Appeared in Quantitative Finance, 1–19. https://doi.org/
10.1080/14697688.2024.2386323

GDP-linked bonds as a new asset class

Ellie Papavassiliou* Nikolas Topaloglou† Stavros A. Zenios‡

January 2024

Abstract

Using stochastic spanning tests without any distributional assumptions on returns, we show that the two classes of GDP-linked bonds, floaters and linkers, are not spanned by a broad benchmark set of stocks, bonds, and cash for a wide range of design specifications. Thus, they provide a new asset class with significant diversification benefits for investors, with proportional investments to these novel instruments estimated in the double digits and an increase in Sharpe ratios by up to 0.37 over the benchmark. The benefits depend on the market risk premium, but they persist for a wide range of premia estimates from existing literature and are robust to a randomized test. Using the generalized method of moments regressions, we document the finance and macro determinants of GDP-linked bond returns.

Keywords: finance; stochastic spanning; contingent debt; risk premium; diversification.

Acknowledgements. We acknowledge comments and suggestions from Alexandros Bechioulis, Claire Economidou, and Matthew Lyle and participants at the 7th Paris Financial Management Conference. We are grateful to Fernando Eguren Martin for providing us with a time series of risk premia for one of our tests.

*Athens University of Economics and Business, 76, Patision Street, GR10434, Athens, Greece; e-mail: papavaselli@aueb.gr.

†IPAG Business School & Athens University of Economics and Business, 76, Patision Street, GR10434, Athens, Greece; e-mail: nikolas@aueb.gr.

‡Durham University Business School, Millhill Ln, Durham DH1 3LB, UK; University of Cyprus, P.O. Box 20537, 1678 Nicosia; Cyprus Academy of Sciences, Letters, and Arts; Bruegel, Brussels; e-mail: zenios.stavros@ucy.ac.cy. Financial support was provided, in part, by the Cyprus Academy of Sciences, Letters, and Arts.

1 Introduction

Sovereign contingent debt instruments received international attention at the 2016 G20 meeting, with the IMF issuing a comprehensive report (IMF, 2017) on their design and the potential benefits for public finance. GDP-linked bonds make debt payments contingent on a country's GDP.[1] They provide sovereign insurance from negative growth shocks (Froot, Scharfstein, and Stein, 1989), allow for taxation smoothing over the economic cycle (Barro, 2003a), and can improve the functioning of the international financial system (Barr, Bush, and Pienkowski, 2014). We take the investors' viewpoint in this paper uniquely among existing literature. We ask whether GDP-linked bonds provide diversification benefits to investor portfolios. GDP-linked bonds would provide investors an opportunity to take an equity-like position on a country's future growth prospects (Kamstra and Shiller, 2009).[2] This raises the question of whether these instruments would indeed provide diversification benefits. We answer affirmatively, supporting the issuance of GDP-linked bonds from a buyers' perspective.

We construct optimal portfolios containing GDP-linked bonds and compare them with broad benchmark portfolios of international stocks, bonds, and cash. We consider the two types of GDP-linked bonds from the literature, *floaters* (Borensztein and Mauro, 2004) and *linkers* (Kamstra and Shiller, 2009).[3] Using in-sample stochastic spanning tests (Arvanitis, Hallam, Post, and Topaloglou, 2019) we construct optimal portfolios with and without GDP-linked bonds. The stochastic spanning tests are model-free, so we can draw inferences without relying on distributional assumptions and accounting for returns' positive skewness and kurtosis. We find that a wide range of designs for floaters or linkers are not spanned. We then use out-of-sample tests to show that an investment possibility set with floaters or linkers contains portfolios that stochastically dominate all benchmark portfolios (Arvanitis, Post, and Topaloglou, 2021), suggesting that GDP-linked bonds are beneficial for investors with convex, non-increasing risk preferences.

Our work is motivated by recent pricing models documenting that the returns of these instruments depend critically on their design (Consiglio and Zenios, 2018). Whereas it has been shown that these instruments can be attractive to issuers for a wide range of designs, it is unclear priori which designs can be attractive to investors or if designs exist that attract both issuers and investors. We, therefore, test different design parameters. Our finding holds for a wide (and reasonable) range of parameters

[1] GDP-linked bonds originated in the works of Froot, Scharfstein, and Stein (1989); Krugman (1988), with more recent supportive arguments from Kamstra and Shiller (2009). For a brief history, see Borensztein and Mauro (2004) and Benford et al. (2018) for a collection of recent works on state-of-the-art GDP-linked bonds.

[2] Potential investors that are or could be interested in GDP-linked bonds are hedge funds, pension funds, and other institutional investors. For pension funds, these instruments could give them a stake in the upside of growth of emerging markets with the benefits of international diversification.

[3] A floater is a coupon-indexed floating rate bond with a coupon rate varying according to the country's GDP, and a linker is a zero-coupon bond that pays at maturity a fraction of the country's GDP (e.g., one trillionth in the reference).

similar to those identified for issuers. The returns of these instruments will depend on the risk premia at which they trade (Benford, Best, and Joy, 2016; Eguren-Martin, Meldrum, and Yan, 2020), and we also test for different market risk premia estimates from the literature and perform a randomized test as well. Our findings are robust to a wide range of premia.

Our answer informs the policy debate initiated by the G20. To the extent that GDP-linked bonds are helpful for public finance, we show that they can find willing investors. One crucial challenge is developing a robust market for such instruments that require, among others, an attractive relation between return and risk (IMF, 2017). The Bank of England put together a prototype term sheet of GDP-linked bonds which they found to enjoy support from potential investor groups.[4] We document that such instruments can enhance the risk-return profile of investor portfolios. Investors would invest proportionately up to double digits in these novel instruments and achieve Sharpe ratios up to 0.37 higher than the benchmark. We also find that floaters seem to have an edge over linkers.

In the final step, we use the generalized method of moments to identify the finance and macro factors that drive the performance of GDP-linked bonds. Financial risk variables (term and default spread, FFR) as well as macro variables (public debt, inflation, and capacity utilization rate) are identified as determinants of GDP-linked bonds returns.

Since GDP shares have not been issued massively, the ongoing debate on their potential benefits relies on inferences from calibrated models. Recent contributions look at the potential benefits from the point of view of issuing sovereigns. Borensztein and Mauro (2004) show that GDP-linked bonds allow countries to avoid pro-cyclical fiscal policies and can reduce the likelihood of crisis; Barr et al. (2014) show how they can raise the maximum sustainable sovereign debt level; Blanchard et al. (2016); Kim and Ostry (2018) show that the introduction of GDP-linked bonds in advanced economies could decrease the tail risks of high debt ratios and create fiscal space; Cabrillac et al. (2018) create a counterfactual for the Greek debt crisis to show that converting half of the public debt into GDP-linked bonds in 2009 would have avoided the need for restructuring in 2012; Demertzis and Zenios (2019) show, using simulations for Germany, Greece, Italy, and a Euro area average, that these bonds could provide a market-based insurance mechanism for the eurozone countries. Overall, there is a consensus on the potential benefits of GDP-linked bonds for issuing sovereigns.

However, the benefits will depend on the market's pricing of these instruments. If they are too expensive (i.e., the markets demand a high risk premium), the benefits for the issuing sovereign will erode. Several papers estimate a *risk premium threshold* that makes these instruments attractive for

[4] See https://www.bankofengland.co.uk/-/media/boe/files/events/2015/november/gdp-linked-bonds-london-term-sheet-2.pdf and https://www.allenovery.com/en-gb/global/news-and-insights/publications/gdp-linked-bonds-one-of-the-greatest-innovations-in-sovereign-debt-in-modern-times for the prototype and the support expressed during BoE road shows, respectively.

sovereigns (Barr et al., 2014; Benford et al., 2016; Blanchard et al., 2016). This literature shows that the GDP-linked bonds would still benefit sovereigns for values up to 350bp (250bp for more conservative estimates). Kamstra and Shiller (2009) suggest that the premium of their proposed linker may be of the order of 150bp. These results serve as thresholds when assessing the viability of financing sovereigns with GDP-linked bonds but do not tell us what the market premia would be. Bowman et al. (2016), who employ the CAPM and downside-CAPM for pricing these new instruments, conclude that model-implied prices are highly uncertain. Consiglio and Zenios (2018) provide more precise estimates by showing how to price and hedge GDP-linked bonds in incomplete markets. They point out that the premium will depend on the design features of the new instrument, and they show that a broad range of designs generate premia within the estimated thresholds of the economic studies.[5] For advanced economies, they find bond designs with spreads in the range 100bp, which is the estimate of the Blanchard et al. (2016) study, to 50bp, which is about the premium of U.S. inflation-linked treasury bonds (IMF, 2017). These estimates are well within the thresholds. We should also recognize that giving GDP-linked bonds a privileged (seniority) status leads to an even lower risk premium.

In summary, current literature argues that premia above 250bp make these instruments too expensive for sovereigns, but reasonable bond designs could carry a premium from 50 to 150bp, so, in general, sovereigns can benefit from these instruments. In this paper, we ask the question from the investors' perspective: Do these instruments provide diversification benefits if they are issued with very low premia, i.e., as low as 50bp? Higher premia make them more beneficial for investors but less attractive for sovereigns.

Answering our research question affirmatively for a low premium is encouraging for the potential benefits to investors. It also shows that the benefits to issuers and investors intersect. However, a potential concern remains that the premia are volatile. A recent no-arbitrage attempt to quantify the GDP risk premium (Eguren-Martin et al., 2020) obtains estimates that vary with market conditions, and we will validate our main finding using this unique time series of premia, in addition to using a fixed premium. We go further to conduct randomized tests and find that our results are robust with very high probability (at least 0.90). While we can not rule out that premia will adjust with time in such a way as to fully integrate the GDP-linked bonds into the markets, thus eroding any diversification benefits, our tests on currently available model-implied or market-driven information show that, at least originally, these instruments will be indeed a new asset class. On the positive side,

[5]The design parameters are the base coupon rate and the target GDP growth, and reasonable designs are those that are plausible given the prevailing market conditions. Excessively large target growths will render the GDP-link ineffective, and large base coupon will render the instrument very expensive.

giving GDP-linked bonds a seniority status would significantly lower the risk premium and increase the diversification benefits beyond what we find in our paper.

We recognize a limitation of our analysis that deserves further study. Specifically, we do not account for liquidity or innovation premia that could concern potential market participants. Sufficient liquidity is important for investors, for these instruments to be actively traded, and for issuers, as higher liquidity could reduce the premium demanded by investors. Likewise, an innovation premium makes these instruments more costly for issuers. Our work shows that if GDP-linked bonds are issued at sufficient volume to reduce liquidity concerns, the resulting instruments will contribute to portfolio diversification. In this sense, the evidence we provide supports the arguments favoring coordination to overcome first-mover disadvantages (Demertzis and Zenios, 2019; IMF, 2017), thus lowering or eliminating the liquidity and innovation premia.

We briefly describe our computational strategies for stochastic spanning and bounding tests in Section 2. In Section 3, we test for spanning using in-sample analysis and in Section 4 we carry out dynamic backtesting experiments to obtain out-of-sample results. Section 5 identifies the finance and macro factors that drive GDP-linked bond returns. Section 6 concludes.

2 Stochastic dominance

Stochastic dominance (SD) is a model-free generalization to mean-variance (MV) dominance criterion, see, e.g., Levy (2016); Perrakis (2019); Whang (2019), without assuming a particular family of distributions.[6] It, therefore allows us to draw inferences accounting for the asymmetric and fat-tailed risk profiles of GDP-linked bonds. Second-order stochastic dominance (SSD) ranks investments based on mild non-parametric regularity conditions on the distributions involved and assuming a class of investors with utilities exhibiting non-satiation and risk aversion. SSD is represented by sets of conditions in the form of lower partial moment inequalities between the compared distributions.

2.1 Preliminaries and definitions

We work with a portfolio space defined as the set of positive convex combinations of N benchmark assets and represented by the set $\left\{\boldsymbol{\lambda} \in \mathbb{R}_+^N : \boldsymbol{\lambda}'\mathbf{1}_N = 1\right\}$. The benchmark assets are the vertices of the portfolio space. The returns of the benchmark assets form the random vector $X := (x_1, \dots, x_N)$, assumed to be of bounded support, $\mathcal{X}^N := [\underline{x}, \overline{x}]^N$, $-\infty < \underline{x} < \overline{x} < +\infty$. Let F denote the continuous CDF of X, and $F(y, \boldsymbol{\lambda}) := \int 1(X^{\mathrm{T}}\boldsymbol{\lambda} \leq y)dF(X)$ the marginal CDF for the portfolio.

[6]Representative applications in finance include Constantinides et al. (2011); Hodder et al. (2015); Jackwerth et al. (2009), among others.

