied by the direct regression of returns at different horizons on the relevant predictors as follows:

$$\sigma_r^2(k) = \psi_1(k) (\sigma_1^2 + \sigma_3^2) + \psi_2(k) \sigma_2^2 \quad (13)$$

$$\psi_1(k) = \frac{1}{k} \sum_{j=1}^k \rho^{2(j-1)}$$

$$\psi_2(k) = \frac{\rho^{2k}}{k} \sum_{j=1}^k \varphi_{22}^{2(j-1)}$$

This term structure is downward sloping as the effect of the noisy component of the dividend-price dies out as the horizon m increases.

Consider now the case in which a VAR is fitted to the data generated by eqs. (2)-(4):

$$\begin{aligned} r_{t+1}^s &= \varphi_{10} + \varphi_{12} dp_t + \varphi_{13} MY_{t+1} + \varepsilon_{1,t+1} + \varepsilon_{3,t+1} - \rho \varepsilon_{2t+1} \\ dp_{t+1} &= \varphi_{20} + \varphi_{22} dp_t + \varphi_{23} MY_{t+1} + \varepsilon_{2,t+1} \end{aligned}$$

As noted by Cochrane (2008b), deliciously, the regression and “structural” model match almost perfectly.¹¹ However, the term structure of stock market risk derived by backward projection of the VAR on the information available at time t delivers a different shape from that obtained by the direct regression:

$$\sigma_r^2(k) = (\sigma_1^2 + \sigma_3^2 + \rho^2 \sigma_2^2) - 2\varphi_{1,2} \rho \sigma_2^2 \psi_1(k) + \varphi_{1,2}^2 \sigma_2^2 \psi_2(k) \quad (14)$$

$$\psi_1(k) = \frac{1}{k} \sum_{l=0}^{k-2} \sum_{i=0}^l \varphi_{22}^i \quad k > 1$$

$$\psi_2(k) = \frac{1}{k} \sum_{l=0}^{k-2} \left(\sum_{i=0}^l \varphi_{22}^i \right)^2 \quad k > 1$$

$$\psi_1(1) = \psi_2(1) = 0$$

Note that, as long-run returns are obtained by aggregating high-frequency returns and the relevant model is solved backward, the effect of the noisy component does not decrease with the forecasting horizon. Direct comparison of (13) with (14) shows that the VAR based backward-looking term structure of risk produces biased estimates of DDG based forward looking term structure of stock market risk.

In order to assess the empirical relevance of the two effects, i.e. the inclusion of the demographic variable in the information set and the use of direct estimation of a forward-looking model rather than iterative multi-step forecast, we compare in Figure 4 the term structure of risk derived by direct estimation of a model including MY_t , with the term structure of risk based on the recursive iteration of two VARs, one with and

¹¹Recall that MY is an exogenous variable and therefore we can use MY_{t+1} in the VAR.

the other without MY_t . The results show clearly that MY_t plays an important role in determining the conditional mean of the system but also that the use of direct estimation of a forward-looking model rather than iterative recursive multi-step forecast is a source of a major shift in the measured term structure of stock market risk. Consistently with the prediction of the dynamic dividend growth model, such shift is far from being a parallel one.

[Insert Figure 4 about here.]

5 Adding two more sources of uncertainty

Pastor and Stambaugh (2008, 2009) illustrate how the possibility of “imperfect predictors” in the predictive system adds two more sources of uncertainty to the VAR based estimation of the term structure of stock market risk and it empirically changes the slope estimated by Campbell-Viceira. The traditional VAR setup is modified by PS by introducing a predictive relationship linking stock market returns to an unobserved variable μ_t , that in turns is only imperfectly related to the observed dividend-price. In this case, the relevant empirical model can be written as follows:

$$\begin{aligned}
 (r_t^s - E_{r^s}) &= (\mu_{t-1} - E_{r^s}) + u_{1,t} \\
 (dp_t - E_{dp}) &= \varphi_{22} (dp_{t-1} - E_{dp}) + u_{2,t} \\
 (\mu_t - E_{r^s}) &= \varphi_{32} (dp_t - E_{dp}) + u_{3,t} \\
 \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix} &\sim \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, & \begin{matrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{matrix} \end{bmatrix}
 \end{aligned} \tag{15}$$

from which the following VAR representation is derived:

$$\begin{aligned}
 (z_t - E_z) &= \Phi_1 (z_{t-1} - E_z) + \nu_t \\
 \nu_t &\sim \mathcal{N}(0, \Sigma_\nu)
 \end{aligned} \tag{16}$$

where

$$z_t = \begin{bmatrix} r_t^s \\ dp_t \\ \mu_t \end{bmatrix}, E_z = \begin{bmatrix} E_{r^s} \\ E_{dp} \\ E_{r^s} \end{bmatrix}$$

$$\Phi_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & \varphi_{22} & 0 \\ 0 & \varphi_{32}\varphi_{22} & 0 \end{bmatrix}$$

$$\begin{bmatrix} v_{1,t} \\ v_{2,t} \\ v_{3,t} \end{bmatrix} \sim \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, & \begin{matrix} \sigma_1^2 & \sigma_{12} & \varphi_{32}\sigma_{12} + \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \varphi_{32}\sigma_2^2 + \sigma_{23} \\ \varphi_{32}\sigma_{12} + \sigma_{13} & \varphi_{32}\sigma_2^2 + \sigma_{23} & \varphi_{32}^2\sigma_2^2 + \sigma_3^2 + 2\varphi_{32}\sigma_{23} \end{matrix} \end{bmatrix}$$

the term structure of stock market risk takes now the form

$$\sigma_r^2(k) = \sigma_1^2 + 2\sigma_{13} + \sigma_3^2 + 2\varphi_{3,2}(\sigma_{12} + \sigma_{23})\psi_1(k) + \varphi_{2,2}^2\sigma_{2,2}^2\psi_2(k)$$

where

$$\psi_1(k) = \frac{1}{k} \sum_{l=0}^{k-2} \sum_{i=0}^l \beta_{22}^i \quad k > 1$$

$$\psi_2(k) = \frac{1}{k} \sum_{l=0}^{k-2} \left(\sum_{i=0}^l \beta_{22}^i \right)^2 \quad k > 1$$

$$\psi_1(1) = \psi_2(1) = 0$$

Note that the specification of the relevant VAR to project the term structure of risk involves an unobservable variable. VAR estimation needs that to deal with this problem that it is best solved within a Bayesian framework. Such a framework in turn generates naturally another source of volatility, namely parameters uncertainty. In the PS framework the term-structure of risk can be decomposed into five components: the original three in eq. (12) plus other two, one reflecting uncertainty around the mean of the process generating returns and one reflecting parameters' uncertainty.

