Article:

Golinski, Adam orcid.org/0000-0001-8603-1171 and Spencer, Peter orcid.org/0000-0002-5595-5360 (2017) The advantages of using excess returns to model the term structure. Journal of Financial Economics. pp. 163-181. ISSN 0304-405X

https://doi.org/10.1016/j.jfineco.2017.05.001

The Advantages of Using Excess Returns to Model the Term Structure *†

Adam Goliński **Peter Spencer**

Abstract

We advocate the use of excess returns rather than yields or log prices in analysing the risk neutral dynamics of the term structure. We show that under standard assumptions, excess returns are affine in the risk neutral innovations in the factors. This framework has several important advantages. First, it allows for an easy estimation of models that are more flexible than the $AR(1)$. Indeed, we estimate models with more general dynamics, like $ARFIMA(p, d, q)$, almost as easily as $AR(1)$. Second, within our framework the dimension of the unrestricted model is the same for the $AR(1)$ as it is for the richer models, and does not expand in line with the state vector as it does in a yield or log price framework. This makes it appropriate to test all of these risk neutral dynamic specifications against the same OLS unrestricted alternative. Our results for the US Treasury bond market show that the unrestricted model is preferred to the $AR(1)$ by the Bayesian Information Criterion, but the opposite conclusion is reached for more flexible models. A final advantage of the excess returns framework is that the pricing errors are much lower than for the equivalent log price system.

JEL classification: G12, C58.

Keywords: term structure, excess return framework, long memory.

---

*We are grateful to Karim Abadir, Richard Baillie, John Campbell, Laura Coroneo, Antonio Diaz Perez, Fabrizio Iacone, Francisco Jareno, Scott Joslin, Peter Phillips, Yongcheol Shin, Rob Taylor, Mike Wickens and to participants at the the 2013 EFMA and WFC conferences for helpful comments. We are also indebted to a referee for exceptionally perceptive and helpful suggestions.

†Goliński (corresponding author): University of York, Department of Economics and Related Studies; E-mail: `adam.golinski@york.ac.uk`;

Spencer: University of York, Department of Economics and Related Studies; E-mail: `peter.spencer@york.ac.uk`.

1. Introduction

This paper proposes a novel approach to the term structure of interest rates, which allows the cross section of returns to be modelled easily without necessarily adopting any specific model of the risk neutral dynamics, other than to assume them to be linear. We exploit the fact that absent arbitrage and measurement error, the forward price of any security for any future period is its risk neutral expectation for that period. Thus it can be used to replace the risk neutral expectation generated by an autoregressive model in the standard specification of the cross section, allowing less restrictive dynamic specifications to be used. It follows that the excess logarithmic return on any security can be viewed as a risk neutral surprise or innovation plus a Jensen term determined by its volatility. Assuming that these excess returns have a factor structure, this approach yields an affine model that relates M excess bond returns that are measured with error to $K < M$ factor *innovations*, which can be represented by excess returns on bond portfolios that are assumed to be measured without error. The loadings of the excess returns on the factor innovations depend upon the risk neutral factor dynamics, allowing these to be identified and estimated.

In the absence of no-arbitrage restrictions, this model can be estimated simply by regressing the M excess bond returns on the K factor innovations using Ordinary Least Squares (OLS).[1] This $M \times K$ model can be restricted by making appropriate assumptions about arbitrage opportunities and the risk neutral dynamics. We show that the excess return framework easily handles $ARFIMA$ processes, which include the various autoregressive (AR), moving average (MA) and long memory (LM) models that have been used in the literature. MA and LM models are notoriously difficult to estimate, but are easily handled in the excess returns framework.[2] The unrestricted OLS excess returns model provides a convenient encompass-

[1] We show that if that the factor dynamics have unrestricted linear representation (allowing them to be represented as an infinite series of current and lagged risk neutral innovations) and the no-arbitrage restriction that follows from the factor structure is neglected, this generates the $M \times K$ unrestricted benchmark OLS model of the cross section of returns. In contrast, it seems hard to interpret the unrestricted OLS yield level benchmark as a model of the factor dynamics.

[2] Goliński and Zaffaroni (2016) model yields with long memory dynamics using infinite-dimensional state vectors and estimate the system using a state-space truncation.

ing model for testing such restrictions and in particular the first order autoregressive $AR(1)$ arbitrage-free specification used extensively in the term structure literature.

As Bams and Schotman (2003) and others have noted, a variant of the excess returns model can be obtained simply by forward-differencing the standard $AR(1)$ affine term structure model ($ATSM$), which relates the M log prices or yields to the values of the K factor portfolios.[3] However, we show that this forward-difference relation only holds for the $AR(1)$ specification; richer dynamic specifications increase the dimensions of the yield $ATSM$ but not the return $ATSM$. For example, if the risk neutral dynamics are second order autoregressive ($AR(2)$) this doubles the number of lagged state variables in the risk neutral dynamics and means that the M yields are not properly explained by the K factor portfolios; another K factors are in principle needed to do this.[4] However, working with innovations removes the second and higher order lags (and, in the case of MA models, lagged error terms) from the risk neutral dynamics and preserves the $M \times K$ structure of the excess returns model.

This is the key message of this paper: the return or forward-difference format provides a very convenient framework for testing risk neutral models of the factor dynamics because it is not necessary to restrict attention to the $AR(1)$ model analyzed by Bams and Schotman (2003), Bauer (forthcoming), Adrian et al. (2013) and others who use returns data. Moreover, we find that empirically the measurement errors in the return framework are much smaller than in the corresponding yield model (see Figure 5). Also, as found by Dai and Singleton (2000), Duffee (2011a) and Adrian et al. (2013), the yield pricing errors exhibit a large degree of autocorrelation, which disappears when working with returns (see Table 3). These results suggest that the latter provides a more powerful framework for asset pricing and model testing. We use this framework to re-assess the performance of $AR(1)$ and other *ATSMs*.

Several researchers have noted that the factor loadings estimated for an $AR(1)$ yield model are typically very close to those of an unrestricted OLS regression and that the

[3] Thus the factor loadings matrices of these systems should in principle be identical.

[4] Duffee (2011b) and Adrian, Crump, and Moench (2013) have suggested that the performance of the $AR(1)$ model might be improved by introducing more factors, but this would also increase the stochastic dimension of the cross section (see Section 3.2.1).

measurement errors are only a little larger. Hamilton and Wu (2014) and Duffee (2011a) compare these specifications and show that the $AR(1)$ restrictions are nevertheless strongly rejected statistically: the differences between the $AR(1)$ and OLS benchmark are numerically small but statistically significant. Balduzzi and Chiang (2012) also find strong evidence against the $AR(1)$ $ATSM$ when testing it against the OLS. These findings reflect the fact that the pricing errors in the Treasury bond market are very small, which means, as Cochrane and Piazzesi (2008) observe, that the risk neutral model parameters are very precisely determined, making data for this market very good at comparing the performance of rival models.

We confirm the rejection of the $AR(1)$ restrictions in our data set but, in a more constructive way, we show that richer dynamic models with parameters that are constrained by asset pricing theory can beat the unrestricted benchmark model. Figure 2 shows the return loadings for the $AR(1)$ model alongside the unconstrained OLS return loadings and their 95 percent confidence interval. As in previous models, the fit looks reasonable, but many of the $AR(1)$ loadings on the first two factors lie outside the confidence interval. More worrying, Figure 1 shows that the intercept coefficients of the $AR(1)$ model are visibly and significantly flatter than those of the OLS benchmark.[5] However, the more flexible models do a better job, generating loadings that largely coincide with the OLS estimates. Reflecting these observations, on the basis of the Bayesian Information Criterion, the $AR(1)$ specification is rejected against the OLS benchmark in both yields and returns models, while the $AR(2)$ returns model is accepted against the benchmark. Further significant improvements in fit can be obtained by introducing MA and LM factor dynamics as suggested by Fenou and Meddahi (2009) and Backus and Zin (1993), respectively.

[**Insert Figure 1 near here**]

[**Insert Figure 2 near here**]

[5] Hamilton and Wu (2014) also find that the badly fitting intercept is the main reason that the $AR(1)$ model is rejected against the OLS model.

Although it is very strong, the statistical rejection of the $AR(1)$ model is very subtle. It is tempting to think that the $AR(1)$ model could be used for pricing other securities without incurring economically significant losses. However, Bams and Schotman (2003) provide a more worrying rejection. They use a fixed effects panel data model to test the three factor $AR(1)$ model in yields and forward differences and find that although the factor loadings should be the same, they are visibly as well as statistically very different (see their Figure 5). Our findings echo this result, showing that the loadings of the OLS log price model differ significantly from those of the OLS returns model. As Bams and Schotman argue, derivative pricing models based on the two different formulations of the $AR(1)$ model would give different results, particularly for longer maturities.

The paper is set out as follows. The next section sets out the theoretical model of excess returns, supported by Appendix A. Section 3 shows how we use conventional yield factor techniques to make this model operational and derive its likelihood function. Section 4 reports the empirical results and compares them with those from conventional yield models. Section 5 offers a conclusion and some suggestions for future research.

2. A Dynamic Term Structure Model for excess returns

This section sets out the theoretical model of excess returns. This is a dynamic term structure model ($DTSM$), which combines an $ATSM$ that specifies the cross section under the risk neutral measure with the return forecasting regression that models the real world dynamics. We follow Duffee (2011b) in assuming in Section 2 that the returns have an exact factor structure, before introducing measurement errors in Section 3. That section shows how we take the model to the data and compares it with existing models.

2.1. Cross section of excess returns

Let $P_{m,t}$ be the price of the $m-$maturity zero coupon bond at time t, $p_{m,t} \equiv \log(P_{m,t})$ its natural logarithm and $\mathbf{p}_t = [p_{1,t},\ p_{2,t},\ \ldots,\ p_{M,t}]'$ an $M-$vector of stacked log prices representing the cross section. The standard *ATSM* specifies a linear relation between these log prices (or equivalently, yields) and a set of $K < M$ linear factors $\mathbf{z}_t = [z_{1,t},\ z_{2,t},\ \ldots,\ z_{K,t}]'$. The spot rate, $r_t \equiv -p_{1,t}$, is modelled as the sum of the factors:

$$r_t = \mu_r + \mathbf{1}'\mathbf{z}_t, \tag{1}$$

where, assuming that the factors are mean-reverting, μ_r is the risk neutral mean of the short rate and $\mathbf{1}$ denotes a vector of ones.

Assumption 1 *There are no arbitrage opportunities and there is a risk neutral probability measure such that each price is equal to the discounted expectation of the price under this measure.*

Assumption 2 *The yield curve is spanned by K factors.*

The risk neutral expectations of the factors must be unbiased:[6]

$$\mathbf{z}_t = E^{\mathcal{Q}}_{t-1}[\mathbf{z}_t] + \mathbf{u}^{\mathcal{Q}}_t, \qquad \mathbf{u}^{\mathcal{Q}}_t \sim i.i.d.N^{\mathcal{Q}}(\mathbf{0}, \mathbf{\Sigma}) \tag{2}$$

where $\mathbf{u}^{\mathcal{Q}}_t$ is a $K-$vector of mean-zero risk neutral innovations, which we additionally assume are *i.i.d.* Gaussian. In this paper we propose an *ATSM* based on affine relation between the risk neutral innovations in the log prices and the factors:

$$p_{m,t} - E^{\mathcal{Q}}_{t-1}[p_{m,t}] = \mathbf{b}'_m \mathbf{u}^{\mathcal{Q}}_t, \qquad m = 1, ..., M. \tag{3}$$

Section 3.2 explores the relation between this and the standard $AR(1)$ *ATSM*.

[6] In this paper expectations and other terms defined under the risk neutral measure are denoted with a $\mathcal{Q}$ superscript. The lack of superscript means that the object is defined with respect to the real-world measure.

We can estimate this model by making a few minimal assumptions, consistent with linearity and the absence of arbitrage. Appendix A shows that because forward prices are risk neutral expectations in this setting, the log excess return on a bond with maturity $m+1$, $rx_{m,t} = p_{m,t} - p_{m+1,t-1} - r_{t-1}$, is equal to the risk neutral innovation in the log price plus a convexity term α_m:

$$\begin{aligned} rx_{m,t} &= \alpha_m + p_{m,t} - E^{\mathcal{Q}}_{t-1}[p_{m,t}] \\ &= \alpha_m + \mathbf{b}'_m \mathbf{u}^{\mathcal{Q}}_t. \end{aligned} \tag{4}$$

The convexity term α_m depends on the loadings on the risk neutral innovations in (3):

$$\alpha_m = -\frac{1}{2}\mathbf{b}'_m \boldsymbol{\Sigma} \mathbf{b}_m \tag{5}$$

and comes in because we are working with log prices.

We assume that the factors are mean-zero and characterised by the mean-independent linear representation under the risk neutral measure:[7]

$$z_{k,t} = \sum_{i=0}^{\infty} \beta_{k,i} u^{\mathcal{Q}}_{k,t-i}, \quad \text{for } k = 1, ..., K, \tag{6}$$

where $u^{\mathcal{Q}}_{k,t}$ are the *i.i.d.* Gaussian innovations. This allows the spot rate to be represented as

$$r_t = \mu_r + \sum_{i=0}^{\infty} \beta'_i \mathbf{u}^{\mathcal{Q}}_{t-i}, \tag{7}$$

where $\beta_i = [\beta_{1,i}, \ldots, \beta_{K,i}]'$.

Assumption 3 *Under the risk neutral probability measure the term structure factors have Gaussian linear dynamics and are mean-independent.*

[7]In other words, as in Joslin, Singleton, and Zhu (2011), the factors have independent dynamics but their innovations may be correlated as indicated in (2).

Appendix A shows that the factor innovation loadings in (4) satisfy

$$\mathbf{b}_m = -\sum\nolimits_{i=0}^{m-1} \beta_i. \tag{8}$$

The cross section of M excess returns (4) can be stacked and represented in vector notation as:

$$\mathbf{rx}_t = \alpha + \mathbf{B}'\mathbf{u}_t^{\mathcal{Q}}, \tag{9}$$

with $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_M]$ and, in the absence of arbitrage,

$$\alpha = -\frac{1}{2} Diag(\mathbf{B}'\mathbf{\Sigma}\mathbf{B}), \tag{10}$$

where for any square matrix $\mathbf{S}$, $Diag(\mathbf{S})$ is the column vector formed from its diagonal elements.

