Title
The Term Structure with Highly Persistent Interest Rates

Author
Valkanov, Rossen

Publication Date
1999-10-30

The Term Structure with Highly Persistent Interest Rates

Rossen Valkanov[1]

10/30/1998

[1] I am extremely thankful to Mark Watson for the original insights that led to this paper and for his encouragement and many helpful ideas along the way. I have also benefitted from comments by Yacine Aït-Sahalia, John Campbell, Gregory Chow, Bo Honoré, and Pierre Perron. All remaining errors are mine. Contact: rossen.valkanov@anderson.ucla.edu or Anderson Graduate School of Management at UCLA, 110 Westwood Plaza, Box 951481, Los Angeles, CA 90095-1481, tel: (310) 825-7246.

Abstract

If yields are assumed to have an exact unit-root, it has previously been shown that the rational expectations hypothesis of the term structure (REHTS) has been rejected by single-equation tests. However, small deviations from exact unit-root produce substantial changes in the small sample distributions of those tests and the normal approximation is no longer satisfactory. We assume that the yield of 1-period zero-coupon bond follows a local-to-unity process with parameter c (c=0 for exact unit root) and use asymptotics to derive alternative distributions, which are far better approximations to finite sample distributions. Those asymptotic distributions depend crucially on c, and that allows us to analyze the impact of small deviations from unit-root on the distribution of the tests. Interestingly, for small values of c, the results obtained in the data do not imply a rejection of the REHTS. The above results are useful only when c is known or consistently estimable. Thus, the REHTS is cast into a triangular representation, where the cointegrating vectors are a function of c. Consistent and asymptotically unbiased estimators of c are proposed. A Wald test for the restrictions imposed by the REHTS on the cointegrating relationship is derived. The relevance of the asymptotic results for samples of practical sizes is investigated with Monte Carlo simulations. The methods are applied to the yield data of McCulloch and Kwon (1993). Although the REHTS is statistically rejected, the results are encouraging and suggest interesting directions for further research.

1 Introduction

The rational expectation hypothesis of the term structure (REHTS) is theoretically simple and appealing, even if it does not have the virtues of fully specified general equilibrium models as in Vasicek (1977) and Cox-Ingersoll-Ross (1985). However, the REHTS has been questioned numerous times on empirical grounds, because it has failed to provide even simple forecasting expressions. Indeed, single-equation and VAR-based tests reject the theory quite consistently (Campbell and Shiller (1987, 1991), Campbell et al. (1997) and references therein). However, the econometric testing of such models in not always straight forward. It is first assumed that yields follow a certain process (usually an integrated process of order one, or I(1)). Statistical tests about the predictions of the model are then constructed, conditional on such an assumption, which is often justified by first-stage pretesting (Campbell and Shiller (1987, 1991)). However, small deviations from the unit root assumption, small enough that they cannot be detected by a statistical pretesting in finite samples, may produce large modifications in the distributions of the REHTS tests. In sum, if the REHTS holds but incorrect assumptions are made about the process driving the yield data, the distributions of the tests change dramatically (see Bekaert et al. (1997) for a Monte Carlo demonstration).

This paper has two main goals. First, we ask the question: Given that the REHTS holds, what process should the short yield follow in order for previously employed single-regression tests to replicate the findings in the data (Campbell and Shiller (1991), Shiller (1990), Campbell et al. (1997))? The short term yield of a zero-coupon bond is parameterized to follow a local-to-unity process (Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), and Phillips (1987)) with the largest root $\phi = 1 + \frac{c}{T}$, where a nuisance parameter c measures deviations from unit-root in a decreasing (at rate T) neighborhood of 1. Since yields are believed to be highly persistent, this parameterization is quite appropriate. It is shown that the limiting distributions of the single-regression tests of the REHTS depend crucially on c. The expressions of these distributions, which are functionals of diffusion processes, will allow us to understand why small changes of the yield process produce large changes in the distribution of the tests. Monte Carlo simulations suggest that the derived asymptotic distributions are very good approximations to the finite sample distributions. More importantly, for some values of the nuisance parameter, previous results obtained from single-regression tests cannot be interpreted as a rejection of the REHTS. The only practical difficulty stems from the dependence of the limiting distributions on the unknown parameter c.

The second goal, and the main contribution of this paper, is to construct a consistent estimator of the nuisance parameter. In general, c cannot be estimated consistently from univariate analysis[1]. We exploit the REHTS to write the term structure as a cointegrated system where the cointegrating vectors are a function of c. T-consistent, asymptotically unbiased, and simple to implement estimators of c are proposed. The asymptotic distribution of those estimators is a mixture of normals. Monte Carlo simulations suggest that the asymptotic distribution is a good approximation to the finite sample distributions, even in samples of 100 observations. The cross-equation restrictions implied by the REHTS are tested with a Wald test, whose asymptotic distribution is shown to be chi-square.

The econometric procedures presented below are quite general in nature. They are applicable to a wide class of rational expectations present value models, where assumptions about the data generating process are critical

[1] However, it is possible to obtain median unbiased estimators and centered confidence intervals of the nuisance parameter as in Stock (1991), Andrews (1993), and Dufour (1990).

and controversial. But the proposed methods are particularly germane to the term structure literature. For example, the data generating process of the short rate is a discrete analogue of an Ornstein-Uhlenbeck process, which is assumed to underlie the general equilibrium model of Vasicek (1977)[2]. Moreover, even though we focus on a term structure driven only by the short rate, results from the paper suggest that there might be gains from using higher dimensional processes, thus paralleling the affine-yield, multi-factor models, analyzed in the general equilibrium literature. The extensions of estimating and testing a matrix of nuisance parameters using the long run restrictions imposed by linear rational expectations models are the focus of current research.

The paper is structured as follows. Section 2 is a brief exposition of the rational expectations theory of the term structure, its forecasting implications and empirical failure. Local-to-unity asymptotics are employed to derive the distributions of some widely used single-equation tests of the REHTS. These distributions are extremely sensitive to the magnitude and sign of the parameter c. In section 3, the REHTS is cast into a triangular cointegrated system (Phillips (1991)). Consistent and asymptotically unbiased estimators of c are derived and a Wald test for the REHTS is also proposed. In section 4, we conduct a Monte Carlo simulation of the distributions derived in Sections 2 and 3. Interestingly, for highly persistent interest rates ($c \in (-1, 0)$), the results observed in the yield data can be explained. Moreover, the simulations suggest that the asymptotic distributions are very satisfactory approximations to the finite sample distributions. In section 5, the above methods are used to estimate c and test the REHTS using the yield data from McCulloch and Kwon (1993). A strictly univari-

[2] More precisely, the short rate follows a discrete version of an Ornstein-Uhlenbeck process with a constant drift parameter c and a constant diffusion parameter of unity.

ate way of constructing median unbiased estimators and centered confidence intervals of c by inverting the Augmented Dickey-Fuller test (ADF) is also used (Stock (1991)). Section 6 concludes.

2 Single-Regression Tests of REHTS

2.1 The rational expectations hypothesis of the term structure

Let the yield of a zero-coupon bond with maturity n at time t be $y_{n,t}$. Assume that the yield of the 1-period bond follows the process:

$$y_{1,t} = \phi y_{1,t-1} + u_{1,t} \tag{1}$$

where $\phi = 1 + \frac{c}{T}$, or ϕ is local to unity with parameter c, T is the sample size, and $u_{1,t}$ is an error term, whose properties will be specified below. The literature often assumes that $y_{1,t} = y_{1,t-1} + u_{1,t}$, or $c = 0$, and this assumption is often justified with unit-root tests. However, unit-root tests have very little power against local-to-unity alternatives.