Consider the CDF integral $L(x, \boldsymbol{\lambda}; F) := \int_{-\infty}^{x} F(y, \boldsymbol{\lambda})dy$. $L(x, \boldsymbol{\lambda}; F)$ equals the first-order lower-partial moment (LPM), or expected shortfall $\int_{-\infty}^{x}(x-y)dF(y, \boldsymbol{\lambda})$, for each return threshold $x \in \mathcal{X}$ (Bawa 1975). Let $D(x, \boldsymbol{\lambda}, \boldsymbol{\kappa}; F) := L(x, \boldsymbol{\lambda}; F) - L(x, \boldsymbol{\kappa}; F)$, denotes the LPM spread between portfolios $\boldsymbol{\lambda}$ and $\boldsymbol{\kappa}$. Then, $\boldsymbol{\lambda}$ stochastically dominates $\boldsymbol{\kappa}$ by SSD, or $\boldsymbol{\lambda} \succeq_F \boldsymbol{\kappa}$, iff $D(x, \boldsymbol{\lambda}, \boldsymbol{\kappa}; F) \leq 0$, $\forall x \in \mathcal{X}^{\mathbb{N}}$. SSD implies that $\boldsymbol{\lambda} \succeq_F \boldsymbol{\kappa}$ iff $\boldsymbol{\lambda}$ achieves a higher expected utility than $\boldsymbol{\kappa}$ for every increasing and concave utility function.

To test the effects of augmenting the set of benchmark assets with GDP-linked bonds, we consider two subsets of the general portfolio space, $\mathrm{K} \subset \Lambda$, where K is the convex hull of the benchmark assets and Λ is the convex hull of the augmented set of benchmark assets and GDP-linked bonds.

2.2 Stochastic spanning

We use, specifically, the *stochastic spanning* test (Arvanitis et al., 2019), which is the model-free alternative to MV spanning (De Roon et al., 2003; Huberman and Kandel, 1987).

Definition 1. (Stochastic Spanning) K spans Λ by SSD iff for every portfolio $\boldsymbol{\lambda} \in \Lambda$, there exists a portfolio $\boldsymbol{\kappa} \in \mathrm{K}$ that dominates it by SSD, i.e., $\forall x \in \mathcal{X}^{\mathbb{N}}$, $D(x, \boldsymbol{\kappa}, \boldsymbol{\lambda}; F) \leq 0$.

Spanning occurs if introducing new securities (or relaxing investment constraints) does not improve the investment possibility set over the broad class of investor preferences. Hence, stochastic spanning is suitable for checking whether portfolios augmented with GDP-linked bonds dominate a broad market benchmark. If we were to add GDP-linked bonds to a portfolio of stocks, bonds, and cash and fail to reject the spanning hypothesis, these additional bonds would be redundant for any risk-averse investor. We test empirically the null hypothesis H_0 vis-á-vis the alternative H_1:

H_0**:** GDP-linked bonds are spanned by a benchmark set of stocks, bonds, and cash.

H_1**:** There exist some portfolios augmented with GDP-linked bonds that are not spanned by the benchmark assets.

Using the continuity properties of $D(\cdot, \cdot, \cdot; F)$ and the compactness of sets Λ, K, $\mathcal{X}$, we can characterize spanning (Arvanitis et al., 2019), by the following scalar-valued functional of F

$$\eta(F) := \sup_{\Lambda} \inf_{K} \sup_{\mathcal{X}} D(x, \boldsymbol{\kappa}, \boldsymbol{\lambda}; F). \tag{1}$$

Spanning occurs iff $\eta(F) = 0$, while some $\boldsymbol{\lambda} \in \Lambda$ exist that are not second order stochastically dominated by any portfolio $\boldsymbol{\kappa} \in K$ (i.e., reject spanning), iff $\eta(F) > 0$.

2.2.1 Hypothesis structure, test statistic, and critical values

F is latent so $\eta(F)$ is unknown, while the analyst has access to a time series sample of realized returns $(X_t)_{t=1}^{T}$, $X_t \in \mathcal{X}$, t = 1, ..., T, for the benchmark assets. Assuming stationarity[7] and mixing for the benchmark asset return process, an empirical analogue of $\eta(F)$ scaled by $\sqrt{T}$ is used as a Kolmogorov-Smirnov type test statistic for the null

$$\eta_T := \sqrt{T} \sup_{\Lambda} \inf_{\mathrm{K}} \sup_{\mathcal{X}} D(x, \boldsymbol{\kappa}, \boldsymbol{\lambda}; F_T),$$

where F_T denotes the empirical CDF (ECDF) associated with the sample.

The asymptotic decision rule is to reject H_0 in favor of H_1 iff $\eta_T > q(\eta_\infty, 1-\alpha)$, the $(1-\alpha)$ quantile of the distribution of η_∞, for significance level $\alpha \in]0, 1[$. Because the distribution of $q(\eta_\infty, 1-\alpha)$ depends on the underlying distribution, we use the subsampling procedure of Arvanitis et al. (2019) to approximate it by feasible decision rules. Specifically, given the choice of the subsampling rate $1 \leq b_T < T$, we generate the maximally overlapping subsamples $(X_s)_{s=t}^{t+b_T-1}$, $t = 1, \cdots, T - b_T + 1$, evaluate the test statistic on each subsample, thereby obtaining $\eta_{b_T;T,t}$ for $t = 1, \cdots, T - b_T + 1$, resulting to the evaluation of $q_{T,b_T}(1-\alpha)$, the $(1-\alpha)$ quantile of the empirical distribution of $\eta_{b_T;T,t}$ across the subsamples. The modified decision rule is to reject H_0 in favor of H_1 iff $\eta_T > q_{T,b_T}(1-\alpha)$.

2.3 Stochastic bounding

We use stochastic bounding to compare the performance of optimal portfolios augmented with GDP-linked bonds to the benchmark portfolios out-of-sample. The stochastic bounding portfolio dominates any portfolio that can be constructed from a given set with respect to the second stochastic dominance criterion. If a bound does not exist, we identify a portfolio that comes as close as possible to being a bound, an 'approximate bound.' Special attention is given to the portfolio, which minimizes the largest deviation from the lowest feasible levels of the low partial moments.

In the empirical application, we identify the bound (or approximate bound) portfolio separately for the benchmark and the augmented set and compare the realized performance of these portfolios out-of-sample, drawing inferences about potential diversification benefits from the augmentation.

In this case, we set $\Lambda = \mathrm{K}$. So, now we search for a portfolio $\lambda \in \Lambda$ that stochastically dominates every other portfolio in Λ. We use the Arvanitis et al. (2021) method to get the portfolio $\lambda \in \Lambda$ that stochastically bounds portfolio set Λ (where Λ is either the benchmark or the augmented set).

When spanning is rejected, the optimal portfolio λ is the portfolio that dominates all portfolios

[7] Augmented Duckey-Fuller tests on asset returns, GDP-growth rates, and risk premia, reject the null hypothesis of unit roots, so all the processes used in the analysis are stationary.

k for at least one increasing concave utility. Using the stochastic bounding methodology, we get the approximate bound λ, which is as close as possible to the portfolio that dominates all portfolios λ in the augmented set for all increasing concave utilities. When spanning is rejected, the portfolio λ resulting from the spanning methodology would not generally coincide with the portfolio λ from the bounding method. In the out-of-sample analysis, we look every month for the portfolio k, which dominates all other portfolios in the benchmark set, and for portfolio λ, which dominates all other portfolios in the augmented set, and compare their realized performance.

2.4 Computational strategies

Here, we give the computational strategies for our two tests.

Computational Strategy for Spanning

The test statistic η can be represented in terms of expected utility as:

$$\eta(F) := \sup_{\boldsymbol{\lambda}\in\Lambda; u\in\mathcal{U}} \inf_{\boldsymbol{\kappa}\in\mathrm{K}} \mathbb{E}_F \left[u\left(X^{\mathrm{T}}\boldsymbol{\lambda}\right) - u\left(X^{\mathrm{T}}\boldsymbol{\kappa}\right)\right]; \tag{2}$$

$$\mathcal{U} := \left\{ u \in \mathcal{C}^0 : u(y) = \int_{\underline{x}}^{\overline{x}} v(x) r(y;x) dx \; v \in \mathcal{V} \right\}; \tag{3}$$

$$\mathcal{V} := \left\{ v : \mathcal{X} \to \mathbb{R}_+ : \int_{\mathcal{X}} v\left(x\right) = 1 \right\} \tag{4}$$

$$r(y;x) := (y-x)1(y \le x), \; (x,y) \in \mathcal{X}^2. \tag{5}$$

$\mathcal{U}$ is comprised of normalized, increasing, and concave utility functions that are constructed as convex mixtures of elementary Russell and Seo (1989) ramp functions $r(y;x)$, $x \in \mathcal{X}$. This implies that K spans Λ, iff for any $\boldsymbol{\lambda} \in \Lambda$ there exists some $\boldsymbol{\kappa} \in \mathrm{K}$, weakly preferred to the former, by every utility in $\mathcal{U}$. Equivalently, spanning occurs iff no risk averter in $\mathcal{U}$ loses expected utility from the excision Λ-K. This representation can be used for the numerical implementation of the associated testing procedure.

The test statistic can be expressed as:

$$\eta_T := \sqrt{T} \sup_{u\in\mathcal{U}} \left(\sup_{\boldsymbol{\lambda}\in\Lambda} \mathbb{E}_{F_T}\left[u\left(X^{\mathrm{T}}\boldsymbol{\lambda}\right)\right] - \sup_{\boldsymbol{\kappa}\in\mathrm{K}} \mathbb{E}_{F_T}\left[u\left(X^{\mathrm{T}}\boldsymbol{\kappa}\right)\right] \right). \tag{6}$$

The computational complexity of evaluating η_T stems from the functional complexity of the set $\mathcal{U}$. Following Arvanitis et al. (2019) we approximate every element of $\mathcal{U}$ with arbitrary prescribed accuracy using a finite set of increasing and concave piecewise linear functions, as we explain next.

Let N_1, N_2 denote integers greater than or equal to 2. First $\mathcal{X}$ is partitioned into N_1 equally spaced values as $\underline{x} = z_1 < \cdots < z_{N_1} = \overline{x}$, where $z_n := \underline{x} + \frac{n-1}{N_1-1}(\overline{x} - \underline{x})$, $n = 1, \cdots, N_1$. Second, $[0,1]$ is partitioned as $0 < \frac{1}{N_2-1} < \cdots < \frac{N_2-2}{N_2-1} < 1$. Using those partitions, consider:

$$\underline{\eta_T} := \sqrt{T} \sup_{u \in \underline{\mathcal{U}}} \left(\sup_{\boldsymbol{\lambda} \in \Lambda} \mathbb{E}_{F_T} \left[u \left(X^{\mathrm{T}} \boldsymbol{\lambda} \right) \right] - \sup_{\boldsymbol{\kappa} \in \mathrm{K}} \mathbb{E}_{F_T} \left[u \left(X^{\mathrm{T}} \boldsymbol{\kappa} \right) \right] \right); \tag{7}$$

$$\underline{\mathcal{U}} := \left\{ u \in \mathcal{C}^0 : u(y) = \sum_{n=1}^{N_1} v_n r(y; z_n) \; v \in V \right\}; \tag{8}$$

$$V := \left\{ v \in \left\{ 0, \frac{1}{N_2-1}, \cdots, \frac{N_2-2}{N_2-1}, 1 \right\}^{N_1} : \sum_{n=1}^{N_1} v_n = 1 \right\}. \tag{9}$$

Every $u \in \underline{\mathcal{U}}$ consists of at most N_2 linear line segments with endpoints at N_1 possible outcome levels. Furthermore $\underline{\mathcal{U}} \subset \mathcal{U}$, it is finite as it has $N_3 := \frac{1}{(N_1-1)!} \prod_{i=1}^{N_1-1} (N_2+i-1)$ elements and $\underline{\eta_T}$ approximates η_T from below as the partitioning scheme is refined $(N_1, N_2 \to \infty)$. Then for every $u \in \underline{\mathcal{U}}$, the two embedded maximization problems in (7) can be solved using linear programming. Consider

$$c_{0,n} := \sum_{m=n}^{N_1} \left(c_{1,m+1} - c_{1,m} \right) z_m; \tag{10}$$

$$c_{1,n} := \sum_{m=n}^{N_1} w_m; \tag{11}$$

$$\mathcal{N} := \{ n = 1, \cdots, N_1 : v_n > 0 \} \bigcup \{ N_1 \}. \tag{12}$$

For any $u \in \underline{\mathcal{U}}$, $\sup_{\boldsymbol{\lambda} \in \Lambda} \mathbb{E}_{F_T} \left[u \left(X^{\mathrm{T}} \boldsymbol{\lambda} \right) \right]$ is the optimal objective function value of the linear program:

$$\max T^{-1} \sum_{t=1}^{T} y_t \tag{13}$$

$$\text{s.t. } y_t - c_{1,n} X_t^{\mathrm{T}} \boldsymbol{\lambda} \leq c_{0,n}, \; t = 1, \cdots, T; n \in \mathcal{N};$$

$$\sum_{i=1}^{M} \lambda_i = 1;$$

$$\lambda_i \geq 0, \; i = 1, \cdots, M;$$

$$y_t \text{ free}, \; t = 1, \cdots, T.$$

The linear programming problem always has a feasible solution and has $\mathcal{O}(T + M)$ variables and constraints and is tractable for typical data dimensions.