2.2. *Return forecasting regressions*

Under the real world measure $\mathcal{P}$ excess returns are not necessarily mean-zero white noise. We allow for this by changing measure and specifying the difference between the risk neutral ($\mathbf{u}_t^{\mathcal{Q}}$) and real world ($\mathbf{u}_t$) innovations as:

$$\mathbf{u}_t^{\mathcal{Q}} = \mathbf{u}_t + \mathbf{\Sigma}'\lambda_{t-1}, \tag{11}$$

where the $K-$vector λ_t contains the 'prices of factor risk'.[8] These are specified following Duffee (2002) as:

$$\mathbf{\Sigma}'\lambda_t = \lambda_0 + \mathbf{\Lambda}_1'\mathbf{x}_t, \tag{12}$$

where $\mathbf{x}_t$ is potentially an $(K+N)-$vector of risk factors. These can include the K principal components of yields that span the term structure as well as N macroeconomic and other

[8]See appendix B for details.

variables that are unspanned, meaning that they affect future returns but not current yields (Joslin, Priebsch, and Singleton (2014)). Substituting (11) and (12) into (9) gives the return forecasting regressions (*RFRs*):

$$\begin{aligned} \mathbf{rx}_t &= \alpha + \mathbf{B}'\lambda_0 + \mathbf{B}'\mathbf{\Lambda}_1'\mathbf{x}_{t-1} + \mathbf{B}'\mathbf{u}_t, \qquad (13) \\ \text{where} \quad \mathbf{u}_t &\sim N(\mathbf{0}, \mathbf{\Sigma}). \end{aligned}$$

The details of the change of measure are set out in Appendix B. (9) and (13) constitute a complete *DTSM*.

Assumption 4 *Under the real world probability measure there can be factors that are not spanned by the term structure of interest rates but are correlated with the term structure factors.*

3. The econometric specification

This section shows how we model the risk neutral dynamics using conventional yield factor techniques to make the excess returns model operational.

3.1. An approximate factor model

Our theoretical model assumes that the excess returns have an exact factor structure and can be explained by K risk neutral shocks (9). However, in practice this model is unlikely to fit perfectly, because bonds are mis-priced or their prices are observed with measurement error. There may also be model misspecification errors, due to arbitrage opportunities or the use of an inappropriate model of the dynamics. To allow for this we follow Duffee (2011b), denoting observed excess returns with a superscript 'o' and assuming these are driven by an approximate factor structure which includes so-called return pricing errors $\mathbf{v_t}$.[9] These are assumed to be Gaussian with zero mean:

$$\mathbf{rx}_t^o = \mathbf{rx}_t + \mathbf{v_t} = \alpha + \mathbf{B}'\mathbf{u_t^{\mathcal{Q}}} + \mathbf{v_t}, \tag{14}$$

$$\text{where} \qquad \mathbf{v}_t \sim i.i.d.N(\mathbf{0}, \Xi)$$

$$\text{and} \qquad E(\mathbf{v}_t | \mathbf{u}_t^{\mathcal{Q}}) = \mathbf{0}.$$

Assumption 5 *Bond returns are measured with independent and identically distributed Gaussian measurement errors that are independent of the factor innovations.*

Because these pricing errors are not part of the formal model of Section 2, they are not priced. Thus, they have the same distribution under both the risk neutral and real world probability measure. The factor model assumes that they are orthogonal to the factor innovations. Also, for convenience, we assume that they have a zero conditional mean.

To estimate the model we assume that excess returns on K bond portfolios are observed without measurement error. The portfolio weights are derived from an analysis of principal components of yields. We select the first K principal components of yields, $\mathbf{q}_{\mathbf{y},t} = \mathbf{W}'\mathbf{y}_t$, where $\mathbf{y}_t$ is a vector of yields with the same maturities as $\mathbf{p}_t$ and $\mathbf{W}$ is a matrix of eigenvectors or holdings that defines these portfolios. Applying this matrix to (14) gives the excess returns on these portfolios $\mathbf{q}_{\mathbf{x},t}$:

$$\mathbf{q}_{\mathbf{x},t} = \mathbf{W}'\mathbf{rx}_t^o = \alpha_{\mathbf{W}} + \mathbf{B}_W'\mathbf{u}_t^{\mathcal{Q}} + \mathbf{W}'\mathbf{v}_t. \tag{15}$$

where $\alpha_W = \mathbf{W}'\alpha$, $\mathbf{B}_W' = \mathbf{W}'\mathbf{B}'$. We assume that these excess return portfolios are measured without error ($\mathbf{W}'\mathbf{v}_t = \mathbf{0}$) and can be used as observable factor innovations.

[9] These measurement errors are small but play a non-trivial role, allowing the cross section to be represented by an approximate factor model with much lower dimension. As Duffee (2011b) notes, they allow other 'knife-edge' properties like the assumption that some factors are unspanned (influence returns but not yields) to be used as an approximation. Similarly, Bauer and Rudebusch (2015) argue that small measurement errors in yields can obscure the effect of macroeconomic variables on the yield curve, helping to explain the empirical finding that they are not 'spanned' by the yield curve.

Assumption 6 *There exist K portfolios of bond log excess returns that are measured without error.*

We can then back out the set of unobservable factor innovations $\mathbf{u}_t^{\mathcal{Q}}$:[10]

$$\mathbf{u}_t^{\mathcal{Q}} = \mathbf{B}_W'^{-1}(\mathbf{q}_{\mathbf{x},t} - \alpha_{\mathbf{W}}), \tag{16}$$

Substituting (16) back into (14) gives the *ATSM*:

$$\mathbf{rx}_t^o = (\mathbf{I} - \mathbf{B}'\mathbf{B}_{\mathbf{W}}'^{-1}\mathbf{W}')\alpha + \mathbf{B}'\mathbf{B}_W'^{-1}\mathbf{q}_{\mathbf{x},t} + \mathbf{v_t}. \tag{17}$$

Pre-multiplying (13) by the principal component weights $\mathbf{W}'$ (and using again $\mathbf{W}'\mathbf{v}_t = \mathbf{0}$) gives our factor *RFRs*:

$$\begin{aligned} \mathbf{q}_{\mathbf{x},t} &= \mathbf{c} + \mathbf{C}\mathbf{x}_{t-1} + \mathbf{e}_t, \\ \mathbf{e}_t &\sim N(\mathbf{0}, \mathbf{\Omega}). \end{aligned} \tag{18}$$

where: $\mathbf{c} = \alpha_W + \mathbf{B}_W'\lambda_0$, $\mathbf{C} = \mathbf{B}_W'\mathbf{\Lambda}_1'$, $\mathbf{e}_t = \mathbf{B}_W'\mathbf{u}_t$ and $\mathbf{\Omega} = \mathbf{B}_W'\mathbf{\Sigma}\mathbf{B}_W$. Since $\mathbf{c}$ and $\mathbf{C}$ are unrestricted (maximally flexible in the terminology of Joslin et al. (2011), henceforth JSZ), the Zellner (1962) theorem allows us to estimate them by *OLS*. We denote these *OLS* estimates as $\hat{\mathbf{c}}$ and $\hat{\mathbf{C}}$. The price of risk parameters λ_0 and $\mathbf{\Lambda}_1$are just-identified and can be backed out separately. Like the model of the cross section, this *RFR* system (18) could be estimated as a stand-alone model. However we estimate it jointly with the model of the cross section (17), exploiting the fact that the variance matrix $\mathbf{\Sigma}$ is the same under both measures. The likelihood function of the *DTSM* is set out in Section 4.2.

[10] Invertibility of $\mathbf{B}_W$ implicitly depends on the assumption that the factors have distinct dynamics. Joslin et al. (2011) show that the problem of repeated autoregressive roots can be dealt with using the Jordan decomposition.

3.2. Restricting the risk neutral factor dynamics

Unrestricted *OLS* estimates of the *ATSM* (9) provide unrestricted estimates of the $b_{k,m}$ coefficients that can be differenced using (8) to retrieve the parameters $\beta_{k,m}$ of the unrestricted impulse responses (6). This model is the most general specification satisfying Assumptions 2-6. We now specialise this using *ARFIMA* models to represent the loadings $b_{k,m}$ and using Eq.(10) to make the model arbitrage-free and thus satisfy Assumption 1.

3.2.1. The autoregressive process of order p

In our model the latent factors $\mathbf{z}_t$ are mean-zero and mean-independent, described by the *MA* representation (6). In the $AR(1)$ model the *MA* coefficients decline exponentially: $\beta_{k,i} = \beta_k^i$. In the absence of repeated roots, each of these components can be integrated to return the risk neutral factor dynamics:

$$z_{k,t} = E^{\mathcal{Q}}_{t-1}[z_{k,t}] + u^{\mathcal{Q}}_{k,t}, \qquad k = 1, ..., K, \tag{19}$$

where:

$$E^{\mathcal{Q}}_{t-1}[z_{k,t}] = \beta_k z_{k,t-1}. \tag{20}$$

Stacking these gives:

$$E^{\mathcal{Q}}_{t-1}[\mathbf{z}_t] = \mathbf{K}_1 \mathbf{z}_{t-1}, \tag{21}$$

where β_k is the $k-$th diagonal element of the $K \times K$ diagonal matrix $\mathbf{K}_1$. This specification makes zero coupon bond prices log-linear in the levels of the factors:

$$\mathbf{p}_t = \mathbf{a} + \mathbf{B}'\mathbf{z_t}. \tag{22}$$

JSZ and others assume that the first K principal components of yields $\mathbf{q}_{\mathbf{y},t} = \mathbf{W}'\mathbf{y}_t$ are measured without error and can be used as observable factors in the cross section, allowing the latent factors to be represented by a rotation similar to ours in (16):[11]

$$\mathbf{z}_t = \mathbf{B}_W'^{-1}(\mathbf{q}_{\mathbf{y},t} - \alpha_{\mathbf{W}}). \tag{23}$$

Forward differencing (22) then gives the excess returns relation (9). The intercept coefficients are lost in this transform but the loadings matrices $\mathbf{B}$ in the log price and excess return representations are in principle identical as noted by Bams and Schotman (2003). The $AR(1)$ model is unique in this respect. For example, consider the $AR(2)$ model:

$$E_{t-1}^{\mathcal{Q}}\left[\mathbf{z}_t\right] = \mathbf{K}_1\mathbf{z}_{t-1} + \mathbf{K}_2\mathbf{z}_{t-2}, \tag{24}$$

where $\mathbf{z}_t$ is still the $K \times 1$ vector of current state variables and $\mathbf{K}_1$ and $\mathbf{K}_2$ are $K \times K$ diagonal matrices. This can be written in a companion form as a restricted 6 factor $AR(1)$ model:

$$\begin{bmatrix} \mathbf{z}_t \\ \mathbf{z}_{t-1} \end{bmatrix} = \begin{bmatrix} \mathbf{K}_1 & \mathbf{K}_2 \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{t-1} \\ \mathbf{z}_{t-2} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_t^{\mathcal{Q}} \\ \mathbf{0} \end{bmatrix}, \tag{25}$$

The second superrow of this system incorporates a dynamic identity, which makes it less flexible than an unrestricted 6 factor $AR(1)$ model. Given our other assumptions, it can be shown that the log bond prices are affine in the state vector $\mathbf{z}_t$ and its lagged value:

$$\mathbf{p}_t = \mathbf{a} + \mathbf{B}_1'\mathbf{z}_t + \mathbf{B}_2'\mathbf{z}_{t-1}, \tag{26}$$

where $\mathbf{B}_1$ and $\mathbf{B}_2$ depend upon $\mathbf{K}_1$ and $\mathbf{K}_2$ (see e.g. Ang and Piazzesi (2003)). Relaxing these restrictions would give a conformable OLS encompassing model with $2K \times M$ free loading parameters. Naturally, the dimension of the model increases in proportion to that of

[11] Although JSZ specify their model in terms of yields rather than returns, our model frameworks are in other respects similar. We both use factors that are mean-independent under the risk neutral measure. Like their *VAR* model of the real world dynamics, the *RFRs* that we use to handle the price of risk are unrestricted and estimated by *OLS* at a preliminary stage, greatly reducing the number of parameters that we need to optimize in the maximum likelihood routine.

the state vector, which is an undesirable feature when it comes to model selection. However, forward-differencing has the effect of removing the higher order lags from any $AR(p)$ system:

$$\mathbf{p}_t - E^{\mathcal{Q}}_{t-1}\left[\mathbf{p}_t\right] = \mathbf{B}'_1\left(\mathbf{z}_t - E^{\mathcal{Q}}_{t-1}\left[\mathbf{z}_t\right]\right) = \mathbf{B}'_1\mathbf{u}^{\mathcal{Q}}_t. \tag{27}$$

Thus, for more general dynamics than $AR(1)$, the loading matrix $\mathbf{B}$ for prices in (22) will be misspecified and will, in principle, differ from the loading matrix for excess returns in (27).

3.2.2. Autoregressive fractionally integrated moving average models

In addition to the AR class of models, we also consider the more general autoregressive fractionally integrated moving average $ARFIMA(p, d, q)$ model (see Granger and Joyeux (1980), Hosking (1981)):

$$\Phi_k(L)\left(1-L\right)^{d_k} z_{k,t} = \Theta_k(L)u^{\mathcal{Q}}_{k,t}, \qquad k = 1, ..., K, \tag{28}$$

where $\Phi_k(L)$ and $\Theta_k(L)$ are polynomials of order p and q, respectively, d is the long memory parameter and L is the lag operator. This allows us to specify (6) as:

$$\beta_k(L)u^{\mathcal{Q}}_{k,t} = \Phi_k(L)^{-1}\left(1-L\right)^{-d_k}\Theta_k(L)u^{\mathcal{Q}}_{k,t}. \tag{29}$$

The process is stationary if the autoregressive roots lie outside the unit circle and the long memory parameter is smaller than 1/2. For example, the $ARFIMA(2, d, 2)$ process with zero mean for $z_{k,t}$ is given by:

$$\left(1-\phi_{k,1}L-\phi_{k,2}L^2\right)\left(1-L\right)^{d_k} z_{k,t} = \left(1+\theta_{k,1}L+\theta_{k,2}L^2\right)u^{\mathcal{Q}}_{k,t}, \tag{30}$$

To specify the values of β_k used in estimation as a function of $ARFIMA$ parameters we first note that:

$$(1-L)^{-d_k} = 1 + d_k L + d_k\left(d_k+1\right)L^2/2! + d_k\left(d_k+2\right)\left(d_k+3\right)/3! + \ldots, \tag{31}$$

which can be substituted (together with similar expansions of $\Phi_k(L)$and $\Theta_k(L)$ in terms of L) into (29). The β's are then determined by matching the coefficients on both sides of (29) at each lag (or power of L). In practice, we only need to compute these coefficients up to the maximum lag given by the highest maturity of the bond (excess return) in the sample. This allows the likelihood of the model to be written in terms of the parameters (β) of the MA process (7) driving the spot rate.

3.3. The dynamic and stochastic dimensions of an AR model

Because the dimension of the $ATSM$ in returns (9) is not affected by the dynamic specification, this allows different specifications to be tested against the same parsimonious $(K+1)\times M$ OLS benchmark. Crucially, we do not lose any information about the risk neutral factor dynamics using this transform. That is because we can use the Yule-Walker equations to rewrite any AR model like (24) as a moving average process (6), letting the $\beta_{k,m}$ parameters and hence the α_m and $\mathbf{b}_m$ coefficients of (4) depend upon AR parameters in $\mathbf{K}_1$ and $\mathbf{K}_2$. [12] Maximum likelihood estimates of the latter can then be obtained by fitting the model to the data and optimizing the likelihood function.

