Constant-Duration Bond Portfolios' Initial (Rolling) Yield Forecasts Return Best at Twice Duration

Gabriel A. Lozada*

March 2016

Abstract. Leibowitz and co-authors showed that with yield paths linear in time, a constant-duration bond portfolio's initial yield equals its mean return at twice duration. We clarify and extend this result to continuously/periodically-compounded yields and arithmetic/geometric mean returns. The difference between initial yield and mean return at twice duration depends on: the concavity/convexity, but not the slope, of the time path of yield; numerical approximations' errors; and changes in the instantaneous yield curve's slope. The first of these explains forecast errors well for a set of bonds over six decades, but not nine.

Keywords: Initial Yield and Bond Portfolio Returns, Forecasting Bond Returns, Constant-Duration Bond Funds, Constant-Maturity Bond Funds
JEL Codes: G12, G17

*Associate Professor, Department of Economics, University of Utah, Salt Lake City, UT 84112, USA; lozada@economics.utah.edu; (801) 581-7650. Many thanks to Richard Fowles and the participants of the Colloquium for the Advancement of Knowledge in Economics at the University of Utah.

A conclusion of literature stemming from Hicks (1939 p. 186), Samuelson (1945 fn. 1 and p. 23), Macauley (1938 p. 48), and Redington (1952 p. 290) is that when a default-free bond is bought and held for a period of time known as its duration, it will earn approximately its initial yield-to-maturity, and thus it constitutes over that horizon a negligible source of risk despite its price's short-run volatility. However, many institutions and many individuals saving for retirement using mutual funds do not engage in a "buy and hold" strategy for bonds, but rather have bonds which are regularly turned over to maintain an approximately constant maturity or constant duration. Some authors have asserted that even in this situation, bonds are still less risky than their short-run volatility suggests, and that their return over some relatively long period will be close to their initial yield. The most common positions are that this "relatively long period" is the bonds' maturity or its duration. Potts and Reichenstein (2004) show that cumulative return of a constant-maturity portfolio gets close to that predicted by initial yield at roughly the bonds' maturity; a similar assertion is made by John C. Bogle (founder of the world's second-largest asset management firm, Vanguard) and others in Gay (2014). William McNabb, Vanguard's current CEO, uses duration (which is reported on most mutual fund sponsors' web sites):

> There is a silver lining to rising [interest] rates. If your time horizon is longer than the duration of the bond funds you are invested in, you actually want interest rates to rise. [McNabb 2014]

By contrast, the early paper of Langeteig, Leibowitz and Kogelman shows that "if interest rates follow a random walk without drift or reversion" then the risk-minimizing holding period is twice the duration (1990 p. 43). The link between twice duration and initial yield was first made in Leibowitz and Bova (2012) and in "Part I: Duration Targeting: A New Look at Bond Portfolios" of Leibowitz, Bova, Kogelman, and Homer (2013, henceforth LBKH); see Leibowitz, Bova, and Kogelman (2014) for a summary of both the theoretical and empirical arguments, and Bova (2013, pp. 4–8) and Leibowitz and Bova (2013) for empirical support.[1]

In Section 1 we rigorously extend the framework of Leibowitz and his collaborators ("Leibowitz et al.") to coupon bonds and give two new proofs (one quite elementary) of the basic result that if one linearizes most of the mathematical relationships including the path of yield through time, then a constant-duration bond portfolio's mean return at twice duration equals its initial yield. Section 2 proves that, at twice duration, mean return minus initial yield will tend to be negative if the path of yield through time is convex and positive if it is concave. The basic result concerns the *arithmetic* mean return, so it pertains to *continuously* compounded yields and returns; Section 2 derives initial-yield-versus-"mean"-return results for periodic compounding and its appropriate mean, the geometric. However, those results involve equations which cannot be solved analytically for the "initial-yield-equaling amount of time," and that amount of time

[1] Fridson and Xu (2014) point out that junk bonds' long-term return will fall short of their initial yield. Thomas and Bosse (2014) explain why initial yield is a poor predictor of foreign bonds' return if the foreign currency exposure is hedged.

could not be found in advance anyway, so although we derive bounds for that amount of time which can be determined in advance, the "periodic compounding/geometric mean" framework is the less useful of the two. Sections 3–5 report empirical results using short, medium, and long US bond yields over more than 60 years. Section 3 treats various horizons while Sections 4 and 5 only use Section 1's theoretically-important horizon of twice duration. Section 5 demonstrates that Section 2's convexity results, together with errors due to the linearity approximations, do provide a good explanation of the gaps between initial yields and mean returns. Sections 1–5 assume that the bonds' yield while it is being held changes only because interest rates in general change, not because the bond grows shorter in term. Section 6 relaxes this assumption, and accordingly uses "rolling yield" and "rolldown return." Whether one uses this more comprehensive comparison, "initial rolling yield" versus "return including rolldown return," or the earlier simpler comparison, "initial yield" versus "return excluding rolldown return," the first element of each pair was almost always within a percent or two of the second element at twice duration, was *ex post* not as good a predictor of the second element over horizons much shorter or longer than twice duration, and is *ex ante* not as good a predictor over any horizon other than twice duration.

1. Framework and Basic Result

Supposing that at dates $t = 1, 2, 3, \ldots$ a bond generates nonnegative payments ("coupons") $C_1, C_2, C_3, \ldots$, denote by $PV(Y)$ the "present value" of the bond's future income flows discounted at rate Y, namely $\sum_{t=1}^{\infty} C_t\, e^{-Yt}$ or $\sum_{t=1}^{\infty} C_t/(1+Y)^t$ depending on whether discounting is, respectively, continuous or periodic. Markets in which bonds are bought and sold determine PV; then Y, which besides "rate" is also called the bond's "yield," follows from PV and the C's. The corresponding "current value" at date t is $PV(Y)\, e^{Yt}$ or, respectively, $PV(Y)\,(1+Y)^t$. This paper is concerned with "straight bonds," in which there is a maturity date "m," all the C's before m are identical and strictly positive, all the C's after m are zero, and at m the payment is the previous dates' C plus the "face value," say, \$1000. For the purpose of this paper there is little loss of generality in our assuming that at the beginning of the first period the bond is priced "at par," namely $PV = 1000$ and $C = Y \cdot PV$ or $C = (e^Y - 1)\, PV$, so that $PV = 1000 = \sum_{t=1}^{m} (1000Y)\, e^{-Yt}$ or $PV = 1000 = \sum_{t=1}^{m} (1000\,(e^Y - 1))/(1+Y)^t$, which determines a relationship $Y(m)$ called the "yield curve."

The bond's "Modified Duration" D is defined as $(-1/PV(Y)) \cdot d\,PV(Y)/dY$. Duration has units of time, and after the passage of D (respectively, $(1+Y)\,D$) periods, the bond's then-current value is, to a first order approximation, the same irrespective of any permanent change in its initial yield:

$$0 = \frac{\partial}{\partial Y}\left[PV(Y)\, e^{Yt}\right] \Rightarrow t = \frac{-1}{PV(Y)} \frac{\partial PV(Y)}{\partial Y} = D \qquad (1)$$

and[2]

$$0 = \frac{\partial}{\partial Y}\left[PV(Y)\,(1{+}Y)^t\right] \Rightarrow t = -\frac{1+Y}{PV(Y)}\frac{\partial PV(Y)}{\partial Y} = (1{+}Y)\,D\ . \tag{2}$$

So if one holds on to the bond until date D (respectively, $(1 + Y)\,D$), the return will be approximately the same as the initial yield. This paper addresses the question of whether, if one periodically sells one's bond holdings before date D, each time buying new bonds with duration D, one can expect the return of this (almost) constant duration strategy over some period of time to equal (or approximately equal) the initial yield of the first bond.

Holding a single bond (a "bullet") at each date is not the only way of implementing a constant-duration strategy; another way would be to hold a set of bonds of differing evenly-spaced durations $D_1 < \cdots < D_N$ at each date (a "ladder"), as periods go by maintaining a constant duration by selling the shortest-duration bond (or taking the proceeds from its maturing) and buying a new one of duration D_N. If bullets and ladders were identically risky, arbitrage would make their returns equal and nothing would be lost by only considering bullets. Different market participants can view these strategies as differing in risk, but not in a universally-agreed way, and details are beyond the scope of this paper. Given that we analyze bullets, the term "bond portfolio" will refer to the set of bonds held over time under this strategy.

Accordingly, suppose an initial investment is made in a bond with duration D and initial yield $Y_1(D)$, and at the end of each period, the bond is sold and the proceeds reinvested into a new bond with duration D, where D is longer than the length of one period. (If D were less than one turnover period, (2) or (1) would apply, so $D > 1$ is the only case of interest and is assumed throughout.) Letting an "e" superscript denote the end of the period, when the bond is sold its yield is $Y_2(D^e)$. Often $D^e \approx D - 1$, and throughout this paper D^e will be less than D because we assume that the bonds were all bought at par.[3] Until Section 6 we will assume that $Y_t(D^e) = Y_t(D)$ for all t, that is, that the duration yield curve $Y_t(D)$ for all t is flat for durations between D^e and D ("flat near D").

Make the following approximation for one-period return as a function of interest income and capital gains:

Proposition 1. *["The Return Approximation"] If the yield curve is flat near D then an approximation of the one-period return R_t for a par bond which is originally priced*

[2]The right-hand side of (2) is equal to the "Macauley Duration of a periodically-compounded bond," and the right-hand side of (1) is equal to the "Macauley Duration of a continuously-compounded bond." Proof for the periodically-compounded case: $\partial PV(Y)/\partial Y = -\sum_{t=1}^{\infty} t\,C_t\,(1{+}Y)^{-t-1} = -(1{+}Y)^{-1}$ times the numerator of (16). (Cf. Bierwag et al. (2000, pp. 127–8).) Proof for the continuously-compounded case: $\partial PV/\partial Y = -\sum_{t=1}^{\infty} t\,C_t\,e^{-Yt}$ is -1 times the numerator of the continuous-compounding version of (16), which is obtained by replacing $1/(1 + Y)$ with e^{-Y} in (16).

[3]Proof of $D^e < D$: maturity at the end of the period is one less than maturity at the beginning of the period. For par bonds (though not for discount bonds), duration is always monotonically increasing in maturity. See Villazón (1991 p. 207) and Pianca (2006).

using yield Y_t but whose yield changes to Y_{t+1} at the end of period t is

$$R_t \approx Y_t - (D_t - 1)\,\Delta Y_t$$

where $D > 1$ is the Modified Duration and $\Delta Y_t = Y_{t+1} - Y_t$.

(Proofs for this and most of Section 1 and 2's results are in the Appendix.) The rest of Sections 1 and 2 use a constant duration, making the Return Approximation

$$R_t \approx Y_t - (D-1)\,\Delta Y_t\,. \tag{3}$$

Any arbitrary path of yield through time can be decomposed into a linear component and a nonlinear component in various ways. The purpose of this section is to show that the linear component of one such decomposition gives rise to a return which is predictable even though the slope of this linear component is unknown in advance. For the rest of this section, assume that yields follow a linear path through time, starting at the actual initial yield. (Yields cannot actually follow linear paths in the long run because that would imply that they linearly rise or fall forever, or never change; nor can they in perfect-foresight equilibrium follow linear paths in the short run because having such linear forward curves for multiple maturities would typically generate arbitrage opportunities.) Once we establish that the linear component of the path of yield through time gives rise to a predictable return over one particular horizon, empirical deviations from that predicted return over that horizon will have to be attributed to the nonlinear component of the yield path, or to Return Approximation errors, or to some combination of those. The empirical sections of this paper will illustrate how large those deviations have been.

Along a linear yield path, ΔY_t is the same "ΔY" for all t. *Ex ante*, the value of ΔY is unknowable. Using the Return Approximation and assuming linear time paths of yields,

$$Y_t = Y_1 + (t-1)\,\Delta Y \quad \text{and therefore} \tag{4}$$

$$R_t = Y_1 + (t-1)\,\Delta Y - (D-1)\,\Delta Y \tag{5}$$

$$= Y_1 + (t-D)\,\Delta Y\,. \tag{6}$$

Setting R_t equal to Y_1 in (6) yields $0 = (t-D)\,\Delta Y$, so irrespective of the magnitude of ΔY, instantaneous return R_t will be equal to Y_1 at $t = D$. At what date will average return be equal to Y_1? In (6), the term $(t-D)\,\Delta Y$ is a disturbance term, pushing R_t away from Y_1. Given the linearity of (6), intuition suggests that the influence of this disturbance term will be zero on average when there are the same number of positive and negative instances of this term. If $\Delta Y > 0$ ($\Delta Y < 0$), there are $D-1$ negative (positive) disturbance terms before the date $t = D$ when the disturbance term is zero; so the conjecture is that there will have to be an additional $D-1$ positive (negative) disturbance terms before the average will be zero, making the total wait time $D-1$ periods before period D, then period D itself, then $D-1$ periods after period D, for a total of $2D-1$ periods. This conjecture is correct:

Proposition 2. *If the yield curve is flat near D, yields are linear in time, returns are approximated by (3), and twice duration is an integer, then the number of periods "N_a" which will make the arithmetic mean return equal to the initial yield is*

$$N_a = 2D - 1\,. \tag{7}$$

Corollary. *Under the conditions of Proposition 2, if returns and yields are continuously compounded then the current value of the bond portfolio at date $2D$ is its original value times e^{2DY_1} independent of the value of ΔY.*

Proof of Proposition 2. Non-elementary proof #1, distilled from LBKH (2013 p. 97) and Leibowitz, Bova, and Kogelman (2014 p. 47 Column 2): Use $\sum_{t=1}^{N} t = N\,(N+1)/2$ to write the arithmetic average return as $\overline{R}(N) = (1/N)\sum_{t=1}^{N} R_t = (1/N)\sum_{t=1}^{N}[Y_1 + (t{-}D)\,\Delta Y] = Y_1 + \Delta Y\,[N+1-2D]/2$, which is Y_1 when $N = 2D-1$. Non-elementary proof #2, by analogy with (1) (this also proves Proposition 2's corollary): use (6) to find the current value of the bond portfolio, then set its derivative with respect to ΔY equal to zero and solve for the t which makes it so. To wit: $PV_{t+1} = PV_t\, e^{R_t} = PV_t\,\exp(Y_1 + (t{-}D)\Delta Y)$, a difference equation whose solution is $PV_t = PV_1\,\exp[\Delta Y(D - t - t\,D + (1/2)\,t^2 + (1/2)\,t) + (t{-}1)Y_1]$; then $\partial PV_t/\partial\,(\Delta Y) = PV_t\,(t-1)(t-2D)/2$, confirming that at $t = 2D$, PV_t does not depend on ΔY. (The $t = 1$ root gives no new information because PV_1 is fixed so it obviously does not depend on ΔY.) Elementary proof: Using (5), the arithmetic average return is

$$\begin{aligned}
\overline{R}_a(N_a) &= \tfrac{1}{N_a}\sum_{t=1}^{N_a} R_t = \tfrac{1}{N_a}\sum_{t=1}^{N_a}\left[Y_1 + (t-1)\Delta Y - (D-1)\,\Delta Y_t\right] \\
&= Y_1 + \tfrac{1}{N_a}\sum_{t=1}^{N_a}(t-1)\Delta Y + \tfrac{1}{N_a}\sum_{t=1}^{N_a}\left[-(D-1)\,\Delta Y_t\right] \\
&= Y_1 + \begin{pmatrix}\text{avg. income} \\ \text{in excess of} \\ Y_1\end{pmatrix} + \begin{pmatrix}\text{avg. capital} \\ \text{gains}\end{pmatrix}.
\end{aligned}$$

Mean return is therefore equal to initial yield when the sum of "average income in excess of Y_1" and "average capital gains" is zero. Table 1 shows "capital gains" and "income in excess of Y_1." Equality between the mean return and Y_1 will occur at the date when the sum of the last two entries in the Table's last row is zero, which is $t = 2D - 1$. ∎

The value of ΔY is irrelevant in determining the date when average return will equal initial yield because ΔY drives both sides of the problem, "income in excess of Y_1" and its oppositely-signed "capital gains." The steady increase in cumulative "income in excess of Y_1" overtakes cumulative capital gains more slowly in the constant-duration case than it would in a buy-and-hold case because the magnitude of the instantaneous capital gains part, duration times ΔY, is constant in the constant-duration case but shrinks with time in a buy-and-hold case (as its duration falls). In (7), "$2D - 1$" means

	Period-by-Period		Arithmetic Average to Date	
Date	Capital Gains	Income in Excess of Y_1	Capital Gains	Income in Excess of Y_1
1	$-(D-1)\,\Delta Y$	0	$-(D-1)\,\Delta Y$	0
2	$-(D-1)\,\Delta Y$	ΔY	$-(D-1)\,\Delta Y$	$\frac{1}{2}\,\Delta Y$
3	$-(D-1)\,\Delta Y$	$2\,\Delta Y$	$-(D-1)\,\Delta Y$	ΔY
4	$-(D-1)\,\Delta Y$	$3\,\Delta Y$	$-(D-1)\,\Delta Y$	$\frac{3}{2}\,\Delta Y$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
t			$-(D-1)\,\Delta Y$	$\frac{t-1}{2}\,\Delta Y$

Table 1: Capital gains and income for linear yield paths. Initial duration is D and yields change at the end of each period.

"twice duration minus one turnover period," so that if bonds of duration two years (730 days) are held for one year before being sold, (7) gives $N_a = 2 \cdot 2 - 1 = 3$ years, but if the same bonds are held for one day before being sold, (7) gives $N_a = 2 \cdot 730 - 1 = 1459$ days $= 3.997$ years: the shorter the turnover period, the closer "$2D - 1$" is to $2D$.

Proof of Corollary. If a sequence of "returns" $\{R_t\}_{t=1}^{T}$ makes one unit of wealth grow to $e^{R_1} e^{R_2} \cdots e^{R_T}$ units then replacing each R_t by the arithmetic mean of the R's would lead to the same final value of wealth, making the use of the arithmetic mean appropriate when returns are continuously compounded. The geometric mean of the R's is defined as $\overline{R}_g = [\prod_{t=1}^{N_g}(1 + R_t)]^{1/N_g} - 1$ (one less than the geometric mean used in mathematics). If $\{R_t\}_{t=1}^{T}$ makes one unit of wealth grow to $(1 + R_1)(1 + R_2) \cdots (1 + R_T)$ units then replacing each $1 + R_t$ by one plus the geometric mean of the R's would lead to the same value of wealth, making the use of the geometric mean appropriate when returns are periodically compounded. Proposition 2 concerns the arithmetic mean return, so under its conditions, after $2D - 1$ periods—that is, at the beginning of period $2D$—the arithmetic mean continuously-compounded return would be Y_1. ∎

Section 2 has detailed results for the case of periodic compounding.

