Reply to Dai, Singleton, and Yang (draft dated June 6, 2004)

John H. Cochrane and Monika Piazzesi

University of Chicago

October 18, 2004

In a recent working paper, titled "Predictability of Bond Risk Premia and Affine Term Structure Models" Dai, Singleton, and Yang (June 6 2004) (DSY) forcefully challenge the analysis and several of the conclusions of our paper "Bond Risk Premia." Here, we respond. Many of their criticisms (points 1–4 below) were already addressed in our 2002 NBER working paper, as well as in our final 2004 draft. We show where. DSY also claim (point 5 below) that our main results are driven by measurement error, and they advocate a different interpolation scheme to produce zero-coupon yields from treasury data, which they claim avoids this measurement error. We show that this claim is not correct.

1. The single-factor model for expected excess returns on bonds of all maturities is statistically rejected.

We agree. Cochrane and Piazzesi (2002, NBER Working paper) stated "The tests all reject the single factor model" on p. 31. Table 12 documented and interpreted additional factors. Cochrane and Piazzesi (2004) discussed the rejection on p. 29 and presented the evidence in Table 6.

2. There are buy-and-hold portfolios with returns that are orthogonal to our forecasting factor and that have a notably large Sharpe ratio.

We agree. Table 12 of Cochrane and Piazzesi (2002, NBER Working paper) and Table 7 in Cochrane and Piazzesi (2004) documented such strategies.

The point of our paper is that a single factor captures the vast majority of variation in expected returns – 99.4% in a factor decomposition–not that it captures *all* such variation. Both drafts of the paper pointed out that there are small idiosyncratic and transitory movements in yields. These very small spreads thus forecast small return differences in ways that violate the single-factor model, with high R^2 values, leading to the appearance of small profitable spread trades, and statistical rejection of the single factor model.

3. The evidence for predictability of excess returns is to a large extent symptom of small-sample biases in estimated R^2.

We disagree. Of course it is not news that R^2 statistics are biased estimates of their population counterparts. This is why we performed an extensive Monte Carlo analysis and computed small-sample confidence bounds for our R^2 measures. These are the "Small-T" intervals in Table 1 and 2 of the 2002 NBER working paper version and Table 1 of the 2004 version. We also clearly state that you test models with χ^2 statistics not R^2 measures. These tests confirm that the predictability we document is significantly stronger than the predictability documented by Fama & Bliss (1987) and significantly stronger than what we get by using the first three principal components of yields. (Tables 1,2, 3 of the 2004 draft, for example.)

4. The predictability documented in our descriptive regression analysis is no greater than the predictability evidenced in three-factor Gaussian term structure models.

We disagree. Section II.A of our paper documented that 3 factors are not enough. This conclusion is not based on increased R^2, (the fact that our factor gives an R^2 of 0.35, while level slope and curvature give 0.28) but done properly, based on χ^2 tests for parameter exclusion reported in Table 4. We clearly rejected the null hypothesis that factors in addition to the first three principal components do not help in forecasting excess returns. DSY do not comment on these rejections.

Coefficient plots in Figure 2 compared the coefficients from a regression on level, slope, and curvature with those obtained from unrestricted regressions on all yields (or, equivalently, forward rates). These plots show the deficiencies of three-factor models.

We do not disagree with affine models as a class of course. In fact Section II.B of our paper writes down a 5-factor Gaussian affine model that generates our regression results exactly. But three factors are not enough to capture yields *and* return forecastability. It is quite possible that three factors capture the vast bulk of variation in the *level* of prices, but do not suffice to describe *expected returns*, because excess returns are based on differences of differences of prices.

5. The predictability and the tent-shaped coefficient patterns result from measurement error introduced by the data construction.

We disagree. Measurement error is always a danger in predictability regressions, since the price at time t is common to left and right hand sides. However, as pointed out in our paper, a large number of observations suggest that the central results – a tent-shaped single factor–are not the result of this effect. DSY do not address any of these observations. Among others,

1. The forecasts work quite well with lagged right hand variables (Figure 3), in which case the same p_t is not on both sides of the regressions.
2. The bond return forecasting factor forecasts stock returns (Table 3), where there is no common price, and with about the right coefficient.
3. Measurement error does not deliver a single-factor structure. Even if the measurement error is correlated across bonds, it delivers a pattern that the n year yield forecasts the n year return (Figure 4).

Specifically, DSL advocate different interpolation schemes to recover zero coupon bond yields from treasury yields. They use smoothed Fama and Bliss (SFB) data, in contrast to our unsmoothed Fama and Bliss (UFB) data from CRSP. They interpret the UFB data as the "true" SFB data plus measurement error, and they view our results as figments of the measurement error. The tent-shaped pattern of coefficients we find in the upper left hand panel of figure 1 disappears in the SFB data shown in the upper right hand panel. This is DSL's main point. (Note that the single factor is still there – the curves all have the same shape – it's just a different shape.)

The bottom panels of Figure 1 run cross-regressions – UFB returns on SFB forwards and SFB returns on UFB forwards. The tent is back! Running the regression without error of *either* left or right hand variable produces an estimate uncontaminated by error, since the "mismeasured" p_t is no longer common to left and right hand sides, so these regressions deny that the tent is a figment of measurement error in the UFB data. Table 1 reports the R^2 from these regressions. The R^2 is the same for cross-regressions as it is for the CP regression. In particular, if the story were measurement error, UFB forwards should not predict SFB returns with undiminished R^2.