Under the linearized REHTS, yields on long and short term bonds are related by the present value expression

$$y_{n,t} = \alpha_n + \frac{1}{n}\sum_{i=0}^{n-1} E_t\left[y_{1,t+i}\right] \tag{2}$$

where $E_t\left[.\right]$ denotes mathematical expectation, given information at time t, and α_n is a premium. Since statistical tests cannot reject the null that yields have a unit root, the literature has investigated two testable expressions, which contain only I(0) variables (provided that the yields are truly a unit root process):$\frac{1}{n-1}s_{nt} = E_t\left[y_{n-1,t+1} - y_{nt}\right]$ and $s_{nt} = E_t\left[\ \sum_{i=1}^{n-1}(1 - \frac{i}{n}) \bigtriangleup y_{1,t+i}\right]$ where $s_{nt} = y_{nt} - y_{1t}$ is the spread. The first expression, in which the high yield spread forecasts increases in long rate, is tested by running a regression (OLS):

$$(y_{n-1,t+1} - y_{nt}) = c_n + \beta \frac{1}{n-1} s_{nt} + \epsilon_{nt} \tag{3}$$

and under the REHTS, $\beta = 1$. This implication of the theory is not supported by the data: the estimated β are always significantly different from 1 and in most cases also significantly different from zero, very often with a negative sign (Shiller et al. (1983), Shiller (1990), Campbell and Shiller (1991), Campbell et al. (1997)).

In the second expression, the high yield spread forecasts long-term increases in short rates. Let's use ex-post short rate changes and define $s^*_{n,t} = \sum_{i=1}^{n-1}(1 - \frac{i}{n}) \triangle y_{1,t+i}$. One way of testing the theory is by running the regression (OLS):

$$s^*_{n,t} = \gamma_n + \psi s_{n,t} + \varepsilon_t \tag{4}$$

where $\psi = 1$ under the REHTS. This regression seems to have a bit more support in the data (Campbell and Shiller (1991), Campbell et al. (1997)).

2.2 Single-Regression Local to Unity Asymptotics

Using Monte-Carlo simulations, Bekaert et al. (1997) show that if ϕ is close to unity, the finite sample distributions of $\hat{\beta}$ and $\hat{\psi}$ are very poorly approximated by the normal distribution for samples as big as 524. In fact, it takes the authors 20,000 observations in order to get normal-looking distributions. Even then, the distributions have a considerable spread and seem to depend on n, the maturity of the bond. All those observations, corroborated by our own simulations, prompted us to look for a more systematic way of analyzing the impact of small deviations from unit-root on the statistics of interest.

In this section, we derive alternative asymptotic distributions of $\hat{\beta}$ and $\hat{\psi}$ that approximate very well the finite sample distributions for $c \neq 0$ (if $c = 0$, standard asymptotic theory is satisfactory). Those distributions, which are functionals of stochastic integrals, depend crucially on the nuisance parameter c, in the sense that a very small change in c, indistinguishable in a finite sample from a statistical point of view, leads to dramatically different distributions. Furthermore, we are able to show how the distributions change when ϕ is close to, but not exactly at unity. Of course, a Monte Carlo approach is also possible (see Bekaert et al. (1997)), but then, no analytic results showing the dependence of the distributions on ϕ would be available. It is known that local-to-unity asymptotics provide a good approximation to the finite sample distributions when ϕ is close to one (see Stock (1994) and references therein), and simulations reaffirm this fact in the present setup. Hence, we adopt the view in Stock (1995) that local-to-unity asymptotics provide "a magnifying glass which focuses on the problematic dependence of the finite-sample distributions" on ϕ.

We use the parameterization $n = [\pi T]$, where $[.]$ is the greatest integer less than πT, $0 < \pi < 1$. All limits are taken as $T \uparrow \infty$, for π fixed and known. Strictly speaking, when $T \uparrow \infty$, then $n \uparrow \infty$, and the analysis is de facto focusing on the short and the infinite-maturity (consol) bonds, not on the entire term structure. However, in practice T is fixed, and for any maturity n we can find a π such that $n = [\pi T]$ holds. In other words, the asymptotic distributions derived below can be interpreted as an approximation of the finite sample distributions, where π is fixed and known. A Monte Carlo simulation demonstrates that such an approximation is satisfactory for 100 or more observations and $\pi \geq 0.05$.

We make the following assumption:

Assumption A [3]Let $\{u_{1,t}\}_0^\infty$ be a random variable that satisfies: $u_{1,t} =$

[3]Assumption A can be relaxed considerably to allow for weakly dependent heterogeneously distributed innovations, but will not add anything to the arguments laid out below. See Phillips (1987) and Hansen (1992).

$b(L)\varepsilon_t$, where ε_t is a martingale difference sequence with $E(\varepsilon_t^2) = \sigma^2$ and finite fourth moment, $b(L) = \sum_{j=0}^{q} b_j L^j$, $b_0 = 1$, all roots of $b(L)$ are outside the unit circle, and $\Sigma_{j=1}^{\infty} j\,|b_j| < \infty$.

Let ω be 2π times the spectral density of $u_{1,t}$ at frequency 0, or $\omega = \sigma^2 b^2(1)$. Also define $\lambda_j = E(u_{1,0}u_{1,j})$, and $\omega_1 = \sum_{j=1}^{\infty} \lambda_j$. Let $W(s)$ be a standard Brownian motion, and $J_c(s)$, a Ornstein-Uhlenbeck process, defined as $dJ_c(s) = cJ_c(s)ds + dW(s),\ J_c(0) = 0$. The demeaned Ornstein-Uhlenbeck process is $J_c^{\mu}(s) = J_c(s) - \int J_c(\tau)d\tau$, where all integrals are from 0 to 1 unless denoted otherwise. Also, let $\Rightarrow$ denote weak convergence on $D\,[0,1]$, and $\equiv$, equality in distribution. All proofs are relegated to the appendix.

Theorem 1 *If $\hat{\beta}$ and $\hat{\psi}$ are the least squares estimators of the parameters β and ψ in equations $(3-4)$, then, under assumption A and $c \neq 0$,*

1. $\hat{\beta} \Rightarrow \zeta_{\,\pi}(c)$
2. $\hat{\psi} \Rightarrow \xi_{\,\pi}(c)$
3. $t_{\hat{\beta}} = \frac{\hat{\beta}-1}{s_{\hat{\beta}}} = O_p(1)$ *and* $T^{-1/2}t_{\hat{\psi}} = \frac{T^{-1/2}(\hat{\psi}-1)}{s_{\hat{\psi}}} = O_p(1)$
   where $\zeta_{\,\pi}(c) \equiv c\kappa_1(c,\pi) + \kappa_1(c,\pi)\left\{\left[\int (J_c^{\mu})^2\right]^{-1}\left[\int J_c^{\mu}dW + \omega_1/\omega\right]\right\}$ *and* $\xi_{\,\pi}(c) \equiv \kappa_2(c,\pi)g_{\,\pi}(c) - \pi\kappa_2(c,\pi)$, $g_{\,\pi}(c) = \{\int_0^1 J_c(s)\int_s^{\pi+s} J_c(\tau)d\tau ds$ $-\left[\int_0^1 J_c(s)ds\right]\left[\int_0^1\int_s^{s+\pi} J_c(\tau)d\tau ds\right]\}/\left\{\int (J_c^{\mu})^2\right\}$, $\kappa_1(c,\pi)$ *and* $\kappa_2(c,\pi)$ *are non-stochastic functions defined as* $\kappa_1(c,\pi) = \frac{(e^{c\pi}-1)\pi}{e^{c\pi}-1-c\pi}$, $\kappa_2(c,\pi) = \frac{c}{e^{c\pi}-1-c\pi}$. *The double integrals are defined as* $\int_0^1\int_s^{s+\pi} J_c(\tau)d\tau ds = \int_0^{1-\pi}\int_s^{s+\pi} J_c(\tau)d\tau ds + \int_{1-\pi}^1\int_s^1 J_c(\tau)d\tau ds$.