Our empirical tests are based on the entire available history of quarterly investment returns to a standard set of benchmark assets discussed in the date section below, with $N = 4$, $T = 112$, $N_1 = 10$ and $N_2 = 5$. This gives $N_3 = \frac{1}{9!} \prod_{i=1}^{9} (4 + i) = 715$ distinct utility functions and $2N_3 = 1,430$ small linear programming problems. The tests are computationally demanding, with each test requiring a few working days of computational time on a desktop PC with a 2.93 GHz quad-core Intel i7 processor and 16GB of RAM, using MATLAB and GAMS with the Gurobi solver.

Computational Strategy for Bounding

The bounding methodology considers all portfolios in Λ, rather than a single portfolio and the joint empirical support generally consists of infinitely many points, introducing the need for discretization.

Let $\hat{\mathcal{X}}_\Lambda := \left[\hat{A}_\Lambda, \hat{B}_\Lambda\right]$, $\hat{A}_\Lambda := \min_{\Lambda,\lambda} \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda}$ and $\hat{B}_\Lambda := \max_{\Lambda,t} \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda}$. The computational strategy partitions $\hat{\mathcal{X}}_\Lambda$ using J equally spaced grid points $\hat{x}_j := \hat{A}_\Lambda + (j-1)\left(\hat{B}_\Lambda - \hat{A}_\Lambda\right)(J-1)^{-1}$, $j = 1, \cdots, J$. For every grid point, let $\hat{L}^*_{\Lambda,j} := \min_\Lambda L(\boldsymbol{\lambda}, \hat{x}_j, \hat{F})$,$j = 1, \cdots, J$.

The approximation $\xi(\Lambda, \Lambda, \sqrt{T}\hat{F}) \approx \sqrt{T}\chi_J(\Lambda, \Lambda, \hat{F})$ is used, where

$$\begin{aligned} \chi_J(\Lambda, \Lambda, \hat{F}) &:= \min_{\Lambda} \max_{\Lambda, j} D(\boldsymbol{\lambda}, \boldsymbol{\lambda}, \hat{x}_j, \hat{F}) \\ &= \min_{\Lambda} \max_{j} \left(L(\boldsymbol{\lambda}, \hat{x}_j, \hat{F}) - \hat{L}^*_{\Lambda,j}\right) \\ &= \min_{\Lambda,\sigma} \left(\sigma : L(\boldsymbol{\lambda}, \hat{x}_j, \hat{F}) - \sigma \leq \hat{L}^*_{\Lambda,j},\ j = 1, \cdots, J\right). \end{aligned} \tag{14}$$

The number of grid points $J = 100$ is chosen to balance accuracy with computer time. It yields very high accuracy for typical applications.

The approximate $\chi_J(\Lambda, \Lambda, \hat{F})$ can be computed by solving a series of linear programs using linear relaxations (Rockafellar et al., 2000). Each value $\hat{L}^*_{\Lambda,j}$, $j = 1, \cdots, J$, can be computed as the optimal value of the objective function of the following program:

$$\begin{aligned} &\min T^{-1} \sum_{t=1}^{T} \eta_{j,t} \\ &-\eta_{j,t} - \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda} \leq -\hat{x}_j,\ t = 1, \cdots, T; \\ &\eta_{j,t} \geq 0,\ t = 1, \cdots, T; \\ &\boldsymbol{\lambda} \in \Lambda. \end{aligned} \tag{15}$$

Given the solutions to the J problems, $\chi_J(\Lambda, \Lambda, \hat{F})$ can be computed by solving the linear program:

$$
\begin{aligned}
&\min \sigma \qquad (16)\\
&T^{-1}\sum_{t=1}^{T}\theta_{j,t} - \sigma \leq \hat{L}^{*}_{\Lambda,j},\ j = 1, \cdots, J;\\
&-\theta_{j,t} - \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\lambda} \leq -\hat{x}_j,\ j = 1, \cdots, J;\ t = 1, \cdots, T;\\
&\theta_{j,t} \geq 0,\ j = 1, \cdots, J;\ t = 1, \cdots, T;\\
&\boldsymbol{\lambda} \in \Lambda;\\
&\sigma \text{ free.}
\end{aligned}
$$

The optimal solution $\lambda \in \Lambda$ identifies the portfolio that stochastically spans but is not spanned by any portfolio in Λ.

3 Are GDP-linked bonds spanned?

We now put stochastic spanning tests to market data, from the perspective of a US investor, to test in-sample whether GDP-linked bonds improve the opportunity set of a broad portfolio of stocks, bonds, and cash. An investment opportunity set that includes GDP-linked bonds provides diversification benefits relative to the set of benchmark assets if adding these bonds to the benchmark leads to a significant leftward shift in the efficient frontier. This is equivalent to testing whether a set of benchmark assets spans the augmented set. If GDP-linked bonds are spanned, including them in the benchmark universe will not increase the portfolio's expected return per unit of risk, and diversification benefits are non-existent.

We test the hypothesis H_0 that the benchmark asset class stochastically spans the augmented with GDP-linked bonds against the alternative H_1 that spanning is rejected. Rejecting spanning suggests that GDP-linked bonds offer diversification benefits.

3.1 Data and GDP-linked bonds

The reference benchmark assets are stocks (S&P 500 Total Return Index), bonds (Barclays US Aggregate Bond Index), and cash (3-Month T-Bill). To ensure the robustness of the results to the choice of the benchmark assets, we additionally include the MSCI World Index, the Fama and French Market Index, the dynamic trading strategies SMB and HML, the Barclays US Corporate bond index, and the US government 5-year, 10-year, and 30-year Benchmark Bond Indices. We use monthly closing prices from Datastream and the Kenneth French Data Library, spanning the 40 years from January

1980 to December 2019, for 480 observations.

We first consider *floaters*, with coupons given as a function of a country's GDP

$$C_t = max[C_{0t} + (g_t - \bar{g}), 0]. \quad (17)$$

C_{0t} is the base coupon which is adjusted by the deviations of the real growth rate g_t from its target $\bar{g}$. We get GDP growth rates from Datastream. The annual real growth rate of GDP, g_t, is calculated in the corresponding quarters, year-on-year, and then is converted to quarterly growth rates. The coupon C_t is linked to GDP growth, increasing from the baseline if the growth exceeds the target. Otherwise, it decreases with a floor at zero.

We consider par bonds with the coupon payments C_t equal to the return of these bonds, i.e., $C_t = R_t$, with a floating base coupon rate given by either the return of US Benchmark 10yr Government Bond Index (labeled, Floaters-10yr) or the return of the Barclays Bond Index (labeled, Floaters-Barclays). We use as a benchmark target GDP growth rate, $\bar{g}$, the average over our test period, which is estimated to be 0.66% for the US using the AMECO database. We also test for values of $\bar{g}$ around this benchmark (namely, 0%, 0.66%, 1%, 1.5%, 2%, 2.5%, 3%). The higher the target, the less likely it is for the increase in coupon to kick in, and the GDP-linked bond becomes a regular bond. Data for calculating the coupon-indexed GDP-linked bonds are seasonally adjusted at constant prices.[8]

We also consider *linkers*, i.e., zero coupon bonds with outstanding principal amount

$$B_t = B_0 \frac{Y_t}{Y_0}, \quad (18)$$

paid at maturity. B_0 is the original amount issued (typically 100), and Y_0 and Y_t are the nominal GDP values at the issuing date and at time t, respectively. We obtain monthly returns of our GDP-linked bonds using linear interpolation of the quarterly data obtained from equations (17)–(18).

Table 1 reports summary statistics for the performance of the benchmark assets and the GDP-linked bonds over the sample period. A Jarque-Berra test rejects the normality of asset return distributions, ruling out the use of MV spanning tests. Given the high positive skewness of GDP-linked bonds, our focus on stochastic dominance instead of parametric methods is well justified. One important factor affecting the yield of floaters is the choice of the target growth level. The Sharpe ratio is considerably greater for floaters with a small target growth rate $\bar{g}$; as the level of $\bar{g}$ increases, the average return decreases along with the volatility, and the Sharpe ratios drop. This observation is in line with Consiglio and Zenios (2018).

[8] We use monthly instead of quarterly data to obtain the long time series required for the asymptotic statistical inference required by our non-parametric tests.

[Insert Table 1 About Here]

3.2 Spanning tests

We first test the null H_0 of stochastic spanning. The uncertainty surrounding the returns of GDP-linked bonds could lead potential investors to demand a risk premium relative to conventional government bonds to compensate for growth risk. Since we do not have a firm empirical estimate of the size of the risk premium,[9] we ask the closely related question whether there exists some risk premium levels at which the benchmark portfolio set does not span the augmented portfolio set. If this were the case for reasonable premia, our tests of the null would be valid. We, therefore, conduct spanning tests using, unfavorably for our test, low but constant premia of zero or 50bp. We also consider several alternative settings for the risk premia. Specifically, we perform tests with growth-dependent and premia that may become negative, as well as a randomized test. All results consistently point in the direction of rejecting the null.

To increase the power and efficiency of our test, we use a bias correction procedure for the quantile estimates $q_{T,b_T}(1-\alpha)$ to mitigate sensitivity on the choice of b_T in finite samples. Following Arvanitis et al. (2019), we choose sample sizes $b_T = \lfloor T^c \rfloor$, with c ranging from 0.6 to 0.9. Using OLS regression on the empirical quantiles $q_{T,b_T}(1-\alpha)$ for significance level $\alpha = 0.05$, we get the estimate q_T^{BC} for the critical value. We reject spanning if the test statistic η_T is higher than q_T^{BC}. This procedure results in an asymptotically exact and consistent test as long as α is appropriately chosen and the subsampling rate b_T diverges to infinity at a slower rate than T.

Recalling the various estimates from the literature of premia in the range of 50bp to 150bp, we test the null for zero risk premium or a fixed premium of 50bp. These are unfavorably low values for our test. We report in Table 2 the test statistics η_T and the regression estimates q_T^{BC}. Panel A is for floaters with C_{0t} set equal to the returns of the 10yr US Benchmark Government Bond Index (Floaters-10yr), Panel B is for the case where C_{0t} is the returns of the Barclays Bond Index (Floaters-Barclays), and Panel C is for linkers. In each panel, we report results with a premium of 0bp and 50bp, and for different levels of $\bar{g}$.

[Insert Table 2 About Here]

In panels A and B, we observe that for a wide range of $\bar{g}$, the test rejects the spanning hypothesis H_0 in favor of the alternative H_1. The Floaters-10yr designed with $\bar{g}$ 3.5% or higher could be spanned. However, adding a 50bp risk premium to these floaters, we see that they could not be spanned for conventional levels of target GDP growth rate. For the Floaters-Barclays, the range of $\bar{g}$ for which

[9] Recall that premia can currently only be inferred from models, such as those of Bowman et al. (2016); Consiglio and Zenios (2018); Eguren-Martin et al. (2020).

the test rejects the spanning hypothesis is limited to up to 2%, whereas when we add a risk premium of 50bp again, we reject the spanning hypothesis for all targets of GDP growth rate.

The characteristics of the bond design are critical in making such bonds attractive to investors. This point, made in Consiglio and Zenios (2018), is usually overlooked in the GDP-linked bonds literature, where one bond design is typically considered. Our results show that these bonds are not spanned for reasonable values of the key parameter $\bar{g}$.

In panel C, we also observe that the null hypothesis is rejected for linkers, with or without a risk premium. Hence, both linkers and floaters constitute a new asset class for investors that can improve their opportunity set. This is a new result in the literature.

3.3 Robustness to changing risk premia

For very large positive premia in our tests above, GDP-linked bonds will be attractive for investors, and the null would be easy to reject. We have shown that for unfavorably low premia, which are reasonable given estimates from the literature, the null is still rejected. We take a step further and consider growth-dependent premia, in which case the results of the hypothesis testing are not obvious. We also consider negative premia with the test stacked in favor of the null (for very large negative premia, the new instruments will trivially not be attractive). Finally, we successfully perform a randomized test.

3.3.1 Growth dependent risk premia

The return of GDP-linked bonds depends on the GDP growth rate. When growth is expected to be high, investors may demand a lower risk premium to add these instruments to their portfolios. Still, the premium can increase when the GDP growth is expected to be relatively low. We use estimates of a time series of premia from market data to add a different risk premium every month, depending on the market conditions. Specifically, we use the risk premia for 2yr GDP-linkers and 7yr GDP-linkers from (Eguren-Martin et al., 2020, Figures 7-8), for the period from January 2010 to June 2017, for a total of 90 monthly observations.[10] For both 2-year and 7-year linkers, the estimated premia are positive, as expected, ranging between 50 and 520bp. The average risk premium over the testing period is 225bp for the 2-year bonds and slightly higher at 270bp for the 7-year instrument.

Given the large average values of the estimated premia, we expect the results to be consistent with those with low constant premia from the previous section. However, the changes in premia add further

[10] The data were provided by the authors of the Bank of England study. We note that the paper estimates premia from the issuers' perspective that has to pay a premium to issue these bonds, but we consider the buyer's perspective of who receives the premium to invest.

volatility to GDP-linked bond returns, which may result in these instruments being dominated by the benchmarks. We carry out the spanning test, and the null is rejected for both maturities. The test statistic for 2yr and 7yr, respectively, is 0.1526 and 0.1930, with the regression estimates at 0.1514 and 0.1892.