Table 3 shows the results from the estimation of the three-variate predictive system. Within this Bayesian framework the prior beliefs on the correlation between innovations in the equation for returns and innovations in the equation for expected returns (i.e. ρ_{ν_1, ν_3} , the Stambaugh Correlation) substantially affects estimates of expected returns as well as various inferences about predictability. In our estimation we impose the belief $\rho_{\nu_1, \nu_3} < 0$ following the evidence of Campbell (1991) and of Van Binsbergen and Koijen (2009). Interestingly, Robertson and Wright (2009) show that for a plausible range of ARMA parameters the Stambaugh Correlation is bounded away from zero and very close

to (minus) unity. This leads us to specify an informative prior on ρ_{ν_1, ν_3} that the implied prior on ρ_{ν_1, ν_3}^2 has 99.9% of its mass above 0.5, with a mean of about 0.77.¹² The Table 3 reports the R^2 in the regression of r_{t+1} on $E[r_{t+1}|D_t]$ for the predictive system and shows that this R^2 is higher than the R^2 in Table 2 because $dp_t \in D_t$ and therefore the estimates of the expected returns from the predictive system are at least as precise as the estimates from the predictive regression and VAR. Moreover the R^2 (not reported) from a regression of μ_t on dp_t larger than 0.5 receive very little posterior probability suggesting that the predictor is not perfectly correlated with the latent expected return and therefore the predictive system has superior ability in extracting information and forming the proxy for the true unobservable μ_t .

[Insert Table 3 about here.]

Figure 5 plots the conditional variance $Var[r_{t,t+k}|D_t]$ and its components for the three-variate predictive system with the dividend price as an observable predictor. It is interesting to note that the three components of $Var[r_{t,t+k}|\mu_t, \Theta, D_t]$ ¹³, namely the i.i.d. (top right panel), the mean reverting (mid left panel) and the uncertainty about future values of μ_t (mid right panel) are fairly similar to the one we compute under the VAR approach (see Figure 3). Therefore the sum of these three contribution, namely $Var[r_{t,t+k}|\mu_t, \Theta, D_t]$ is almost identical under the CV and the PS approaches. Nevertheless Pastor and Stambaugh (2009) show that other two important blocks affect the conditional variance $Var[r_{t,t+k}|D_t]$: one is the predictor imperfection that reflects the uncertainty about the current conditional expected returns and the other is the estimation risk that reflects uncertainty about parameters in Φ_1 . In particular, when we add the predictor imperfection component to the $Var[r_{t,t+k}|\mu_t, \phi, D_t]$ we obtain the dashed line in top-left panel of Figure 5 which shows no evidence for a downward sloping term structure of stock-market risk.

[Insert Figure 5 about here.]

In the light of these results it important to evaluate the effect of the introduction of the new sources of risk on the term structure measured by direct regression within our small structural model. The imperfect predictors problem would affect the interpretation of the error term in the ten estimated equations for returns at horizon from 1-year to 10-year but it would not change their estimated standard deviations. In fact in the direct regression approach the term structure of stock market risk is obtained by projecting returns at different horizon $t + m$ on observable variables at time t . If these observable

¹²As noted in PS 2008 this prior reflects the belief that at least half of the variance of market returns is due to discount rate news.

¹³We indicate with Θ the full set of parameters.

variables are imperfect predictors, then the variance of the direct regression residual will reflect this feature of the data.

Parameter uncertainty could be an issue, as in the previous section we derived the term structure of stock market risk by keeping the estimated parameters fixed without considering uncertainty surrounding them. We have therefore recomputed the term structure by allowing for parameter uncertainty both in the restricted and unrestricted version of our model. The results, reported in Figure 6, show that parameters uncertainty add very little uncertainty to the conditional distribution of stock market returns in our framework. This evidence is not surprising, as the estimation results reported in Table 1 document that the term structure of risk derived from our structural system is based on the estimation of very few parameters, all them very well determined. To provide further evidence on the relative performance of our small structural model with demographics and the three-variable VAR with one imperfect predictor proposed by PS we report in Figure 7 actual and predicted returns at the 1-year and 10-year horizon. At the short-end of the term structure the information from fundamentals is totally blurred by the "noise" and the structural model with demographics does not overperform the three-variate VAR. However, at the long-end the "information" generated by using the predictor determined by demographics in the structural model prevails and generates a sizeable overperformance. More precisely at the horizon of 10 years, the mean square error (MSE) and mean absolute error (MAE) for the predictive system are equal to 0.0023 and 0.0409, respectively, whereas the MSE and MAE for the direct regression are equal to 0.0012 and 0.0280, respectively.

[Insert Figures 6 and 7 about here.]

6 Conclusions

We started this paper by arguing that the emergence of relative importance of "information" versus "noise" in the determination of stock market returns at different horizon could generate a downward sloping term structure of stock market risk. We have shown that this is indeed the case when i) a demographic variable is used to capture the slow-moving information component in the dividend-price ratio and in stock market returns and ii) direct regressions based on the structural estimation of a forward-looking specification consistent with the dynamic dividend growth model is adopted. We have also shown that the use of backward-looking iterated multi-step forecasts to derive the term structure of risk leads to an underestimation of the importance of the emergence of "information" as the horizon increases.

