# Not-for-Publication Appendix to "A New Linear Estimator for Gaussian Dynamic Term Structure Models"

**Antonio Diez de los Rios**
Bank of Canada
diez@bankofcanada.ca

February 2014

## A Generalized inverse of $\mathbf{V}_g$

As noted in the paper, one has to be careful not to delete any of the identity equations in $\mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbf{0}$ that define $\boldsymbol{\theta}_2 = \boldsymbol{\pi}_2$ and $\boldsymbol{\theta}_3 = \boldsymbol{\pi}_3$ when computing the generalized inverse of $\mathbf{V}_g$. In particular, $\mathbf{g}_1(\boldsymbol{\pi}, \boldsymbol{\theta}) = vec[\mathbf{G}(\boldsymbol{\pi}, \boldsymbol{\theta})]$ with $\mathbf{G}(\boldsymbol{\pi}, \boldsymbol{\theta})$ in equation (16) in the main text of the paper does not identify $\boldsymbol{\theta}_2$, and only weakly identifies $\boldsymbol{\theta}_3$ (i.e., the innovation parameters of the VAR dynamics under $\mathbb{P}$ only appear in the pricing equations through a (small) Jensen's inequality term). Thus, eliminating any of the identities in $\mathbf{g}_2(\boldsymbol{\pi}, \boldsymbol{\theta}) = \boldsymbol{\pi}_2 - \boldsymbol{\theta}_2$ or $\mathbf{g}_3(\boldsymbol{\pi}, \boldsymbol{\theta}) = \boldsymbol{\pi}_3 - \boldsymbol{\theta}_3$, might lead to numerical instabilities of the CGLS estimates.

To solve this problem, we note that

$$\sqrt{T}\begin{bmatrix} \mathbf{g}_1(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \\ \mathbf{g}_2(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \\ \mathbf{g}_3(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \end{bmatrix} \xrightarrow{d} N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V}_{g_{11}} & \mathbf{0} & \mathbf{V}_{g_{13}} \\ \mathbf{0} & \mathbf{V}_{g_{22}} & \mathbf{0} \\ \mathbf{V}_{g_{13}} & \mathbf{0} & \mathbf{V}_{g_{33}} \end{pmatrix}\right] \tag{A.1}$$
$$\sqrt{T}\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_g)$$

where $\mathbf{V}_{g_{12}}, \mathbf{V}_{g_{21}}, \mathbf{V}_{g_{23}}$ and $\mathbf{V}_{g_{32}}$ are equal to blocks of zeros given the structure of $\partial\mathbf{g}/\partial\boldsymbol{\pi}'$ and $\mathbf{V}_{\boldsymbol{\pi}}$. We further note that $\mathbf{V}_{g_{22}}$ and $\mathbf{V}_{g33}$ are both invertible, and the source of the singularity of $\mathbf{V}_g$ comes from the reduced rank structure of $\mathbf{V}_{g_{11}}$.

In particular, we propose to orthogonalize $\mathbf{g}_1(\boldsymbol{\pi}, \boldsymbol{\theta})$ with respect to $\mathbf{g}_3(\boldsymbol{\pi}, \boldsymbol{\theta})$, and let $\mathbf{g}_{1|3}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) = \mathbf{g}_1(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) - \mathbf{V}_{g_{13}}(\boldsymbol{\theta}^0)\left[\mathbf{V}_{g_{33}}(\boldsymbol{\theta}^0)\right]^{-1}\mathbf{g}_3(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0)$ be the residual of such a regression. Further, the asymptotic normality of $\sqrt{T}\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0)$ allows us to write

$$\sqrt{T}\begin{bmatrix} \mathbf{g}_{1|3}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \\ \mathbf{g}_2(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \\ \mathbf{g}_3(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \end{bmatrix} \xrightarrow{d} N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V}_{g_{1|3}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{g_{22}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_{g_{33}} \end{pmatrix}\right] \tag{A.2}$$
$$\sqrt{T}\widetilde{\mathbf{g}}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{\widetilde{g}})$$

where $\mathbf{V}_{g_{1|3}} = \mathbf{V}_{g_{11}} - \mathbf{V}_{g_{13}}\mathbf{V}_{g_{33}}^{-1}\mathbf{V}_{g_{31}}$ is the asymptotic residual variance of the limiting least squares projection of $\sqrt{T}\mathbf{g}_1(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0)$ on $\sqrt{T}\mathbf{g}_3(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0)$. Note that $\widetilde{\mathbf{g}}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0)$ and $\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0)$ are linearly related trough:

$$\begin{bmatrix} \mathbf{g}_{1|3}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \\ \mathbf{g}_2(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \\ \mathbf{g}_3(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \end{bmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & -\mathbf{V}_{g_{13}}\mathbf{V}_{g_{33}}^{-1} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{bmatrix} \mathbf{g}_{1|3}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \\ \mathbf{g}_2(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \\ \mathbf{g}_3(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) \end{bmatrix} \tag{A.3}$$
$$\widetilde{\mathbf{g}}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) = \mathbf{A}\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0)$$

Given this linear relationship, we propose the following generalized inverse of $\mathbf{V}_g$:

$$\mathbf{V}_g^+ = \mathbf{A}'\mathbf{V}_{\widetilde{g}}^+\mathbf{A} \tag{A.4}$$

where

$$\mathbf{V}_{\widetilde{g}}^+ = \begin{pmatrix} \mathbf{V}_{g_{1|3}}^+ & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{g_{22}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_{g_{33}}^{-1} \end{pmatrix}$$

where $\mathbf{V}_{g_{1|3}}^+$ is a generalized inverse of $\mathbf{V}_{g_{1|3}} = \mathbf{V}_{g_{11}} - \mathbf{V}_{g_{13}}\mathbf{V}_{g_{33}}^{-1}\mathbf{V}_{g_{31}}$ such that $\mathbf{V}_{g_{1|3}}\mathbf{V}_{g_{1|3}}^+\mathbf{V}_{g_{1|3}} = \mathbf{V}_{g_{1|3}}$. Expanding equation (A.4), we have

$$\mathbf{V}_g^+ = \begin{pmatrix} \mathbf{V}_{g_{1|3}}^+ & \mathbf{0} & -\mathbf{V}_{g_{1|3}}^+\mathbf{V}_{g_{13}}\mathbf{V}_{g_{33}}^{-1} \\ \mathbf{0} & \mathbf{V}_{g_{22}}^{-1} & \mathbf{0} \\ -\mathbf{V}_{g_{33}}^{-1}\mathbf{V}_{g_{31}}\mathbf{V}_{g_{1|3}}^+ & \mathbf{0} & \mathbf{V}_{g_{33}}^{-1} + \mathbf{V}_{g_{33}}^{-1}\mathbf{V}_{g_{31}}\mathbf{V}_{g_{1|3}}^+\mathbf{V}_{g_{13}}\mathbf{V}_{g_{33}}^{-1} \end{pmatrix}.$$

After some tedious but straightforward algebra, it is easy to show that $\mathbf{V}_g\mathbf{V}_g^+\mathbf{V}_g = \mathbf{V}_g$ so that $\mathbf{V}_g^+$ is a generalized inverse of $\mathbf{V}_g$.

