POOS Experiment: This appendix contains the more detailed discussion of the POOS experiment that was contained in the working paper version of the paper:

6.4 A pseudo-out-of-sample forecasting experiment

Typically, pseudo-out-of-sample (POOS) forecasting experiments are of limited use for evaluating long-horizon forecasts because of the limited number of independent long-horizon POOS time-series observations. However, in our context each of the n = 113 countries provides some independent POOS information about the validity of the predictive distribution. We have carried out a POOS experiment that focuses on this cross-sectional information.

Specifically, in the first experiment we estimated the complete model through time T_1 = 1977, and computed joint predictive distributions for the average growth rate of f_t and $y_{i,t}$ for each of the 113 countries over the subsequent h = 20, 30 and 40 years. The realized values of $y_{i,t}$ are known over these POOS forecast periods; moreover, the realized value of f_t is well approximated by full-sample estimates $f_{t|T}$ (see Figure 6a). Thus, $c_{i,t|T} = y_{i,t} - f_{t|T}$ provides an accurate estimate for the POOS out-of-sample realized value of $c_{i,t}$. We therefore used $f_{t|T}$ and $c_{i,t|T}$ to evaluate the POOS predictive distributions. Specifically, as is standard for evaluating predictive distributions (see Diebold, Gunther and Tsay (1998)), sample values of the predictive distributions probability integral transform (*PIT*s) were computed by evaluating the predictive distributions at the realized POOS values of $f_{t|T}$ and $c_{i,t|T}$. Recall that for a correctly specified predictive distribution, the sample values of the *PIT* are distributed as a U(0,1) random variables.

Table A1 summarizes the resulting *PIT*s for the experiment, and two other experiments use T_1 = 1987 (with a forecast horizon h = 20 and 30 years) and T_1 = 1997 (with a forecast horizon of h = 20 years). Table 7 summarizes the results. These result in six forecasts for f_t and with *PIT* values shown in the first column of the table. This is a very small sample of dependent observations, but the *PIT*s provide no evidence of misspecification in the predictive distributions for f_t.

There are 113 forecasts $c_{i,t}$ for each POOS experiment and forecast horizon, so these forecasts are more informative about their predictive distributions. The table summarizes the $c_{i,t}$ forecasts by showing the fraction of *PIT*s in each of the quartiles or the U(0,1) distribution: ideally, 25% of the *PIT*s should be less than 0.25, 25% should be between 0.25 and 0.50, and so forth. The table also shows the Kolmogorov-Smirnov (*KS*) statistics testing the null hypothesis that the actual values were draws from the predictive distributions (i.e., that the *PIT*s were U(0,1) draws). The *p*-values for the *KS* statistic were computed treating the posterior predictive distribution as the truth, which incorporates the model-implied cross-sectional dependence in $c_{i,t}$.

The results suggest that the predictive distributions for T_1 = 1977 were somewhat too optimistic: roughly half of the realized values of $c_{i,t}$ lie in the lower quartile of the predictive distributions. The predictive distributions for T_1 = 1987 and T_1 = 1997 seem to be reasonably well-calibrated.

Table A1. Summary of probability integral transforms for pseudo-out-of-sample forecasts

Forecast Horizon	**PITs**						
	f		**Fraction of *c*-PITs between**				
			0 – 0.25	**0.25-0.50**	**0.50-0.75**	**0.75-1.00**	***KS-c* (*p-value*)**
(a) T_1 = 1977							
20 years	0.41		0.57	0.14	0.18	0.11	0.36 (0.01)
30 years	0.38		0.50	0.15	0.19	0.16	0.28 (0.06)
40 years	0.30		0.46	0.17	0.17	0.20	0.23 (0.14)
(b) T_1 = 1987							
20 years	0.30		0.30	0.31	0.22	0.17	0.15 (0.42)
30 years	0.20		0.27	0.26	0.28	0.19	0.09 (0.83)
(c) T_1 = 1997							
20 years	0.24		0.12	0.29	0.29	0.29	0.14 (0.47)