References

- [1] Ang, Andrew and Geert Bekaert, 2007, Stock Return Predictability: Is It There? *The Review of Financial Studies*, 20, 651–707.
- [2] Ang, A. and A. Maddaloni. "Do Demographic Changes Affect Risk Premiums? Evidence from International Data." *Journal of Business*, 78 (2005), 341-380.
- [3] Bakshi, Gurdip S., and Zhiwu Chen, 1994, Baby Boom, Population Aging, and Capital Markets, *Journal of Business*, 67, 2, 165-202.
- [4] Van Binsbergen J.H. and R.S.Koijen, 2009, Predictive Regression:a Present Value Approach, mimeo
- [5] Boudoukh, Jacob, Richardson, Matthew and Robert F. Whitelaw, 2008, The Myth of Long-Horizon Predictability, *The Review of Financial Studies*, 21, 4, 1577-1605.
- [6] Campbell, J. Y., 1991. A variance decomposition for stock returns. *Economic Journal* 101, 157–179.
- [7] Campbell, J. Y., Lo, A. W., MacKinlay, C., 1997. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ.
- [8] Campbell, John Y., and Samuel B. Thomson, 2008, Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, *The Review of Financial Studies*, 21, 1509-1531.
- [9] Campbell, John Y., and Robert Shiller, 1988, Stock Prices,Earnings, and Expected Dividends, *Journal of Finance*, 43, 661–676.
- [10] Campbell, John Y., and Luis M. Viceira, 2002, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors* (Oxford University Press).
- [11] Campbell, John Y., and Luis M. Viceira, 2005, The Term Structure of the Risk-Return Tradeoff, *Financial Analysts Journal*, 61, 1, 34-44
- [12] Cochrane, J. H., 2001. *Asset Pricing*. Princeton University Press, Princeton, NJ.
- [13] Cochrane, John H., 2008, The Dog that Did Not Bark: A Defense of Return Predictability, *Review of Financial Studies*, 20, 5.
- [14] Cochrane J.H., 2008, State-Space versus VAR models of Stock Returns, mimeo
- [15] DellaVigna S., and J.Pollet. "Demographics and Industry Returns." *American Economic Review*, 97(2007), 1167-1702.

- [16] Erb, C. B.; C. R. Harvey, and T. E. Viskanta. "Demographics and International Investment." *Financial Analysts Journal*, 53 (1996), 4,14-28.
- [17] Fama, Eugene and Kenneth R. French, 1988, Dividend Yields and Expected Stock Returns, *Journal of Financial Economics*, 22, 3-26.
- [18] Favero C.A., A.Gozluklu and A.Tamoni (2010) "Demographic Trends, the Dividend-Price Ratio and the Predictability of Long-Run Stock Market Returns" , CEPR working paper 7734, forthcoming in the *Journal of Financial and Quantitative Analysis*
- [19] Geanakoplos, John, Magill, Michael and Martine Quinzii, 2004, Demography and the Long Run Behavior of the Stock Market, *Brookings Papers on Economic Activities*, 1: 241-325.
- [20] Goyal, A., Welch, I., May 2003. Predicting the equity premium with dividend ratios. *Management Science* 49 (5), 639–654.
- [21] Goyal, A. "Demographics, Stock Market Flows, and Stock Returns." *Journal of Financial and Quantitative Analysis*, 39 (2004), 1, 115-142.
- [22] Goyal, Amit, and Ivo Welch, 2008, A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, 21-4, 1455-1508
- [23] Hodrick, Robert, 1992, Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, *Review of Financial Studies*, 5, 357–386.
- [24] Lacerda, F., and Santa-Clara, P. 2010, Forecasting Dividend Growth to Better Predict Returns, Discussion paper.
- [25] Lettau, Martin, and Sydney Ludvigson, 2005, Expected Returns and Expected Dividend Growth, *Journal of Financial Economics*, 76, 583-626.
- [26] Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, Aggregate Wealth and Expected Stock Returns, *Journal of Finance*, 56, 3, 815-849.
- [27] Lettau, Martin, and Stijn Van Nieuwerburgh, 2008, Reconciling the Return Predictability Evidence, *Review of Financial Studies*, 21, 4, 1607-1652.
- [28] Lewellen, Jonathan, 2004, Predicting Returns with Financial Ratios. *Journal of Financial Economics* 74:209–35.
- [29] Nelson, C. C., Kim, M. J., 1993. Predictable stock returns: The role of small sample bias. *Journal of Finance* 43, 641–661.

- [30] Neely C.J., Weller P.A.(2000). Predictability in international asset returns: a reexamination. *Journal of Financial and Quantitative Analysis*, 35, 601-620
- [31] Newey, Whitney K. and Kenneth D. West, 1987, A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, 3, 703-08.
- [32] Newey, Whitney K. and Kenneth D. West, 1994, Automatic Lag Selection in Covariance Matrix Estimation, *Review of Economic Studies*, 61, 631-653.
- [33] Pastor, Lubos, and Robert F. Stambaugh, 2008, Predictive systems: Living with imperfect predictors, *Journal of Finance*
- [34] Pastor, Lubos and Robert F. Stambaugh, 2009, Are Stocks Really Less Volatile in the Long Run?, Working Paper 14757, <http://www.nber.org/papers/w14757>
- [35] Paye, B.S., Timmermann A.(2006), Instability of return prediction models. *Journal of Empirical Finance*, 13, 274-315
- [36] Pesaran M.H., A.Pick, and A.G.Timmermann(2010), Variable Selection Estimation and Inference for Multi-period forecasting problems, CEPR working paper 7139
- [37] Poterba, J. M. "Demographic Structure and Asset Returns." *The Review of Economics and Statistics*, 83 (2001), 4, 565-584.
- [38] Rapach D.E., Wohar M.E., 2006, In-sample vs. out-of-sample tests of stock return predictability in the context of data-mining, *Journal of Empirical Finance*, 13, 231-247
- [39] Robertson D. and S.Wright, 2009, The Limits to Stock Return Predictability, mimeo
- [40] Oleg Rytchkov, 2008, Filtering Out Expected Dividends and Expected Returns, Working paper.
- [41] Siegel Jeremy.J.(1998), *Stocks For the Long-Run*, 2nd edition, McGraw-Hill
- [42] Shiller, Robert J., 2005, *Irrational Exuberance*, second edition, Princeton University Press.
- [43] Schotman, Peter C., Rolf Tschernig, and Jan Budek, Long Memory and the Term Structure of Risk, *Journal of Financial Econometrics*, Fall 2008, 6 (4), 459-495.
- [44] Stambaugh, R. F., 1999. Predictive regressions. *Journal of Financial Economics* 54, 375–421.