# B Adrian, Crump and Moench (2013) from an ALS perspective

Adrian, Crump and Moench (2013) focus on bond excess holding period returns rather than on yields themselves. In particular, the one-period excess return on a bond of maturity $n$ is the gain from buying an $n$-period bond and selling it one year later, financing the position at the short rate:

$$rx_{t+1,n} \equiv \log\left(\frac{P_{t+1,n-1}}{P_{t,n}}\right) - r_t = ny_{t,n} - (n-1)y_{t+1,n-1} - r_t. \tag{B.1}$$

Substituting (1), (6), (7) and (8) into this last expression, we can show that the one-period excess return for holding an $n$-period zero-coupon bond is given by

$$rx_{t+1,n} = -\frac{1}{2}\mathbf{B}_{n-1}'\boldsymbol{\Sigma}\mathbf{B}_{n-1} + \mathbf{B}_{n-1}'(\boldsymbol{\lambda}_0 + \boldsymbol{\lambda}_1\mathbf{f}_t) + \mathbf{B}_{n-1}'\mathbf{v}_{t+1},$$

and, adding an *iid* pricing error $\varepsilon_{t,n}$ to this last equation, we have that

$$rx_{t+1,n} = D_n + \mathbf{E}_n'\mathbf{f}_t + \mathbf{F}_n'\mathbf{v}_{t+1} + \varepsilon_{t,n}, \tag{B.2}$$

where

$$D_n = -\frac{1}{2}\mathbf{F}'_n\boldsymbol{\Sigma}\mathbf{F}_n + \mathbf{F}'_n\boldsymbol{\lambda}_0, \tag{B.3}$$

$$\mathbf{E}'_n = \mathbf{F}'_n\boldsymbol{\lambda}_1, \tag{B.4}$$

for $n = 2, ..., N$.

By a similar argument to the one proposed in section 2.2, we have that if the innovation covariance matrix $\boldsymbol{\Sigma}$ and the set of coefficients $D_n$, $\mathbf{E}_n$ and $\mathbf{F}_n$'s were observed directly, one could easily estimate the price of risk parameters of the model using a set of (cross-sectional) OLS regressions. In particular, one could recover an estimate of $\boldsymbol{\lambda}_1$ as

$$\widehat{\boldsymbol{\lambda}}_1 = \left(\sum_{n=2}^{N}\mathbf{F}_n\mathbf{F}'_n\right)^{-1}\left(\sum_{n=2}^{N}\mathbf{F}_n\mathbf{E}'_n\right), \tag{B.5}$$

while an estimate of $\boldsymbol{\lambda}_0$ from

$$\widehat{\boldsymbol{\lambda}}_0 = \left(\sum_{n=2}^{N}\mathbf{F}_n\mathbf{F}'_n\right)^{-1}\left[\sum_{n=1}^{N}\mathbf{F}_n(D_n - \frac{1}{2}\mathbf{F}'_n\boldsymbol{\Sigma}\mathbf{F}_n)\right]. \tag{B.6}$$

However, this estimator is (again) infeasible, because the innovation covariance matrix $\boldsymbol{\Sigma}$, and the set of coefficients $D_n$, $\mathbf{E}_n$ and $\mathbf{F}_n$'s are, in practice, unknown. Instead, we could follow the same principles used to develop the linear estimator in section 2.2, and replace these unknown quantities by some consistent estimates. Specifically, we could first estimate the VAR(1) process in equation (1) to obtain $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\Phi}}$ and $\widehat{\boldsymbol{\Sigma}}$, as well as an estimate of the innovation, $\widehat{\mathbf{v}}_{t+1}$. Second, we could use equation (B.1) to run a regression of $rx_{t,n}$ on a constant, the lagged pricing factors, $\mathbf{f}_{t-1}$, and the contemporaneous pricing factor innovations $\widehat{\mathbf{v}}_t$ to obtain a set of estimates of $D_n$, $\mathbf{E}_n$ and $\mathbf{F}_n$'s for $n = 2, ..., N$. Finally, we could recover an estimate for the market prices of risk parameters, $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$ from the cross-sectional regressions in (B.5) and (B.6) by simply replacing the unknown objects in these equations by the consistent estimates obtained in the previous step. In fact, such an approach is exactly the three-step linear regression method proposed by Adrian, Crump and Moench (2013). In addition, the short-rate parameters, $\delta_0$ and $\boldsymbol{\delta}_1$, can be obtained by running an OLS regression of one-period yield on the pricing factors.

Along the same lines of section 3.2, we can now interpret the ACM estimator within the ALS framework. In particular, we have that the vector of reduced-form parameters is given by

$$\boldsymbol{\pi}_{ACM} = \left((\delta_0 \ \boldsymbol{\delta}'_1), \left\{vec\left[(\mathbf{D}\ \mathbf{E}\ \mathbf{F})'\right]\right\}', \left\{vec\left[(\boldsymbol{\mu}\ \boldsymbol{\Phi})'\right]\right\}', \left[vech\left(\boldsymbol{\Sigma}^{1/2}\right)\right]'\right)',$$

where $\mathbf{D}$ is a vector that stacks the corresponding elements of $D_n$, and $\mathbf{E}$ and $\mathbf{F}$ are matrices that stack the corresponding elements of $\mathbf{E}'_n$ and $\mathbf{F}'_n$.