The Return Approximation, with its assumption of small ΔY, is relatively innocuous for one period, but we will be using it for a large number of successive periods. Over such long periods, the right-hand side of (8) shows that the Return Approximation subdivides average return into average capital gains and average income, and calculates the capital gains using the "change in yield, $Y_1 - Y_{N+1}$, times duration" approximation even though $Y_1 - Y_{N+1}$ could be very large because those dates are very far apart. When yield changes are large it is inappropriate to ignore the second-order term (the "convexity") of bond price's sensitivity to yield, especially for bonds of long duration, so the Return Approximation's way of calculating capital gains is likely to introduce non-negligible errors; Section 5 will return to this topic. The proposition's corollary gives conditions under which, in the long-run, return is just equal to average yield so capital gains become unimportant; it is used in Section 3.

Proposition 3. *Assuming returns are approximated by (3),*

$$\overline{R}_a(N) = \frac{(Y_1 - Y_{N+1}) \cdot (D-1)}{N} + \frac{1}{N}\sum_{t=1}^{N} Y_t \tag{8}$$

at date N.

Corollary. *Assuming yields are linear in time and returns are approximated by (3),*

$$\lim_{N\to\infty} \overline{R}_a(N) = \lim_{N\to\infty} \frac{\sum_{t=1}^{N} Y_t}{N} \quad \textit{if and only if } \Delta Y = 0\,. \tag{9}$$

The explicit exact formulas for R and D used in the empirical part of this paper are:

Proposition 4. *With no approximations, for continuous compounding*

$$R_t = \ln\left\{ e^{Y_t} - 1 + \frac{e^{Y_t} - 1 + e^{-(m-1)Y_{t+1}}[e^{Y_{t+1}} - e^{Y_t}]}{e^{Y_{t+1}} - 1} \right\} \tag{10}$$

and

$$D = \frac{1 - e^{-mY_t}}{1 - e^{-Y_t}}\,. \tag{11}$$

2. Quadratic Paths, Non-integral Durations, and the Geometric Mean

For yield paths which are not linear but quadratic, return at $2D-1$ will not equal initial yield. Intuitively, since if yields follow a linear path then after $2D-1$ periods positive (or negative) different-from-Y_1 yields just balance negative (or positive) capital gains, resulting in an average return of Y_1, if yield falls below this linear path before it catches up—that is, if yield follows a convex path through time—then these lower yields cause overall return to be lower, below Y_1. Similarly, if yield rises above the linear path before it slows down to meet it—that is, if yield follows a concave path through time—then these higher yields cause overall return to be higher, above Y_1. This intuition is correct:

Proposition 5. *If yields follow the quadratic function $z_1t^2 + z_2t + z_3$ where t is time, if returns are approximated by (3), and if twice duration is an integer, then over a time period of length $N_a = 2D - 1$ the "forecast error"*

$$\overline{R}_a - Y_1 = \frac{2z_1}{3}(1-D)\,D\,. \tag{12}$$

Corollary. *Under the conditions of Proposition 5, forecast error $\overline{R}_a - Y_1$ at period $2D-1$ is negative if the yield path is convex and positive if the yield path is concave.*

Proof of Proposition 5. $\overline{R}_a = (1/N_a)\sum_{t=1}^{N_a}[Y_t - (D-1)\Delta Y_t]$. Writing Y_t as $z_1t^2 + z_2t + z_3$, one has: Y_t is convex if $z_1 > 0$ and concave if $z_1 < 0$; $Y_1 = z_1 + z_2 + z_3$; and $\Delta Y_t = Y_{t+1} - Y_t = 2z_1t + z_1 + z_2$, so $\overline{R}_a = (1/N_a)\sum_{t=1}^{N_a}[z_1t^2 + z_2t + z_3 - (D-1)(2z_1t + z_1 + z_2)]$. It can then be shown, either by tedious calculations using at one point $\sum_{t=1}^{T} t^2 = T^3/3 + T^2/2 + T/6$ and $\sum_{t=1}^{T} t = T(T+1)/2$, or by using a computer algebra system, that if $N_a = 2D-1$ then $\overline{R}_a - (z_1 + z_2 + z_3) = \frac{2}{3}z_1(1-D)\,D$. ∎

	A	G
1. $\Delta Y > 0$ and $2D$ an integer	$N_a = 2D - 1$	$2D - 1 < N_g^+ < \infty$
2. $\Delta Y > 0$ and $2D$ not an integer	$N_a^+ = \lceil 2D - 1 \rceil$	$2\lfloor D \rfloor - 1 < N_g^+ < \infty$
3. $\Delta Y < 0$ and $2D$ an integer	$N_a = 2D - 1$	$D < N_g^+ \leq 2D - 1$
4. $\Delta Y < 0$ and $2D$ not an integer	$N_a^+ = \lceil 2D - 1 \rceil$	$D < N_g^+ \leq 2\lceil D \rceil - 1$

Table 2: Theoretical Results for linear yield paths under the Return Approximation.

Proof of Corollary. Since $D > 1$, the right-hand side of (12) has the opposite sign of z_1. ∎

Like ΔY, the convexity parameter z_1 is unknown in advance. Unlike ΔY, its value does affect $\overline{R}_a - Y_1$ at N_a, so Proposition 5 is not useful for predicting $\overline{R}_a$ but only for *ex post* analysis of forecast errors. (Proposition 5's "convexity" describes d^2Y/dt^2 and is completely unrelated to what is commonly meant by "bond convexity," which was discussed before Proposition 3 and is $d^2PV/dY^2 = \sum_{t=1}^{\infty} t^2 C_t\, e^{-Yt} > 0$ using footnote 2.)

The rest of this section reverts to assuming linear yield paths.

Proposition 2 proved Column A Rows 1 and 3 of Table 2; the next extension is to prove its 2A and 4A, since if $2D$ is not an integer then (7) cannot describe the integer N_a. The proof is relegated to the Appendix; in the Table, the notation $\lfloor x \rfloor$ is the standard notation for "the largest integer smaller than x," the "floor function," and $\lceil x \rceil$ stands for "the smallest integer larger than x," the "ceiling function."

The proof of the corollary to Proposition 2 explained that that proposition was useful for continuously-compounded returns, whereas results for periodically-compounded returns would require knowing the behavior of the geometric mean. (Throughout, Γ means the gamma function of mathematics, not the Gamma of option price theory.)

Proposition 6. *Assuming yields are linear in time and returns are approximated by (3), the number of periods "N_g" which will make the geometric mean return equal to the initial yield satisfies*

$$1 = \prod_{t=1}^{N_g} \left(1 + (t - D)\frac{\Delta Y}{1 + Y_1}\right) \tag{13}$$

$$= \begin{cases} \left(\frac{\Delta Y}{1+Y_1}\right)^{N_g} \dfrac{\Gamma\left(\frac{1+Y_1}{\Delta Y} - D + N_g + 1\right)}{\Gamma\left(\frac{1+Y_1}{\Delta Y} - D + 1\right)} & \text{if } \Delta Y > 0 \\ \left(\frac{-\Delta Y}{1+Y_1}\right)^{N_g} \dfrac{\Gamma\left(-\frac{1+Y_1}{\Delta Y} + D\right)}{\Gamma\left(-\frac{1+Y_1}{\Delta Y} + D - N_g\right)} & \text{if } \Delta Y < 0\ . \end{cases} \tag{14}$$

In the proof of this proposition the gamma function arises due to the following result from Finite Calculus:

Lemma. *If a and b are positive real numbers and $b - a$ is a positive integer then*

$$\sum_{j=a}^{b} \ln j = \ln a + \ln(a{+}1) + \cdots + \ln(b) = \ln\Gamma(b{+}1) - \ln\Gamma(a)\,.$$

Proof. Since $\Gamma(n) = (n{-}1)!$ when n is a positive integer, if a and b are both positive integers then this is merely the claim that $\sum_{j=a}^{b} \ln j = \ln[b!] - \ln[(a{-}1)!]$, which follows from the expansion of $b!$. To construct a proof for non-integer a and b, recall that a basic property of the gamma function is $\Gamma(x{+}1) = x\,\Gamma(x)$. Letting Δ denote the difference operator in this paragraph only, $\Delta \ln\Gamma(x) = \ln\Gamma(x{+}1) - \ln\Gamma(x) = \ln\big(x\,\Gamma(x)\big) - \ln\Gamma(x) = \ln x$. If in general $\Delta f(x) = g(x)$ then[4]

$$\begin{aligned}\sum_{x=a}^{b} g(x) &= \sum_{x=a}^{b} \Delta f(x) = \sum_{x=a}^{b} f(x{+}1) - \sum_{x=a}^{b} f(x) = \sum_{x=a+1}^{b+1} f(x) - \sum_{x=a}^{b} f(x)\\ &= \sum_{x=a+1}^{b} f(x) + f(b{+}1) - f(a) - \sum_{x=a+1}^{b} f(x) = f(b{+}1) - f(a)\,.\end{aligned}$$

Identify $\ln\Gamma(x)$ with f and $\ln x$ with g. ∎

It is not possible to solve (13) or even the closed-form (14) for N_g as a function of D, ΔY, and Y_1. Furthermore, although $N_a = 2D - 1$ did not depend on Y_1 or ΔY, N_g does depend on them, which means it is time-varying and it is not possible to calculate in advance.[5] There may be no integer N_g making the right-hand side of (13) exactly equal to one, but if not, there will exist some integer N_g^+ such that the right-hand side of (13) switches from being less than one to being more than one (or vice versa) when the upper limit of the product switches from being $N_g^+ - 1$ to being N_g^+. Denote $N_g^+ - 1$ by N_g^-. In numerical examples, N_g^+ and N_g^- are easily found by trial and error, calculating the right-hand side of (13) with $N_g = 2, 3, \ldots$ until it crosses one.[6]

[4]This method is called "additive telescoping" on p. 5 of Naik (no date), who writes: "This is just like the fundamental theorem of calculus. Here, f is the discrete analogue of an antiderivative for g, and to add the g-values over an interval, we evaluate f at the endpoints and take the difference. However, the discrete nature of the situation makes things slightly different: instead of $f(b) - f(a)$, we get $f(b{+}1) - f(a)$." Gleich (2005, pp. 7–8) refers to these ideas as the definition of the discrete antiderivative, the definition of the discrete definite integral, and the Fundamental Theorem of Finite Calculus.

[5]If, for fixed N, N-period periodically-compounded returns *exactly* equalled initial yields period after period, an implausible periodicity occurs. Assuming an arbitrary (not necessarily linear) time path of yields, if $1{+}R_1 = \prod_{t=1}^{N}(1{+}R_t)^{1/N}$ and $1{+}R_2 = \prod_{t=2}^{N+1}(1{+}R_t)^{1/N}$ then $(1{+}R_1)/(1{+}R_2) = ((1{+}R_1)/(1{+}R_{N+1}))^{1/N}$, i.e., R_{N+1} is completely determined by R_1 and R_2. Similarly R_{2N+1} would be completely determined by R_{N+1} and R_{N+2}, which in turn would be determined by R_1, R_2 and R_3; and R_{N+3} would be completely determined by R_1, R_2, R_3 and R_4; and so on *ad infinitum.* A similar argument applies to continuous compounding.

[6]Proposition 6 attempts to solve the problem by analogy with the first non-elementary proof of Proposition 2. Alternatively, one could reason by analogy with the second non-elementary proof of Proposition 2: use (6) to find the current value of the bond portfolio under periodic discounting, set its derivative with respect to ΔY equal to zero and solve for the t which makes it so. One has $PV_{t+1} = PV_t\,(1 + R_t) = PV_t\,(1 + Y_1 + (t{-}D)\Delta Y)$, a difference equation whose solution is $PV_t = PV_1\,(\Delta Y)^{t-1}\Gamma(((1{+}Y_1)/\Delta Y) + t - D)/\Gamma(((1{+}Y_1)/\Delta Y) + 1 - D)$. This PV_t can be differentiated with re-

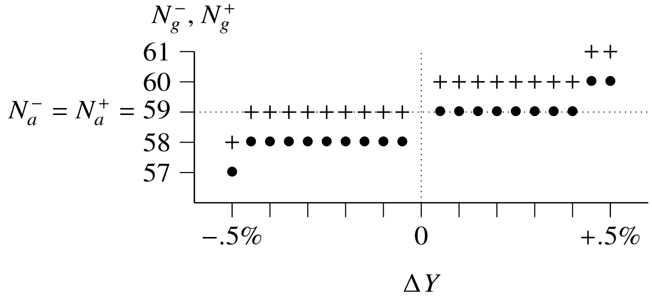

Figure 1. For $D = 30$, $Y_1 = 30\%$, and linear yield paths with different values of ΔY, the plus signs denote N_g^+ and the bullets denote N_g^-.

$\overline{R}_a$ equals $\overline{R}_g$ to first order, so N_a will be close to N_g in most cases.[7] Figure 1 uses exaggerated values of Y_1 and D to prove that it is possible for neither N_g^+ nor N_g^- to be equal to N_a.[8] In Figure 1, when D is an integer and $\Delta Y > 0$, $N_g^+ > 2D - 1$, whereas when $\Delta Y < 0$, $N_g^+ \leq 2D - 1$. We need to prove the claims in Table 2 Column G Rows 1 and 3 that this is a general asymmetry: that it would take longer to get the geometric mean return up to Y_1 after suffering a capital loss resulting from $\Delta Y = +x\%$ than it would take to bring the geometric mean return down to Y_1 after enjoying a capital gain resulting from $\Delta Y = -x\%$. Intuitively, this is because the geometric mean is less than or equal to the arithmetic mean.

Proposition 7. *Assuming yields are linear in time and returns are approximated by (3), N_g^+ satisfies 1G, 2G, 3G, and 4G of Table 2.*

Overall, it is best to use continuously-compounded returns because since their relevant mean is the arithmetic, the results of Column A of Table 2 govern, and the date $N_a = 2D - 1$ at which mean return will be equal to initial yield can be calculated in advance (because it does not depend on ΔY, as N_g does) and is not time-varying (because it does not depend on Y_1, as N_g does).

---

spect to ΔY but the resulting expression, containing not only the gamma function but also the polygamma function, does not permit analytically finding the value of t at which the expression is equal to zero, and solving such a problem numerically would be more complicated than numerically solving (13).

[7] $(1 + \overline{R}_g)^N = \prod_{t=1}^N (1 + R_t) \iff \ln(1 + \overline{R}_g)^N = \sum_{t=1}^N \ln(1 + R_t)$. Applying the first-order Taylor Series approximation $\ln(1 + x) \approx x$ to both sides and dividing by N, $\overline{R}_g \approx (1/N) \sum_{t=1}^N R_t = \overline{R}_a$. For higher-order approximations of $\overline{R}_g$, see Mindlin (2011) and Yogaranpan (2005).

[8] This example is not completely unrealistic because it satisfies the sufficient conditions of the next two sentences: The linear yield path satisfies $R_t > -1$ for all $t \in [1, N_a]$ if $(D - 1) \cdot |\Delta Y| - Y_1 < 1$. The linear path satisfies $Y_t > 0$ for all $t \in [1, N_a]$ if $Y_1 > 0$ when $\Delta Y > 0$ and $Y_1 + 2(D - 1)\Delta Y > 0$ when $\Delta Y < 0$.

Proof of first claim: from (6), if $\Delta Y > 0$ then R_t is increasing in t, so $R_t > -1$ for all t if $R_1 > -1$; imposing that leads to $(D-1)\,\Delta Y - Y_1 < 1$. If $\Delta Y < 0$, R_t is decreasing in t; insisting that $R_{N_a} = R_{2D-1}$ be greater than -1 leads to $(D - 1)(-\Delta Y) - Y_1 < 1$. Combine these results.

Proof of second claim: if $\Delta Y > 0$ the smallest Y_t is Y_1, proving the first part; and if $\Delta Y < 0$ the smallest Y_t in $[1, N_a]$ occurs at $t = 2D - 1$, which, substituted into (4), proves the second part.

3. Empirical Results for Various Horizons

From Section 1 we know in advance that the linear component of the unknowable future yield path will generate a mean return that at $2D-1$ is forecastable with certainty (under the Return Approximation). The nonlinear components of the future yield path generate unforecastable return components, as do errors introduced by the Return Approximation. These unpredictable idiosyncracies will cause Section 1's *ex ante* forecast of mean return to be wrong *ex post.* The history studied in this section shows how inaccurate Section 1's predictions have been. Put another way, given Section 1's model for forecasting (arithmetic mean return at date $2D-1$ is equal to $1 \cdot Y_{mt}$ plus a zero constant plus an error term), the only role remaining for empirical analysis is to describe the error term. In this section the focus is on how errors can cause the dependent variable which is best *ex ante*, return at $2D-1$, to historically have been worse *ex post* than models whose dependent variable is return at a different horizon. We study fifteen different horizons and six types of bonds. Sections 4 and 5 will instead focus only on the error term of the $2D-1$ horizon model. For the reasons given in Section 2, from now on all computed yields and returns are continuously compounded and all means are arithmetic.

Section 1 and 2's theoretical results were for constant-duration portfolios, but it is of interest to see if they have relevance for constant-maturity portfolios, which are more widely discussed among and available to retail investors[9] and which sometimes do not differ much from constant-duration portfolios. We use the following constant-maturity bond yield series from the "Federal Reserve Economic Data" ("FRED") web site of the St. Louis branch of the Federal Reserve System (`https://research.stlouisfed.org`):

- 3-Year Treasury Constant Maturity Rate (GS3). This and all the Treasury bonds are "straight."
- 5-Year Treasury Constant Maturity Rate (GS5)
- 10-Year Treasury Constant Maturity Rate (GS10)
- 20-Year Treasury Constant Maturity Rate (GS20)[10]
- Moody's Seasoned Baa Corporate Bond Yield (BAA). This series does not have a completely constant maturity; the calculations below assume it has a constant maturity which is equal to its average maturity, which is approximately 25 years

[9] For example, the phrase `"constant maturity" bond` gets three times more hits on a Google web search than `"constant duration" bond`, and a search for "constant duration" on the web site of Fidelity Investments returns no hits while "constant maturity" has several.

[10] The 20-year series has 81 months of missing data, from January 1987 until September 1993. The average 20-year yield for the previous 40 months and succeeding 40 months was 8.801%; the average 30-year yield for the previous 40 months and succeeding 40 months was, from another FRED data set, almost the same, at 8.691% (a yield curve that was just barely inverted at the long end). We approximated the 20-year yield for the missing months as 8.801/8.691 = 1.013 times the corresponding 30-year yield. (Otherwise we did not use FRED's 30-year constant-maturity Treasury data because it only starts in February 1977 and 2003–2006 is missing.)

(Hall (2001 p. 1200); Ivaschenko (2003 p. 17)). Some of these bonds are not literally "straight," but their options are unimportant (Hall, Ivaschenko, op.cit.).

These are monthly series and we will assume that their bonds are held for one month before being exchanged for new bonds. The first four series have data starting in April 1953. In this section we use data for all five from April 1953 until April 2014, a total of 733 months (62 years 1 month). Since for each month and each Treasury series we can calculate duration, it is possible to synthesize constant-duration yield time series. We synthesized a 75-month constant-duration Treasury yield series by linear interpolation using the duration and yield data for the 5-year and 10-year series when 75 months fell between their durations, and for the 10-year and 20-year series during the much less frequent times when 75 months fell between their durations. Seventy-five months was almost the longest-duration series we could synthesize by interpolation because the minimum duration of the 20-year series was 78.3 months. We did not synthesize other, shorter-duration yield series because the difference between "constant maturity" and "constant duration" is most evident with longer-term bonds (see for example how the range of the durations varies with maturity in the fifth row of Table 4 below).