When working with prices rather than returns higher order risk neutral factor dynamics mean that, besides contemporaneous factor values, their lags are reflected in prices. For example, the $AR(2)$ model implies that bond prices are affected by both contemporaneous and once-lagged factor values (26). With $K=3$, its companion form (25) is of dimension 6×6, which (after rotation) means that prices could be spanned by six noiseless principal

[12]For example in the case of the mean-independent $AR(2)$ model (24), we write these as: $\beta_{k,1} = K_{1,k}\beta_{k,0}$, $\beta_{k,m} = K_{1,k}\beta_{k,m-1} + K_{2,k}\beta_{k,m-2}$ for $m>1$, where $K_{i,k}$ is the kth diagonal element of $\mathbf{K}_i$.

components. However, there is an important distinction from the richer $AR(1)$ model, which relaxes the dynamic identity: the six factor $AR(1)$ model has a full stochastic span because it incorporates six factor innovation terms rather than the three incorporated into $\mathbf{u}_t^{\mathcal{Q}}$ in (25). It would generally take six principal components or other factor portfolios to hedge any price in this model, whereas three portfolios (held with the weights $\mathbf{b}'_m\mathbf{B}'^{-1}_W$ defined in (4) and (16)), are sufficient in the 3 factor $AR(2)$ model.

3.4. Identification

We have adopted a series of normalization restrictions to identify the model. First, the short rate has unit factor loadings, as in JSZ. Second, as in Dai and Singleton (2000), Singleton (2006) and JSZ the factor means are assumed to be zero under $\mathcal{Q}$, allowing the central tendency to be determined by the short rate equation (1). Finally, the factors are ranked according to their persistence, as in Hamilton and Wu (2012).[13] Other equivalent identification schemes are possible by invariant transformation (rotation) of the factor dynamics.

We also assume that the factors are mean-independent, as in JSZ. This model is maximally flexible in the sense of Dai and Singleton (2000)) for $AR(1)$ dynamics (21). However, it is an over-identifying restriction in richer *ARFIMA* specifications.[14] In other words, it does not allow for the most general (maximally flexible) specification of the factor dynamics. For the $VAR(2)$ with $K = 3$ dynamic model in (24), a maximally flexible specification would allow 6 dynamic restrictions. For example, we could follow the JSZ rotation and assume that $\mathbf{K}_1$ was unrestricted and $\mathbf{K}_2$ diagonal. Although we can handle this in our excess return framework by allowing in the factor dynamics in (6) to be interdependent, this additional flexibility would come at a cost of a significant increase in computational complexity. Adding long memory component would further complicate modelling the dynamics of the system. As such, we use our *ARFIMA* specifications with mean-independent factors as a tractable,

[13]Specifically, the factors in long memory models are ranked in terms of their long memory and in short memory models in terms of their highest autoregressive root.

[14]We are grateful to a referee for pointing this out.

although not maximally flexible, generalization of the popular $AR(1)$ model.

3.5. The relationship with previous models of excess returns

The excess return framework (14) has also been used by Bams and Schotman (2003), Bauer (forthcoming) and Adrian et al. (2013) to model the cross section under the $AR(1)$ assumption.[15] However, the framework developed in this paper is novel in two respects. First and foremost, we use the excess return representation to handle higher order autoregressive models as well as $ARFIMA$ models that encompass the AR and $ARMA$ models used previously in this literature. [16] Second, we use excess returns to model the innovations on the right as well as the left hand side of (14), mimicking the JSZ assumption that some portfolios of yields are observed without error by assuming that the returns on these portfolios are measured without error. This allows us to exploit the consistency between excess returns that follows from the assumption of no-arbitrage when these have an approximate factor structure. In contrast, the standard approach completely ignores the information in the forward rate structure, using the $AR(1)$ specification (21) to model the risk neutral expectations instead. Bams and Schotman (2003) model the forward difference system on the right hand side of (14) but use a specific effects panel data model to represent the innovations $\mathbf{u}_\mathbf{t}^\mathcal{Q}$, while Bauer (forthcoming) uses macro 'news' to do this.

Like Bams and Schotman (2003) and Bauer (forthcoming) we model the cross section under the risk neutral measure, exploiting the precision given by the small pricing errors. In contrast, Adrian et al. (2013) estimate the parameters of the system under the real world probability measure $\mathcal{P}$. Under this measure the excess return on the $m-$period bond in (13)

[15] Bauer models innovations in forward rates rather than forward log prices, but his specification is similar to (9).

[16] Unlike the AR model, these more general specifications do not allow a representation of the cross section in terms of levels of yields or log prices like (22). However, as we shall see, it is much easier to work with the forward difference equation when working with more general dynamic specifications (13).

can be written as:

$$rx_{m,t} = \frac{1}{2}\mathbf{b}_m'\mathbf{\Sigma}\mathbf{b}_m + \mathbf{b}_m'\left(\lambda_0 + \Lambda_1'\mathbf{x}_{t-1}\right) + \mathbf{b}_m'\mathbf{u}_t + v_{m,t} \tag{32}$$

Adrian et al. (2013) adopt the $AR(1)$ assumption, modelling the factors $\mathbf{x}_{t-1}$ under $\mathcal{P}$ using a vector autoregression (VAR). They fit (32) using residuals from this VAR to represent the priced return innovations $\mathbf{u}_t$. The regression coefficients on this vector give an estimate of $\mathbf{b}_m$ and the other parameters can be estimated using cross-section regressions.[17] These can be used to construct a model of the $AR(1)$ risk neutral dynamics in (21) and hence estimates of the factor loadings in the *ATMS* (22).

4. Estimation method and results

This section presents our results, starting with a description of the data set. It shows how we use conventional yield factor techniques to make the excess returns model operational and derive its likelihood function. Our empirical work starts by examining the cross-sectional behaviour of returns, yields and log prices using (17) and (22). This strongly suggested the use of return data in subsequent analysis. We next develop a set of *RFRs* that model the behaviour of returns under the real world measure $\mathcal{P}$. Then we combine the cross-section and *RFR* systems and estimate them jointly as the *DTSM*.

4.1. Data

The empirical models are fitted to a monthly data sample extracted from two well-known data sets. The first is that of Gurkaynak, Sack, and Wright (2007), henceforth GSW. This has the attractive feature that it allows prices and yields to be generated at any maturity,

[17] The price of risk parameters λ_0 and $\Lambda_1$are just-identified in our framework, but are over-identified in Adrian et al. (2013) (since the number of bonds in the sample exceeds the number of factors). Thus they need to use a series of cross-section regressions to estimate these parameters.

which greatly facilitates the calculation of the monthly log forward prices and returns used in this study.[18] Adrian et al. (2013) also use this source to calculate the monthly returns used in their *RFRs*. Because yield data at the long end are thin (issuance of the 30 year Treasury bond was suspended between 2002 and 2006), we restrict the analysis to maturities up to 15 years. The number of actively traded securities is also sparse at the short end. Moreover, GSW only employ Treasury bonds in their sample (excluding Treasury bills and notes). They do not use any bonds with a maturity of less than four months and it is sometimes hard to reconcile their yield estimates for maturities of less than one year with data for yields on Treasury bills and notes (which are excluded from their sample). To allow for this we follow Duffee (2010) and others in taking our short maturity observations from the Fama Treasury Bill Term Structure Files available from the Center for Research in Security Prices.[19] We also took the data for our one month spot interest rate from this source.

Our cross section of yields and log prices starts at the short end with the 2, 4 and 6 month maturities extracted from the Fama Treasury Bill Term Structure files and adds those for the annual maturities 1 through 15 from GSW, giving a total of 18 maturities. Yield portfolios, $\mathbf{q}_{\mathbf{y},t}$, were estimated as the first three principal components of these yields with weights $\mathbf{W}$ given by the eigenvectors corresponding to the three largest eigenvalues of the yields covariance matrix ($\mathbf{q}_{\mathbf{y},t} = \mathbf{W}'\mathbf{y}_t$). The excess returns on these 18 maturities were then calculated ($rx_{n,t}$, with $n = 1/6, 1/3, 1/2, 1, 2, ..., 15$ years) and the weights $\mathbf{W}$ were used to find the noiseless portfolios of returns ($\mathbf{q}_{\mathbf{x},t} = \mathbf{W}'\mathbf{rx}_t$). In view of a suspected structural break in the time series behaviour following the end of the Volker experiment in 1982, we used the estimation period January 1983 to December 2011.[20]

[18]Conveniently, this gives the parameters of the interpolated yield curve, allowing us to find bond returns on a consistent time period/maturity basis. The parameters are available from from `http://www.federalreserve.gov/pubs/feds/2006/200628/feds200628.xls`.

[19]These estimates are obtained using the bootstrap method rather than curve fitting approximations, which is generally thought to produce more robust estimates at the short end. For a detailed description please see the US Treasury Database Guide at: `www.crsp.com/documentation`.

[20]Garcia and Perron (1996) using a regime switching model and structural change test found shifts in the ex-post real rate in 1973 and 1981. Bai (1997) using daily data for changes in 3 month Treasury bills found breaks in October 1979, December 1981, and August 1982.

Figure 3 presents time series for the 1 month, 3 and 15 year maturity yields. It shows the well-known downward trend in interest rates over the three decades. This trend underpinned the handsome realized returns on bonds over this period. In Figure 4 we plot the average annualized 1 month excess returns on zero coupon bonds over the 1 month rate. It reveals that over our sample period the excess return on the 15 year zero coupon bond was close to 7.7%.[21]

[**Insert Figure 3 near here**]

[**Insert Figure 4 near here**]

4.2. The likelihood function

The set of variables observable by an econometrician consists of a time series of M excess returns observable with errors, $\mathbf{rx}_t^o$, and K portfolios of these returns that are assumed to be observed without error, $\mathbf{q}_{\mathbf{x},t}$. Also, an econometrician observes a number of economy-wide variables, both spanned and unspanned by the bond market, summarized in a vector of variables $\mathbf{x}_t$. The spanned yield curve factors are the first three principal components of yields, while the unspanned factors are the fourth and fifth principal components of yields and macro variables. The significance of the conditioning variables in $\mathbf{x}_t$ is assessed in Section 4.4.

Under these assumptions the parameters of our model can be estimated by joint nonlinear optimization of the likelihood of (17) and (18). Let $\Theta \equiv (\beta, \mathbf{\Sigma}, \Xi, \lambda_0, \mathbf{\Lambda}_1)$ denote the parameters of the full model where β is a vector of parameters specifying the risk neutral dynamics β_i. Since the model is Gaussian, we estimate it by maximum likelihood (*MLE*).

[21] In this time period the average 1 month rate was 4.15%.

Conditional upon the lagged factors $\mathbf{x}_{t-1}$, the log-likelihood function is:

$$\log\left(\mathcal{L}(\Theta)\right) = \sum_{t=1}^{T} \log f(\mathbf{rx}_t^o, \mathbf{q}_{\mathbf{x},t}|\mathbf{x}_{t-1}; \Theta), \tag{33}$$

where the joint conditional density of the observed period t data can be decomposed into the likelihood (g) of observing the excess returns given the spanned factors (which is given by (17)) and the likelihood (h) of observing the latter given the lagged factors (18):

$$\begin{aligned} f(\mathbf{rx}_t^o, \mathbf{q}_{\mathbf{x},t}|\mathbf{x}_{t-1}; \Theta) &= g(\mathbf{rx}_t^o|\mathbf{q}_{\mathbf{x},t}; \Theta) \times h(\mathbf{q}_{\mathbf{x},t}|\mathbf{x}_{t-1}; \Theta) \\ &= g(\mathbf{rx}_t^o|\mathbf{q}_{\mathbf{x},t}; \beta, \mathbf{\Sigma}, \Xi) \times h(\mathbf{q}_{\mathbf{x},t}|\mathbf{x}_{t-1}; \lambda_0, \mathbf{\Lambda}_1, \beta, \mathbf{\Sigma}). \end{aligned} \tag{34}$$

Since the unspanned variables are not traded (either individually or as portfolios) they do not enter the likelihood function. Our estimation strategy is based on a direct specification of the risk neutral dynamics and the price of risk. The real world factor dynamics can then be determined subsequently using the change of measure technique described in Appendix C.

As usual, the parameters of the measurement error covariance Ξ can be concentrated out of the first (cross-section) component of (34) and, as noted above, the reduced form estimates $\hat{\mathbf{c}}$ and $\hat{\mathbf{C}}$ can be substituted for λ_0, $\mathbf{\Lambda}_1$ and β in the second (time-series) component:

$$f(\mathbf{rx}_t^o, \mathbf{q}_{\mathbf{x},t}|\mathbf{x}_{t-1}; \Theta) = g^*(\mathbf{rx}_t^o|\mathbf{q}_{\mathbf{x},t}; \beta, \mathbf{\Sigma}) \times h(\mathbf{q}_{\mathbf{x},t}|\mathbf{x}_{t-1}; \hat{\mathbf{c}}, \hat{\mathbf{C}}, \mathbf{\Sigma},), \tag{35}$$

Thus, we use numerical methods only to find the maximum likelihood estimates of $(\beta, \mathbf{\Sigma})$ using the conditional densities given by (35). Moreover, as in JSZ, we can use a transformation of the covariance matrix of the residuals from (18) to obtain a good starting value for $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \mathbf{B}_W^{-1} \widehat{\mathbf{\Omega}} \mathbf{B}_W'^{-1}. \tag{36}$$

The consistency of the cross-section parameter estimates in (17) depends crucially upon our assumption that the excess portfolio returns $\mathbf{q}_{\mathbf{x},t}$ are measured without error. This leaves open the possibility that the portfolios of log bond prices (or yields), $\mathbf{q}_{\mathbf{y},t}$, could be measured with error. Because we use them as regressors in (18), this would induce measurement error bias in the *RFRs* in our framework. However, this is much less of a problem than errors in $\mathbf{q}_{\mathbf{x},t}$ would be. These *RFRs* are only used in a forecasting guise and *OLS* regressions give consistent predictions even if the regressors are observed with error.[22] The *RFR* parameters are only used in a subsidiary exercise to estimate the prices of risk and hence the $\mathcal{P}-$dynamics, as explained in Appendix C. Moreover, any parameter bias in the *RFRs* should be tiny since it depends upon the ratio of the measurement errors to the prediction errors, which is negligibly small in models of the Treasury market.

4.3. Preliminary cross-section tests

Our empirical tests start by examining the cross-sectional behaviour of yields, log prices and returns using a conventional three factor model. At this stage we only consider cross-sectional restrictions on the risk neutral dynamics in an *ATSM* framework, without imposing restrictions on the physical dynamics of the factors. Recall that these restrictions mean that each model of the cross section of returns is linear in variables but subject to nonlinear restrictions across its coefficients. These models are therefore nested within an unrestricted *OLS* benchmark model. Table 1 shows the basic results from these tests. It reports the log-likelihood for each model, the number of parameters and the Bayesian Information Criterion value (*BIC*), which makes a correction for the number of parameters and sample size and is the basic model selection criterion used in this paper.[23,24] Our tables show the negative

[22] Johnston (1963), pages 290-291. He also notes on page 283 that decision makers are likely to use the same mis-measured observations as the econometrician, making it reasonable to suppose this is the relevant model, giving a plausible split of the returns into forecastable and unforecastable components.