On the other hand, using that $\boldsymbol{\mu}^{\mathbb{Q}} = \boldsymbol{\mu} - \boldsymbol{\lambda}_0$ and $\boldsymbol{\Phi}^{\mathbb{Q}} = \boldsymbol{\Phi} - \boldsymbol{\lambda}_1$ with equations (B.3) and (B.4) and stacking, it is possible to express the restrictions implied by the no-arbitrage model as

$$\mathbf{H}(\boldsymbol{\pi}_{ACM}, \boldsymbol{\theta})' = \mathbf{Y}_{ACM}(\boldsymbol{\pi}_{ACM}) - \mathbf{X}_{ACM}(\boldsymbol{\pi}_{ACM})\boldsymbol{\Upsilon}' = \mathbf{0}, \tag{B.7}$$

where

$$\mathbf{Y}_{ACM}(\boldsymbol{\pi}_{ACM}) = \begin{pmatrix} \delta_0 & \boldsymbol{\delta}_1' \\ D_2 - \frac{1}{2}\mathbf{F}_2'\boldsymbol{\Sigma}\mathbf{F}_2 & \mathbf{E}_2' \\ \vdots & \vdots \\ D_N - \frac{1}{2}\mathbf{F}_N'\boldsymbol{\Sigma}\mathbf{F}_N & \mathbf{E}_N' \\ \boldsymbol{\mu} & \boldsymbol{\Phi} \end{pmatrix}, \quad \mathbf{X}_{ACM}(\boldsymbol{\pi}_{ACM}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\mathbf{F}_2' & \mathbf{F}_2' \\ \vdots & \vdots & \vdots \\ 0 & -\mathbf{F}_N' & \mathbf{F}_N' \\ 0 & 0 & \mathbf{I} \end{pmatrix},$$

and where $\boldsymbol{\Upsilon}$ satisfies

$$\boldsymbol{\Upsilon}' = \begin{pmatrix} \delta_0 & \boldsymbol{\delta}_1' \\ \boldsymbol{\mu}^{\mathbb{Q}} & \boldsymbol{\Phi}^{\mathbb{Q}} \\ \boldsymbol{\mu} & \boldsymbol{\Phi} \end{pmatrix}.$$

Note that $\boldsymbol{\theta} = \left\{ vec(\boldsymbol{\Upsilon}'), \left[ vech\left(\boldsymbol{\Sigma}^{1/2}\right)\right]' \right\}$, so by vectorizing equation (B.7) and adding a set of identities, it is possible to arrive at the following distance function for the case of ACM estimation:

$$\mathbf{h}(\boldsymbol{\pi}_{ACM}, \boldsymbol{\theta}) = vec\left[\mathbf{H}(\boldsymbol{\pi}_{ACM}, \boldsymbol{\theta})\right] = \boldsymbol{\gamma}_{ACM}(\boldsymbol{\pi}_{ACM}) - \boldsymbol{\Gamma}_{ACM}(\boldsymbol{\pi}_{ACM})\boldsymbol{\theta},$$

where

$$\boldsymbol{\gamma}_{ACM}(\boldsymbol{\pi}_{ACM}) = \left[\mathbf{Y}_{ACM}(\boldsymbol{\pi}_{ACM}), vech\left(\boldsymbol{\Sigma}^{1/2}\right)\right]',$$
$$\boldsymbol{\Gamma}_{ACM}(\boldsymbol{\pi}_{ACM}) = \begin{pmatrix} \mathbf{X}_{ACM}(\boldsymbol{\pi}_{ACM}) \otimes \mathbf{I} & 0 \\ 0 & \mathbf{I} \end{pmatrix}.$$

With this notation, the ACM estimator is equivalent to the estimator that minimizes a quadratic form in the distance function $\mathbf{h}(\boldsymbol{\pi}_{ACM}, \boldsymbol{\theta})$, evaluated at the estimates of the reduced-form parameters, $\widehat{\boldsymbol{\pi}}_{ACM}$, where the weighting matrix has been chosen to be the identity matrix, $\mathbf{W}_T = \mathbf{I}$. Therefore, it is possible to achieve efficiency gains by selecting an appropriate weighting matrix and imposing the self-consistency of the model. However, note that that system of implicit relationships $\mathbf{h}(\boldsymbol{\pi}_{ACM}, \boldsymbol{\theta})$ is not complete (the number of reduced-form parameters is larger than the dimension of the distance function), so, even if we were using an optimal weighting matrix and impose self-consistency, the ACM approach would still not be asymptotically equivalent to maximum likelihood estimation.

# C Additional extensions

## C.1 GDTSMs with lags

In a recent paper, Joslin, Le and Singleton (2013a) extend the family of GDTSMs to accommodate higher-order dynamics (i.e., beyond the VAR(1) model in equation (1)) in the parameterization of the distribution of yields under $\mathbb{P}$, while preserving the parsimonious factor representation of yields. These authors assume that the factors $\mathbf{f}_t$ follow a VAR($p$) under the physical distribution, and a VAR(1) under the risk-neutral measure. Since this can be achieved by assuming that the lags of $\mathbf{f}_t$ are, in essence, unspanned from the cross-section of interest rates, our linear estimator can still be used to estimate this new class of GDTSMs with lags.

## C.2 Measurement errors

The approach in this paper can be extended to the case of observable factors with measurement errors. In such a case, and given the dimensionality of the problem, one needs to estimate the reduced-form parameters using the computationally efficient (Kalman filtering) techniques of Jungbacker and Koopman (2008). Still, it is important to recall that Joslin, Le and Singleton (2013b) show that, in practice, the (fitting) gain from assuming that all observable factors are subject to measurement errors is minimal when one uses the first $M$ principal components of yields as factors.

## C.3 Autocorrelation of the residuals

Our framework can be easily adapted to handle autocorrelation in the measurement errors and/or overlapping in the dynamics under the physical measure. For example, both Cochrane and Piazzesi (2008), and Bauer and Diez de los Rios (2012) focus on annual dynamics of yields estimated using monthly data, which induces a moving average structure on the residuals of the VAR dynamics in equation (1). Similarly, and as noted above, ACM show that uncorrelated pricing errors on excess returns deliver autocorrelated pricing errors on yields. As a difference with maximum likelihood estimation, our framework can naturally handle the presence of autocorrelation in the residuals as long as we estimate the covariance of the reduced-form parameters using a method that is robust to autocorrelation, i.e., using Newey and West (1987).

## C.4 Temporal aggregation

Interest rates evolve on a much finer time scale than the frequency of observations typically employed by empirical researchers. While the sampling frequency is often given because collecting data is very expensive in terms of time and money (e.g., output or labor force statistics), this is no longer the case for financial prices. In fact, for interest rates (i.e., bond prices), currently the sampling frequency is, to a large extent, chosen by the researchers. Marcellino (1999) and Diez de los Rios and Sentana (2011), among others, show in the context of the estimation of time-series models that this choice has an impact on the properties of the estimators/tests considered.