- [45] Taleb, N.N., 2001, Fooled By Randomness. The Hidden Role of Chance in the Markets and in Life, TEXERE XLC
- [46] Valkanov, R., 2003. Long-horizon regressions: Theoretical results and applications. *Journal of Financial Economics* 68, 201–232. 33

A Robustness

The effect of overlapping observations in the system described in equation (7) is that the error term is serially correlated, as it features a moving average structure, and therefore conventional OLS standard errors are incorrect. In Table 1 the reported standard errors are corrected for the presence of serial correlation, using a Generalised Method of Moments (GMM) estimator. As an additional test, we examined non-overlapping regression for the unrestricted system. This approach of course does not capture all of the information in the data but it has the advantage of being transparent and providing a baseline estimate. The results, available upon request from the corresponding author, show that the parameters δ_{1m} and δ_{2m} are all significant and very close to the values reported in Table 1. The problem now is that, e.g. at horizon $m = 10$ we have only ten observations. To deal with this problem we have adopted the approach proposed by Pesaran, Pick, and Timmermann (2010) that allows to capture all of the information in the data by choosing the appropriate weighting matrix to accommodate overlapping observations. In particular, following the notation of Pesaran, Pick, and Timmermann (2010), for the fixed horizon m we can estimate $\beta = \{\delta_{1m}, \delta_{2m}\}$ from pooled regressions of m non-overlapping regressions:

$$y_{j+(i-1)m} = \beta x_{j+(i-1)m-m} + v_{i,j} \quad i = 1, 2, \dots, \frac{T}{m} \text{ and } j = 1, 2, \dots, m$$

where $y_{j+(i-1)m}$ is the series of m -period cumulated returns sampled every m period with offset j and analogously $x_{j+(i-1)m-m}$ are the sampled regressors MY and dp_t . Now we can view the m regressions as a set of seemingly unrelated regression equations, allowing for cross dependence of the errors. For each horizon m we estimate $\hat{\beta} = (\delta_{0m}, \delta_{1m}, \delta_{2m})$ as $[W'\Sigma_u^{-1}W]^{-1}W'\Sigma_u^{-1}\tilde{y}$ and $Var(\hat{\beta}) = [W'\Sigma_u^{-1}W]^{-1}$ where \tilde{y} and W are the reordered matrices of regressand and regressors and Σ_u is the appropriate GLS covariance matrix whose analytic form is reported in Pesaran, Pick, and Timmermann (2010). Importantly as Pesaran, Pick, and Timmermann (2010) highlight, since the direct regression is derived from the forward solution of the Campbell-Shiller return log-linearization, the GLS covariance matrix and the MA coefficients of the errors term can be linked to a set of deeper parameters, namely $\{\varphi_{22}, \rho\}$ by just recalling the following

$$u_{t+m} = \sum_{j=1}^m \rho^{j-1} (\varepsilon_{1,t+j}) - \rho^m \sum_{j=1}^m \varphi_{22}^{j-1} \varepsilon_{2,t+m+1-j}$$

The corrected t-statistics obtained by imposing the above error terms structure when we estimate the GLS covariance matrix are reported in Table 1 in the line GLS-PPT. Alternatively we also estimate the GLS covariance matrix Σ_u without these restrictions. Results are almost identical.

Table 1: System Estimation (1910-2009)

