GMM estimation of the Long Run Risks model

February 14, 2016

Abstract

In this paper, we propose a Gmm estimation of the structural parameters of the Long Run Risk model that allows for the separation between the consumer optimal decision's frequency and the frequency by which the econometrician observes the data. Our inference procedure is also robust to weak identification. The key finding is that the Long Run Risk model adapts well to the data but could not be so good at forecasting or telling the true story about what drives the evolution of asset prices. Indeed, the model is able to reproduce the qualitative behavior of targeted moments in the long run when the corresponding estimates of the structural parameters are used for simulations, but it also faces a urge tension in keeping in track with all the observed moments considered.

1 Introduction

One of the long lasting challenges in asset pricing is to provide a model that could explain the observed level of asset prices and reconcile model's predictions with the observed macro-finance data. Three main competing models, which focus on the decision problem of a representative agent, have received a great attention in this domain : The Habit-formation model (Abel, 1990; Constantinides, 1990; Cochrane and Campbell, 1999), the long run risks model (Bansal and Yaron, 2004) and more recently, the rare disasters model (Rietz, 1988, Barro 2006, 2009; Wachter, 2013). Those alternative models are usually able to account for the observed level of equity premium and risk free rate with a reliable value of the representative agent risk aversion coefficient.

However, each of this model has received some criticisms related to their capacity to yield the same predictions as what have been documented in previous studies from the observed data, especially when we look at higher order moments of the models predictions. Indeed, just to take some few examples, the main intuition behind the habit model is that the drop of consumption toward the habit level rises the risk aversion of the representative agent who asks for a higher equity premium as compensation. So this model implies that past consumption growth should predict future prices but that has not been observed in the data (Bansal et al., 2012a). In the case of the rare disasters model, the eventuality of a rare but large drop in consumption induces a higher equity premium. However, the implied consumption growth process necessary to generate the level of the U.S historical average equity premium is much more volatile and skewed than what we observe in the historical data[1] ([Julliard and Ghosh, 2012]).

More recently, there have also been some criticisms toward the Long run risks (henceforth LRR) model concerning the predictive power that the model confers to asset prices toward cash flows variables. Indeed, as pointed by Beeler and Campbell [2012], the initial calibration of the model implies that the price-dividend ratio has a huge predictive power on the consumption and dividend growth and that is at odd with what is observed in the data. As we know from Cochrane, 2001 Chap. 20, the observed variability in log price-dividend ratio implies that it must predict future excess returns or future dividend growth; but in the data, the log price-dividend ratio seems to predict future excess return and it does not predict future dividend

[1]This is mainly because the size of the annual drop in consumption growth considered in this model corresponds in the data to a pluri-annual summation of consecutive drops. In other words, the probability that such a drop (the one considered in the model) would be observed in the data is more smaller than the rare disaster probability considered to generate the observed equity premium.

growth. A subsequent calibration of the model in Bansal et al. [2012a][2] has weakened this predictive power and done a great job in reconciling the model with the data; hence showing that the model implications could be improved just by changing the parameters calibration.

Even though this new calibration of the model has brought the model implications closer to the data by reducing the predictability of future dividend and consumption growth by the asset prices, it has failed to captured the observed predictability of future excess returns by the log price-dividend ratio. Some other mismatches between the model implications and the observed data have also been pointed out and they call for the seek of at least a better calibration of the model in order to reconcile the model predictions, especially in terms of moments matching, with what is observed in the data as advocated by Bansal et al. [2012a], Beeler and Campbell [2012].

The aim of this paper is to estimate the structural parameters that will enable the model to well fit a set of empirical moments derived from the observed data. While this estimation is of particularly importance to test some key assumptions of the model like for example the Elasticity of Inter-temporal Substitution (EIS) being greater than 1, very few studies have attempted to estimate the fundamental parameters of the LRR model. Indeed, the simulation of the model appears to be pretty easy, but the estimation of the fundamental parameters is more challenging due to the unobserved latent variables, the preference structure (Epstein-Zin recursive utility function) and the large number of parameters to be estimated. However, contrary to the calibration exercise which can look like a "fishing expedition", the estimation allows to determined the set of parameters that will bring the model predictions closer to the data in terms of moments we considered, which could also be easily extended to take into account some new stylized facts. Bansal et al. [2012a] recommend the use of a Simulated Method of Moment (SMM) for the estimation of the model parameters instead of a successive regressions approach which produces a downward bias of some coefficients like the EIS and a upward bias of others like the risk aversion coefficient. We try to go further in this direction by deriving the analytical formulas for theoretical moments in order to use the Generalized Method of Moment (GMM) which is more efficient than the SMM.

In this paper, we extend the parametric estimation of the LRR model as done

[2]In the Bansal and Yaron, 2004 (henceforth, BY) model the emphasize is put on the persistent of the expected component of the consumption growth process as the main driver of asset prices, while in the Bansal et al. [2012a] (henceforth, BKY) formulation it is the stochastic volatility of the consumption growth process that determines the equity premium level Beeler and Campbell [2012].

by Constantinides and Ghosh [2011], to take into account some key failures of the previous calibrations of the model pointed by Bansal et al. [2012a], Beeler and Campbell [2012]. Our procedure is closely related to their; the difference being that firstly, we consider an extended version of the LRR model as in BKY where we allow the dividend growth process to be correlated to the short run risk; secondly, we extend the set of moments to be matched, in order to consider some critics formulated against the previous calibrations of the model; and thirdly, we disentangle the agent decision frequency from the one at which the econometrician observes the data through temporal aggregation. Indeed, the LRR model is usually simulated assuming that the agent optimal allocation is made at the monthly frequency while the econometrician observe the data at the annual frequency. We assume that this mapping between the agent decision frequency and the data observation frequency by the econometrician is unknown and could be estimated. Our approach is closely related to Bansal et al. [2012b]who introduced the time aggregation in the estimation of the LRR model, but contrary to them we used a weak identification robust method for the inference. Other approaches have also been used to estimate modified versions of the LRR model. Bansal et al. [2007a] assumed that consumption and dividend growth were co-integrated and used the Efficient Method of Moments (EMM) for the estimation. Schorfheide et al. [2014] disentangled the volatilities of the cash flow processes (consumption and dividend growth), introduced a shock to preferences and used a Bayesian Markov Chain Monte Carlo (MCMC) methods for the estimation[3]. Calvet and Czellar [2015] used an indirect inference method with a restricted version of the LRR model that has a closed form solution as auxiliary model and in their estimation procedure they do not used the market return.

The remaining of the paper is organized as follow: In section 2 we provide the model statement and derive the asset prices formulas in term of underlying parameters. Section 3 provides some comparisons of the model predictions and the data for two common used calibrations of the model. It also highlight the successes and failures of those calibrations and points out the problem we will try to solve by estimating a new set of parameters. Section 4 presents the estimation procedure and some result of the simulations we did in order to verify the reliability of our method before applying it on the observed data. In section 5, we present the results of our estimations and discuss the results. Section 6 concludes. An appendix is provided in section 7 with time aggregation derivations and the analytical formulas of the theoretical moments.

[3]Contrary to them, we prefer to remain parsimonious in terms of model's parameters as our objective is to improve the basic LRR model results through our estimation procedure.

2 The model and the solution

The long run risks model initiated by Bansal and Yaron [2004] (BY) provides an explanation to many asset pricing puzzles in the context of the U.S. Economy and the calibrations of the model considered in BY and subsequently in Bansal et al. [2007b] seems to well fit many first and second order moments of the U.S. historical macro-finance data. The success of the model lies in particular in the use of the Epstein and Zin [1989] utility function[4] with an elasticity of inter-temporal substitution (EIS) greater than 1, which features a preference for an early resolution of time uncertainty and the introduction of two predictable latent variables (a persistent component and a time varying volatility) to describe the evolution of the cash flows (consumption and dividend growth) and asset prices processes.

2.1 Model statement

We consider a representative agent with Epstein-Zin recursive preferences. For these preferences the Euler equation satisfied by the representative agent's optimal allocation plan is given by:

$$E_t[exp(\theta \log \delta - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r_{c,t+1} + r_{i,t+1})] = 1 \tag{1}$$

with $\theta = \frac{1-\gamma}{1-\frac{1}{\psi}}$, γ is the coefficient of risk aversion and ψ is the elasticity of inter-temporal substitution (EIS); g_t is the consumption growth process, δ is the discount rate, $r_{i,t}$ is the log-return on a given asset i and $r_{c,t+1}$ is the log-return on wealth portfolio that delivers aggregate consumption as its dividend each period. The evolution of the growth process is driven by x_t the persistent component of the consumption growth process which is unobservable but seems to play a big role in the model andσ_t is the stochastic volatility process.

$$g_{t+1} = \mu + x_t + \sigma_t \eta_{t+1} \tag{2}$$

$$x_{t+1} = \rho x_t + \varphi_e \sigma_t e_{t+1}$$

$$\sigma^2_{t+1} = \bar{\sigma}^2 + \nu(\sigma^2_t - \bar{\sigma}^2) + \sigma_w w_{t+1}$$

$$g_{d,t+1} = \mu_d + \phi x_t + \pi \sigma_t \eta_{t+1} + \varphi_d \sigma_t u_{t+1}$$

$$w_{t+1}, e_{t+1}, u_{t+1}, \eta_{t+1} \sim N.i.id(0,1)$$

$e_{t+1}, \eta_{t+1}, w_{t+1}$represent the three sources of risk in the aggregate consumption

[4]The Epstein-Zin recursive utility function allows to disentangle the time preference (characterized by the elasticity of inter-temporal substitution) from the risk aversion. Indeed, the fact that someone is highly risk averse does not necessarily imply that she will also want to smooth consumption over time [Mehra, 2003].

dynamics that are priced in the long run risks model and correspond respectively to the short run risk in consumption, the long run risks and the volatility risk. π represents the magnitude of the influence of short run consumption shocks on the dividend growth process. In The Bansal and Yaron [2004] calibration, this coefficient is equal to zero, while it is equal to 2.6 in Bansal et al. [2012a]; furthermore it has not been considered in the Constantinides and Ghosh [2011] estimation of the model parameters and so, this omission could be damaging for the estimation of the other parameters of the model or could implied a rejection of the model. The exposure of the dividend growth process to the long run consumption shocks is captured by ϕ.

2.2 Solution

To solve the model, Bansal and Yaron [2004] assume that the price-consumption ratio, the price-dividend ratio and the risk free rate are affine functions of the states variables.

$$\begin{aligned} z_t &= A_0 + A_1 x_t + A_2 \sigma_t^2 \\ z_{m,t} &= A_{0,m} + A_{1,m} x_t + A_{2,m} \sigma_t^2 \\ r_{f,t} &= A_{0,f} + A_{1,f} x_t + A_{2,f} \sigma_t^2 \end{aligned} \tag{3}$$

They combined this with the Campbell and Shiller [1988] approximation of the returns on the wealth and market portfolios; $r_{c,t+1} = k_0 + k_1 z_{t+1} - z_t + g_{t+1}$ and $r_{m,t+1} = k_{0,m} + k_{1,m} z_{m,t+1} - z_{m,t} + g_{d,t+1}$ with $k_0 = \log(1+e^{\bar{z}}) - k_1 \bar{z}$, $k_1 = \frac{e^{\bar{z}}}{1+e^{\bar{z}}}$,$k_{0,m} = \log(1 + e^{\bar{z}_m}) - k_1 \bar{z}_m$ and $k_{1,m} = \frac{e^{\bar{z}_m}}{1+e^{\bar{z}_m}}$. Plugging the system of equations (1) and (2) in the Euler equation, we can easily derive the analytical formulas for the coefficients A_0, A_1, A_2, A_{om}, A_{1m}, A_{2m}, A_{0f}, A_{1f} and A_{2f} conditionally on k_0,k_1,$k_{0m}$and k_{1m}which also depend on $\bar{z}$ and $\bar{z}_m$[5]. So, the model is solved by firstly determining the mean values of the log price-consumption and log price-dividend ratios through the resolution of a fixed point problem, secondly the $k's$ are computed and finally the $A's$ coefficients are determined. All this is done using a fixed point algorithm which converges very quickly when the fixed point is feasible. Indeed, the fixed point feasibility has not been emphasized in the previous literature. However, when dealing with an optimization algorithm, this feasibility becomes very relevant because when not satisfied it leads to a crash of the routine. To obtained the fixed point feasibility

[5]In many studies of the LRR involving the Campbell and Shiller [1988] approximation, the values of $\bar{z}$ and $\bar{z}_m$are calibrated instead of been determined as the solutions of the fixed point problem. This could lead to a bias estimation of the true parameters of the model. Indeed, we run a simulation experiment showing that calibrating$\bar{z}$ and $\bar{z}_m$to the wrong values leads to a rejection of the true parameters of the model and this happened even when the calibration error is less than 5%.

condition, we substitute k_0, k_1,A_0, A_1, A_2 in the expression of $\bar{z} = A_0 + A_2\bar{\sigma}^2$ and obtain the following equation:

$$f(\bar{z}) = g(\bar{z}) \tag{4}$$

where

$$f(\bar{z}) = \bar{z} - \log(\delta) - (1 - \tfrac{1}{\psi})\mu_c - \log(1 + e^{\bar{z}}) \text{ and}$$
$$g(\bar{z}) = \lambda_1 \left(1 + U(\bar{z}, \varphi_e, \rho)^2\right) + \lambda_1 \left(1 + U(\bar{z}, \varphi_e, \rho)^2\right)^2 U(\bar{z}, \lambda_0, \nu)^2$$

with $\lambda_0 = 0.5(1-\gamma)\frac{\sigma_w}{\bar{\sigma}}$, $\lambda_1 = 0.5(1-\gamma)(1-\frac{1}{\psi})\bar{\sigma}^2$ and $U(z, a, b) = \frac{ae^z}{1+e^z(1-b)}$

The function f is increasing : $\mathbb{R} \to]-\infty; F[$ and the function

$$g \text{ is } \begin{cases} decreasing : \mathbb{R} \to]G_2; G_1[& if \quad \lambda_1 < 0 \\ \\ increasing : \quad \mathbb{R} \to]G_1; G_2[& if \quad \lambda_1 > 0 \end{cases}$$

where
$F = \lim\limits_{z \to +\infty} f(z) = -\log(\delta) - (1 - \frac{1}{\psi})\mu_c$, $G_1 = \lim\limits_{z \to -\infty} g(z) = \lambda_1$ and

$$G_2 = \lim_{z \to +\infty} f(z) = \lambda_1 \left(1 + \left(\frac{\varphi_e}{1-\rho}\right)^2\right)\left(1 + \left(1 + \left(\frac{\varphi_e}{1-\rho}\right)^2\right)\left(\frac{\lambda_0}{1-\nu}\right)^2\right)$$

Equation 4 put a restriction on the preference parameters (δ, γ, ψ) of the representative agent for a given set of consumption growth parameters. For a given set of parameters to be feasible, the pure discount factor should satisfy the following conditions:

$$\begin{cases} \delta \leq e^{-(1-\frac{1}{\psi})\left(\mu_c + \frac{G_2}{1-\frac{1}{\psi}}\right)} & if\ \gamma > 1, \psi > 1\ or\ \gamma < 1, \psi < 1 \\ \\ \delta \leq e^{-(1-\frac{1}{\psi})\left(\mu_c + \frac{G_1}{1-\frac{1}{\psi}}\right)} & if\ \gamma > 1, \psi < 1\ or\ \gamma < 1, \psi > 1 \end{cases}$$

For some vector of parameters, for example if $F < G_2$ while $\lambda_1 < 0$ or $F < G_1$ while $\lambda_1 > 0$, there is no fixed point. A similar condition also derives from the log price dividend ration equilibrium condition (see Appendix). As illustrate in figure 1, the increase of the unconditional mean of consumption growth restricts the feasible combination of agent risk aversion and elasticity of inter-temporal substitution; and the increase of unconditional mean of the dividend growth process further restrict it. Furthermore, once some parameters are kept fixed, the comparison between F and G's determines the fixed point feasible set for the remaining parameters. In

the case where $\gamma > 1$and $\psi > 1$ for some values of the unconditional mean of the consumption growth process (μ_c), the upper bound of the pure discount factor will be lower than 1 and in this case, it is an increasing function of the relative risk aversion coefficient (γ). If $\delta = 1$ then there exist $\bar{\gamma}$ such that there is a solution to the fixed point problem only if ($\gamma > \bar{\gamma}$ and $\psi > 1$) or ($\gamma < \bar{\gamma}$ and $\psi < 1$).

This condition is the same as the ones obtained by Burnside [1998] with a CRRA utility function and a Gaussian consumption growth process and extended in de Groot [2015] to a stochastic volatility consumption growth process. They obtained a **sufficient** condition for the price dividend ratio to exist[6], but contrary to them our condition is a **necessary**[7] condition for an equilibrium finite value of the price dividend ratio to exist and it is determined after applying the Campbell and Shiller [1988] approximation. In a model with CRRA utility function ($\gamma = \frac{1}{\psi}$), our condition is necessary and sufficient (analytically it is the same as the one obtained by de Groot [2015]) for the existence of the price-consumption ratio, but with the Epstein-Zin recursive utility function, we show that it is much more less restrictive compare to the one obtained through the series convergence criterion. Creal and Wu [2015] also studied the fixed problem in a consumption based model with recursive preferences, where they introduced some shocks to preferences and affine dynamics of the state variables. There conclusions are similar to the ones we obtained here.

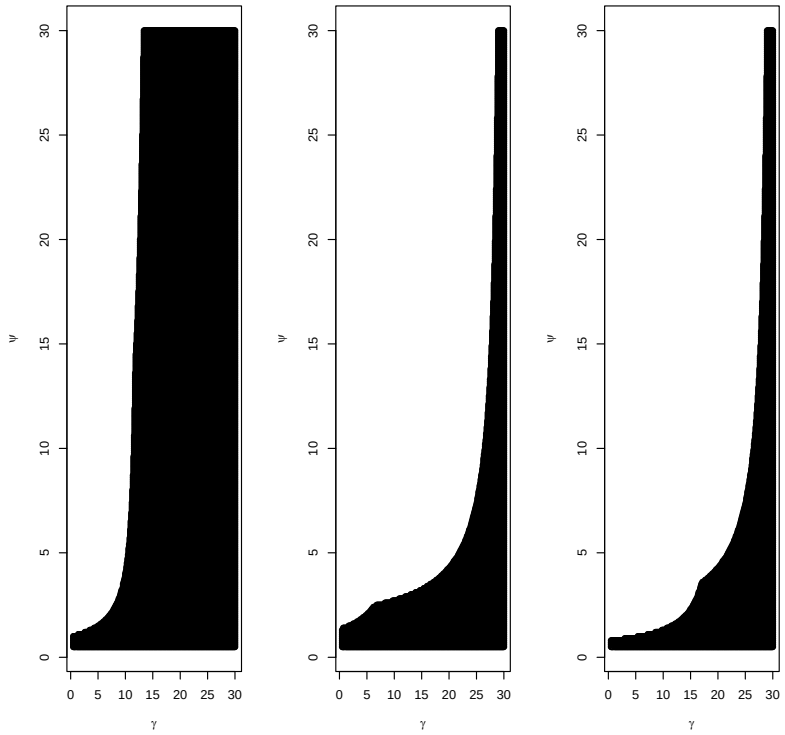
To handle this fixed point existence problem in the estimation, we set a ruling out condition in the optimization algorithm for all the parameter's vector for which the fixed point does not exist; in the case where the algorithm is trying such a vector of parameters, the objective function is automatically set to a very high value.This penalty value imposes the optimizers to seek the parameters in other regions (Holly et al. [2011]). By trying various combinations of parameters, the optimization algorithm allows to expand the set of possible parameter values compared to a calibration exercise where some parameters are given less attention (like for example, the ones driving the consumption and dividend processes) and are preliminary fixed, hence constraining the set of feasible parameters. This fixed point feasibility might also cause difficulties to optimization method based on the gradient if its analytical expression is not provided. Indeed, as the numerical approximation of the gradient

[6]When the stochastic volatility is shut down, Their condition is is given by $\delta \leq e^{-\left((1-\gamma)\mu_c + 0.5\left(\frac{1-\gamma}{1-\rho}\right)^2\sigma^2\right)}$, which is obtained from the convergence criterion of a given series

[7]When the sufficient condition is satisfied, we are sure that the price dividend ratio exist, but it does not mean that it will not exist when the condition is not satisfied, instead the sufficient condition can be too much restrictive and should not be used to construct the set of unfeasible parameters. On the other hand, when the necessary condition is satisfied, the price dividend ratio could possibly exist and we are sure that it does not exist when that condition is not verified, so this condition in more appropriate to construct the set of unfeasible parameters.

value needs two points where the objective function will be evaluated, one of the two might fall in the infeasible region. That is why we preferred an optimization method based uniquely on the objective function (The Nelder and Mead [1965]'s algorithm).

Figure 1: Feasible set forγ and ψ with the remaining parameters fixed



This figure shows the feasible values (for which the model can be simulated) of RRA (γ) and EIS (ψ) when all the other parameters are kept fixed. We restrict the values of γ and ψ in $[0.5, 30]$. In the baseline picture, the value of the other parameter are as in the BYK monthly calibration: $\delta = 0.9989, \mu_c = 0.0015, \mu_d = 0.0015, \phi = 2.5, \varphi = 5.96, \rho = 0.975, \varphi_e = 0.038, \sigma = 0.0072, \nu = 0.999, \sigma_w = 0.0000028, \pi = 2.6$. in the middle picture the value of the average consumption growth is put at $\mu_c = 0.0020$ and in the right picture: $\mu_c = 0.0020, \mu_d = 0.0025$.

2.3 Temporal aggregation

The structural parameters used to simulate the LRR model are usually assumed to generate the data at the monthly frequency. Then these data are aggregated at the annual frequency and the moment are computed to be compared to those computed in the real annual data. We choose to follow this tradition here. But we could also, as done by Bansal et al. [2012b], assume that the frequency at which the model evolves is unknown but we still observed the data at the annual frequency. Then

the aggregation frequency becomes another parameter that could be determined optimally in order to match the observed data. This implies that the annual theoretical moments should be computed while taking the temporal aggregation from an unknown frequency, which would be within a year, to the annual frequency. For example, it could be that the representative agent in the model takes its decisions about the optimal consumption allocation every 2 months; then the annual data will be the results of 6 period's temporal aggregation. Appendix B. presents the analytical formulas for cash flows and asset prices allowing to move from the agent decision frequency to the annual frequency while assuming that there are h-periods within a year. The theoretical moment derived from the aggregate processes will be used to match the corresponding ones observed in the data and this is how our Gmm procedure will estimate the structural parameters of the model to match the observed data.

3 Moments matching

In this section we simulate the model using the standard calibrations as in Bansal and Yaron [2004], Bansal et al. [2012a] and compare the model predictions to the observed annual data. This is done in order to highlight the successes and failures of the previous calibrations of the model relative to the data. Later after the estimation, we will re-do this simulation exercise using the estimated parameters to calibrate the model and see whether the model predictions have been improved. Simulated data from the model are generated at the monthly frequency, then aggregated at the annual frequency to make the comparison with the observed annual data. Our observed annual data set goes from 1926 to 2009, which corresponds to 1008 months.

3.1 Data and Basics moments

We use the U.S annual real data on consumption growth per capita of non durable goods and services from Robert Shiller Website. The financial data (risk free rate, market return, dividend growth and log price dividend ratio) are also from Shiller's Website, the real values are computed using data on inflation from the same source. The log price dividend ratio is obtained by dividing the real prices by the real dividend and taking the log. The real market return is obtained by subtracting the inflation rate from the nominal market return; the same procedure is applied to obtained the real risk free rate. the growth rates are obtained by taking the differences in the log of the corresponding variables.

The simulation procedure is done as described in appendix A.1. The moments computed from the simulated data will give us an idea of how far the model is close to the data by comparing their values with what can be computed using the recorded data. We also report the results obtained by Constantinides and Ghosh [2011]using the same calibration from Bansal and Yaron [2004] as we do. In general the model is very close to the data as far as we consider moments of the consumption growth, dividend growth and market return processes; the moments computed using the observed data are within the 95 % confidence intervals of the model predictions. But for the risk free rate and the log price-dividend ratio, the model predictions are less convincing and in many cases, the observed data moments fall out of the 95 % confidence interval of the model prediction.