In our case, researchers also need to choose which bonds to use in the estimation of GDTSMs. Specifically, since the $a_n$ and $\mathbf{b}_n$'s can be considered as time-series processes indexed by maturity, selecting which bonds to use in the estimation is essentially equivalent to choosing a "cross-sectional" or risk-neutral frequency of observation. Thus, paralleling the case of time-aggregation under $\mathbb{P}$, it can be shown that this second choice might affect the statistical properties of the GDTSM estimators, specifically, the efficiency of the parameter estimates. In particular, we note that using the full set of maturities (as prescribed by our methodology) versus using only a sparse selection of yields (as usually done in the literature), or increasing the number of bonds in the estimation, might deliver efficiency gains. To see why this is the case, we can partition $\mathbf{g}(\boldsymbol{\pi},\boldsymbol{\theta}) = [\mathbf{g}_1'(\boldsymbol{\pi},\boldsymbol{\theta})\ \mathbf{g}_2'(\boldsymbol{\pi},\boldsymbol{\theta})]' = \mathbf{0}$, and note that, in general, $\mathbf{V}_g(\boldsymbol{\theta}^0) = avar\left[\sqrt{T}\mathbf{g}(\widehat{\boldsymbol{\pi}},\boldsymbol{\theta}^0)\right]$ and $\mathbf{V}_{g_1}(\boldsymbol{\theta}^0) = avar\left[\sqrt{T}\mathbf{g}_1(\widehat{\boldsymbol{\pi}},\boldsymbol{\theta}^0)\right]$

are not necessarily the same. Hence, and consistent with the Monte Carlo evidence presented in section F, the estimator that uses only a subset of bonds (and therefore, just a subset of the distance functions) and a weighting matrix given by $\mathbf{V}_{g_1}^{+}$ will not be optimal in this set-up, even if it imposes the self-consistency restrictions.

Furthermore, our framework can be adapted to handle situations where the physical and risk-neutral sampling frequencies do not coincide. Specifically, it is possible to exploit that the VAR(1) model is closed under temporal aggregation to derive an alternative regression-based estimation approach that deals with this mismatch of sampling frequencies. In particular, iterating the pricing equations for bond loadings in (7), we have:

$$\mathbf{B}_n' = \mathbf{B}_{n-j}' \left(\mathbf{\Phi}^{\mathbb{Q}}\right)^j + \mathbf{B}_1' \left[\sum_{i=0}^{j-1} \left(\mathbf{\Phi}^{\mathbb{Q}}\right)^i\right],$$

which, given that

$$\mathbf{B}_j' = \mathbf{B}_1' \left[\sum_{i=0}^{j-1} \left(\mathbf{\Phi}^{\mathbb{Q}}\right)^i\right], \tag{C.1}$$

allows us to write:

$$\mathbf{B}_n' = \mathbf{B}_{n-j}' \left(\mathbf{\Phi}^{\mathbb{Q}}\right)^j + \mathbf{B}_j'. \tag{C.2}$$

While this equation is not linear in $\mathbf{\Phi}^{\mathbb{Q}}$, it is linear in $\mathcal{F}^{\mathbb{Q}} = \left(\mathbf{\Phi}^{\mathbb{Q}}\right)^j$. Thus, if a set of factor loadings $\{\mathbf{B}_{nj}\}_{n=1}^{N/j}$ were observed directly, the structure of the model would allow us to recover an estimate of $\mathcal{F}^{\mathbb{Q}}$ from the cross-sectional regression of $\left(\mathbf{B}_n' - \mathbf{B}_j'\right)$ on $\mathbf{B}_{n-j}'$. Assuming that $\mathcal{F}^{\mathbb{Q}}$ is diagonalizable, i.e. $\mathcal{F}^{\mathbb{Q}} = \mathbf{T}\left(\mathbf{\Psi}^{\mathbb{Q}}\right)^j \mathbf{T}^{-1}$ where $\left(\mathbf{\Psi}^{\mathbb{Q}}\right)^j$ is a diagonal matrix, an estimate of $\mathbf{\Phi}^{\mathbb{Q}}$ can be obtained as $\mathbf{\Phi}^{\mathbb{Q}} = \mathbf{T}\mathbf{\Psi}^{\mathbb{Q}}\mathbf{T}^{-1}$ where $\mathbf{\Psi}^{\mathbb{Q}}$ is a diagonal matrix that contains the $j$th-roots of the diagonal elements of $\left(\mathbf{\Psi}^{\mathbb{Q}}\right)^j$.

It is important to bear in mind that, when $\mathbf{\Phi}^{\mathbb{Q}}$ has complex eigenvalues, there are several matrices $\mathbf{\Phi}^{\mathbb{Q}}$ that deliver the same $\mathcal{F}^{\mathbb{Q}}$ (i.e., there is no bijection between $\mathbf{\Phi}^{\mathbb{Q}}$ and $\mathcal{F}^{\mathbb{Q}}$). This problem, known as aliasing (see, e.g., Phillips, 1973; Hansen and Sargent, 1981; Bergstrom, 1984), implies that it is not possible to distinguish between parameter structures generating oscillations under $\mathbb{Q}$ at frequencies higher than the interval chosen for the maturities of bonds (i.e., the "cross-sectional" frequency). For example, it will not be possible to identify the parameters driving the $\mathbb{Q}$-dynamics at the monthly frequency if we only have interest rates with quarterly maturities.[1] In this case, we locally identify $\mathbf{\Phi}^{\mathbb{Q}}$ by restricting our attention to complex values inside the unit circle only.

Once an estimate of $\mathbf{\Phi}^{\mathbb{Q}}$ is obtained, it is straightforward to get an estimate of $\boldsymbol{\delta}_1$. In

[1] To see this point, let $\psi = \rho\left(\cos\omega \pm i\sin\omega\right)$ denote a complex eigenvalue of $\mathbf{\Phi}^{\mathbb{Q}}$ and its conjugate. By De Moivre's theorem, we have that $\psi^n = \rho^n\left(\cos n\omega \pm i\sin n\omega\right)$. Now consider the following complex number in polar form $\widetilde{\psi} = \rho\left(\cos\widetilde{\omega} \pm i\sin\widetilde{\omega}\right)$ with $\widetilde{\omega} = \omega + 2\pi/n$. Using standard trigonometry results, we have that $\psi^n = \widetilde{\psi}^n$. Thus, two matrices $\mathbf{\Phi}^{\mathbb{Q}}$ and $\widetilde{\mathbf{\Phi}}^{\mathbb{Q}}$ with the same eigenvectors, and eigenvalues given by $\psi$ and $\widetilde{\psi}$, respectively, will deliver the same $\mathcal{F}^{\mathbb{Q}}$.

particular, we can substitute $\mathbf{B}_1 = -\boldsymbol{\delta}_1$, and solve for $\boldsymbol{\delta}_1$ in equation (C.1) and get:

$$\boldsymbol{\delta}_1' = -\mathbf{B}_j' \left[ \sum_{i=0}^{j-1} \left( \boldsymbol{\Phi}^{\mathbb{Q}} \right)^i \right]^{-1}. \tag{C.3}$$