Our main finding is that GDP-linked bonds are not spanned holds under growth-dependent risk premia from the literature. Investors can still benefit from including these bonds in their portfolios.

3.3.2 Negative risk premia

We have tested the spanning hypothesis under the reasonable assumption that investors will receive a risk premium to add GDP-linked bonds to their portfolios. An interesting question, though, is what happens if these instruments become attractive (Blanchard et al., 2016; Demertzis and Zenios, 2019; Kamstra and Shiller, 2009) so that investors would pay a premium to acquire them. We test for this unlikely but plausible eventuality by repeating the analysis with negative premia and report the results in Table 3. We consider negative risk premia of -25bps or -50bps for the floaters and -5bps and -10bps for the linkers. For more significant negative premia, all GDP-linked bonds are trivially spanned.

[Insert Table 3 About Here]

In Panel A (Floaters-10yr) and B (Floaters-Barclays), we observe that if investors have to pay a premium, GDP-linked bonds are no longer attractive except for very low $\bar{g}$. The Floaters-10yr designed with values of $\bar{g}$ up to the average of 0.66% are not spanned for a risk premium of up to -25bps. With a -50bps premium, spanning is rejected only for bonds designed with zero growth target. The floaters-Barclays are attractive only for zero growth rate.

Likewise, in Panel C (linkers), we reject the null hypothesis of spanning for linkers only for a minimal negative risk premium of -5bps.

In conclusion, for small negative risk premia, these bonds are not spanned for low values of the key parameter $\bar{g}$, so investors may even pay a premium to add them to their investment opportunity set.

3.3.3 Randomized risk premia

To further test the robustness of our main finding, we take another step with the (unfavorable) negative risk premia and add noise. This randomized test is also stacked in favor of the null.

We start with a risk premium of -25bps and add noise with a mean of zero and a standard deviation equal to the standard deviation of GDP growth over the testing period. We generate 480 random numbers, one for each month of the testing period, and repeat the spanning tests. For Floaters-10yr, we use a $\bar{g} = 0.66\%$, the maximum $\bar{g}$ for which we reject spanning with a fixed negative premium of 25bps, and, likewise, for Floaters-Barclays, we use a $\bar{g} = 0\%$.

We create 50 samples of random risk premia and test for spanning each sample. We reject the spanning hypothesis for both Floaters-10yr and Floaters-Barclays in all cases. We repeat the analysis with a more volatile distribution of risk premia, with noise equal to twice the historical standard deviation of GDP growth. We generate another 50 independent samples of random risk premiums, run the spanning test, and reject the null in 90% of the cases for both Floaters-10yr and Floaters-Barclays.

3.3.4 A Garch(1,1) process for risk premia

Finally, we test the robustness of our main finding using a framework of conditional heteroskedasticity for the risk premia. The $(X_t)_{t\in\mathbb{Z}}$ process of the premia is constructed as a vector GARCH(1,1) process that also contains an appropriately transformed element. This allows for both temporal and cross-sectional dependence between the random premia that constitute the vector process. Below, we describe the process. We generate processes for three risk premia with different Garch parameters.

Suppose that

$$z_t \overset{\text{iid}}{\sim} N(0,1), \; t \in \mathbb{Z}.$$

Furthermore, for all $t \in \mathbb{Z}$, for $i = 1, 2, 3$, $\omega_i, \alpha_i, \beta_i \in \mathbb{R}_{++}$, $\mu_i \in \mathbb{R}_+$ define

$$\begin{aligned} x_{i_t} &= \mu_i + z_t h_{i_t}^{1/2}, \\ h_{i_t} &= \omega_i + \left(\alpha_i z_{t-1}^2 + \beta_i\right) h_{i_{t-1}}, \; \mathbb{E}\left(\alpha_i z_0^2 + \beta_i\right)^{1+\epsilon} < 1, \end{aligned}$$

for some $\epsilon > 0$.

Suppose that $X_t = (x_{1_t}, x_{2_t}, x_{3_t})'$. We set $\mu_i = 0$ for $i = 1, 2, 3$, $\omega_1 = 0.5$, $\omega_2 = 0.5$, and $\omega_3 = 0.5$, $a_1 = 0.4$, $a_2 = 0.45$, and $a_3 = 0.5$ and $\beta_1 = 0.5$, $\beta_2 = 0.45$, and $\beta_3 = 0.4$. For Floaters-10yr we use a $\bar{g} = 0.66\%$. The test statistics are 0.1835, 0.2109 and 0,1965 with the regression estimates at 0.1746, 0.2075 and 0.1892. For Floaters-Barclays, we use a $\bar{g} = 0\%$. The test statistics are 0.1745, 0.1985 and 0,1834 with the regression estimates at 0.1701, 0.1894 and 0.1796. Thus, in all cases, we reject the spanning hypothesis for both Floaters-10yr and Floaters-Barclays.

Overall, our results show that GDP-linked bonds can be attractive to investors for a wide range of target growth parameters when even a small risk premium is added to the return of these bonds to compensate them for the risk of low GDP growth. The premium for which spanning is rejected (50bp or less) is at the low range of current estimates from the literature on potential premia for these instruments, lending confidence that our findings will likely be true when such bonds are issued. Investors may even be willing to pay a premium to include these bonds in their investment opportunity set.[11]

4 Diversification benefits of GDP-linked bonds

We further validate our stochastic spanning results with out-of-sample testing. Although the in-sample tests have provided strong evidence to reject the null, the out-of-sample tests mimics the investor behavior and further documents the diversification benefits from GDP-linked bonds.

We construct optimal portfolios from the set of benchmark assets and the set augmented with GDP-linked bonds, and conduct backtesting experiments on a rolling horizon basis. The rolling horizon covers 480 months from January 1980 to December 2019, with the first 300 months used to start the calibration and the remaining 180 from January 2005 to carry out the backtesting. At month t, we use the previous 300 observations to solve the stochastic spanning model and get the optimal portfolio weights. We use these weights to compute the out-of-sample realized return of the portfolio over the period $[t, t+1]$. We repeat this process by advancing t by one month and dropping the first observation from the previous window until we reach the end of the sample. With this approach, we obtain two series of monthly ex-post optimal portfolio returns and use these time series to evaluate the out-of-sample performance of the two derived optimal portfolios. First, we illustrate the results and compare the portfolio cumulative returns for different designs of the GDP-linked bonds. Then, we compare the portfolio performance using, again, nonparametric stochastic dominance tests.

In both cases, we assume zero risk premia. We are not interested in coming up with the best estimate of returns in each portfolio. Rather, our goal is to maintain a uniform approach consistent across asset classes and minimizes the pernicious effects of data snooping. As such, if data snooping can be avoided, our results may understate the true performance gains for the GDP-linked portfolios that can be achieved with more thoughtful choices of risk premia.[12]

[11]We repeat the analysis when short selling of the benchmark assets is allowed, as well as for different sub-periods, and find consistent results. These tests are available from the authors.

[12]Some of the tests of the next section were also carried out with positive risk premia, up to 50bp, confirming empirically the expectations for even better performance.

4.1 Out-of-sample cumulative returns

Figure 1 illustrates the out-of-sample cumulative returns of the benchmark and the augmented optimal portfolios during the backtesting period (January 2005 - December 2019) with either Floaters-10yr (Panel A) or Floaters-Barclays (Panel B) for different levels of $\bar{g}$. Panel C illustrates the out-of-sample cumulative returns of the benchmark portfolio and the portfolio augmented with linkers, priced at par.

[Insert Figure 1 About Here]

From Panel A, we observe that for Floaters-10yr, the augmented portfolios dominate the performance of the benchmark portfolios for even high levels of the target of GDP growth rate, with $\bar{g}$ ranging from 0% to 3%. The augmented portfolio with Floaters-Barclays (Panel B) outperforms the benchmark but for a lower range of $\bar{g}$ up to 1%. As $\bar{g}$ increases, the cumulative return of the augmented portfolio converges to the benchmark. This is expected since, for high levels of target GDP, the investor is unlikely to receive any upside potential; the returns of the GDP-linked bonds are low, with the optimal portfolios consisting primarily of benchmark assets.

For linkers (Panel C), the difference between the performance of the two portfolios is marginal.

In conclusion, the out-of-sample cumulative returns of portfolios that include GDP-linked bonds can be higher than the cumulative returns of the benchmarks. However, this analysis does not account for the volatility (risk) of the returns. We compare the benchmark and augmented portfolios next. The two optimal portfolios formed by the respective two asset universes, using a non-parametric stochastic (non-)dominance test as well as some well-known parametric performance measures.

4.2 Non-parametric stochastic dominance performance test

We use a pairwise (non-)dominance test for a risk-adjusted comparison of the out-of-sample performance of the benchmark and augmented portfolios. The most common approach to test for stochastic dominance (SD) is to posit the null hypothesis of dominance. But If we run a test to reject the null of SD by one distribution, this would not imply SD by the other distribution since it can also happen that the test fails to rank the distributions. This suggests that it is more desirable to posit instead the null of non-dominance.

The definition of second order stochastic non-dominance is the following:

Definition 2. (Stochastic non-dominance): *The augmented portfolio λ does not strictly second order stochastically dominate the benchmark portfolio κ*, say $\lambda \not\succ_F \kappa$, iff

$$\exists x \in \mathcal{X}^{\mathbb{N}} : \, D(x, \lambda, \kappa, F) > 0, \text{ or } \forall x \in \mathcal{X}^{\mathbb{N}} : \, D(z, \lambda, \kappa, F) = 0.$$

The convexity assumption of Λ allows for an equivalent formulation in terms of expected utility. Strict second-order stochastic non-dominance holds iff κ achieves a higher expected utility for some non-decreasing and concave utility function or achieves the same expected utility for every non-decreasing and concave utility function. Equivalently, strict stochastic non-dominance holds iff κ is strictly preferred to λ by some risk averter, or every risk averter is indifferent between them.

We test the null hypothesis H_0' vis-á-vis the alternative (Anyfantaki et al., 2021):

H_0': Portfolio λ with GDP-bonds does not strictly second-order stochastically dominate the benchmark portfolio κ,

H_1': Portfolio λ with GDP-bonds stochastically dominates the benchmark portfolio κ,

To calculate p-value, we use block-boostrapping. The p-value is approximated by $\tilde{p}_j = \frac{1}{R}\sum_{r=1}^{R}\{\xi^{\star}_{T,r} > \xi_T\}$, where ξ_T is the test statistic, $\xi^{\star}_{T,r}$ is the bootstrap test statistic, averaging over R replications.

Table 4 reports the p-values for the null from the distribution of monthly portfolio returns over the backtesting period. We observe that for both Floaters-10yr and Floaters-Barclays, we reject the null at conventional levels for a reasonable range of target GDP growth rates, including the historical average ($\bar{g} = 0.66 - 3$ for Floaters-10yr, and $\bar{g} = 0.66 - 1$ for Floaters-Barclays). For linkers, we reject the null at the 0.10 level.

In conclusion, the out-of-sample performance of portfolios that include GDP-linked bonds can be better than the performance of the benchmarks.

[Insert Table 4 About Here]

4.3 Parametric performance tests

Finally, we run a battery of tests using commonly accepted parametric performance measures to strengthen our results further and corroborate the evidence suggested in Figure 1. We compare the performance of the portfolios for different levels of $\bar{g}$, using the Sharpe ratio (Sharpe, 1994), the downside Sharpe ratio (Ziemba, 2005), and the upside potential (UP) ratio (Sortino and Van Der Meer, 1991). Given the asymmetric return distribution of GDP-linked bonds, the downside Sharpe and UP ratios are more appropriate for our tests than the Sharpe ratio. Failing these tests would not necessarily refute our main findings of GDP-linked bonds as a new asset class. However, passing these tests corroborates our findings even when we make some distributional assumptions on asset returns.

Let R_p be the portfolio return, with realised monthly returns R_t at the tth rolling window, $t = 1, ..., T$, and let R_{ft} be the return of the risk-free rate for the same period. The Sharpe ratio is given by $SR_p = \frac{E[R_p - R_f]}{(Var[R_p - R_f])^{1/2}}$. The downside Sharpe ratio is given by $S_{p_-} = \frac{\bar{R}_p - R_f}{\sqrt{2}\sigma_{P_-}}$, where the downside risk

σ_{p-} is calculated by $\sigma_{p_-}^2 = \frac{\sum_{t=1}^{T}(R_{p,t}-\bar{x})_-^2}{T-1}$, where the benchmark $\bar{x}$ is set to zero, and the numerator is positive for the months t for which the portfolio p has losses ($R_{p,t}$ is negative). The UP ratio is given by $\text{UP} = \frac{\frac{1}{K}\sum_{t=1}^{K}\max[0,R_{p,t}-R_{ft}]}{\sqrt{\frac{1}{K}\sum_{t=1}^{K}(\max[0,R_{ft}-R_{p,t}])^2}}$.