[23] The weighting matrix used to calculate observable portfolios is the same for the all models: the three eigenvectors associated with the first three principal components of yields.

[24] Naturally, yields are a simple transformation of log prices, $y_{m,t} = (-1/m)p_{m,t}$ and, as such, it should not matter if the model is estimated in yields or prices. The difference in our exercise comes from applying the same weighting matrix $\mathbf{W}$ to find observable portfolios of yields and prices, respectively. We checked that if we adjust the entries of $\mathbf{W}$ by appropriate maturities, then the estimated models are identical.

of the BIC statistic, making it appropriate to select the model with the highest value of $(-)BIC = 2\ln\mathcal{L} - k\ln(TM)$, where, $\mathcal{L}$ is the value of the maximized log-likelihood function, k is the number of parameters and T and M are the numbers of observations in the time series and cross-sectional dimension, respectively.

[Insert Table 1 near here]

The first panel of the table shows that the yields strongly reject the $AR(1)$ restrictions, reflecting the results of Hamilton and Wu (2014).[25] However the second panel shows that the difference in the test statistics for log prices is smaller than in the yields model. The final panel of the table shows that the difference between the $AR(1)$ and OLS models is even smaller in the returns data, but that the latter is still preferred on the basis of the BIC selection criterion. Further improvements can be obtained by introducing moving average or long memory factor dynamics but we investigate these models in using the full $DTSM$ framework, reporting the results in Sections 4.6 and 4.7 below.

We subsequently re-run these tests using a two factor model. Figure 5 plots the standard deviations of the residuals of the two factor (Panel A) and three factor models (Panel B) estimated in yields, log prices and returns.[26] Again, it is evident that the restricted $AR(1)$ yield and log price models have difficulty in matching their unrestricted OLS counterparts. Although the residuals for the two factor $AR(1)$ yields and log prices are visibly higher than for the OLS, these differences being much higher than for the three factor tests, the difference remains tiny for the returns model. Moreover, these residuals are surprisingly small, of the order of $2-3$ basis points in the $1-10$ year maturities, hardly any higher than those for the three factor analogue.

[Insert Figure 5 near here]

[25] They test these restrictions within a $DTSM$ framework while our test is for the $ATSM$ cross-section model.

[26] The standard deviations for the yields are multiplied by maturity to make them comparable with those for log prices and returns. All quantities are multiplied by 100.

Despite the statistical preference for the *OLS* over the $AR(1)$ models, the response coefficients for these pairs of models are numerically close in each data set, as we would expect from previous findings reported in the literature. The standard errors are also small. However, the coefficients of the models fitted to log prices and returns data are visibly different, as Bams and Schotman (2003) found for their yield and forward-differenced yield models. As can be inferred from (3), if the $AR(1)$ models replicate the *OLS* coefficients accurately then the *OLS* coefficients of the log price in (22) and returns (9) models should be the same. This relation holds also for rotated factors:

$$\mathbf{p}_t = \mathbf{A}_{\mathbf{qp}} + \mathbf{B}'_{\mathbf{qp}}\mathbf{q}_{\mathbf{p},t} + \mathbf{w}_t, \tag{37}$$

and

$$\mathbf{rx}_t = \mathbf{A}_{\mathbf{qx}} + \mathbf{B}'_{\mathbf{qx}}\mathbf{q}_{\mathbf{x},t} + \mathbf{v}_t, \tag{38}$$

where $\mathbf{q}_{\mathbf{p},t} = \mathbf{W}'\mathbf{p}_t$, since the weighting matrix $\mathbf{W}$ is the same. In Table 2 we test the null hypothesis that $\mathbf{B}_{\mathbf{qp}} = \mathbf{B}_{\mathbf{qx}}$. Under the alternative hypothesis ($\mathbf{B}_{\mathbf{qp}} \neq \mathbf{B}_{\mathbf{qx}}$) the joint likelihood is the sum of the likelihoods of the two *OLS* models shown in Table 1 for the model for prices and returns ($39,068.08$). We compare this value with that of a joint model that uses the same slope parameters in both log price and returns data ($38,820.81$). The likelihood ratio test rejects the null hypothesis at any conventional significance level. As noted in Section 3.2, this could indicate dynamic misspecification. The table also shows that the *BIC* value for the model under the alternative hypothesis is better than that for the model under the null.

[Insert Table 2 near here]

It is also worth noting that the volatility of the residuals from the yield and log price regressions shown in Figure 5, which can be interpreted as pricing errors, is much higher than in the return regressions. Further analysis of these residuals (see Table 3) shows that this could largely be due to the first order autocorrelation in the levels data noted previously by other authors (Dai and Singleton (2000), Duffee (2011a), Adrian et al. (2013), Hamilton and Wu (2014)). The autocorrelation in pricing errors can be corrected by first differencing the data or fitting a time series model to the errors as proposed by Hamilton and Wu. Switching to a returns or forward difference framework as we propose surely offers a more natural way of dealing with this problem. Table 3 shows that the return pricing errors are not significantly autocorrelated, which is consistent with our model assumption. If this were literally true, this would mean that the bond pricing errors were close to a random walk.[27]

[**Insert Table 3 near here**]

4.4. Modelling the price of risk

Our *DTSM* amalgamates the models of the returns under the risk neutral and real world measures (respectively the cross-section (17) and *RFR* (18) models) and estimates them jointly subject to the restriction that the innovation variances ($\mathbf{\Sigma}$) are the same in both models (see Section 4.2). This section discusses the *RFRs* for the excess returns on the three noiseless portfolios $\mathbf{q}_{\mathbf{x},t}$ (i.e. those that have returns observed without error, (18)) and the variables driving the price of risk. The evidence on price of risk factors in the term structure literature is mixed. Cochrane and Piazzesi (2005, 2008) claim that there is only one pricing factor driving the risk premium that is not spanned by the first three principal components of the term structure, which is nonetheless a combination of different yields. Duffee (2011b) shows that the forth and fifth principal components help to forecast excess bond returns and similar conclusion is reached by Adrian et al. (2013). Based on this evidence, we initially included the first five principal components of yields in the $\mathbf{x}_t$ vector in (18). The work of Ludvigson and Ng (2009) and Joslin et al. (2014) also suggested the use of macro variables

[27] We are grateful to an anonymous referee for pointing this out.

like industrial production growth (IP) and expected inflation (EI).[28]

Our preliminary estimation results suggested that the second and third principal components of yields as well as expected inflation were insignificant in the $RFRs$ and we therefore drop these regressors from our final model. Table 4, panel A, reports the estimates of the RFR parameters for the remaining variables and their significance. There is no obvious pattern of signs and significance across regressors. Panel B of Table 4 reports the $F-$test of joint significance of each variable in the three regressions. It is interesting to note that although the pricing variables are mostly individually insignificant in particular regressions their overall effect is very significant.

[Insert Table 4 near here]

4.5. The risk neutral dynamics in the DTSM

We estimate eight models that maintain Assumptions 1-6, in which the risk neutral dynamics are restricted using the following specifications: $AR(1)$, $ARMA(1,1)$, $AR(2)$, $ARFIMA(1,d,0)$, $ARFIMA(1,d,1)$, $ARFIMA(2,d,0)$, $ARFIMA(2,d,1)$ and $ARFIMA(2,d,2)$. In these models we parametrize all three factors in the same way, e.g. for the first model all three factors have $AR(1)$ dynamics. Table 5 reports the estimates of the risk neutral parameters of these models. The asymptotic standard errors are reported in a small font.

[Insert Table 5 near here]

All models exhibit very high short memory persistence. The highest autoregressive root is typically higher than 0.99. The models with $AR(2)$ dynamics also display very high short memory persistence, with roots larger than 0.99, in the second factor ($ARFIMA(2,d,1)$ and $ARFIMA(2,d,2)$) or in all three factors ($AR(2)$ and $ARFIMA(2,d,0)$). The models with

[28]IP is the monthly logarithmic growth in the seasonally adjusted industrial production index and EI is a monthly series for the median expected price change over the next 12 months compiled by the, University of Michigan Survey Research Center. Both series are available from the website of Federal Reserve Bank of St. Louis.

long memory, except for $ARFIMA(1, d, 0)$, exhibit 'dual persistence' under the $\mathcal{Q}$ measure: long-lasting short and long memory dynamics. The values of the long memory parameter are different for different models. All long memory models, however, have at least one non-stationary (i.e. with $d > 0.5$) and one stationary ($d < 0.5$) factor. For the $ARFIMA(1, d, 0)$ model one factor has a very persistent long memory parameter, just below 1, but its $AR(1)$ coefficient is negative. The most persistent factor in other models with fractionally integrated component has long memory coefficient between 0.6 (in the $ARFIMA(2, d, 0)$ model) and 0.9 (in the $ARFIMA(1, d, 1)$ and $ARFIMA(2, d, 1)$ models). Generally, the second and third factors in long memory models also display a significant degree of long range persistence, between 0.23 and 0.55 (except the third factor in the $ARFIMA(2, d, 1)$ model for which the long memory estimate is close to zero). The factor dynamics under the $\mathcal{Q}$ measure are shown in Figure 6, which plots the moving average representation of the risk neutral factor dynamics in these models (the β coefficients in (7)). These effects are depicted in the form of an impulse response function, showing the response over time of the spot rate to a unit shock spanned by the yield curve.

[**Insert Figure 6 near here**]

4.6. The real world dynamics

Appendix C shows how the model of the risk premium (12) is used to obtain the real world short rate dynamics ((C-8) in this appendix) from the risk neutral short rate dynamics (7). Figure 7 shows the effect of shocks backed out from the three noiseless portfolios used as factors (the $\mathbf{\Phi}$'s in (C-8)) and is comparable with Figure 6 showing the impulse response functions under $\mathcal{Q}$. While the general pattern is similar to that of Figure 6, these responses tend to be smaller and their 'humps' more muted, making them look more exponential than the impulse response functions under $\mathcal{Q}$. Figure 8 shows the effect of the standardized macro and other unspanned factors (the $\mathbf{\Upsilon}$'s in (C-8)), which affect the risk premium and the evolution of the yield curve but not the current yield curve. As expected, an increase in industrial production increases the interest rate.

[**Insert Figure 7 near here**]

[**Insert Figure 8 near here**]

4.7. Relative model performance

4.7.1. Likelihood comparisons

The standard errors of the cross-sectional parameters shown in Table 5 are very small. There are two reasons for this. The first is that the likelihood of the cross section depends upon measurement errors ($\mathbf{v}_t$) that are, as Cochrane and Piazzesi say in their 2008 paper, 'tiny' empirically, so that restrictions that generate tiny perturbations in the yield curve estimates are often rejected. The second is the large sample of the panel data, which tends to bias classical statistics such as likelihood ratio statistics toward rejection of model simplifications (Hendry (1995), Canova (2007)). The *BIC* statistics adjust for this effect.

Table 6 reports the number of parameters, log-likelihood and *BIC* values for each model. The most striking fact that emerges from this comparison is the very poor explanation of the data provided by the $AR(1)$ model. Despite the fact that it saves 57 degrees of freedom, it is the only model that is not preferred to the *OLS* model on the *BIC*. Relaxing the tight $AR(1)$ dynamics results in a big increase in the likelihood function which in turn results in a better *BIC* statistic. For instance, adding the moving average component $MA(1)$ or allowing for $AR(2)$ dynamics increases the likelihood by 130.50 and 194.96, respectively. However, these short memory models are still inferior to the models with long memory dynamics. The long memory models perform much better than the rest in this comparison. According to *BIC*, the $ARFIMA(2, d, 0)$, $ARFIMA(2, d, 1)$ and $ARFIMA(2, d, 2)$ have very similar performance and are preferred to all the other models.

[Insert Table 6 near here]

4.7.2. OLS comparisons

Hamilton and Wu (2014) note that the rejection of the $AR(1)$ $DTSM$ against the OLS benchmark is to a large extent due to its tight restrictions on the intercept terms. If these are freely estimated the likelihood of the model improves significantly, though not enough to beat the benchmark. These restrictions stem from Assumption 1 and Jensen's inequality and are functions of the moving average coefficients of the factors under the risk neutral measure (10). Figure 1 plots the intercepts (17) estimated for the $AR(1)$, $AR(2)$, and $ARFIMA(2, d, 1)$ models against those of the unrestricted OLS benchmark and its 95% confidence interval. The estimated coefficients depend on the assumption about the weighting matrix $\mathbf{W}$ which selects the noiseless portfolios. For our particular choice of the first three eigenvectors of yields, the intercepts have a sinusoidal shape. Evidently, all of these models find it difficult to replicate the OLS intercepts. However, the $AR(2)$ and $ARFIMA(2, d, 1)$ model intercepts, display a visibly better fit than those of the $AR(1)$ ones, especially at the long end. The $AR(1)$ model produces intercepts that are too flat to fit the OLS model.

Figure 2 plots the loadings (slope coefficients) on the noiseless portfolios in these models. These loadings also have a sinusoidal shape. As noted by Cochrane and Piazzesi (2008), the very small errors give a very tight confidence intervals for the loadings. This allows for a powerful discrimination between particular models, as noted in the previous section. As is evident from the figures, despite the fact that the $AR(1)$ model yield loadings seem to be close to those of the OLS model, these coefficients lie outside the 95% confidence interval for most maturities. The $AR(2)$ and $ARFIMA(2, d, 1)$ models are able to replicate the OLS estimates much better, lying within the confidence interval for most maturities. Also, the $ARFIMA(2, d, 1)$ model fits the slope coefficients slightly better than the $AR(2)$ model, especially for the first and second portfolios, which is the most likely the reason for the improved performance in terms of the value of the likelihood function and BIC value.

4.7.3. Low frequency characteristics

It is well known that interest rates are highly persistent. Usually the highest autoregressive root estimated for a *DTSM* is very close to one (see e.g. Bauer, Rudebusch, and Wu (2014) and references therein). Other researchers attribute the high persistence in interest rates to fractionally integrated or long memory processes (see for example Gil-Alana (2004), Iacone (2009), Osterrieder (2013), Goliński and Zaffaroni (2016)).

The long range behaviour of yields is best described by the shape of the periodogram near frequency zero. In Figure 9 we plot the logarithm of the periodogram ordinates of the standardized first principal component of yields for the first 25 Fourier frequencies. The periodogram of a series x_t is defined as:

$$I(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t e^{\imath t\lambda}\right|, \tag{39}$$

where $\imath$ denotes a complex unit and the Fourier frequencies are $\lambda_j = 2\pi j/T$. In the neighborhood of frequency zero the periodogram displays a steep spike, a tell-tale characteristic of a long memory processes. This is consistent with one of the definitions of the long memory process, which defines it as having an unbounded spectral density at frequency zero (see e.g. Baillie (1996), Diebold and Inoue (2001)). On the other hand, a stationary *ARMA* process has a bounded spectral density at frequency zero.