Table 3 summarizes returns and yield changes for the full data set, and separately for the "early period" before the September 1981 peak of decades of generally rising interest rates, and for the "late period" of generally falling interest rates after September 1981. The early period is 28 years 6 months long and the late period is 32 years 7 months long. In the early period, with rising yields, average (nominal) returns were lower than average yields, while the opposite was true in the late period; over the full period, average returns were quite close to average yields. This may not be unexpected over a sufficiently long time period because of the Corollary of Proposition 3, which says (accepting Section 1's conditions) that the limit of $\overline{R}_a$ is average yield (and capital gains approach zero) if and only if ΔY is equal to zero, which is close to what happened over the full period. Similarly, over the early and late periods, ΔY was not close to zero, and returns were not similar to average yields. Although Table 3 presents some information about inflation and real returns, all the analysis in this paper is done in nominal terms since we use nominal bonds. This paper's theory does imply that the initial real yield on a constant-duration strategy for inflation-indexed bonds best predicts their real return over twice their duration minus one turnover period, but in the US, inflation-indexed bonds have existed only since 1997, so there is not much data, and analyzing UK data is beyond the scope of this paper.

The model of Section 1 predicts that average return will come closest to equaling initial yield over a horizon of length $2D - 1$. To examine a wide variety of horizons on both sides of that prediction, we used horizons of length "F" times duration, with fifteen values for F: $\{0.75, 1, 1.25, 1.5, 1.6, 1.7, 1.75, 1.8, 1.9, 2, 2.1, 2.2, 2.3, 2.4, 2.5\}$. A horizon of $2D - 1$ implies an F of $2 - 1/D$. Using the smallest and largest values of D observed in the data (Table 4), 2.4 years and 16.8 years, converted to months since that is our turnover period, leads to an $F = 2 - 1/D$ of between $2 - 1/(2.4 \cdot 12) = 1.965$ and $2 - 1/(16.8 \cdot 12) = 1.995$, a small range; comparing F for the *average* durations of

	3 year Treasury	5 year Treasury	75 month duration Treasury	10 year Treasury	20 year Treasury	Long-Term Corporate
FULL PERIOD (3.6% avg. inflation)						
Change in yields	−1.6%	−0.9%	−0.6%	−0.1%	0.2%	1.2%
Avg. ann. yield	5.5%	5.7%	5.8%	6.0%	6.2%	7.8%
Avg. ann. return and its standard deviation	5.5% ±3.2%	5.7% ±4.4%	6.0% ±6.0%	5.9% ±6.3%	6.1% ±8.2%	7.4% ±6.7%
Real avg. ann. return	1.9%	2.1%	2.4%	2.3%	2.5%	3.8%
EARLY PERIOD, April 1953 to September 1981 (4.4% avg. inflation)						
Change in yields	13.1%	12.7%	11.9%	12.0%	11.5%	12.6%
Avg. ann. yield	5.5%	5.6%	5.7%	5.7%	5.7%	7.0%
Avg. ann. return and its standard deviation	4.4% ±3.2%	3.9% ±4.2%	3.1% ±5.2%	2.9% ±5.3%	1.7% ±6.5%	2.6% ±4.9%
Real avg. ann. return	0.0%	−0.5%	−1.3%	−1.5%	−2.7%	−1.8%
LATE PERIOD, October 1981 to April 2014 (2.9% avg. inflation)						
Change in yields	−14.1%	−13.2%	−12.5%	−11.9%	−11.3%	−11.6%
Avg. ann. yield	5.4%	5.7%	6.0%	6.2%	6.7%	8.5%
Avg. ann. return and its standard deviation	6.5% ±3.1%	7.3% ±4.5%	8.5% ±6.5%	8.6% ±6.9%	9.9% ±9.3%	11.7% ±7.7%
Real avg. ann. return	3.6%	4.5%	5.6%	5.7%	7.1%	8.8%

Table 3: All rates including inflation are continuously-compounded annual percentages. Inflation is calculated from FRED's "Consumer Price Index for All Urban Consumers: All Items, Index 1982–84 = 100, Monthly, Not Seasonally Adjusted" (CPIAUCNS).

	3 year Treasury	5 year Treasury	75 month duration Treasury	10 year Treasury	20 year Treasury	Long-Term Corporate
FULL PERIOD (starts 4/1/53)						
Last Date of 2.5 $*$ initial dur.	4/1/14	3/1/14	4/1/14	2/1/14	2/1/14	10/1/13
Last Date of purchase	4/1/07	8/1/02	8/1/98	8/1/95	6/1/89	1/1/91
No. of Observations	649	593	545	509	435	454
Range of initial dur. (years)	2.4–2.9	3.5–4.8	6.3	5.3–8.9	6.5–15.7	6.0–16.8
EARLY PERIOD (starts 4/1/53)						
Last Date of 2.5 $*$ initial dur.	9/1/81	8/1/81	9/1/81	8/1/81		
Last Date of purchase	12/1/74	11/1/70	1/1/66	11/1/60		
No. of Observations	261	212	154	92		
Range of initial dur. (years)	2.7–2.9	4.2–4.8	6.3	8.0–8.9		
LATE PERIOD (starts 10/1/81)						
No. of Observations	307	251	203	167	93	112
Range of initial dur. (years)	2.4–2.9	3.6–4.6	6.3	5.3–7.8	6.5–10.7	6.0–9.7
EARLY PERIOD plus LATE PERIOD						
No. of Observations	568	463	357	259	93	112

Table 4: Timing details and durations for the different time periods. Observations are only between the starting date and the last date of purchase. The late period has the same "last dates" as the full period.

the 3-year and 20-year or long-term corporate would generate an even smaller range. Rounding to one decimal place, the theory thus predicts that the best horizon over which initial yield forecasts return will be at $F = 2$ for all of the bond series.

Throughout this section comparisons are only made using the *same* bonds with different values of F. Therefore the "early period" data will only include bonds whose "purchase date plus 2.5 times their initial duration" occurred on or before the September 1981 (2.5 being the largest F used), and data for the "late period" and the "full period" will only include bonds whose "purchase date plus 2.5 times their initial duration" occurred on or before the April 2014. Since bonds whose purchase dates are before 9/1/81 but whose "purchase date plus 2.5 $*$ initial duration" are after 9/1/81 are excluded from the early and late periods but are included in the full period, the full period has more purchase dates, and thus has more observations, than the union of the early period and the late period, as shown in the last line of Table 4.

To be consistent with the framework of Sections 1 and 2, the bonds were modeled as if they paid interest monthly instead of semiannually. Return and Modified Duration were calculated according to Proposition 4. The Modified Duration of each month's bond was next multiplied by the factor F and rounded to the nearest integer; then the annual continuously-compounded return was calculated for that forward time span. "Forecast Error" is defined as this value minus the annual continuously-compounded initial yield, i.e., if $\overline{R}_{amFt}$ denotes the arithmetic mean realized annual continuously-compounded return for "a strategy of purchasing at date t a bond with maturity (or duration) m, then at the end of the month selling this bond and using the proceeds to buy a new one with maturity (or duration) m" over the forward period of length "F $*$ initial duration," then since its initial yield Y_{mt} is its predicted annual return, its forecast error is $\overline{R}_{amFt} - Y_{mt}$.

With fifteen values of F and sixteen series of bonds (six each for the full and late periods and four for the early period) there were 240 time series of forecast errors. The number of individual (largely overlapping) bond paths whose forecast errors were calculated was 75,555 (fifteen times the sum of the number of observations given in Table 4). Section 4 presents more complete results for the $F = 2 - 1/D \approx 2$ horizon, but due to space constraints the results for all the values of F here are described only by the following summary measures of goodness of fit:

Centered R^2: The R^2 measure of goodness of fit for $\overline{R}_{amFt} = 1 * Y_{mt} + 0$. Many software packages[11] report "centered R^2" as their goodness-of-fit measurement for regressions with a constant term and "uncentered R^2" for regressions without a constant term. We conduct no regression estimations because we are only interested in the model which has slope one and intercept zero; nevertheless, both R^2 measures can be calculated. We agree with Wooldridge (2012 p. 237) that it is better to use centered R^2, which is the scale in most people's minds when thinking about R^2 because most regressions have constant terms. Uncentered R^2 is always greater than centered R^2 and could make the reader think the fit is better than it actually is. For an opposite opinion and survey of this "long dispute" in statistics, see Eisenhauer (2003). Centered R^2 can be negative if there is no constant term and the fit is poor, as occurs several times below.

Root mean square ('RMS') forecast error: Given n purchase dates for bonds of a fixed maturity or duration, and a fixed choice of F, this is $(\sum_t(\overline{R}_{amFt} - Y_{mt})^2/n)^{1/2}$ over the relevant purchase dates.[12]

Average forecast error: $\sum_t(\overline{R}_{amFt} - Y_{mt})/n$. While all our other measures of goodness of fit would rank forecast errors of $\{-2, +2, -2, +2\}$ worse than $\{+1/2, +1/2, +1/2, +1/2\}$, this one will rank the latter worse than the former, and it is possible investors would have such a preference (for example, that they would care about some moving average of the errors).

Frequency of Absolute Value of Forecast Error ("FFE") less than $x\%$: This is the value of the cumulative distribution function of the absolute value of forecast errors for various arbitrary values of x.

Initial yield predicts future return better the smaller the RMS forecast error and the smaller the absolute value of the average forecast error; Figure 2 reports those results.

[11]E.g., Excel: `http://office.microsoft.com/en-us/excel-help/linest-HP005209155.aspx`

[12]LBKH (2013, p. 4-13) call this the "tracking error" ("TE"); so do Leibowitz, Bova, and Kogelman (2014 p. 49).

Initial yield predicts future return better the larger the centered R^2 and the larger "frequency of absolute value of forecast error $< x\%$" is for any given x; Figure 3 reports those results.

Bond returns were somewhat predictable at all of the fifteen F values used, and eliminating $F = 0.75$, only three of the remaining 70,518 forecast errors were not within $\pm 5\%$ of initial yield (the outliers were the May, June, and July 1975 75-month constant-duration Treasurys under $F = 1$, meaning horizons ending in August, September, and October 1981). Nevertheless the graphs show that overall, the predictions were better for F values near the middle of the given range, and an horizon of initial duration ($F = 1$) did not do very well. Section 1 and 2's theoretical model's $F \approx 2$, while certainly better than $F = 1$, did more poorly overall than F's closer to $1\,{}^3\!/_4$. The only bonds for which $F > 2$ did very well, the early period's 75-month and 10-year, had among the smallest number of observations; but F's as low as 1.25 did better than $F = 2$ for the full period 20-year and Long-term Corporate, whose number of observations was not very small.[13]

Since we know that despite these *ex post* results, $F = 2$ is the best value to use *ex ante* when using initial yield to predict return, Sections 4 and 5 treat only that case.

4. Empirical Results for an Horizon of Twice Duration

Table 5 gives details for $F = 2$. The centered R^2 values were quite good, 0.82 or higher, except for the longest bonds of the full and early periods. For the full period the root mean square forecast error at $F = 2$ was around one percent per year for 10-year and shorter bonds, and between 1.6 and 2.1 percent for longer bonds. In the early period, initial yield was $\pm 1\%$ of realized return more than 93% of the time. In the late period, initial yield was $\pm 1\%$ of realized return more than 83% of the time except for the 3-year Treasury at 63%. Predictions for the full period were not as precise, but initial yield was $\pm 1\%$ of realized return more than 70% of the time except for the two longest bonds, for which even $\pm 2\%$ accuracy was only obtained 74% of the time (20 year) and 50% of the time (long-term corporate). In summary, almost all the time, initial yield was within a percent or two of average annual realized return with a horizon of twice initial duration. As Figure 4 illustrates, with long-term bonds, experiencing a constant annual forecast error of a few percent for a period that could be longer than two decades sometimes lead to a substantial difference in final value.

[13] The empirical work of Leibowitz et al. differs from ours in that while we study individual bonds of disparate maturities, they mostly study one intermediate-term bond index, the Barclays US Aggregate Government/Credit Index, or its two components (abbreviating the citations, LBK '14, LB '13, LB '12, LBKH Ch. 4). The exceptions are Ch. 3 of LBKH, using 3-year and 5-year Treasuries, and Ch. 6 of LBKH, which studies a laddered portfolio of Treasuries; in both exceptions, the turnover periods are annual not monthly. Also, their work mostly studies $F \approx 1$ (LBK '14, LB '13, LB '12, LBKH); the exceptions are LBK '14 Fig. 16, LB '13 Displays 5a–c, LBKH Ex. 4.6, 4.14–4.17, and 5.10, which briefly present results for multiple horizons but hold each horizon fixed instead of tailoring it to a constant multiple of each newly purchased bond's duration. In addition, they use much shorter data sets than ours (to limit how much duration varies), and while their averages are arithmetic, they do not specify whether their yields and returns are periodically- or continuously-compounded. Finally, they do not decompose forecast errors as in Section 5 below.

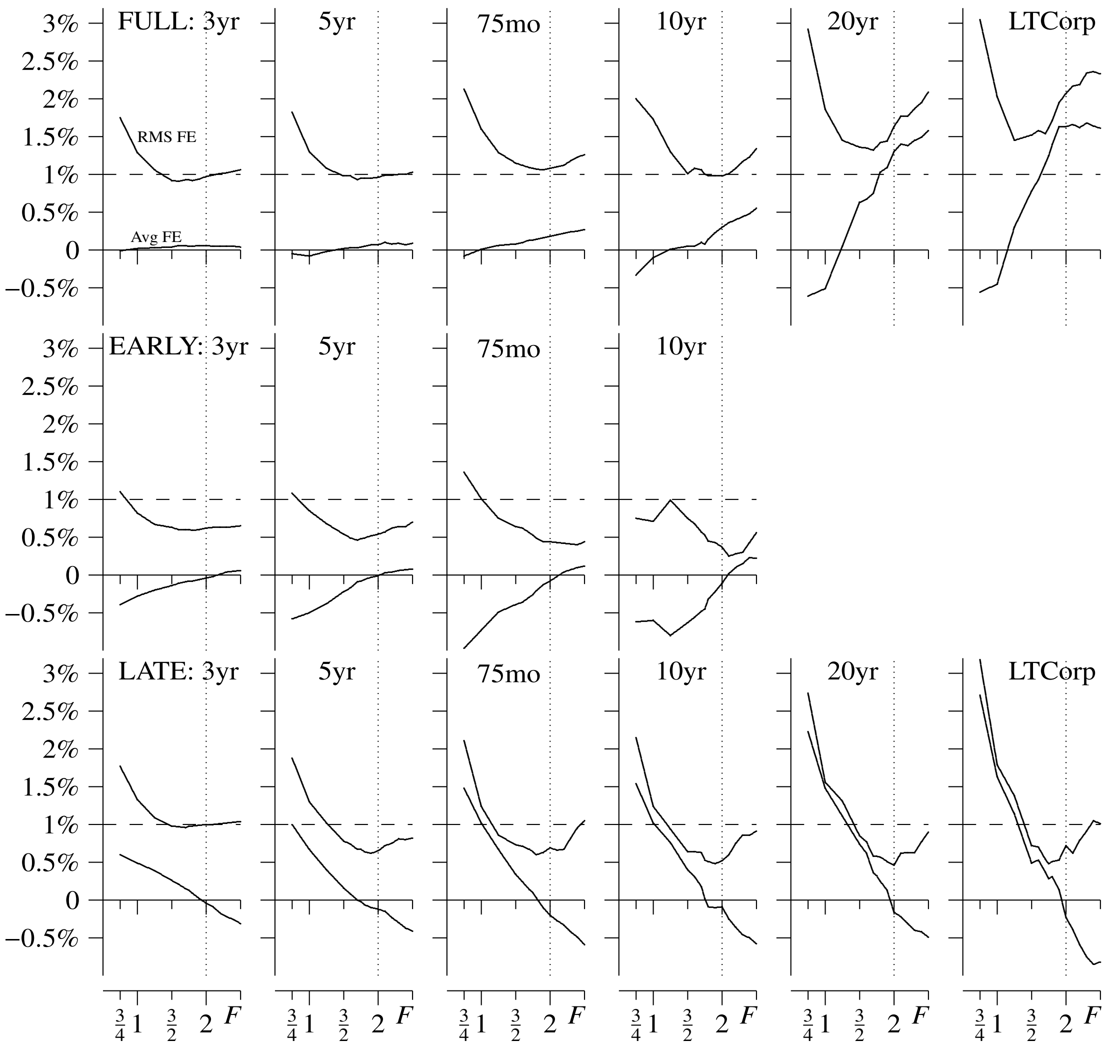

Figure 2. Measures of goodness-of-fit where being closer to zero is better, graphed versus F, the multiple of initial duration over which errors are computed. Abbreviations: "RMS FE," root mean square forecast error (top line in each graph); "Avg FE," average forecast error (bottom line in each graph).

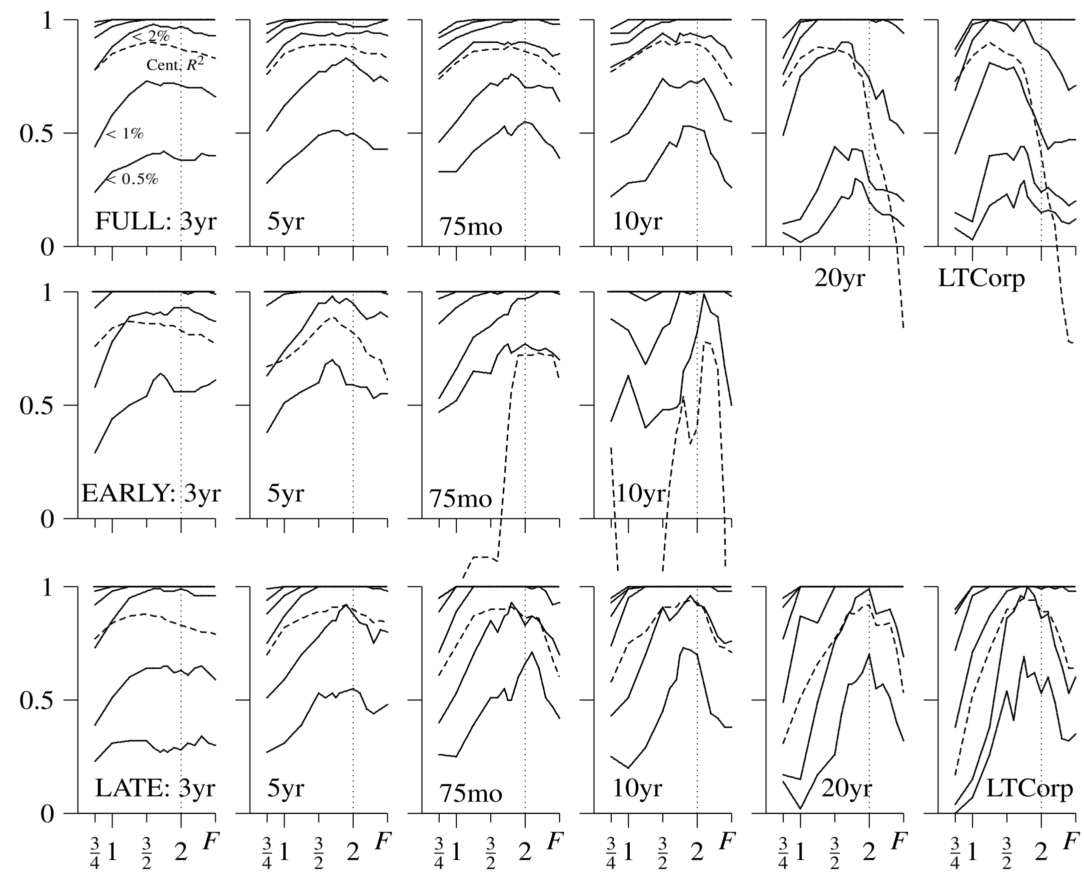

Figure 3. Measures of goodness-of-fit where being closer to one is better, graphed versus F. Abbreviations: "Cent. R^2," centered R^2 of $\overline{R}_{amFt} = 1Y_{mt} + 0$ (dashed line in each graph, with some negative values in the early period truncated); "$< x\%$," frequency of absolute value of forecast error less than the given percent. The lines for "less than 3%," "less than 4%," and "less than 5%" are never labeled.