To assess the economic significance of any performance difference between the two optimal portfolios, we also report the opportunity cost (Simaan, 1993). Let R_A and R_B be the realized returns of the augmented and the benchmark portfolios, respectively. The opportunity cost θ is defined as the return that needs to be added to (or subtracted from) the benchmark portfolio return so that the investor is indifferent (in utility terms) between the two optimal portfolios, i.e., $E[U(1+R_B+\theta)] = E[U(1+R_A)]$. A positive (negative) θ implies that the investor is better (worse) off if the investment opportunity set includes GDP-linked bonds. The opportunity cost takes into account the entire probability density function of asset returns without assuming normally distributed returns. To calculate the opportunity cost, we use exponential and power utility functions, consistent with second-degree stochastic dominance, for different levels of risk aversion.

We also compute the portfolio turnover (PT), defined as the average of the absolute change of weights over the T rebalancing points for the N assets, $PT = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}(|w_{i,t+1} - w_{i,t}|)$, where $w_{i,t+1}$ and $w_{i,t}$ are the optimal weights of asset i at time t and $t+1$, respectively. PT indicates the amount of rebalancing required to implement the strategy.

Finally, we evaluate the performance of the two portfolios under the risk-adjusted return measure net of transaction costs (DeMiguel et al., 2009), to account for proportional transaction costs due to portfolio turnover. Let tc_i be the proportional transaction cost for asset i, and R_{t+1} the realized return of the portfolio at time $t+1$. The portfolio wealth net of transaction costs, NW, is given by:

$$NW_{t+1} = NW_t(1 + R_{t+1})[1 - \sum_{i=1}^{N}(tc_i \times |w_{i,t+1} - w_{i,t}|)], \tag{19}$$

and the return, net of transaction costs, is defined as

$$RTC_{t+1} = \frac{NW_{t+1}}{NW_t} - 1. \tag{20}$$

Let μ_B and μ_A be the out-of-sample mean of monthly RTC with the benchmark and the augmented opportunity set, respectively, and σ_B and σ_A be the corresponding standard deviations. Then, the return-loss measure is

$$RLM = \frac{\mu_A}{\sigma_A} \times \sigma_B - \mu_B. \tag{21}$$

This is the additional return required on the benchmark portfolio so that it performs as well as the augmented portfolio. Following the literature, we use 50bp for the transaction cost of stocks, 35bp for bonds, including GDP-linked bonds, and zero for the risk-free asset.

Table 5, Panels A-C, report the parametric performance measures for the benchmark and floaters portfolios for different levels of $\bar{g}$. These performance measures supplement the evidence from the non-parametric stochastic (non-)dominance tests above. The higher the value of each one of these measures, the greater the investment opportunities for GDP-linked bonds.

Comparing the benchmark portfolio (Panel A) with the Floaters-10yr (Panel B) and Floaters-Barclays (Panel C), we observe that including coupon-indexed GDP-linked bonds into the investment set increases both the Sharpe and the downside Sharpe ratios, for a large range of $\bar{g}$, up to 3%. This reflects an increase in the augmented portfolios' risk-adjusted performance; hence, floaters expand the investment opportunity set of risk averse investors. The equality test of Sharpe ratios (Leung and Wong, 2006) rejects the null hypothesis of equality of the Sharpe ratios of the benchmark with any augmented portfolio. The same results hold when comparing UP ratios. Moreover, we can see that the return-loss measure (RLM) accounting for transaction costs is positive for all cases, again confirming the out-of-sample superiority of portfolios that include floaters. The opportunity cost θ is consistently positive for all levels of risk aversion, providing further evidence in favor of floaters. This result indicates that a risk-averse investor needs an additional return equal to θ to be indifferent between investing in the benchmark and the augmented portfolio. The opportunity cost definition relies on the computation of the expected utility or, equivalently on the probability density function of portfolio returns. Thus, the opportunity cost takes into account higher-order moments in contrast to the Sharpe ratios.

Finally, we observe that including linkers into the investment set (Panel D) also increases the Sharpe ratio, the downside Sharpe ratio, and the UP- ratio. The increase, however, is small, and the return-loss and the opportunity cost are positive but close to zero for all levels of risk aversion. The results indicate that investors are better off if the investment opportunity set includes linkers priced at par. Still, the performance benefits are marginal.[13]

[Insert Table 5 About Here]

In conclusion, all out-of-sample results are consistent with the in-sample spanning tests. The stochastic (non-)dominance test and the parametric performance measures we use indicate that GDP-linked bonds provide new investment opportunities and yield diversification benefits for risk-averse investors.

[13]We also performed the out-of-sample test allowing for short selling of the benchmark assets. Our finding that including GDP-linked bonds improves the opportunity set of risk averse investors holds also under this more demanding test. The results are available from the authors.

To shed further light on using GDP-linked bonds in investment portfolios, we report the average portfolio compositions in Table 6. The optimal benchmark portfolios invest in Barclays US Aggregate index (70.36%), in Fama and French Market portfolio (17.39%) and Barclays US Corporate index (5.36%). The optimal augmented portfolio includes mainly GDP-linked bonds, especially when such bonds are available with low levels of target GDP. When the Floaters-10yr ($\bar{g} = 0.66$) are included in the available assets, the augmented portfolios include 87% the floaters and 13% benchmark assets. The holdings in floaters decrease as the level of target GDP increases, to 27% for $\bar{g} = 3$. The same is true when Floaters-Barclays are included in the optimal portfolios. The optimal weight of Floaters-Barclays ranges from 73.27% for $\bar{g} = 0.66$, to 2.72% for $\bar{g} = 2$. For linkers, the average composition of the augmented portfolio is 88.38% benchmark assets and 11.68% linkers.

[Insert Table 6 About Here]

5 Factors of return

Having established that GDP-linked bonds would provide diversification benefits to investors, we investigate the determinants of returns. We expect factors that drive GDP growth and interest rates to determine the returns of GDP-linked bonds. However, since the same factor may drive GDP growth and interest rates, but with different signs, it is not apparent a priori how the factors will drive GDP-linked bond returns. An investigation of the net effects deserves attention, and we provide a first estimation in this direction.

We use two sets of factors to identify the determinants of return of floaters. The first refers to financial factors documented to explain and predict bond market returns, thus potentially explaining the base coupon C_{0t}. We consider the default spread (Bessembinder and Chan, 1992) defined as the excess of the yield on long-term corporate BAA-rated (Moody's) over the yield on AAA-rated bonds, and the term spread (Fama and French, 1989) defined as the difference between the yield on AAA-rated bonds and the one-month T-Bill rate.

The second set refers to macro factors that could explain the GDP growth rate g_t. We do not aim at an exhaustive test of the vast list of factors that can explain GDP-linked bond returns through the GDP channel but select the most relevant ones from the literature. The output growth rate is determined by the growth rate of production inputs (physical capital and labor) and a component that captures the productivity growth (Mankiw et al., 1992; Solow, 1956, 1957). We employ a proxy of capital stock, the capacity (or economic) utilization (or operation) rate (Shaikh and Moudud, 2004), which measures the proportion of potential economic output that is realized given existing capital equipment. As a proxy for labor, we use the unemployment rate (Pissarides, 2000; Rodríguez-Pose and Di Cataldo, 2014). Additional factors from the literature are used as controls, relating to fiscal (public debt) and monetary (interest rates, inflation) factors (Banerjee and Marcellino, 2006; Barro, 2003b; Checherita-Westphal and Rother, 2012; Saint-Paul, 1992).[14]

The financial and macroeconomic factors considered in our analysis and the data sources are presented in Appendix Table A1. All the macro factors are in real terms (constant prices) and seasonally adjusted, with the factor correlations reported in Appendix Table A2.

For floaters, the variable underestimation is from eqn. (17), which we rewrite as

$$C_t - C_{0t} = max[(g_t - \bar{g}), -C_{0t}]. \tag{22}$$

We use this equation to calculate the excess floater return R_t, and run a linear regression of the excess return for both Floaters-10yr and Floaters-Barclays on the factors

$$R_t = \alpha + \beta_i f_{it-1} + \epsilon_{t-1}. \tag{23}$$

The f_{it} is the level of factor i at time t, and ϵ_t is the error term. All regressors are lagged to alleviate potential endogeneity issues. From (18), we also calculate the linker excess returns and run the same linear regression.

A potential concern is that there may be unobserved characteristics of the economy that correlate with our GDP-linked bonds and interest rates but are not explicitly taken into account and are part of the error term. We apply a two-step efficient generalized method of moments (GMM) estimator to address potential omitted bias and endogeneity issues. We use exogenous variables (instruments) that are strongly correlated with the potentially endogenous regressors and check that the instruments only influence the dependent variable through the presumed endogenous independent variables.

Appendix Tables A3 and A4 present the GMM estimates of eqn. (23) for Floaters-10yr (Panel A) and Floaters-Barclays (Panel B), for different target GDP growth $\bar{g} = 0.66\%, 2\%, 3\%$, and $\bar{g} = 0.66\%, 1\%, 2\%$, respectively, as well as for Linkers (Panel C), for quarterly and monthly data.[15]

Overall, our findings support that when using monthly data, the financial risk measures (term and default spread, FFR) are highly statistically significant and with correct signs, whereas using quarterly data, we obtain statistically significant macro factors. This is reasonable because the macro factors react better in the long term than the financials.

[14]See Checherita-Westphal and Rother (2012); Saint-Paul (1992) for the relationship between debt and economic growth, Barro (2003b) for a discussion between inflation and growth, and Banerjee and Marcellino (2006) for the relationship between economic growth and various structures of interest rates.

[15]The dataset in this section starts from 1986 due to the BAA corporate Moody's bond used to calculate the default spread.

6 Conclusions

GDP-linked bonds are attracting increasing attention mostly for their potential countercyclical benefits to the issuing sovereigns, but it has also been argued that they allow individuals to invest in the future prosperity of their country. We have asked if these innovative instruments are a distinct asset class with diversification benefits for investors. We employ a stochastic spanning methodology to answer this research question both in and out of sample. We construct and compare optimal portfolios derived from an asset universe that includes a benchmark set of equities, bonds, and cash assets, and one that is augmented with floaters or linkers GDP-linked bonds.

We find that both floaters and linkers are not spanned by a broad set of benchmark assets, thus constituting a new asset class. This finding holds for a reasonable range of bond design parameters and risk premia. This aspect of our contribution is important since GDP-linked bonds are not currently traded, and their design and pricing have to be inferred from models in the literature. In this sense, it is important to use a non-parametric approach so that our findings do not rely on any assumptions about the return distributions of these instruments. Out-of-sample testing confirms significant performance gains from including GDP-linked bonds in a broad-based benchmark set of assets.

We go further to identify the finance and macro determinants of GDP-linked bond returns. Given their nature as bond instruments linked to a country's GDP growth, we find that finance (term and default spread, FFR) and macro variables (public debt, inflation, capacity utilization rate) matter.

Overall, our results document significant advantages for investors in GDP-linked bonds, complementing the extensive literature on the benefits for issuing sovereigns. This contribution fills a gap in the literature by considering the demand side.

One area for future work would be to explore further the determinants of GDP-linked bond returns, given the very large list of variables that explain GDP growth. Such work is complicated by the need to distinguish the factors explaining potential growth from cyclical factors linked to fiscal policy and the spillover effects of the volume of GDP-linked bonds issued on growth in addition to the effects through the default rate. Also, we face the challenge that several macro variables may only be available on a yearly basis. Another direction worthy of investigation would be to obtain estimates of the liquidity and innovation premia that may prevail at the early stages of launching such instruments. Equipped with such estimates, one can easily repeat the analysis of this paper with premia that encompass the risk of GDP volatility, liquidity, and innovation.

Disclosures and declarations

Stavros Zenios received funding for this work from the Cyprus Academy of Sciences, Letters, and Arts.

The authors have no competing interests to declare relevant to this article's content.

References

Anyfantaki, S., E. Maasoumi, J. Ren, and N. Topaloglou (2021): "Evidence of Uniform Inefficiency in Market Portfolios Based on Dominance Tests," *Journal of Business & Economic Statistics*, 0, 1–13.

Arvanitis, S., M. Hallam, T. Post, and N. Topaloglou (2019): "Stochastic spanning," *Journal of Business & Economic Statistics*, 37, 573–585.

Arvanitis, S., T. Post, and N. Topaloglou (2021): "Stochastic bounds for reference sets in portfolio analysis," *forthcoming in Management Science*, available on https://doi.org/10.1287/mnsc.2020.3838.

Banerjee, A. and M. Marcellino (2006): "Are there any reliable leading indicators for US inflation and GDP growth?" *International Journal of Forecasting*, 22, 137–151.

Barr, D., O. Bush, and A. Pienkowski (2014): "GDP-linked bonds and sovereign default," Working Paper 484, Bank of England.

Barro, R. (2003a): "Optimal Management of Indexed and Nominal Debt," *Annals of Economics and Finance*, 4, 1–15.

Barro, R. J. (2003b): "Determinants of economic growth in a panel of countries," *Annals of Economics and Finance*, 4, 231–274.