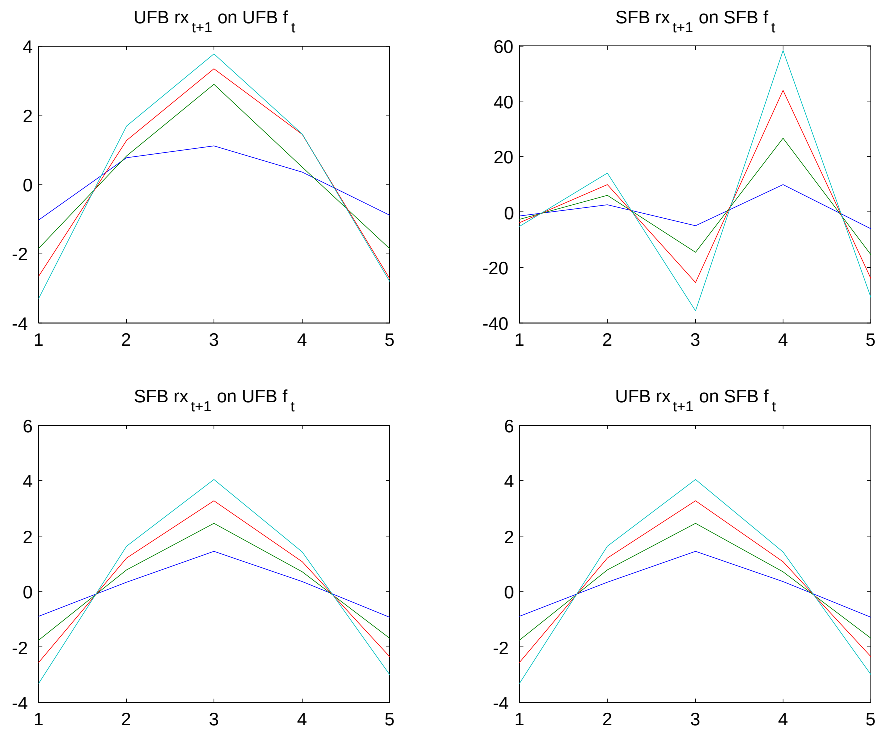

Figure 1: Cross-regressions of UFB excess returns on SFB forwards and SFB returns on UFB forwards.

Table 1: Cross-Regression R^2s

	maturity			
	2	3	4	5
(CP) UFB on UFB	0.35	0.37	0.39	0.36
(DSY) SFB on SFB	0.30	0.31	0.31	0.32
SFB on UFB	0.35	0.35	0.36	0.36
UFB on SFB	0.35	0.35	0.36	0.36

Well, so much for measurement error, but what *should* we make of the wavy line in top right panel of Figure 1? Strong alternating + and − coefficients are a standard signal of multicollinearity, which is exactly what smoothing across maturities induces. Following this idea, Figure 2 presents regressions that drop one variable at a time, a standard cure for multicollinearity. We have the tent again – *even in DSL's SFB on SFB regressions*. This calculation denies the claim that there is a difference in shape between SFB and UFB forecasts *at all*.

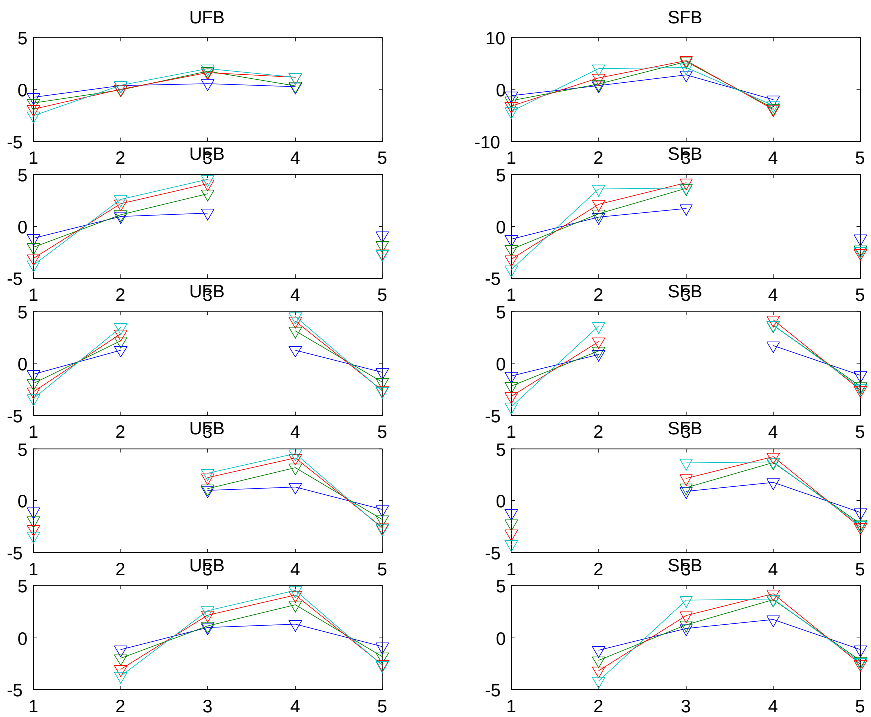

Figure 2: Multicollinearity in the smoothed Fama and Bliss (SFB) data. Each regression drops one maturity from the right hand side, running excess returns on four of the five available forward rates.

Consistent with the multicollinearity interpretation, when DSL add noise in their Figure 2 panel d, the recover something more like the tent shape. This change amounts to adding a diagonal element to the $X^\top X$ matrix. This is another standard "cure" for multicollinearity. Panel d thus also confirms to us that the smoothed data are so utterly multicollinear that the coefficients in c are meaningless.

In sum, we suspect that DSL's smoothing procedure throws out most of the baby with the bathwater. Certainly, if you smooth across maturities by fitting a three factor model, then you *cannot* find the extra forecastability due to a fourth factor, which is the heart of our paper. Their smoothing procedure seems to have come close to this result.