| horizon m in years | | | | | | | | | | |
|--|--|--------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| $dp_{t+1} = \varphi_{20} + \varphi_{22}dp_t + \varphi_{23}MY_{t-j} + \varepsilon_{2t+1}$ | | | | | | | | | | |
| UM: | $\frac{1}{\sqrt{m}} \sum_{j=1}^m r_{t+j}^s = \delta_{0m} + \frac{\delta_{1m}}{\sqrt{m}} dp_t + \frac{\delta_{2m}}{\sqrt{m}} \left(\sum_{j=1}^m \varphi_{22}^{j-1} MY_{t+m+1-j} \right) + u_{t+m}$ | $m = 1, \dots, 10$ | | | | | | | | |
| RM: | $\frac{1}{\sqrt{m}} \sum_{j=1}^m r_{t+j}^s = \delta_{0m} + \frac{1}{\sqrt{m}} (1 - \varphi_{22}^m) dp_t - \frac{\varphi_{23}}{\sqrt{m}} \left(\sum_{j=1}^m \varphi_{22}^{j-1} MY_{t+m+1-j} \right) + u_{t+m}$ | | | | | | | | | |
| UM | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| δ_{1m} | 0.18 | 0.38 | 0.48 | 0.61 | 0.70 | 0.73 | 0.78 | 0.86 | 0.89 | 0.89 |
| (<i>t-stat</i>) | (3.87) | (5.57) | (5.61) | (6.96) | (7.99) | (7.17) | (7.25) | (8.73) | (7.63) | (6.15) |
| (<i>GLS-PPT</i>) | (-) | (4.54) | (5.21) | (5.77) | (7.43) | (5.19) | (7.03) | (8.61) | (7.57) | (2.67) |
| δ_{2m} | 0.41 | 0.52 | 0.56 | 0.64 | 0.69 | 0.70 | 0.74 | 0.78 | 0.81 | 0.83 |
| (<i>t-stat</i>) | (3.23) | (3.69) | (3.85) | (4.28) | (4.55) | (4.69) | (4.78) | (4.93) | (4.94) | (5.01) |
| (<i>GLS-PPT</i>) | (-) | (3.85) | (3.50) | (3.87) | (4.21) | (4.10) | (4.06) | (4.10) | (4.19) | (4.84) |
| φ_{22} | 0.61 | | | | | | | | | |
| (<i>t-stat</i>) | (9.21) | | | | | | | | | |
| φ_{23} | -0.83 | | | | | | | | | |
| (<i>t-stat</i>) | (-3.79) | | | | | | | | | |
| RM | | | | | | | | | | |
| φ_{22} | 0.76 | | | | | | | | | |
| (<i>t-stat</i>) | (19.31) | | | | | | | | | |
| φ_{23} | -0.53 | | | | | | | | | |
| (<i>t-stat</i>) | (4.41) | | | | | | | | | |
| χ_{12}^2 | 13.45 | χ_{20}^2 | 17.19 | | | | | | | |
| | (0.34) | | (0.64) | | | | | | | |
| σ_{DepVar} | 0.195 | 0.198 | 0.187 | 0.185 | 0.181 | 0.174 | 0.172 | 0.173 | 0.171 | 0.168 |
| $\sigma_{u_{t+m}}$ | UM | 0.188 | 0.179 | 0.164 | 0.152 | 0.140 | 0.131 | 0.125 | 0.118 | 0.109 |
| $\sigma_{u_{t+m}}$ | RM | 0.189 | 0.179 | 0.164 | 0.152 | 0.141 | 0.133 | 0.127 | 0.120 | 0.112 |
| $adjR^2$ | UM | - | 0.18 | 0.24 | 0.33 | 0.41 | 0.44 | 0.48 | 0.54 | 0.57 |
| $adjR^2$ | RM | 0.06 | 0.18 | 0.23 | 0.32 | 0.40 | 0.42 | 0.46 | 0.52 | 0.55 |

Table 1: This table compares the univariate OLS long-horizon regression coefficients, to the GMM estimates that impose the restrictions suggested by the present-value model with demographics. The estimation is by GMM, where the moments are the OLS normal conditions. Standard errors are by Newey-West with optimal bandwidth selection. The first-stage weighting matrix is the identity matrix. *GLS - PPT* is the t-stat that explicitly accounts for the MA(m-1) errors structure by using a GLS covariance matrix as suggested in Pesaran, Pick and Timmermann (2010). σ_{DepVar} is the annualized unconditional standard deviation. $\sigma_{u_{t+m}}$ is the annualized conditional standard deviation of the compounded (over m periods) returns, i.e. our measure of stock market risk. The effective sample period is 1910-2009.

Table 2: A simple bivariate VAR (1910-2009)

| $(r_{t+1}^s - E_{r^s}) = \varphi_{12} (dp_t - E_{dp}) + \nu_{1,t+1}$ $(dp_{t+1} - E_{dp}) = \varphi_{22} (dp_t - E_{dp}) + \nu_{2,t+1}$ | | | | | | | |
|--|----------------|----------------------------------|------------|------------|--|-----------------------|----------------------|
| φ_{12} | φ_{22} | χ_2^2 | σ_1 | σ_2 | $\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}$ | $adj R_{r_{t+1}^s}^2$ | $adj R_{dp_{t+1}}^2$ |
| (t-stat) | (t-stat) | $\varphi_{11}=0, \varphi_{21}=0$ | | | | | |
| 0.067 | 0.892 | 5.673 | 0.194 | 0.219 | -0.856 | 0.02 | 0.78 |
| (1.70) | (18.80) | (0.06) | | | | | |

Table 2: The table reports coefficient estimates (with t-statistics in parentheses) and the R^2 statistic for each equation. We also report the standard deviations and correlations of residuals.

Table 3: A three-variate VAR with imperfect predictors (1910-2009)

| $(r_{t+1}^s - E_{r^s}) = (\mu_t - E_{r^s}) + \nu_{1,t+1}$ $(dp_{t+1} - E_{dp}) = \varphi_{22} (dp_t - E_{dp}) + \nu_{2,t+1}$ $(\mu_{t+1} - E_{r^s}) = \varphi_{33} (\mu_t - E_{r^s}) + \nu_{3,t+1}$ | | | | |
|---|----------------|--|--|--|
| φ_{22} | φ_{33} | σ_1 | σ_2 | σ_3 |
| (low-upp) | (low-upp) | | | |
| 0.847 | 0.939 | 0.187 | 0.285 | 0.171 |
| [0.757 0.934] | [0.811 0.993] | [0.166 0.214] | [0.251 0.327] | [0.088 0.335] |
| Pred R^2 | | $\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}$ | $\frac{\sigma_{13}}{\sigma_{11}\sigma_{33}}$ | $\frac{\sigma_{23}}{\sigma_{22}\sigma_{33}}$ |
| 0.04 | | -0.631 | -0.655 | 0.451 |
| | | [-0.732 -0.502] | [-0.841 -0.325] | [0.205 0.623] |

Table 3: This table shows the posterior median and [0.025 0.975] quantile obtained with the predictive system described in Pastor and Stambaugh (2008). The sample period is 1910-2009.