In a similar fashion, we can iterate the pricing equations for the constants in (8) and, after some tedious yet straightforward algebra, we have that:

$$\begin{aligned} A_n &= A_{n-j} + \mathbf{B}_{n-j}' \left[ \sum_{i=0}^{j-1} \left( \boldsymbol{\Phi}^{\mathbb{Q}} \right)^i \right] \boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2} \mathbf{B}_{n-j}' \left[ \sum_{i=0}^{j-1} \left( \boldsymbol{\Phi}^{\mathbb{Q}} \right)^i \boldsymbol{\Sigma} \left( \boldsymbol{\Phi}^{\mathbb{Q}\prime} \right)^i \right] \mathbf{B}_{n-j} \\ &+ \frac{1}{2} \mathbf{B}_{n-j}' \left[ \sum_{i=1}^{j-1} \left( \boldsymbol{\Phi}^{\mathbb{Q}} \right)^i \boldsymbol{\Sigma} \mathbf{B}_i \right] + \frac{1}{2} \left[ \sum_{i=1}^{j-1} \mathbf{B}_i' \boldsymbol{\Sigma} \left( \boldsymbol{\Phi}^{\mathbb{Q}\prime} \right)^i \right] \mathbf{B}_{n-j} + A_j. \end{aligned} \tag{C.4}$$

We note that this equation is linear in $\boldsymbol{\mu}^{\mathbb{Q}}$. Thus, provided that the innovation covariance matrix, $\boldsymbol{\Sigma}$; a set of constants, $\{A_{nj}\}_{n=1}^{N/j}$ ; and a set of factor loadings $\{\mathbf{B}_{nj}\}_{n=1}^{N/j}$ were observed directly, we could estimate $\boldsymbol{\mu}^{\mathbb{Q}}$ using cross-sectional regressions using our previous estimates of $\boldsymbol{\delta}_1$ and $\boldsymbol{\Phi}^{\mathbb{Q}}$. In particular, such a regression requires estimates of $\mathbf{B}_i$ for $i = 1, \ldots, j$, which can be easily obtained from equation (C.1).

Finally, iterating (8), we have that:

$$A_j = \left( \sum_{i=1}^{j-1} \mathbf{B}_i' \right) \boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2} \left( \sum_{i=1}^{j-1} \mathbf{B}_i' \boldsymbol{\Sigma} \mathbf{B}_i \right) + j A_1,$$

which, by using that $A_1 = -\delta_0$, allows to obtain an estimate of $\boldsymbol{\delta}_0$ as

$$\boldsymbol{\delta}_0 = \frac{1}{j} \left[ \left( \sum_{i=1}^{j-1} \mathbf{B}_i' \right) \boldsymbol{\mu}^{\mathbb{Q}} + \frac{1}{2} \left( \sum_{i=1}^{j-1} \mathbf{B}_i' \boldsymbol{\Sigma} \mathbf{B}_i \right) - A_j \right],$$

given $\boldsymbol{\mu}^{\mathbb{Q}}$, $A_j$, and the set $\{\mathbf{B}_i\}_{i=1}^{j}$ obtained above.

Again, this linear estimator is infeasible given that it depends on objects that are, in practice, unknown. Still, along the same lines of section of 2.2, we can replace these unknown objects by consistent estimates (which are readily available from the reduced-form representation of the model). More importantly, this linear estimator can, again, be interpreted as an ALS estimation problem with an identity weighting matrix and, thus, we could use equation (14) to build standard errors.

# D Interpolating yields: an illustration

As noted in the main text of the paper, there are some practical situations where the availability of zero-coupon yields of adjacent maturities is not readily available. In such situations, where the researcher has only access to a subset of zero-coupon bonds, we propose to (i) interpolate the bonds yields of the remaining set of maturities and (ii) assume that such interpolated yields are subject to larger measurement errors than the set of original yields.

In order to deal with such situations, we propose to suitably model $\mathbf{\Omega}$, i.e. the covariance matrix of the measurement errors $\boldsymbol{\eta}_t$ in equation (11), such that the interpolated yields are subject to larger measurement errors than the set of original yields. In particular, let $\mathbf{y}_{1t}$ be the original set of $N_1$ bond yields that are readily observable, and let $\mathbf{H}_1$ be a $N_1 \times N$ selector matrix of one and zeroes such that $\mathbf{y}_{1t} = \mathbf{H}_1\mathbf{y}_t$ where $\mathbf{y}_t$ is the vector containing the full set of consecutive maturities from $n = 1$ to $N$. Let $\mathbf{y}_{2t} = \mathbf{H}_2\mathbf{y}_t$ be the set of $N_2 = N - N_1$ (interpolated) bond yields where $\mathbf{H}_2$ be a $N_2 \times N$ selector matrix such that $\mathbf{H}_1'\mathbf{H}_2 = \mathbf{0}$ (which implies that there is no overlap between the bonds in the first and second group). Finally, further assume that the bond state variables are linear combinations of the original set of yields only, $\mathbf{f}_t = \mathbf{P}_1'\mathbf{y}_{1t}^o = \mathbf{P}_1'\mathbf{H}_1\mathbf{y}_t$, so that the full-rank matrix of weights $\mathbf{P} = \mathbf{P}_1'\mathbf{H}_1$.

Specifically, we propose to model the covariance matrix $\mathbf{\Omega}$ such that (i) $\mathbf{\Omega}_1 = \mathbf{H}_1\mathbf{\Omega}\mathbf{H}_1' = \boldsymbol{\sigma}_{\eta_1}^2 \times (\mathbf{P}_{1\perp}\mathbf{P}_{1\perp}')$ where $\mathbf{P}_{1\perp}'$ is a basis for the orthogonal component of the row span of $\mathbf{P}_1'$, (ii) $\mathbf{\Omega}_2 = \mathbf{H}_2\mathbf{\Omega}\mathbf{H}_2' = \boldsymbol{\sigma}_{\eta_2}^2 \times \mathbf{I}$ where $\boldsymbol{\sigma}_{\eta_2}^2/\boldsymbol{\sigma}_{\eta_1}^2 = \omega > 1$, and (iii) $\mathbf{H}_1\mathbf{\Omega}\mathbf{H}_2' = \mathbf{0}$. Again, $\boldsymbol{\sigma}_{\eta_1}^2$ can be concentrated from the likelihood function.

Using the Canadian zero-coupon bond yields described in section 6 of the main text of the paer, we now illustrate (i) how, by appropriately choosing the parameter $\omega$, a researcher can control how much noisier the interpolated set of bond yields are thought to be when compared to the original one, and (ii) how sensitive is the fit of the model to this choice.