The least squares estimators $\hat{\beta}$ and $\hat{\psi}$ are not consistent, and their distributions depend on c through the Ornstein-Uhlenbeck processes and through the non-stochastic functions. Using similar arguments, it can be shown that

s^2 converges in probability to $\left(\left(e^{c\pi}-1\right)/c\pi\right)^2\lambda_0$ in the first regression and diverges in the second. Furthermore, the usual t-statistic, testing for unity of a parameter, converges to a functional of Ornstein-Uhlenbeck processes in the first regression and diverges in the second. The exact functional forms of the t-statistics yield little insight but can be readily derived from the proofs in the appendix.

Monte Carlo simulations for various $T's$, $c's$ and $\pi's$ show that the above asymptotic distributions approximate closely the finite sample distributions of $\widehat{\beta}$ and $\widehat{\psi}$. More details on the simulations are presented in section 4. The exact density of $\widehat{\beta}$ and the simulated asymptotic distributions (smoothed with a normal kernel) are shown in figures 1a–1c. For $T = 100$ and above, the asymptotic and small-sample distributions are very similar for all $\pi's$ and $c's$. The distributions are extremely sensitive to small changes in c as expected from Theorem 1. Similar results are obtained for $\hat{\psi}$ and thus omitted. To put this in perspective, recall that for ϕ close to one, sample sizes of tens of thousands of observations are needed in order for the usual asymptotically normal approximations to be tenable (Bekaert et al. (1997) for an example). Moreover, the asymptotic distributions in Theorem 1 are extremely easy to simulate (they are nothing but rescaled sums of $y_{1,t}$).

Small deviations from $c = 0$ have a big impact on the distributions above. Note that $\lim_{c\to 0}\kappa_1(c,\pi)$ and $\lim_{c\to 0}\kappa_2(c,\pi)$ are undefined. For the location parameter in $\zeta_\pi(c)$, $\lim_{c\to 0}\left(c\kappa_1(c,\pi)\right) = 2$. To demonstrate the dependence of the distributions above on the nuisance parameter, we simulated them 5,000 times. The 10th and 90th quantiles as well as the mean are plotted in figure 2, for $c = (-10,-3,-1,1,3)$, $T = 500$, $\pi = 0.1$. On both sides of the asymptote $c = 0$, the intervals widen and the entire distribution changes. Interestingly, for values of c=1 or c=-1, which are essentially indistinguishable with unit-root tests, the distributions of $\hat{\beta}$ and

$\hat{\psi}$ are very different.

Using the results above, we can ask the question: Given that (1) and (2) hold, what value of c is likely to yield results similar to those observed in the data[4]? The simulated distributions of $\hat{\beta}$ and $\hat{\psi}$ for different $c's$ and $\pi's$ can provide an answer to this question (section 4). Recalling the empirical results of Campbell and Shiller (1991, tables 1-2), least squares estimates in the vicinity of -4 and above were observed in (3) and in the vicinity of 0.7 in (4). Those estimates are contained within the intervals graphed in figure 2 for a small (positive or negative) c. This might suggest that the expectations hypothesis is not false, but instead, the short yield is very persistent, with c contained in $(-1, 0)$.

3 Multivariate Estimation and Testing

We have argued that if (1) and (2) hold, the least squares estimators in (3) and (4) will have distributions shown in Theorem 1. Furthermore, the distributions above depend crucially on c, and for some values of this parameter close to 0, the results obtained from the yield data do not necessarily contradict the theory.

The natural question to ask is: Can we estimate the nuisance parameter c consistently? In general, the answer is no[5]. In the univariate case, the best we can do is find a median unbiased estimate of c, by inverting a statistic (Andrews(1993), Dufour (1990), Stock (1991)). In section 4, we use Stock's (1991) method of inverting the Augmented Dickey Fuller (ADF) test and obtain centered confidence intervals and a median unbiased estimate of c.

[4] See Campbell and Shiller (1991) and Campbell et al. (1997).

[5] To see why consider regression (1) and let $\hat{\phi}$ be the OLS estimator of ϕ. Then $T(\widehat{\phi} - \phi) \Rightarrow \left[\int (J_c^{\mu}(s))^2 \, ds\right]^{-1} \left[\int J_c^{\mu}(s) dW(s) + \omega_1/\omega\right]$ and since $\widehat{c} = T(\hat{\phi} - 1)$, we have $(\widehat{c} - c) \Rightarrow \left[\int (J_c^{\mu}(s))^2 \, ds\right]^{-1} \left[\int J_c^{\mu}(s) dW(s) + \omega_1/\omega\right]$.

If the REHTS holds, a consistent estimator of c can be constructed as shown below. First, the term structure is cast into a triangular representation, where the cointegrating vectors are a function of c. The feasible estimators of the cointegrating vectors are consistent but asymptotically biased. Using delta-method arguments, a consistent (but asymptotically biased) estimator of c is constructed. A parametric and a non-parametric methods for eliminating the bias are proposed. The implications of the REHTS in the triangular representation and the restrictions on c can be tested with a Wald test.

3.1 Triangular Representation of the Term Structure

It is useful to think of the term structure as a sequence of yields $\{y_{n,t}\}_{n\in\mathbb{N}}$ at date t. However, as a practical matter, we often have yields of only a selected number of maturities. Therefore, it would be convenient to define the subsequence $\left\{y_{n_j,t}\right\}_{j\in\mathbb{A}}$, for $\mathbb{A}=\{1,2,...,q\}$ and q fixed and finite, to be a selection of some of the yields $y_{n,t}$ taken in order and with maturity less than the sample size, or $n_j<T$. To simplify notation, we will refer to the subsequence of available yields $y_{n_j,t}$ as $y_{j,t}$, keeping in mind that their corresponding maturities are n_j, and $j=1,2,...q$.

For a sample of size T and a given yield $y_{j,t}$, let π_j be such that $n_j=[\pi_jT]$, $\pi_j>0$, as above. From equations $(1-2)$, $y_{j,t}=\alpha_j+\left(\frac{1}{[\pi_jT]}\sum_{i=0}^{n_j-1}\phi^i\right)y_{1,t}+E_t\left\{\frac{1}{n_j}\sum_{i=0}^{n_j-1}\sum_{k=1}^{i}\phi^{i-k}u_{1,t+k}\right\}\approx\alpha_j+\frac{e^{c\pi_j}-1}{c\pi_j}y_{1,t}+E_t\left\{\frac{1}{n_j}\sum_{i=0}^{n_j-1}\sum_{k=1}^{i}\phi^{i-k}u_{1,t+k}\right\}$[6]. If the expectations hypothesis holds, $y_{j,t}$ and $\frac{e^{c\pi_j}-1}{c\pi_j}y_{1t}$ must differ only because of $u_{j,t}$, or

$$y_{j,t}=\alpha_j+\widetilde{\gamma}_j(c)y_{1,t}+u_{j,t} \tag{5}$$

[6] Strictly speaking, $y_{j,t}=\alpha_j+\left(\frac{1}{[\pi_jT]}\sum_{i=0}^{n_j}\phi^i\right)y_{1,t}+E_t\{\frac{1}{n_j}\sum_{i=0}^{n_j-1}\sum_{k=1}^{i}\phi^{i-k}u_{1,t+k}\}-\frac{\phi^j}{[\pi_jT]}y_{1,t}$, but since all expressions involving the last term are $o_p(1)$, we omit it from this point on, for clarity of exposition.