Indeed, the average level of the risk free rate predicted by the model is above what has been recorded in the data, meaning that the representative agent in the model is less risk adverse than the real one, or the preference for the future is higher (a lower time preference) in the reality than what the calibrated model assumes. It means that, the risk aversion coefficient or the preference for the future coefficients (the EIS or the pure discount rate) should be increased in the model. As we see in the BKY calibration, the pure discount rate has been increased and the model yields a 95 % confidence interval which now contains the small observed value of the average risk free rate. So increasing the pure discount factor, and by the way lowering the time preference for the representative agent in the model, reduces the demand for saving and allows the model to generate a low average risk free rate that is closer to what is observed in the data. However, some improvement still need to be done. Indeed, the observed volatility of the risk free rate, the volatility and auto-correlation of the log-price dividend ratio are all out of the 95 % confidence intervals predicted by the model for all the calibrations considered.

Table 3.1 also reveals that the BY and the BYK calibrations deliver almost the same results in term of basic moments even though they are quite different. So, by using only the basic moments, we could identify many sets of parameters that could fit the model and in order to discriminate among then we need to add other moment conditions and select among the candidate sets of parameters allowing to fit the basic moments, those that will also be able to fit the added moments.

These observations clearly advocate for the use of an estimation procedure in order to select the parameters that will bring the model closer to the observed data. This exercise will be very important also because the mechanism at play in the model depends on the calibration. Indeed, as it has be pointed by Beeler and Campbell [2012], the relative importance of movements in consumption growth and

volatility is not the same in the BY and BKY calibrations. In the former calibration, time variation in the volatility almost plays no role while in the later, it is a key determinant of the level of the equity premium and the volatility of the stock prices.

Table 1: Basic moments implied by the model

Moments	Data	BY	BKY
$E(\Delta c)$	0.020	0.018 [0.005; 0.031]	0.018 [0.006; 0.030]
$\sigma(\Delta c)$	0.03	0.028 [0.023; 0.035]	0.028 [0.015;0.044]
$AC1(\Delta c)$	0.359	0.473 [0.245; 0.668]	0.402 [0.168;0.608]
$E(\Delta d)$	0.008	0.010 [-0.034;0.054]	0.018 [-0.032;0.066]
$\sigma(\Delta d)$	0.110	0.112 [0.091;0.137]	0.155 [0.084;0.247]
$AC1(\Delta d)$	0.146	0.367 [0.149;0.570]	0.261 [0.034;0.466]
$E(r_m)$	0.075	0.067 [0.027;0.109]	0.065 [0.019;0.121]
$\sigma(r_m)$	0.200	0.167 [0.139;0.198]	0.205 [0.118;0.321]
$AC1(r_m)$	-0.003	0.017 [-0.203;0.236]	-0.002 [-0.227;0.230]
$E(r_f)$	0.005	0.026 [0.018;0.034]	0.001 [-0.004;0.018]
$\sigma(r_f)$	0.039	0.012 [0.009; 0.017]	0.011 [0.005;0.018]
$AC1(r_f)$	0.563	0.807 [0.671;0.900]	0.806 [0.660;0.908]
$E(p-d)$	3.320	3.007 [2.894;3.118]	3.05 [2.593;3.319]
$\sigma(p-d)$	0.441	0.188 [0.140;0.250]	0.204 [0.106;0.331]
$AC1(p-d)$	0.936	0.654 [0.441;0.811]	0.628 [0.314;0.850]

The table displays the results from 10.000 simulations of 948 months each (to mimic the period 1930-2009) of the long run risks model based on the following BY calibration: $\delta = 0.998, \gamma = 10, \psi = 1.5, \mu_c = 0.0015, \mu_d = 0.0015, \phi = 3, \varphi = 4.5, \rho = 0.979, \varphi_e = 0.044, \sigma = 0.0078, \nu = 0.987, \sigma_w = 0.0000023, \pi = 0$. The statistics are for annually aggregated variables (consumption growth rate, dividend growth rate, market return, risk free rate, log price-dividend ratio). Aggregation has been done through a 12 months summation for growth rates and returns. For the log price-dividend ratio, the annual values have been computed using the year's December values of the price divided by the summation of monthly dividends within the year. The last column reports the result obtained using Bansal et al. [2012a] calibration: $\delta = 0.9989, \gamma = 10, \psi = 1.5, \mu_c = 0.0015, \mu_d = 0.0015, \phi = 2.5, \varphi = 5.96, \rho = 0.975, \varphi_e = 0.038, \sigma = 0.0072, \nu = 0.999, \sigma_w = 0.0000028, \pi = 2.6$.

3.2 Stylized facts and Model predictions

Beeler and Campbell [2012] have extended the basic set of moments matching to verified that the long run risks model is coherent with some stylized facts that have been documented by earlier studies. Specifically, they tested 2 stylized facts related to the predictability of cash flows by stock prices:

- log price-dividend ratio predicts excess stock returns and not dividend growth or real interest rate (Campbell and Shiller [1988]; Fama and French [1988]; Hodrick [1992]);
- there is a strong relationship between stock prices and consumption volatility but no link between stock prices and return volatility [Beeler and Campbell, 2012];

As shown in Table 3.2, in the data the log price-dividend ratio predicts excess stock returns with a negative and significant sign at 5% confidence level; meaning that an increase in stock prices predicts a drop in future returns. This predictability increases with the horizon; at one year horizon, log price-dividend ratio explained only 6 % of future excess return. This explanation increases to 29 % at 5 year horizon. The predictability of future excess return by the log price-dividend ratio is due to the persistence in the later variable (Cochrane [2001], Chap. 20), but it is not captured by both the BK and the BKY calibrations of the long run risks model: at 5 year horizon, the log price-dividend ratio explains less than 5% on the excess return and the regression's slope even sometime has the wrong sign or is not significant.

For the consumption process, the picture is inverted, the model overstates the predictability of future consumption growth by stock prices : in the data, there is no evidence of predictability of future consumption growth by the log price-dividend ratio while in the BK and BYK calibrations, log price dividend ratio predicts future consumption growth with a significant slope coefficient. Even though the power of the predictability of the future consumption growth by the log price dividend ratio has been greatly reduced from BK to BYK calibration, the slope coefficients still remain significant at 5 % confidence level; what is at odd with the data.

As we know from (Cochrane [2001], sect.20.1.2), log price-dividend ratio only moves if it forecasts excess return or dividend growth or if there is a bubble. In the data, log price dividend ratio rather predicts excess return as we can see in Table 2. The BKY calibration does well in this respect compare to the BK calibration where the log price dividend ratio seems to predict future dividend growth.

Table 2: Predictability of excess returns, consumption, and dividends.

	$\hat{\beta}$	$\widehat{R^2}$	$\beta(50\%)$	$\beta(50\%)$	$R^2(50\%)$		$\%\left(\widehat{R^2}\right)$	
	data	data	BY	BKY	BY	BKY	BY	BKY
$\sum_{j=1}^{J} r^a_{m,t+j} - r^a_{f,t+j} = const. + \beta(p^a_t - d^a_t) + \varepsilon_{t+J}$								
1Y	-0.107 / -3,446	0,059	0.0614 / 0.712	0.079 / 0.925	0.006	0.010	0.944	0.908
3Y	-0.304 / -4.078	0.222	-0.171 / -1.082	-0.219 / -1.315	0.024	0.026	0.950	0.884
5Y	-0.440 / -4.420	0.293	-0.299 / -1.298	-0.447 / -1.365	0.021	0.040	0.960	0.862
$\sum_{j=1}^{J}(\triangle c^a_{t+j}) = const. + \beta(p^a_t - d^a_t) + \varepsilon_{t+J}$								
1Y	0.001 / 0.165	0.0003	0.067 / 4.845	0.056 / 5.694	0.225	0.150	0	0.009
3Y	-0.002 / -0.116	0.0002	0.175 / 5.361	0.108 / 1.713	0.267	0.114	0.001	0.027
5Y	-0.009 / -0.562	0.005	0.220 / 4.648	-0.089 / -1.269	0.219	0.089	0.010	0.086
$\sum_{j=1}^{J}(\triangle d^a_{t+j}) = const. + \beta(p^a_t - d^a_t) + \varepsilon_{t+J}$								
1Y	-0,003 / -0,133	0.000	0.262 / 3.217	0.317 / 4.898	0.118	0.198	0.002	0.001
3Y	0,024 / 0,436	0.003	0.663 / 3.955	0.351 / 2.301	0.165	0.085	0.015	0.068
5Y	0,043 / 0,612	0.007	0.673 / 3.622	0.737 / 1.581	0.145	0.062	0.047	0.157

There is a great improvement in the model implications from BK calibration to BKY own. However, the BKY calibration of the model failed to capture excess return predictability by the log price dividend ratio that is observed in the data. Each column represent the slope and the t-stat of the predictive regression of future excess return, consumption and dividend growth on the log price dividend ratio for different horizons (1 year, 3 years and 5 years). The R-square are also provided. The last column represents the percentage of sample simulated R-squared that are below the R-squared observed in the data; a percentile below 5% or above 95 % correspond to a one side test rejection of the model at 5% significant level.

Some of these stylized facts have been shown to be a failure of the BY and BYK calibrations to match the data. But, as the matching of the data has been greatly improved from the BY calibration to the BYK own, there is a hope that the failure highlighted by Beeler and Campbell [2012] and Garcia et al. [2008] could be just a failure of finding the good set of parameters for the model rather than a rejection of the economic mechanism presented in the model as the explanation of many asset pricing puzzles. So, the addition of the predictive moment constraints implied by these stylized facts in the GMM procedure should improved the model prediction through a better selection of the set of parameters. The derivation of the theoretical

annual moment involving the log price-dividend ratio is not an easy task because the annual values of the log price-dividend ratio are obtained by dividing the end of year (December) monthly prices by the sum of the monthly dividends within the year. So,

$$p_t^a - d_t^a = \log P_{ht} - \log \sum_{j=0}^{h-1} D_{ht-j} \tag{5}$$

Then using the following approximation presented in Bansal et al. [2007b]. [8]:

$$\log \sum_{j=0}^{h-1} D_{h(t+1)-j} \approx \log D_{ht} + \log h + \sum_{j=0}^{h-1} \frac{j+1}{h} \triangle d_{h(t+1)-j} \tag{6}$$

Equation (5) can be rewritten as follow[9]:

$$p_t^a - d_t^a = \mu_d^a + x_t^a + A_{2m}\sigma_{ht}^2 - \pi\eta_t^a - \varphi_d u_t^a \tag{7}$$

Where

$$\mu_d^a = A_{0m} - \log h - \sum_{j=0}^{h-1} \left(\frac{j+1}{h}\right) \mu_d, \quad x_t^a = A_{1m} x_{ht} - \sum_{j=0}^{h-1} \left(\frac{j+1}{h}\right) \phi x_{ht-j-1}$$

$$\eta_t^a = \sum_{j=0}^{h-1} \left(\frac{j+1}{h}\right) \sigma_{ht-j-1}\eta_{ht-j}, \quad u_t^a = \sum_{j=0}^{h-1} \left(\frac{j+1}{h}\right) \sigma_{ht-j-1} u_{ht-j}$$

Equation (5)) allows to write the annual values of the log price-dividend ratio in terms of monthly values of the log price-dividend ratio and monthly dividend growth rate. This formula will enable to express the annual log price-dividend ratio in term of monthly shocks and later to compute the theoretical moments mentioned in the stylized facts that will be added to the previous usual set of moments. Equation (7) shows that the annual value at year t of the log price-dividend ratio depend on the monthly values of the persistent component of the consumption growth process, on the December volatility, on the monthly short run consumption shocks and monthly dividend growth shocks. We can see that negative shocks will increased asset prices while past month positive shocks will lower the level log price-dividend ratio.

To compute the theoretical value of the slope in the predictive regressions, we need to express the annual value of the cash flows in term of monthly processes. Fu-

[8] This approximation is exact when the arithmetic mean and the geometric mean are equal; it is also assumed in Hansen et al. [1996]. It could also be obtained through a Taylor approximation around the mean of the one period (log) consumption growth and assuming that mean to be equal to zero, as shown in Ghattassi and Meddahi [2012] .

[9] This expression of the aggregate log price-dividend ratio is mainly used to compute the theoretical moments, but for simulations we rather used the following intermediary formula: $p_t^a - d_t^a = z_{m,ht} + \sum_{j=0}^{h-1} \triangle d_{ht-j} - \log h - \sum_{j=0}^{h-1} \left(\frac{j+1}{h}\right) \triangle d_{ht-j}$

ture annual excess returns could be expressed in term of present year last month states variables (x_{ht} and σ_{ht}) and weighted sums of within years future monthly consumption and dividend shocks ($\sigma_{h(t+j)-p-1}\eta_{h(t+j)-p}$, $\sigma_w w_{h(t+j)-p}$,$\sigma_{h(t+j)-p-1}e_{h(t+j)-p}$, $\sigma_{h(t+j)-p-1}u_{h(t+j)-p}$) as follow:

$$\sum_{j=1}^{J}\left(r^a_{m,t+j} - r^a_{f,t+j}\right) = \mu_r^{a,J} + \phi_x^{a,J} x_{h(t-1)} + \phi_\sigma^{a,J}\sigma^2_{h(t-1)} + e_t^{a,J} + w_t^{a,J} + \varphi_d u_t^{a,J} + \pi\eta_t^{a,J} \quad (8)$$

where $\mu_r^{a,J}$, $\phi_x^{a,J}$, $\phi_\sigma^{a,J}$, $e_t^{a,J}$,$w_t^{a,J}$,$u_t^{a,J}$,η_t^a, Λ_0, Λ_1,Λ_2, $\beta_{m,w}$ and $\beta_{m,e}$ are given in Appendix A.3.

We choose here to focus on the predictive regression of the future excess return on the log price-dividend ratio because this is the key mismatch that both the BK and BKY calibrations failed to capture. For the other predictive regressions (with future consumption or dividend growth as the dependent variables), the BKY calibration delivered a pattern that seems to be close to the data.

Concerning the second stylized fact, the derivation of analytical formulas is more tricky because, as in Beeler and Campbell [2012], we need to compute the realized volatility suggested by Bansal et al. [2012a]. For that, the first step is to run an AR(1) regression of each variable of interest y_t (consumption or dividend growth or excess return):

$$y_{t+1} = b_0 + b_1 y_t + u_{t+1} \quad (9)$$

Then we can compute the K-period realized volatility as the sum of the absolute values of the residuals over the K periods:

$$Vol_{t+1,t+K} = \sum_{k=1}^{K} \mid u_{t+k} \mid \quad (10)$$

and finally we regress the log of K period realized volatility on the present log price-dividend ratio:

$$log[Vol_{t+1,t+K}] = \alpha_c + \beta_c(p_t^a - d_t^a) + \xi_{t+K} \quad (11)$$

Keeping track of analytical formulas along this procedure becomes difficult; that is why we will use it (this procedure) as an out of sample test to see whether the new calibration derived from estimation performs better than the previous one as far as this predictability is concerned.

Table 3: Predictability of the volatility of excess returns , consumption, and dividends.

	$\hat{\beta}$	$\widehat{R^2}$	$\beta(50\%)$	$\beta(50\%)$	$R^2(50\%)$		$\%\left(\widehat{R^2}\right)$	
	data	data	BY	BKY	BY	BKY	BY	BKY
Excess return volatility predictability								
1Y	0.015	0.003	-0.040	-0.090	0.007	0.027	0.351	0.173
	0.553		-0.5732	-1.756				
3Y	0.008	0.0003	-0.149	-0.256	0.015	0.057	0.077	0.038
	0.106		-1.102	-1.480				
5Y	-0.012	0.0004	-0.225	-0.415	0.022	0.073	0.067	0.040
	-0.114		-0.812	-2.292				
Consumption growth volatility predictability								
1Y	-0.013	0.120	-0.006	-0.016	0.007	0.033	0.996	0.864
	-3.615		-1.019	-1.474				
3Y	-0.043	0.306	0.012	-0.034	0.016	0.072	0.9995	0.928
	-4.138		1.010	-1.996				
5Y	-0.072	0.397	0.0280	-0.049	0.021	0.091	0.999	0.9403
	-4.119		1.065	-1.899				
Dividend growth volatility predictability								
1Y	-0.025	0.020	0.022	-0.118	0.007	0.033	0.763	0.386
	-1.409		0.837	-1.541				
3Y	-0.069	0.031	-0.095	-0.316	0.0147	0.069	0.667	0.327
	-1.158		-1.060	-2.192				
5Y	-0.141	0.054	-0.103	-0.318	0.021	0.088	0.720	0.385
	-1.165		-1.035	-1.425				

There is a great improvement in the model implications from BK calibration to BKY concerning the consumption growth volatility predictability. However, the BKY calibration of the model overstated the excess return volatility predictability by the log price dividend ratio that is observed in the data. Each column represent the slope and the t-stat of the predictive regression of future excess return, consumption and dividend growth on the log price dividend ratio for different horizons (1 year, 3 years and 5 years). The R-square are also provided. The last column represents the percentage of sample simulated R-squared that are below the R-squared observed in the data; a percentile below 5% or above 95 % correspond to a one side test rejection of the model at 5% significant level.

4 Estimation procedure and Simulations

In the data, the latent variables (x_t and σ_t^2) are not observable but we observe the log price-dividend ratio (z_m), the real market return (r_m), the real risk free rate (r_f), the market dividend growth rate (g_d) and the consumption growth rate (g). We use the Euler equation and the affine link between the observed variables and the latent ones to set up some moment restrictions that will be used for the GMM procedure.

Our approach is similar to Constantinides and Ghosh [2011] who solved for the 12 parameters of the model by imposing 15 unconditional moment restrictions. But contrary to them we choose to remain closer to how the original LRR model works by expressing the theoretical moments taking temporal aggregation into account to match the annual data ($h = 12$). Indeed it is usually assume during calibrations that the long run risks model operates at the monthly frequency, but the monthly data are reputed to be of poor quality especially because of seasonality and other measurement problems Wilcox [1992]. We also estimate the model assuming that the agent decisions are made at the same frequency as the econometrician observes the data (meaning at the annual frequency, $h = 1$), but we reject the model in that case. This shows that the temporal aggregation also play an important role for the LRR model to match the data.

The long run risks model we consider here has 13 parameters[10] The moment restrictions are made of 16 basic moment restrictions whose empirical counterparts are summarized in Table 1. and are known to be matched pretty well by the previous calibrations of the model; except the variances of the risk free rate and of the log price dividend ratio. In a first step, we will only use those restrictions for the estimation. In a second step, we add 3 other moment restrictions to capture the predictability (at 1, 3 and 5 years horizon) of the cash flows (future excess returns) on asset prices (log price-dividend ratio) summarized in Table 2. and which have been presented as a failure of the previous calibrations of the model to match the data. Finally, in a third step we extended the previous 19 moment restrictions to 22 (resp. 25) moment restrictions by adding 3 other moment restrictions for the predictive regressions of consumption (resp. dividend) growth on log price-dividend ratio.

4.1 Weak identification evidence in the LRR model

In this section, we will show that some parameters in the long run risk model are weakly identified, mainly because of the sample size, and thus the standard inference procedure, which relies on the delta method to obtain the confidence intervals and in testing is not valid. The type of weak identification that we observe in the LRR features the special case described by [Andrews and Cheng 2012, 2014].

Following Newey and McFadden [1986] (Sec. 2.2.3), Stock and Wright [2000],

[10] In the model considered by Constantinides and Ghosh [2011], the parameter π which captures the exposure of dividend growth to short run consumption shocks is ignored. The time aggregation parameter h could also be estimated, that will extend the number of structural parameters to be estimated to 14.

we defined a parameter θ as being identified if $\theta = \theta_0$ is the unique solution of the moment condition $E(g(y_t, \theta)) = 0$[11] for $\theta \in \Theta$, where the expectation is taken with respect to the true distribution (for which $\theta = \theta_0$). Some parameters can be weakly identified while others are strongly identified; we can then split the parameter vector as follow $\theta = (\alpha, \beta)$, where α is the sub-vector of weakly identified parameters and β is the sub-vector of strongly identified parameters. This means that the moment conditions are zero at $\theta_0 = (\alpha_0, \beta_0)$, but are also very nearly zero for $\alpha \neq \alpha_0$: In other words, the population objective function is steep in β around β_0, but nearly flat in α (Stock and Wright [2000], Sect 2.3).

Andrews and Cheng [2012] presented another specific situation where some parameters of the model might be weakly identified when the true value of a given parameter is close to 0[12]. Then, they split the vector of parameters into three components : $\theta = (\beta, \xi, \pi)$ where β and ξ are strongly identified but π becomes weakly identified when $\beta_0 = 0$ and there is a drifting sequence of parameters, depending on the sample size ($\theta_n = (\beta_n, \xi_n, \pi_n)$ that converges to θ_0 such that $\beta_n \neq 0$ and $n^{1/2}\beta_n \to b \in R^{d_\beta}$(category I(b) in their paper). The situation we encounter in the LRR model really looks like that one but we should be cautious about this categorization. Indeed, if we assume that the true value of volatility of the volatility (σ_w) is equal to 0 then the persistence of the volatility (ν) will not be identified in population; and from the literature on the LRR model we know that the true value of the volatility of the volatility must be different from 0; if not the volatility of the consumption growth will be constant and thus the equity premium will not be time varying. But from the subsequent calibrations of the model[13], it seems that it is close to 0, which implies the weak identification of the persistence of the volatility (ν) in finite sample even if it could be identified in population.

By the same argument, we can deduce that the closeness to 0 of the true value on the vector of parameters (σ, σ_w) induces the weak identification of the vector of parameters $(\phi, \varphi_d, \varphi_e, \rho, \nu, \pi)$ and the remaining parameters $(\delta, \gamma, \psi, \mu_c, \mu_d)$ are strongly identified. Hence we have the decomposition of our vector of parameters which satisfies the Assumption A in Andrews and Cheng [2012], and we can rely on their robust procedure for estimation and inference. Furthermore, we observe that when the EIS become greater (let say more than 10), the objective function becomes

[11] The vector function $g(y_t, \theta)$ can be seen as the difference between the empirical moment from the data evaluated at the true value of the parameter θ_0 and the population moment evaluated at θ: $g(y_t, \theta) = \bar{m}(y_t, \theta_0) - m(Y_t, \theta)$, where y_t denotes the sample observation of the random variable Y_t

[12] The value 0 has been used without lost of generality and can be achieved by using an appropriate re-parametrization

[13] It was 2.3 e-06 in the BY calibration and 2.8 e-06 in the BYK calibration

almost insensitive with respect to that parameter changes[14] and thus the jacobian matrix is of reduced rank. So, this parameter might also be weakly identified in a part of the parameters space.

4.2 Simulations: finite sample approximation of the asymptotic distribution

As we know that commonly used testing procedures, in particular those robust to weak identification (on which we will come back in later), are derived asymptotically (Dufour, 2009), to evaluate the accuracy of our unconditional moment matching procedure in finite sample, we run a Monte Carlo experiment, where we simulate the model at various sample sizes and plot kernel density of the Gmm objective function. On the same plot (7) , we also represent the corresponding chi-squared limiting distribution. When the sample size is small (e.g: T=84) as what we have in the data, the finite sample approximation of the asymptotic distribution of the Gmm objective function is very poor; the right tail is very huge and that will lead to an over-rejection of the model. It means that while using the asymptotic distribution to evaluate the model in this case, we will end up wrongly rejecting the model more often that usual. So the rejection probability of a 95 % confidence test will be much more greater than the size of the test and the true size of the obtained confidence set will be much lower than 95 %. However, for a large sample (e.g: T=4200) , we see that the asymptotic test becomes conservative because the asymptotic distribution of the Gmm objective distribution is dominated by Chi-squared limiting distribution; meaning that rejection rate for a true model is lower than the significant level of the test. this conservativeness is the usual drawback of this against-weak identification robust methods[15].

5 Data and results

5.1 Estimation results

We use the Continuously Update GMM Estimator (CUE) for the estimation because it is partially robust to weak identification Stock et al. [2002]. We ran the estimations for four sets of moment conditions: (i) The first one is made of 16 moment restrictions

[14] More precisely, the moment conditions depend on the EIS only through the coefficient of the affine functions and it is those coefficients which are almost insensitive to the changes of ψ as it becomes high

[15] See Stock et al. [2002]

mainly targeting the basic moments summarized in Table 1 and does not include the predictive moments; we call it the restricted Gmm (RCUGmm). (ii) The second one, named CUGmm, is made of the moment conditions considered in the restricted Gmm and the excess returns predictive moments (covariance between excess returns at different horizons and the log price dividend ratio). (iii) The third is made of all the previous moment conditions and the consumption growth predictive moments (covariance between consumption growth at different horizons and the log price dividend ratio); we call it the Extended Gmm (ECUGmm). (iv) The last one is made of all the moment restrictions in (iii) and 3 predictive moments of the dividend growth by the log-price dividend ratio, we call it EECUGmm .