Benford, J., T. Best, and M. Joy (2016): "Sovereign GDP-linked bonds," Financial Stability Paper 39, Bank of England, London, UK.

Benford, J., J. D. Ostry, and R. Shiller (2018): *Sovereign GDP-Linked Bonds: Rationale and Design*, CEPR Press.

Bessembinder, H. and K. Chan (1992): "Time-varying risk premia and forecastable returns in futures markets," *Journal of Financial Economics*, 32, 169–193.

Blanchard, O., P. Mauro, and J. Acalin (2016): "The Case for Growth-Indexed Bonds in Advanced Economies Today," Policy Brief PB16–2, Peterson Institute for International Economics.

Borensztein, E. and P. Mauro (2004): "The case for GDP-indexed bonds," *Economic Policy*, 165–216.

Bowman, J., P. Naylor, et al. (2016): "GDP-linked Bonds," *Reserve Bank of Australia Bulletin, September Quarter*, 61–68.

Cabrillac, B., L. Gauvin, J.-B. Gossé, and F. Lalanne (2018): "GDP-indexed bonds: a solution to debt crises?" Eco Notepad Post 62, Banque de France, Available at https://www.banque-france.fr/index.php/en/publications-and-statistics/publications/gdp-indexed-bonds-solution-debt-crises.

Checherita-Westphal, C. and P. Rother (2012): "The impact of high government debt on economic growth and its channels: An empirical investigation for the euro area," *European Economic Review*, 56, 1392–1405.

Consiglio, A. and S. Zenios (2018): "Pricing and hedging GDP-linked bonds in incomplete markets," *Journal of Economic Dynamics and Control*, 88, 137–155.

Constantinides, G. M., M. Czerwonko, J. Carsten Jackwerth, and S. Perrakis (2011): "Are Options on Index Futures Profitable for Risk-Averse Investors? Empirical Evidence," *The Journal of Finance*, 66, 1407–1437.

De Roon, F. A., T. E. Nijman, and B. J. Werker (2003): "Currency hedging for international stock portfolios: The usefulness of mean–variance analysis," *Journal of Banking & Finance*, 27, 327–349.

Demertzis, M. and S. Zenios (2019): "State Contingent Debt as Insurance for Euro-Area Sovereigns," *Journal of Financial Regulation*, 5, 64–90.

DeMiguel, V., L. Garlappi, and R. Uppal (2009): "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" *The Review of Financial Studies*, 22, 1915–1953.

Eguren-Martin, F., A. Meldrum, and W. Yan (2020): "No-arbitrage pricing of GDP-linked bonds," *Bank of England*.

Fama, E. F. and K. R. French (1989): "Business conditions and expected returns on stocks and bonds," *Journal of Financial Economics*, 25, 23–49.

Froot, K., D. Scharfstein, and J. Stein (1989): "LDC Debt: Forgiveness, Indexation, and Investment Incentives," *Journal of Finance*, 44, 1335–1350.

Hodder, J. E., J. C. Jackwerth, and O. Kolokolova (2015): "Improved portfolio choice using second-order stochastic dominance," *Review of Finance*, 19, 1623–1647.

Huberman, G. and S. Kandel (1987): "Mean-variance spanning," *The Journal of Finance*, 42, 873–888.

IMF (2017): "State-contingent debt instruments for sovereigns," Staff report, International Monetary Fund.

Jackwerth, J. C., G. M. Constantinides, and S. Perrakis (2009): "Mispricing of S&P 500 Index Options," *Review of Financial Studies*, 22, 1247–1277.

Kamstra, M. and R. Shiller (2009): "The case for trills: Giving the people and their pension funds a stake in the wealth of the nation," Discussion Paper 1717, Cowles Foundation for Research in Economics, Yale University, New Haven, CT.

Kim, J. and J. Ostry (2018): "Boosting fiscal space: the roles of GDP-linked debt and longer maturities," Research Departmental Paper 18/04, International Monetary Fund.

Krugman, P. (1988): "Financing vs. Forgiving a Debt Overhang," *Journal of Development Economics*, 29, 253–268.

Leung, P.-L. and W.-K. Wong (2006): "On testing the equality of the multiple Sharpe Ratios, with application on the evaluation of iShares," *Available at SSRN 907270.*

Levy, H. (2016): *Stochastic dominance: Investment Decision Making Under Uncertainty*, 3rd Edition, Springer.

Mankiw, N. G., D. Romer, and D. N. Weil (1992): "A contribution to the empirics of economic growth," *The Quarterly Journal of Economics*, 107, 407–437.

Perrakis, S. (2019): *Stochastic Dominance Option Pricing: An Alternative Approach to Option Market Research*, ISBN: 9783030115906, Springer.

Pissarides, C. A. (2000): *Equilibrium unemployment theory*, 2nd Edition, The MIT Press.

Rockafellar, R. T., S. Uryasev, et al. (2000): "Optimization of conditional value-at-risk," *Journal of Risk*, 2, 21–42.

Rodríguez-Pose, A. and M. Di Cataldo (2014): "Quality of government and innovative performance in the regions of Europe," *Journal of Economic Geography*, 15, 673–706.

Russell, W. R. and T. K. Seo (1989): "Representative sets for stochastic dominance rules," in *Studies in the Economics of Uncertainty*, Springer, 59–76.

Saint-Paul, G. (1992): "Fiscal policy in an endogenous growth model," *The Quarterly Journal of Economics*, 107, 1243–1259.

Shaikh, A. M. and J. K. Moudud (2004): "Measuring capacity utilization in OECD countries: A cointegration method," Tech. rep., Working papers, No. 415, Levy Economics Institute of Bard College, Annandale-on-Hudson, NY, also available as http://hdl.handle.net/10419/31467.

Sharpe, W. (1994): "The Sharpe ratio," *The Journal of Portfolio Management*, 21, 49–58.

Simaan, Y. (1993): "What is the opportunity cost of mean-variance investment strategies?" *Management Science*, 39, 578–587.

Solow, R. M. (1956): "A contribution to the theory of economic growth," *The Quarterly Journal of Economics*, 70, 65–94.

——— (1957): "Technical change and the aggregate production function," *The Review of Economics and Statistics*, 312–320.

Sortino, F. A. and R. Van Der Meer (1991): "Downside risk," *Journal of Portfolio Management*, 17, 27.

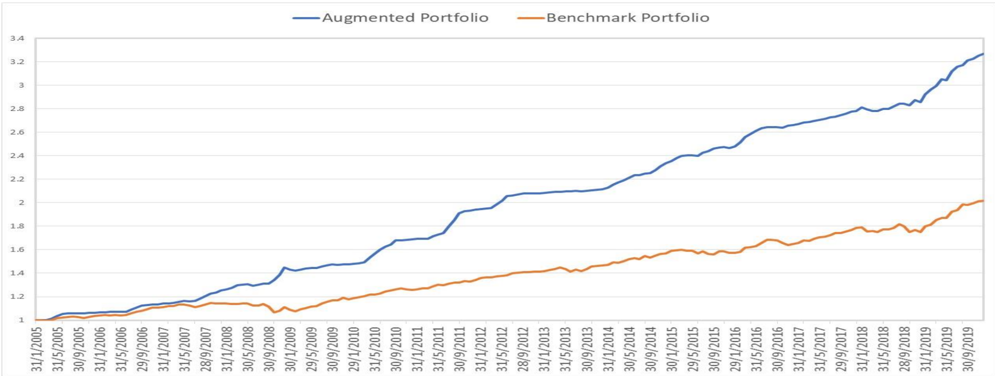
Whang, Y.-J. (2019): *Econometric Analysis of Stochastic Dominance: Concepts, Methods, Tools, and Applications*, Cambridge University Press.

Ziemba, W. T. (2005): "The symmetric downside-risk Sharpe ratio," *The Journal of Portfolio Management*, 32, 108–122.

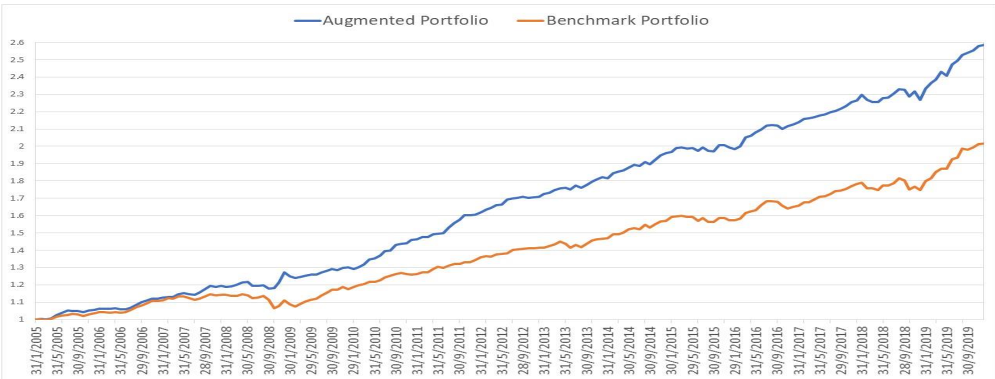
Figure 1: Cumulative out-of-sample returns of the benchmark and augmented optimal portfolios

This figure plots the out-of-sample cumulative returns of the benchmark optimal portfolio and the optimal augmented portfolio with Floaters-10yr (Panel A) and Floaters-Barclays (Panel B), for different levels of $\bar{g}$, and with Linkers (Panel C). The dataset spans the period Jan. 1980–Dec. 2019, for a total of 480 monthly returns, with out-of-sample testing conducted over the period Jan. 2005 - Dec. 2019.