Table 4: A bi-variate VAR with MY (1910-2009)

$$r_{t+1}^s = \varphi_{10} + \varphi_{12}dp_t + \varphi_{13}MY_{t+1} + \nu_{1,t+1}$$

$$dp_{t+1} = \varphi_{20} + \varphi_{22}dp_t + \varphi_{23}MY_{t+1} + \nu_{2,t+1}$$

| φ_{12} (<i>t-stat</i>) | φ_{13} (<i>t-stat</i>) | φ_{22} (<i>t-stat</i>) | φ_{23} (<i>t-stat</i>) | χ_2^2 $\varphi_{11}=0, \varphi_{21}=0$ | σ_1 | σ_2 | $\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}$ | $adj R_{r_{t+1}^s}^2$ | $adj R_{dp_{t+1}}^2$ |
|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|--|------------|------------|--|-----------------------|----------------------|
| 0.179 | 0.410 | 0.728 | -0.603 | 4.74 | 0.188 | 0.207 | -0.846 | 0.07 | 0.80 |
| (3.07) | (2.67) | (11.33) | (-3.56) | (0.09) | | | | | |

Table 4: The table reports coefficient estimates (with t-statistics in parentheses) and the R^2 statistic for each equation. We also report the standard deviations and correlations of residuals.

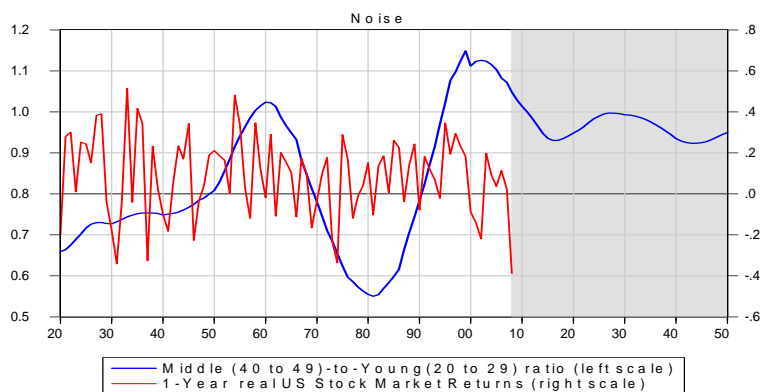


Figure 1.1: 1-year real US stock market returns and demographic trends

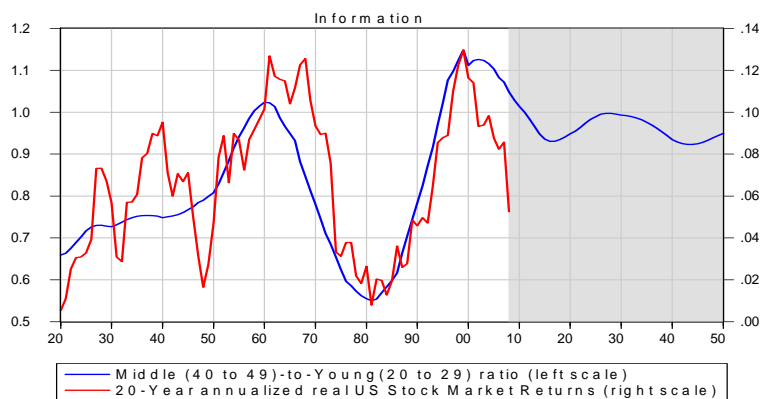


Figure 1.2: 20-year real US stock market returns and demographic trends (ex post-returns over the period $t - (t + j)$ are plotted along MY_{t+j}).