In a first place, we follow the usual procedure in a Gmm estimation, meaning that we minimize the objective function and then for the achieve minimum we test the model by comparing that minimum to a 95 % quantile of a chi-square distribution with the number of moment equations minus the number of estimated parameters as degree of freedom. In a second place, we follow the procedure describe in Stock and Wright [2000] when there is weak identification, which consist at fixing the weak identified parameters and estimating the remaining ones using continuously updated Gmm estimator with the optimal weighting matrix. The objective function value could then be compared to a chi-squared distribution and the model will not be rejected if we are able to find some fixed values of the weakly identified parameters for which the objective function at the minimum will be lower to a threshold given by the chi-square quantile at 95% level and the degree of freedom being equal to the number of moment restrictions minus the number of parameters estimated[16]. This method also allows to construct the confidence interval of the fixed parameter, by collecting all the values for which the model is not rejected. We repeated the procedure by fixing one parameter each time.

Table 4 shows that the model is not rejected at 5 % confidence level for two of the four set of moment conditions that we considered. In fact, the model is not rejected when we consider the basic moments made of the first and second order moments of the cask flows and stock prices variables, and the covariance between excess return and the log price dividend ratio. But it is rejected when we add the other predictive restrictions. So the LRR model is able to match the first and second order moment of asset prices and cash flows data, hence to solve the equity premium and risk free rate puzzles. It can also reproduce the predictability of excess return by the price-dividend ratio. But, it fails at matching simultaneously the observed level predictability of cash flows (consumption and dividend growth) by the price dividend

[16]See appendix B.

ratio. In fact, there is a huge tension for the model to capture the excess return predictability without overstating the cash flows (consumption growth and dividend growth) predictabilities compare to what is observed o real data. We tried man different starting values and still in the last two cases (ECUGmm and EECUGmm) we where not able to achieve a set of parameters for which the model could not be rejected at the standard confidence level.

Some parameters (δ, μ_c, μ_d,ρ,ν,σ) are estimated very precisely and their values are stable whatever the moments restriction we considered. The value of the pure discount rate is very closed to 1, the monthly consumption growth is around 0.19 %. For the RCUGmm, where there is not predictability moment in the restrictions, the model is not rejected at the 5 % critical error probability. The estimates of γ and ψ are very large but also not significant. this is also the case for many parameters describing the dividend growth process (ϕ,φ_d,π). When the model is not rejected, the estimated level of risk aversion coefficient (2.7) is quite below the one that is usually used for calibration (10). The estimate of the EIS (1.27) is above 1 as usual assumed in the model, even though the 95 % confidence interval contains values below 1.

We find some difficulties while computing the confidence intervals for the minimands we obtained. Indeed, applying the formula in equation 17, the matrix we computed at the minimum was not invertible ($G(\hat{\theta})'\Omega(\hat{\theta})^{-1}G(\hat{\theta})$). The singularity of this matrix comes from the fact that the estimate of the jacobian matrix is not of full rank. We think that this lack of full rankness does not come from the non identification of the model, if not this would have happened for every vector of parameters. But instead, it comes from the fixed point feasibility condition we highlighted in section 2.2. Indeed, the numerical computation of the Jacobian requires two points where the objective function will be evaluated and when the minimand is close to the feasibility bound this can be difficult do[17].

Figures 8 and 9 show respectively the confidence intervals of γ and ψ obtained with the CUGmm restriction setup, following the weak identification robust procedure described in D.2. We see that those confidence intervals are different from the ones obtained in 4. The confidence intervals for γ and ψ are included in $[1; 20]$. No value below 1 is admitted in the 95% confidence interval of ψ; this is in adequacy with the hypothesis made in the LRR model that the value of the EIS is above 1. However, this confidence sets can still be extended because for computational reason[18], we took values of γ and ψ ranging between 0.1 and 20, but we can see

[17] We still need to work on this point.

[18] For each value of the fixed parameter, we run several optimizations restarting from the previous obtained minimand until convergence. This can take more than several minutes for each value of

from figures 8 and 10, that some values of ψ even above 50 can still be admitted.

5.2 Predictability

For reminding, one of the goals we had at the beginning was to find through an efficient procedure, a set of parameters that could enable (if possible) to restore the predictability failure highlighted by Beeler and Campbell [2012] in the LRR model. So, we now examine the model implied moments and the predictability of future cash flows by the log price dividend ratio when the model is calibrated at the value we find in the previous estimations.

Table 6 shows that concerning the basic moments, all the four restrictions estimates did equivalently good job at generating data whose first and second order moments meet the observed ones. The mean of the consumption growth is well around 2% with a volatility around 3%. The market return is at 6% with a risk free rate below 1 %.

Table 6 also highlights the tension that the model faces while attempting to match all the data characteristics. Indeed, we see that the estimate of the first moment restrictions setup (RCUGmm) does almost the same job as the BYK calibration, mainly on one hand it fails at capturing the the volatility of the risk free rate and its auto-correlation. On the other hand, the volatility of the market return is a bit overstate, but the volatility of the price-dividend ratio and its auto-correlation are captured. When we move to the second moment restrictions setup (CUGmm), we observed that the volatility of the risk free rate and the volatility of the market return have been improved, but then the risk free rate auto-correlation is still not captured and the dividend volatility is a bit overstated. For the last two moment restrictions set up (ECUGmm and EECUGmm), we see that the volatility of the risk free rate has been improved again and its auto-correlation is now captured. But the consumption and dividend growth auto-correlations are overstated and the price dividend auto-correlation is understated.

The results concerning the predictability implied by the model are even more interesting. Indeed as Table 5 shows, for the first restriction set up (RCUGmm), there is a slight predictability observed for excess return and almost no predictability for the consumption growth; both can hardly but still be rejected as being equal to the ones observed in the data at the 95 % confidence level. For dividend growth, there is some predictability in the simulated data, which is strongly rejected as being equal to the one observed in the data. While in the second restriction setup (CUGmm),

the fixed parameter.

Table 4: Results of the Gmm estimations

	RCUGmm	CUGmm	ECUGmm	EECUGmm
δ	0.99738 [0.915 ;1.0798]	9.999999e-01 [0.994 ; 1.006]	0.999815 [0.9998146 ; 0.9998164]	9.997412e-01 [0.9994 ;1.0001]
γ	21.659 [-132.336 ;175.654]	2.685 [-7.346e-02 ; 5.443]	5.374 [-10.554 ; 21.302]	5.128 [4.671 ; 5.586]
ψ	6.812 [-632.82 ;646.44]	1.250 [-0.922 ; 3.423]	1.475 [3.938e-01 ; 2.557]	1.4816 [1.480; 1.484]
μ_c	1.93e-03 [1.567e-03 ;2.292e-03]	2.082e-03 [1.788e-03; 2.375e-03]	1.905e-03 [1.903e-03 ; 1.907e-03]	1.900e-03 [1.657e-03 ;2.142e-03]
μ_d	2.066e-03 [4.298e-04 ;3.703e-03]	2.398e-03 [1.129e-03 ;3.667e-03]	2.824e-03 [1.563e-03 ;4.085e-03]	2.437e-03 [1.770e-03 ;3.104e-03]
ϕ	20.631 [-121.87 ;163.132]	2.648 [0.751 ;4.545]	3.880 [-0.880 ;8.640]	3.439 [3.020 ;3.859]
φ_d	2.379 [-1.232 ;5.990]	-4.053 [-6.388 ;-1.718]	17.699 [-72.341 ;107.740]	17.471 [17.011 ;17.930]
ρ	0.9944 [0.988 ; 1.001]	0.9965 [0.988 ; 1.005]	0.957 [0.879 ; 1.034]	9.475e-01 [9.005e-01 ; 9.945e-01]
φ_e	2.342e-03 [-1.141e-02 ; 1.609e-02]	1.664e-02 [-9.714e-03 ; 4.300e-02]	0.355 [-1.642 ; 2.353]	6.368e-01 [2.942e-01 ; 9.794e-01]
σ	7.865e-03 [4.655e-03 ; 1.108e-02]	4.805e-03 [2.504e-03 ; 7.106e-03]	8.136e-04 [0 ;4.277e-03]	5.333e-04 [1.461e-04 ; 9.205e-04]
ν	9.208e-01 [8.867e-01 ; 9.548e-01]	9.900e-01 [0.975 ;1.005]	0.997 [0.995 ;0.998]	9.972e-01 [9.965e-01 ; 9.980e-01]
σ_w	1.071e-04 [0 ;1.426e-03]	2.922e-05 [0 ;1.494 e-04]	5.4872e-07 [2.6126e-07 ; 8.3617e-07]	2.005e-07 [0 ; 2.336e-07]
π	3.956e-01 [-1.711 ;2.502]	0.8172 [-2.872 ;4.506]	3.422 [-209.713 ;216.557]	1.823 [1.363 ;2.283]
$TJ_T(\widehat{\theta_p^u})$	6.715 0.0816	10.862 0.0927	23.252 0.00565	38.626 0.00012

This table presents the parameter's estimates and their 95 % level confidence intervals derived from the minimization of the GMM criterion function. The last row gives the value of the minimum achieved and the corresponding p-value. 13 parameters are estimated. For the RCUGmm we have 16 moment restrictions; for the CUGmm, 19 moment restrictions; for the ECUGmm, 22 moment restrictions; and for the EECUGmm, 25 moment restrictions. So the p-values are computed respectively with respect to a $\chi^2(3)$, $\chi^2(6)$, $\chi^2(9)$ and $\chi^2(12)$.

where the predictability of the excess return by the log price dividend ratio has been targeted, there is indeed a predictability pattern emerging, which mimics the one observed in the data. However, we see also that the consumption growth and the dividend growth both became highly predictable, which is not the case in the data. For the predictability of the excess return by the log price-dividend ratio, the equality of the model's R-squared to the sample counterpart is not empirically rejected at the 5 % confidence level. When the model is calibrated at the CUGmm estimates, at 1 year horizon, the log price-dividend ratio explains 2 % of the future excess return variability and this explanation increases to 7 % at the 5 years horizon. Furthermore, as in the observed data, the slope in the predictive regression has a negative sign and its significance increases with the horizon of prediction. Similarly, in the third restriction setup (ECUGmm) where the predictability of the excess returns and the consumption growth has been targeted, we see that the predictability of the excess return is almost similar to the one obtained in the previous restriction setup and the predictability of the consumption growth is more closer two the one observed in the data (it is declining with the horizon). There is also an improvement for the predictability of the dividend growth. In the last restriction setting (EECUGmm), The predictability of excess return by the log price dividend ratio has continue to improve, moving from 3 % part in the explanation of the future excess return by the log price dividend ratio at one year horizon to 11 % of that explanation at 5 years horizon. the predictabilities of the consumption growth and the dividend growth by the log price dividend ratio are both declining with the horizon.

We also look at the volatility predictability to see whether the model could reproduce the predictability of the volatility of the consumption growth by the log price dividend ratio observed in the data. As we can see from Table D.2, there is an improvement when we move from the basic moment set up to the extended Gmm estimation set up. The median r-squared has increased compare to the BYK calibration. However, there is an over-state of the predictability of excess return volatility by the log price-dividend ratio.

These results say that the long run risk model is very good at reproducing many features of the observed data when they are targeted. But as we saw this could come at the cost of introducing more bias in mimicking the non targeted ones, and thus increasing the targeted moment does not seem to be the solution. This cast some doubt about the reliability of this model to be used for forecasting and policy recommendations.

To resume, the results obtained in the empirical part still need to be examined and could be improved by the choice of a better starting value in the optimization

program. Indeed, because of the complexity of the objective GMM criterion, the presence of non-feasible regions of parameters and also due to the identification problem highlighted, the finally selected set of parameters is obviously a local minimum and depends on the starting point. The choice of another starting point could yield to a better parameter vector, for which the moments restriction will be validated and the moments predictability improved. Another possible source of bias, could be the approximations on which we relied. Indeed, there are two main approximations that we made in order to solve the model and later to make a bridge between the monthly simulated data in the model and the annualized observed data. The first approximation is the Campbell-Shiller approximation of log return that allows to expressed asset log returns in terms of log price dividend ratio and therefore as linear function of states variables. By comparing the model implications using this approximation and the computational power, Bansal et al. [2007b] argued that this approximation has little impact on the results. The second approximation of the arithmetic mean by the geometric mean in order to derived analytical expressions of annualized variables in terms of monthly counterparts. Since the one-month growth rate of cash flows or asset prices are usually small in magnitude, this approximation based on a Taylor expansion could be quite reliable to compare first order moments but could break down when the higher order terms in the Taylor expansion are no more negligible (Campbell et al., 1996-sect.1.4).

6 Conclusion

The Long Run Risks model explains the long run behavior of asset prices by the presence of a persistence movement in consumption growth, a time varying volatility, and a consumer preference structure described by the Epstein-Zin recursive utility function with an EIS greater than 1. The calibration of this model has remarkably well performed as far as the matching of the macro-finance first and second order moments is concerned. However, this model has recently received some criticism concerning the ability of the model to match the predictability of cash flows by asset prices observed in the data. More specifically it has failed to match the predictability that price-dividend has on the future returns. The correction of this failure and the rehabilitation of the model require to go beyond the calibration and to provide an estimation procedure that will find the parameter's set which brings the model closer to the data and restore the predictability. The difficulties being that on one hand, the return on the aggregate consumption asset embedded in the Euler equation of a representative agent with an Epstein-Zin recursive utility function is not observable;

on the other hand the states variables driving the risk of consumption growth are not observable and even difficult to estimated in a reduced form model because of bias induced by the time aggregation. However, by assuming an affine link between stock prices and the states variables, and using the Euler equation, we can express the cash flows and the stock prices moments in terms of the model underlying parameters. We apply this procedure to derive the analytical expression of a set of moment conditions and apply the GMM method to estimate the structural model's parameters.

Empirically, we find that using the standard asymptotic theory, the model succeeded to pass the test of validation of the moment conditions when we target only the basic moments or extended the basic moments to capture predictability of future excess returns by the price dividend ratio, but it became rejected when the other predictability moments are added. We also saw that the LRR model faces a huge tension while trying to match all the stylized facts highlighted in the literature. Furthermore, even considering the best fitting of the model achieved while taking into account the predictability concerns, it is still very difficult for the model to match the quantitative levels observed in the data, even though the qualitative behavior of the model is very similar in the long run to what appears in the data. However, the parameters estimates achieved using extended moment conditions allow to improve the model predictions and to bring them closer to the observed data.

A Model solution

A.1 Case of $EIS \neq 1$

- For the wealth portfolio,

$$A_1 = \frac{1 - \frac{1}{\psi}}{1 - k_1\rho},\ A_2 = \frac{0.5[(-\frac{\theta}{\psi} + \theta)^2 + (\theta k_1 A_1 \varphi_e)^2]}{\theta(1 - k_1\nu)}$$

$$A_0 = \frac{\log\delta + (1 - \frac{1}{\psi})\mu_c + k_0 + k_1 A_2\sigma^2(1-\nu) + 0.5\theta k_1^2 A_2^2\sigma_w^2}{1 - k_1}$$

For the price-consumption ratio to produce a positive response to an increase in the long run consumption growth, A_1needs to be positive; that means the EIS (ψ) should be greater than one. By the same way, for the price-consumption ratio to decrease with an increase in the consumption volatility, we need that θshould be negative; which correspond to an EIS greater than 1 given that the risk aversion is greater than 1.

- For the market portfolio,

$$A_{1,m} = \frac{\phi - \gamma + (\theta - 1)(k_1\rho - 1)A_1}{1 - k_{1,m}\rho}$$

$$A_{2,m} = \frac{(\theta-1)(k_1\nu - 1)A_2 + 0.5[(\pi - \gamma)^2 + \varphi_d^2 + ((\theta-1)k_1A_1 + k_{1,m}A_{1,m})^2\varphi_e^2]}{1 - k_{1,m}\nu}$$

$$A_{0,m} = \frac{\theta\log\delta - \gamma\mu_c + (\theta-1)[k_0 + (k_1 - 1)A_0 + k_1A_2\sigma^2(1-\nu)]}{1 - k_{1,m}}$$

$$\frac{+k_{0,m} + \mu_d + k_{1,m}A_{2,m}\sigma^2(1-\nu) + 0.5[(\theta-1)k_1A_2 + k_{1,m}A_{2,m}]^2\sigma_w^2}{1 - k_{1,m}}$$

- For the risk free asset,

$$A_{1,f} = \gamma + (\theta - 1)(1 - k_1\rho)A_1$$

$$A_{2,f} = -[(\theta-1)(k_1\nu - 1)A_2 + 0.5\left(\gamma^2 + (\theta-1)^2k_1^2A_1^2\varphi_e^2\right)]$$

$$A_{0,f} = -\theta\log\delta + \gamma\mu_c - (\theta-1)[k_0 + (k_1-1)A_0 + k_1A_2(1-\nu)\sigma^2] - 0.5(\theta-1)^2k_1^2A_2^2\sigma_w^2$$

Because the price consumption ratio is not observable, the values of the affine function coefficients are obtained numerical by solving a fixed point problem to get the value of $\bar{z}$ and then compute the k's values and later solve for the coefficient values.

A.2 Case of $EIS = 1$

The value function is given by:

$$V_t = C_t^{1-\delta} \left(E_t \left(V_{t+1}^{1-\gamma} \right) \right)^{\frac{\delta}{1-\gamma}}$$

which implies the following log price-consumption ratio:

$$vc_t = log(V_t/C_t) = \frac{\delta}{1-\gamma} log \left(E_t \left[exp \left((1-\gamma) \left(vc_{t+1} + \Delta c_{t+1} \right) \right) \right] \right)$$

We conjecture that the log price-consumption ratio can be expressed as an affine function of the state variable, meaning that : $vc_t = \tilde{A}_0 + \tilde{A}_1 x_t + \tilde{A}_2 \sigma_t^2$. Given the conditional normality of all the process involved, we have that:

$$\begin{aligned} vc_t &= \frac{\delta}{1-\gamma} \left\{ E_t \left[\left((1-\gamma) \left(vc_{t+1} + \Delta c_{t+1} \right) \right] + 0.5(1-\gamma)^2 Var_t \left(vc_{t+1} + \Delta c_{t+1} \right) \right\} \right. \\ &= \delta \left(\tilde{A}_0 + \tilde{A}_2 (1-\nu) \bar{\sigma}^2 + \mu_c + 0.5(1-\gamma)(\tilde{A}_2 \sigma_w)^2 \right) + \delta \left(1 + \tilde{A}_1 \rho \right) x_t \\ &+ \delta \left(\nu \tilde{A}_2 + 0.5(1-\gamma)(\varphi_e \tilde{A}_1)^2 + 0.5(1-\gamma) \right) \sigma_t^2 \end{aligned}$$

By identification, we get that:

$$\begin{aligned} \tilde{A}_0 &= \delta \left(\tilde{A}_0 + \tilde{A}_2 (1-\nu) \bar{\sigma}^2 + \mu_c + 0.5(1-\gamma)(\tilde{A}_2 \sigma_w)^2 \right) \\ \tilde{A}_1 &= \delta \left(1 + \tilde{A}_1 \rho \right) \\ \tilde{A}_2 &= \delta \left(\nu \tilde{A}_2 + 0.5(1-\gamma)(\varphi_e \tilde{A}_1)^2 + 0.5(1-\gamma) \right) \end{aligned}$$

Then,

- For the wealth portfolio,

$$\tilde{A}_1 = \frac{\delta}{1-\rho\delta}, \; \tilde{A}_2 = \frac{0.5\delta (1-\gamma) \left(1 + (\tilde{A}_1 \varphi_e)^2 \right)}{1-\nu\delta}$$

$$\tilde{A}_0 = \frac{\delta}{1-\delta} \left[\mu_c + \tilde{A}_2 (1-\nu) \bar{\sigma}^2 + 0.5(1-\gamma) \tilde{A}_2^2 \sigma_w^2 \right]$$

- For the market portfolio (we still use the Campbell and Shiller [1988] approximation for the log return on the market portfolio)

$$\tilde{A}_{1,m} = \frac{\tilde{\Gamma}_1 + \phi}{1 - k_{1,m}\rho}$$

$$\tilde{A}_{2,m} = \frac{\tilde{\Gamma}_2 + 0.5[(\pi-\gamma)^2 + \varphi_d^2 + [k_{1,m}\tilde{A}_{1,m} + (1-\gamma)\tilde{A}_1]^2\varphi_e^2]}{1 - k_{1,m}\nu}$$

$$\tilde{A}_{0,m} = \frac{\tilde{\Gamma}_0 + k_{0m} + \mu_d + k_{1m}\tilde{A}_{2m}(1-\nu)\bar{\sigma}^2 + 0.5[k_{1m}\tilde{A}_{2m} + (1-\gamma)\tilde{A}_2]^2\sigma_w^2}{1 - k_{1,m}}$$

$\tilde{\Gamma}_1 = -\gamma + (1-\gamma)\tilde{A}_1\frac{(\delta\rho-1)}{\delta}$, $\tilde{\Gamma}_2 = (1-\gamma)\tilde{A}_2\frac{(\delta\nu-1)}{\delta}$,$\tilde{\Gamma}_0 = \log\delta + (1-\gamma)\tilde{A}_0\frac{(\delta-1)}{\delta} - \gamma\mu_c + (1-\gamma)\tilde{A}_2(1-\nu)\bar{\sigma}^2$,

- For the risk free asset,

$$\tilde{A}_{1,f} = \gamma + (1-\gamma)\tilde{A}_1\frac{(1-\delta\rho)}{\delta}$$

$$\tilde{A}_{2,f} = (1-\gamma)\tilde{A}_2\frac{(1-\delta\nu)}{\delta} - 0.5\left(\gamma^2 + [(1-\gamma)\tilde{A}_1\varphi_e]^2\right)]$$

$$\tilde{A}_{0,f} = -\log\delta - (1-\gamma)\tilde{A}_0\frac{(\delta-1)}{\delta} + \gamma\mu_c - (1-\gamma)\tilde{A}_2(1-\nu)\bar{\sigma}^2 - 0.5(\gamma-1)^2\tilde{A}_2^2\sigma_w^2$$

B Fixed point feasibility condition for the log price-dividend ratio

The fixed point condition is determined by the equilibrium value of the log price dividend ratio, which should satisfy: $\bar{z}_m = A_{0m} + A_{2m}\bar{\sigma}^2$

Substituting the affine coefficient A_{0m} and A_{2m} by their following expressions:

$$A_{0m} = \frac{R_1 + k_{0,m} + k_{1,m}A_{2,m}\sigma^2(1-\nu) + 0.5[R_2 + k_{1,m}A_{2,m}]^2\sigma_w^2}{1 - k_{1,m}}$$

Where $R_1 = \theta\log\delta - \gamma\mu_c + (\theta-1)[k_0 + (k_1-1)A_0 + k_1A_2\sigma^2(1-\nu)] + \mu_d$ and $R_2 = (\theta-1)k_1A_2$

$$A_{2m} = \left[\Delta_{123} + \frac{e^{2\bar{z}_m}}{(1+e^{\bar{z}_m}(1-\rho))^2}\left(\phi - \frac{1}{\psi}\right)^2\varphi_e^2 + \frac{2e^{\bar{z}_m}}{(1+e^{\bar{z}_m}(1-\rho))}\Delta_3\left(\phi - \frac{1}{\psi}\right)\varphi_e^2\right]\frac{(1+e^{\bar{z}_m})}{(1+e^{\bar{z}_m}(1-\nu))}$$

Where $\Delta_1 = \frac{-(\theta-1)\frac{1}{2}\left[(-\frac{\theta}{\psi}+\theta)^2 + (\theta k_1A_1\varphi_e)^2\right]}{\theta}$, $\Delta_2 = (\pi-\gamma)^2 + \varphi_d^2$, $\Delta_3 = (\theta-1)k_1A_1$ and $\Delta_{123} = \Delta_1 + 0.5\Delta_2 + \varphi_e^2\Delta_3^2$