where $u_{j,t} = E_t\left\{\frac{1}{n_j}\sum_{i=0}^{n_j-1}\sum_{k=1}^{i}\phi^{i-k}u_{1,t+k}\right\}$ is an I(0) process and $\widetilde{\gamma}_j(c) = \frac{e^{c\pi_j}-1}{c\pi_j}$. Thus, $y_{j,t}$ and y_{1t} are cointegrated[7] and the cointegrating vector is a function of c. However, $\widetilde{\gamma}_j(c)$ is not defined at $c = 0$, but $\lim_{c\to 0}\widetilde{\gamma}_j(c) = 1$, and we can define the continuous function

$$\gamma_j(c) = \begin{cases} 1 & , c = 0 \\ \widetilde{\gamma}_j(c) & , \text{otherwise} \end{cases}$$

The function $\widetilde{\gamma}_j(c)$ is differentiable everywhere except at $c = 0$. Define $\frac{d\gamma_j(c)}{dc}|_{c=0} = \lim_{c\to 0}\frac{d\widetilde{\gamma}_j(c)}{dc}$. Hence $\gamma_j(c)$ is continuous and differentiable everywhere. Furthermore $\gamma_j(c)$ is a strictly increasing function of c and its inverse exists and is continuous. To clearly understand the relationship between c, π_j and $\gamma_j(c)$, notice that $\gamma_j(c) \approx 1 + \frac{c\pi_j}{2}$, or $c \approx 2(\gamma_j(c) - 1)/\pi_j$. For $c = 0$, we have the usual result that the long and short rates are cointegrated, with cointegrating vector $(1\ -1)$ (Campbell and Shiller (1987)).

It is important to emphasize that the parameterization $n_j = [\pi_j T]$ was only used to motivate the arguments leading to the cointegrating vectors. However, π_j is fixed. The reasoning is justified using the same arguments as in section 2: for arbitrarily large T, we can always find π_j such that $n_j = [\pi_j T]$. In other words, the finite sample arguments are used to show that, if the REHTS holds, there should be a cointegrating relationship between the short and long rates. The asymptotics ($T \uparrow \infty$) will be used to derive the distributions of the estimators of the cointegrating vectors and of the nuisance parameter c, assuming π_j is fixed and known. The large sample distributions can be interpreted as approximations for the finite sample distributions.

The following structure is assumed for u_t:

Assumption B Let $\{u_t\}_0^\infty$ be a $q \times 1$ vector of random variables that satisfy: $u_t = B(L)\varepsilon_t$, where $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}..., \varepsilon_{q,t})'$ is a martingale differ-

[7] Here, the term "cointegrated" is used loosely, since $y_{1,t}$ is local-to-unity.

ence sequence with $E(\varepsilon_t \varepsilon_t^{'}) = \Sigma$ and finite fourth moments, $B(L) = \sum_{i=0}^{r} B_i L^i$, $B_0 = I_q$, all roots of $B(L)$ are outside the unit circle, and $\Sigma_{i=1}^{\infty} i\,|B_i| < \infty$.

Partition u_t and ε_t after their first elements as $u_t = (u_{1,t}\ \overline{u}_{2,t})$ and $\varepsilon_t = (\varepsilon_{1,t},\ \overline{\varepsilon}_{2,t})$ and partition $B(L)$ and Σ appropriately. Let Ω be 2π times the spectral density of u_t at frequency 0, or $\Omega = B(1)\Sigma B(1)'$ and partition Ω conformably with $(u_{1,t}\ \overline{u}_{2,t})$. Also define $\Lambda_i = E(u_{1,0}u_{1,i})$,and $\widetilde{\Omega} = \sum_{i=1}^{\infty} \Lambda_i$.

If the REHTS holds, the term structure must be generated by a triangular system (Phillips (1991)). Stack the equations in (5) for $2 \leq j \leq q$ and let $Y_{2,t}^0 = (y_{2,t}, y_{3,t}, ..., y_{q,t})'$, $\alpha = (\alpha_2, \alpha_3, ..., \alpha_q)'$, $\underline{c} = (c_2, c_3, ..., c_q)$ $\Gamma(\underline{c}) = \left(\gamma_2(c_2), \gamma_3(c_3), ..., \gamma_q(c_q)\right)'$ be $(q-1)$ dimensional vectors and e is a $(q-1)$ vector of $1's$. The triangular representation of the term structure is:

$$y_{1,t} = \phi y_{1,t-1} + u_{1,t} \tag{6a}$$

$$Y_{2,t}^0 = \alpha + \Gamma(\underline{c}_0) y_{1,t} + \overline{u}_{2,t} \tag{6b}$$

where $\underline{c}_0 = c * e$. It must be emphasized that, if the REHTS holds, then $c_2 = c_3 = ... = c_q = c$.

Campbell and Shiller (1987) also use a triangular representation to study the term structure and they make an assumption similar to Assumption B[8]. However, the triangular representation above differs from Campbell and Shiller's in two important ways.

First, we do not constrain $y_{1,t}$ to be unit root, although the unit root case is nested in $(6a)$. Elliott (1994, 1998) demonstrates that if $y_{1,t}$ is local to unity $(c \neq 0)$, the estimators of $\Gamma(\underline{c})$, constructed under the assumption that $c = 0$, are consistent but have a bias, resulting in size distortions of the

[8] In Campbell and Shiller (1987), u_t is a finite order VAR.

usual t and Wald tests involving $y_{1,t}$. The distortions can be quite severe if the covariance between $u_{1,t}$ and $\overline{u}_{2,t}$ is high[9]. In the term structure setup, it is reasonable to suspect that $u_{1,t}$ and $\overline{u}_{2,t}$ are very highly correlated.

From a methodological viewpoint, our use of the triangular representation also differs from Campbell and Shiller's. They assume that $c = 0$ (and known) and perform various volatility tests. Here, we want to find a consistent estimator of $\underline{c}$ and test whether all its elements are equal, since this is a direct implication of the REHTS. Keeping in mind the results from Part 2, we also want to know if c is statistically different from 0.

3.2 Consistent and Asymptotically Unbiased Estimators of $\Gamma(\underline{c})$, $\underline{c}$ and c

Numerous estimators have been proposed for estimating $\Gamma(\underline{c})$. For a review of the literature, see Watson (1994). Although the least squares estimator of $\Gamma(\underline{c})$ in ($6b$) is consistent, it is also asymptotically biased (Stock(1987)). There are several ways to deal with this bias, and here we follow the approach suggested by Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993). The idea is to make the errors in ($6a$) independent of the errors in ($6b$) and apply the results in Park and Phillips (1988) and Sims et al. (1990). Following Stock and Watson (1993), let $Proj(\overline{u}_{2,t}|\{u_{1,t}\}_{-\infty}^{\infty}) = Proj\left(\overline{u}_{2,t}|\{(1-\phi L)y_{1,t}\}_{-\infty}^{\infty}\right) = D(L)\,(1-\phi L)\,y_{1,t}$, where $Proj(r|\{s_1...s_k\})$ is the linear projection of r onto $\{s_1...s_k\}$ and $D(L)$ is a two sided polynomial. We assume that $D(L)$ is a polynomial of finite leads and lags of equal length, $D(L) = \sum_{i=-k}^{k} d_i L^i$.

Assumption C [10]Assume $D(L) = \sum_{i=-k}^{k} d_i L^i$ where k is a finite and

[9]Elliot (1994) shows that the size of the Wald test approaches 1 as the long-run covariance increases, provided $c \neq 0$.

[10]If an infinite number of leads and lags is necessary in the projection, but we use a

known integer.