	3 year Treasury	5 year Treasury	75 month duration Treasury	10 year Treasury	20 year Treasury	Long-Term Corporate
FULL PERIOD						
Avg. ann. return over 2 ∗ initial duration	6.0% ±2.7%	6.5% ±2.7%	6.8% ±2.9%	7.0% ±3.0%	8.0% ±2.5%	9.8% ±2.7%
RMS FE	0.97%	0.96%	1.08%	0.98%	1.63%	2.07%
Avg FE	0.06%	0.07%	0.18%	0.30%	1.30%	1.63%
Cent. R^2	0.87	0.88	0.86	0.89	0.56	0.41
FFE < 0.5%	38%	50%	55%	52%	20%	15%
FFE < 1%	71%	81%	70%	72%	29%	24%
FFE < 2%	97%	94%	90%	93%	74%	50%
FFE < 3%	100%	97%	100%	100%	100%	88%
FFE < 4%	100%	100%	100%	100%	100%	100%
Correl. Coeff.	0.94	0.94	0.93	0.95	0.96	0.94
EARLY PERIOD						
RMS FE	0.62%	0.54%	0.44%	0.37%		
Avg FE	-0.04%	-0.01%	-0.08%	-0.11%		
Cent. R^2	0.83	0.82	0.72	0.40		
FFE < 0.5%	56%	59%	77%	82%		
FFE < 1%	93%	95%	97%	100%		
FFE < 2%	100%	100%	100%	100%		
Correl. Coeff.	0.93	0.92	0.85	0.86		
LATE PERIOD						
RMS FE	1.00%	0.65%	0.69%	0.52%	0.46%	0.72%
Avg FE	-0.04%	-0.12%	-0.20%	-0.09%	-0.16%	-0.23%
Cent. R^2	0.83	0.90	0.86	0.93	0.92	0.89
FFE < 0.5%	28%	55%	66%	70%	70%	53%
FFE < 1%	63%	88%	83%	92%	99%	86%
FFE < 2%	99%	100%	100%	100%	100%	99%
FFE < 3%	100%	100%	100%	100%	100%	100%
Correl. Coeff.	0.93	0.97	0.97	0.97	0.98	0.95

Table 5: Results for $F = 2$; "FFE" is "frequency of absolute value of forecast error" and numbers after "±" denote standard deviation.

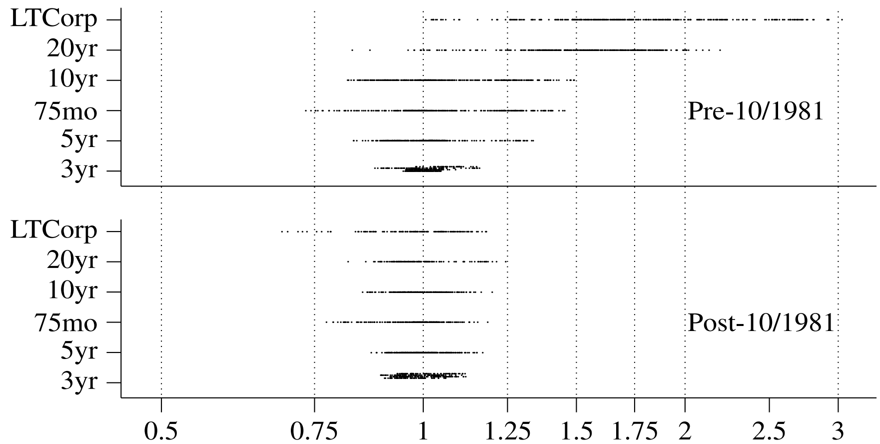

Figure 4. Ratios of actual to predicted (nominal) wealth after $2D$ periods (log scale). Top/bottom: bonds purchased before/after October 1981.

Some investors measure the risk of a constant-duration bond portfolio by its annual variance, a measure of return fluctuation. Since the important source of risk for a long-term investor may not be return fluctuation, but rather having realized return turn out different from initial yield, a better measure of risk might be the square of the root mean square forecast error. To compare RMS FE to more traditional measures, Table 3 reports the annualized standard deviation of monthly returns, and for the full period Table 5 reports the standard deviation of twice-duration average annual returns.

Table 5 reports the Correlation Coefficient between initial yield and subsequent return just to show what an inappropriate measure it is (since if the data *exactly* followed $\overline{R}_{amFt} = aY_{mt} + b$ but a was not equal to one or b was not equal to zero, then the correlation coefficient would be a perfect 1.0 but the hypothesis proposed in this paper would have failed). For example, the "full period" 20-year Treasury in Table 5 had the highest correlation coefficient at 0.96 but, as shown in Figures 2 and 3 (and Figure 6 below), one of the worst fits by all of the legitimate criteria.

The forecast error of return for each date when using $F = 2$ is illustrated in Figure 5. For example, the 10-year bond bought in February 1976 had "two times initial duration" ending on February 1990; its initial yield was 7.64%, but its actual annualized return over that period was 9.23%; so its black solid line is 1.59% higher than its red dotted line, illustrated by the vertical dashed line. The yield at the end of the period was 8.30%; the figure's dashed sloped line joins the initial yield to the end-of-period yield. This bond did better than initially expected because despite the fact that the yield at the end was higher than at the beginning, which caused a small capital loss, the sharply higher yields in the middle of the period caused interest earnings to be much higher than initial yield. Section 5 systematically analyzes such explanations for forecast errors.

The gap between each pair of lines in Figure 5 is graphed with a solid line in Figure 6

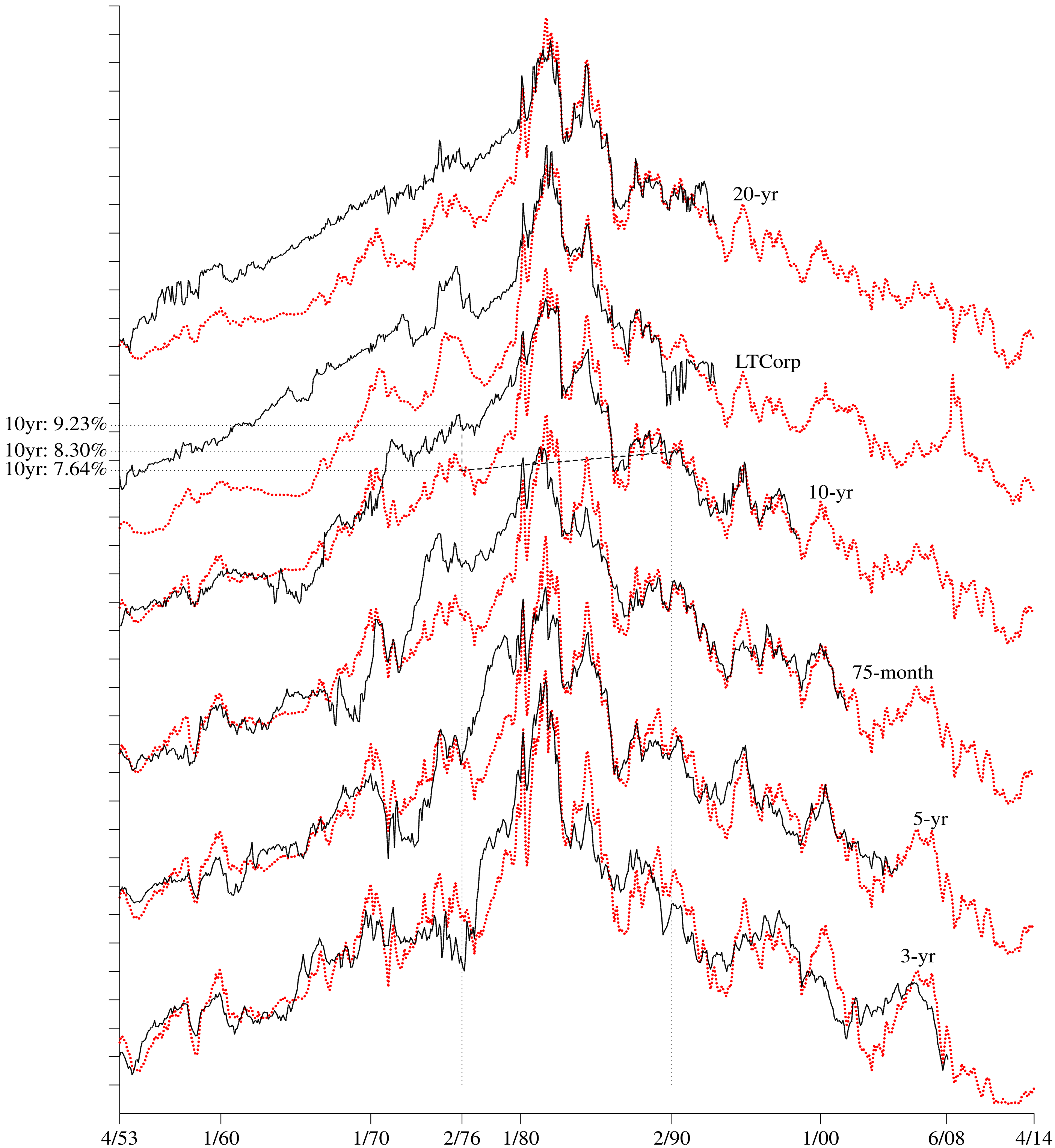

Figure 5. Initial yield (dotted red line) and forward return (black solid line) for $F = 2$. Horizontal axis: purchase date. Vertical axis: continually-compounded annualized return, with absolute yields not shown so that the graphs can be separated for legibility, but vertical tick marks given every 1%.

	Actual Path	Linear Path
Return Approximation	$R_1 \approx 1\% - (D{-}1)(+2\%)$ $R_2 \approx 3\% - (D{-}1)(+0\%)$ (a) $\overline{R}_a \approx 3\% - 0.01D$	$R_1 \approx 1\% - (D{-}1)(+1\%)$ $R_2 \approx 2\% - (D{-}1)(+1\%)$ (b) $\overline{R}_a \approx 2.5\% - 0.01D$
No Return Approximation	$R_1 = \ln\left[e^{1\%} - 1 + \frac{PV^e(3\%)}{PV^e(1\%)}\right]$ $R_2 = \ln\left[e^{3\%} - 1 + \frac{PV^e(3\%)}{PV^e(3\%)}\right]$ (c) $\overline{R}_a = 1.5\% +$ $\frac{1}{2}\ln\left[e^{1\%} - 1 + \frac{PV^e(3\%)}{PV^e(1\%)}\right]$	$R_1 = \ln\left[e^{1\%} - 1 + \frac{PV^e(2\%)}{PV^e(1\%)}\right]$ $R_2 = \ln\left[e^{2\%} - 1 + \frac{PV^e(3\%)}{PV^e(2\%)}\right]$ (d) $\overline{R}_a = (R_1 + R_2)/2$

Table 6: Sources of error in continuously-compounded returns for the example in Figure 7. Returns with no approximation use Section 1's equation (18).

(ignore the figure's dots until Section 5).[14] The sharp drop near the end of the Long-term Corporate series reflects the late-2008 market disruptions.[15] The fit for the 10-year bonds' late period was particularly good. For overall performance, the series having 10 years' maturity or less were all similar to each other and better than the 20-year and the long-term corporate, whose bad performances were due to persistent positive forecast errors for bonds purchased before the early 1980's. Post-1981, forecast errors changed sign at least three times a decade, often much more frequently.

5. *Ex Post* Analysis of Sources of Forecasting Errors

Consider a path of yield given by the solid dots in Figure 7, and consider the linear approximation to it which begins and ends at the same points as the actual path and goes through the open circle. Such a linear path will in general not be a first-order Taylor Series approximation to the original path (not a best-fit trendline). This linear approximation, together with the Return Approximation, gives rise for this example to the $\overline{R}_a$ of Table 6's cell "(b)." The actual return comes about by dropping the linear path assumption and the Return Approximation, resulting for this example in the Table's "(c)." In this section we decompose the gap, (c) minus (b), into two components:

1. a correction "*NL*" for the nonlinearity of the yield path (in the example above, "(a) minus (b)") ; and
2. a correction "*CRA*" for the Return Approximation (in the example above, "(d) minus (b)").

[14]It is impossible to calculate forecast errors for purchase dates to the right of the dashed vertical line when $F = 2.5$ and so those dates are not included in any of this Section's analyses, but some of them are feasible for $F = 2$ and are depicted in Figure 6. They bring Figure 6's number of observations, listed in the order given in Table 4, to 664, 624, 583, 545, 478, and 478.

[15]Long-term Corporate data for (purchase date, last date of twice initial duration, forecast error) for twice-duration periods ending from October 2008 to January 2009 is: (10/89, 11/08, −1.66%); (11/89, 12/08, −1.27%); (12/89, 12/08, −1.28%); (1/90, 11/08, −1.74%); (6/90, 12/08, −1.51%); (7/90, 1/09, −1.35%); (8/90, 11/08, −2.03%); (11/90, 10/08, −1.99%).

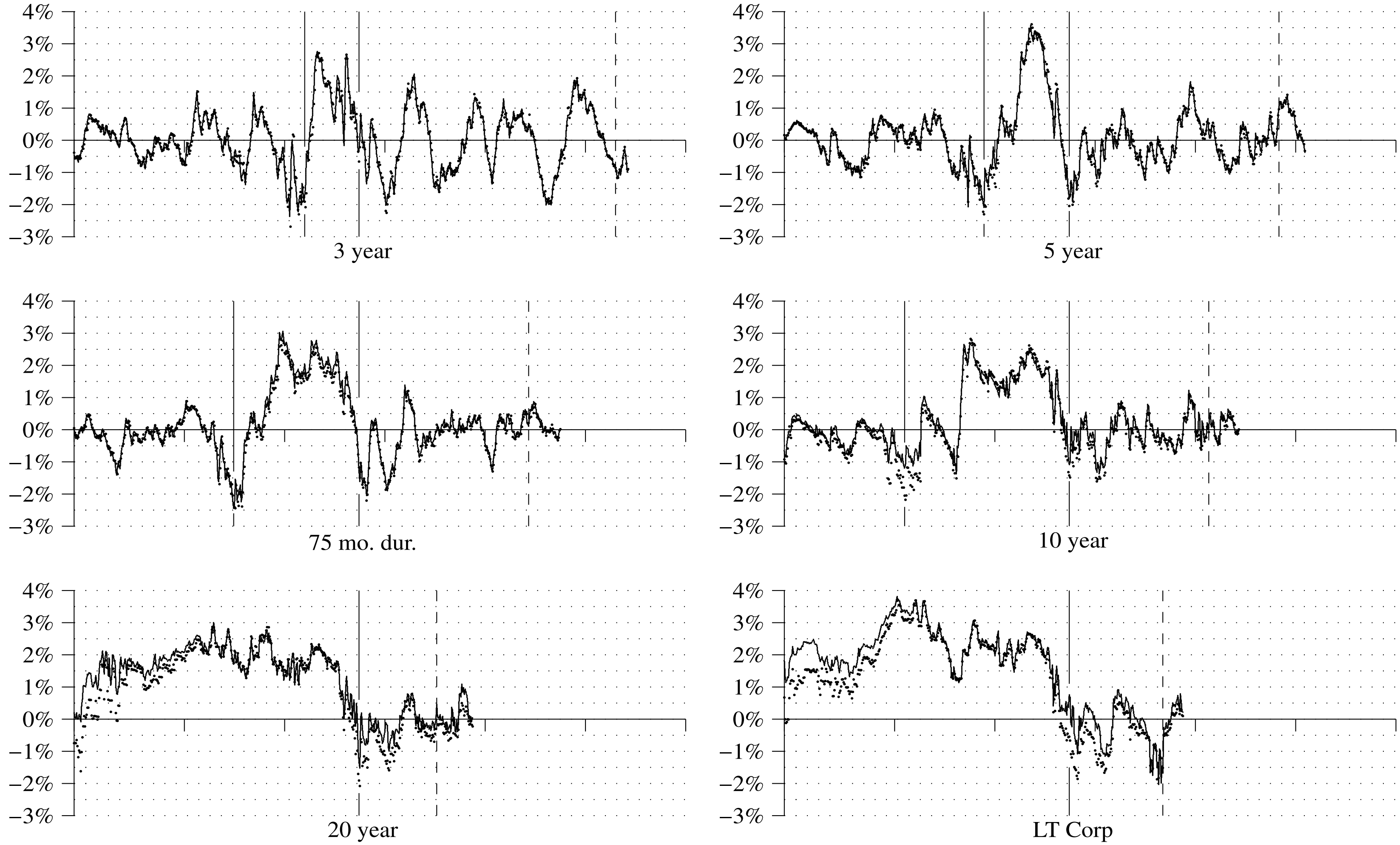

Figure 6. The solid lines are the forecast errors for each series when $F = 2$, graphed with identical axes. The dots are Section 5's "Nonlinearity Measure" for each "two times initial duration" period. The horizontal axes are purchase dates. They start on 4/1/53, end on 4/1/14, have a tick mark every decade, and have a solid vertical line marking 9/1/81. There is another solid vertical line marking each "last date of purchase" of the early period for $F = 2$. The solid vertical lines are sometimes interrupted to improve legibility. The dashed vertical lines show the last observation possible when $F = 2.5$.

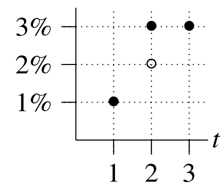

Figure 7. A fictional yield path (black dots) and our *ex post* choice of linear approximation to it $((1, 1\%), (2, 2\%), (3, 3\%))$.