We get that the fixed point condition is equivalent to :

$$f(\bar{z}_m) = g(\bar{z}_m)$$

where

$$f(\bar{z}_m) = \bar{z}_m - log(1+e^{\bar{z}_m}) - \left[\Delta_{123} + \frac{e^{2\bar{z}_m}}{(1+e^{\bar{z}_m}(1-\rho))^2}\left(\phi - \frac{1}{\psi}\right)^2\varphi_e^2 + \frac{2e^{\bar{z}_m}}{(1+e^{\bar{z}_m}(1-\rho))}\Delta_3\left(\phi - \frac{1}{\psi}\right)\varphi_e^2\right]\bar{\sigma}^2$$

$$g(\bar{z}_m) = R_1 + 0.5\sigma_w^2 \left[R_0 + \frac{e^{\bar{z}_m}}{(1+e^{\bar{z}_m}(1-\nu))} \left(\Delta_{123} + \frac{e^{2\bar{z}_m}}{(1+e^{\bar{z}_m}(1-\rho))^2} \left(\phi - \frac{1}{\psi}\right)^2 \varphi_e^2 + \frac{2e^{\bar{z}_m}}{(1+e^{\bar{z}_m}(1-\rho))} \Delta_3 \left(\phi - \frac{1}{\psi}\right) \varphi_e^2 \right) \right]^2$$

$$\lim_{\bar{z}_m \to +\infty} f(\bar{z}_m) = -\left[\Delta_{123} + \frac{\left(\phi - \frac{1}{\psi}\right)^2}{(1-\rho)^2} \varphi_e^2 + \frac{2\Delta_3 \left(\phi - \frac{1}{\psi}\right) \varphi_e^2}{(1-\rho)} \right] \bar{\sigma}^2 = B^I$$

$$f(0) = -\left[\Delta_{123} + \frac{\left(\phi - \frac{1}{\psi}\right)^2}{(2-\rho)^2} \varphi_e^2 + \frac{2\Delta_3 \left(\phi - \frac{1}{\psi}\right) \varphi_e^2}{(2-\rho)} \right] \bar{\sigma}^2 = B^S$$

$$\lim_{\bar{z}_m \to +\infty} g(\bar{z}_m) = R_1 + 0.5\sigma_w^2 \left(R_2 - \frac{B^S}{(1-\nu)\bar{\sigma}^2} \right)^2 = G^S$$

$$g(0) = R_1 + 0.5\sigma_w^2 \left(R_2 - \frac{B^S}{(2-\nu)\bar{\sigma}^2} \right)^2 = G^I$$

When ($\gamma > 1$ and $\psi < 1$) or ($\gamma < 1$ and $\psi > 1$) (meaning that $\theta < 0$), f is decreasing from $[0, +\infty[\to]B^I, B^S]$ and g is increasing from $[0, +\infty[\to [G^I, G^S[$

C Temporal aggregation

- **Annual consumption growth**

As shown in Bansal et al. [2007b], the h-period aggregated consumption growth rate can be expressed as a weighted average of monthly consumption growth, with the weight taking a $\Lambda - shape$:

$$\Delta c_{t+1}^a = log \frac{\sum_{j=0}^{h-1} C_{h(t+1)-j}}{\sum_{j=0}^{h-1} C_{ht-j}} \approx \sum_{j=0}^{2h-2} \tau_j \Delta c_{h(t+1)-j}$$

where the index t is used to count the aggregated time and $h(t-1)+1$ to ht are the corresponding month within the aggregate period t

$$\tau_j = \frac{j+1}{h} \quad if \quad j < h \qquad and \qquad \tau_j = \frac{2h-j-1}{h} \quad if \quad j \geqslant h$$

It follows that :

$$\Delta c_{t+1}^a = \mu_c^a + \phi^a x_{h(t+1)-2h-1} + \eta_{t+1}^a$$

with

$$\mu_c^a = \sum_{j=0}^{2h-2} \tau_j \mu_c, \quad , \phi^a = \sum_{j=0}^{2h-2} \tau_j \rho^j$$

and

$$\eta^a_{t+1} = \sum_{j=0}^{2h-2} \tau_j \sigma_{h(t+1)-j-1} \eta_{h(t+1)-j} + \sum_{j=0}^{2h-3} \tau_j \left[\sum_{k=0}^{2(h-1)-j-1} \rho^k \varphi_e \sigma_{h(t+1)-j-k-2} e_{h(t+1)-j-k-1} \right]$$

As,

$$\sum_{j=0}^{n} \tau_j \rho^{n-j} = \begin{cases} \frac{\rho^{(n+2)} - (n+2)\rho + n + 1}{h(1-\rho)^2} & n < h \\ \\ \frac{\rho^{(n+2)} - 2\rho^{(n-h+2)} + (2(h-1)-n)(1-\rho)+1}{h(1-\rho)^2} & n \geq h \end{cases}$$

Δc^a_{t+1}can also be written the following way which enables a compact writing of the analytical formulas:

$$\Delta c^a_{t+1} = h\mu_c + \rho b_{h-2} x_{h(t-1)} + \sum_{j=0}^{h-1} a_j \varphi_e v_{h(t+1)-j-1} + \sum_{j=0}^{h-2} b_j \varphi_e v_{ht-j-1}$$

$$+ \sum_{j=0}^{2(h-1)} \tau_j \sigma_{h(t+1)-j-1} \eta_{h(t+1)-j}$$

with

$$\begin{cases} a_j = \sum_{k=0}^{j} \left(\frac{k+1}{h} \right) \rho^{j-k} = \frac{\rho^{j+2} - (j+2)\rho + j + 1}{h(1-\rho)^2} & if \quad j < h \\ \\ b_j = \sum_{k=0}^{j+h} \tau_k \rho^{j+h-k} = \frac{\rho^{j+h+2} - 2\rho^{j+2} + (h-2-j)(1-\rho)+1}{h(1-\rho)^2} & if \quad j \leq h-2 \end{cases}$$

and $v_{h(t+1)-j-1} = \sigma_{h(t+1)-j-2} e_{h(t+1)-j-1}$.

Notice that $\forall h \geq 1$, $b_{h-2} = \frac{\left(1-\rho^h\right)^2}{h(1-\rho)^2}$, but in the summations from 0 to $h-2$, $b_j = 0$ if $j < 0$ (for example when $h = 1$). So out of the summation, $b_{h-2} = \frac{\left(1-\rho^h\right)^2}{h(1-\rho)^2}$ to avoid writing this ugly fraction every time but in the summations from 0 to $h-2$ $b_j = 0$, if $h = 1$

Therefore,

$$E(\Delta c^a_{t+1}) = E(\sum_{j=0}^{2(h-1)} \tau_j \mu_c) = h\mu_c$$

$$var(\Delta c^a_{t+1}) = var(\sum_{j=0}^{2(h-1)} \tau_j \Delta c_{h(t+1)-j})$$

$$= var\left(\sum_{j=0}^{2(h-1)} \tau_j x_{h(t+1)-j-1} + \sum_{j=0}^{2(h-1)} \tau_j \sigma_{h(t+1)-j-1}\eta_{h(t+1)-j}\right)$$

$$= \sum_{j=0}^{2(h-1)} \tau_j^2 var(x_{ht}) + 2\sum_{j=0}^{2(h-1)-1}\sum_{k=j+1}^{2(h-1)} \tau_j\tau_k cov(x_{h(t+1)-J-1}, x_{h(t+1)-k-1}) + \sum_{j=0}^{2(h-1)} \tau_j^2\bar{\sigma}^2$$

$$var(\Delta c_{t+1}^a) = \left[\sum_{j=0}^{2(h-1)} \tau_j^2 + 2\sum_{j=0}^{2h-3}\sum_{k=j+1}^{2(h-1)} \tau_j\tau_k\rho^{k-j}\right]\left(\frac{\varphi_e^2\bar{\sigma}^2}{1-\rho^2}\right) + \left(\sum_{j=0}^{2h-2}\tau_j^2\right)\bar{\sigma}^2$$

$$var(\Delta c_{t+1}^a) = (\rho b_{h-2})^2\left(\frac{\varphi_e^2\bar{\sigma}^2}{1-\rho^2}\right) + \left(\sum_{j=0}^{h-1} a_j^2 + \sum_{j=0}^{h-2} b_j^2\right)\varphi_e^2\bar{\sigma}^2 + \left(\sum_{j=0}^{2(h-1)}\tau_j^2\right)\bar{\sigma}^2$$

$$cov(\Delta c_t^a, \Delta c_{t+1}^a) = cov(\sum_{j=0}^{2(h-1)} \tau_j \Delta c_{ht-j}, \sum_{j=0}^{2(h-1)} \tau_j \Delta c_{h(t+1)-j})$$

$$= cov\left(\sum_{j=0}^{2(h-1)} \tau_j x_{ht-j-1} + \sum_{j=0}^{2(h-1)} \tau_j \sigma_{ht-j-1}\eta_{ht-j}, \sum_{m=0}^{2(h-1)} \tau_m x_{h(t+1)-m-1} + \sum_{m=0}^{2(h-1)} \tau_m \sigma_{h(t+1)-m-1}\eta_{h(t+1)-m}\right)$$

$$= \left(\sum_{j=0}^{h-2}\left[\sum_{m=0}^{j+h-1} \tau_j\tau_m\rho^{h-m+j} + \sum_{m=j+h}^{2(h-1)} \tau_j\tau_m\rho^{m-j-h}\right] + \sum_{j=h-1}^{2(h-1)}\sum_{m=0}^{2(h-1)} \tau_j\tau_m\rho^{h-m+j}\right)\left(\frac{\varphi_e^2\bar{\sigma}^2}{1-\rho^2}\right)$$

$$+ \left(\sum_{m=0}^{h-2}\tau_{m+h}\tau_m\right)\bar{\sigma}^2$$

$$cov(\Delta c_t^a, \Delta c_{t+1}^a) = \rho^h(\rho b_{h-2})^2\left(\frac{\varphi_e^2\bar{\sigma}^2}{1-\rho^2}\right) + \sum_{j=0}^{h-2}\left[\rho^{j+2}b_j b_{h-2} + a_j b_j\right]\varphi_e^2\bar{\sigma}^2 + \rho b_{h-2}a_{h-1}\varphi_e^2\bar{\sigma}^2$$

$$+ \left(\sum_{m=0}^{h-2}\tau_{m+h}\tau_m\right)\bar{\sigma}^2$$

More generally, we can show that for $s \geq 1$,

$$cov(\Delta c_t^a, \Delta c_{t+s}^a) = \rho^{hs} \left(\rho b_{h-2}\right)^2 \left(\frac{\varphi_e^2 \bar{\sigma}^2}{1-\rho^2}\right) + \sum_{j=0}^{h-2} \left[\rho^{j+h(s-1)+2} b_j b_{h-2}\right] \varphi_e^2 \bar{\sigma}^2$$

$$+ \mathbb{1}\{s=1\} \left\{\rho b_{h-2} a_{h-1} + \sum_{j=0}^{h-2} a_j b_j\right\} \varphi_e^2 \bar{\sigma}^2 + \mathbb{1}\{s=1\} \left(\sum_{m=0}^{h-2} \tau_{m+h} \tau_m\right) \bar{\sigma}^2$$

$$+ \mathbb{1}\{s>1\} \left\{\sum_{j=0}^{h-1} \left[\rho^{j+h(s-2)+2} a_j b_{h-2}\right] \varphi_e^2 \bar{\sigma}^2\right\}$$

- **Annual dividend growth**

Using the same formulation as for the consumption growth rate, the annual dividend growth rate can be expressed in term of monthly dividend growth rates as follow:

$$\Delta d_{t+1}^a = log \frac{\sum_{j=0}^{h-1} D_{h(t+1)-j}}{\sum_{j=0}^{h-1} D_{ht-j}} \approx \sum_{j=0}^{2(h-1)} \tau_j \Delta d_{h(t+1)-j}$$

$$\Delta d_{t+1}^a = h\mu_d + \phi\rho b_{h-2} x_{h(t-1)} + \phi\varphi_e \sum_{j=0}^{h-1} a_j v_{h(t+1)-j-1} + \phi\varphi_e \sum_{j=0}^{h-2} b_j v_{ht-j-1}$$

$$+\pi \sum_{j=0}^{2(h-1)} \tau_j \sigma_{h(t+1)-j-1} \eta_{h(t+1)-j} + \varphi_d \sum_{j=0}^{2(h-1)} \tau_j \sigma_{h(t+1)-j-1} u_{h(t+1)-j}$$

Then

$$E(\Delta d_t^a) = E(\sum_{j=0}^{2(h-1)} \tau_j \Delta d_{ht-j}) = h\mu_d$$

$$var(\Delta d_t^a) = var\left(\phi \sum_{j=0}^{2(h-1)} \tau_j x_{ht-j-1}\right) + var\left(\varphi_d \sum_{j=0}^{2(h-1)} \tau_j \sigma_{ht-j-1} u_{ht-j}\right) + var(\pi \sum_{j=0}^{2(h-1)} \tau_j \sigma_{ht-j-1} \eta_{ht-j})$$

$$= \phi^2 \left[\sum_{j=0}^{2(h-1)} \tau_j^2 + 2 \sum_{j=0}^{2h-3} \sum_{k=j+1}^{2(h-1)} \tau_j \tau_k \rho^{k-j}\right] \left(\frac{\varphi_e^2 \bar{\sigma}^2}{1-\rho^2}\right) + \left(\varphi_d^2 + \pi^2\right) \sum_{j=0}^{2(h-1)} \tau_j^2 \bar{\sigma}^2$$

$$var(\Delta d_t^a) = (\phi\rho b_{h-2})^2 \left(\frac{\varphi_e^2\bar{\sigma}^2}{1-\rho^2}\right) + \left(\sum_{j=0}^{h-1} a_j^2 + \sum_{j=0}^{h-2} b_j^2\right) \phi^2\varphi_e^2\bar{\sigma}^2 + \left(\pi^2+\varphi_d^2\right) \left(\sum_{j=0}^{2(h-1)} \tau_j^2\right) \bar{\sigma}^2$$

$$cov(\Delta d_t^a, \Delta d_{t+1}^a) = cov\left(\sum_{j=0}^{2(h-1)} \phi\tau_j x_{ht-j-1} + \sum_{j=0}^{2(h-1)} \tau_j\varphi_d\sigma_{ht-j-1}u_{ht-j} + \sum_{j=0}^{2(h-1)} \tau_j\pi\sigma_{ht-j-1}\eta_{ht-j}\right.,$$

$$\left.\sum_{m=0}^{2(h-1)} \phi\tau_m x_{h(t+1)-m-1} + \sum_{m=0}^{2(h-1)} \tau_m\varphi_d\sigma_{h(t+1)-m-1}u_{h(t+1)-m} + \sum_{m=0}^{2(h-1)} \tau_m\pi\sigma_{h(t+1)-m-1}\eta_{h(t+1)-m}\right)$$

$$= \phi^2 \left(\sum_{j=0}^{h-2} \left[\sum_{m=0}^{j+h-1} \tau_j\tau_m\rho^{h-m+j} + \sum_{m=j+h}^{2(h-1)} \tau_j\tau_m\rho^{m-j-h}\right] + \sum_{j=h-1}^{2(h-1)} \sum_{m=0}^{2(h-1)} \tau_j\tau_m\rho^{h-m+j}\right) \left(\frac{\varphi_e^2\bar{\sigma}^2}{1-\rho^2}\right)$$

$$+ \left(\varphi_d^2+\pi^2\right) \left(\sum_{m=0}^{h-2} \tau_{m+h}\tau_m\right) \bar{\sigma}^2$$

$$cov(\Delta d_t^a, \Delta d_{t+1}^a) = \rho^h (\phi\rho b_{h-2})^2 \left(\frac{\varphi_e^2\bar{\sigma}^2}{1-\rho^2}\right) + \sum_{j=0}^{h-2} \left[\rho^{j+2} b_j b_{h-2} + a_j b_j\right] \phi^2\varphi_e^2\bar{\sigma}^2$$

$$+ (\rho b_{h-2} a_{h-1}) \phi^2\varphi_e^2\bar{\sigma}^2 + \left(\varphi_d^2+\pi^2\right) \left(\sum_{m=0}^{h-2} \tau_{m+h}\tau_m\right) \bar{\sigma}^2$$

$$cov(\Delta c_t^a, \Delta d_t^a) = cov\left(\sum_{j=0}^{2(h-1)} \tau_j x_{ht-j-1} + \sum_{j=0}^{2(h-1)} \tau_j\sigma_{ht-j-1}\eta_{ht-j}, \sum_{j=0}^{2(h-1)} \phi\tau_j x_{ht-j-1}\right.$$

$$\left. + \sum_{j=0}^{2(h-1)} \tau_j\varphi_d\sigma_{ht-j-1}u_{ht-j} + \sum_{j=0}^{2(h-1)} \tau_j\pi\sigma_{ht-j-1}\eta_{ht-j}\right)$$

$$= \phi var(\sum_{j=0}^{2(h-1)} \tau_j x_{ht-j-1}) + \pi var(\sum_{j=0}^{2(h-1)} \tau_j\sigma_{ht-j-1}\eta_{ht-j})$$

$$= \phi \left[\sum_{j=0}^{2(h-1)} \tau_j^2 + 2 \sum_{j=0}^{2h-3} \sum_{k=j+1}^{2(h-1)} \tau_j \tau_k \rho^{k-j} \right] \left(\frac{\varphi_e^2 \bar{\sigma}^2}{1-\rho^2} \right) + \pi \left(\sum_{j=0}^{2(h-1)} \tau_j^2 \right) \bar{\sigma}^2$$

$$cov(\Delta c_t^a, \Delta d_{t+1}^a) = \phi \left(\rho b_{h-2}\right)^2 \left(\frac{\varphi_e^2 \bar{\sigma}^2}{1-\rho^2} \right) + \left[\sum_{j=0}^{h-1} a_j^2 + \sum_{j=0}^{h-2} b_j^2 \right] \phi \varphi_e^2 \bar{\sigma}^2$$

$$+\pi \left(\sum_{j=0}^{2(h-1)} \tau_j^2 \right) \bar{\sigma}^2$$

- **Annual market return**

Let's denote

$$\Gamma_0 = k_{0m} + (k_{1m} - 1) A_{0m} + k_{1m} A_{2m} (1-\nu) \bar{\sigma}^2 + \mu_d$$

$$\Gamma_{1m} = A_{1m}(k_{1m}\rho - 1) + \phi \quad \Gamma_{2m} = (k_{1m}\nu - 1) A_{2m}$$

$$\beta_{m,e} = k_{1m} A_{1m} \varphi_e;\ \beta_{m,w} = k_{1m} A_{2m} \sigma_w$$

Then

$$r_{m,t+j}^a = \sum_{k=0}^{h-1} \big(\Gamma_0 + \Gamma_{1m} x_{h(t+j)-k-1} + \Gamma_{2m} \sigma_{h(t+j)-k-1}^2 + \beta_{m,e} v_{h(t+j)-k}$$

$$+ \beta_{m,w} w_{h(t+j)-k} + \varphi_d \sigma_{h(t+j)-k-1} u_{h(t+j)-k} + \pi \sigma_{h(t+j)-k-1} \eta_{h(t+j)-k} \big)$$

$$= h\Gamma_0 + \Gamma_{1m} \left(\sum_{k=0}^{h-1} \rho^{h(j+1)-1-k} x_{h(t-1)} + \sum_{k=0}^{h-1} \sum_{r=0}^{h(j+1)-2-k} \varphi_e \rho^r v_{h(t+j)-k-r-1} \right)$$

$$+\Gamma_{2m} \left(\sum_{k=0}^{h-1} (1-\nu) \left(\sum_{r=0}^{h(j+1)-2-k} \nu^r \right) \bar{\sigma}^2 + \sum_{k=0}^{h-1} \nu^{h(j+1)-1-k} \sigma_{h(t-1)}^2 + \sigma_w \sum_{k=0}^{h-1} \sum_{r=0}^{h(j+1)-2-k} \nu^r w_{h(t+j)-k-r-1} \right)$$

$$+ \sum_{k=0}^{h-1} \beta_{m,e} v_{h(t+j)-k} + \sum_{k=0}^{h-1} \beta_{m,w} w_{h(t+j)-k}$$

$$+\varphi_d \sum_{k=0}^{h-1} \sigma_{h(t+j)-k-1} u_{h(t+j)-k} + \pi \sum_{k=0}^{h-1} \sigma_{h(t+j)-k-1} \eta_{h(t+j)-k}$$

$$= h\Gamma_0 + \Gamma_{2m} \left(h - \nu^{hj} \left(\frac{1-\nu^h}{1-\nu} \right) \right) \bar{\sigma}^2 + \Gamma_{1m} \rho^{hj} \left(\frac{1-\rho^h}{1-\rho} \right) x_{h(t-1)} + \Gamma_{2m} \nu^{hj} \left(\frac{1-\rho^h}{1-\rho} \right) \sigma_{h(t-1)}^2$$

$$+ \sum_{k=0}^{h-1} \beta_{m,e} v_{h(t+j)-k} + \varphi_e \Gamma_{1m} \left(\sum_{r=0}^{h-1} \left(\frac{1-\rho^{r+1}}{1-\rho} \right) v_{h(t+j)-r-1} + \left(\frac{1-\rho^h}{1-\rho} \right) \sum_{r=h}^{h(j+1)-2} \rho^{(r-h+1)} v_{h(t+j)-r-1} \right)$$

$$+\sum_{k=0}^{h-1}\beta_{m,w}w_{h(t+j)-k}+\sigma_w\Gamma_{2m}\left(\sum_{r=0}^{h-1}\left(\frac{1-\nu^h}{1-\nu}\right)w_{h(t+j)-r-1}+\left(\frac{1-\nu^h}{1-\nu}\right)\sum_{r=h}^{h(j+1)-2}\nu^{(r-h+1)}w_{h(t+j)-r-1}\right)$$

$$+\varphi_d\sum_{k=0}^{h-1}\sigma_{h(t+j)-k-1}u_{h(t+j)-k}+\pi\sum_{k=0}^{h-1}\sigma_{h(t+j)-k-1}\eta_{h(t+j)-k}$$

and

$$E(r^a_{m,t})=h\Gamma_0+h\Gamma_{2m}\sigma^2$$

$$var(r^a_{m,t})=var(\sum_{j=0}^{h-1}r_{m,ht-j})=var(\sum_{j=0}^{h-1}\Gamma_0+\Gamma_{1m}x_{ht-j-1}+\Gamma_{2m}\sigma^2_{ht-j-1}+\beta_{m,e}\sigma_{ht-j-1}e_{ht-j}$$

$$+\beta_{m,w}w_{ht-j}+\varphi_d\sigma_{ht-j-1}u_{ht-j}+\pi\sigma_{ht-j-1}\eta_{ht-j})$$

$$=\sum_{j=0}^{h-1}(\Gamma^2_{1m}var(x_{ht-j-1})+\Gamma^2_{2m}var(\sigma^2_{ht-m-1})+\beta^2_{m,e}\bar{\sigma}^2+\beta^2_{m,w}+\varphi^2_d\bar{\sigma}^2)+2\sum_{j=0}^{h-2}\sum_{k=j+1}^{h-1}cov(\Gamma_{1m}x_{ht-j-1},\Gamma_{1m}x_{ht-k-1})$$

$$+\Gamma^2_{2m}cov(\sigma^2_{ht-j-1},\sigma^2_{ht-k-1})+\Gamma_{1m}\beta_{m,e}cov(x_{ht-j-1},\sigma_{ht-k-1}e_{ht-k})+\Gamma_{2m}\beta_{m,w}cov(\sigma^2_{ht-m-1},w_{ht-k})$$

$$=\left[h+\frac{2h\rho}{1-\rho}-\frac{2\rho(1-\rho^h)}{(1-\rho)^2}\right]\frac{\varphi^2_e\bar{\sigma}^2}{1-\rho^2}\Gamma^2_{1m}+\left[h+\frac{2h\nu}{1-\nu}-\frac{2\nu(1-\nu^h)}{(1-\nu)^2}\right]\frac{\sigma^2_w}{1-\nu^2}\Gamma^2_{2m}+h[(\beta^2_{m,e}+\varphi^2_e+\pi^2)\bar{\sigma}^2+\beta^2_{m,w}]$$

$$+2\Gamma_{1m}\beta_{m,e}\varphi_e\bar{\sigma}^2\left(\frac{h}{1-\rho}-\frac{1-\rho^h}{(1-\rho)^2}\right)+2\Gamma_{2m}\beta_{m,w}\sigma_w\left(\frac{h}{1-\nu}-\frac{1-\nu^h}{(1-\nu)^2}\right)$$

$$cov(r^a_{m,t},r^a_{m,t+1})=\rho\left(\frac{\varphi^2_e\bar{\sigma}^2}{1-\rho^2}\right)\left(\frac{1-\rho^h}{1-\rho}\right)^2\Gamma^2_{1m}+\nu\left(\frac{\sigma^2_w}{1-\nu^2}\right)\left(\frac{1-\nu^h}{1-\nu}\right)^2\Gamma^2_{2m}+\beta_{m,e}\Gamma_{1m}\varphi_e\bar{\sigma}^2\left(\frac{1-\rho^h}{1-\rho}\right)^2$$

$$+\beta_{m,w}\Gamma_{2m}\sigma_w\left(\frac{1-\nu^h}{1-\nu}\right)^2$$

- **Annual risk free rate**

$$r^a_{f,t+j}=\sum_{k=0}^{h-1}A_{0,f}+A_{1,f}x_{12(t+j)-k}+A_{2,,f}\sigma^2_{h(t+j)-k}$$

$$=hA_{0f}+\sum_{k=0}^{h-1}\left[A_{1f}\left(\rho^{h(j+1)-k}x_{h(t-1)}+\sum_{r=0}^{h(j+1)-k-1}\rho^r\varphi_e v_{h(t+j)-k-r}\right)\right.$$

$$\left.+A_{2f}\left(\nu^{h(j+1)-k}\sigma^2_{h(t+1)}+\sum_{r=0}^{12j+11-k}\nu^r\sigma_w w_{h(t+j)-k-r}+(1-\nu^{h(j+1)-k})\bar{\sigma}^2\right)\right]$$

$$=hA_{0f}+A_{1f}\rho^{hj+1}\left(\frac{1-\rho^h}{1-\rho}\right)x_{h(t-1)}+\sum_{k=0}^{h-1}A_{1f}\varphi_e\left(\sum_{r=0}^{h(j+1)-k-1}\rho^r v_{h(t+j)-k-r}\right)$$

$$+A_{2f}\nu^{hj+1}\left(\frac{1-\nu^h}{1-\nu}\right)\sigma^2_{h(t-1)}+\sum_{k=0}^{h-1}(1-\nu^{h(j+1)-k})A_{2,f}\bar{\sigma}^2+\sum_{k=0}^{h-1}A_{2,f}\sigma_w\left(\sum_{r=0}^{12j+11-k}\nu^r w_{h(t+j)-k-r}\right)$$

$$=hA_{0f}+A_{1f}\rho^{hj+1}\left(\frac{1-\rho^h}{1-\rho}\right)x_{h(t-1)}+A_{2f}\nu^{hj+1}\left(\frac{1-\nu^h}{1-\nu}\right)\sigma^2_{h(t-1)}+A_{2,f}\left(h-\nu^{hj+1}\left(\frac{1-\nu^h}{1-\nu}\right)\right)\bar{\sigma}^2$$

$$+A_{1f}\varphi_e\left[\sum_{r=0}^{h-1}\left(\frac{1-\rho^{r+1}}{1-\rho}\right)v_{h(t+j)-r}+\sum_{r=12}^{h(j+1)-1}\rho^{r-h+1}\left(\frac{1-\rho^h}{1-\rho}\right)v_{h(t+j)-r}\right]$$

$$+A_{2f}\sigma_w\left[\sum_{r=0}^{h-1}\left(\frac{1-\nu^{r+1}}{1-\nu}\right)w_{h(t+j)-r}+\sum_{r=12}^{h(j+1)-1}\nu^{r-h+1}\left(\frac{1-\nu^h}{1-\nu}\right)w_{h(t+j)-r}\right]$$