Let $v_t = \overline{u}_{2,t} - \sum_{i=-k}^{k} d_i \widetilde{\triangle} y_{1,t-i}$, where $\widetilde{\triangle} = (1-\phi L)$ is the quasi-differencing operator. We can augment $(6b)$ as $Y_{2,t}^0 = \alpha + \Gamma(\underline{c}) y_{1,t} + \sum_{i=-k}^{k} d_i \widetilde{\triangle} y_{1,t-i} + v_t$. Note that v_t is independent of all right hand side variables. The generalization of the dynamic OLS (DOLS) estimator of Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993) in the local-to-unity case is:

$$Y_{2,t} = \Gamma(\underline{c}) y_{1,t} + \sum_{i=-k}^{k} d_i \widetilde{\triangle} y_{1,t-i} + v_t \tag{7}$$

where d_i is an $(q-1)$ dimensional vector of coefficients and $Y_{2,t} = Y_{2,t}^0 - \frac{1}{T}\sum_t Y_{2,t}^0$. If c is known, estimating (7) by least squares yields a T-consistent estimator of $\Gamma(\underline{c})$ with an asymptotic distribution, which is a mixture of normals. This result, shown in Elliott (1994) in the bivariate case, is a straight forward extension of Stock and Watson (1993) and is provided here for completeness.

Theorem 2 *Suppose the data is generated by the process* $(6a - 6b)$ *and assumptions B and C hold. Let* $T_{q-1} = T * I_{q-1}$, I_{q-1} *is the identity matrix, and* $\Gamma(\underline{c}_0)$ *is the true value of the parameters. Then, the least squares estimator of* $\Gamma(\underline{c})$ *in* (7) *has the following asymptotic distribution*

$$T\left(\widehat{\Gamma(\underline{c})} - \Gamma(\underline{c}_0)\right) \Rightarrow \Omega_{11}^{-1/2}\Omega_{2.1}^{1/2}\int J_c^\mu dW_{2.1}\left(\int (J_c^\mu)^2\right)^{-1}$$

where $T^{-1}\sum_{t=1}^{[sT]} v_t \Rightarrow \Omega_{2.1}^{1/2} W_{2.1}(s)$ *and* $\Omega_{2.1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$.

Not surprisingly, the results are a straight forward generalization of Stock and Watson (1993), with the Brownian Motion functionals replaced by the diffusion process $J_c{}^\mu$.

truncated polynomial instead, the error from the truncation vanishes asymptotically if the number of included leads and lags increases at T^r, $0 < r < 1/3$ (Saikkonen (1991)).

The distribution of $\widehat{\Gamma(\underline{c})}$ depends on c. However, c is precisely the unknown parameter, and the estimator above is unfeasible. One might be tempted to assume that $\underline{c} = 0 * e$ (i.e. $\widetilde{\triangle} = \triangle = (1 - L)$) and use the usual DOLS

$$Y_{2,t} = \Gamma(\underline{c})y_{1,t} + \sum_{i=-k}^{k} d_i \triangle y_{1,t-i} + v_t \tag{8}$$

Elliott (1994, 1998) shows that if $y_{1,t}$ follows $(6a)$, the least squares estimator of $\Gamma(\underline{c})$ in (8), call it $\widehat{\widehat{\Gamma(\underline{c})}}$, is T-consistent, but asymptotically biased, with bias $B = -\Omega_{11}^{-1}\Omega_{21}c$, and a Wald test, involving parameters of $y_{1,t}$ will have incorrect size.

In our case, since $\Gamma(\underline{c})$ is a consistently estimable function of $\underline{c}$, delta-method arguments are used to find a consistent estimator of $\underline{c}$, $\widehat{\widehat{\underline{c}}}$.

Theorem 3 *If $\widehat{\widehat{\Gamma(\underline{c})}}$ is the least squares estimator of $\Gamma(\underline{c})$ in (8), $G(.)$: $R^{q-1} \to R^{q-1}$ is such that $G(\Gamma(\underline{c})) = \underline{c}$ with Jacobian $D(\underline{c}) = \frac{\partial G}{\partial \Gamma'}|_{\Gamma(\underline{c})}$ then,*

$$T\left(\widehat{\widehat{\underline{c}}} - \underline{c}_0\right) \Rightarrow \left(D_0^{-1}\Omega_{11}\int (J_c^{\mu})^2 D_0^{-1}\right)^{-1} D_0^{-1}\left(\Omega_{11}^{1/2}\Omega_{2.1}^{1/2}\int J_c^{\mu} dW_{2.1}\right) - \Omega_{11}^{-1}D_0\Omega_{21}c$$

where $G\left(\widehat{\widehat{\Gamma(\underline{c})}}\right) = \widehat{\widehat{\underline{c}}}$, $G(\Gamma(\underline{c}_0)) = \underline{c}_0$, $D(\underline{c}_0) = D_0$.

Moreover, using the results from Park and Phillips (1988), we can make the following observation:

Remark 4 *Let $\widehat{V} = \hat{\Omega}_{2.1}\left(T^{-2}\sum_{t=1}^{T} y_{1,t}\right)^{-1}$, where $\hat{\Omega}_{2.1} \xrightarrow{p} \Omega_{2.1}$ and $\hat{D} = D(\widehat{\widehat{\underline{c}}})$. Then,*

$$\left(\hat{D}\widehat{V}\hat{D}\right)^{-1/2}\left[T\left(\widehat{\widehat{\underline{c}}} - \underline{c}_0\right) + \Omega_{11}^{-1}D\Omega_{21}c\right] \overset{a}{\sim} N(0, I_{n-1})$$

Two issues need to be addressed. First, any of the $q-1$ elements of the vector $\widehat{\widehat{\underline{c}}}$ is a consistent estimator of the scalar c, but which one should we choose? A natural way is to take a convex combination of the elements of $\widehat{\widehat{\underline{c}}}$ such that its variance is as small as possible. More precisely,

Corollary 5 *If $\hat{D}$ and $\widehat{V}$ are as above and $\widetilde{c} = \hat{a}'\widehat{\underline{c}}$, where the $q-1$ vector $\hat{a} = [\hat{a}_1, ..., \hat{a}_{q-1}]'$ is a solution of the minimization problem* $\min_a a' \left\{ \hat{D}\widehat{V}\hat{D} \right\} a$, *s.t.* $0 \leq a_i \leq 1$ *and* $\sum_{i=1}^{q-1} a_i = 1$, *then*

$$T(\widetilde{c} - c) \Rightarrow \hat{a}' \left(D_0^{-1}\Omega_{11} \int (J_c^\mu)^2 D_0^{-1} \right)^{-1} D_0^{-1} \left(\Omega_{11}^{1/2}\Omega_{2.1}^{1/2} \int J_c^\mu dW_{2.1} \right) - \Omega_{11}^{-1}\hat{a}' D_0 \Omega_{21} c$$

Second, the estimators $\widehat{\underline{c}}$ and $\widetilde{c}$ are consistent but asymptotically biased. We propose a parametric and a non-parametric methods to eliminate the bias.