There will be a residual gap between (c) and "(b) plus *NL* plus *CRA*" for a theoretical reason, namely that "(b) plus *NL* plus *CRA*" does not equal (c) although it might be close; and also for several empirical reasons (the data used in Sections 3 and 4 has nonconstant duration or, in the case of the constant-75-month-duration series, imperfect interpolation methods (and in the Long-term Corporate series, nonconstant duration and imprecisely known maturity); horizons rounded to an integer number of months; and possible imperfections in the Federal Reserve's method for calculating constant-maturity yields). The theoretical reason means that modeling forecast error as equaling *NL* + *CRA* is a misspecification. That would be troubling if we were using the model to discover what the true sources of error were, but we already know what the true sources of error are and stated them in the sentence before last; we are only using the model to decompose the sources of errors in an illuminating albeit imperfect way. This section will show that the residual gap was small, so *NL* and *CRA* do explain most of the error between (b) and (c) in our data.

If $Y^{\ell}_{m\tau tF}$ is the value at date t of a straight line joining $Y_{m\tau}$ and $Y_{m(\tau+FD_{\tau})}$ then the measure of nonlinearity "*NL*" we use is[16]

$$NL_{m\tau} = \tfrac{1}{F \cdot D_{\tau}+1} \sum_{t=\tau}^{\tau+FD_{\tau}} (Y_{mt} - Y^{\ell}_{m\tau tF}) \, .$$

For example, *NL* for the 2/76 10-year Treasury is found by constructing a linear path between its actual 2/76 and 2/90 yields—the sloped dashed line drawn in Figure 5—and then calculating the (discrete-time analog of) the average area between the actual time path of yields and the chosen linear path, counting as negative the areas generated when actual yields are below the straight line. By allowing positive and negative deviations from linearity to cancel, this measure ensures that "nonlinearity" will be furthest from zero when yields mostly deviate from linearity in a single direction. If the yield path is convex, the yield path will be below the linear path and $NL < 0$; if the yield path is concave, $NL > 0$. According to Proposition 5's corollary, if the yield path is convex and quadratic, forecast error is negative, and if the yield path is concave and quadratic, forecast error is positive, so if the actual yield paths are sufficiently close to being quadratic, *NL* will have the same sign as forecast error.

[16]This expression for *NL* has the same form as the measure of nonlinearity in equation (2) of Emancipator and Kroll (1993) except that they take the absolute value of the differences.

Note that *NL* in Table 6, 0.5%, is also equal to the difference between the actual point $(2, 3\%)$ and the approximate point $(2, 2\%)$ in Figure 7, divided by its number of periods (two). This reflects a general result: since the actual path and our linear path share the same initial and final yields, the average capital gains along the two paths are the same, so the difference in return is just due to the difference in average income earned.[17]

The dots in Figure 6 show *NL*. To the extent that *NL* moved together with the forecast error, the former explains the latter. The graphs clearly suggest that forecast errors were largely due to nonlinearity as measured by *NL*; for short bonds, it is hard to even see the *NL* dots. One formal measure of how well *NL* fit forecast error is the correlation coefficient between them; another is the centered R^2 of $\overline{R}_{amFt} - Y_{mt} = 1 * NL_{mFt} + 0$. Both of these are reported in Table 7, the latter using the abbreviation "R^2*NLFE*."[18] The correlation coefficients were quite high, ranging from 0.92 to 0.99, except for Long-term Corporate bonds in the late period, which had 0.85. The R^2*NLFE* statistics were almost as good, except for the two longest bonds in the later period. In Figure 6, these bonds did not seem to have a very bad fit between *NL* and forecast errors in the late period except for a few years. Overall, it would seem that *NL* explained most of the forecast errors of interest.

This is incomplete. Figure 6 shows that the greatest deviations between *NL* and forecast errors were for long bonds in the early and mid-1950's, which is not apparent from Table 7 because that period is only analyzed in combination with all other dates (there is no "early period" for those bonds). This makes the period before April 1953 of interest, and while there is no data for the Treasury bonds before April 1953, for the long-term corporate bonds there is data going back to 1919. The yield, forward return, and duration calculated for these 1144 months (95 years 3 months) are plotted in Figure 8. (The sharp interest rate peak of 1932 was due to a rise in long-term Treasury rates (from FRED's "Long-term U.S. Government Securities (Discontinued Series)")

[17] To prove that *NL* is the difference between the average yield along the realized path and the average yield along the linear path, denote the slope of the linear path $(Y_{m(\tau+FD_\tau)} - Y_{m\tau})/(F \cdot D_\tau)$ by s. Then since $Y^{\ell}_{m\tau tF} = Y_{m\tau} + s(t-\tau)$ for $\tau \le t \le \tau + FD_\tau$,

$$NL = \frac{1}{F \cdot D_\tau + 1} \sum_{t=\tau}^{\tau+FD_\tau} (Y_{mt} - Y^{\ell}_{m\tau tF}) = \frac{1}{F \cdot D_\tau + 1} \sum_{t=\tau}^{\tau+FD_\tau} Y_{mt} - \frac{1}{F \cdot D_\tau + 1} \sum_{t=\tau}^{\tau+FD_\tau} Y_{m\tau} - \frac{s}{F \cdot D_\tau + 1} \sum_{t=\tau}^{\tau+FD_\tau} (t-\tau)$$

$$= \frac{1}{F \cdot D_\tau + 1} \sum_{t=\tau}^{\tau+FD_\tau} Y_{mt} - Y_{m\tau} - \frac{s}{F \cdot D_\tau + 1} \frac{(F \cdot D_\tau)(F \cdot D_\tau + 1)}{2} = \frac{1}{F \cdot D_\tau + 1} \sum_{t=\tau}^{\tau+FD_\tau} Y_{mt} - \frac{Y_{m\tau} + Y_{m(\tau+FD_\tau)}}{2}.$$

[18] An alternative to *NL* as an explanation of forecast errors would be an estimate of the right-hand side of (12), where for a purchase date of t_1 one would set $z_1 = \arg\min \sum_{t=t_1}^{t_1+F*D_{t_1}} [Y(t_1) - z_1 t^2 - z_2 t - z_3]^2$ given the constraints $Y(t_1) = z_1 t_1^2 + z_2 t_1 + z_3$ and $Y(t_1 + F * D_{t_1}) = z_1 (t_1 + F * D_{t_1})^2 + z_2 (t_1 + F * D_{t_1}) + z_3$, which can be rewritten so that z_2 and z_3 drop out of the minimization problem, whose only unknown then is z_1. This measure captures only quadratic nonlinearity, not all nonlinearity, and using it on the post-4/53 long-term corporate data gave a predictor which was highly correlated with *NL* but had wider swings, which made it worse than *NL*. This measure is also much harder to calculate than *NL* because it requires solving an optimization problem for each purchase date, whereas *NL* just requires summing up differences.

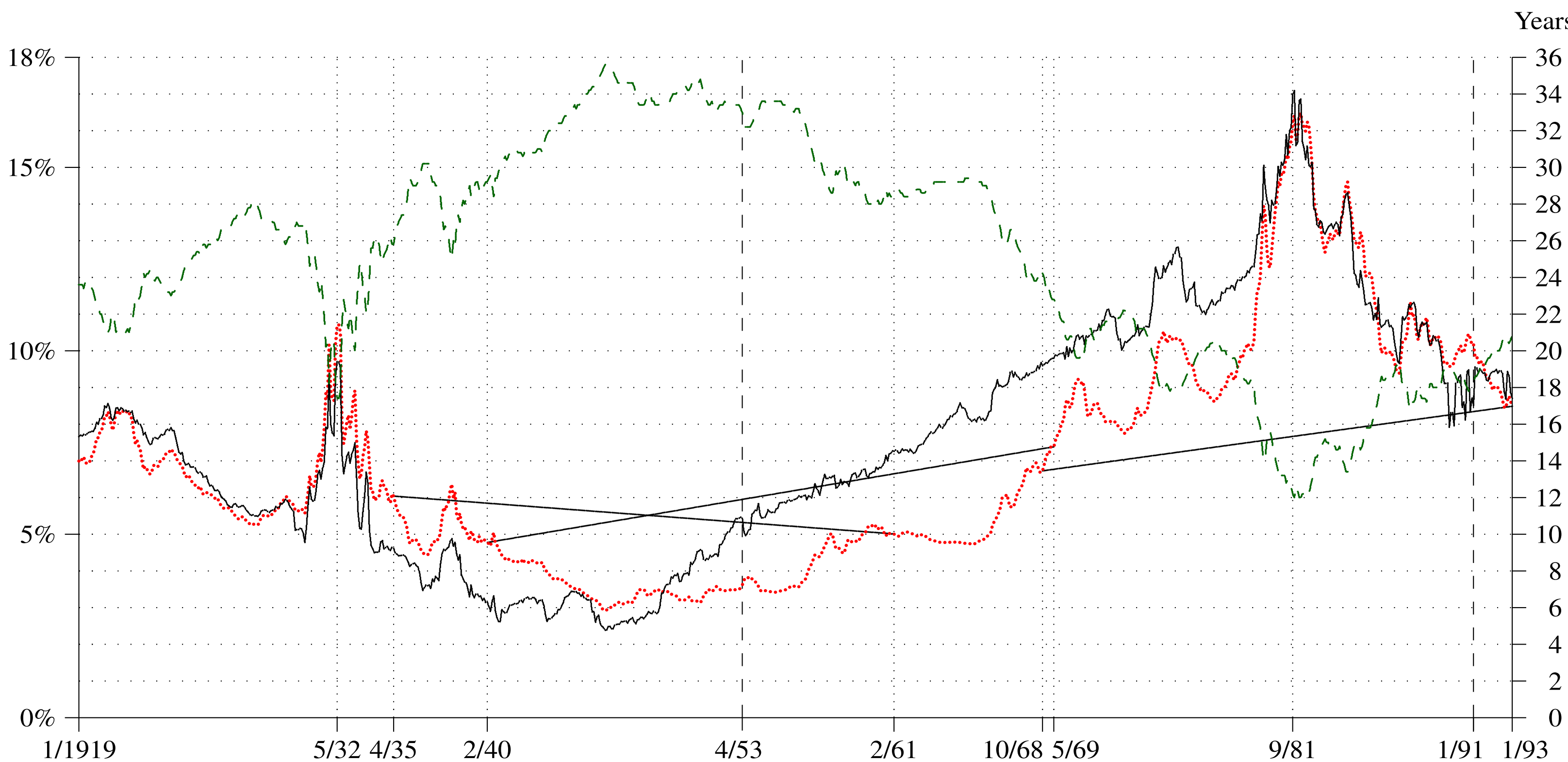

Figure 8. Results for $F = 2$ for the Long-term Corporate series. The left vertical axis is continually-compounded annualized yield or return. The horizontal axis is the initial purchase date. The scale of these axes is the same as in Figure 5. The dotted red line is the yield and the solid black line is the forward realized return over a period twice its initial duration. The right-hand axis is used to read the dashed green line, which is two times duration. Goodness-of-fit measurements are RMS FE, 1.62%; Avg. FE, 0.72%; Cent. R^2, 0.76; FFE < 0.5%, 25%; < 1%, 46%; < 2%, 74%; < 3%, 94%, and < 4%, 100%.

	3 year Treasury	5 year Treasury	75 month duration Treasury	10 year Treasury	20 year Treasury	Long-Term Corporate
FULL PERIOD						
Average *NL*	0.05%	0.06%	0.11%	0.16%	1.04%	1.26%
NL & Forec. Err. Corr. Coeff.	0.98	0.98	0.99	0.98	0.97	0.97
$R^2 NLFE$	0.97	0.97	0.97	0.91	0.80	0.84
$R^2 NLCRAFE$	0.97	0.97	0.96	0.96	0.98	0.99
EARLY PERIOD						
Average *NL*	–0.04%	–0.02%	–0.07%	–0.25%		
NL & Forec. Err. Corr. Coeff.	0.96	0.98	0.97	0.96		
$R^2 NLFE$	0.92	0.96	0.94	0.76		
$R^2 NLCRAFE$	0.92	0.96	0.93	0.90		
LATE PERIOD						
Average *NL*	–0.03%	–0.12%	–0.24%	–0.21%	–0.54%	–0.69%
NL & Forec. Err. Corr. Coeff.	0.99	0.98	0.98	0.94	0.92	0.85
$R^2 NLFE$	0.98	0.96	0.96	0.80	–0.14	0.26
$R^2 NLCRAFE$	0.98	0.96	0.96	0.87	0.91	0.96

Table 7: Results for $F = 2$. "$R^2 NLFE$" is defined in the text; "*NL* & Forec. Err. Corr. Coeff." stands for the correlation coefficient between *NL* and the forecast errors; and "$R^2 NLCRAFE$" is defined later.

of slightly more than 1% coupled with an increase in the spread between these two rates from approximately 2% in early 1929 to more than 7% in mid-1932.)

This graph mainly supports Table 7's conclusion that nonlinearity is the primary explanation for forecast errors. From 1932 until the late-1950's, the yield path on a scale appropriate to long-term bonds—approximately two decades—had an overall convex shape; then it reached an inflection point and became concave. From the Corollary to Proposition 5, this would give rise to negative forecast errors during the earlier period, which are observed at least from 1934 until the late 1940's, and positive errors afterwards, which are observed until the early-1980's. More concretely, Figure 8 has three straight lines, each of which has a length equal to twice the duration of the bond on the left-hand endpoint of the line. One of these lines connects the 4/35 yield with the 2/61 yield, which was lower. The intervening yields lie mostly below this line, so the shape is generally convex. That explains the 4/35 negative forecast error in accordance with Proposition 5's corollary. The next line connects the 2/40 yield with the 5/69 yield, which was higher. The intervening yields lie mostly below this line, so the shape is again generally convex, again leading to a negative forecast error. The yields between the start and end of the line from 10/68 to 1/93 lie mostly above the line, so the shape is generally concave, and the forecast error positive. The lines going to the right from 2/40 and from 10/68 have similar positive slopes, but they have very different forecast errors—the first negative, the second positive—because their associated yield paths differ in convexity. These results are all in complete accordance with Proposition 5's corollary: on this time

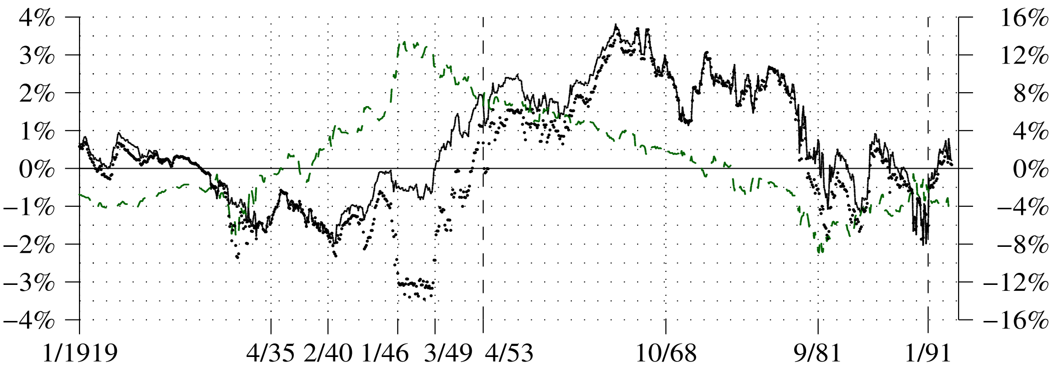

Figure 9. The dots are the nonlinearity measure *NL*, and the solid line is the forecast error, of Figure 8. The right-hand axis corresponds to the dashed green line, which is final yield minus initial yield with $F = 2$. Using the abbreviations of Table 7, *NL* & Forec. Err. Corr. Coeff. is 0.92 and R^2*NLFE* is 0.70. The scale is the same as in Figure 6.

scale it is not the direction of the yield path but rather its concavity or convexity that determines the sign of the forecast error.[19]

Despite these successes in using only the nonlinearity of the yield path to explain this series' forecasting error, when in Figure 9 *NL* is graphed (as the dots) together with the forecast error (the solid line), it becomes clear that the poor fit observed in Figure 6 at the beginning of the post-4/53 period is just the tail end of a period of unprecedented poor fit from 1945 until the mid-1950's. (The fit is also somewhat poor circa 1932 and 1981.) The largest gap in the more than six decades post-April 1953 was 0.71%, so if one only had that data one would hardly expect there to have been a more-than-three-year time period during which the *smallest* gap was 2.27%, as occurred between 1/46 and 3/49. The likely explanation is that since 1945–1952 was an interest rate trough between the 1932 and 1981 peaks, and since the "twice duration" periods of the 1/46–3/49 bonds ended during 4/81–10/82 which is the second peak, those portfolios had extraordinarily large differences between their paths' initial and final yields—illustrated by the green line dashed line in Figure 9—which Section 1's discussion of Proposition 3 warned are just the circumstances, together with long duration which also exists here, in which errors in the Return Approximation are most likely to be important. The absolute value of the difference between initial and final yields for 5/1932 was 7.4%, and for 10/1981 was 8.9%; these were not as large as between 1/46 and 3/49 (when the differences ranged from 10.8% to 12.0%), but they do suggest that the Return Approximation might be helpful in explaining forecast errors in 1932 and 1981 as well as for the period from the late 40's to the mid-50's.

[19]The direction of the yield path did play a role in Column G of Table 2, which showed that forecast error would be lowered by using larger F's when $\Delta Y > 0$ and smaller F's when $\Delta Y < 0$, but that was a consequence of using the geometric mean, which is inapplicable to the continuously-compounded yields and returns of this Section.

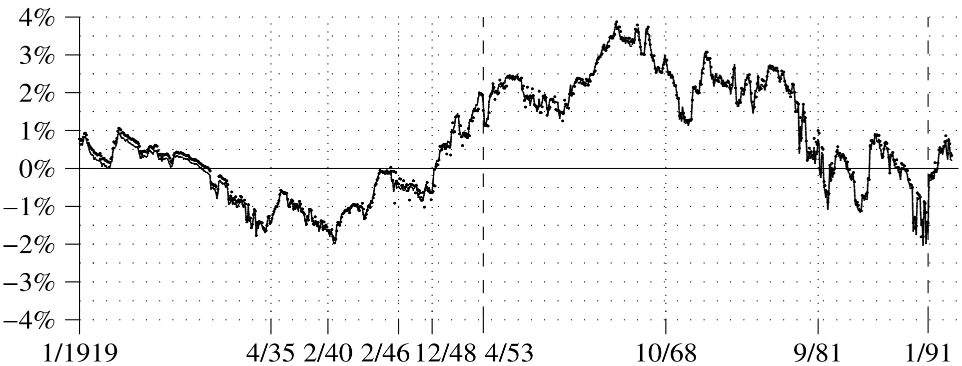

Figure 10. The dots are the nonlinearity measure *NL* added to the correction for the Return Approximation *CRA*, and the solid line is the forecast error, of Figure 8. Using abbreviations resembling Table 7's, (*NL*+*CRA*) & Forec. Err. Corr. Coeff. is 0.997 and $R^2 NLCRAFE$ is 0.99. The scale is the same as in Figure 6.