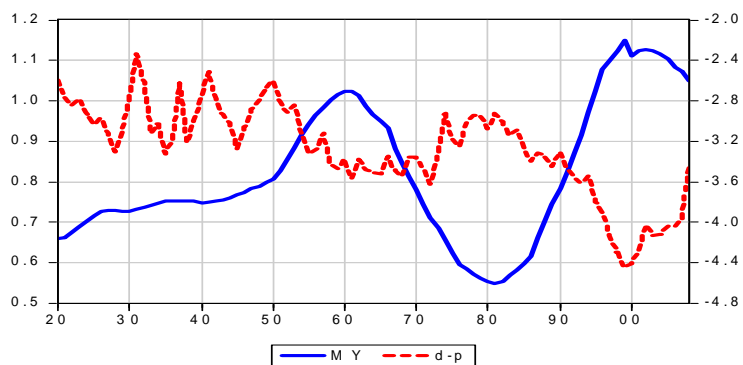


Figure 1.3: MY and the log dividend-price ratio

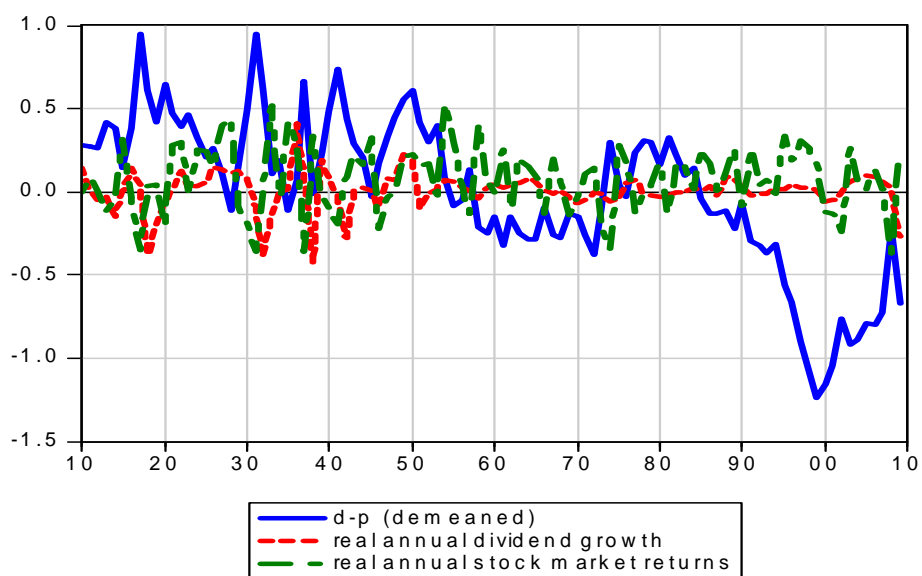


Figure 2.1: the three endogenous variable in our structural model

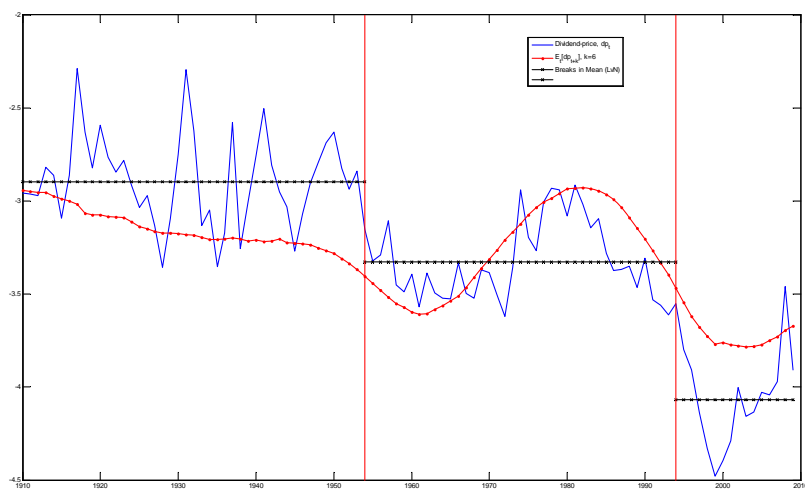


Figure 2.2: dp_t , the time varying linearization point used in our model, and the breaks identified by LVN.