For the sake of simplicity, we focus on a two-factor model with no risk-premia restrictions and no bias corrections. In particular, we will assume that the original data set consists of bonds maturities two, four, eight, twelve, twenty, twenty eight, forty and sixty quarters[2]. Therefore, we will use only these bonds when computing the principal components. On the other hand, the remaning bond yields will be treated as interpolated from this initial set of maturities, and therefore, subject to large measurement errors.[3]

Figure A1 presents the root mean squared pricing error (RMSPE) of the original set of bonds yields as a function of the parameter $\omega$. We include the RMSPE of the model implied by the reduced-form model and the OLS estimates (which do not depend on $\omega$) as well for comparison. Specifically, the RMSPE of the model implied by the CGLS estimates is close to 12.70 bps which is only marginally worse than the 12.50 bps RMSPE of the reduced-form model. The effect of $\omega$ is small as the model RMSPE stabilizes around 12.70 bps very quickly. Even with a small value of $\omega$ we find that the CGLS approach improve the fit of the model with respect to the OLS estimates. However, the reduction in RMSPE is small. Similar results are found when focusing on mean absolute pricing errors (MAPEs).

Further we note that the estimates of the parameters driving the risk-neutral dynamics of the factors are rather robust to the choice of the parameter $\omega$, where only the estimated

[2]This is the same set of maturities considered by JSZ, augmented by the sixty-quarter (fifteen-year) bond yield.

[3]This distinction is clearly artificial and made only for the point of illustrating the methods presented in the paper given that all bond yields are actually interpolated from price data on Treasury coupon bonds with time-varying maturities

$r_\infty^{\mathbb{Q}}$ seems to be affected by slightly the choice of $\omega$. Again, this is due to the fact that the process is extremely persistent under $\mathbb{Q}$ which makes inference about the risk-neutral long-run mean of the short rate very difficult (see Hamilton and Wu, 2012).

# E A return-forecasting factor for Canada

Cochrane and Piazzesi (2005) (CP) show that (i) a linear combination of forward rates predicts annual bond excess holding period returns with $R^2$ values as high as 0.44, (ii) this single factor has a tent-shaped structure, (iii) this factor captures all of the economically interesting variation in one-year excess returns for bonds of all maturities.

In order to investigate the existence of a similar factor in the Canadian term structure of interest rates, we start by regressing the average (across maturity) annual excess returns at time $t+4$ on forward rates at time $t$:

$$\frac{1}{14}\sum_{n=2}^{15} rx_{t\to t+4,n} = \gamma_0 + \boldsymbol{\gamma}'\mathbf{g}_t + \epsilon_{t+4}, \tag{E.1}$$

where $rx_{t\to t+4,n} \equiv \log\left(P_{t+4,n-4}/P_{t,n}\right) - y_{t,4}$ is the annual bond excess holding period returns, and $\mathbf{g}_t$ is a vector of log forward (annual) interest rates, $g_t^{(n\to n+4)} = p_{t,n} - p_{t,n+4}$.[4] Given the overlapping nature of the regression equation (E.1), we follow CP in computing Newey and West (1987) standard errors with six lags.[5]

The first row in Table A2 reports the estimated values of $\gamma_0$ and $\boldsymbol{\gamma}$ for the original choice of five forwards in CP ($g_t^{(n\to n+4)}$ for $n = 0, 4, 8, 12, 16$). While the predictability is weaker than in the original CP paper (i.e., $R^2 = 0.20$ versus 0.44), the Wald test for the hypothesis that $\boldsymbol{\gamma} = \mathbf{0}$ cannot be rejected at standard confidence levels. However, the regression coefficients present an M shape that is suggestive of multicollinearity. For this reason, we follow Sekkel (2011), who tests the robustness of the CP factor across several international markets, in using only the one-, three-, and five-year forwards ($g_t^{(n\to n+4)}$ for $n = 0, 8, 16$) when estimating equation (E.1). Such results are reported in the second row of Table A2. In this case, the M pattern in the estimated coefficients disappears, and we recover a tent-shaped forecasting factor. Yet, the $R^2$ decreases to 0.17 given the loss of information from reducing the number of forecasting instruments.

Note that neither the CP nor the Sekkel (2011) specifications incorporate the information that long-dated forwards potentially contain. However, rather than using the full set of forward rates, which could lead again to potential collinearity issues, we focus on one specification of equation (E.1) that has only five forward rates as regressors. In particular, we use the one-, two-, five-, ten- and fifteen-year forward rates ($g_t^{(n\to n+4)}$ for $n = 0, 4, 16, 36, 56$) as our set of regressors. By doing so, the $R^2$ increases to 0.46 and the estimated regression coefficients have the desired tent-shaped structure.[6] Thus, we

[4] In particular, $g_t^{(n\to n+4)}$ is the interest rate at time $t$ for loans between time $t+n$ and $t+n+4$.

[5] The results remain the same when we use Hansen and Hodrick (1980) standard errors with three lags (i.e., the order of the MA process of the error term $\epsilon_{t+4}$ induced by the overlapping problem).

[6] Following CP, we conduct two robustness exercises. First, in order to address the concern that forward

conclude that the information contained in long-dated forwards seems to be important for explaining time variation in Canadian bond premia.

We also verify that this new version of the CP factor captures all of the economically interesting variation in one-year excess returns for Canadian bonds. We do so by following CP once more in comparing the $R^2$ of an unrestricted regression of individual bond excess holding period returns with the $R^2$ of the predictive regression that imposes the one-factor structure in expected excess returns. While unreported for the sake of space, we find that the $R^2$'s from both regressions are essentially the same, which indicates that the single-factor structure in expected returns does little damage to its forecast ability. In other words, this Canadian CP factor captures all of the economically interesting variations in one-year excess returns for bonds of all maturities.

# F Monte Carlo simulations

In this section, we carry out a Monte Carlo study to assess the finite-sample properties of the proposed ALS estimators of the parameters of GDTSMs. In addition, we also compare our proposed OLS and CGLS estimators to two of the main approaches to the estimation of GDTSMs described above: the maximum likelihood estimator of Joslin, Singleton and Zhu (2011) and the minimum chi-square estimator of Hamilton and Wu (2012). We leave for further research a comparison with the linear estimator of Adrian, Crump and Moench (2013), because parameter estimates obtained using this approach are not directly comparable to the JSZ, HW and ALS estimators given the different distributional assumption on the measurement errors.

## F.1 Design

We simulate 10,000 samples of 25 years of quarterly interest rates ($T = 100$), ranging from one quarter to 15 years ($N = 60$), from a one-factor model using equations (1), (2) and (3) in the main text of the paper. We focus on a one-factor model for its simplicity. For example, note that in the one-factor case there is no need to consider complex or repeated eigenvalues when estimating the model subject to the JSZ normalization. Still, we can illustrate most of the properties of our estimators within this simple framework.