[Insert Figure 9 near here]

Although we directly estimate the risk neutral dynamics rather than the physical dynamics, Goliński and Zaffaroni (2016), Theorem 4.4, show that in the affine setting, the largest order of (fractional) integration under the $\mathcal{P}$ and $\mathcal{Q}$ measures coincide, and this determines the spectral density of yields at the zero, i.e.:

$$s(\lambda) \sim c\lambda^{-2\underline{d}} \qquad \text{as} \quad \lambda \to 0^+, \tag{40}$$

where $\sim$ denotes asymptotic equivalence and $\underline{d} = max\,(d_1, \ldots, d_K)$.

Thus, to examine the ability of particular models to emulate the behavior of the periodogram of yields at ultra-low frequencies, Figure 9 plots the periodogram alongside the spectral densities of the autoregressive series:

$$s_{AR(1)} = \frac{(1-\phi^2)}{2\pi} \left|1 - \phi e^{\imath\lambda}\right|^{-2} \tag{41}$$

with $AR1$ coefficient $\phi = 0.996$, and the long memory processes:

$$s_{LM}\,(\lambda) = c\lambda^{-2d}, \tag{42}$$

where c is some generic constant. Based on our estimates of long memory parameters, we plot the spectral density for $d = 0.9$, which is close to the estimate of the largest long memory parameter for models $ARFIMA(1, d, 1)$ and $ARFIMA(2, d, 1)$, and $d = 0.7$, which is close to the estimate in the $ARFIMA(2, d, 2)$ model. The size of the autoregressive coefficient corresponds to the typical degree of persistence found in interest rates. It is close to the highest autoregressive root estimated under the $\mathcal{Q}$ measure for the $AR(1)$ model (0.9964) and slightly higher than the estimate of the autoregressive coefficient for the first principal component (0.9945). Figure 9 shows that despite the very high short memory persistence the autoregressive model is not able to match the periodogram ordinates at the lowest frequencies. The long memory models perform much better in this respect. The model with $d = 0.7$ matches the lowest frequencies, but it flattens out too quickly for the intermediate frequencies. It seems that the model with $d = 0.9$ generates the spectral density that most closely represents the long range characteristics of these data.

5. Conclusion

The term structure literature has almost invariably modelled the dynamics of the yield curve as the autoregressive process of order one despite mounting statistical evidence that this model is misspecified (Duffee (2011a), Hamilton and Wu (2014)). It may be tempting to think that the $AR(1)$ model is still acceptable since the measurement errors are small and the factor loadings are similar to those of an unrestricted OLS model. However the poor performance of the model in Bams and Schotman (2003) type tests would caution against that, revealing that estimating this model in levels and forward-differences has very different implications for the pricing parameters. The fact that the pricing errors exhibit strong serial correlation is also a concern. Hamilton and Wu circumvent this problem using an $AR(1)$ error model, while Adrian et al. (2013) show that formulating the model in terms of returns eliminates serial correlation.

We find similar evidence against the $AR(1)$ yield model in our data. Using a set of 18 maturities we strongly reject the $AR(1)$ model against the OLS benchmark, regardless of how it is formulated. Following Adrian et al. (2013), we also find that changing the framework from yields to returns eliminates the serial correlation in the pricing errors. An additional and novel result is that we find that this change greatly reduces the size of the pricing errors, motivating the use of returns rather than yields or log prices for both research and pricing. Although we adopted a conventional three factor approach, subsequent work on a two factor variant suggests that the factor structure of the returns data may be of lower order than for the yield data. We intend to investigate this further in future work. At a practical level, the observation of Cochrane and Piazzesi (2008) that the tiny pricing error variances seen in the Treasury bond market make the cross section very good at discriminating between different models holds *a fortiori* for the returns framework on account of its lower variance structure.

These empirical findings nicely complement our observation that the return framework is much more flexible than the affine log price or yield specification and opens the way to the estimation and testing of a wide range of affine models. We can easily analyse higher order autoregressive, moving average and long memory processes, without increasing the dimension of the model and the OLS test benchmark. It is particularly difficult to handle the long memory feature in the level framework since this class of processes does not have a finite state space representation. In our return framework, however, estimating long memory models is hardly more difficult than estimating other time series models. Our results suggest that allowing for richer risk neutral dynamics leads to a significant improvement in fit, generating models with coefficients that are constrained by asset pricing theory and yet are preferred to the unrestricted benchmark. Log-likelihood and other comparisons strongly support the use of $ARFIMA$ models, which include AR, LM and MA effects.

In sum, our results suggest that the returns framework has the twin advantages of accuracy and flexibility, important features that have been overlooked in the literature. This will surely increase its appeal, minimizing the risk of misspecification of the risk neutral dynamics. Besides the government bond market, this forward market based framework can contribute to better understanding of the mechanics of other financial markets, including corporate bond and credit derivative markets. Forward rates and prices are extensively used in other markets like foreign exchange and commodities, but their full potential for research and pricing in our view has yet to be exploited.

Appendix A Forward differences and the coefficients of the ATSM

Absent arbitrage and measurement error, the one-period-ahead risk neutral expectation of a security price like $P_{m,t}$ equals its one-period-ahead forward price, $F_{m,t} = e^{r_t} P_{m,t}$. The price revisions are affine in the Gaussian innovations in (2). Thus prices are lognormal and their expectations are:

$$E_{t-1}^{\mathcal{Q}}[P_{m,t}] = \exp\left(E_{t-1}^{\mathcal{Q}}[p_{m,t}] + \frac{1}{2} Var_{t-1}[p_{m,t}] \right), \tag{A-1}$$

where Var_t denotes the period t conditional variance. Setting this equal to $F_{m,t}$, taking logs and rearranging gives:

$$E_{t-1}^{\mathcal{Q}}[p_{m,t}] = \alpha_m + p_{m+1,t-1} + r_{t-1}, \tag{A-2}$$

where $f_{m,t} \equiv \ln(F_{m,t})$ and $\alpha_m = Var_{t-1}[p_{m,t}]/2$. From (3) we have $Var_{t-1}[p_{m,t}] = \mathbf{b}_m' \mathbf{\Sigma} \mathbf{b}_m$ and thus the convexity adjustment term is given by (5). The term $(\alpha_m + p_{m+1,t-1} + r_{t-1})$ represents a synthetic forward log price: the value at which a risk neutral agent would price a contract at $t-1$ paying $p_{m,t}$ at t. Subtracting this from the outturn $p_{m,t}$ gives the true (convexity-adjusted) excess return to this agent. Given (3) and (A-2), this depends upon the convexity-adjusted excess returns on the factor portfolios: $p_{m,t} - \alpha_m - p_{m+1,t-1} - r_{t-1} = \mathbf{b}_m' \mathbf{u}_t^{\mathcal{Q}}$. However, we follow the convention of defining the excess return gross of the convexity adjustment to get (4).

The remaining coefficients of this relation follow from the model (7) of the risk neutral spot rate dynamics. Advancing this to period $t+m$:

$$r_{t+m} = \mu_r + \sum_{i=0}^{\infty} \beta_i' \mathbf{u}_{t+m-i}^{\mathcal{Q}} \tag{A-3}$$

and taking conditional expectations gives:

$$\begin{aligned}
E_{t-1}^{\mathcal{Q}}[r_{t+m}] &= \{\mu_r + \beta'_{m+1}\mathbf{u}_{t-1}^{\mathcal{Q}} + \beta_{m+2}\mathbf{u}_{t-2}^{\mathcal{Q}} + \ldots\}, \\
E_{t}^{\mathcal{Q}}[r_{t+m}] &= \{\mu_r + \beta'_{m}\mathbf{u}_{t}^{\mathcal{Q}} + \beta'_{m+1}\mathbf{u}_{t-1}^{\mathcal{Q}} + \beta_{m+2}\mathbf{u}_{t-2}^{\mathcal{Q}} + \ldots\} \\
&= E_{t-1}^{\mathcal{Q}}[r_{t+m}] + \beta'_m\mathbf{u}_t^{\mathcal{Q}}. \qquad \text{(A-4)}
\end{aligned}$$

Given these expectations and revisions, the parameters of the $ATSM$ of returns follow as usual from the well-known implication of no-arbitrage for the price of a an m period zero-coupon bond:

$$P_{m,t} = E_t^{\mathcal{Q}} \exp\left[-\sum_{i=0}^{m-1} r_{t+i}\right]. \qquad \text{(A-5)}$$

Substituting this back into (A-1) and taking logs gives:

$$p_{m,t} = -r_t - E_t^{\mathcal{Q}}\left[\sum_{i=1}^{m-1} r_{t+i}\right] + \frac{1}{2}Var_t\left[\sum_{i=1}^{m-1} r_{t+i}\right], \qquad \text{(A-6)}$$

The $i.i.d.$ assumption means that the last component is independent of t. It can be shown that (A-3) implies:

$$Var\left[\sum_{i=1}^{m-1} r_{t+i}\right] = \sum_{i=1}^{m-1} \mathbf{b}_i'\mathbf{\Sigma}\mathbf{b}_i. \qquad \text{(A-7)}$$

Thus by the tower property of expectations:

$$E_{t-1}^{\mathcal{Q}}[p_{m,t}] = -E_{t-1}^{\mathcal{Q}}\left[\sum_{i=0}^{m-1} r_{t+i}\right] + \frac{1}{2}Var\left[\sum_{i=1}^{m-1} r_{t+i}\right]. \qquad \text{(A-8)}$$

Subtracting this from (A-6) and substituting (A-3) gives:

$$\begin{aligned}
p_{m,t} - E_{t-1}^{\mathcal{Q}}[p_{m,t}] &= -\sum_{i=0}^{m-1}(E_t^{\mathcal{Q}}[r_{t+i}] - E_{t-1}^{\mathcal{Q}}[r_{t+i}]) \\
&= -\left(\sum_{i=0}^{m-1}\beta_i'\right)\mathbf{u}_t^{\mathcal{Q}} \\
&= \mathbf{b}_m'\mathbf{u}_t^{\mathcal{Q}}. \qquad \text{(A-9)}
\end{aligned}$$

Finally, equating this with (3) gives the expression (8) for the loadings on the innovations in the returns ATSM.

Appendix B The change of measure

The change of measure from $\mathcal{Q}$ to $\mathcal{P}$ depends on the Radon-Nikodym derivative or $\frac{d\mathcal{P}}{d\mathcal{Q}}(\mathbf{x}_t|\mathbf{x}_{t-1})$ which transforms the conditional densities multiplicatively:[29]

$$f^{\mathcal{P}}(\mathbf{x}_t|\mathbf{x}_{t-1}) = \frac{d\mathcal{P}}{d\mathcal{Q}}(\mathbf{x}_t|\mathbf{x}_{t-1}) \times f^{\mathcal{Q}}(\mathbf{x}_t|\mathbf{x}_{t-1}), \tag{B-1}$$

where $f^{\mathcal{Q}}$ and $f^{\mathcal{P}}$ are densities under $\mathcal{Q}$ and $\mathcal{P}$ respectively conditional upon the information set $\mathbf{x}_{t-1}$. We assume that this multiplier (and hence the distribution of $\mathbf{u}_t$ under $\mathcal{P}$) is exponential affine:

$$\frac{d\mathcal{P}}{d\mathcal{Q}}(\mathbf{x}_t|\mathbf{x}_{t-1}) = \exp\left(\frac{1}{2}\lambda'_{t-1}\lambda_{t-1} - \lambda'_{t-1}\mathbf{\Sigma}'^{-1}\mathbf{u}_t\right), \tag{B-2}$$

where λ_{t-1} is the price of risk vector.[30] Assuming Gaussianity of the risk neutral innovations as in (2), the conditional density of the factor innovations under the risk neutral measure $\mathcal{Q}$ is:

$$f^{\mathcal{Q}}(\mathbf{u}_t^{\mathcal{Q}}) = (2\pi)^{-\frac{3}{2}}\left|\mathbf{\Sigma}\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\mathbf{u}_t^{\mathcal{Q}\prime}\mathbf{\Sigma}^{-1}\mathbf{u}_t^{\mathcal{Q}}\right). \tag{B-3}$$

[29] Although there are two kinds of innovations in our model, shocks to the factors and measurement errors, these are independent and serially uncorrelated. Consequently the latter are not priced and do not affect this transform.

[30] In other words, the Stochastic Discount Function (SDF) $\frac{dP}{dQ}(\mathbf{x}_{t+1}|\mathbf{x}_t)e^{-r_t}$ is specified as $\exp[-r_t - \frac{1}{2}\lambda'_t\lambda_t - \lambda'_t\mathbf{P}'^{-1}\mathbf{u}_{t+1}]$.

Substituting these two relations into (B-1) and consolidating the exponents gives the density of the factor innovations under $\mathcal{P}$ conditional upon the information set $\mathbf{x}_{t-1}$:

$$
\begin{aligned}
f^{\mathcal{P}}(\mathbf{u}_t) &= (2\pi)^{-\frac{3}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\mathbf{u}_t - \mathbf{\Sigma}'\lambda_{t-1}\right)' \mathbf{\Sigma}^{-1} \left(\mathbf{u}_t - \mathbf{\Sigma}'\lambda_{t-1}\right) - \lambda_{t-1}'\mathbf{\Sigma}'^{-1}\mathbf{u}_t + \frac{1}{2}\lambda_{t-1}'\lambda_{t-1}\right] \\
&= (2\pi)^{-\frac{3}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{u}_t'\mathbf{\Sigma}^{-1}\mathbf{u}_t\right), \qquad \text{(B-4)}
\end{aligned}
$$

which satisfies $E_{t-1}[\mathbf{u}_t] = \mathbf{0}$ as reported in (13).

Appendix C Derivation of the dynamics under the $\mathcal{P}$ measure

This appendix shows how the model of the risk premium (11) is used to obtain the model of the real world short rate dynamics ((C-8), below) from the model of the risk neutral short rate dynamics (7).

We focus on the relation between the moving average coefficients (or impulse responses) under the two measures. Thus we ignore intercept terms and assume that λ_0 is zero. We use the method of undetermined coefficients to derive the relation between the moving average coefficients of the spot interest rate under the two measures, assuming that the system is initially in equilibrium.

The risk neutral dynamics of the vector $\mathbf{z}_t$ follow from (6):

$$
\mathbf{z}_t = \sum_{j=0}^{t-1} \mathcal{B}_j' \mathbf{u}_{t-j}^{\mathcal{Q}}, \qquad \text{(C-1)}
$$

where $\mathcal{B}_j$ is a diagonal matrix with the diagonal equal to β_j. From (A-4) and (A-6) in

Appendix A we find the price process:

$$p_{m,t} = \mu_{p,m} + \sum_{j=0}^{\infty} \left(\mathbf{b}'_{m+j} - \mathbf{b}'_j \right) \mathbf{u}^{\mathcal{Q}}_{t-j}, \tag{C-2}$$

where

$$\mu_{p,m} = -m\mu_r + \frac{1}{2} \sum\nolimits_{j=1}^{m-1} \mathbf{b}'_j \boldsymbol{\Sigma} \mathbf{b}_j, \tag{C-3}$$

$\mathbf{b}_j$ is defined in (8) and $\mathbf{b}_0 = \mathbf{0}$. The yields are then:

$$y_{m,t} = -\frac{1}{m}\mu_{p,m} + \frac{1}{m} \sum_{j=0}^{\infty} \left(\mathbf{b}'_j - \mathbf{b}'_{m+j} \right) \mathbf{u}^{\mathcal{Q}}_{t-j}. \tag{C-4}$$

It is convenient to stack these relations in a vector $\mathbf{y}_t$. It follows that the vector that spans the term structure of yields, $\mathbf{q}_{\mathbf{y},t} = \mathbf{W}'\mathbf{y}_t$, has the moving average representation:

$$\mathbf{q}_{\mathbf{y}_t} = \mu_{\mathbf{qy}} + \sum_{j=0}^{\infty} \boldsymbol{\Gamma}'_j \mathbf{u}^{\mathcal{Q}}_{t-j}, \tag{C-5}$$

where the row entries of $\mu_{\mathbf{qy}}$ and $\boldsymbol{\Gamma}'_j$ are determined by the product of $\mathbf{W}'$ and the corresponding terms in (C-4).