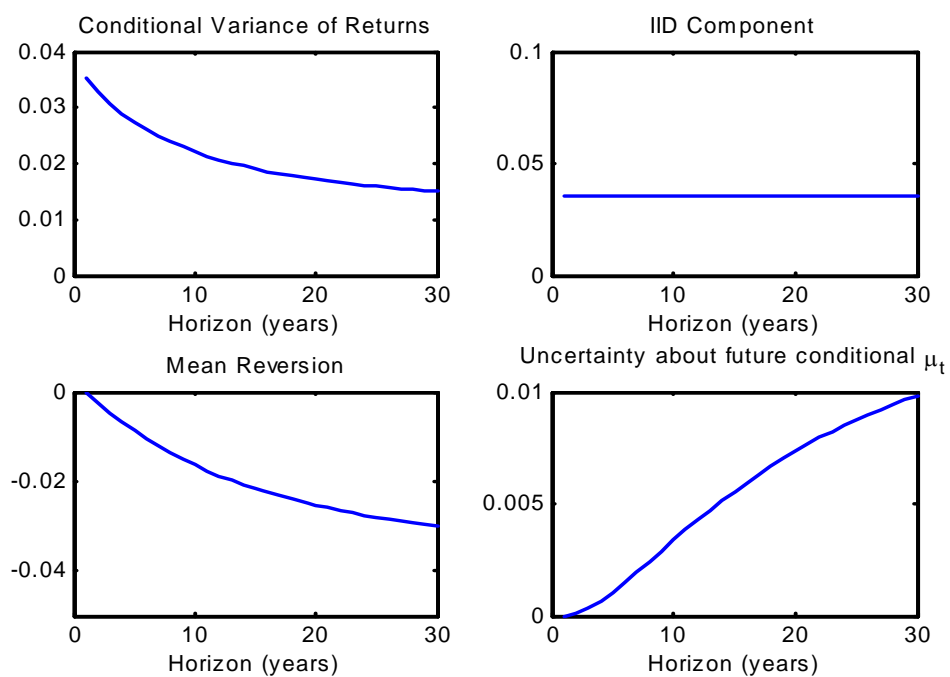


Figure 3: The TS of stock market risk from a bi-variate VAR

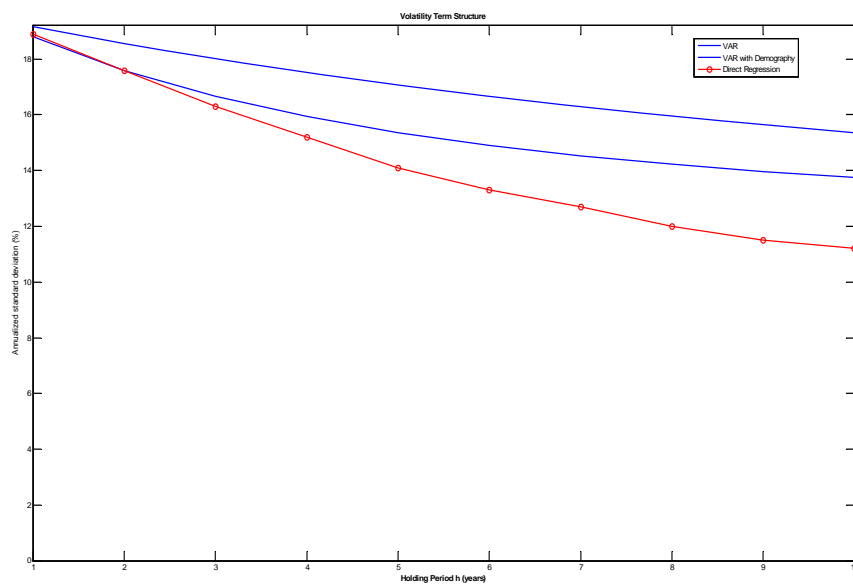


Figure 4: Three alternative measures of the TS of stock market risk

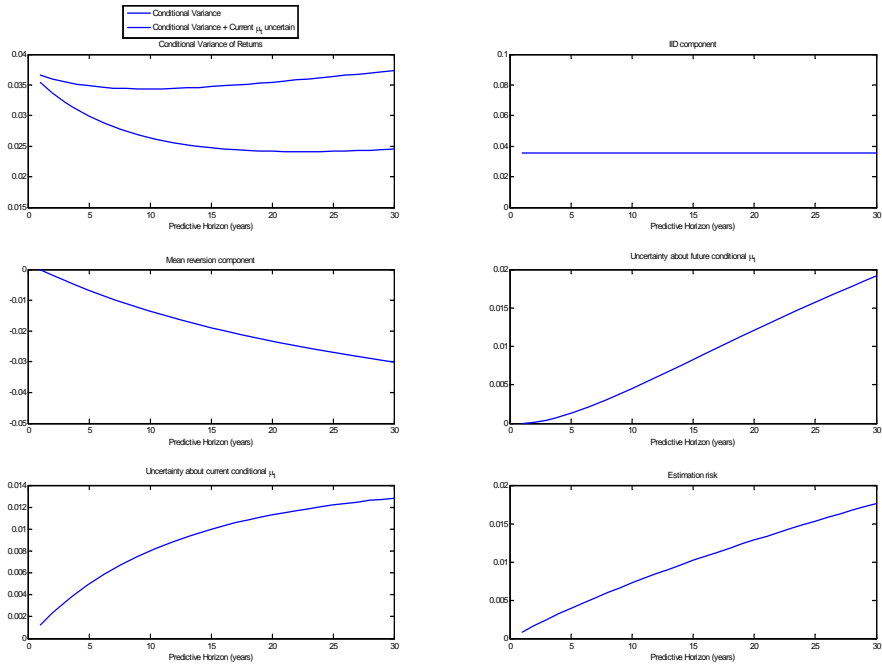


Figure 5: The TS of US stock market risk from a three-variate VAR with an unobservable component

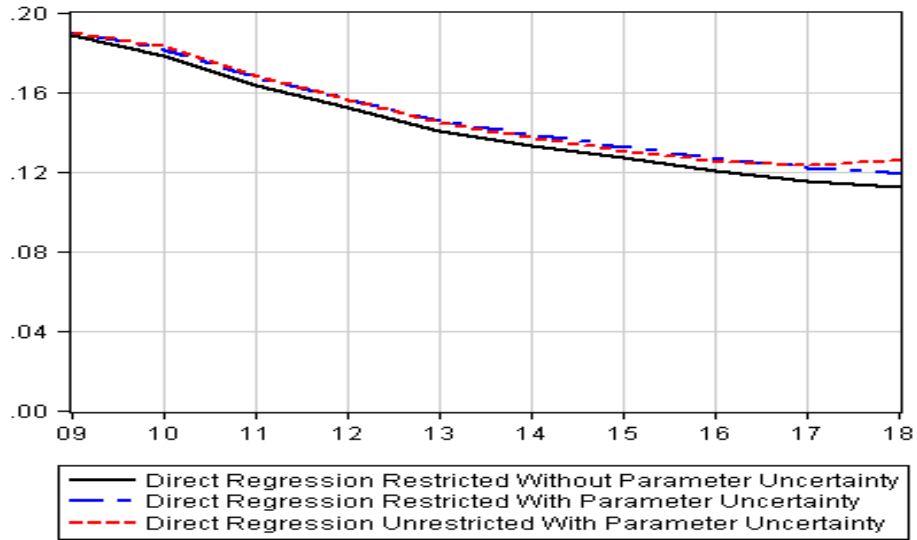


Figure 6: The term structure of stock market risk in a structural model with and without parameters' uncertainty.

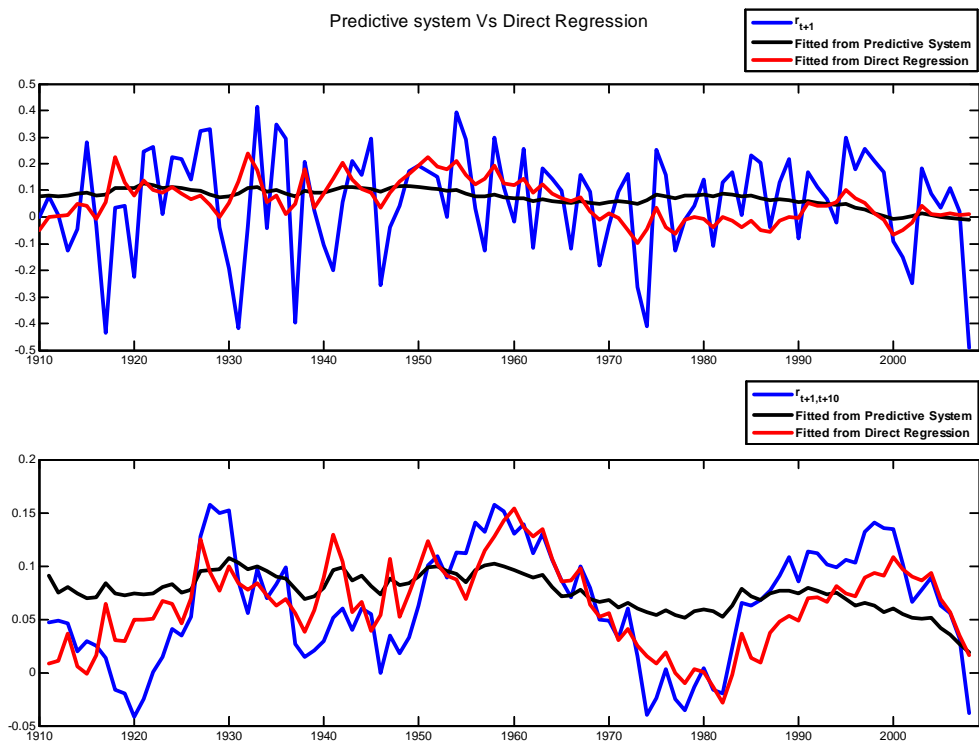


Figure 7: Noise and Information reconsidered