In particular,

$$r^a_{f,t}=\sum_{j=0}^{h-1}r_{f,ht-j}=\sum_{j=0}^{h-1}A_{0f}+A_{1f}x_{ht-j}+A_{2f}\sigma^2_{ht-j}$$

$$=hA_{0f}+A_{1f}\rho\left(\frac{1-\rho^h}{1-\rho}\right)x_{h(t-1)}+A_{2f}\nu\left(\frac{1-\nu^h}{1-\nu}\right)\sigma^2_{h(t-1)}+\left[h-\nu\left(\frac{1-\nu^h}{1-\nu}\right)\right]A_{2f}\bar{\sigma}^2$$

$$+A_{1f}\varphi_e\sum_{j=0}^{h-1}\left(\frac{1-\rho^{(j+1)}}{1-\rho}\right)v_{ht-j}+A_{2f}\sigma_w\sum_{j=0}^{h-1}\left(\frac{1-\nu^{(j+1)}}{1-\nu}\right)w_{ht-j}$$

$$E(r^a_{f,t})=hA_{0f}+hA_{2f}\bar{\sigma}^2$$

$$var(r^a_{f,t})=\left[h+\frac{2h\rho}{1-\rho}-2\rho\frac{\left(1-\rho^h\right)}{(1-\rho)^2}\right]\frac{\varphi_e^2\bar{\sigma}^2}{1-\rho^2}A^2_{1f}+\left[h+\frac{2h\nu}{1-\nu}-2\nu\frac{\left(1-\nu^h\right)}{(1-\nu)^2}\right]\frac{\sigma^2_w}{1-\nu^2}A^2_{2f}$$

$$cov(r^a_{f,t},r^a_{f,t+1})=\left(\frac{\varphi_e^2\bar{\sigma}^2}{1-\rho^2}\right)\rho\left(\frac{1-\rho^h}{1-\rho}\right)^2A^2_{1f}+\left(\frac{\sigma^2_w}{1-\nu^2}\right)\nu\left(\frac{1-\nu^h}{1-\nu}\right)^2A^2_{2f}$$

C.1 Expression of asset prices and cash flows in terms of the latent processes

- **Annual price-dividend ratio**

Form (3) and (4), we get that

$$p^a_t-d^a_t=\log P_{ht}-\log\sum_{j=0}^{h-1}D_{ht-j}$$

$$=\log P_{ht}-\log D_{ht}+\sum_{j=0}^{h-1}(\log D_{ht-j}-\log D_{ht-j-1})-\log h-\sum_{j=0}^{h-1}\frac{j+1}{h}\triangle d_{ht-j}$$

$$= z_{m,ht} + \sum_{j=0}^{h-1} \triangle d_{ht-j} - \log h - \sum_{j=0}^{h-1} \frac{j+1}{h} \triangle d_{ht-j}$$

$$p_t^a - d_t^a = z_{m,ht} - \log h - \sum_{j=0}^{h-1} \left(\frac{j+1}{h} - 1 \right) \triangle d_{ht-j}$$

$$= A_{0m} + A_{1m} x_{ht} + A_{2m} \sigma_{ht}^2 - \log(h) - \sum_{j=0}^{h-1} \left(\frac{j+1}{h} - 1 \right) (\mu_d + \phi x_{ht-j-1} + \pi \sigma_{ht-j-1} \eta_{ht-j} + \varphi_d \sigma_{ht-j-1} u_{ht-j})$$

$$= [A_{0m} - \log(h) - \sum_{j=0}^{h-1} \left(\frac{j+1}{h} - 1 \right) \mu_d] + A_{1m} x_{ht} - \sum_{j=0}^{h-1} \left(\frac{j+1}{h} - 1 \right) \phi x_{ht-j-1}$$

$$- \sum_{j=0}^{h-1} \left(\frac{j+1}{h} - 1 \right) \pi \sigma_{ht-j-1} \eta_{ht-j} + A_{2m} \sigma_{ht}^2 - \sum_{j=0}^{h-1} \left(\frac{j+1}{h} - 1 \right) \varphi_d \sigma_{ht-j-1} u_{ht-j}$$

$$= \left[A_{0m} + 0.5(h-1)\mu_d - \log(h) + A_{2m} \left(1 - \nu^h \right) \bar{\sigma}^2 \right] + \sum_{j=1}^{h-1} \left[A_{1m} \rho^j - \phi a'_{j-1} \right] \varphi_e v_{ht-j}$$

$$+ \left[A_{1m} \rho^h - \phi a'_{h-1} \right] x_{h(t-1)} + A_{2m} \nu^h \sigma_{h(t-1)}^2 - \varphi_d \sum_{j=0}^{h-1} \left(\frac{j+1}{h} - 1 \right) \sigma_{ht-j-1} u_{ht-j}$$

$$+ A_{1m} \varphi_e v_{ht} - \pi \sum_{j=0}^{h-1} \left(\frac{j+1}{h} - 1 \right) \sigma_{ht-j-1} \eta_{ht-j} + A_{2m} \sigma_w \sum_{j=0}^{h-1} \nu^j w_{ht-j}$$

where

$$a'_j = a_j - \left(\frac{1 - \rho^{(j+1)}}{1 - \rho} \right) \qquad for \quad j \in \{0, ..., h-1\}$$

$$E(p_t^a - d_t^a) = [A_{0m} - \log(h) + 0.5(h-1)\mu_d] + A_{2m} \bar{\sigma}^2$$

$$var(p_t^a - d_t^a) = \left[A_{1m} \rho^h - \phi a'_{h-1} \right]^2 \left(\frac{\varphi_e^2 \bar{\sigma}^2}{1 - \rho^2} \right) + \sum_{j=1}^{h-1} \left[A_{1m} \rho^j - \phi a'_{j-1} \right]^2 \varphi_e^2 \bar{\sigma}^2$$

$$+ (A_{1m} \varphi_e)^2 \bar{\sigma}^2 + \sum_{j=0}^{h-1} \left(\frac{j+1}{h} - 1 \right)^2 \left(\pi^2 + \varphi_d^2 \right) \bar{\sigma}^2 + \frac{A_{2m}^2 \sigma_w^2}{1 - \nu^2}$$

$$cov(p_t^a - d_t^a, p_{t+1}^a - d_{t+1}^a) =$$

$$cov \left(A_{2m} \sigma_{ht}^2 + A_{1m} \varphi_e v_{12t} + \left(A_{1m} \rho^h - \phi a'_{h-1} \right) x_{h(t-1)} + \sum_{j=1}^{h-1} \left(A_{1m} \rho^j - \phi a'_{j-1} \right) \varphi_e v_{ht-j} \right.$$

$$\left. - \sum_{j=0}^{h-1} \left(\frac{j+1}{h} - 1 \right) \pi \sigma_{ht-j-1} \eta_{ht-j} - \sum_{j=0}^{h-1} \left(\frac{j+1}{h} - 1 \right) \varphi_d \sigma_{ht-j-1} u_{ht-j}, A_{2m} \sigma_{h(t+1)}^2 + A_{1m} \varphi_e v_{h(t+1)} \right.$$

$$+\left(A_{1m}\rho^h-\phi a_{h-1}^{'}\right)x_{ht}+\sum_{j=1}^{h-1}\left(A_{1m}\rho^j-\phi a_{j-1}^{'}\right)\varphi_e v_{h(t+1)-j}-\sum_{j=0}^{h-1}\left(\frac{j+1}{h}-1\right)\pi\sigma_{h(t+1)-j-1}\eta_{h(t+1)-j}$$

$$-\sum_{j=0}^{h-1}\left(\frac{j+1}{h}-1\right)\varphi_d\sigma_{h(t+1)-j-1}u_{h(t+1)-j}\Bigg)$$

$$cov(p_t^a-d_t^a,p_{t+1}^a-d_{t+1}^a)=\left[A_{1m}\rho^h-\phi a_{h-1}^{'}\right]^2\rho^h\left(\frac{\varphi_e^2\bar{\sigma}^2}{1-\rho^2}\right)+A_{1m}\left[A_{1m}\rho^h-\phi a_{h-1}^{'}\right]\varphi_e^2\bar{\sigma}^2$$

$$+\sum_{j=1}^{h-1}\left[A_{1m}\rho^j-\phi a_{j-1}^{'}\right]\left[A_{1m}\rho^h-\phi a_{h-1}^{'}\right]\rho^j\varphi_e^2\bar{\sigma}^2+A_{2m}^2\nu^h\left(\frac{\sigma_w^2}{1-\nu^2}\right)$$

- **Forward annual excess return**

Combining the expressions obtained for the annual market return and the annual risk free rate, we obtain that for $j\geq 1$,

$$r_{m,t+j}^a-r_{f,t+j}^a=h(\Gamma_0-A_{0,f})+h(\Gamma_{2m}-A_{2,f})\bar{\sigma}^2-\nu^{hj}\left(\frac{1-\nu^h}{1-\nu}\right)(\Gamma_{2m}-\nu A_{2,f})\,\bar{\sigma}^2$$

$$+\rho^{hj}\left(\frac{1-\rho^h}{1-\rho}\right)(\Gamma_{1m}-A_{1,f}\rho)\,x_{h(t-1)}+\nu^{hj}\left(\frac{1-\nu^h}{1-\nu}\right)(\Gamma_{2m}-A_{2,f}\nu)\,\sigma_{h(t-1)}^2$$

$$+\sum_{k=0}^{h-1}\left[\beta_{m,e}+\Gamma_{1m}\varphi_e\left(\frac{1-\rho^k}{1-\rho}\right)-A_{1,f}\varphi_e\left(\frac{1-\rho^{k+1}}{1-\rho}\right)\right]v_{h(t+j)-k}$$

$$+\sum_{k=0}^{h-1}\left[\beta_{m,w}+\Gamma_{2m}\sigma_w\left(\frac{1-\nu^k}{1-\nu}\right)-A_{2,f}\sigma_w\left(\frac{1-\nu^{k+1}}{1-\nu}\right)\right]w_{h(t+j)-k}$$

$$+\sum_{k=h}^{hj+h-1}\left[\Gamma_{1m}-\rho A_{1,f}\right]\varphi_e\left(\frac{1-\rho^h}{1-\rho}\right)\rho^{k-h}v_{h(t+j)-k}$$

$$+\sum_{k=h}^{hj+h-1}\left[\Gamma_{2m}-A_{2,f}\nu\right]\sigma_w\nu^{k-h}\left(\frac{1-\nu^h}{1-\nu}\right)w_{h(t+j)-k}$$

$$+\pi\sum_{k=0}^{h-1}\sigma_{h(t+j)-k-1}\eta_{h(t+1)-k}+\varphi_d\sum_{k=0}^{h-1}\sigma_{h(t+j)-k-1}u_{h(t+j)-k}$$

So,

$$\sum_{j=1}^{J}\left(r_{m,t+j}^a-r_{f,t+j}^a\right)=\mu_r^{a,J}+\phi_x^{a,J}x_{h(t-1)}+\phi_\sigma^{a,J}\sigma_{h(t-1)}^2+e_t^{a,J}+w_t^{a,J}+\varphi_d u_t^{a,J}+\pi\eta_t^{a,J}$$

where $\mu_r^{a,J}=hJ(\Gamma_0-A_{0,f})+hJ(\Gamma_{2m}-A_{2,f})\bar{\sigma}^2-\nu^h\left(\frac{1-\nu^{hJ}}{1-\nu}\right)(\Gamma_{2m}-\nu A_{2,f})\,\bar{\sigma}^2$,
$\phi_x^{a,J}=\rho^h\left(\frac{1-\rho^{hJ}}{1-\rho}\right)(\Gamma_{1m}-A_{1,f}\rho)$, $\phi_\sigma^{a,J}=\nu^h\left(\frac{1-\nu^{hJ}}{1-\nu}\right)(\Gamma_{2m}-A_{2,f}\nu)$,

$$e_t^{a,J} = \sum_{j=1}^{J} \left(\sum_{k=0}^{h-1} \left[\beta_{m,e} + \Gamma_{1m}\varphi_e \left(\frac{1-\rho^k}{1-\rho} \right) - A_{1,f}\varphi_e \left(\frac{1-\rho^{k+1}}{1-\rho} \right) \right] v_{h(t+j)-k} \right.$$
$$\left. + \sum_{k=h}^{hj+h-1} \left[\Gamma_{1m} - \rho A_{1,f} \right] \varphi_e \left(\frac{1-\rho^h}{1-\rho} \right) \rho^{k-h} v_{h(t+j)-k} \right)$$

remind that $v_{h(t+1)-j-1} = \sigma_{h(t+1)-j-2} e_{h(t+1)-j-1}$.

$$w_t^{a,J} = \sum_{j=1}^{J} \left(\sum_{k=0}^{h-1} \left[\beta_{m,w} + \Gamma_{2m}\sigma_w \left(\frac{1-\nu^k}{1-\nu} \right) - A_{2,f}\sigma_w \left(\frac{1-\nu^{k+1}}{1-\nu} \right) \right] w_{h(t+j)-k} \right.$$
$$\left. + \sum_{k=h}^{hj+h-1} \left[\Gamma_{2m} - A_{2,f}\nu \right] \sigma_w \nu^{k-h} \left(\frac{1-\nu^h}{1-\nu} \right) w_{h(t+j)-k} \right)$$

$u_t^{a,J} = \sum_{j=1}^{J} \sum_{k=0}^{h-1} \sigma_{h(t+j)-k-1} u_{h(t+j)-k}$ and $\eta_t^{a,J} = \sum_{j=1}^{J} \sum_{k=0}^{h-1} \sigma_{h(t+j)-k-1} \eta_{h(t+1)-k}$

C.2 Theoretical moments for the predictive regression

- **Prediction of future excess return by the log price-dividend ratio**

$$cov(r^a_{m,t+j} - r^a_{f,t+j}, p^a_t - d^a_t) = \left[A_{1m}\rho^h - \phi a'_{h-1} \right] \rho^{hj} \left(\frac{1-\rho^h}{1-\rho} \right) (\Gamma_{1m} - \rho A_{1,f}) \left(\frac{\varphi_e^2 \bar{\sigma}^2}{1-\rho^2} \right)$$
$$+ A_{2m}\nu^{h(j+1)} \left(\frac{1-\nu^h}{1-\nu} \right) (\Gamma_{2m} - A_{2,f}\nu) \left(\frac{\sigma_w^2}{1-\nu^2} \right) + A_{1m}\rho^{12(j-1)} \left(\frac{1-\rho^h}{1-\rho} \right) (\Gamma_{1m} - \rho A_{1,f}) \varphi_e^2 \bar{\sigma}^2$$
$$+ \sum_{k=1}^{h-1} \rho^{h(j-1)+k} \left[A_{1m}\rho^k - \phi a'_{k-1} \right] \left(\frac{1-\rho^h}{1-\rho} \right) (\Gamma_{1m} - \rho A_{1,f}) \varphi_e^2 \bar{\sigma}^2$$
$$+ \sum_{k=0}^{h-1} A_{2m}\nu^{h(j-1)+2k} \left(\frac{1-\nu^h}{1-\nu} \right) (\Gamma_{2m} - \nu A_{2,f}) \sigma_w^2$$

$$cov\left(\sum_{j=1}^{J} (r^a_{m,t+j} - r^a_{f,t+j}), p^a_t - d^a_t \right) = \sum_{j=1}^{J} cov(r^a_{m,t+j} - r^a_{f,t+j}, p^a_t - d^a_t)$$

$$= A_{1m} \left(\frac{1-\rho^{hJ}}{1-\rho} \right) (\Gamma_{1m} - \rho A_{1,f}) \left(\frac{\varphi_e^2 \bar{\sigma}^2}{1-\rho^2} \right) + A_{2m} \left(\frac{1-\nu^{hJ}}{1-\nu} \right) (\Gamma_{2m} - \nu A_{2,f}) \left(\frac{\sigma_w^2}{1-\nu^2} \right)$$
$$- \phi a'_{h-1}\rho^h \left(\frac{1-\rho^{hJ}}{1-\rho} \right) (\Gamma_{1m} - \rho A_{1,f}) \left(\frac{\varphi_e^2 \bar{\sigma}^2}{1-\rho^2} \right) - \sum_{k=1}^{h-1} \phi a'_{k-1}\rho^k \left(\frac{1-\rho^{hJ}}{1-\rho} \right) (\Gamma_{1m} - \rho A_{1,f}) \varphi_e^2 \bar{\sigma}^2$$

- **Prediction of future consumption growth by the log price-dividend ratio**

$$cov\left(\sum_{j=1}^{J}\Delta c_{t+j}^{a},p_{t}^{a}-d_{t}^{a}\right)=\left(\rho b_{h-2}A_{1m}\right)\left[\frac{1-\rho^{h(J-1)}}{1-\rho^{h}}\right]\left(\varphi_{e}^{2}\bar{\sigma}^{2}\right)\mathbb{1}\left\{J\geq 2\right\}$$

$$+a_{h-1}A_{1m}\left(\varphi_{e}^{2}\bar{\sigma}^{2}\right)+\left(\rho b_{h-2}\right)\left[\frac{1-\rho^{h(J-1)}}{1-\rho^{h}}\right]\left(\sum_{k=1}^{h-1}\rho^{k}\left(A_{1m}\rho^{k}-\phi a_{k-1}^{'}\right)\right)\left(\varphi_{e}^{2}\bar{\sigma}^{2}\right)\mathbb{1}\left\{J\geq 2\right\}$$

$$+\left(\rho b_{h-2}\right)\left(A_{1m}\rho^{h}-\phi a_{h-1}^{'}\right)\left[\frac{1-\rho^{hJ}}{1-\rho^{h}}\right]\left(\frac{\varphi_{e}^{2}\bar{\sigma}^{2}}{1-\rho^{2}}\right)+\sum_{k=1}^{h-1}b_{k-1}\left(A_{1m}\rho^{k}-\phi a_{k-1}^{'}\right)\left(\varphi_{e}^{2}\bar{\sigma}^{2}\right)$$

$$-\pi\sum_{k=0}^{h-2}\left(\frac{k+1}{h}-1\right)\tau_{k+h}\bar{\sigma}^{2}$$

- **Prediction of future dividend growth by the log price-dividend ratio**

$$cov\left(\sum_{j=1}^{J}\Delta d_{t+j}^{a},p_{t}^{a}-d_{t}^{a}\right)=\phi\left(\rho b_{h-2}A_{1m}\right)\left[\frac{1-\rho^{h(J-1)}}{1-\rho^{h}}\right]\left(\varphi_{e}^{2}\bar{\sigma}^{2}\right)\mathbb{1}\left\{J\geq 2\right\}$$

$$+\phi a_{h-1}A_{1m}\left(\varphi_{e}^{2}\bar{\sigma}^{2}\right)+\phi\left(\rho b_{h-2}\right)\left[\frac{1-\rho^{h(J-1)}}{1-\rho^{h}}\right]\left(\sum_{k=1}^{h-1}\rho^{k}\left(A_{1m}\rho^{k}-\phi a_{k-1}^{'}\right)\right)\left(\varphi_{e}^{2}\bar{\sigma}^{2}\right)\mathbb{1}\left\{J\geq 2\right\}$$

$$+\phi\left(\rho b_{h-2}\right)\left(A_{1m}\rho^{h}-\phi a_{h-1}^{'}\right)\left[\frac{1-\rho^{hJ}}{1-\rho^{h}}\right]\left(\frac{\varphi_{e}^{2}\bar{\sigma}^{2}}{1-\rho^{2}}\right)+\phi\sum_{k=1}^{h-1}b_{k-1}\left(A_{1m}\rho^{k}-\phi a_{k-1}^{'}\right)\left(\varphi_{e}^{2}\bar{\sigma}^{2}\right)$$

$$-\left(\pi^{2}+\varphi_{d}^{2}\right)\sum_{k=0}^{h-2}\left(\frac{k+1}{h}-1\right)\tau_{k+h}\bar{\sigma}^{2}$$

D Short note on Gmm estimation and inference

D.1 Strong identification case

The Gmm estimation procedure has been very popular since the seminal paper by Hansen [1982]. Let's call $\theta \in R^{p}$ the vector of parameters of the model and θ_0 its true value. The set of unconditional moment restrictions can be summarized by :

$$E(g(y_t,\theta_0))=0 \tag{12}$$

θ is identified at θ_0 if $\theta=\theta_0$ is the unique solution of (12) for $\theta\epsilon\Theta$ (Stock and Wright, 2000). Let's denote $g_{_T}(y_t,\theta)=\frac{1}{T}\sum_{t=1}^{T}g(y_t,\theta)$ and $\bar{g}(\theta)=E(g(y_t,\theta)$

The Gmm estimator of θ minimizes the criterion function :

$$S_T(\theta;\bar{\theta}_T(\theta))=T.\left(g_T(y_t,\theta)\right)^{'}W_T(\bar{\theta}_T(\theta))\left(g_T(y_t,\theta)\right)$$

Where $W_T(\bar{\theta}_T(\theta))$ is a positive semi-definite weighting matrix and $\bar{\theta}_T(\theta)$ is the value of the parameters used to compute W_T. For the 2-step Gmm estimator, $W_T = I$ in the first step and for the second step,$\bar{\theta}_T(\theta)$ will be equal to the value estimated in the first step. For the CUE Gmm estimator, $\bar{\theta}_T(\theta) = \theta$. The optimal weighting matrix is the inverse of the asymptotic variance-covariance matrix of $\sqrt{T}g_T(y_t,\theta)$[19] which can be consistently estimated using a kernel based estimator (Newey and West [1987]; Andrews [1991]) by:

$$\widehat{\Omega}(\theta) = \widehat{\Omega}_0 + \sum_{j=1}^{m} \bar{k}_j \left(\widehat{\Omega}_j + \widehat{\Omega}_j^{'}\right)$$

where $\widehat{\Omega}_j = T^{-1}\sum_{t=j+1}^{T}\left(g(y_t,\theta) - g_T(y_t,\theta)\right)\left(g(y_{t-j},\theta) - g_T(y_t,\theta)\right)^{'}$ and $\bar{k}_j$is a kernel function (e.g: Bartlett, Quadratic Spectral, Parzen, Truncated, etc.). For the Bartlett kernel with bandwidth m,$\bar{k}_j = \begin{cases} (1-\frac{j}{m+1})\ if\ 0 \leq j \leq m \\ 0 \quad if \quad Not \end{cases}$ where $g(y_t,\theta) \in R^q \times R^T$ [20]. A part from being consistent, this estimator is also positive semi-definite. Newey and West [1994] also showed how to automatically select the number of lag ($m = parameter \times T^{1/3}$)[21]

From the Central Limit Theorem, we know that:

$$\sqrt{T}\bar{g}_T(y_t,\theta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(y_t,\theta_0) \rightarrow N(0, Avar\left(\sqrt{T}\bar{g}_T(y_t,\theta_0)\right)) \tag{13}$$

So

$$TJ_T(\theta_0) = T.\bar{g}(y_t,\theta_0)^{'}W_T(\theta_0)\bar{g}(y_t,\theta_0) \rightarrow \chi(q) \tag{14}$$

Where $W_T(\theta_0) = \widehat{\Omega}(\theta_0)^{-1}$, $\widehat{\Omega}(\theta_0)$ is a consistent estimate of the asymptotic variance-covariance matrix $\Omega(\theta_0)$.