Joslin et al. (2014) note that the dynamics under $\mathcal{P}$ can be richer than under $\mathcal{Q}$ in the sense that they can be driven by the factors that span the term structure $(\mathbf{z}_t)$ as well as macroeconomic and other unspanned price of risk factors $(\mathbf{m}_t)$. Suppose that there are N unspanned factors modelled by an *AR OLS* regression system for these additional variables:[31]

$$\mathbf{m}_t = \boldsymbol{\Theta}'_q \mathbf{q}_{\mathbf{y},t-1} + \boldsymbol{\Theta}'_m \mathbf{m}_{t-1} + \eta_t \qquad t \geq 1, \tag{C-6}$$

where we use mean-adjusted data and η_t is an error vector.

Similarly, we assume that the price of risk is driven by both the observable factors driving the term structure of bond prices, $\mathbf{q}_{\mathbf{y},t}$, and other unspanned variables, $\mathbf{m}_t$. Splitting Λ_1 into

[31] This is not reported but available upon request.

parts corresponding to term structure factors and unspanned macro factors, the price of risk can be written following (12) as:

$$\mathbf{\Sigma}'\lambda_t = \mathbf{\Lambda}'_{\mathbf{q}}\mathbf{q}_{\mathbf{y},t} + \mathbf{\Lambda}'_{\mathbf{m}}\mathbf{m}_t. \tag{C-7}$$

Together with (C-6), this implies that the dynamics of the factors under the $\mathcal{P}$ measure are of the form:

$$\mathbf{z}_t = \sum_{j=0}^{t-1} \mathbf{\Phi}'_j \mathbf{u}_{t-j} + \sum_{j=0}^{t-1} \mathbf{\Upsilon}'_j \eta_{t-j}, \tag{C-8}$$

where the risk neutral and real world innovations are related by

$$\begin{aligned} \mathbf{u}_t^{\mathcal{Q}} &= \mathbf{u}_t + \mathbf{\Sigma}'\lambda_{t-1} \\ &= \mathbf{u}_t + \mathbf{\Lambda}'_{\mathbf{q}}\mathbf{q}_{\mathbf{y},t-1} + \mathbf{\Lambda}'_{\mathbf{m}}\mathbf{m}_{t-1}. \end{aligned} \tag{C-9}$$

Substituting recursively for $\mathbf{u}_t^{\mathcal{Q}}$ in (C-1) equations (C-5), (C-6), (C-7) and (C-9) we can express the dynamics of $\mathbf{z}_t$ by innovations $\mathbf{u}_t$ and η_t. Thus, matching the coefficients at corresponding terms we find the $\mathcal{P}$ coefficients in (C-8) as

$$\begin{aligned} &\mathbf{\Phi}'_0 = \mathcal{B}'_0, \\ &\mathbf{\Phi}'_j = \mathcal{B}'_j + \sum_{i=0}^{j-1} \left(\mathbf{\Phi}'_i \mathbf{\Lambda}'_q + \mathbf{\Upsilon}'_i \mathbf{\Theta}'_q\right) \mathbf{\Gamma}_{j-1-i}, \quad \text{for} \quad j \geq 1 \end{aligned} \tag{C-10}$$

and

$$\begin{aligned} &\mathbf{\Upsilon}'_0 = \mathbf{0}, \\ &\mathbf{\Upsilon}'_j = \mathbf{\Phi}'_{j-1}\mathbf{\Lambda}'_m + \mathbf{\Upsilon}'_{j-1}\mathbf{\Theta}'_m, \quad \text{for} \quad j \geq 1. \end{aligned} \tag{C-11}$$

We can find the impulse response of the short rate to the $k-$th latent factor $x_{k,t}$ at lag j as $\mathbf{e}'_k\mathbf{\Phi}'_j$, where $\mathbf{e}_k$ is a vector of zeros with 1 at the $k-$th position. Similarly the impulse response of the short rate to the $k-$th unspanned variable is given by $\mathbf{e}'_k\mathbf{\Upsilon}'_j$.

References

Adrian, T., Crump, R. K., Moench, E., 2013. Pricing the term structure with linear regressions. Journal of Financial Economics 110, 110–138.

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 745–787.

Backus, D. K., Zin, S. E., 1993. Long-memory inflation uncertainty: Evidence from the term structure of interest rates. Journal of Money Credit and Banking 25, 681–700.

Bai, J., 1997. Estimation of a change point in multiple regression models. Review of Economics and Statistics 79, 551–563.

Baillie, R. T., 1996. Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5–59.

Balduzzi, P., Chiang, I.-H. E., 2012. A simple test of the affine class of term structure models. Review of Asset Pricing Studies 2, 203–244.

Bams, D., Schotman, P. C., 2003. Direct estimation of the risk neutral factor dynamics of Gaussian term structure models. Journal of Econometrics 117, 179–206.

Bauer, M. D., forthcoming. Restrictions on risk prices in dynamic term structure models. Journal of Business & Economic Statistics .

Bauer, M. D., Rudebusch, G. D., 2015. Resolving the spanning puzzle in macro-finance term structure models. CESifo Working Paper Series 5187, CESifo Group Munich.

Bauer, M. D., Rudebusch, G. D., Wu, J. C., 2014. Term premia and inflation uncertainty: Empirical evidence from an international panel dataset: Comment. American Economic Review 104, 323–337.

Canova, F., 2007. Methods for Applied Macroeconomic Research. Princeton University Press, Princeton NJ.

Cochrane, J. H., Piazzesi, M., 2005. Bond risk premia. American Economic Review 95, 138–160.

Cochrane, J. H., Piazzesi, M., 2008. Decomposing the yield curve. Unpublished working paper .

Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 415–41.

Diebold, F. X., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105, 131–159.

Duffee, G. R., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance 62, 405–443.

Duffee, G. R., 2010. Sharpe ratios in term structure models. Unpublished working paper, Johns Hopkins University .

Duffee, G. R., 2011a. Forecasting with the term structure: The role of no-arbitrage restrictions. Unpublished working paper, Johns Hopkins University .

Duffee, G. R., 2011b. Information in (and not in) the term structure. Review of Financial Studies 24, 2895–2934.

Fenou, B., Meddahi, N., 2009. Generalised affine models. Unpublished working paper .

Garcia, R., Perron, P., 1996. An analysis of the real interest rate under regime shifts. Review of Economics and Statistics 78, 111–125.

Gil-Alana, L. A., 2004. Modelling the U.S. interest rate in terms of I(d) statistical models. Quarterly Review of Economics and Finance 44, 475–486.

Goliński, A., Zaffaroni, P., 2016. Long memory affine term structure models. Journal of Econometrics 191, 33–56.

Granger, C. W. J., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis 1, 15–29.

Gurkaynak, R. S., Sack, B., Wright, J. H., 2007. The U.S. Treasury yield curve: 1961 to the present. Journal of Monetary Economics 54, 2291–2304.

Hamilton, J. D., Wu, J. C., 2012. Identification and estimation of Gaussian affine term structure models. Journal of Econometrics 168, 315–331.

Hamilton, J. D., Wu, J. C., 2014. Testable implications of affine term structur models. Journal of Econometrics 178, 231–242.

Hendry, D. F., 1995. Dynamic Econometrics. Oxford University Press.

Hosking, J. R. M., 1981. Fractional differencing. Biometrica 68, 165–176.

Iacone, F., 2009. A semiparametric analysis of the term structure of the U.S. interest rates. Oxford Bulletin of Economics and Statistics 71, 475–490.

Johnston, J., 1963. Econometric Methods. McGraw-Hill, Kogakusha.

Joslin, S., Priebsch, M., Singleton, K. J., 2014. Risk premiums in dynamic term structure models with unspanned macro risks. Journal of Finance 69, 1197–1233.

Joslin, S., Singleton, K. J., Zhu, H., 2011. A new perspective on Gaussian dynamic term structure models. Review of Financial Studies 24, 926–970.

Ludvigson, S. C., Ng, S., 2009. Macro factors in bond risk premia. Review of Financial Studies 22, 5027–5067.

Osterrieder, D., 2013. Interest rates with long memory: A generalized affine term structure model. CREATES Research Papers, Aarhus University .

Singleton, K. J., 2006. Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment. Princeton University Press.

Zellner, A., 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association 57, 348–368.

Tables

Model	$\log \mathcal{L}$	k	(-)BIC
	Panel A		
Yields:	$\mathbf{y}_t = \mathbf{A_{qy}} + \mathbf{B'_{qy}}\mathbf{qy}$		
OLS	$40,780.54$	60	$81,167.67$
AR(1)	$40,533.20$	10	$80,978.97$
	Panel B		
Log prices:	$\mathbf{p}_t = \mathbf{A_{qp}} + \mathbf{B'_{qp}}\mathbf{q}_t$		
OLS	$18,103.61$	60	$35,813.81$
AR(1)	$17,823.89$	10	$35,560.36$
	Panel C		
Returns:	$\mathbf{xr}_t = \mathbf{A_{qx}} + \mathbf{B'_{qx}}\mathbf{qx}_t$		
OLS	$20,964.47$	60	$41,404.57$
AR(1)	$20,703.37$	9	$41,328.10$

Table 1: Preliminary cross section tests of the standard *AR*(1) *ATSM*. The table presents tests of this model formulated in yields (Panel A), log prices (Panel B) and excess returns (Panel C) against the corresponding *OLS* model. For each model the table presents the value of the log-likelihood function ($\log \mathcal{L}$), the number of parameters (k), and the (negative of the) Bayesian Information Criterion (*BIC*). The equation standard errors at different maturities are depicted in Figure 5(b). The sample period is January 1983 to December 2011.

	$\log \mathcal{L}$	LR	$p - value$ $(df = 18 \times 3)$	$(-)BIC$
H_0	$38,820.81$	494.55	0	$76,855.05$
H_1	$39,068.08$			$76,877.65$

Table 2: Test of equality of the slope parameters in equations (37) as estimated by *OLS*. The table reports the log-likelihood ($\log \mathcal{L}$), the Likelihood Ratio test (LR), its $p-$value and the (negative) Bayesian Information Criterion (*BIC*). The sample period is January 1983 to December 2011.

	Maturity (in years):							
Model	1/6	1/3	1/2	1	3	5	10	15
Log prices								
$\rho(1)$	0.7177	0.7224	0.4563	0.7795	0.7820	0.8457	0.8460	0.8581
$\rho(6)$	0.3736	0.4898	0.2529	0.4725	0.3783	0.4514	0.5027	0.5262
$\rho(12)$	0.2630	0.2613	0.2952	0.2330	0.3282	0.2575	0.4464	0.4629
Returns								
$\rho(1)$	0.5389	0.0340	0.0403	0.1414	−0.0809	0.0878	−0.0544	−0.0244
$\rho(6)$	0.2554	0.1061	0.1404	0.1542	−0.1093	0.0494	−0.1670	−0.1638
$\rho(12)$	0.4677	0.0682	0.1447	0.1159	0.0691	0.1138	0.0945	0.1190

Table 3: Autocorrelation of residuals of the $AR(1)$ $ATSM$ for yields, log prices and returns for selected maturities. The 95 percent confidence interval is $(-0.1072, 0.1072)$. The sample period is January 1983 to December 2011.

	c	$PCyield_1$	$PCyield_4$	$PCyield_5$	IP_t	R^2
Regressand			Panel A			
$\widetilde{xr}_1$	0.0428 0.2503	0.4077 0.5091	30.1896 27.62	−66.067 49.70	6.4091 8.240	0.0122
$\widetilde{xr}_2$	0.0768 0.2018	−0.157 0.4104	−33.614 22.27	30.7493 40.06	−8.262 6.643	0.0129
$\widetilde{xr}_3$	−0.1863 0.0735	0.0103 0.1495	25.5201 8.112	13.9002 14.60	5.1634 2.420	0.0432
Joint significance test			Panel B			
$F(3,340)$ $p-value$	20.3900 0.000	30.0011 0.000	17.2739 0.000	6.5194 0.001	7.1040 0.001	

Table 4: Monthly excess return forecasting regressions (18). These regress the noiseless portfolio returns on likely risk indicators and determine the price of risk. Standard errors of the parameters are reported in small font. In Panel B the joint significant test is reported with the corresponding p-value. The sample period is January 1983 to December 2011.