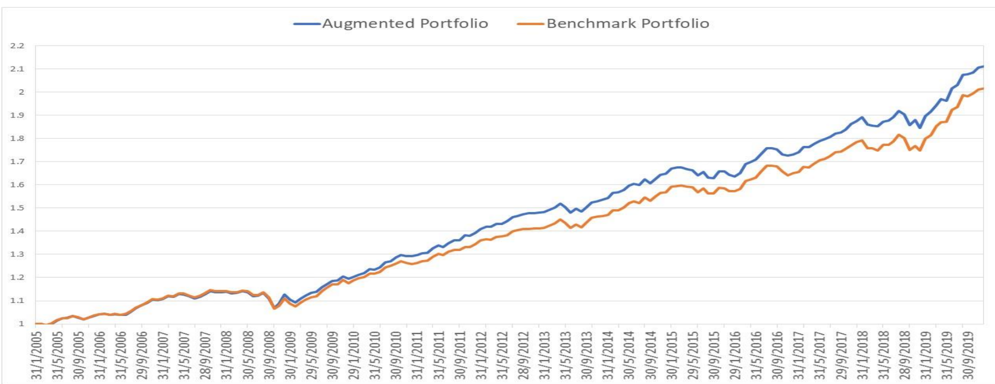
(a) Floaters-10yr



(i) $\bar{g} = 0.66\%$

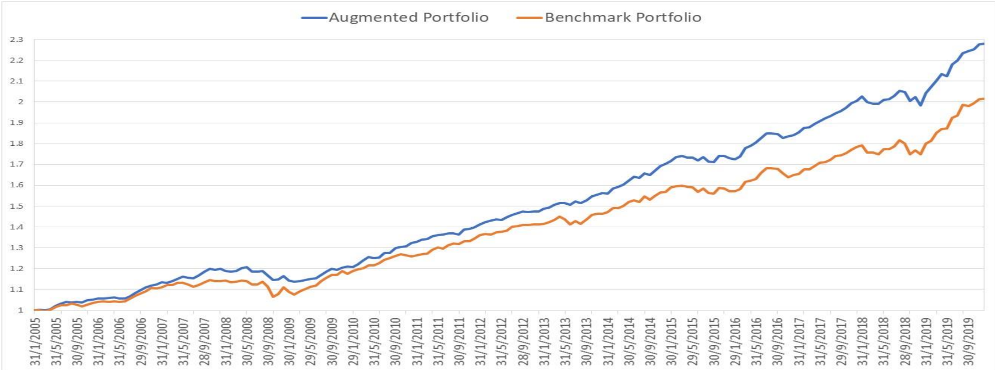


(ii) $\bar{g} = 2\%$

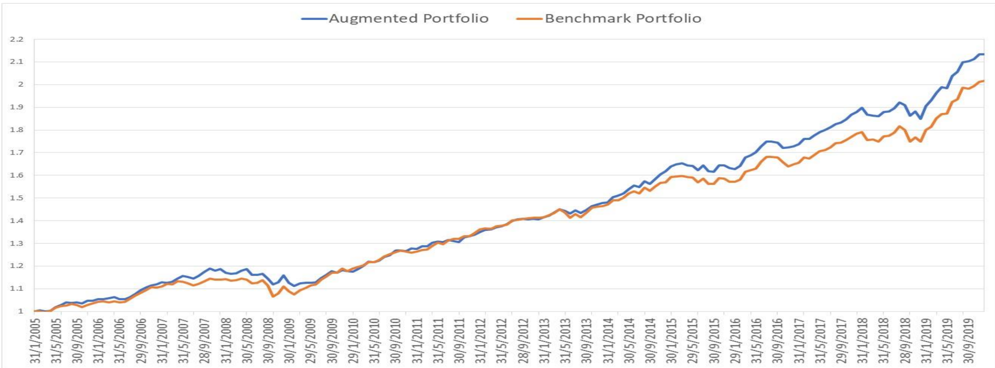


(iii) $\bar{g} = 3\%$

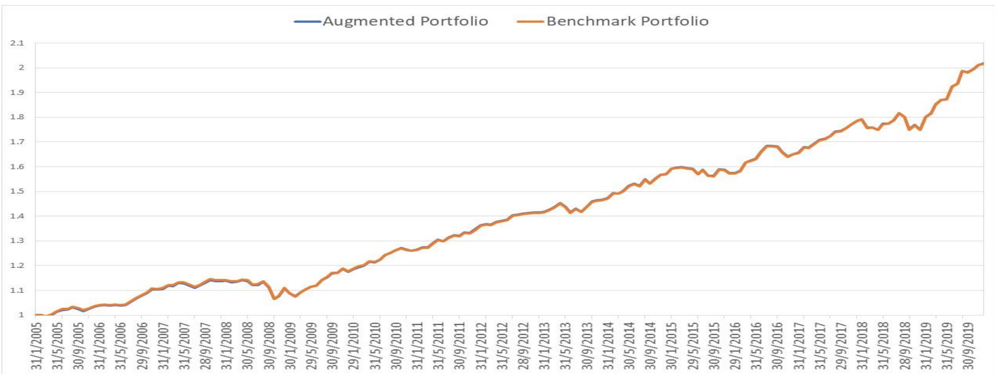
(b) Floaters-Barclays



(i) $\bar{g} = 0.66\%$



(ii) $\bar{g} = 1\%$



(iii) $\bar{g} = 2\%$

(c) Linkers

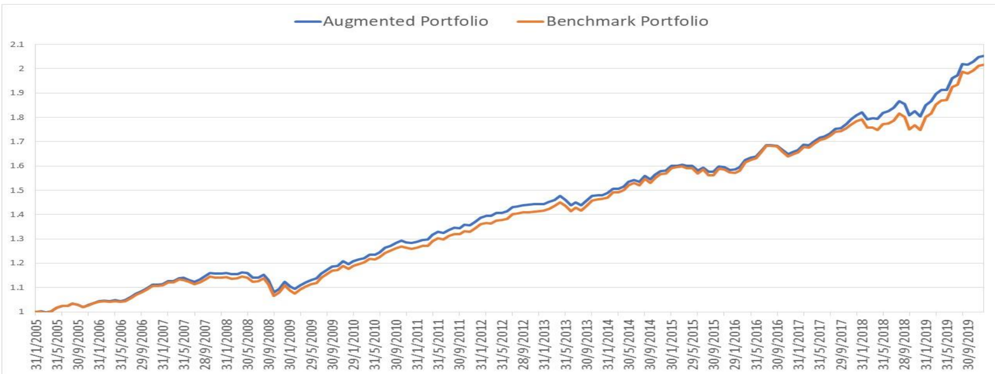


Table 1: Descriptive statistics of monthly returns

Entries report the descriptive statistics on monthly returns for the alternative asset classes that proxie the benchmark asset universe in Panel A and the augmented with GDP-linked bonds (Floaters-10yr and Floaters-Barclays for different levels of $\bar{g}$, and Linkers) in Panel B. The last column presents the p-values of the Jarque-Bera test in order to test the null hypothesis that the data are normally distributed at the 1% significance level. Data spans the period 31/1/1980 - 31/12/2019.

	Mean	St.Dev.	Skewness	Kurtosis	Sharpe Ratio	JB test
(a) Benchmark Assets						
S&P 500	0.008	0.043	-0.659	5.243	0.108	0.001
MSCI World	0.007	0.042	-0.649	4.679	0.084	0.001
Fama-French Market	0.010	0.044	-0.743	5.350	0.156	0.001
Barclays U.S. Aggregate	0.006	0.015	0.979	10.18	0.181	0.001
Barclays U.S. Corporate	0.007	0.020	0.537	8.957	0.175	0.001
US Bench. 5Yr Govt.Index	0.006	0.016	0.717	7.183	0.136	0.001
US Bench. 10Yr Govt.Index	0.006	0.024	0.392	4.702	0.118	0.001
US Bench. 30Yr Govt.Index	0.007	0.038	0.383	4.782	0.104	0.001
Fama-French SMB	0.000	0.030	0.663	10.99	-0.086	0.001
Fama-French HML	0.002	0.029	0.155	4.947	-0.038	0.001
3M US T-Bill	0.003	0.003	0.751	3.230	-	0.001
(b) GDP-linked Bonds						
Floaters-10yr						
$\bar{g} = 0$	0.011	0.013	1.559	5.434	0.583	0.001
$\bar{g} = 0.66$	0.009	0.012	1.720	5.975	0.488	0.001
$\bar{g} = 1$	0.009	0.012	1.807	6.295	0.437	0.001
$\bar{g} = 1.5$	0.008	0.012	1.941	6.837	0.369	0.001
$\bar{g} = 2$	0.007	0.011	2.099	7.526	0.311	0.001
$\bar{g} = 2.5$	0.006	0.010	2.465	8.302	0.251	0.001
$\bar{g} = 3$	0.005	0.010	2.441	9.187	0.193	0.001
Floaters-Barclays						
$\bar{g} = 0$	0.009	0.009	2.219	10.81	0.658	0.001
$\bar{g} = 0.66$	0.007	0.009	2.467	12.03	0.470	0.001
$\bar{g} = 1$	0.007	0.009	2.642	13.00	0.380	0.001
$\bar{g} = 1.5$	0.005	0.008	2.933	14.76	0.225	0.001
$\bar{g} = 2$	0.004	0.008	3.263	16.96	0.136	0.001
$\bar{g} = 2.5$	0.004	0.007	3.615	19.55	0.021	0.001
$\bar{g} = 3$	0.003	0.007	3.988	22.60	-0.083	0.001
Linkers	0.004	0.003	0.415	7.094	0.351	0.001

Table 2: Spanning tests for GDP-linked bonds without and with risk premium

Stochastic Spanning tests for the GDP-linked bonds. Panel A is for coupon-indexed GDP-linked bonds in case where C_{0t} is the returns of the US Benchmark 10 Years Government Bond Index without and with 50 basis points risk premium, for different levels of $\bar{g}$, and Panel B is for the case where C_{0t} is the returns of the Barclays Bond Index without and with 50 basis points risk premium, for different levels of $\bar{g}$, respectively. Different levels of $\bar{g}$ are given in %. Panel C is for principal-indexed GD-linked bonds without and with 50 basis points. Entries report the test statistics η_T as well as the regression estimates q_T^{BC} in order to test in-sample the null hypothesis. NR means that the model was not run to obtain the test statistic and the regression estimate because the result will not change from the last run for which data are reported. The dataset spans the period 31/1/1980 - 31/12/2019, for a total of 480 monthly returns.

(a) Floaters-10yr

	(i) 0bp premium			(ii) 50bp premium		
$\bar{g}$	η_T	q_T^{BC}	Result	η_T	q_T^{BC}	Result
0.00	NR	NR	Reject Spanning	NR	NR	Reject Spanning
0.66	0.0583	0.0450	Reject Spanning	NR	NR	Reject Spanning
1.00	0.0487	0.0373	Reject Spanning	NR	NR	Reject Spanning
1.50	0.0379	0.0286	Reject Spanning	NR	NR	Reject Spanning
2.00	0.0288	0.0226	Reject Spanning	NR	NR	Reject Spanning
2.50	0.0202	0.0173	Reject Spanning	NR	NR	Reject Spanning
3.00	0.0125	0.0124	Reject Spanning	0.0829	0.0655	Reject Spanning
3.50	0.0056	0.0084	Spanning	0.0758	0.0591	Reject Spanning
4.00	0.0024	0.0059	Spanning	0.0691	0.0544	Reject Spanning

(b) Floaters-Barclays

	(i) 0bp premium			(ii) 50bp premium		
$\bar{g}$	η_T	q_T^{BC}	Result	η_T	q_T^{BC}	Result
0.00	NR	NR	Reject Spanning	NR	NR	Reject Spanning
0.66	0.0383	0.0325	Reject Spanning	NR	NR	Reject Spanning
1.00	0.0255	0.0216	Reject Spanning	NR	NR	Reject Spanning
1.50	0.0136	0.0102	Reject Spanning	NR	NR	Reject Spanning
2.00	0.0042	0.0037	Reject Spanning	0.0792	0.0544	Reject Spanning
2.50	0.0000	0.0003	Spanning	0.0673	0.0466	Reject Spanning
3.00	0.0000	0.0000	Spanning	0.0572	0.0414	Reject Spanning
3.50	NR	NR	Spanning	0.0491	0.0381	Reject Spanning
4.00	NR	NR	Spanning	0.0432	0.0357	Reject Spanning

(c) Linkers

(i) 0bp premium			(ii) 50bp premium		
η_T	q_T^{BC}	Result	η_T	q_T^{BC}	Result
0.0177	0.0152	Reject Spanning	0.1070	0.0731	Reject Spanning

Table 3: Spanning tests for GDP-linked bonds with negative risk premia

Stochastic Spanning tests for the GDP-linked bonds. Panel A is for coupon-indexed GDP-linked bonds in case where C_{0t} is the returns of the US Benchmark 10 Years Government Bond Index with negative risk premia, -25bps and -50bps, for different levels of $\bar{g}$, and Panel B is for the case where C_{0t} is the returns of the Barclays Bond Index with negative risk premia, -25bps and -50bps, for different levels of $\bar{g}$, respectively. Different levels of $\bar{g}$ are given in %. Panel C is for principal-indexed GD-linked bonds with negative risk premia, -5bps and -10bps. Entries report the test statistics η_T as well as the regression estimates q_T^{BC} in order to test in-sample the null hypothesis. NR means that the model was not run to obtain the test statistic and the regression estimate because the result will not change from the last run for which data are reported. The dataset spans the period 31/1/1980 - 31/12/2019, for a total of 480 monthly returns.

(a) Floaters-10yr

	(i) -25bp premium			(ii) -50bp premium		
$\bar{g}$	η_T	q_T^{BC}	Result	η_T	q_T^{BC}	Result
0.00	0.0401	0.0295	Reject Spanning	0.0051	0.0046	Reject Spanning
0.66	0.0191	0.0160	Reject Spanning	0.0002	0.0007	Spanning
1.00	0.0092	0.0099	Spanning	NR	NR	Spanning
1.50	NR	NR	Spanning	NR	NR	Spanning
2.00	NR	NR	Spanning	NR	NR	Spanning
2.50	NR	NR	Spanning	NR	NR	Spanning
3.00	NR	NR	Spanning	NR	NR	Spanning

(b) Floaters-Barclays

	(i) -25bp premium			(ii) -50bp premium		
$\bar{g}$	η_T	q_T^{BC}	Result	η_T	q_T^{BC}	Result
0.00	0.0256	0.0238	Reject Spanning	0.0000	0.0017	Spanning
0.66	0.0025	0.0058	Spanning	NR	NR	Spanning
1.00	NR	NR	Spanning	NR	NR	Spanning
1.50	NR	NR	Spanning	NR	NR	Spanning
2.00	NR	NR	Spanning	NR	NR	Spanning
2.50	NR	NR	Spanning	NR	NR	Spanning
3.00	NR	NR	Spanning	NR	NR	Spanning

(c) Linkers

(i) -5bp premium			(ii) -10bp premium		
η_T	q_T^{BC}	Result	η_T	q_T^{BC}	Result
0.0107	0.0089	Reject Spanning	0.0030	0.0058	Spanning

Table 4: Out-of-sample performance with non-parametric stochastic dominance test

Entries report test statistics and p-values for stochastic non-dominance test of the augmented portfolios with respect to the benchmark portfolio. The dataset spans the period 31/1/1980 - 31/12/2019, for a total of 480 monthly returns, with out-of-sample testing conducted over the period 31/1/2005 - 31/12/2019.

	Floaters-10yr			Floaters-Barclays			Linkers
	$\bar{g} = 0.66$	$\bar{g} = 2$	$\bar{g} = 3$	$\bar{g} = 0.66$	$\bar{g} = 1$	$\bar{g} = 2$	
Test statistic	-0.0034	-0.0047	-0.0012	-0.0031	-0.0014	-0.0008	-0.0019
p-value	(0.042)**	(0.049)**	(0.075)*	(0.049)**	(0.050)*	(0.12)	(0.088)*

Table 5: Out-of-sample performance with parametric measures

Entries report the performance measures (Sharpe ratio, downside Sharpe ratio, UP ratio, portfolio turnover, return-loss and opportunity cost) for the benchmark (Panel A) and the augmented portfolio with coupon-indexed GDP-linked bonds in case where the C_{0t} is the returns of the US Benchmark 10 Years Government Bond Index (Panel B), and the returns of the Barclays Bond Index (Panel C), for different values of target GDP growth rate $\bar{g}$ as well as the augmented portfolio with principal-indexed GDP-linked bonds (Panel D). The test for the equality of Sharpe ratios reports the p-values of the null hypothesis that the difference of Sharpe ratios of the benchmark from the augmented portfolio is zero. Turnover, return-loss, and opportunity cost (θ) are in %. The opportunity cost is reported for different degrees of absolute risk aversion (ARA) for the exponential utility function, and different degrees of relative risk aversion (RRA) for the power utility function. The dataset spans the period 31/1/1980 - 31/12/2019, for a total of 480 monthly returns, with out-of-sample testing conducted over the period 31/1/2005 - 31/12/2019.