If k parameters θ_k^u in the vector θ_0 are estimated while the others are known then under regularity conditions (see Newey and McFadden [1986],section 9.5 for proof):

[19] $\Omega(\theta) = \lim_{T\rightarrow+\infty} T.E\left(\left([g_T(y_t,\theta) - \bar{g}(\theta)]\right)\left([g_T(y_t,\theta) - \bar{g}(\theta)]\right)^{'}\right) = \sum_{j=-\infty}^{\infty} E\left(g(y_t,\theta)g(y_{t-j},\theta)'\right)$

[20] p is the number of parameters in the model, q is the number of non redundant moment restrictions; it is equal here to 16 when we consider only the basic moments and to 19 if we add the moment restrictions from the predictive regressions of the future excess return on the log price dividend ratio and 22 (resp. 25) if we add the moment restrictions from the predictive regressions of the future consumption (resp. future dividend) growth on the log price dividend ratio . T is the length of the data series)

[21] The choice of m is made by minimizing the asymptotic mean squared error ($\lim_{T\rightarrow\infty} E\{normalised\, w'(\hat{S}-S)w\}^2$). For the Bartlett kernel that we used here, $m(T) = \hat{\gamma}T^{1/3}$ with $\hat{\gamma} = c_\gamma\{\hat{s}^{(1)}/\hat{s}^{(0)}\}^{2/3}$,$c_\gamma \equiv \left(1/\int_{-1}^{1}(1-\mid x \mid)^2\, dx\right)^{1/3} = \sqrt[3]{3/2} \backsimeq 1.1447$,$\hat{s}^{(1)} = 2\sum_{j=1}^{n} j\hat{\sigma}_j$, $\hat{s}^{(0)} = \hat{\sigma}_0 + 2\sum_{j=1}^{n}\hat{\sigma}_j$,$\hat{\sigma}_j = (T-1)^{-1}\sum_{t=j+2}^{T}\{(w'g(y_t,\theta_0))(w'g(y_{t-j},\theta_0))\}$, $j = 0,...,n$ with $n = [4(T/100)^{2/9}]$ and $w = (1,...,1)'$

$$TJ_T(\widehat{\theta_k^u}, \theta_{-k}^0) = T.\bar{g}(y_t, \widehat{\theta_k^u}, \theta_{0,-k})\widehat{W}_T\bar{g}(y_t, \widehat{\theta_k^u}, \theta_{0,-k}) \rightarrow \chi\,(q-k) \tag{15}$$

where θ_k^u is the sub-vector of strongly identified parameters, $\theta_{0,-k}$ is the vector of weakly identified parameters and the weighting matrix $\widehat{W}_T$ is evaluated at the $(\widehat{\theta_k^u}, \theta_{-k}^0)$. This statistic is the same that we also use in the case of weak identification in the next section; the only difference being that in the weak identification case, because of the inconsistency of the weakly identified parameters, the weighting matrix needs to be evaluated at the same value of the parameters as the moment conditions (in other words we must use the continuously update estimator).

It follows that if we choose to fix $p-k$ parameters in the p-vector at their true values and to estimate k parameters, then :

$$TJ_T(\widehat{\theta_k^u}, \theta_{-k}^0) - TJ_T(\widehat{\theta_p^u}) \rightarrow \chi\,(p-k) \tag{16}$$

The three statistics $TJ_T(\widehat{\theta_p^u})$, $TJ_T(\theta_0)$,$TJ_T(\widehat{\theta_k^u}, \theta_{-k}^0) - TJ_T(\widehat{\theta_p^u})$ are used in Hansen et al. [1996] to study the finite sample properties of alternative GMM estimators and respectively denominated “Minimized", “True" and “Constrained-Minimized". This last statistic correspond to the quasi-likelihood (LR) statistic which is usually used to test $\theta = \theta_0$ and could be used to construct confidence region for the set of constrained parameters (θ_{-k}^0) based on the increment of the objective function from its unconstrained value. The second statistic corresponds to the Gmm criterion function computed using a consistent estimator of the weighting matrix evaluated at the parameter and it also correspond to the S-statistic of Stock and Wright [2000].

From a Taylor expansion of equation 13, the confidence intervals for the estimates are obtained using the fact that for the efficient Gmm (as our weighting matrix is a consistent estimate of the asymptotic variance-covariance matrix $\Omega(\theta_0)$), we have (see Hayashi (2000), Prop. 7.10):

$$\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow N(0, \left[G'\Omega(\theta_0)^{-1}G\right]^{-1}) \tag{17}$$

Where $G = E\left[\frac{\partial g(y_t,\theta_0)}{\partial \theta'}\right]$

D.2 Weak identification case

We further assumed that $\theta = (\alpha, \beta)$, where α is a p_α sub-vector of weakly identified parameters and β is the p_β sub-vector of strongly identified parameters ($p_\alpha + p_\beta = p$). This means that the moment conditions are zero at $\theta_0 = (\alpha_0, \beta_0)$, but are also very nearly zero for $\alpha \neq \alpha_0$: In other words, the population objective function is steep

in β around β_0, but nearly flat in α (Stock and Wright, 2000, Sect 2.3). In a structural economic model, α could be also seen as a vector of "less influential" or "nuisance" parameters, because their changes have no or only a slight influence on the equilibrium characterized by (7). Detecting the weakly identified parameters is not an easy task in a model with non linear equations. If $\bar{g}(y_t, \theta)$ is continuously differentiable and $\nabla_\theta E(\bar{g}(y_t, \theta)) = E(\nabla_\theta \bar{g}(y_t, \theta))$, a necessary and sufficient condition for local identification (which is also necessary for global identification), is that $E(\nabla_\theta \bar{g}(y_t, \theta))$ has a full column rank (Newey and McFadden, 1986), which is no more the case under weak identification.

When some parameters are weakly identified, the limiting distribution of the Gmm estimator is no more a standard normal distribution (Stock and Wright [2000]; Section 2.4). Indeed, the weakly identified parameter α is no more consistently estimated; it converges to the solution α^* of a non-quadratic rather than a local quadratic minimization problem, and thus this solution might be different to the true value of the parameter α_0 which put the objective function (when $\beta = \beta_0$) to 0 (the minimum could be negative and achieved at a parameter different from the true one). This inconsistency in the estimation of the weakly identified parameter affects the limiting distribution of the strongly identified parameters. In the end, *because the limiting distributions are nonstandard, confidence intervals for β constructed by inverting the quasi-likelihood ratio (LR) statistic $S_T(\theta_0; \bar{\theta}_T(\theta_0)) - S_T(\hat{\theta}; \bar{\theta}_T(\hat{\theta}))$ or the conventional Wald statistic will not be valid* (Stock and Wright, 2000).

However, if we assume that $\Psi_T(\theta_0) = \sqrt{T}([g_T(y_t, \theta_0) - \bar{g}(\theta_0)]) \xrightarrow{d} N(0, \Omega(\theta_0, \theta_0))$ and $W_T(\theta_0) \xrightarrow{p} W(\theta_0) = \Omega(\theta_0, \theta_0)^{-1}$ then $S_T(\theta_0; \theta_0) \xrightarrow{d} \chi^2_m$ (Stock and Wright [2000]; Theorem 2) and this could be used to construct the $(1-\lambda)\%$ confidence region of θ_0 by collecting all the parameters θ such that $S_T(\theta; \theta) \leq \chi^2_{(1-\lambda)}(m)$. Under stronger assumptions (Assumptions B,C,D in Stock and Wright, 2000), $S_T(\alpha_0, \hat{\beta}(\alpha_0); \alpha_0, \hat{\beta}(\alpha_0)) \xrightarrow{d} \chi^2_{m-p_2}$ and this could be used to construct the $(1-\lambda)\%$ confidence region of α_0 by collecting all the parameters α such that $S_T(\alpha, \hat{\beta}(\alpha); \alpha, \hat{\beta}(\alpha)) \leq \chi^2_{(1-\lambda)}(m-p_2)$. Notice that for all those statistic the criterion function used is for the CUE Gmm and the weighting matrix must be a consistent estimator of the inverse of the variance covariance matrix of $\sqrt{T}\bar{g}(y_t, \theta)$. Stock and Wright [2000] recommend to be cautious in the interpretation of the confidence set derived by inverting the S-statistic because the confidence region is the result of a joint test of two hypothesis (that $\theta = \theta_0$ and that the over-identifying conditions are valid). *So the S-set could be small either because the model is correctly specified and precisely estimated or because the model is misspecified but the evidence is two weak to reject it entirely.*

Kleibergen [2005] proposed a decomposition of this joint test into its two part using

two independent statistics. The S-statistic can be decomposed as:

$$S_T(\theta_0, \theta_0) = J(\theta_0, \theta_0) + K(\theta_0, \theta_0)$$

Where

$$J(\theta_0, \theta_0) = T.\bar{g}(y_t, \theta)' W_T(\theta_0)^{-1/2} M_{W_T(\theta_0)^{-1/2} D_T(\theta_0, Y)} W_T(\theta_0)^{-1/2} \bar{g}(y_t, \theta)$$

and

$$K(\theta_0, \theta_0) = \frac{1}{4} \left(\frac{\partial S_T(\theta, \theta)}{\partial \theta'} |_{\theta_0} \right) \left[D_T(\theta_0, Y)' W_T(\theta_0)^{-1} D_T(\theta_0, Y) \right]^{-1} \left(\frac{\partial S_T(\theta, \theta)}{\partial \theta'} |_{\theta_0} \right)'$$

$D_T(\theta_0, Y) = T.[q_{1,T}(y_t, \theta_0) - \hat{V}_{\theta g,1}(\theta_0) W_T(\theta_0)^{-1} \bar{g}_T(y_t, \theta_0) ... q_{p,T}(y_t, \theta_0) - \hat{V}_{\theta g,p}(\theta_0) W_T(\theta_0)^{-1} \bar{g}_T(y_t, \theta_0)]$ is $m \times p$ dimensional matrix,

$q_{i,T}(y_t, \theta_0) = \frac{\partial \bar{g}_T(y_t, \theta)}{\partial \theta_i} |_{\theta_0} \quad i = 1, .., p$ is $m \times 1$ dimensional vector and $\hat{V}_{\theta g}(\theta_0) = (\hat{V}_{\theta g,1}(\theta_0) ... \hat{V}_{\theta g,p}(\theta))$ is a $mp \times m$ dimensional matrix that consistently estimate

$$\lim_{T \to \infty} \frac{1}{T} E \left[\left(\underset{m \times T}{(q_1(y_t, \theta_0) - E(q_1(y_t, \theta_0)))'} ... (q_p(y_t, \theta_0) - E(q_p(y_t, \theta_0)))' \right) \left(\underset{T \times m}{g(y_t, \theta_0) - E(g(y_t, \theta_0))} \right) \right]$$

It is extracted from the variance-covariance matrix estimate of

$$V(\theta) = \lim_{T \to +\infty} var \left[\sqrt{T} \begin{pmatrix} g_T(\theta) \\ q_T(\theta) \end{pmatrix} \right] = \begin{pmatrix} V_{gg}(\theta) & V_{g\theta}(\theta) \\ V_{\theta g}(\theta) & V_{\theta\theta}(\theta) \end{pmatrix}$$

In the case of the moment conditions, we consider here, since there is a separation between the theoretical moment evaluated at θ and the empirical moment computed from the data (the difference between the two forms the moment conditions), $D_T(\theta_0, Y)$ is equal to the jacobian of the moment conditions which only depends on θ, such that $V_{\theta g}$ and $V_{\theta\theta}$ are both null matrices, and

$$V(\theta) = \begin{pmatrix} V_{gg}(\theta) & 0 \\ 0 & 0 \end{pmatrix}$$

Under the null hypothesis, which is $H_0 : \theta = \theta_0$, for the K-statistic, it converges to $\chi^2(p)$ and which is $H_m : E(g(y_t, \theta_0)) = 0$ for the J-statistic, it converges to $\chi^2(m-p)$ More specifically,

$$J(\theta_0, \theta_0) = T.\bar{g}(y_t, \theta)' \left[W_T(\theta_0)^{-1} - W_T(\theta_0)^{-1} D_T(\theta_0, Y) \left(D_T(\theta_0, Y)' W_T(\theta_0)^{-1} D_T(\theta_0, Y) \right)^{-1} \right.$$

$$\left. \times D_T(\theta_0, Y)' W_T(\theta_0)^{-1} \right] \bar{g}(y_t, \theta)$$

$$K(\theta_0, \theta_0) = T.\bar{g}(y_t, \theta)' W_T(\theta_0)^{-1} D_T(\theta_0, Y) \left[D_T(\theta_0, Y)' W_T(\theta_0)^{-1} D_T(\theta_0, Y) \right]^{-1}$$

$$\times D_T(\theta_0, Y)' W_T(\theta_0)^{-1} \bar{g}(y_t, \theta)$$

Testing the set of strongly identified parameters (β) is more complicated because α^*could be different from α_0 and then the limiting distribution under the null will be unconventional. The Dufour [2003] approach is to construct a valid $(1-\lambda)$ % confidence set for the all vector of parameters or for the weakly identified ones and to project out the other elements. This projection method becomes difficult to apply when the number of parameters is greater than 3[22] . Kleibergen and Mavroeidis [2009] also propose some subset weak instrument robust statistics (S-statistic, KLM, JKLM, MQLR) to test hypothesis and build confidence intervals for all subset of parameters (even if they are not weakly identified). In there setting, if you want to test $H_0 : \beta = \beta_0$ (here there is no more distinction between, strongly and weakly identified parameters), then you can compute the following statistics:

$$S(\beta_0) = S_T(\tilde{\alpha}(\beta_0), \beta_0; \tilde{\alpha}(\beta_0), \beta_0) \underset{a}{\preccurlyeq} \varphi_{p_\beta} + \varphi_{m-p}$$

$$KLM(\beta_0) = K(\tilde{\alpha}(\beta_0), \beta_0; \tilde{\alpha}(\beta_0), \beta_0) \underset{a}{\preccurlyeq} \varphi_{p_\beta}$$

$$JKLM(\beta_0) = S(\beta_0) - KLM(\beta_0) \underset{a}{\preccurlyeq} \varphi_{m-p}$$

$$MQLR(\beta_0) = \frac{1}{2}[KLM(\beta_0) + JKLM(\beta_0) - rk(\beta_0)$$

$$+(\{KLM(\beta_0) + JKLM(\beta_0) + rk(\beta_0)\}^2 - 4JKLM](\beta_0) rk(\beta_0))^{1/2}]$$

$$\underset{a}{\preccurlyeq} \frac{1}{2}[\varphi_{p_\beta} + \varphi_{m-p} - rk(\beta_0) + (\{\varphi_{p_\beta} + \varphi_{m-p} + rk(\beta_0)\}^2 - 4\varphi_{m-p} rk(\beta_0))^{1/2}]$$

where $rk(\beta_0)$ is a statistic that test the lower rank of the jacobian matrix of the objective function at $(\tilde{\alpha}(\beta_0), \beta_0)$ and which is given by :

$$rk(\beta_0) = \min_{v \in R^{p-1}} T \begin{pmatrix} 1 \\ v \end{pmatrix}' D_T(\tilde{\alpha}(\beta_0), \beta_0)' \left[\left(\begin{pmatrix} 1 \\ v \end{pmatrix} \otimes I_m \right)' V_{\theta\theta}(\tilde{\alpha}(\beta_0), \beta_0) \left(\begin{pmatrix} 1 \\ v \end{pmatrix} \otimes I_m \right) \right]^{-1}$$

$$\times D_T(\tilde{\alpha}(\beta_0), \beta_0) \begin{pmatrix} 1 \\ v \end{pmatrix}$$

[22]For example with a vector of $p = 3$ parameters, if we assume each parameter to belong initially to an interval that we discretize into only 10 values, then our grid will be made of 10^ppoints at which the objective function will be evaluated. So as p increases, the number of points in our grid growth exponentially and it becomes very difficult to settle the S-set.

They show that under the null hypothesis $H_0 : \beta = \beta_0$, the limiting distributions of those statistics is asymptotically bounded by the usual chi-squared distributions with the corresponding degrees of freedom: φ_{p_β} and φ_{m-p} are independent chi-squared distributed random variables with p_β and $m - p$ degrees of freedom, respectively. The drawback being that using the corresponding degree of freedom $(1 - \lambda)$ chi-squared quantile to build the confidence intervals will result in conservative confidence intervals when the unrestricted parameters are weakly identified, meaning that the maximum rejection probability of the tests over all values of the nuisance parameters is lower than the size of the test or equivalently the level of the confidence interval will be greater than $(1 - \lambda)$ %.

References

A. Abel. Asset prices under habit formation and catching up with the joneses. *American Economic Review*, 80(2):38–42, 1990. URL http://EconPapers.repec.org/RePEc:aea:aecrev:v:80:y:1990:i:2:p:38-42.

D. W. Andrews and X. Cheng. Gmm estimation and uniform subvector inference with possible identification failure. *Econometric Theory*, 30:287–333, 4 2014. ISSN 1469-4360. doi: 10.1017/S0266466613000315. URL http://journals.cambridge.org/article_S0266466613000315.

D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):pp. 817–858, 1991. ISSN 00129682. URL http://www.jstor.org/stable/2938229.

D. W. K. Andrews and X. Cheng. Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, 80(5):2153–2211, 2012. ISSN 1468-0262. doi: 10.3982/ECTA9456. URL http://dx.doi.org/10.3982/ECTA9456.

R. Bansal and A. Yaron. Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59(4):1481–1509, 2004. ISSN 1540-6261. doi: 10.1111/j.1540-6261.2004.00670.x. URL http://dx.doi.org/10.1111/j.1540-6261.2004.00670.x.

R. Bansal, A. R. Gallant, and G. Tauchen. Rational pessimism, rational exuberance, and asset pricing models. (13107), May 2007a. doi: 10.3386/w13107. URL http://www.nber.org/papers/w13107.

R. Bansal, D. Kiku, and A. Yaron. Risks for the long run: Estimation and inference. *Working paper*, 2007b.

R. Bansal, D. Kiku, and A. Yaron. An empirical evaluation of the long-run risks model for asset prices. *Critical Finance Review*, 1(1):183–221, 2012a. ISSN 2164-5744. doi: 10.1561/104.00000005. URL http://dx.doi.org/10.1561/104.00000005.

R. Bansal, D. Kiku, and A. Yaron. Risks for the long run: Estimation with time aggregation. Working Paper 18305, National Bureau of Economic Research, August 2012b. URL http://www.nber.org/papers/w18305.

R. J. Barro. Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics*, 121(3):823–866, 2006. doi: 10.1162/qjec.121.3.823. URL http://qje.oxfordjournals.org/content/121/3/823.abstract.

R. J. Barro. Rare disasters, asset prices, and welfare costs. *American Economic Review*, 99(1):243–64, 2009. doi: 10.1257/aer.99.1.243. URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.1.243.

J. Beeler and J. Y. Campbell. Appendix for "the long-run risks model and aggregate asset prices: An empirical assessment". *Critical Finance Review*, 2012.

C. Burnside. Solving asset pricing models with gaussian shocks. *Journal of Economic Dynamics and Control*, 22(3):329 – 340, 1998. ISSN 0165-1889. doi: http://dx.doi.org/10.1016/S0165-1889(97)00075-4. URL http://www.sciencedirect.com/science/article/pii/S0165188997000754.

L. E. Calvet and V. Czellar. Through the looking glass: Indirect inference via simple equilibria. *Journal of Econometrics*, 185(2):343 – 358, 2015. ISSN 0304-4076. doi: http://dx.doi.org/10.1016/j.jeconom.2014.11.003. URL http://www.sciencedirect.com/science/article/pii/S0304407614002681.

J. Y. Campbell and R. J. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1:195–228, 1988. doi: 10.1093/rfs/1.3.195.

J. Cochrane. *ASSET PRICING Revisited Edition*. Princeton University Press, 41 William street Princeton, New Jersey 08540, 2001.

J. H. Cochrane and J. Campbell. By force of habit: A consumption-based explanation of aggregate stock market behavior. (3119444), 1999. URL http://ideas.repec.org/p/hrv/faseco/3119444.html.

G. Constantinides. Habit formation: a resolution of the equity premium puzzle. 1990. URL http://EconPapers.repec.org/RePEc:cla:levarc:1397.

G. M. Constantinides and A. Ghosh. Asset pricing tests with long-run risks in consumption growth. *Review of Asset Pricing Studies*, 1(1):96–136, 2011. doi: 10.1093/rapstu/rar004. URL http://raps.oxfordjournals.org/content/1/1/96.abstract.

D. D. Creal and J. C. Wu. Bond risk premia in consumption-based models. *Working paper*, 4 2015.

O. de Groot. Solving asset pricing models with stochastic volatility. *Journal of Economic Dynamics and Control*, 52:308 – 321, 2015. ISSN 0165-1889. doi: http://dx.doi.org/10.1016/j.jedc.2015.01.001. URL http://www.sciencedirect.com/science/article/pii/S0165188915000020.

J.-m. Dufour. Identification, weak instruments, and statistical inference in econometrics. *JOURNAL OF ECONOMICS*, pages 767–808, 2003.

J.-m. Dufour. Comments on âweak instrument robust tests in gmm and the new keynesian phillips curveâ by f. kleibergen and s. mavroeidis. *Journal of Business & Economic Statistics*, 27(3):318–321, 2009.

L. G. Epstein and S. E. Zin. *Substitution, Risk A version, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework*, volume 57. July 1989. URL http://www.jstor.org/stable/1913778.

E. F. Fama and K. R. French. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3 – 25, 1988. ISSN 0304-405X. doi: http://dx.doi.org/10.1016/0304-405X(88)90020-7. URL http://www.sciencedirect.com/science/article/pii/0304405X88900207.

R. Garcia, N. Meddahi, and R. TÃ©dongap. An analytical framework for assessing asset pricing models and predictability. *Unpublished Paper, Edhec Business School, Imperial College London, and Stockholm School of Economics*, 2008. ISSN 0165-1889. doi: http://dx.doi.org/10.2139/ssrn.1109080. URL http://ssrn.com/abstract=1109080.

I. Ghattassi and N. Meddahi. Time-aggregation effects on estimating asset pricing models. Working paper, 2012. URL http://neeo.univ-tlse1.fr/3450/.

L. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–54, 1982. URL http://EconPapers.repec.org/RePEc:ecm:emetrp:v:50:y:1982:i:4:p:1029-54.