Since c and Ω can be estimated consistently (the latter follows from the consistency of the least squares estimators of ϕ and $\Gamma(\underline{c})$ in 6a–6b), then a consistent estimator of the bias is $\widehat{B} = -\widehat{\Omega}_{11}^{-1}\widehat{\Omega}_{21}\widetilde{c}$. Correcting $\widehat{\Gamma(\underline{c})}$ for the estimated bias produces an asymptotically unbiased estimator of $\Gamma(\underline{c})$. The delta method is used to obtain an asymptotically unbiased estimator of $\underline{c}$, that we call the nonparametric quasi DOLS (NQDOLS):

Theorem 6 *Suppose* $\widetilde{c} \xrightarrow{p} c$ *and* $\widehat{\Omega} \xrightarrow{p} \Omega$. *Let* $\widehat{B} = -\widehat{\Omega}_{11}^{-1}\widehat{\Omega}_{21}\widetilde{c}$, *and* $\widehat{\Gamma}^1(\underline{c}) = \widehat{\Gamma(\underline{c})} - \frac{\widehat{B}}{T}$. *Then, using the notation of Theorem 3,*

$$T\left(\widehat{\Gamma}^1(\underline{c}) - \Gamma(\underline{c}_0)\right) \Rightarrow \left(\Omega_{11} \int (J_c^\mu)^2\right)^{-1} \left(\Omega_{11}^{1/2}\Omega_{2.1}^{1/2} \int J_c^\mu dW_{2.1}\right)$$

and the NQDOLS estimator of $\underline{c}$ is $\widehat{\underline{c}}^1 = G\left(\widehat{\Gamma}^1(\underline{c})\right)$ and:

$$T\left(\widehat{\underline{c}}^1 - \underline{c}_0\right) \Rightarrow \left(D_0^{-1}\Omega_{11} \int (J_c^\mu)^2 D_0^{-1}\right)^{-1} D_0^{-1} \left(\Omega_{11}^{1/2}\Omega_{2.1}^{1/2} \int J_c^\mu dW_{2.1}\right)$$

Another way of correcting for the bias is to use $\widetilde{c}$ to define the estimated quasi-differencing operator $\overset{\wedge}{\triangle} = (1 - (1 + \frac{\widetilde{c}}{T})L)$. The least squares estimator of $\Gamma(\underline{c})$ in the regressions

$$Y_{2,t} = \Gamma(c)y_{1,t} + \sum_{i=-k}^{k} \overset{\wedge}{\triangle} d_i y_{1,t-i} + v_t \tag{9}$$

will be asymptotically unbiased. We can consequently derive a parametric quasi DOLS estimator of $\underline{c}_0$, called the QDOLS.

Theorem 7 *Let* $\widehat{\Gamma}^2(c)$ *be the least squares estimator of* $\Gamma(c)$ *in* (9) *and* $\widetilde{c} \xrightarrow{p} c$*. Then, using the notation of Theorem 3,*

$$T\left(\widehat{\Gamma}^2(\underline{c}) - \Gamma(\underline{c}_0)\right) \Rightarrow \left(\Omega_{11} \int (J_c^\mu)^2\right)^{-1} \left(\Omega_{11}^{1/2}\Omega_{2.1}^{1/2} \int J_c^\mu dW_{2.1}\right)$$

and the QDOLS estimator of $\underline{c}$ *is* $\widehat{\underline{c}}^2 = G\left(\widehat{\Gamma}^2(\underline{c})\right)$ *and:*

$$T\left(\widehat{\underline{c}}^2 - \underline{c}_0\right) \Rightarrow \left(D_0^{-1}\Omega_{11} \int (J_c^\mu)^2 D_0^{-1}\right)^{-1} D_0^{-1} \left(\Omega_{11}^{1/2}\Omega_{2.1}^{1/2} \int J_c^\mu dW_{2.1}\right)$$

It is possible to iterate on the previous procedures, but simulations suggest that the benefit is small to negligible. Finally, we can use the minimization in Corollary 5 to find an asymptotically unbiased estimator of c.

3.3 Testing $R\underline{c} = r$

Given the results in the previous subsection, the asymptotic distribution of the Wald test for the null hypothesis $R\underline{c} = r$, where $rank(R) = s$, can be obtained using the results from Park and Phillips (1988).

Corollary 8 *For the asymptotically unbiased estimators* $\widehat{c}^1$ *and* $\widehat{c}^2$ *above,* $\widehat{\Omega}_{2.1} \xrightarrow{p} \Omega_{2.1}$ *and under assumptions B and C, define*

$$W = (R\widehat{\underline{c}} - r)' \left[RD\left(\widehat{\underline{c}}\right)\left(\widehat{\Omega}_{2.1}\left\{\sum (y_{1,t})^2\right\}^{-1}\right) D\left(\widehat{\underline{c}}\right) R'\right]^{-1} (R\widehat{\underline{c}} - r)$$

Under the null,

$$W \Rightarrow \chi_s^2$$

As mentioned above, the REHTS implies that $c_2 = c_3 = ... = c_q = c$. A rejection of the implication of the REHTS would imply a rejection of the theory itself.

4 Monte Carlo Simulations

4.1 Single-Regression Tests

We investigate whether the asymptotic distributions derived in section 2 adequately approximate the finite sample distributions of $\hat{\beta}$ and $\hat{\psi}$. For tractability, the system $(1-2)$ is simulated for two yields, $y_{1,t}$ and $y_{n,t}$. For any value of the triplet (c,π,T), the data is simulated $5,000$ times as in $(1-2)$, $y_{1,0}=0$, and different specifications of $u_{1,t}$.

The results for $c=(-10,-5,-1,1,3)$, $\pi=(0.05,0.1,0.25)$, $T=(100,500)$ and $u_{1,t}\sim NIID(0,1)$ are presented in tables 1 and 2. Various AR and MA processes for $u_{1,t}$ were also simulated, but the results were similar and hence omitted. Tables 1a and 2a present percentiles, mean and standard deviation of the slope coefficients in $(3-4)$ estimated by least squares, for various values of (c,π,T). The functionals of stochastic integrals, $\zeta_{\pi}(c)$ and $\xi_{\pi}(c)$, are simulated using scaled partial sums of $y_{1,t}$, with $y_{1,0}=0$. Their percentiles, mean and standard deviation are in tables 1b and 2b, respectively, for various (c,π,T). Comparing tables 1a with 1b, and 2a with 2b, we see that the exact and the asymptotic distributions are very close. To illustrate this point even further, figure 1 shows the finite and asymptotic distributions of $\hat{\beta}$. They are very similar, even for T=100. In regressions $(3-4)$ the local-to-unity asymptotics provide a far better approximation to the distributions of $\hat{\beta}$ and $\hat{\psi}$ than does the normal distribution. Bekaert et al. (1997) conduct a similar experiment, with $\phi=0.986$ and T=524, which corresponds roughly to our case $c=-10$, $T=500$. Not surprisingly, the results here coincide with theirs (up to a simulation error and smoothing). However, the local-to-unity parameterization used in this paper, offers a way of understanding how small changes in the data generating process can result in big changes in the distribution of the coefficients.

It is evident from figures 1a–1c and tables 1–2 that the distributions of $\hat{\beta}_n$ and $\hat{\psi}_n$ change dramatically for small changes in c. Different values of π have very little impact, confirming the results from Theorem 1. Moreover, for values of c in the range of $(-1, 0)$, the estimates $\hat{\beta}_n$ and $\hat{\psi}_n$ might be in the range observed in the yield data. Therefore, if we are willing to accept the possibility that the short yield is highly persistent (in finite samples), the results presented by Shiller et al. (1983), Shiller (1990), Campbell and Shiller (1991), Campbell et al. (1997) are not necessarily in conflict with the REHTS.

4.2 Median Unbiased Simulations

Following Stock (1991), a median unbiased estimator of c might be obtained. Since $(1-\phi L)y_{1t} = u_{1t}$ and $(1-\phi L)Y^0_{2,t} = (1-\phi L)\alpha + (1-\phi L)\Gamma(c,\tau)y_{1y} + (1-\phi L)\overline{u}_{2,t}$, we can rewrite

$$Y^0_{2,t} = \widetilde{\alpha} + \phi Y^0_{2,t-1} + \overline{v}_t \tag{10}$$

where $\overline{v}_t = \Gamma(c,\tau)u_{1,t} + (1-\phi L)\overline{u}_{2,t}$ and $\overline{v}_t = (v_{2,t}, v_{3,t}, ..., v_{q,t})$, $v_{j,t} = \gamma(c_j, \tau_j)u_{1,t} + (1-\phi L)u_{j,t}$. Taking the jth equation in (10) and inverting the augmented Dickey Fuller (ADF) test produces a median unbiased estimator, $\widehat{c}^{MU}_j$. Note that the error terms $v_{j,t}$ are autocorrelated. The autocorrelated structure depends on the cointegrating vector and might be fairly persistent. As shown recently, the size and power of the augmented Dickey Fuller (ADF) test depends on the number of augmenting lags included in the test and this in turn impacts the precision of our confidence intervals (Ng and Perron (1998)). As discussed in Ng and Perron (1996), there is no satisfactory way of finding the appropriate lag structure. Information-based methods (AIC, Schwartz) under-parameterize the test and hence, do not entirely correct for serial correlation. The real size of the test is bigger than the nominal one. On the other hand, a t-test for the significance of the last lag tends to over-parameterize the test, resulting in loss of power. In this particular application we are more concerned with size distortions[11] and choose the lag structure with sequential t-tests. In the notation above we have allowed the $c's$ to vary from equation to equation, but according to the Expectations Hypothesis, c should not change with j. This is a testable implication of the REHTS.