	ϕ_1	ϕ_2	d	θ_1	θ_2	Cholesky factor of Ω f_1	f_2	f_3
				Panel $A : AR(1)$				
factor 1	0.9964					0.9291		
	0.0002					0.0003		
factor 2	0.9804					−0.7389	0.0974	
	0.0005					0.0002	0.0000	
factor 3	0.9305					0.2315	−0.1286	0.0278
	0.0017					0.0001	0.0000	0.0000
				Panel $B : ARMA(1,1)$				
factor 1	0.9965			−0.0312		0.9166		
	0.0000			0.0519		0.0003		
factor 2	0.9682			−0.3431		−0.7334	0.0976	
	0.0000			0.0000		0.0002	0.0000	
factor 3	0.9661			−0.3550		0.2345	−0.1290	0.0278
	0.0000			0.0000		0.0001	0.0000	0.0000
				Panel $C : AR(2)$				
factor 1	1.9915	−0.9916				0.9510		
	0.0000	0.0000				0.0003		
factor 2	0.4778	0.5219				−0.7696	0.1018	
	0.0002	0.0002				0.0002	0.0000	
factor 3	1.8999	−0.9004				0.2549	−0.1353	0.0279
	0.0000	0.0000				0.0001	0.0001	0.0000
				Panel $D : ARFIMA(1,d,0)$				
factor 1	−0.5265		0.9908			0.9508		
	0.0439		0.0064			0.0003		
factor 2	0.9679		0.5431			−0.7707	0.1022	
	0.0006		0.0079			0.0002	0.0000	
factor 3	0.9904		0.3855			0.2568	−0.1357	0.0279
	0.0002		0.0059			0.0001	0.0000	0.0000
				Panel $E : ARFIMA(1,d,1)$				
factor 1	0.9101		0.8829	−0.8658		0.9501		
	0.0010		0.0019	0.0013		0.0003		
factor 2	0.9915		0.3192	0.9743		−0.7727	0.1019	
	0.0000		0.0002	0.0073		0.0002	0.0000	
factor 3	0.9922		0.3006	0.5875		0.2598	−0.1350	0.0278
	0.0000		0.0002	0.0042		0.0001	0.0000	0.0000
				Panel $F : ARFIMA(2,d,0)$				
factor 1	1.9752	−0.9756	0.5923			0.9429		
	0.0000	0.0000	0.0070			0.0003		
factor 2	0.9818	0.0103	0.3269			−0.7663	0.1021	
	0.0001	0.0002	0.0023			0.0003	0.0000	
factor 3	0.5861	0.4068	0.2317			0.2573	−0.1354	0.0278
	0.0001	0.0001	0.0083			0.0001	0.0000	0.0000
				Panel $G : ARFIMA(2,d,1)$				
factor 1	1.9765	−0.9771	0.8977	−0.8984		0.9477		
	0.0000	0.0000	0.0243	0.0034		0.0003		
factor 2	0.2785	0.6546	0.5133	−0.0511		−0.7709	0.1021	
	0.0003	0.0003	0.0122	0.0494		0.0003	0.0000	
factor 3	1.8714	−0.8720	0.0443	−0.6406		0.2597	−0.1352	0.0278
	0.0000	0.0000	0.0051	0.0102		0.0001	0.0000	0.0000
				Panel $H : ARFIMA(2,d,2)$				
factor 1	1.9823	−0.9828	0.6921	0.0176	−0.8452	0.9502		
	0.0000	0.0000	0.0093	0.0190	0.0000	0.0003		
factor 2	0.9670	0.0003	0.4840	0.1965	0.0932	−0.7738	0.1021	
	0.0001	0.0000	0.0006	0.0019	0.0000	0.0002	0.0000	
factor 3	1.3082	−0.3296	0.4818	−0.0601	−0.0475	0.2616	−0.1352	0.0278
	0.0000	0.0000	0.0006	0.0000	0.0000	0.0001	0.0000	0.0000

Table 5: The risk neutral dynamic parameters of the *DTSM*s. The standard errors are reported in small font and the likelihood statistics are shown in Table 6. The sample period is January 1983 to January 2011.

Model	k	$\log \mathcal{L}$	$(-)BIC$
OLS	81	$21,526.09$	$42,344.26$
$AR(1)$	24	$21,165.92$	$42,122.09$
$ARMA(1,1)$	27	$21,296.42$	$42,356.87$
$AR(2)$	27	$21,360.88$	$42,485.78$
$ARFIMA(1,d,0)$	27	$21,370.67$	$42,505.38$
$ARFIMA(1,d,1)$	30	$21,379.96$	$42,497.73$
$ARFIMA(2,d,0)$	30	$21,398.08$	$42,533.97$
$ARFIMA(2,d,1)$	33	$21.410.71$	$42,533.01$
$ARFIMA(2,d,2)$	36	$21,424.04$	$42,533.45$

Table 6: Likelihood statistics for different *ARFIMA* specification of the *DTSM* and their unrestricted *OLS* benchmark. The columns show respectively: the number of parameters, the loglikelihood and the (negative) Bayesian Information Criterion.

Figures

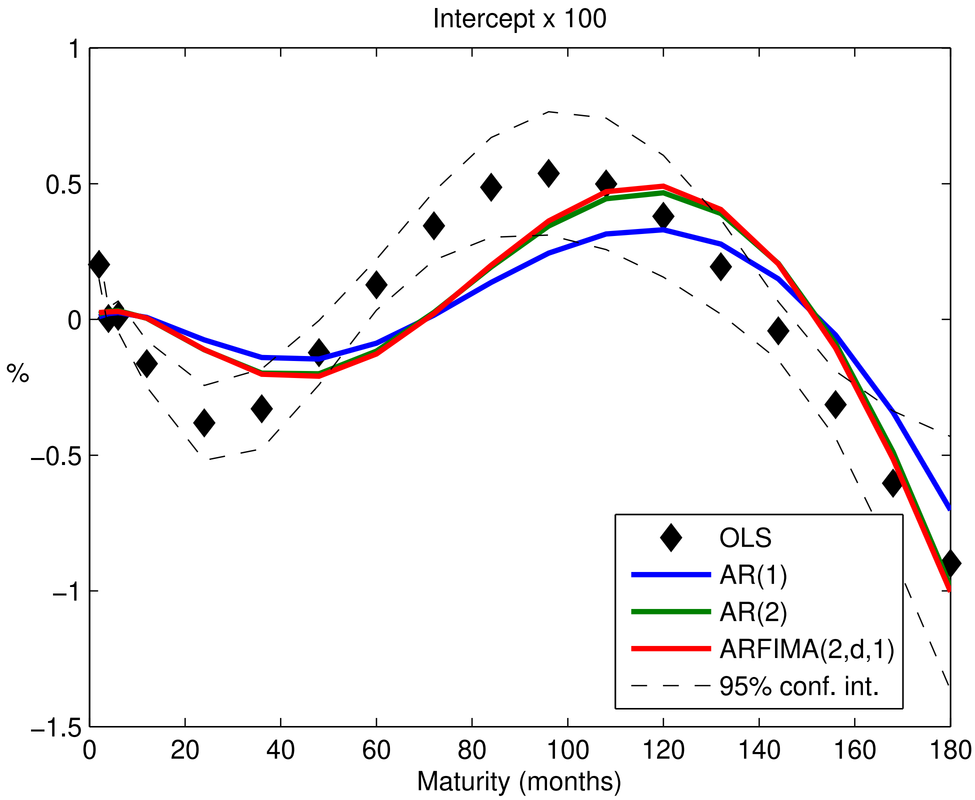
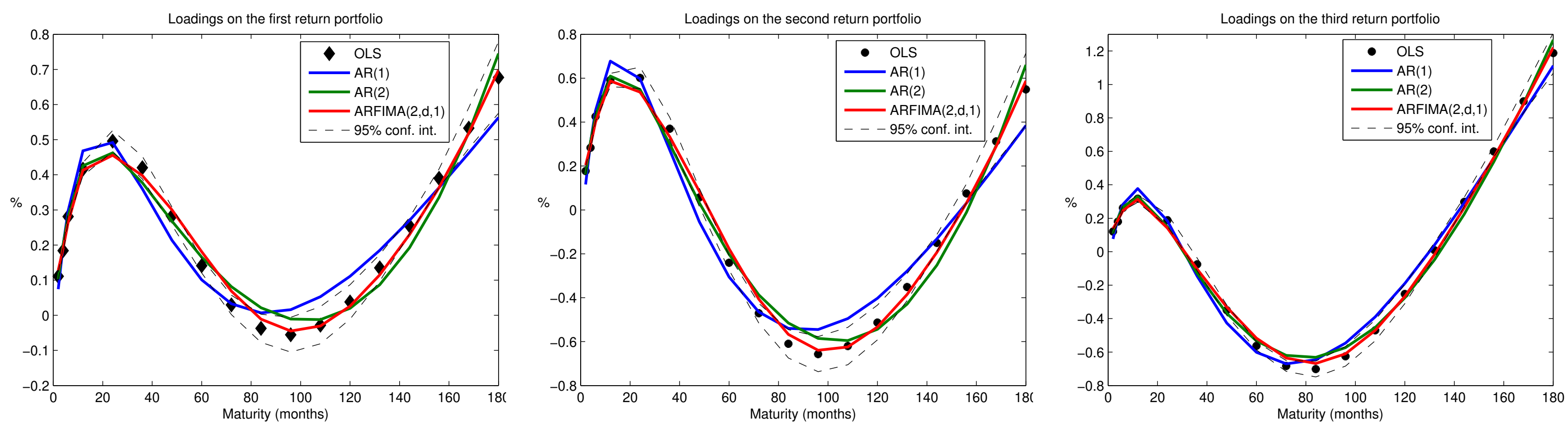


Fig. 1. Intercept coefficients by maturity of the $AR(1)$, $AR(2)$ and $ARFIMA(2, d, 1)$ excess return $DTSM$ models and the unrestricted OLS benchmark. The dashed lines represent the 95 percent OLS confidence interval.



(a) First return portfolio (b) Second return portfolio (c) Third return portfolio

Fig. 2. Slope estimates of the excess returns $DTSM$ for the three returns portfolios as estimated by the $AR(1)$, $AR(2)$ and $ARFIMA(2, d, 1)$ models and their unrestricted OLS benchmark. The dashed lines show the 95 percent confidence intervals for the OLS model.

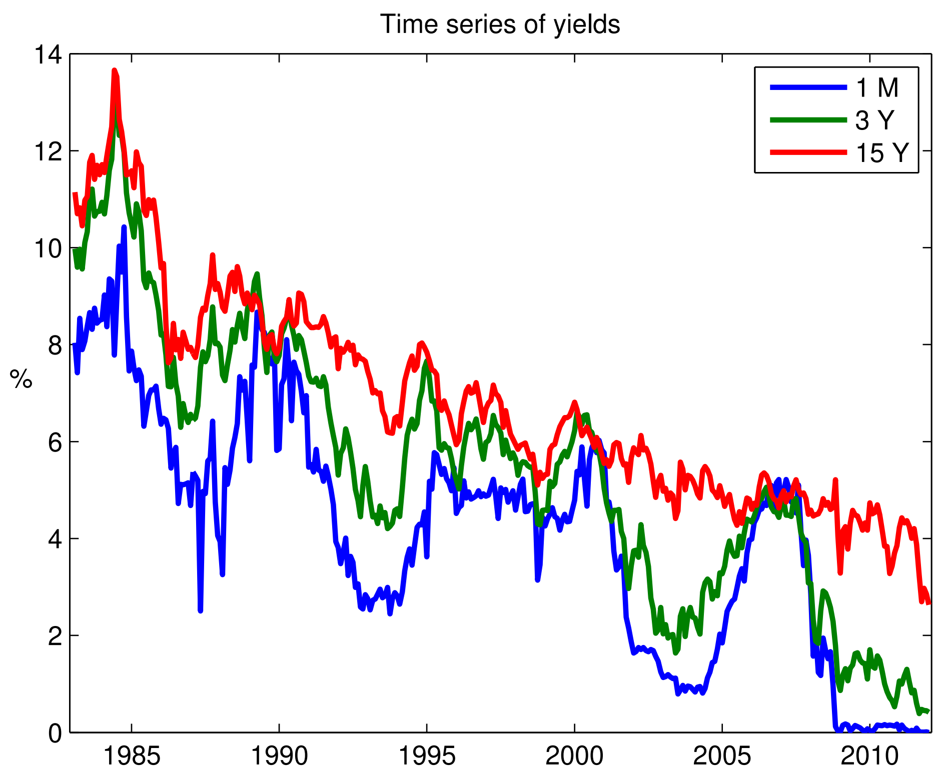


Fig. 3. Time series of 1 month, 3 year and 15 year zero coupon yields. The sample period is January 1983 to December 2011.

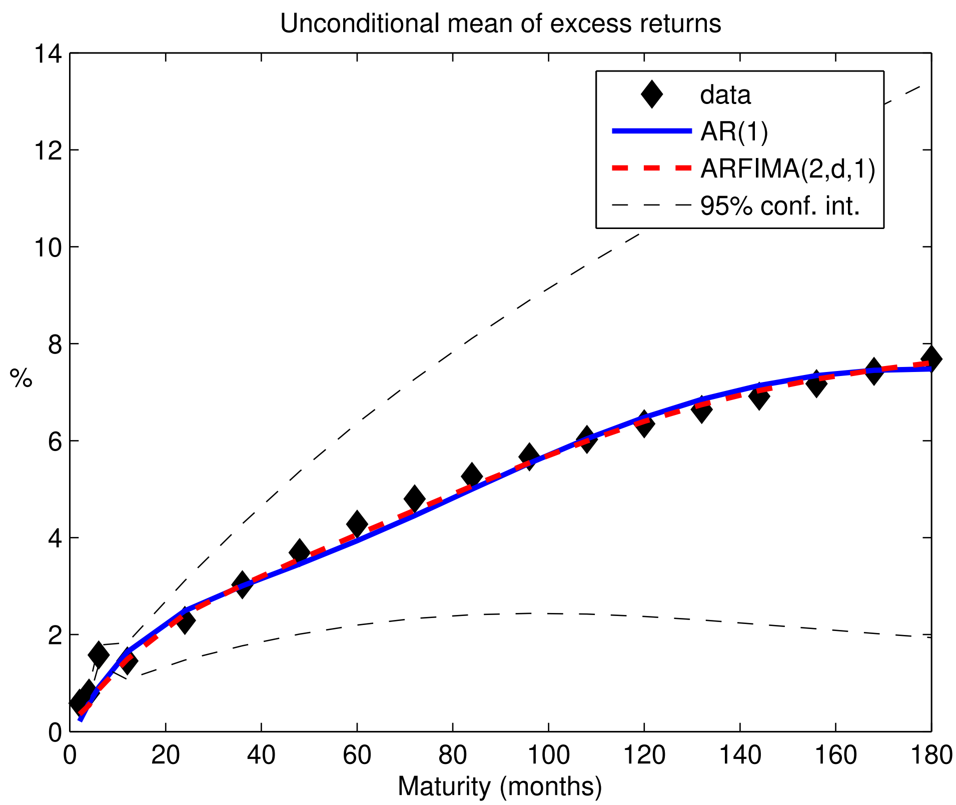
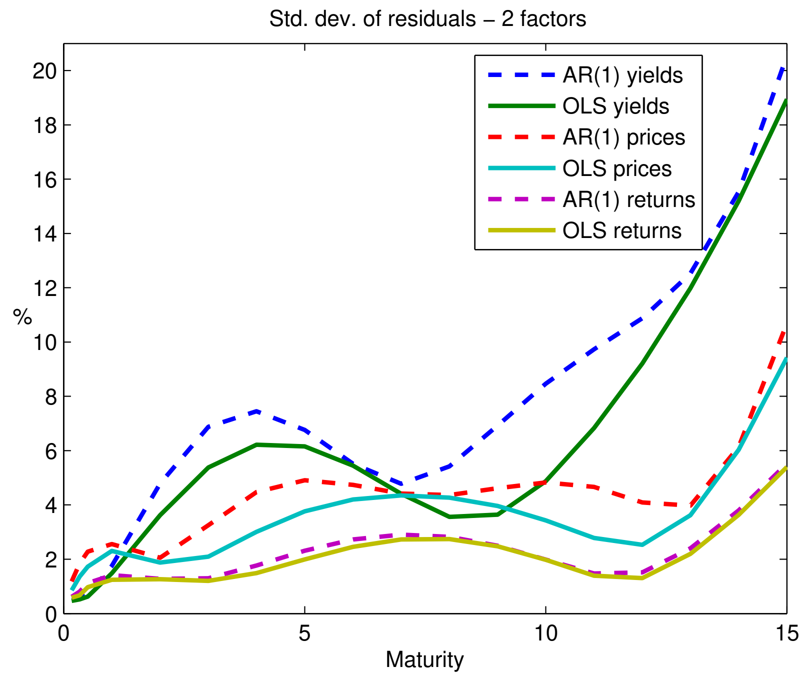


Fig. 4. Unconditional mean of excess returns, observed and implied by the $AR(1)$ and $ARFIMA(2, d, 1)$ models. The dashed lines represent the 95 percent OLS confidence interval.