L. Hansen, J. Heaton, and A. Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business Economic Statistics*, 14(3):262–80, 1996. URL http://EconPapers.repec.org/RePEc:bes:jnlbes:v:14:y:1996:i:3:p:262-80.

R. J. Hodrick. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial studies*, 5(3):357–386, 1992.

A. Holly, A. Monfort, and M. Rockinger. Fourth order pseudo maximum likelihood methods. *Journal of Econometrics*, 162(2):278–293, June 2011. URL https://ideas.repec.org/a/eee/econom/v162y2011i2p278-293.html.

C. Julliard and A. Ghosh. Can rare events explain the equity premium puzzle? *Review of Financial Studies*, 2012. doi: 10.1093/rfs/hhs078. URL http://rfs.oxfordjournals.org/content/early/2012/09/01/rfs.hhs078.abstract.

F. Kleibergen. Testing parameters in gmm without assuming that they are identified. *Econometrica*, 73(4):1103–1123, 2005. ISSN 1468-0262. doi: 10.1111/j.1468-0262.2005.00610.x. URL http://dx.doi.org/10.1111/j.1468-0262.2005.00610.x.

F. Kleibergen and S. Mavroeidis. Inference on subsets of parameters in gmm without assuming identification. 2009.

R. Mehra. The Equity Premium: Why is it a Puzzle? NBER Working Papers 9512, National Bureau of Economic Research, Inc, Feb. 2003. URL http://ideas.repec.org/p/nbr/nberwo/9512.html.

J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7(4):308–313, 1965.

W. Newey and K. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–08, 1987. URL http://EconPapers.repec.org/RePEc:ecm:emetrp:v:55:y:1987:i:3:p:703-08.

W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4 of *Handbook of Econometrics*, chapter 36, pages 2111–2245. Elsevier, 1986. URL http://ideas.repec.org/h/eee/ecochp/4-36.html.

W. K. Newey and K. D. West. Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, 61(4):631–53, October 1994. URL http://ideas.repec.org/a/bla/restud/v61y1994i4p631-53.html.

T. A. Rietz. The equity risk premium a solution. *Journal of Monetary Economics*, 22(1):117 – 131, 1988. ISSN 0304-3932. doi: 10.1016/0304-3932(88)90172-9. URL http://www.sciencedirect.com/science/article/pii/0304393288901729.

F. Schorfheide, D. Song, and A. Yaron. Identifying long-run risks: A bayesian mixed-frequency approach. (20303), July 2014. doi: 10.3386/w20303. URL http://www.nber.org/papers/w20303.

J. Stock, M. Yogo, and J. Wright. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics*, 20:518 – 529, 2002.

J. H. Stock and J. H. Wright. Gmm with weak identification. *Econometrica*, 68(5): 1055–1096, 2000. ISSN 1468-0262. doi: 10.1111/1468-0262.00151. URL http://dx.doi.org/10.1111/1468-0262.00151.

J. A. Wachter. Can time-varying risk of rare disasters explain aggregate stock market volatility? *The Journal of Finance*, 68(3):987–1035, 2013. ISSN 1540-6261. doi: 10.1111/jofi.12018. URL http://dx.doi.org/10.1111/jofi.12018.

D. Wilcox. The construction of u.s. consumption data: Some facts and their implications for empirical work. *American Economic Review*, 82(4):922–41, 1992. URL http://EconPapers.repec.org/RePEc:aea:aecrev:v:82:y:1992:i:4:p:922-41.

Table 5: Predictability of excess returns, consumption, and dividends

The predictive regressions in the data show that log price-dividend predicts excess return but does not predict consumption growth and dividend growth. When using the estimated parameters to calibrate the LRR model, we see that the values obtained under basic moment restrictions (RCUGmm) do not allow to reproduce the level of the observed predictability of excess return of excess returns. Furthermore the implied predictability of the dividend growth is highly rejected. If we use the values obtained in the extended restrictions (ECUGmm and EECUGmm) to calibrate the model, the implied predictability becomes closer to the observed one. Furthermore, implied consumption growth predictability is closer to the data for EECUGmm restrictions (where this moment conditions have been included) compare to CUGmm restrictions

Horizon	$\beta(50\%)$ RCU	$R^2(50\%)$ RCU	$\%\left(\widehat{R^2}\right)$ RCU	$\beta(50\%)$ CU	$R^2(50\%)$ CU	$\%\left(\widehat{R^2}\right)$ CU	$\beta(50\%)$ ECU	$R^2(50\%)$ ECU	$\%\left(\widehat{R^2}\right)$ ECU	$\beta(50\%)$ EECU	$R^2(50\%)$ EECU	$\%\left(\widehat{R^2}\right)$ EECU
	Excess return predictability : $\sum_{j=1}^{J} r^a_{m,t+j} - r^a_{f,t+j} = const. + \beta(p^a_t - d^a_t) + \varepsilon_{t+J}$											
1Y	-0.049	0.010	0.924	-0.068	0.018	0.850	-0.123	0.022	0.797	-0.115	0.028	0.7971
	-0.996			-1.331			-1.198			-1.12		
3Y	0.099	0.029	0.976	-0.156	0.047	0.945	-0.309	0.056	0.931	-0.247	0.074	0.900
	1.160			-1.385			-1.915			-1.995		
5Y	-0.268	0.048	0.961	-0.264	0.072	0.928	-0.649	0.076	0.926	-0.618	0.108	0.888
	-1.289			-1.675			-1.335			-1.903		
	Consumption growth predictability : $\sum_{j=1}^{J}(\triangle c^a_{t+j}) = const. + \beta(p^a_t - d^a_t) + \varepsilon_{t+J}$											
1Y	0.008	0.013	0.077	0.053	0.245	0.0016	0.059	0.410	0.0001	0.044	0.325	0.002
	0.882			4.182			4.749			4.164		
3Y	-0.025	0.025	0.042	0.072	0.363	0.0019	0.1248	0.247	0.003	0.134	0.182	0.008
	-1.186			5.856			4.522			4.513		
5Y	0.031	0.037	0.207	0.134	0.400	0.013	0.0684	0.1436	0.063	0.073	0.104	0.104
	1.428			4.179			2.422			2.468		
	Dividend growth predictability : $\sum_{j=1}^{J}(\triangle d^a_{t+j}) = const. + \beta(p^a_t - d^a_t) + \varepsilon_{t+J}$											
1Y	0.084	0.338	0.000	0.127	0.194	0.003	0.231	0.302	0.0005	0.206	0.277	0.002
	5.032			5.132			5.133			4.792		
3Y	0.352	0.432	0.002	0.353	0.241	0.034	0.417	0.190	0.045	0.242	0.159	0.077
	4.967			3.792			3.506			3.019		
5Y	0.384	0.449	0.004	0.310	0.270	0.034	0.433	0.119	0.091	0.348	0.095	0.120
	6.270			4.377			2.164			2.641		

The column named $\hat{\beta}$ represents the slope and (below) the t-stat of the predictive regression of future excess return, consumption and dividend growth on the log price dividend ratio for different horizons (1 year, 3 years and 5 years). The R-square are also provided in columns named $\widehat{R^2}$ and $R^2(50\%)$. The first one is the R-squared obtained for the observed data, while the second one is the median of the R-squared obtained for regressions on 10.000 simulated data set using the corresponding estimates of parameters (under RCUGmm, CUGmm , ECUGmm and EECUGmm restrictions) to calibrate the model. The column named $\%\left(\widehat{R^2}\right)$ represents the percentage of sample simulated R-squared that are below the R-squared observed in the data; a percentile below 5% or above 95 % correspond to a one side test rejection of the model at 5% significant level.

Table 6: Basic moments implied by the model

Moments	Data	RCUGmm	CUGmm	ECUGmm	EECUGmm
$E(\Delta c)$	0.020	0.023 [0.010; 0.036]	0.025 [-0.016; 0.066]	0.0229 [0.008; 0.037]	0.0229 [0.008 ; 0.038]
$\sigma(\Delta c)$	0.03	0.044 [0.035 ; 0.053]	0.043 [0.031 ; 0.059]	0.030 [0.019 ;0.045]	0.032 [0.020 ; 0.049]
$AC1(\Delta c)$	0.359	0.233 [0.018; 0.429]	0.414 [0.133; 0.678]	0.713 [0.555 ; 0.834]	0.694 [0.535 ; 0.819]
$E(\Delta d)$	0.008	0.024 [-0.073 ;0.121]	0.029 [-0.085 ;0.142]	0.0343 [-0.029 ; 0.097]	0.030 [-0.023 ; 0.082]
$\sigma(\Delta d)$	0.110	0.126 [0.098; 0.163]	0.1640 [0.120; 0.219]	0.164 [0.107 ; 0.241]	0.134 [0.086 ; 0.202]
$AC1(\Delta d)$	0.146	0.427 [0.172 ; 0.666]	0.3190 [0.062 ; 0.561]	0.471 [0.246; 0.664]	0.509 [0.298 ; 0.692]
$E(r_m)$	0.075	0.059 [-0.015; 0.134]	0.066 [-0.025 ; 0.157]	0.073 [0.012 ; 0.140]	0.065 [0.013 ; 0.121]
$\sigma(r_m)$	0.200	0.333 [0.272 ;0.400]	0.326 [0.246 ;0.421]	0.290 [0.205 ; 0.397]	0.243 [0.174 ; 0.334]
$AC1(r_m)$	-0.003	-0.015 [-0.245 ;0.214]	-0.024 [-0.269 ;0.221]	0.009 [-0.227 ;0.250]	0.020 [-0.218 ;0.262]
$E(r_f)$	0.005	-0.0255 [-0.032; -0.02]	0.002 [-0.032; 0.033]	0.009 [-0.0019 ; 0.0183]	0.0115 [0.0004 ; 0.0215]
$\sigma(r_f)$	0.039	0.021 [0.0160 ;0.026]	0.0197 [0.011 ;0.034]	0.021 [0.013 ; 0.031]	0.0232 [0.014 ; 0.036]
$AC1(r_f)$	0.563	0.322 [0.093; 0.530]	0.906 [0.780; 0.977]	0.683 [0.511 ; 0.818]	0.650 [0.474 ; 0.791]
$E(p-d)$	3.320	3.364 [2.403;4.339]	3.298 [2.194;4.353]	3.204 [2.808 ; 3.449]	3.274 [2.832 ; 3.524]
$\sigma(p-d)$	0.441	0.715 [0.423 ; 1.155]	0.641 [0.350 ; 1.114]	0.371 [0.218 ; 0.593]	0.336 [0.192 ; 0.554]
$AC1(p-d)$	0.936	0.873 [0.713; 0.958]	0.857 [0.659; 0.960]	0.691 [0.466 ; 0.855]	0.727 [0.501 ; 0.883]

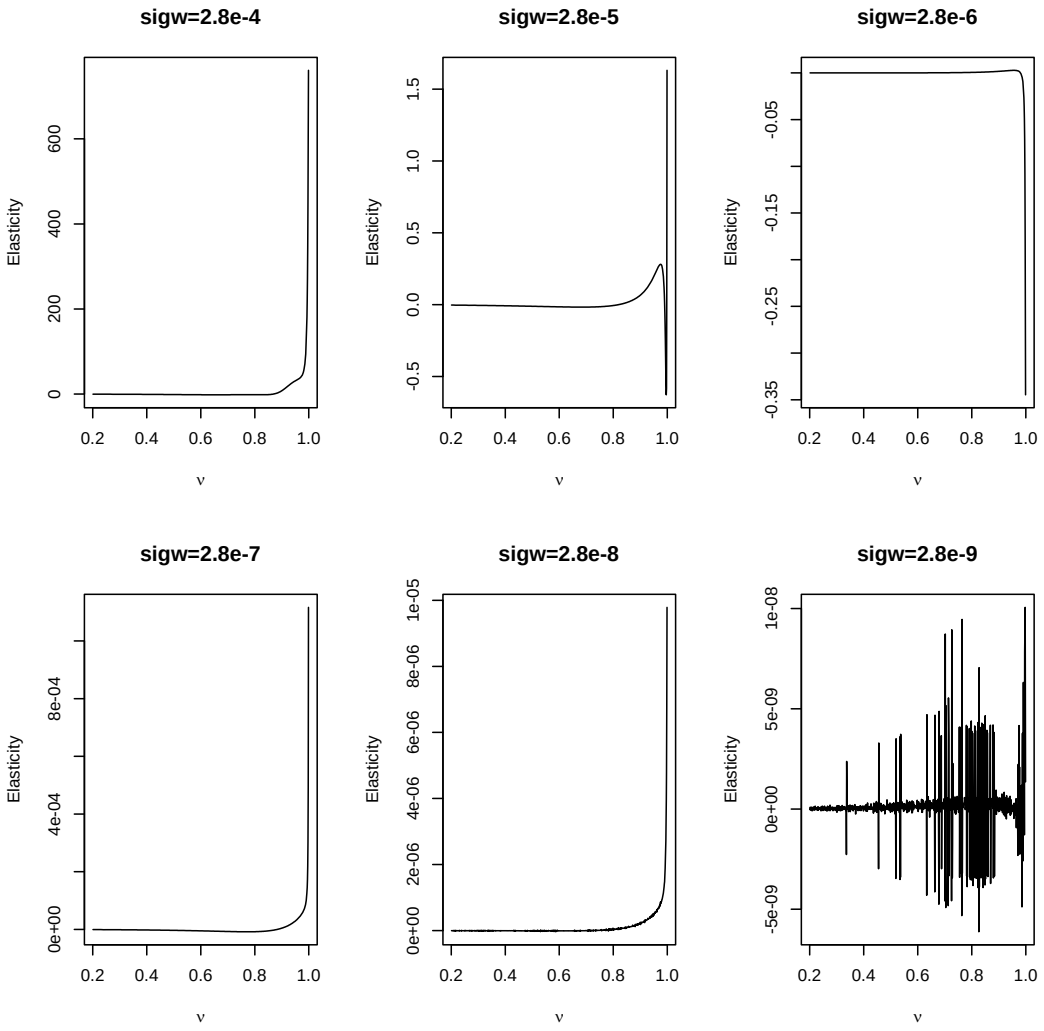
The table displays the results from 10.000 simulations of 948 months each (to mimic the period 1930-2009) of the long run risks model based on the CUE estimates. The statistics are for annually aggregated variables (consumption growth rate, dividend growth rate, market return, risk free rate, log price-dividend ratio). Aggregation has been done through a 12 months summation for growth rates and returns. For the log price-dividend ratio, the annual values have been computed using the year's December values of the price divided by the summation of monthly dividends within the year.

Table 7: Predictability of the volatility of excess returns, consumption, and dividends

Horizon	$\beta(50\%)$ RCU	$R^2(50\%)$ RCU	$\%\left(\widehat{R^2}\right)$ RCU	$\beta(50\%)$ CU	$R^2(50\%)$ CU	$\%\left(\widehat{R^2}\right)$ CU	$\beta(50\%)$ ECU	$R^2(50\%)$ ECU	$\%\left(\widehat{R^2}\right)$ ECU	$\beta(50\%)$ EECU	$R^2(50\%)$ EECU	$\%\left(\widehat{R^2}\right)$ EECU
	Excess return volatility predictability											
1Y	-0.0248	0.006	0.362	-0.033	0.012	0.2712	-0.114	0.032	0.134	-0.079	0.035	0.123
	-0.952			-1.153			-1.709			-1.420		
3Y	-0.046	0.018	0.069	0.090	0.027	0.0585	-0.266	0.061	0.034	-0.125	0.069	0.029
	-0.769			0.822			-1.633			-1.760		
5Y	-0.128	0.029	0.060	0.107	0.037	0.0531	-0.494	0.068	0.036	-0.288	0.082	0.034
	-0.938			1.119			-1.580			-1.825		
	Consumption growth volatility predictability											
1Y	-0.002	0.007	0.9960	-0.003	0.013	0.9656	-0.009	0.061	0.773	-0.013	0.076	0.701
	-0.843			-0.853			-2.344			-2.648		
3Y	-0.010	0.018	0.998	0.010	0.028	0.9852	-0.0296	0.106	0.908	-0.025	0.140	0.861
	-0.873			1.321			-1.800			-3.240		
5Y	0.015	0.029	0.996	-0.019	0.038	0.9859	-0.0384	0.113	0.943	-0.052	0.152	0.906
	1.080			-1.107			-1.692			-3.178		
	Dividend growth volatility predictability											
1Y	-0.012	0.007	0.753	-0.018	0.013	0.5960	-0.074	0.057	0.224	-0.048	0.074	0.162
	-0.968			-0.842			-1.840			-3.279		
3Y	0.019	0.018	0.623	0.0281	0.028	0.5202	-0.197	0.101	0.215	-0.144	0.137	0.151
	1.045			1.056			-2.303			-2.694		
5Y	-0.032	0.029	0.651	-0.064	0.039	0.5752	-0.252	0.108	0.325	-0.164	0.149	0.239
	-1.030			-1.032			-1.815			-3.939		

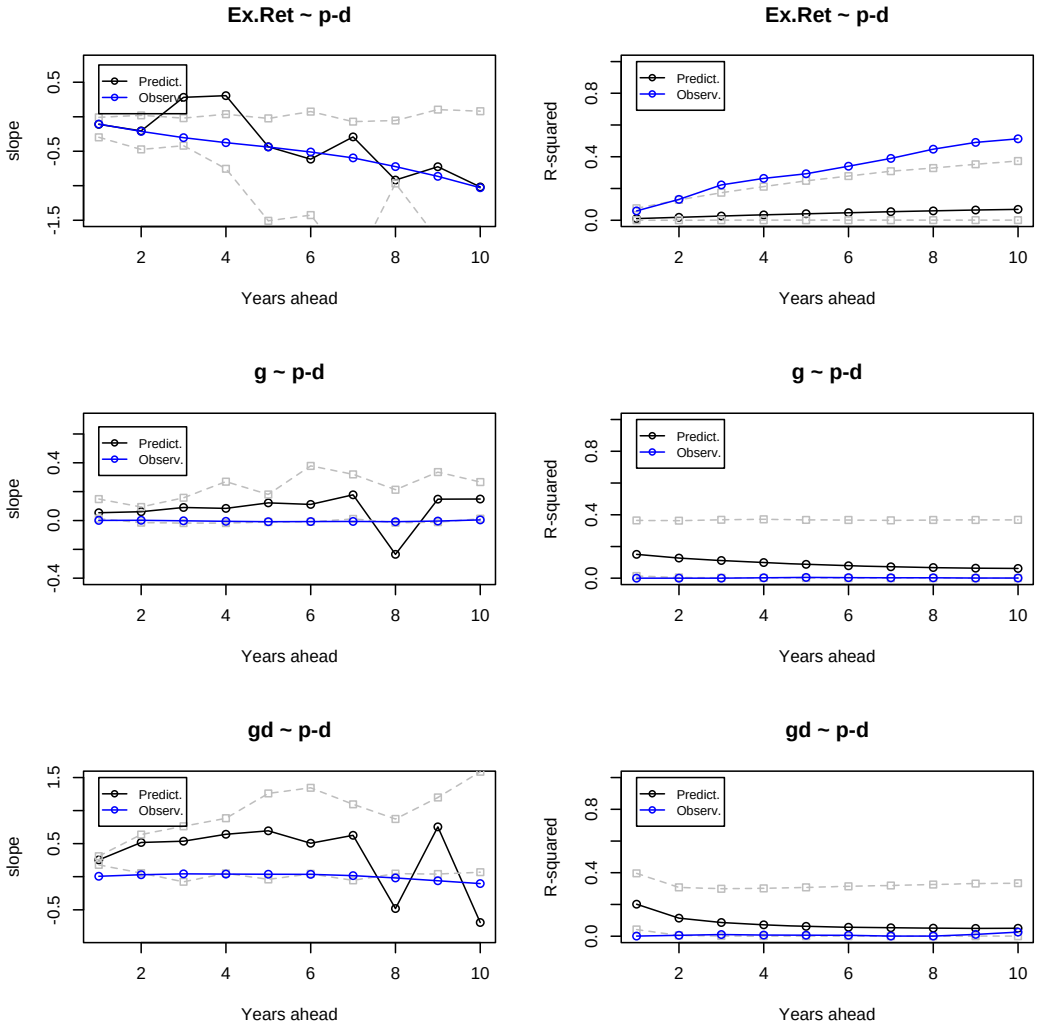
The table displays the results of model predictive regressions of future excess return volatility, consumption growth volatility and dividend growth volatility on the log price-dividend ratio as in eq. 11. We run 10.000 simulations of 948 months each (to mimic the period 1930-2009) of the long run risks model when calibrated at the parameters obtained from the estimations. Each column represents the slope and below the t-stat of the predictive regression for different horizons (1 year, 3 years and 5 years). The median R-square are also provided. The last column represents the percentage of sample simulated R-squared that are below the R-squared observed in the data; a percentile below 5% or above 95 % correspond to a one side test rejection of the model at 5% significant level.

Figure 2: Illustration of weak identification of ν driven by σ_w



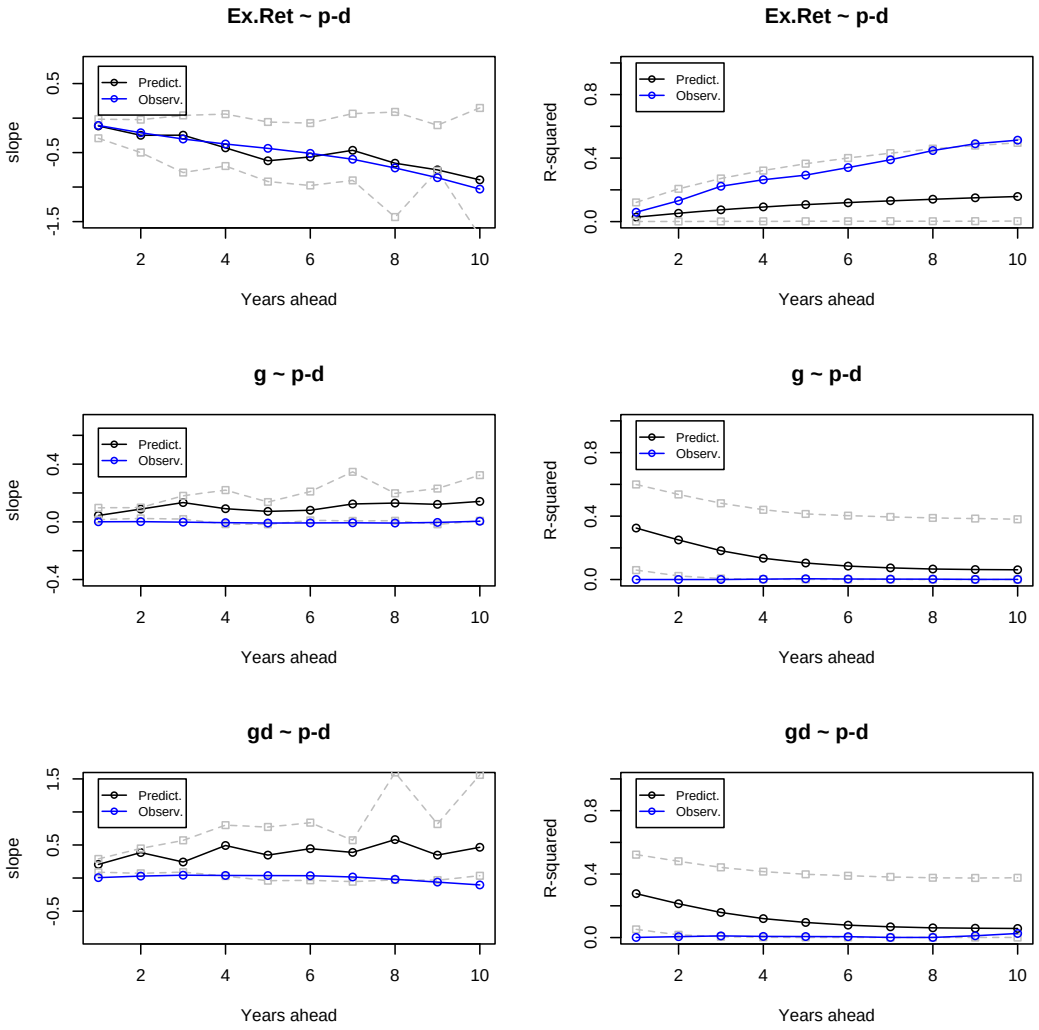
This figure shows the elasticity of the continuously updated Gmm function with respect to the persistence of the volatility (ν) as the volatility of the volatility (σ_w) goes to 0. As σ_w goes to zero, the objective function becomes almost insensitive to changes in ν. All the parameters (except ν and σ_w) have been calibrated as in BYK for the simulation and the sample length is 840 years.