To confirm this requires calculating a correction for the Return Approximation, *CRA*, which is $CRA_{mF\tau} = \overline{R}^{\ell}_{amF\tau e} - \overline{R}^{\ell}_{amF\tau RA}$ where the first is the arithmetic mean of exact return, starting at time τ and ending at time $\tau + F \cdot D_\tau$, of bonds following a linear yield path from $Y_{m\tau}$ to $Y_{m(\tau+FD_\tau)}$, and the second is the arithmetic mean of the Return Approximation of return over the same linear yield path. Because the formula for exact bond price changes is nonlinear, calculating the first quantity requires generating a counterfactual linear yield path between months τ and $\tau + F \cdot D_\tau$ for each starting month τ (and each bond series and each F of interest) in the data set, then calculating each exact monthly price change along each of the counterfactual paths (increasing the number of monthly returns to be calculated for all the bond series from less than 6000 to more than 2.2 million). Performing the calculations and then adding *CRA* to *NL* should provide a close match to forecast error, and Figure 10, whose dots are $NL + CRA$ instead of just *NL*, shows that it does. Incorporating *CRA* eliminates the unexplained forecast error around the interest rate peaks and trough.

We calculated *CRA* for all the other bond series, but space does not permit inclusion of graphs like Figure 9 and 10 for them. Letting "$R^2 NLCRAFE$" be the centered R^2 of $\overline{R}_{amFt} - Y_{mt} = 1 * NL_{mFt} + 1 * CRA_{mFt} + 0$, incorporating *CRA* results in the important improvements for the longer-term bonds reported in Table 7. Another way to see the improvements is in Table 8, which compares the size of root mean square forecast error ("RMS FE"), root mean square forecast error minus *NL*, and root mean square forecast error minus *NL* minus *CRA*. Using FE − *NL* − *CRA*, $F = 2$ gave the smallest errors. One cannot calculate *NL* and *CRA* in advance but if one could, then taking them into account, even in the theoretically incomplete way of *NL* + *CRA*, lead to $F = 2$ giving the best match between initial yield and realized return. This is the clearest empirical reflection yet of the theoretical argument that initial yield best predicts return over a horizon of $F \approx 2$.

	RMS FE			RMS (FE − *NL*)			RMS (FE − *NL* − *CRA*)		
		off by	min. F		off by	min. F		off by	min. F
3 Year	0.97%	0.06%	1.6	0.18%	0.00%	2	0.18%	0.00%	2
5 Year	0.96%	0.02%	1.7	0.18%	0.00%	2	0.18%	0.00%	2
75 Month	1.08%	0.01%	1.9	0.19%	0.00%	2	0.21%	0.00%	2
10 Year	0.98%	0.01%	1.9	0.28%	0.02%	1.9	0.20%	0.00%	2
20 Year	1.63%	0.31%	1.7	0.43%	0.13%	2.5	0.13%	0.00%	2
LTCorp	2.07%	0.62%	1.25	0.51%	0.17%	2.4	0.11%	0.00%	2
LTCorp 1919–	1.62%	0.37%	1.7	0.79%	0.13%	1.7	0.12%	0.00%	2

Table 8: RMS of forecast error and of forecast error minus various corrections, for $F = 2$, for the full period except for the last row, which is 1/1919–4/2014; how much greater that is than if one were to use the F which minimizes the respective RMS (from the fifteen F's described earlier); and that minimizing F.

Overall, the three factors of initial yield, *NL*, and *CRA* captured almost all aspects of the constant-duration (or constant-maturity) bond portfolio returns over approximately $2D - 1$ periods. Of the three factors, initial yield is the only one knowable in advance, confirming its significance as a tool for forecasting returns.

6. Non-flat Yield Curves

Above we assumed that the yield curve was "flat near D": $Y_t(D^e) = Y_t(D)$ for all t. However, $D^e < D$ and yield curves are usually upward-sloping, so usually $Y_t(D^e) < Y_t(D)$ (for example, 35-month bonds usually have a lower yield than 36-month bonds). Although the difference in yields between 35-month bonds and 36-month bonds is usually small, if it affects return twelve times a year, its cumulative effect may not be small; so in this section we take it into account. The change in the bond's yield can be rewritten as

$$Y_{t+1}(D^e) - Y_t(D) = [Y_{t+1}(D^e) - Y_{t+1}(D)] + [Y_{t+1}(D) - Y_t(D)]$$

where the second term was already present in previous sections (where it was written simply as $Y_{t+1} - Y_t$ since D is fixed) and where the first term is new. The first term is typically negative, giving rise to a capital gain called "rolldown return" in a process called "rolling down the yield curve." We show that the main results of Section 1 and the beginning of Section 2 hold true in this more general setting if initial yield is replaced with the sum of initial yield and rolldown return. We only consider the continuously-compounded case. Omitted proofs are in the appendix.

Proposition 1′. *Under the assumptions of Proposition 1 except permitting the yield curve to be non-flat near D, (3) is replaced by*

$$R_t \approx RY_t - (D-1)\Delta Y_t \tag{3'}$$

where

$$RY_t = Y_t(D) - (D-1)[Y_{t+1}(D-1) - Y_{t+1}(D)]$$

denotes the "rolling yield" and $\Delta Y_t = Y_{t+1}(D) - Y_t(D)$.

Proposition 2′. *Make the assumptions of Proposition 2, except permitting the duration yield curve near D to have a non-zero but constant slope s for all t. Then the conclusions of Proposition 2 and its corollary hold, replacing the initial yield Y_1 with the initial rolling yield RY_1.*

Proof. If ΔY_t is constant for all t and if $Y_{t+1}(D-1) - Y_{t+1}(D) \equiv s$ for all t then from (3′),

$$\begin{aligned} R_t &\approx Y_1 + (t-1)\Delta Y - (D-1)s - (D-1)\Delta Y \\ &= RY_1 + (t-1)\Delta Y - (D-1)\Delta Y && (5') \\ &= RY_1 + (t-D)\Delta Y && (6') \end{aligned}$$

which have the same form as (5) and (6) with RY_1 replacing Y_1. Hence RY_1 replaces Y_1 in the expression for $\overline{R}_a(N_a)$ in the proof of Proposition 2. ▮

Leibowitz and Bova (2014) contains an assertion of Proposition 1′ and a proof, using the method of their earlier papers, of Proposition 2′. To make this result operational, RY_1's term $Y_2(D-1) - Y_2(D)$ has to be replaced with $Y_1(D-1) - Y_1(D)$ because the former is unknown at $t = 1$.

Proposition 4 holds as before if, in (10), Y_t means $Y_t(D)$ and Y_{t+1} means $Y_{t+1}(D^e)$, and if in (11), Y_t means Y_t^m. Note that (11) can be thought of as defining m as a function of D and t, so given the yield curve Y_t^m at each date one can find $Y_t(D) = Y_t^{m(D,t)}$ and $Y_{t+1}(D^e) = Y_{t+1}^{m(D,t)-1}$.

Finally:

Proposition 5′. *Make the assumptions of Proposition 5, except permitting the duration yield curve near D to have a non-zero but constant slope s for all t. Then the conclusions of Proposition 5 and its corollary hold, replacing the initial yield Y_1 with the initial rolling yield RY_1.*

Proof. If $Y_{t+1}(D-1) - Y_{t+1}(D) \equiv s$ for all t then from (3′),

$$\begin{aligned} \overline{R}_a &= \frac{1}{N_a}\sum_{t=1}^{N_a}[Y_t(D) - (D-1)s - (D-1)\Delta Y_t] \quad \Rightarrow \\ \overline{R}_a + (D-1)s &= \frac{1}{N_a}\sum_{t=1}^{N_a}[Y_t(D) - (D-1)\Delta Y_t]\,. \end{aligned}$$

The proof of Proposition 5 showed that if $Y_t = z_1t^2 + z_2t + z_3$ then at $N_a = 2D-1$, the right-hand side of this equation is equal to $(z_1+z_2+z_3) + \frac{2}{3}z_1(1-D)D$. So $\overline{R}_a - Y_1 + (D-1)s = \frac{2}{3}z_1(1-D)D$; but the left-hand side of this equation is $\overline{R}_a - RY_1$. ▮

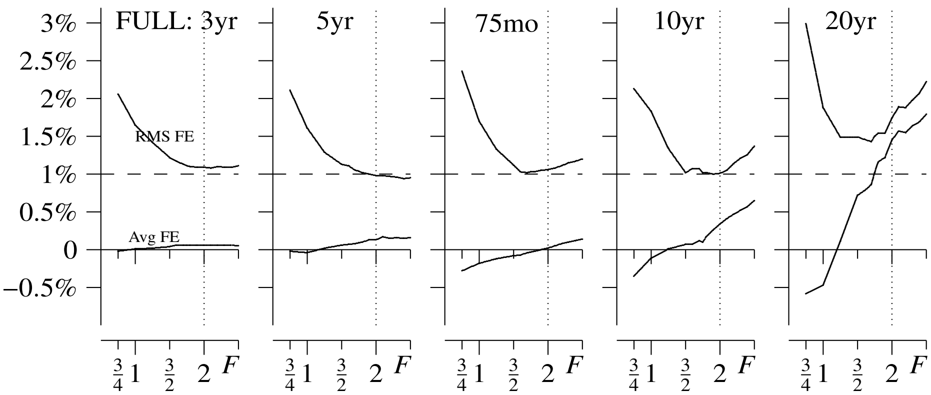

Figure 2′. The rolling-yield version of Figure 2 (only full period shown).

Given these results, the empirical analysis done in Sections 3–5 under the assumption that the yield curves were "flat near D" can be redone to determine how well initial rolling yield predicts return including rolldown return. The rest of this section summarizes the results for the "full period."

Rolldown return was calculated using piecewise linear interpolation to estimate each date's yield curve; this is the simplest possible method and echoes our use of piecewise linear estimation in synthesizing the constant-duration bond series and in footnote 10. (There is a large literature on more sophisticated methods of constructing yield curves, and for example the FRED constant-maturity Treasury "data" we use is actually not raw data, but the result of yield curve interpolation using quasi-cubic hermite spline functions.[20]) No shorter-maturity series exists corresponding to the long-term corporate series used above, so long-term corporates had to be omitted. In interpolating 35-month Treasurys, FRED's 1-year constant-maturity Treasury series was used.

Prediction accuracy was expected to be worse here than in Sections 3–5 because here there is an additional source of error: earlier sections assumed the yield curve was locally flat and then calculated realized return as if that were true, whereas this section assumes the yield curve locally has a constant slope and then calculates realized return using the actual non-constant slopes. However, an analysis like Section 3's results in Figures 2′ and 3′, which have no important differences from Figures 2 and 3 except perhaps for slightly worse fits for the 3-year bonds, and no perceptible differences at all for many of the longer-term bonds.

The conclusion is the same if, as in Section 4, we turn to only the $F = 2$ case. Comparing Table 9 and Table 5 shows that average return was greater here, due to rolldown, by 0.6%, 0.4%, 0.1%, 0.3%, and 0.2%, respectively, so yield curves were usually upward-sloping, as expected; but despite standard deviation being uniformly 0.1% higher here, RMS FE, centered R^2, and the cumulative distribution function for forecast error show no important worse fit here compared with Section 4, except perhaps

[20] `http://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/yieldmethod.aspx`

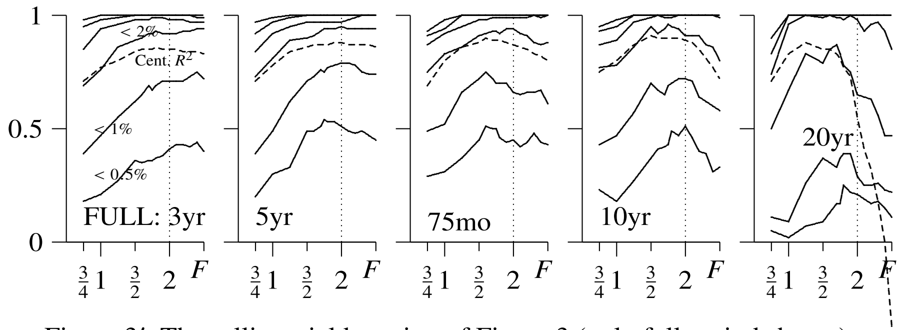

Figure 3′. The rolling-yield version of Figure 3 (only full period shown).

again for the 3 year. Fit at each date is shown in Figure 5′. It may be hard to discern the differences between the lines in Figure 5′ and Figure 5, but the general upward displacement of the former can be detected by comparing their ending positions with the position of their labels, which is the same in both figures, and by contrasting the starting positions of the 3-year and 5-year bonds. A graph (not supplied) superimposing the solid black lines of Figures 5 and 5′ shows that they were much more similar before the early 1980's than afterwards for the shorter bonds, and before the early 1970's for the 10-year and 20-year. Overall, using initial rolling yield to predict return including rolldown return gave results which had essentially the same accuracy as using initial (non-rolling) yield to predict return excluding rolldown return.

To conduct an analysis as in Section 5, the nonlinearity measure should be defined as before, as nonlinearity of the path of yield, not of rolling yield, since constancy of ΔY_t pertains to yield. Given the presence of the extra source of error, it is unsurprising that *NL* in Table 9 explained much less of forecasting error than in Table 7; *NL* still explained much of it. *CRA* is now the difference between: exact calculation (equation (18) as interpreted in the paragraph before Proposition 5′) with linear yield paths through time and, in addition, "*s*" constant; and the Return Approximation with linear yield paths and "*s*" constant.

To obtain the correction for assuming a constant yield curve slope, "*CCYCS*," start from (3′), assume the yield path is linear ((4)), but do not assume that the slope of the duration yield curve is constant. Then

$$R_t = Y_1(D) + (t - D)\Delta Y(D) - (D - 1)\,[Y_{t+1}(D{-}1) - Y_{t+1}(D)]\,.$$

The difference between this and assuming that $[Y_{t+1}(D{-}1) - Y_{t+1}(D) \equiv s = Y_1(D{-}1) - Y_1(D)$ is $\Delta R_t = -(D{-}1)\,[Y_{t+1}(D{-}1) - Y_{t+1}(D)] + (D{-}1)\,s$, and *CCYCS* was set to the average of these values, $(1/N)\sum_{t=1}^{N} \Delta R_t$ (translated to a starting date of τ and an ending date of $\tau + F \cdot D_\tau$). If the bond series was constant maturity instead of constant duration, R_t was set to $-(D_{t+1}{-}1)\,[Y_{t+1}(D{-}1) - Y_{t+1}(D)] + (D_1{-}1)\,s$. As shown in the last line of Table 9 and the right-hand part of Table 8′, adding *CCYCS* to *NL* and *CRA* always

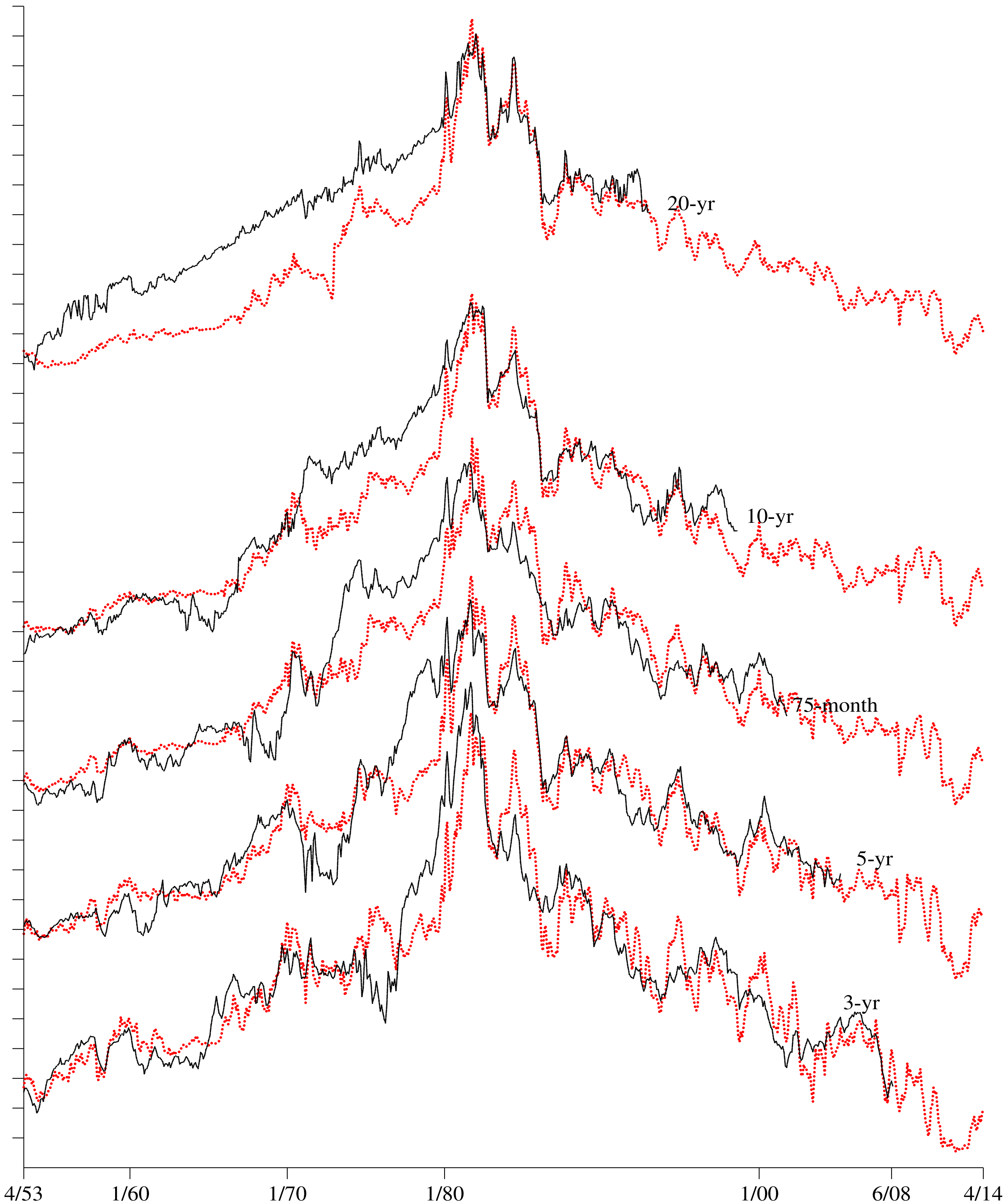

Figure 5′. Initial rolling yield (dotted red line) and forward return including rolldown return (solid black line) for $F = 2$; compare with Figure 5.

	3 year Treasury	5 year Treasury	75 month duration Treasury	10 year Treasury	20 year Treasury
Avg. ann. return over 2 * initial duration	6.6% ±2.8%	6.9% ±2.8%	6.9% ±3.0%	7.3% ±3.1%	8.2% ±2.6%
RMS FE	1.09%	0.98%	1.06%	1.01%	1.74%
Avg FE	0.06%	0.13%	0.02%	0.34%	1.45%
Cent. R^2	0.85	0.88	0.87	0.89	0.55
FFE < 0.5%	41%	51%	45%	51%	21%
FFE < 1%	71%	79%	66%	72%	29%
FFE < 2%	92%	95%	94%	91%	65%
FFE < 3%	98%	97%	99%	100%	98%
FFE < 4%	100%	100%	100%	100%	100%
$R^2 NLFE$	0.60	0.69	0.72	0.73	0.67
$R^2 NLCRAFE$	0.60	0.70	0.71	0.84	0.92
$R^2 NLCRACCYCSFE$	0.97	0.97	0.93	0.96	0.98

Table 9: Results of using initial rolling yield to predict return including rolldown return: $F = 2$ for the full period. The last three lines give centered R^2 of FE versus *NL*, *NL* + *CRA*, and $NL + CRA + CCYCS$.