Using simulated data for various specifications of $\pi's$, $c's$ and error specification, the ADF test is inverted using the tables published in Stock (1991), to find 95%, 90%, 80%, 70% centered confidence intervals as well as a median unbiased estimate of c. The real coverage of the intervals is verified by reporting the fraction of times that the calculated confidence interval contains the true value of c, for different specifications. The results are reported in Table 8.

The case $\pi = 0$ corresponds to Stock's simulations, and indeed our results concur. For $\pi > 0$, the coverage is still acceptable even for high $\pi's$. The missing (c, π) specifications in table 8 could not be computed, because the values of the tests were often outside the range of the tables in Stock (1991). If more than one percent of the tests could not be inverted, the specification was eliminated altogether. Since the Monte Carlo simulations yield satisfactory results, we will use the median unbiased estimator as a staring point in the analysis.

[11] This choice is made on the basis that in the term structure example, we have a fair amount of observations and power will be satisfactory. Furthermore, obtaining precise confidence intervals of c relies on the test being correctly sized.

4.3 Multivariate Tests

In a Monte Carlo experiment, we simulate the bivariate series $(y_{1,t}, y_{n,t})$ as in $(1-2)$ with $\pi = (0.05, 0.1, 0.2, 0.3)$, $c = (-5, -3, -1, 1, 3)$, $T = (200, 400)$, and $y_{1,0} = 0$.[12] For each experiment (c, π, T), u_t follows one of two processes, specified in the tables below, and $\varepsilon_t \sim NIID(0, \Sigma)$, drawn from the Matlab pseudo-random number generator "randn". Each simulation is repeated $5,000$ times. The mean from the OLS, DOLS, QDOLS, iterated QDOLS, and nonparametric QDOLS estimators of c is tabulated for each (c, π, T). OLS is the original estimator of the cointegrating vector, proposed by Engle and Granger (1986). The other estimators were defined in section 3. The augmenting leads and lags for each procedures are chosen with sequential t-tests.

Table 3a shows the results for $u_t = \varepsilon_t$, $\varepsilon_t \sim NIID(0, \Sigma)$. The matrix Σ is specified in the notes under the table. As expected, the OLS estimates are very biased even for $T = 400$. The DOLS is also biased, but considerably less than the OLS. The QDOLS and the nonparametric QDOLS are almost unbiased. As we can see, the iterated QDOLS does not necessarily perform better than the QDOLS.

Similar results for $u_t = Au_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, \Sigma)$ are shown in table 3b. The matrices A and Σ are specified in the notes the tables. Admittedly, the long-run variance of u_t was chosen to yield OLS estimates with extreme biases. The bias in the DOLS is smaller, but still sizeable even for $T = 400$. The QDOLS estimators (parametric and nonparametric) are less biased. The difference in biases between the DOLS and QDOLS estimators is best seen for $T = 400$.

The empirical distributions (smoothed with a normal kernel) of the QDOLS estimator of c for the case $u_t = \varepsilon_t$, are plotted in figures 3a-3e

[12] We abstract from the non-zero initial condition issues.

for various (c, π, T). The distributions are standardized by subtracting the mean and dividing by the standard deviations. As we can see, the estimator is purged from the bias. From remark 1 in section 3, we know that the standardized distributions should approximate the standard normal distribution. For comparison, the standard normal density is also plotted, and we notice that the approximation is quite good for all values of c. By normalizing the QDOLS distributions, we are not able to observe the rate of convergence. Table 4 intends to demonstrate the convergence at rate T of the estimators. The entries in the table are the ratios of the variances for $T = 200$ versus $T = 400$ of the various estimators for a given experiment (c, π). If the estimators converge at rate T, the entries should be close to 4. This is indeed the case.

Figures 3a-3e suggest that the standardized distributions of the QDOLS estimators is well approximated by the normal distribution. But in order to conduct inference, we must know if the confidence intervals, suggested by the standard normal distribution have an adequate real coverage. The real coverage of the, say, 95% confidence interval, is verified by reporting the fraction of times that the true value of c is contained within 1.96 standard deviations from the calculated mean. The results for the parametric and nonparametric QDOLS are reported in tables 5a and 5b, for various specifications of (c, π, T) and error terms. As we can observe, the nominal and real coverage are remarkably similar even for positive values of c.

One last useful fact is worth mentioning. The variance of the QDOLS estimators depends on π only through the Jacobian, $D(c)$. For bivariate series $(y_{1,t}, y_{n,t})$, $D(c)$ is a scalar. To analyze the effect of π on the variance, $D^2(c)$ is plotted as a function of c and π in figure 4. For π close to zero, the variance increases exponentially, for all c because in the bivariate case $D^2(c) \approx \frac{4}{\pi^2}$. In other words, rates at the short end of the term structure will have a much higher variance than rates at the long end. The same argument is re-enforced in table 6. The ratio of variances of the QDOLS estimators for various $\pi's$ is computed from the Monte Carlo experiments and compared the theoretical ratio. Once again, the variance decreases dramatically as π increases.

5 Estimations and Testing of the REHTS

Monthly data of continuously compounded yields to maturity of US Government securities from McCulloch and Kwon (1993) is used in this section. Bills, notes and bonds with maturities from 1 month to 13 years, spanning the period 1946:12–1991:2 are available, but pre-1952:1 numbers are discarded so that no calculations (including lags) use data prior to the Treasury Accord of 1951. This dataset has been used by Campbell and Shiller (1991) and Campbell et al. (1997) among others to study the term structure of zero-coupon bonds. We take $y_{1,t}$ to be the one-month yield, and maturities of 2, 3,...,18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, 120, 144, 156 months are used in $Y_{2,t}$. The sample size varies from the lag and lead configurations of the tests, but is never below 460 observations.

5.1 Median Unbiased Estimation of c

The median unbiased estimates of c are only used as a starting point in the investigation. The ADF test for each equation in (10) is inverted, using the number of lags chosen by sequential t-tests.

The median unbiased estimate for c from $y_{1,t}$ is -5.99 and the 90% centered confidence interval is $(-14.84, 2.97)$. As the maturity increases, so do the median unbiased estimates. The results from the estimation are plotted in figure 5a. In the high-end of the term structure, the estimates are positive. However, the confidence intervals are too wide for a conclusion to be drawn. For example, the range $(-14.84, 2.97)$ covers processes with an autoregressive root anywhere from 0.97 to 1.01. A few alternative lag specifications are plotted in figure 5b as a way of verifying the robustness of the results to the lag structure of the test.

A positive median unbiased estimate of c in the yield data should not come as a complete surprise. Using a more aggregated dataset, Stock (1991) finds that among several macroeconomic series, only the bond yield has a 90% confidence interval above unity. Based on all this evidence, the possibility that c is in the neighborhood of 1 cannot be ruled out. But neither can the possibility that c is, say, -3.