	(a) Benchmark	(b) Floaters-10yr			(c) Floaters-Barclays			(d) Linkers
		$\bar{g} = 0.66$	$\bar{g} = 2$	$\bar{g} = 3$	$\bar{g} = 0.66$	$\bar{g} = 1$	$\bar{g} = 2$	
Performance Measures								
Sharpe ratio	0.291	0.667	0.461	0.317	0.432	0.347	0.294	0.323
Sharpe equality test		(0.000)	(0.000)	(0.028)	(0.000)	(0.036)	(0.196)	(0.000)
Downside Sharpe	0.348	2.580	0.761	0.394	0.610	0.443	0.353	0.385
UP ratio	0.864	3.178	1.357	0.927	1.180	0.987	0.873	0.898
Portfolio turnover	6.952	2.913	5.107	16.12	3.071	4.364	8.923	6.969
Return-loss (RLM)		0.808	1.581	1.185	1.531	1.349	1.110	1.278
Opportunity cost (θ)								
Exponential utility								
ARA=2		0.273	0.140	0.026	0.071	0.032	0.000	0.011
ARA=4		0.276	0.142	0.026	0.074	0.034	0.000	0.012
ARA=6		0.280	0.144	0.026	0.077	0.035	0.001	0.013
Power utility								
RRA=2		0.273	0.140	0.026	0.071	0.032	0.000	0.011
RRA=4		0.276	0.142	0.026	0.074	0.034	0.000	0.012
RRA=6		0.280	0.144	0.026	0.077	0.035	0.001	0.013

Table 6: Out-of-sample average portfolio composition

Entries report the average composition of the benchmark optimal portfolio (Panel A) and the augmented optimal portfolio with coupon-indexed GDP-linked bonds in case where the C_{0t} is the returns of the US Benchmark 10 Years Government Bond Index (Panel B), and the returns of the Barclays Bond Index (Panel C), for different values of target GDP growth rate $\bar{g}$ as well as the optimal augmented portfolio with principal-indexed GDP-linked bonds (Panel D). All entries are given in %. The dataset spans the period 31/1/1980 - 31/12/2019, for a total of 480 monthly returns, with out-of-sample testing conducted over the period 31/1/2005 - 31/12/2019.

	(a) Benchmark	(b) Floaters-10yr			(c) Floaters-Barclays			(d) Linkers
		$\bar{g} = 0.66$	$\bar{g} = 2$	$\bar{g} = 3$	$\bar{g} = 0.66$	$\bar{g} = 1$	$\bar{g} = 2$	
S&P 500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MSCI World	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fama-French Market	17.39	9.721	23.01	19.03	21.35	23.27	17.34	16.54
Barclays US Aggr.	70.36	0.000	0.000	28.73	0.000	0.000	69.34	61.42
Barclays US Corp.	5.362	0.000	3.417	17.97	0.000	1.747	5.570	4.601
US Bench. 5Yr Govt.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
US Bench. 10Yr Govt.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
US Bench. 30Yr Govt.	2.752	0.000	3.580	5.430	5.380	9.908	2.668	3.457
Fama-French SMB	0.000	0.431	0.000	0.000	0.000	0.000	0.000	0.000
Fama-French HML	2.583	1.133	0.000	1.847	0.000	0.000	2.350	2.304
3M US T-Bill	1.558	1.715	0.002	0.000	0.004	0.000	0.013	0.000
GDP-linked Bond		87.00	69.99	26.99	73.27	65.07	2.719	11.68

Appendix: Factors of return and GMM regressions

Table A1: Financial and Macro Factors

List of financial and macroeconomic factors and transformations used. All the macro data are in constant prices and the SA indicates that they have been seasonally adjusted at source. All data selected from Thomson Reuters Datastream except to the 1M T-Bill that is selected from Keneth French data library.

US Financial & Macro Factors	Transformations Used	Variable Name/Description	Source
Default Spread	BAA-AAA	Excess of the yield on long-term corporate BAA-rated bonds over the yield on AAA-rated bonds by Moody's	Federal Reserve
Term Spread	AAA-1M T-Bill	Difference between the yield on AAA-rated bond and the one-month T-Bill rate	Federal Reserve & Kenneth French Data Library
FFR	Levels	Federal Funds Rate (%)	Main Economic Indicators, OECD
Debt-to-GDP	Levels	Total Debt-to-GDP Ratio (%, SA)	Federal Reserve of Bank St. Louis
Inflation	Levels	Growth Rate of CPI index (SA)-Inflation Rate	Bureau of Labor Statistics, U.S. Department of Labor
CapUtil	Levels	Capacity Utilization Rate (%, SA)	Federal Reserve
Unempl	Levels	Unemployment Rate (%, SA)	Bureau of Labor Statistics, U.S. Department of Labor

Table A2: Correlation Matrix of the Factors

Entries report the correlation coefficients between all the factors that we use to explain the performance of the GDP-linked bonds, in quarterly data.

	Default Spread	Term Spread	FFR	Debt-to-GDP	Inflation	CapUtil	Unempl
Default Spread	1						
Term Spread	-0.05	1					
FFR	-0.23	0.71	1				
Debt-to-GDP	0.02	-0.81	-0.77	1			
Inflation	-0.16	0.51	0.63	-0.52	1		
CapUtil	-0.51	0.48	0.78	-0.50	0.58	1	
Unempl	0.32	0.13	-0.37	0.13	-0.18	-0.47	1

Table A3: Results of GMM Regressions using Quarterly Data

The table reports the GMM estimates of the returns of Floaters-10yr (Panel A), Floaters-Barclays (Panel B) and Linkers (Panel C) on the financial and macro factors. In case of the floaters the dependent variable for the regression is the quarterly excess return of the difference $C_t - C_{0t}$ that is used as return of the coupon-indexed GDP-linked bond with base coupon,C_{0t}, the returns of the US Benchmark 10 Years Government Bond Index (Panel A), as well as the returns of the Barclays Bond Index (Panel B), and with three different levels of target growth $\bar{g}$ given in %. The independent variables are the financial and macro factors. The abbreviations for the macro explanatory variables are explained in Table A1. The instrumental variables used are Default Spread(t-2), ΔDefault Spread(t-2), ΔDefault Spread(t-3), ΔDefault Spread(t-4), Term Spread(t-2), ΔTerm Spread(t-2), ΔTerm Spread(t-3), ΔTerm Spread(t-4). Coefficients estimates with p-values in parentheses are reported. The dataset spans 1986Q_1-2019Q_4, the number of observations is 136 and after adjustments is 131. The *, ** and *** asterisks indicate that the coefficient estimates are statistically significant at 10%, 5% and 1% significance level, respectively. The R^2 is reported, as well as, a number of diagnostic tests such as the Hansen J-test with null hypothesis that the over-identifying restrictions are valid and the Arellano-Bond AR(1) to AR(3) tests with null hypothesis of no autocorrelation.

	Dependent variable for different $\bar{g}$ (%)						
	(a) Floaters-10yr			(b) Floaters-Barclays			(c) Linkers
$\bar{g}$	0.66	2	3	0.66	1	2	
Default Spread	0.864**	1.202**	1.419***	-0.599***	-0.521***	-0.260	-0.551***
	(0.028)	(0.013)	(0.008)	(0.000)	(0.004)	(0.279)	(0.000)
Term Spread	-0.504***	-0.783***	-0.956***	-0.080	-0.143	-0.383***	0.023
	(0.008)	(0.001)	(0.001)	(0.369)	(0.154)	(0.005)	(0.741)
FFR	-0.060	-0.010	-0.013	-0.220***	-0.215***	-0.176**	-0.178***
	(0.529)	(0.932)	(0.925)	(0.000)	(0.000)	(0.011)	(0.000)
Debt-to-GDP	-0.034***	-0.046***	-0.054***	-0.009*	-0.012**	-0.021***	-0.001
	(0.002)	(0.003)	(0.004)	(0.069)	(0.028)	(0.008)	(0.841)
Inflation	-1.842***	-2.109***	-2.184***	-0.432*	-0.457*	-0.624**	-0.159
	(0.000)	(0.000)	(0.001)	(0.063)	(0.066)	(0.047)	(0.374)
CapUtil	0.055***	0.064**	0.070**	0.023**	0.025**	0.034***	0.016**
	(0.004)	(0.010)	(0.014)	(0.012)	(0.013)	(0.010)	(0.024)
Unempl	0.265***	0.364***	0.427***	0.057	0.074	0.136*	0.049
	(0.005)	(0.005)	(0.006)	(0.136)	(0.102)	(0.051)	(0.125)
Observations	131	131	131	131	131	131	131
R^2	0.268	0.269	0.313	0.589	0.637	0.685	0.629
p-value	0.434	0.521	0.588	0.104	0.121	0.399	0.250
Hansen J-test							
p-value	0.755	0.518	0.430	0.331	0.336	0.334	0.327
Arrellano-Bond AR-1							
p-value	0.541	0.493	0.576	0.774	0.984	0.769	0.323
Arrellano-Bond AR-2							
p-value	0.335	0.407	0.870	0.382	0.372	0.339	0.338
Arrellano-Bond AR-3							

Table A4: Results of GMM Regressions using Monthly Data

The table reports the GMM estimates of the returns of Floaters-10yr (Panel A), Floaters-Barclays (Panel B) and Linkers (Panel C) on the financial and macro factors. In case of the floaters the dependent variable for the regression is the quarterly excess return of the difference $C_t - C_{0t}$ that is used as return of the coupon-indexed GDP-linked bond with base coupon, C_{0t}, the returns of the US Benchmark 10 Years Government Bond Index (Panel A), as well as the returns of the Barclays Bond Index (Panel B), and with three different levels of target growth $\bar{g}$ given in %. The independent variables are the financial and macro factors. The abbreviations for the macro explanatory variables are explained in Table A1. The instrumental variables used are lags from (t-2) to (t-12) for Default Spread and Term Spread. Coefficients estimates with p-values in parentheses are reported. The dataset spans 31/1/1986 - 31/12/2019, the number of observations is 408 and after adjustments is 396. The *, ** and *** asterisks indicate that the coefficient estimates are statistically significant at 10%, 5% and 1% significance level, respectively. Year-fixed effects are included in all cases of GDP-linked bonds. The R^2 is reported, as well as, a number of diagnostic tests such as the Hansen J-test with null hypothesis that the over-identifying restrictions are valid and the Arellano-Bond AR(1) to AR(3) tests with null hypothesis of no autocorrelation.

	Dependent variable for different $\bar{g}$ (%)						
	(a) Floaters-10yr			(b) Floaters-Barclays			(c) Linkers
$\bar{g}$	0.66	2	3	0.66	1	2	
Default Spread	-0.015	-0.069	-0.097	-0.100***	-0.124***	-0.202***	-0.290***
	(0.902)	(0.657)	(0.588)	(0.003)	(0.003)	(0.003)	(0.000)
Term Spread	-0.447***	-0.691***	-0.890***	-0.201***	-0.254***	-0.404***	-0.059**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.046)
FFR	-0.315***	-0.422***	-0.455***	-0.128***	-0.163***	-0.272***	-0.097***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Debt-to-GDP	-0.097	-0.108	-0.096	-0.056***	-0.054**	-0.056	-0.056**
	(0.142)	(0.236)	(0.363)	(0.005)	(0.031)	(0.188)	(0.010)
Inflation	1.087*	1.564*	2.052**	0.184	0.339	1.170***	0.296*
	(0.083)	(0.062)	(0.033)	(0.377)	(0.191)	(0.005)	(0.070)
CapUtil	0.202	0.073	0.006	0.271***	0.268***	0.208	0.104
	(0.353)	(0.802)	(0.985)	(0.000)	(0.002)	(0.148)	(0.119)
Unempl	-0.119	-0.229	-0.305*	0.009	-0.013	-0.126*	-0.023
	(0.252)	(0.105)	(0.065)	(0.808)	(0.757)	(0.059)	(0.593)
Observations	396	396	396	396	396	396	396
R^2	0.094	0.096	0.088	0.187	0.186	0.165	0.259
Time-fixed effects	year	year	year	year	year	year	year
p-value	0.407	0.446	0.405	0.659	0.571	0.318	0.166
Hansen J-test							
p-value	0.316	0.316	0.317	0.316	0.317	0.317	0.314
Arrellano-Bond AR-1							
p-value	0.318	0.327	0.334	0.317	0.318	0.319	0.298
Arrellano-Bond AR-2							
p-value	0.327	0.326	0.325	0.323	0.323	0.328	0.327
Arrellano-Bond AR-3							