	RMS FE	off by	min. F	RMS (FE − *NL*)	off by	min. F	RMS (FE − *NL* − *CRA*)	off by	min. F	RMS (FE − *NL* − *CRA* − *CCYCS*)	off by	min. F
3 Year	1.09%	0.01%	2.1	0.69%	0.06%	2.4	0.69%	0.05%	2.3	0.18%	0.00%	2
5 Year	0.98%	0.04%	2.4	0.54%	0.04%	2.2	0.54%	0.03%	2.2	0.18%	0.00%	2
75 Month	1.06%	0.04%	1.7	0.56%	0.00%	2	0.58%	0.00%	2	0.27%	0.01%	1.9
10 Year	1.01%	0.01%	1.9	0.50%	0.01%	1.9	0.39%	0.00%	2	0.18%	0.00%	2
20 Year	1.74%	0.30%	1.7	0.55%	0.14%	2.5	0.26%	0.00%	2	0.14%	0.00%	2

Table 8′: The analogue of Table 8 for non-flat yield curves.

improved the amount of forecast error that was explained. Furthermore, Table 8′ shows that adding all the corrections revealed the *ex ante* theoretical superiority of $F = 2$ slightly better than not including all of them. In the right-hand part of that table, all the bonds achieved lowest RMS at $F = 2$ except one, which was off by merely one basis point from the best F, 1.9, which is adjacent to 2.0 in the set of F's tested. One cannot calculate *NL*, *CRA*, or *CCYCS* in advance but if one could, then taking them into account, even in the theoretically incomplete way of $NL + CRA + CCYCS$, leads to $F = 2$ giving the best match or almost the best match between initial rolling yield and "realized return including rolldown return."

Tables 8 and 8′ can be interpreted as in Figure 11. For example, from the last line of Table 8, RMS forecast error is 1.62%, which shrinks by 1.62% − 0.79% = 0.83% due to *NL*, and shrinks by a further 0.79% − 0.12% = 0.67% due to *CRA*. The height of the corresponding bar in Figure 11 is 1.62%. The height of its *NL* portion would be 0.83%, and of its *CRA* portion would be 0.67%, were it not for the complication that Table 8

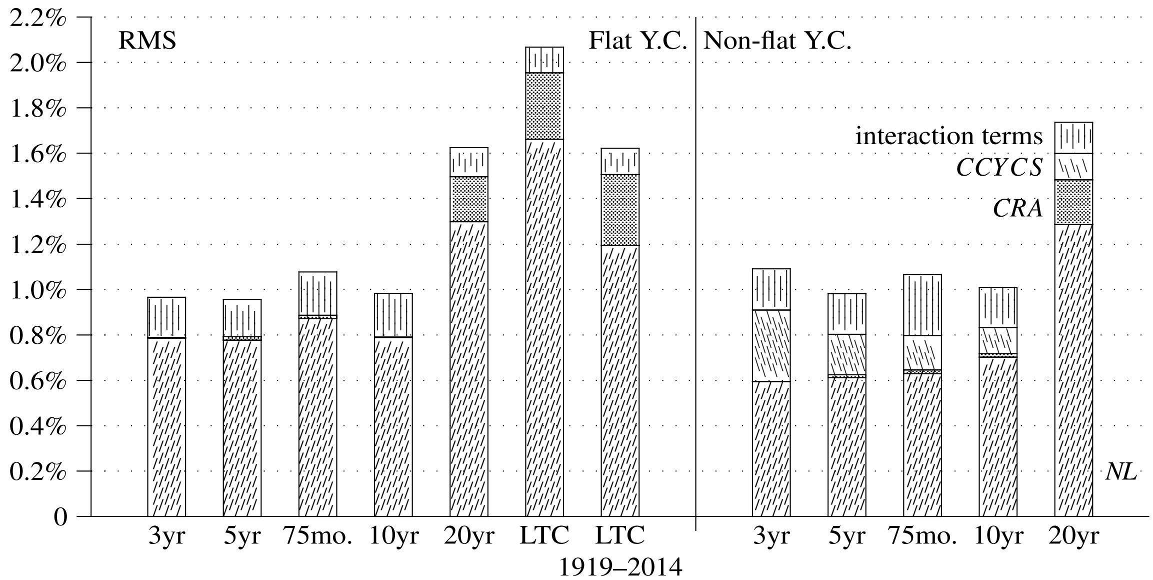

Figure 11. RMS forecast error when $F = 2$, and how it shrinks when taking into account *NL* (bottom portion), *CRA* (next to bottom portion), and for non-flat yield curves *CCYCS* (next to top portion); error remaining is due to uncharacterized interaction terms (top portion).

only analyzes what happens when *NL* is included first and *CRA* is included afterwards, but if one were to do the calculation in the other order the values would change. In other words, although there is only one way of calculating *NL* and *CRA*, there are two ways of calculating their influence on RMS error: for example, for *NL*, the way Table 8 makes possible, RMS(FE) − RMS(FE − *NL*), but also RMS(FE − *CRA*) − RMS(FE − *CRA* − *NL*). Figure 11 follows an approach which Bring (1996 Sec. 5.2) attributes to Chevan and Sutherland (1991) and which Grömping (2012 p. 139) attributes to Lindeman, Merenda, and Gold (1980) (though in the context of R^2 instead of RMS error): calculate the *NL* and *CRA* contributions for both orders and depict the average *NL* and *CRA* contribution; and for the non-flat yield curve part, calculate the *NL*, *CRA*, and *CCYCS* contributions for all permutations of the three variables and depict the average contribution of each. For non-flat yield curves, *CCYCS* was most important for the shorter bonds, where the yield curve typically has the greatest slope (and evidently varies the most). For both locally-flat and for non-flat yield curves, incorporating *CRA* greatly improved the fit of the longest-term bonds but had little effect on the shorter bonds, as expected from Section 1's discussion of Proposition 3. Figure 11 is a useful summary but lacks fine detail; contrasting its "LTC" and "LTC 1919–2014" does not reveal the great importance of *CRA* to the late 1940's shown in Figures 9 and 10.

Conclusion

Over a horizon of twice duration minus one turnover period, whether yields have gone up or down does not affect realized return of a constant-duration portfolio. What does affect realized return is mostly the yield path's convexity or concavity through time, with negative forecast errors occurring when the yield path was generally convex and positive forecast errors occurring when the yield path was generally concave. Other factors affecting realized return are the inaccuracies in simple return approximations (which especially affect long-term bonds) and shifts in the steepness of the yield curve (which especially affect short-term bonds). If the yield path had no convexity or concavity, there was no inaccuracy in simple return approximations, and yield curves were always flat, then the arithmetic mean continuously-compounded return of the portfolio excluding rolldown return would, over this horizon, be equal to its initial continuously-compounded yield. If the yield curve's steepness was not zero but was unchanging, the arithmetic mean continuously-compounded return of the portfolio including rolldown return would, over this horizon, be equal to its initial continuously-compounded rolling yield. A horizon of twice duration gave relatively low annualized forecasting errors in most of our samples of historical yield paths with bonds turned over monthly, despite all but one of the paths being constant-maturity instead of constant-duration, although the corresponding cumulative errors for long-term bonds were sometimes considerable.

The extent of bond returns' predictability demonstrated in this paper may make investment-grade constant-duration or constant-maturity portfolios appear less risky to long-term investors who thought that the portfolios' short-term variance of return was a good measure of risk; it probably will make such portfolios appear more risky to long-term investors who thought that it was only necessary to wait until the bonds' duration had passed in order for initial yield to be a good forecast of average return.

APPENDIX

Proof of Proposition 1. Denote end-of-period quantities with a superscript e. For the periodically-compounded case, if the one-period return is denoted by R_t then after one period wealth is $(1+R_t)\,PV(Y_t)$. Also wealth after one period is equal to principal $PV(Y_t)$ plus interest C plus capital gains $PV^e(Y_{t+1}) - PV^e(Y_t)$. Assuming straight bonds, without loss of generality consider only bonds bought "at par," that is, with the rate of discount equal to the yield; thus $C = Y_t \cdot PV(Y_t)$, where the first Y_t is the yield and the second Y_t is the rate of discount. It follows from this and $PV^e(Y_t) = (1+Y_t)\,PV(Y_t) - C = (1+Y_t)\,PV(Y_t) - Y_t \cdot PV(Y_t) = PV(Y_t)$ that

$$R_t = Y_t + \left[\frac{PV^e(Y_{t+1})}{PV^e(Y_t)} - 1\right]\frac{PV^e(Y_t)}{PV(Y_t)} = Y_t + \frac{PV^e(Y_{t+1})}{PV^e(Y_t)} - 1\,. \tag{15}$$

Expanding $PV^e(Y_{t+1})$ in a first-order Taylor Series around Y_t and using the definition of D^e gives $PV^e(Y_{t+1})/PV^e(Y_t) \approx 1 - D^e\,\Delta Y_t$ for small ΔY, which leads to $R \approx Y - D^e\,\Delta Y_t$. All that remains to show is $D^e \approx D - 1$.

The "Macauley duration for periodic compounding" at the beginning of the period, which is defined to be

$$\frac{\frac{C_1}{1+Y} + \frac{2C_2}{(1+Y)^2} + \frac{3C_3}{(1+Y)^3} + \cdots}{\frac{C_1}{1+Y} + \frac{C_2}{(1+Y)^2} + \frac{C_3}{(1+Y)^3} + \cdots}\quad, \tag{16}$$

will be approximately equal to "one plus the Macauley duration for periodic compounding" at the end of the period,

$$1 + \frac{\frac{C_2}{1+Y} + \frac{2C_3}{(1+Y)^2} + \cdots}{\frac{C_2}{1+Y} + \frac{C_3}{(1+Y)^2} + \cdots} = \left[\frac{\frac{C_2}{1+Y} + \frac{C_3}{(1+Y)^2} + \cdots}{\frac{C_2}{1+Y} + \frac{C_3}{(1+Y)^2} + \cdots} + \frac{\frac{C_2}{1+Y} + \frac{2C_3}{(1+Y)^2} + \cdots}{\frac{C_2}{1+Y} + \frac{C_3}{(1+Y)^2} + \cdots}\right] \cdot \frac{\frac{1}{Y}}{\frac{1}{Y}}$$

$$= \frac{\frac{2C_2}{(1+Y)^2} + \frac{3C_3}{(1+Y)^3} + \cdots}{\frac{C_2}{(1+Y)^2} + \frac{C_3}{(1+Y)^3} + \cdots} \quad ,$$

as long as, writing the above in shorthand, $\frac{x+A}{x+B} \approx \frac{A}{B}$ with $x = C_1/(1+Y)$. Note that $A > B$ since it cannot be true that "for $t \geq 2$ all the C_t's are zero" because that would imply $D = 1$. Define $f(x+A, x+B) = (x+A)/(x+B)$, then expand

$$f(x+A, x+B) \approx f(A,B) + \left.\frac{\partial f(A,B)}{\partial A}\right|_{A,B} x + \left.\frac{\partial f(A,B)}{\partial B}\right|_{A,B} x$$
$$= \frac{A}{B} + \frac{x}{B} - \frac{Ax}{B^2} = \frac{A}{B}\big(1 - \frac{x}{B}\big) + \frac{x}{B} \, .$$

The approximation is good when $1 - x/B \approx 1$, which is equivalent to $x \ll B$, and when $A/B \gg x/B$, which is equivalent to $x \ll A$. Since $A > B$, only $x \ll B$ needs to be satisfied. For straight par bonds, one has $C = Y \cdot PV(Y)$, so

$$\frac{x}{B} = \frac{C_1/(1+Y)}{\frac{C_2}{1+Y} + \frac{C_3}{(1+Y)^2} + \cdots} = \frac{C_1}{PV(Y) - \frac{C_1}{1+Y}} = \frac{Y \cdot PV(Y)}{PV(Y) - \frac{Y \cdot PV(Y)}{1+Y}} = Y\,(1+Y)$$

and therefore $(x+A)/(x+B) \approx A/B$ is true when $1 \approx 1 - x/B = 1 - Y - Y^2$, which is true to zeroth order when Y is small. Given that the "Macauley Duration for periodic compounding" changes by approximately one, the Modified Duration for this periodically-compounded bond will by footnote 2 change by approximately $1/(1+Y)$, which is $1 - Y$ to first order but simply 1 to zeroth order, which again is applicable for small Y.

For the continuously-compounded case, after one period wealth is $e^{R_t}\, PV(Y_t)$ and it is also equal to principal $PV(Y_t)$ plus interest $C = (e^{Y_t} - 1) \cdot PV(Y_t)$ (assuming a par bond) plus capital gains $PV^e(Y_{t+1}) - PV^e(Y_t)$. It follows from this and $PV^e(Y_t) = e^{Y_t}\, PV(Y_t) - C = e^{Y_t}\, PV(Y_t) - (e^{Y_t} - 1) \cdot PV(Y_t) = PV(Y_t)$ that

$$e^{R_t} = e^{Y_t} + \frac{PV^e(Y_{t+1}) - PV^e(Y_t)}{PV(Y_t)} = e^{Y_t} + \frac{PV^e(Y_{t+1})}{PV^e(Y_t)} - 1 \Rightarrow \tag{17}$$

$$R_t = \ln\left[e^{Y_t} - 1 + \frac{PV^e(Y_{t+1})}{PV^e(Y_t)}\right] \tag{18}$$
$$\approx \ln\left[Y_t + \frac{PV^e(Y_{t+1})}{PV^e(Y_t)}\right]$$

using for the last line the first-order approximation $e^x \approx 1 + x$. Expanding $PV^e(Y_{t+1})$ in a first-order Taylor Series around Y_t as before gives $PV^e(Y_{t+1})/PV^e(Y_t) \approx 1 - D^e\, \Delta Y$ for small ΔY, which leads to $R \approx \ln[1 + Y_t - D^e\, \Delta Y_t]$. Using the first-order Taylor Series expansion $\ln(1+x) \approx x$ for small "x" gives $R \approx Y_t - D^e\, \Delta Y_t$. All that remains to show is $D^e \approx D - 1$. This can be shown similarly to the periodically-compounded case, but it is easier than there because, as shown in footnote 2, for continuous compounding Macauley Duration is equal to Modified Duration, so there is no need in the continuous-compounding case to use the $1/(1+Y) \approx 1$ approximation.[21] ∎

[21] In their theoretical work Leibowitz et al. use zero-coupon bonds, for which all the C's but one are zero. This simplifies the derivation of "$D - 1$" because with $C_1 = 0$ the passage of one period reduces Macauley Duration by exactly one period, meaning there is no need to make the first zeroth-order approximation of small Y. In their post-2000 papers Leibowitz et al. do not systematically treat the continuously-compounded case.

Proof of Proposition 3. From (3),

$$\overline{R}_a(N) = \frac{1}{N}\sum_{t=1}^{N} R_t = \frac{1}{N}\sum_{t=1}^{N}\Big(Y_t - (D-1)\,\Delta Y_t\Big)\,. \tag{19}$$

Thus

$$\begin{aligned} N \cdot \overline{R}_a &= \sum_{t=1}^{N}[Y_t - (D-1)\Delta Y_t] = \sum_{t=1}^{N} Y_t - (D-1)\sum_{t=1}^{N} Y_{t+1} + (D-1)\sum_{t=1}^{N} Y_t \\ &= -D\sum_{t=1}^{N} Y_{t+1} + \sum_{t=1}^{N} Y_{t+1} + D\sum_{t=1}^{N} Y_t\,; \end{aligned}$$

expanding the first and third terms and cancelling,

$$= D\,(Y_1 - Y_{N+1}) + \sum_{t=1}^{N} Y_{t+1}\,.$$

Add and subtract Y_1 and expand the last term, then divide by N. ▮

Proof of Corollary. From (8),

$$\lim_{N\to\infty} \overline{R}_a(N) = -\lim_{N\to\infty}\frac{Y_{N+1}}{N} + \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N} Y_t\,.$$

(4) implies $\lim_{N\to\infty} Y_{N+1}/N = \lim_{N\to\infty}(Y_1/N + \Delta Y) = \Delta Y$. ▮

Proof of Proposition 4. Let m be the maturity at the initial date of purchase t. Assume without loss of generality that the bond has a par value of 1000. Its present value then is $PV(Y,m) = \sum_{t=1}^{m} Ce^{-tY} + 1000e^{-mY} = C\,(1-e^{-mY})/(e^{Y}-1) + 1000e^{-mY}$. Originally the bonds are purchased at par, making $C = (e^{Y_t}-1)\,PV(Y_t,m) = (e^{Y_t}-1)\cdot 1000$. Thus $PV(Y,m) = 1000\,(1-e^{-mY})(e^{Y_t}-1)/(e^{Y}-1) + 1000e^{-mY}$. That leads to $PV^e(Y_t) = PV(Y_t,m-1) = 1000$ and $PV^e(Y_{t+1}) = PV(Y_{t+1},m-1) = 1000\,(1-e^{-(m-1)Y_{t+1}})(e^{Y_t}-1)/(e^{Y_{t+1}}-1) + 1000e^{-(m-1)Y_{t+1}}$. Use (18).

The Modified Duration was found by differentiating this proof's expression for $PV(Y,m)$ with respect to Y, forming $-(1/PV)\,\partial PV/\partial Y$, and evaluating it at $C = (e^{Y_t}-1)\,PV(Y_t)$ and $Y = Y_t$. ▮

Proof of Table 2's Column A Rows 1 and 3. From (19), $\overline{R}_a - Y_1$ has the same sign as $\sum_{t=1}^{N}(t-D)\,\Delta Y$. Thus for small N and $\Delta Y > 0$, $\overline{R}_a < Y_1$, whereas for small N and $\Delta Y < 0$, $\overline{R}_a > Y_1$. In case 2A, $\Delta Y > 0$, so $\overline{R}_a$ will be less than Y_1 if N_a is set to $\lfloor 2D-1\rfloor$ and larger than Y_1 if N_a is set to $1 + \lfloor 2D-1\rfloor$, which is $\lceil 2D-1\rceil$. Therefore $\overline{R}_a$ will never be exactly equal to initial yield, but $\overline{R}_a$ will be slightly less than initial yield in period $\lfloor 2D-1\rfloor$ and slightly more than initial yield in period $\lceil 2D-1\rceil$. Defining N_a^- as the largest date when $\overline{R}_a$ lies on the same side of Y_1 as it did at $t=1$, and N_a^+ as the smallest date when $\overline{R}_a$ does not lie on the same side of Y_1 as it did at $t=1$, we have $N_a^- = \lfloor 2D-1\rfloor$ and $N_a^+ = \lceil 2D-1\rceil$, proving 2A. In case 4A, $\Delta Y < 0$, so $\overline{R}_a$ will similarly never be exactly equal to initial yield, but it will be slightly more than initial yield in period $\lfloor 2D-1\rfloor$ and slightly less than initial yield in period $\lceil 2D-1\rceil$. Therefore in case 4A, as in case 2A, $N_a^- = \lfloor 2D-1\rfloor$ and $N_a^+ = \lceil 2D-1\rceil$. ▮

Proof of Proposition 6. The geometric mean return is

$$\overline{R}_g = \left[\prod_{t=1}^{N_g}(1+R_t)\right]^{1/N_g} - 1 = \left[\prod_{t=1}^{N_g}\Big(1 + Y_1 + (t-D)\,\Delta Y\Big)\right]^{1/N_g} - 1\,. \tag{20}$$

Setting $\overline{R}_g = Y_1$ and adding one to both sides of the equation means that N_g satisfies

$$1 + Y_1 = \left[\prod_{t=1}^{N_g} \Big(1 + Y_1 + (t - D)\, \Delta Y \Big) \right]^{1/N_g} \quad \text{or}$$

$$(1 + Y_1)^{N_g} = \prod_{t=1}^{N_g} \big(1 + Y_1 + (t - D)\, \Delta Y \big) \tag{21}$$

which leads to (13).