(a)

(b)

Fig. 5. Standard errors of measurement errors for the $AR(1)$ $ATSM$s formulated in yields, log prices and excess returns, and their unrestricted OLS counterparts for a two factor model (a) and three factor model (b). The standard errors for returns and log prices are directly comparable. The standard errors from the yields model were multiplied by their yield maturities to make them comparable. All standard deviations are then multiplied by 100.

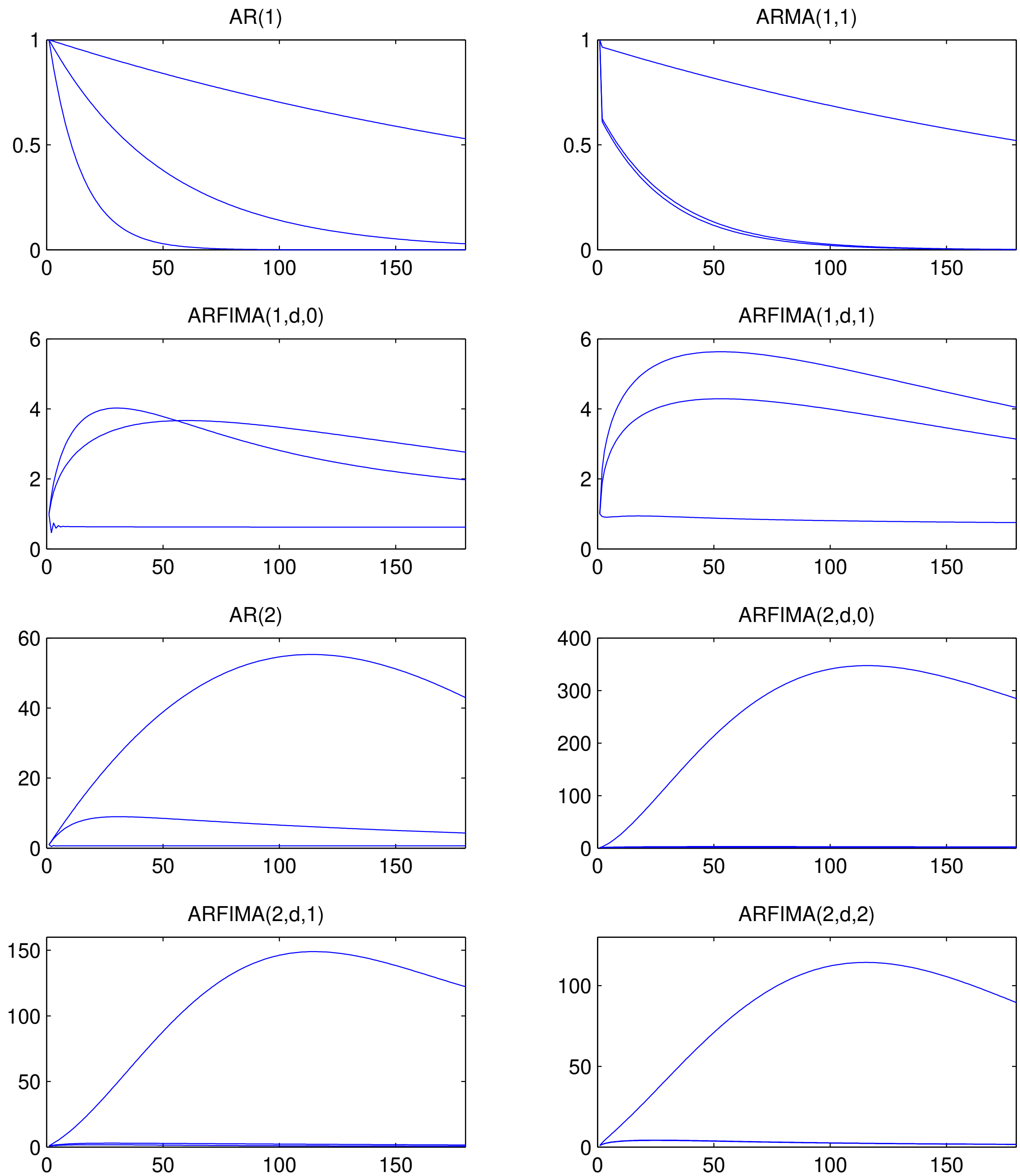


Fig. 6. Moving average representations of the factor dynamics under the risk neutral measure. These show the response over time (measured in months) of the spot rate to innovations in the three spanned factors.

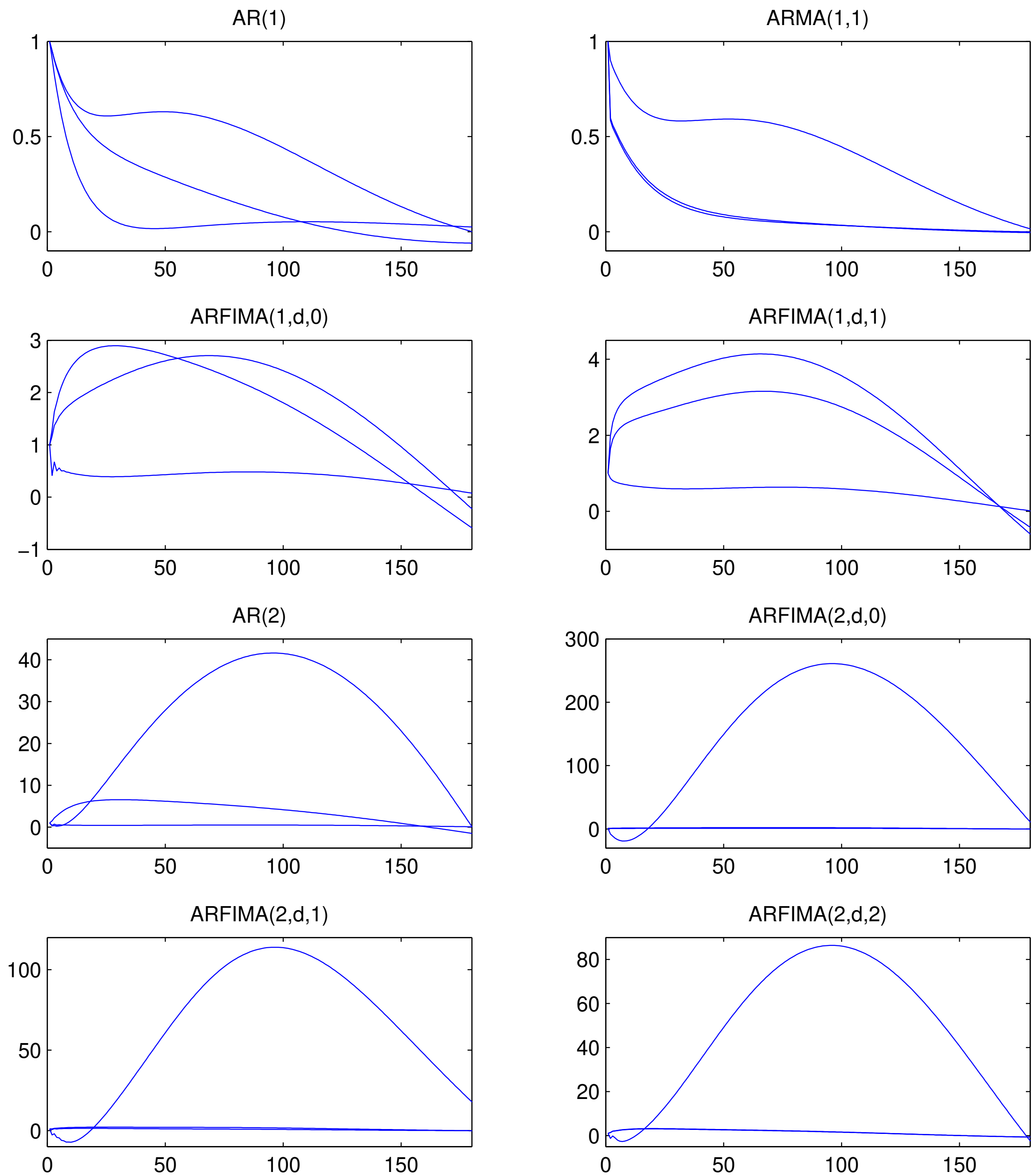


Fig. 7. Impulse responses showing the response over time (measured in months) of the spot rate to unit innovations in the three spanned factors under the real world measure. The responses are given by column sums of $\mathbf{\Phi}'_j$ defined in C-10.

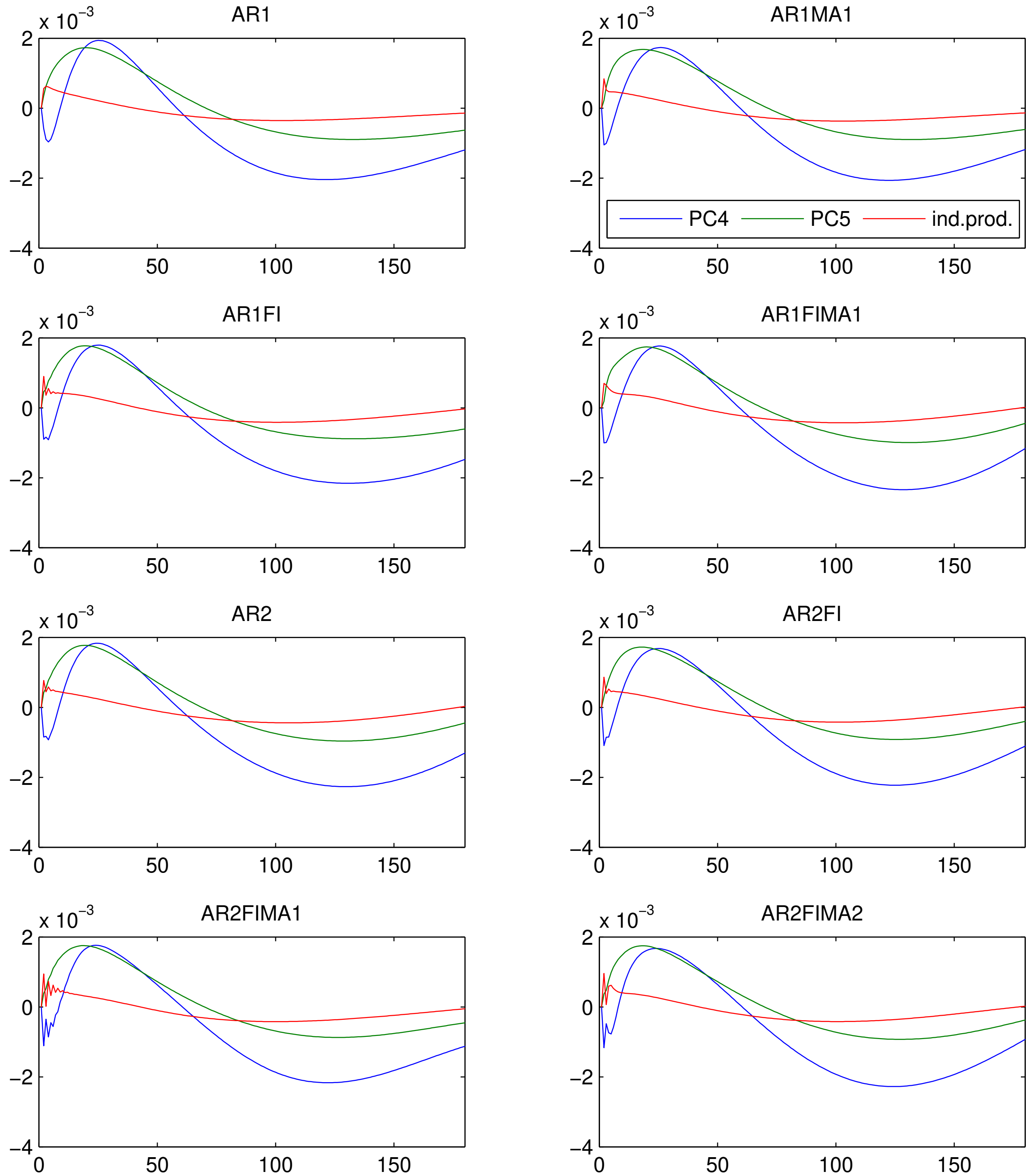


Fig. 8. Impulse responses showing the response over time (measured in months) of the short rate to unit innovations in the standardized unspanned factors: industrial production, fourth and fifth principal components of yields. The responses are given by column sums of $\boldsymbol{\Upsilon}_j'$ defined in C-11.

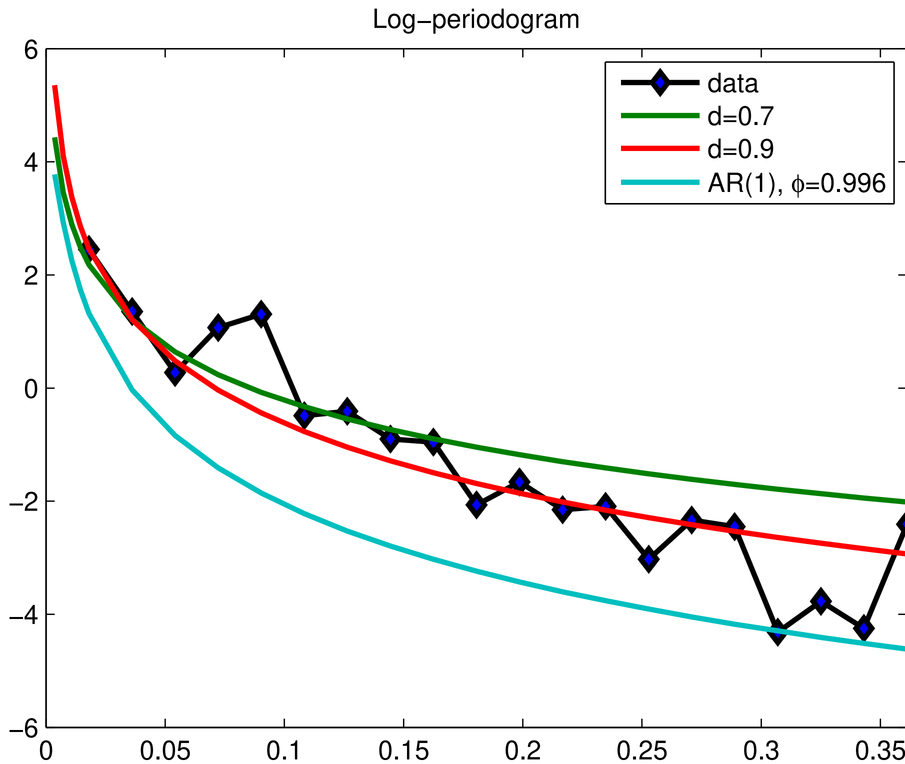


Fig. 9. Logarithm of the periodogram of the first principal component of yields for the first 25 Fourier frequencies together with spectral densities for fractionally integrated processes with long memory parameter equal to 0.7 and 0.9 and $AR(1)$ with an autoregressive coefficient equal to 0.996.