To ensure that self-consistency is satisfied in the simulated data, we focus on the JSZ canonical representation of a GDTSM. That is, we use $\delta_0$, $\boldsymbol{\delta}$, $\boldsymbol{\mu}^{\mathbb{Q}}$ and $\boldsymbol{\Phi}^{\mathbb{Q}}$ defined in equation (27), and the following parameter values: $r_\infty^{\mathbb{Q}} = 0.03$, $\boldsymbol{\Psi}^{\mathbb{Q}} = 0.975$, $\boldsymbol{\mu} = 0.0015$, $\boldsymbol{\Phi} = 0.9$, $\boldsymbol{\Sigma} = 0.003$, and $\sigma_\eta = 0.0015$. Such parameter values are chosen to match the empirical characteristics of our data set. In order to capture the level factor that characterizes yield curve data, we choose $\mathbf{f}_t$ to be equal to the one-year yield, which is

rates on the right-hand side show secular decline over the sample studied, we analyze a specification that uses spread information, $g_t^{(n\to n+4)} - r_t$. Second, in order to address the concern that the price at $t$ is common to both the left- and right-hand sides of the regression, we run the regression in equation (E.1) using forecasting instruments measured at time $t-1$. In both cases, the forecasting power and the tent-shaped pattern are preserved.

assumed to be observed without measurement error. That is, we take $\mathbf{P} = \mathbf{e}_4$ where $\mathbf{e}_j$ is a $N \times 1$ vector with a one in the $j$th position and zeroes in the other. Finally, we draw starting values for the one-year yield (i.e., the factor) from its stationary distribution.

## F.2 Results of the simulations

Table A1 reports the results of our Monte Carlo exercise. In this table, OLS refers to the linear ALS estimator defined in equation (19), and CGLS refers to the iterated version of the ALS estimator defined in equation (26), respectively; HW-ei and HW-oi refer to the exactly identified and overidentified minimum-chi-square estimators in HW; and, finally, ML refers to the maximum likelihood estimator proposed in JSZ. In the exactly identified case, the HW estimator uses only two points of the yield curve. In particular, we choose to match both the one- and ten-year yields. In addition, as is customary in the literature, we use only a sparse set of bond yields when computing the HW and JSZ estimators. In particular, we use yields of maturities two, four, eight, twelve, twenty, twenty eight, forty and sixty quarters when estimating the model using these two methods. On the other hand, we note that our linear estimator uses all the information available in the term structure of interest rates. Therefore, in order to provide a relevant bechmark for our linear framework and to gauge the efficiency loss from discarding data, as traditionally done in the literature, we also compute the ML estimates that use the full span of maturities (ML-all).

The results reported in Table A1 are as follows: *Mean,* the mean (across Monte Carlo replications) of the estimate; *Std,* the standard deviation of the estimate; *EStd,* the sample mean of the estimated asymptotic standard error; *RMSE*, the root-mean-squared error of the estimate; and *CINT-95,* the proportion of times that the true parameter value lies within the 95 per cent asymptotic confidence interval.

Our simulations show that, in terms of bias, the CGLS approach accurately estimates the parameters describing the dynamic evolution of the factor under the risk-neutral measure, $\mathbb{Q}$. On the other hand, the OLS estimates of the canonical parameters seem to be estimated subject to a small downward bias. The bias in the estimated $\mathbb{P}$-parameters, $\boldsymbol{\mu}$ and $\boldsymbol{\Phi}$, is sizable, which is consistent with Bauer, Rudebusch and Wu (2012) and their study of the properties of the HW and JSZ estimators. We note that this is not a problem exclusive to our methodology. In fact, the six estimation methods considered in this Monte Carlo study recover exactly the same OLS estimates of the parameters driving the $\mathbb{P}$-dynamics of the factors. In consequence, our linear estimators will also suffer from the well-known problem that OLS estimates of autoregressive parameters tend to underestimate the persistence of the system in finite samples.

Choosing an optimal weighting matrix and imposing the self-consistency restrictions clearly matters. First, the CGLS estimator has lower variability than the OLS estimator. For example, the standard error of the CGLS estimate of $\boldsymbol{\Phi}^{\mathbb{Q}}$ is smaller than the one corresponding to the OLS estimate by a factor of five. Second, the coverage rate of the CGLS estimator is very close to the 95 per cent nominal rate. In contrast, the fact that the OLS estimated standard errors slightly understate the true variability of the estimate,

combined with the slight bias in this estimator, results in non-trivial differences between the empirical coverage rate and the nominal rate of 95 per cent.

Similarly, discarding bonds when estimating the model has an important effect on the efficiency of the estimator. For example, the standard error of the HW-ei (HW-oi) estimator of $r_\infty^{\mathbb{Q}}$ more than triples (doubles) the standard error of the corresponding CGLS (or ML-full) estimate. A similar pattern can be observed when considering the ML estimator of the GDTSM parameters. In fact, the loss of efficiency incurred from focusing on the HW exactly identified estimator is similar to the loss of efficiency incurred from using an identity matrix (versus using the optimal estimator) within our linear framework. On the other hand, and as predicted by asymptotic theory, the properties of our CGLS estimator are almost identical to the properties of the ML estimator that uses the full set of bonds: a result that we find particularly reassuring.

Finally, we briefly analyze the finite-sample properties of the ALS overidentification test. Specifically, we have that the optimized value of the ALS criterion function has an asymptotic $\chi^2$ distribution with degrees of freedom equal to the number of overidentifying restrictions $(G - K)$. We find that this test has a slight tendency to under-reject. In particular, we find that, in our simulations, the empirical rejection rate of a 5 per cent (10 per cent) confidence-level overidentification test is 2.8 per cent (6.9 per cent).

# References


[1] Adrian, T., R.K. Crump and E. Moench (2013): "Pricing the term structure with linear regressions," *Journal of Financial Economics*, 110, 110–138.

[2] Bergstrom, A. R. (1984): "Continuous time stochastic models and issues of aggregation over time," in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics: Vol. 2*, Elsevier Science Press (Amsterdam), 1145-1212.

[3] Cochrane, J. and M. Piazzesi (2005): "Bond risk premia," *American Economic Review*, 95, 138-60.

[4] Cochrane, J. and M. Piazzesi (2008): "Decomposing the yield curve," mimeo, University of Chicago.

[5] Diez de los Rios, A. and E. Sentana (2011): "Testing uncovered interest parity: a continuous-time approach," *International Economic Review*, 52, 1215-1251.