Define $A = 1 - D\,\Delta Y/(1 + Y_1)$ and $B = \Delta Y/(1 + Y_1)$, so that (13) becomes $1 = \prod_{t=1}^{N_g}(A + Bt)$. Since using (6), $A + Bt = (1 + R_t)/(1 + Y_t)$, we know that $0 < A + Bt$ for all t. The latter implies $0 < A + Bt = B \cdot (\frac{A}{B} + t)$, so either

$\frac{A}{B} + t > 0$ for all t and $B > 0$, i.e., $\Delta Y > 0$, or
$\frac{A}{B} + t < 0$ for all t and $B < 0$, i.e., $\Delta Y < 0$.

For the $\Delta Y > 0$ case, (22) follows by using the identity $\sum_{t=1}^{T} f(t{+}C) = \sum_{j=1+C}^{T+C} f(j)$ and setting its C equal to $\frac{A}{B}$ and its f equal to ln.

$$\exp \ln \prod_{t=1}^{N_g}(A + Bt) = \exp \sum_{t=1}^{N_g} \ln(A + Bt) = \exp \sum_{t=1}^{N_g} \Big(\ln B + \ln(\tfrac{A}{B}{+}t) \Big)$$

$$= \exp\{N_g \ln B + \sum\nolimits_{j=\frac{A}{B}+1}^{\frac{A}{B}+N_g} \ln j\,\}\,; \tag{22}$$

from the Lemma given in the main body of the paper,

$$= \exp\{N_g \ln B + \ln \Gamma(\tfrac{A}{B}{+}N_g{+}1) - \ln \Gamma(\tfrac{A}{B}{+}1)\,\} \tag{23}$$

$$= \exp \ln \big[B^{N_g}\, \Gamma(\tfrac{A}{B}{+}N_g{+}1) \big/ \Gamma(\tfrac{A}{B}{+}1) \big]$$

$$= B^{N_g}\, \Gamma(\tfrac{A}{B}{+}N_g{+}1) \big/ \Gamma(\tfrac{A}{B}{+}1)\,, \tag{24}$$

as was to be shown. For the $\Delta Y < 0$ case, using the identity $\sum_{t=1}^{T} f(-t{+}C) = \sum_{j=-T+C}^{-1+C} f(j)$ one has

$$\exp \ln \prod_{t=1}^{N_g}(A + Bt) = \exp \sum_{t=1}^{N_g} \ln(A + Bt) = \exp \sum_{t=1}^{N_g} \Big(\ln(-B) + \ln(-\tfrac{A}{B} - t) \Big)$$

$$= \exp\{N_g \ln(-B) + \sum\nolimits_{j=-\frac{A}{B}-N_g}^{-\frac{A}{B}-1} \ln j\,\}$$

$$= \exp\{N_g \ln(-B) + \ln \Gamma(-\tfrac{A}{B}) - \ln \Gamma(-\tfrac{A}{B} - N_g)\,\}$$

$$= \exp \ln \big[(-B)^{N_g}\, \Gamma(-\tfrac{A}{B}) \big/ \Gamma(-\tfrac{A}{B} - N_g) \big]$$

$$= (-B)^{N_g}\, \Gamma(-\tfrac{A}{B}) \big/ \Gamma(-\tfrac{A}{B} - N_g)\,.$$

∎

Proof of Proposition 7. Denote the geometric mean return up to time N as $\overline{R}_g(N) = [\prod_{t=1}^{N}(1 + R_t)]^{1/N} - 1$ and define "one plus R" $OPR(N) = \overline{R}_g(N) + 1$.

Case 1G's first inequality: From Proposition 2, $N_a = 2D - 1$; that is, $\overline{R}_a = Y_1$ when $t = 2D - 1$. Since the geometric mean is less than or equal to the arithmetic mean,[22] as long as we can show

[22] A straightforward way to show that the famous "arithmetic mean-geometric mean inequality" of mathematics, $[\prod_{i=1}^{n} x_i]^{1/n} \leq \sum_{i=1}^{n} x_i/n$, applies also to the geometric mean of finance, $[\prod_{t=1}^{N_g}(1 + x_t)]^{1/N_g} - 1$, is to rewrite the latter's " $-$ 1" term as $-\sum_{i=1}^{n} 1/n$ and add its opposite to both sides of the inequality.

that the geometric mean is increasing in N, it is clear that the geometric mean will take longer than $2D - 1$ periods to reach Y_1.

According to (6), R_t is strictly increasing in t. Hence

$$OPR(N) = \prod_{t=1}^{N}(1 + R_t)^{1/N} < \prod_{t=1}^{N}(1 + R_N)^{1/N} = 1 + R_N \,. \tag{25}$$

To prove that $\overline{R}_g(N)$ is increasing in N, it suffices to show that the following is positive:

$$\begin{aligned}
\ln OPR(N+1) - \ln OPR(N) &= \tfrac{1}{N+1} \sum_{t=1}^{N+1} \ln(1 + R_t) - \tfrac{1}{N} \sum_{t=1}^{N} \ln(1 + R_t) \\
&= \tfrac{1}{N+1} \ln(1 + R_{N+1}) + \left[\tfrac{1}{N+1} - \tfrac{1}{N}\right] \sum_{t=1}^{N} \ln(1 + R_t) \\
&= \tfrac{1}{N+1} \ln(1 + R_{N+1}) + \tfrac{-1}{N(N+1)} \sum_{t=1}^{N} \ln(1 + R_t) \\
&= \tfrac{1}{N+1} \left[\ln(1 + R_{N+1}) - \tfrac{1}{N} \sum_{t=1}^{N} \ln(1 + R_t)\right] \\
&= \tfrac{1}{N+1} \left[\ln(1 + R_{N+1}) - \ln OPR(N)\right] \qquad (26) \\
&> \tfrac{1}{N+1} \left[\ln(1 + R_N) - \ln OPR(N)\right] .
\end{aligned}$$

This is positive by (25).

Case 1G's second inequality: When $\Delta Y > 0$, $\overline{R}_g(1)$ is less than $(1 + Y_1)$; if we can show that $\overline{R}_g(N)$ becomes larger than Y_1 as $N \to \infty$, this will prove that $N_g^+ < \infty$. To prove that $\overline{R}_g(N)$ becomes larger than Y_1 as $N \to \infty$, recall that $R_D = Y_1$ and R_t is increasing. Pick a date τ such that $\tau > D$, and let $Y_2 = R_\tau$. We know that $Y_2 > Y_1$; let their difference be K. The left-hand side of the following inequality is the OPR of R_t up to time $N > \tau$; the right-hand side is the OPR of a return path which equals R_t up to time τ, then becomes Y_2 forever (the "kinked path"):

$$\prod_{t=1}^{N}(1 + R_t)^{1/N} > \prod_{t=1}^{\tau}(1 + R_t)^{1/N} \prod_{t=\tau+1}^{N} (1 + Y_2)^{1/N} \,. \tag{27}$$

The ratio of the OPR of the kinked path to the OPR of an $R_t \equiv Y_2$ path is

$$\left(\prod_{t=1}^{\tau} \tfrac{1+R_t}{1+Y_2}\right)^{1/N} \tag{28}$$

since for $t > \tau$ the two paths are the same. The fraction in (28) is less than one. The limit of (28) as $N \to \infty$ is one. Hence for sufficiently large N, the geometric mean of the kinked path can be made arbitrarily close to the geometric mean of the "always Y_2" path—and in particular, for sufficiently large N the geometric mean of the kinked path can be made closer than K to the geometric mean of the "always Y_2" path, which is Y_2. For such N, the geometric mean of the kinked path must be larger than Y_1; and hence from (27), the geometric mean of the original path must be larger than Y_1.

Case 3G's second inequality: From Proposition 2, $N_a = 2D - 1$; that is, $\overline{R}_a = Y_1$ when $t = 2D - 1$. Since the geometric mean is less than or equal to the arithmetic mean, as long as we can show that the geometric mean is decreasing in N, it is clear that the geometric mean will be equal to Y_1 at an earlier date than $2D - 1$.

According to (6), R_t is strictly decreasing in t. Hence

$$OPR(N) = \prod_{t=1}^{N}(1+R_t)^{1/N} > \prod_{t=1}^{N}(1+R_N)^{1/N} = 1 + R_N\ . \tag{29}$$

To prove that $\overline{R}_g(N)$ is decreasing in N, it suffices to show that the following is negative (using (26)):

$$\begin{aligned}\ln OPR(N+1) - \ln OPR(N) &= \tfrac{1}{N+1}\left[\ln(1+R_{N+1}) - \ln OPR(N)\right] \\ &< \tfrac{1}{N+1}\left[\ln(1+R_N) - \ln OPR(N)\right] .\end{aligned}$$

This is negative by (29).

($N_g^+ = 2D-1$ is possible in the latter case because although $\overline{R}_g(2D-1)$ is too small, $\overline{R}_g(2D-2)$ may be too large, making $N_g^+ = 2D-1$ and $N_g^- = 2D-2$.)

Case 3G's first inequality: When $\Delta Y < 0$, $\overline{R}_g(D)$ is greater than Y_1:

$$OPR(D) = \sum_{t=1}^{D}(1+R_t)^{1/D} > \sum_{t=1}^{D}(1+R_D)^{1/D} = 1 + R_D = 1 + Y_1\ .$$

Since $\overline{R}_g(N)$ is decreasing in N, more time must pass before $\overline{R}_g$ is equal to Y_1.

Case 2G's first inequality: Note that D cannot be an integer when $2D$ is not an integer. Thus $OPR(2\lfloor D\rfloor-1)^{2\lfloor D\rfloor-1}$ is equal to

$$\prod_{t=1}^{2\lfloor D\rfloor-1}[1+Y_1+(t-D)\Delta Y] < \prod_{t=1}^{2\lfloor D\rfloor-1}[1+Y_1+(t-\lfloor D\rfloor)\Delta Y] < (1+Y_1)^{2\lfloor D\rfloor-1}$$

using Case 1G. Similarly for 4G:

$$\prod_{t=1}^{2\lceil D\rceil-1}[1+Y_1+(t-D)\Delta Y] < \prod_{t=1}^{2\lceil D\rceil-1}[1+Y_1+(t-\lceil D\rceil)\Delta Y] < (1+Y_1)^{2\lceil D\rceil-1}\ .$$

▮

Proof of Proposition 1′. (17) means explicitly in this case

$$e^{R_t} = e^{Y_t(D)} + \frac{PV(Y_{t+1}(D^e), m-1) - PV(Y_t(D), m)}{PV(Y_t(D), m)}\ . \tag{30}$$

As in the proof of (18), for par bonds $PV^e(Y_t) = PV(Y_t) \iff PV(Y, m-1) = PV(Y, m)$. The second term on the right-hand side of (30) thus can be approximated to first order by

$$\begin{aligned}&\frac{1}{PV(Y_t(D), m-1)}\ \frac{\partial PV}{\partial Y}\bigg|_{(Y_t(D), m-1)} \cdot [Y_{t+1}(D^e) - Y_t(D)] \\ &\quad = -D(Y_t, m-1)\cdot\Big\{[Y_{t+1}(D^e) - Y_{t+1}(D)] + [Y_{t+1}(D) - Y_t(D)]\Big\}\ .\end{aligned}$$

Approximating $D(Y_t, m-1)$ and D^e by $D-1$ as in the proof of Proposition 1 and $e^{Y_t(D)}$ by $1 + Y_t(D)$, (30) implies

$$\begin{aligned}R_t &\approx \ln\big[1 + Y_t(D) - (D-1)[Y_{t+1}(D-1) - Y_{t+1}(D)] - (D-1)\Delta Y_t\big] \\ &= \ln\left[1 + RY_t - (D-1)\Delta Y_t\right]\ ;\end{aligned}$$

use $\ln(1+x) \approx x$. ▮

BIBLIOGRAPHY

Bierwag, Gerald O., Iraj J. Fooladi, and Gordon S. Roberts (2000), "Risk Management with Duration: Potential and Limitations," *Canadian Journal of Administrative Sciences* **17/2**: 126–142.

Bring, Johan (1996), "A Geometric Approach to Compare Variables in a Regression Model," *The American Statistician* **50/1**: 57–62.

Bova, Anthony (2013), "Hypothetical Equity Returns Models under Rising Rates." Society of Actuaries Investment Symposium, March 14–15, Session R5, "A Fresh Look at the Equity Risk Premium." At `http://www.soa.org/Files/Pd/2013/investment-sym/2013-ny-invest-sym-r5.pdf`.

Chevan, Albert, and Michael Sutherland (1991), "Hierarchical Partitioning," *The American Statistician* **45/2**: 90–96.

Eisenhauer, Joseph G. (2003), "Regression Through the Origin," *Teaching Statistics* **25/3**: 76–80.

Emancipator, Kenneth, and Martin H. Kroll (1993), "A Quantitative Measure of Nonlinearity," *Clinical Chemistry* **39/5**: 766–772.

Fridson, Martin S., and Xiaoyi Xu (2014), "Duration Targeting: No Magic for High-Yield Investors," *Financial Analysts Journal* **70/3**: 28–33.

Gay, Chris (March 3, 2014), "How to Predict the Next Decade's Bond Returns; Current Yields Are a Good Indicator of What You Will Earn Over Time." *Wall Street Journal*.

Gleich, David (2005), "Finite Calculus: A Tutorial for Solving Nasty Sums." At `https://www.cs.purdue.edu/homes/dgleich/publications/finite-calculus.pdf`.

Grömping, Ulrike (2012), "Estimators of Relative Importance in Linear Regression Based on Variance Decomposition," *The American Statistician* **61/2**: 193–147.

Hall, Robert E. (2001), "The Stock Market and Capital Accumulation," *American Economic Review* **91/5**: 1185–1202.

Hicks, John R. (1939), *Value and Capital*. Oxford: Oxford University Press.

Ivaschenko, Iryna (January 2003), "How Much Leverage is Too Much, or Does Corporate Risk Determine the Severity of a Recession? (IMF Working Paper)." At `www.imf.org/external/pubs/ft/wp/2003/wp0303.pdf`.

Langeteig, Terence C., Martin L. Leibowitz and Stanley Kogelman (1990), "Duration Targeting and the Management of Multiperiod Returns," *Financial Analysts Journal* **46/5**: 35–45.

Leibowitz, Martin, and Anthony Bova (December 18, 2012), "Duration Targeting: A New Look at Bond Portfolios (Morgan Stanley Research)." At `www.q-group.org/wp-content/uploads/2014/01/2013sp-Leibowitz-Bova-Duration-Targeting-121812.pdf`.

Leibowitz, Martin, and Anthony Bova (2014), "Portfolio Strategy: Rolling Yields and Return Convergence," *Morgan Stanley Investment Management Journal* **4/2**: 19–30.

Leibowitz, Martin, and Anthony Bova (2013), "Duration Targeting and Index Convergence," *Morgan Stanley Investment Management Journal* **3/1**: 73–80.

Leibowitz, Martin L., Anthony Bova, and Stanley Kogelman (2014), "Long-Term Bond Returns under Duration Targeting," *Financial Analysts Journal* **70/1**: 31–51.

Leibowitz, Martin L., Anthony Bova, Stanley Kogelman, and Sidney Homer (2013), *Inside the Yield Book: The Classic That Created the Science of Bond Analysis (3rd Edition)*. Somerset, New Jersey: John Wiley & Sons.

Lindeman, Richard Harold, Peter Francis Merenda, and Ruth Z. Gold (1980), *Introduction to Bivariate and Multivariate Analysis*. Glenview, Illinois: Scott, Foresman.

Macaulay, Frederick R. (1938), *Some theoretical problems suggested by the movements of interest rates, bond yields, and stock prices in the US. since 1856*. New York: National Bureau of Economic Research. At http://www.nber.org/authors/frederick_macaulay.

Mindlin, Dimitry, (August 14, 2011), "On the Relationship between Arithmetic and Geometric Returns." At http://ssrn.com/abstract=2083915 or http://dx.doi.org/10.2139/ssrn.2083915.

McNabb, William (August 5, 2014), "Taking Stock: A Conversation with Vanguard CEO Bill McNabb." At https://institutional.vanguard.com/VGApp/iip/site/institutional/researchcommentary/article/InvComTakingStock.

Naik, Vipaul (no date), "Summation Notation, Series, and Convergence Basics." At files.vipulnaik.com/math-153-sequence/seriesandconvergence.pdf

Pianca, Paolo (2006), "Maximum Duration of below par Bonds: a Closed-form Formula," *Mathematical Methods in Economics and Finance* **1**:57–64.

Potts, Tom, and William Reichenstein (November 2004), "Predictability of Fixed-Income Fund Returns," *Journal of Financial Planning* **17/11**:62–71.

Redington, F.M. (1952), "Review of the Principles of Life-Office Valuations," *Journal of the Institute of Actuaries (1886-1994)* **78/3**:286–340.

Samuelson, Paul A. (March 1945), "The Effect of Interest Rate Increases on the Banking System," *American Economic Review* **35/1**: 16–27.

Thomas, Charles, and Paul M. Bosse (July 2014), "Understanding the 'Hedge Return': The Impact of Currency Hedging in Foreign Bonds." Vanguard Research. At https://advisors.vanguard.com/VGApp/iip/site/advisor/researchcommentary/article/IWE_InvResForeignBondCurrencyHedging.

Villazón, César (1991), "Bond Duration, Yield to Maturity and Bifurcation Analysis." Second AFIR Colloquium. At www.actuaries.org/AFIR/Colloquia/Brighton/Villazon.pdf.

Wooldridge, Jeffrey (2012), *Introductory Econometrics: A Modern Approach (5th edition)*. Mason, Ohio: South-Western Cengage Learning.

Yogaranpan, Yanagan (July 2005), "Expanding the Geometric Mean Return." Research Paper No. 2005/04, Division of Economic and Financial Studies, Macquarie University. At www.businessandeconomics.mq.edu.au/our_departments/Applied_Finance_and_Actuarial_Studies/acst_docs/research_papers/2005/200504_Geometric_Mean_Return.pdf.