Figure 3: Predictability in the BYK calibration of the LRR model



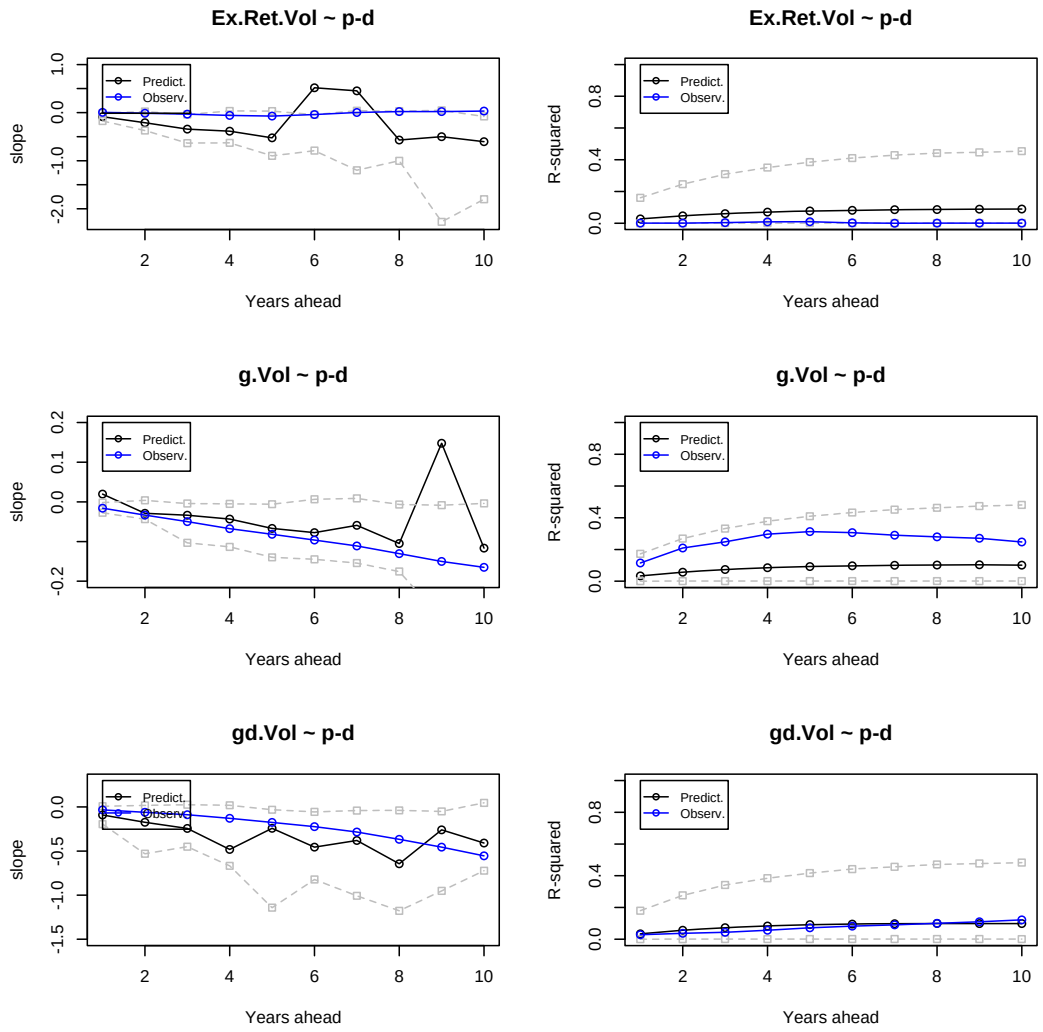
This figure shows the slopes (on the left) and the R-squared (on the right) of the predictive regression of future excess returns (Ex.Ret), consumption growth (g) and dividend growth (gd) by log price-dividend ratio. The horizontal axis represents the number of ear ahead. The blue line correspond to the observed data (the sample is from 1930 to 2009). The black lines correspond to simulated data using the BYK calibration. The grey lines correspond to the confidence intervals of the black line. We run 10.000 simulations of sample with the same length as the observed data.

Figure 4: Predictability in the EECUGmm calibration of the LRR model



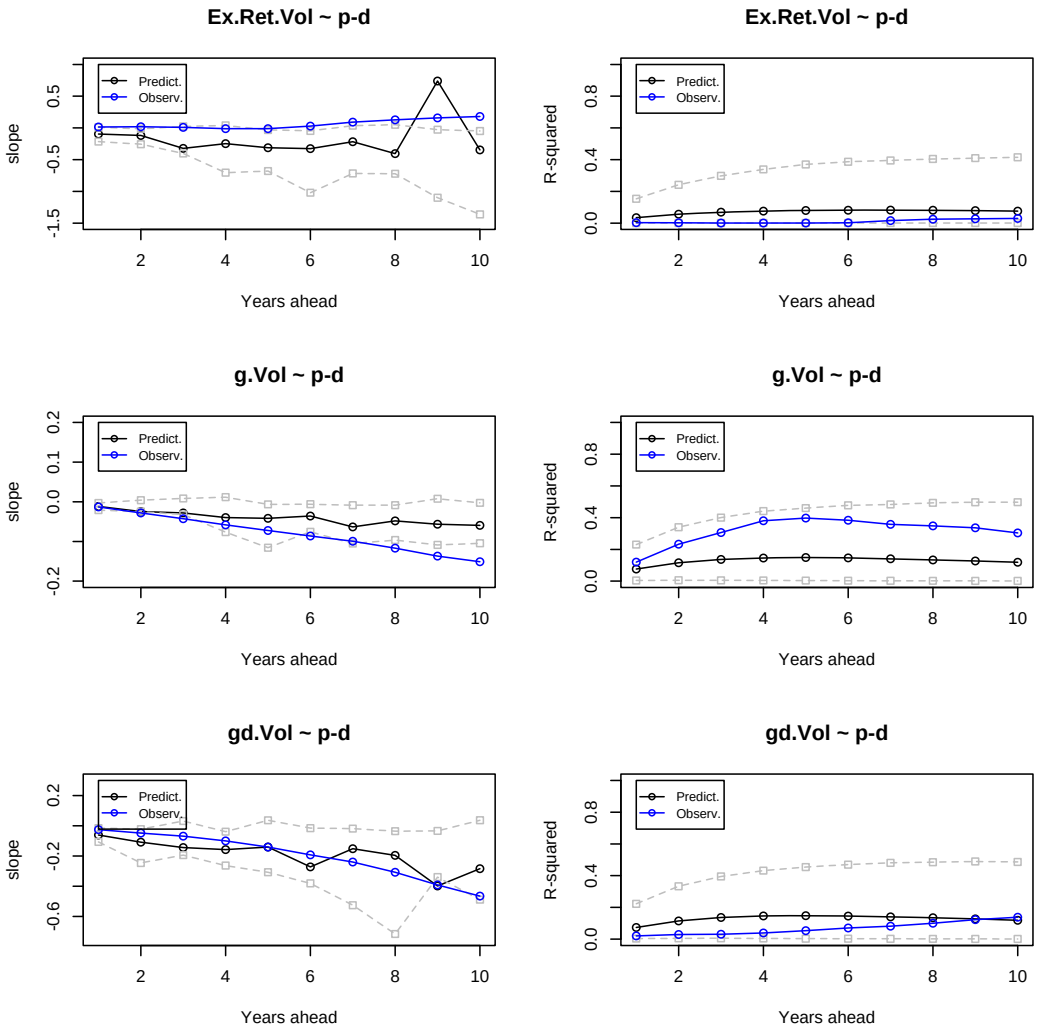
This figure shows the slopes (on the left) and the R-squared (on the right) of the predictive regression of future excess returns (Ex.Ret), consumption growth (g) and dividend growth (gd) by log price-dividend ratio. The horizontal axis represents the number of ear ahead. The blue line correspond to the observed data (the sample is from 1930 to 2009). The black lines correspond to simulated data using the estimates obtained from EEUGmm setup for calibration. The grey lines correspond to the confidence intervals of the black line. We run 10.000 simulations of sample with the same length as the observed data.

Figure 5: Volatility predictability in the BYK calibration of the LRR model



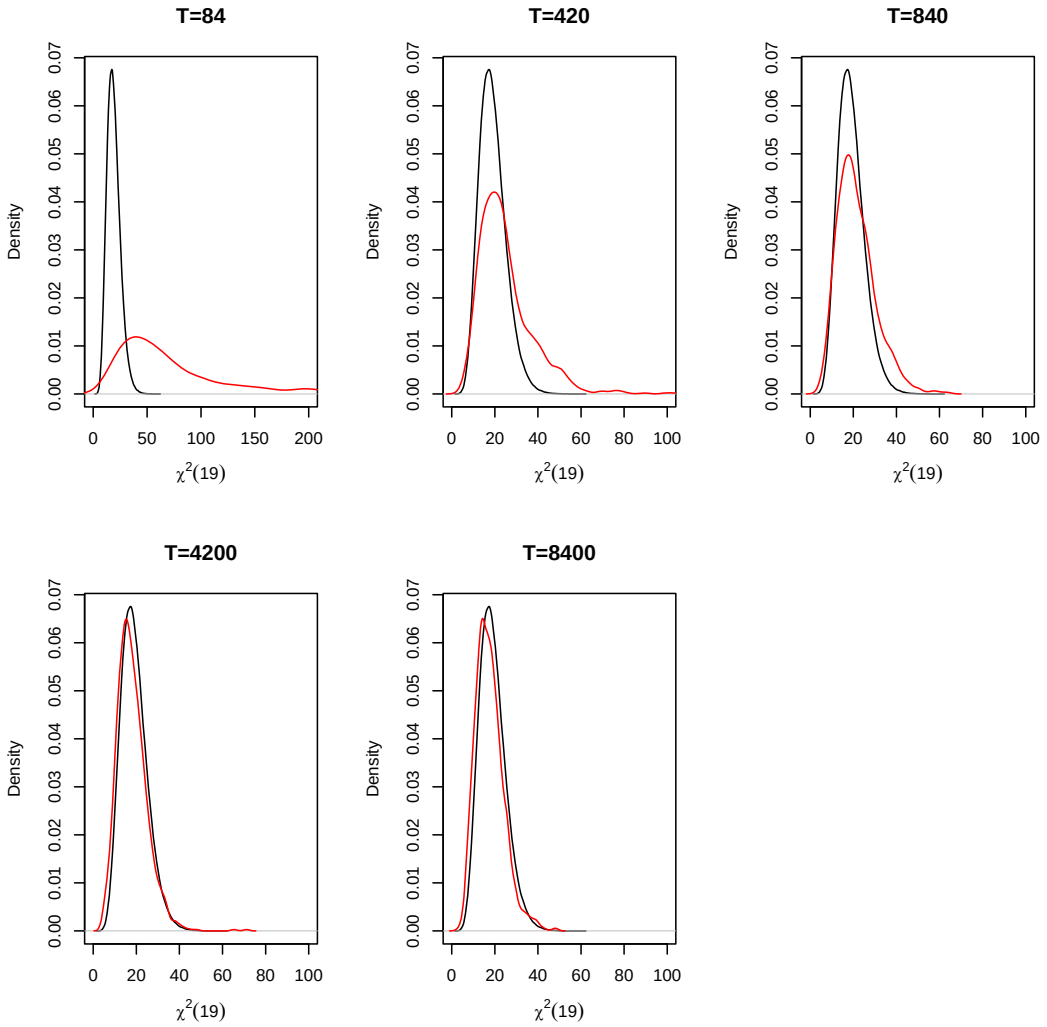
This figure shows the slopes (on the left) and the R-squared (on the right) of predictive regression of future excess returns volatility (Ex.Ret.Vol), consumption growth volatility (g.Vol) and dividend growth volatility (gd.Vol) by log price-dividend ratio.The horizontal axis represents the number of ear ahead. The blue lines correspond to the observed data (the sample is from 1930 to 2009). The black lines correspond to simulated data using the BYK calibration. The grey lines correspond to the confidence intervals of the black line. We run 10.000 simulations of sample with the same length as the observed data.

Figure 6: Volatility predictability in the EECUGmm calibration of the LRR model



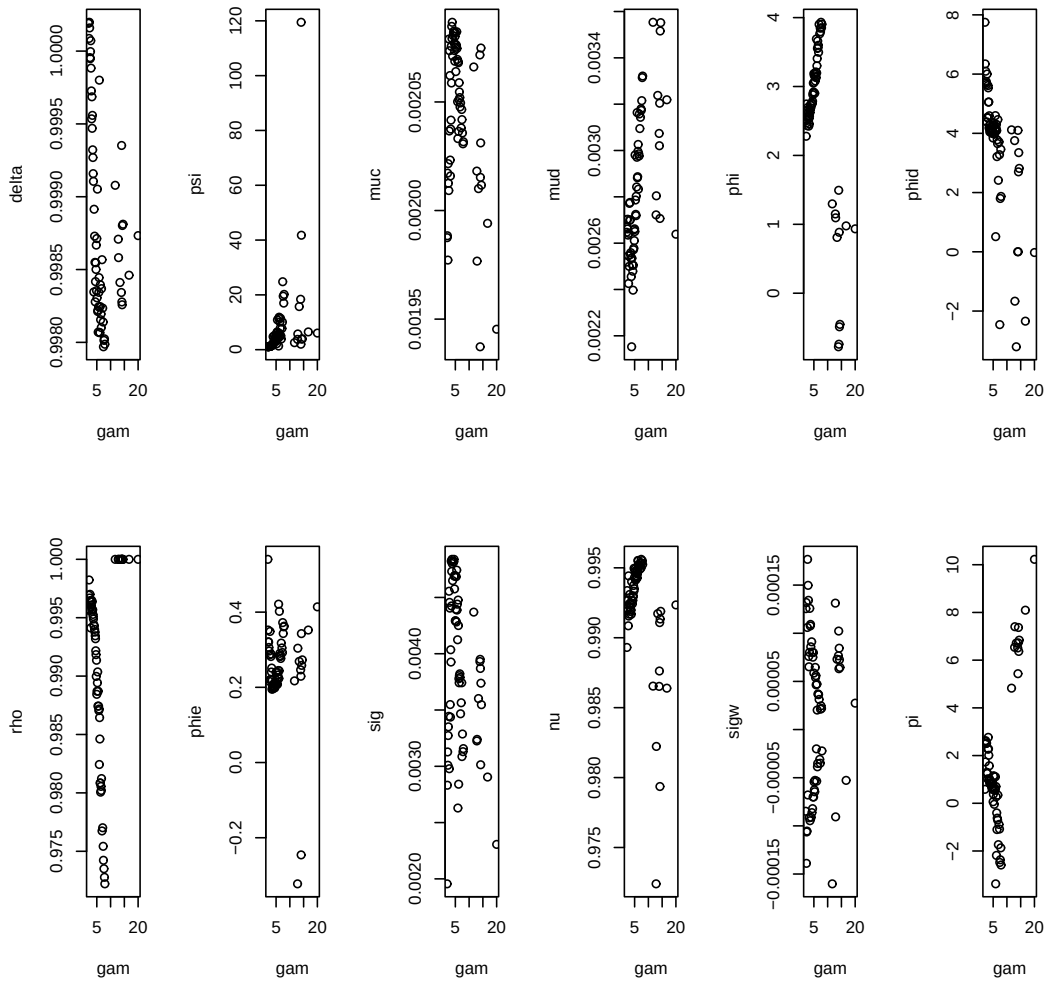
This figure shows the slopes (on the left) and the R-squared (on the right) of the predictive regression of future excess returns volatility (Ex.Ret.Vol), consumption growth volatility (g.Vol) and dividend growth volatility (gd.Vol) by log price-dividend ratio. The horizontal axis represents the number of ear ahead. The blue lines correspond to the observed data (the sample is from 1930 to 2009). The black lines correspond to simulated data using the estimates obtained from EEUGmm setup for calibration. The grey lines correspond to the confidence intervals of the black line. We run 10.000 simulations of sample with the same length as the observed data.

Figure 7: Finite sample distribution of the CUGmm objective function in the Model



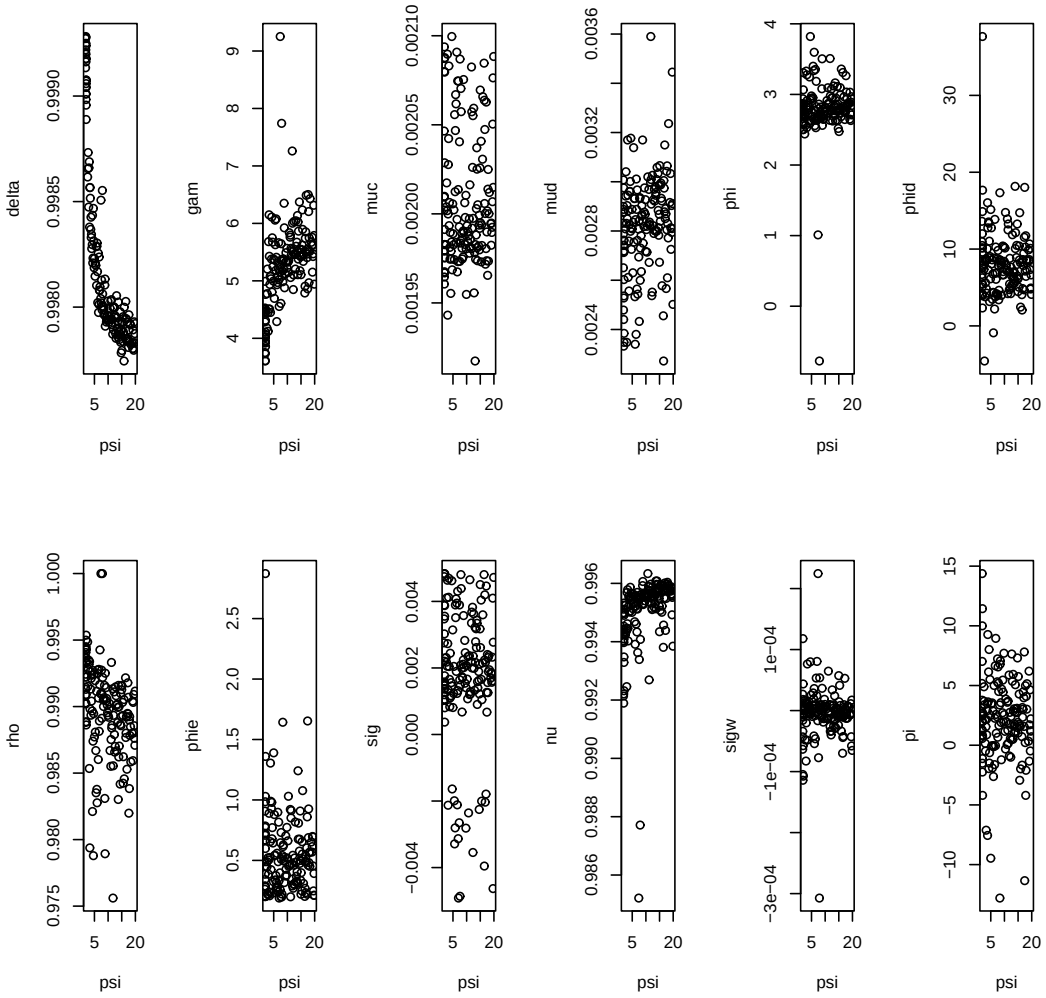
This figure shows the finite sample distribution of the objective function evaluated at the true value of the parameter for the model simulated with various sample sizes. The number of simulation used to compute the kernel densities (red plots) is the same for all the plots (N=1000), but the sample size of the draws varies: T (in years) =84, 420, 840, 4200, 8400. The black lines represent the density of a chi-squared random variable with 19 degrees of freedom (which correspond to the number of moment conditions). When the sample size is small (e.g: T=84) as what we have in the data, the finite sample approximation of the asymptotic distribution of the CUGmm objective is very poor and that will lead to an over-rejection of the model.

Figure 8: 95% Confidence region for γ with respect to the other parameters



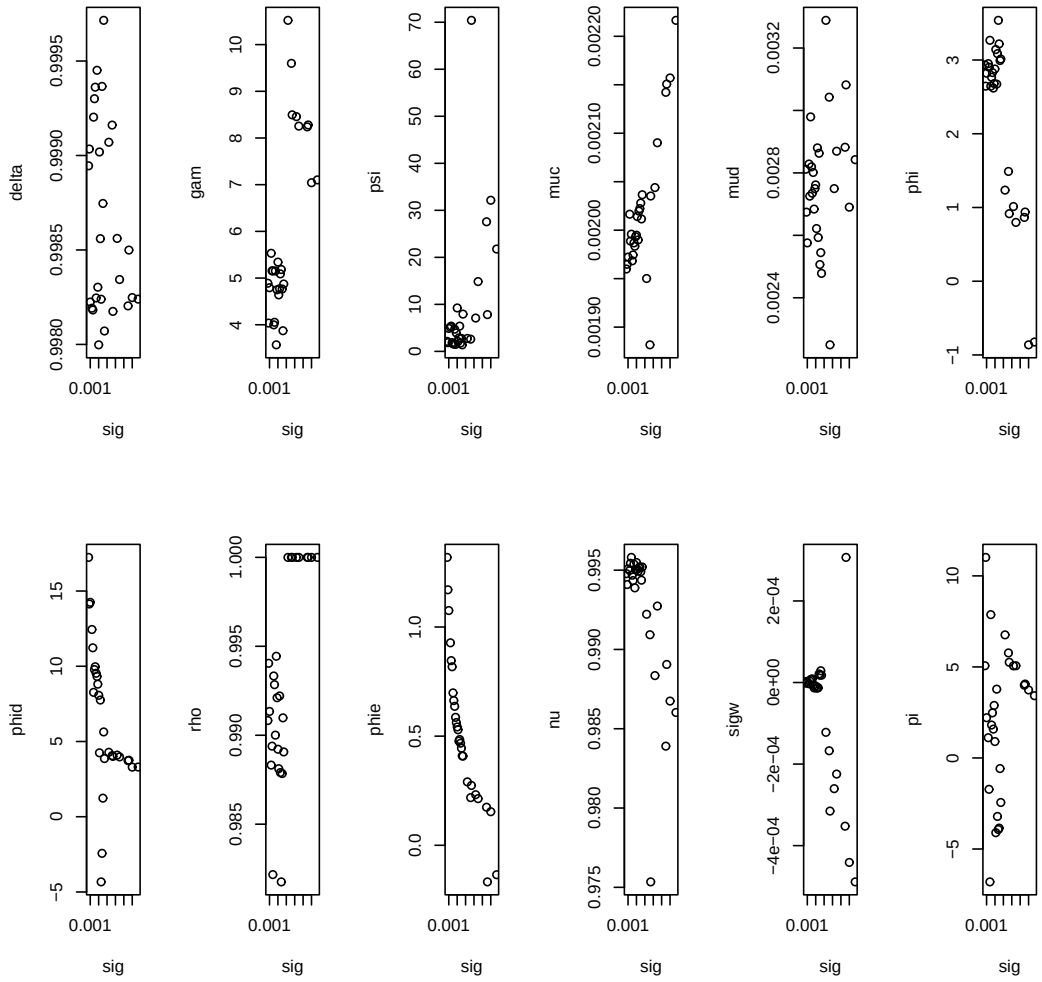
This plot shows the joint 95 % confidence region for γ and others parameters. It has been constructed by selecting all the vectors of parameters for which the CUGmm objective function value is below the 95 % quantile of $\chi^2(7)$. In the CUGmm objective function, we have 19 moment conditions, γ is fixed here and the remaining parameters are estimated.

Figure 9: Confidence region for ψ with respect to the other parameters



This plot shows the joint 95 % confidence region for ψ and others parameters. It has been constructed by selecting all the vectors of parameters for which the CUGmm objective function value is below the 95 % quantile of $\chi^2(7)$. In the CUGmm objective function, we have 19 moment conditions, ψ is fixed here and the remaining parameters are estimated.

Figure 10: Confidence region for σ with respect to the other parameters



This plot shows the joint 95 % confidence region for σ and others parameters. It has been constructed by selecting all the vectors of parameters for which the CUGmm objective function value is below the 95 % quantile of $\chi^2(7)$. In the CUGmm objective function, we have 19 moment conditions, σ is fixed here and the remaining parameters are estimated.