[6] Hansen, L.P. and R.J. Hodrick (1980): "Forward exchange rates as optimal predictors of future spot rates, an econometric analysis," *Journal of Political Economy,* 88, 829-53.

[7] Hansen, L.P. and T.J. Sargent (1981): "Identification of continuous-time rational expectations models from discrete-time data," in L.P. Hansen and T.J. Sargent (eds.), *Rational Expectations Econometrics*, Westview Press (Boulder), 219-236.

[8] Hamilton, J.D. and J.C. Wu (2012): "Identification and estimation of Gaussian affine term structure models," *Journal of Econometrics*, 168, 315-331.

[9] Joslin, S., A. Le and K.J. Singleton (2013a): "Gaussian macro-finance term structure models with lags," *Journal of Financial Econometrics,* 11, 581-609.

[10] Joslin, S., A. Le and K.J. Singleton (2013b): "Why Gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs," *Journal of Financial Economics*, 109, 604–622.

[11] Joslin, S., K.J. Singleton and H. Zhu (2011): "A new perspective on Gaussian DTSMs," *Review of Financial Studies,* 24, 926-970.

[12] Jungbacker, B. and S.J. Koopman (2008): "Likelihood-based analysis for dynamic factor models," Tinbergen Institute Discussion Paper No. TI 2008-0007/4.

[13] Marcellino, M. (1999): "Some consequences of temporal aggregation in empirical analysis," *Journal of Business and Economic Statistics*, 17, 129-36.

[14] Newey, W.K. and K.D. West (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica,* 55, 703–708.

[15] Phillips, P.C.B. (1973): "The problem of identification in finite parameter continuous time models," *Journal of Econometrics*, 1, 351–262.

[16] Sekkel, R. (2011): "International evidence on bond risk premia," *Journal of Banking and Finance*, 35, 174-181.

**Table A1**
**Finite-sample properties of GDTSMs estimators**

| | | $100 \times r_\infty^{\mathbb{Q}}$ | $\boldsymbol{\Psi}^{\mathbb{Q}}$ | $100 \times \boldsymbol{\mu}$ | $\boldsymbol{\Phi}$ | $100 \times \boldsymbol{\Sigma}$ |
|---|---|---|---|---|---|---|
| | True values | 3.00 | 0.9750 | 0.15 | 0.900 | 0.300 |
| **OLS** | Mean | 2.91 | 0.9734 | 0.21 | 0.860 | 0.296 |
| | Std | 0.14 | 0.0022 | 0.09 | 0.056 | 0.021 |
| | EStd | 0.13 | 0.0021 | 0.08 | 0.050 | 0.021 |
| | RMSE | 0.17 | 0.0027 | 0.11 | 0.070 | 0.022 |
| | CINT-95 | 82.5% | 88.8% | 91.1% | 90.7% | 93.2% |
| **CGLS** | Mean | 2.99 | 0.9750 | 0.21 | 0.860 | 0.297 |
| | Std | 0.04 | 0.0004 | 0.09 | 0.056 | 0.013 |
| | EStd | 0.04 | 0.0004 | 0.08 | 0.050 | 0.013 |
| | RMSE | 0.04 | 0.0004 | 0.11 | 0.070 | 0.013 |
| | CINT-95 | 93.8% | 95.6% | 91.1% | 90.7% | 94.9% |
| **HW-ei** | Mean | 3.00 | 0.9749 | 0.21 | 0.860 | 0.296 |
| | Std | 0.15 | 0.0026 | 0.09 | 0.056 | 0.021 |
| | EStd | 0.15 | 0.0025 | 0.08 | 0.050 | 0.021 |
| | RMSE | 0.15 | 0.0026 | 0.11 | 0.070 | 0.021 |
| | CINT-95 | 94.5% | 94.8% | 91.1% | 90.7% | 93.2% |
| **HW-oi** | Mean | 3.00 | 0.9750 | 0.21 | 0.863 | 0.297 |
| | Std | 0.09 | 0.0014 | 0.09 | 0.056 | 0.019 |
| | EStd | 0.09 | 0.0013 | 0.08 | 0.050 | 0.019 |
| | RMSE | 0.09 | 0.0014 | 0.11 | 0.070 | 0.019 |
| | CINT-95 | 93.9% | 95.3% | 91.1% | 90.7% | 93.1% |
| **ML** | Mean | 3.00 | 0.9750 | 0.21 | 0.860 | 0.297 |
| | Std | 0.09 | 0.0014 | 0.09 | 0.056 | 0.019 |
| | EStd | 0.09 | 0.0014 | 0.08 | 0.050 | 0.019 |
| | RMSE | 0.09 | 0.0014 | 0.11 | 0.070 | 0.019 |
| | CINT-95 | 94.1% | 95.4% | 91.1% | 90.7% | 93.2% |
| **ML-all** | Mean | 3.00 | 0.9750 | 0.21 | 0.863 | 0.299 |
| | Std | 0.04 | 0.0004 | 0.09 | 0.053 | 0.013 |
| | EStd | 0.04 | 0.0004 | 0.08 | 0.050 | 0.012 |
| | RMSE | 0.04 | 0.0004 | 0.11 | 0.065 | 0.013 |
| | CINT-95 | 94.6% | 95.1% | 91.1% | 90.7% | 94.5% |

**Table A2**
**Cochrane and Piazzesi regressions**

| | constant | $g_t^{(0\to4)}$ | $g_t^{(4\to8)}$ | $g_t^{(8\to12)}$ | $g_t^{(12\to16)}$ | $g_t^{(16\to20)}$ | $g_t^{(36\to40)}$ | $g_t^{(56\to60)}$ | $R^2$ | Wald |
|---|---|---|---|---|---|---|---|---|---|---|
| **Original CP (2005)** | -0.90 | -5.00 | 10.70 | -21.72 | 26.65 | -10.28 | | | 0.20 | 25.34 |
| | (3.00) | (1.42) | (5.21) | (10.26) | (11.03) | (5.16) | | | | [<0.001] |
| **Sekkel (2011)** | -1.22 | -2.78 | | 3.00 | | 0.19 | | | 0.17 | 10.22 |
| | (2.91) | (0.92) | | (2.55) | | (2.05) | | | | [0.02] |
| **Long-dated forwards** | -0.25 | -2.95 | 0.79 | | | 5.94 | 2.55 | -5.48 | 0.46 | 82.92 |
| | (2.97) | (1.17) | (1.90) | | | (1.64) | (0.87) | (1.12) | | [<0.001] |

**Note:** Data are sampled quarterly from 1986Q1 to 2012Q2. Newey and West (1987) asymptotic standard errors are given in parentheses.

**Figure A1: Model fit (RMSPE) in bps as a function of ω**