5.2 Consistent and Asymptotically Unbiased Estimators of $\underline{c}$ and c

The methods developed in section 3 and tested in section 4 are now applied to the US yield data. Since the parametric and non-parametric estimator performed similarly in the Monte Carlo experiments, we use the parametric one (QDOLS) only. The lead-lag specification is chosen with sequential t-tests. The long-run variance of u_t is estimated with an autoregressive spectral estimator, whose truncation is also determined by sequential t-tests.

We estimate three separate systems. In the first one, $Y_{2,t}$ is comprised of yields with one or more years to maturity. The second system includes yields with 18 months and over, and the third one, 3 years and over. In all three systems, $y_{1,t}$ is the one-month rate. Yields between two months and one year are not included in the analysis, because the variance of their estimates is too big (π is too small). However, the results do not change considerably if all available yields are included in $Y_{2,t}$.

The sequential t-tests selected the same 3 leads and lags specification in all three specifications. For each system, the following algorithm is implemented. First, we find the consistent but biased estimate of $\underline{c}$. Second, we minimize the quadratic criterion in Corollary 5 to find a consistent but biased estimate of c. Third, a consistent and asymptotically unbiased estimate of $\underline{c}$ is obtained with the QDOLS. Finally, a consistent and asymptotically unbiased estimate of c is produced. This estimate is -0.37 for the first system, -0.35 for the second one, and -0.35 for the third one. The estimates of $\underline{c}$ are plotted in figure 6a, along with confidence bands at plus and minus two standard deviations. As seen in section 4, if the REHTS holds, those bands provide quite an adequate coverage of the 95% confidence interval.

The last plot in figure 6a represents the unbiased estimates of $\underline{c}$ for the system with maturities of 3 years and more. First, we notice that the 95% error bands are very small, compared to those provided by inverting the ADF test. More interestingly, the estimates at the long end of the term structure seem to be very close to each other, just as implied by the REHTS. Looking at the corresponding plots in figures 6b-6d, the results seem quite robust to various lead/lag specifications. However, the Wald test, suggested in section 3.3, rejects the REHTS at any significance level. The results are reported in table 7. This rejection might come as a surprise considering the very tight confidence intervals around each estimate. However, it is known that GMM-based Wald tests tend to have a small-sample size that exceeds the asymptotic one. Burnside and Eichenbaum (1996) find that the size discrepancy is very severe particularly when multiple restrictions are imposed, and suggest that a big part of the problem is in the estimates of the weighing matrix. Since our regressions can be viewed within the GMM framework, we expect that the same size distortions would be present even more so because we impose many restrictions in all three systems and the covariance matrix is particularly difficult to estimate. A small Monte-Carlo study, not reported in the paper, confirms our suspicions. A more complete study of the small sample properties of Wald tests with I(1) variables needs to be performed.

The first two plots in figure 6a show the estimates of $\underline{c}$ for the first two systems. The implications of the REHTS do not seem to hold at the short end of the term structure. The Wald test rejects the REHTS at any significance level. Various systems with different lead/lag specifications were tested. Some of the computed Wald statistics are reported in table 7. We could not reject the REHTS only for very small systems, with only a few long-end yields. The robustness of the results for different lead and lag specifications are reported in figures 6b-6d; the results change very little with different lag/lead specifications.

In sum, the plots in figures 6a–6d bear little similarity to the ones from figures 5a and 5b. In both sets of figures, the estimates are more tightly estimated at the long end of the term structure, where the estimates are close to zero. But the similarities stop here. The median unbiased estimates at the long end seem to be slightly positive, although the centered confidence intervals are much too wide for a conclusion to be drawn. On the contrary, the confidence bands around the negative QDOLS estimates are very tight. Moreover, at the short end, the DOLS yields positive estimates when the median unbiased estimates are negative (but again the confidence intervals of the latter are very wide). The DOLS estimates of c at the long end of the term structure are very similar, although formal Wald tests reject the REHTS. Estimates at the short end are very different from those at the long end, suggesting that some of the dynamics of the system are not well captured by a single local-to-unity process. There might be gains from using a multi-dimensional driving process with more than one nuisance parameter.

6 Conclusion

The methods presented above can be applied to various other econometric problems. Indeed, it is often assumed that the variables of interest have been reduced to I(0) processes after some linear transformation. If the assumption is untenable (and it probably always is), asymptotic normality might not be an adequate approximation of the finite distributions of the statistics of interest. A link between the distributions of those statistics and the local-to-unity parameter c can be established, using standard local-to-unity asymptotics.

Consistent estimators of c are not available for a general local-to-unity process. In this paper, we exploit the structure of the REHTS to construct consistent and asymptotically unbiased estimators of c. In that sense, our estimators are "structural." Monte Carlo simulations demonstrate the validity of the procedures even in samples of reasonable size. Moreover, when the proposed methods are applied to the US yield data, we obtain very accurate estimates of the nuisance parameter. However, even if the estimates at the long end of the term structure seem to follow the implications of the theory, a formal Wald test rejects the REHTS. The estimates at the short end of the term structure are very different from those at the long end. This result might suggest that the rejections of the expectations hypothesis are a consequence of using unrealistic assumptions about the driving process. One might speculate that a multi-dimensional local-to-unity process is needed to capture the dynamics underlying the term structure, much as in the affine multi-factor general-equilibrium literature of the term structure.

The REHTS is just one example of rational expectations present value models in economics where the data follows a very persistent process. The estimators above can be cast more elegantly into a GMM framework. Obtaining an estimate of c and testing the restrictions of the model can also be carried out more readily in that fashion. Lastly, the driving process $y_{1,t}$ might be a vector instead of a scalar. In that case, c will be a square matrix, thereby increasing the complexity of the problem. All these non-trivial extensions are the focus of current research.

Appendix

Lemma 9 *Under assumption A,*

1. $T^{-1/2}y_{n,t} \Rightarrow e^{c\pi}\omega^{1/2}J_c(s),$ *where* $n = [\pi T]$
2. $T^{-1}\sum_{t=1}^{T} y_{1,t-1}u_{1,t} \Rightarrow \omega \int_0^1 J_c(s)dW(s) + \omega_1$
3. $T^{-1/2}s_{nt} \Rightarrow \omega^{1/2}\left(\frac{e^{c\pi}-1-c\pi}{c\pi}\right) J_c(s)$
4. $T^{-1/2}s_{nt}^{*} \Rightarrow \omega^{1/2}\left\{\frac{1}{\pi}\int_s^{\pi+s} J_c(\tau)d\tau - J_c(s)\right\}$

Proof. Results (1-4) follow by applying local-to-unity asymptotics, such as in Phillips (1987), Lemma 1. Starting with (1), then $y_{1,t+k} = \phi^k y_{1,t} + \sum_{j=1}^{k}\phi^{k-j}u_{1,t+j}$, and since $u_{1,t} = b(L)\varepsilon_t$, $y_{1,t+k} = \phi^k y_{1,t} + \sum_{j=1}^{k}\phi^{k-j}\left(\frac{b(L)}{L^j}\right)\varepsilon_t$. Taking conditional expectations at time t and using $[.]_+$, the annihilation operator to obtain: $E_t\left(y_{1,t+k}\right) = \phi^k y_{1,t} + \sum_{j=1}^{k}\phi^{k-j}\left[\frac{b(L)}{L^j}\right]_+ \varepsilon_t = \phi^k y_{1,t} + \sum_{i=0}^{\infty}\sum_{j=1}^{k}\varepsilon_{t-i}\phi^{k-j}b_{i+j} = \phi^k y_{1,t} + \sum_{i=0}^{\infty}\varepsilon_{t-i}\left(\sum_{j=1}^{k}\phi^{k-j}b_{i+j}\right) = \phi^k y_{1,t} + \sum_{i=0}^{\infty}q_{i,k}\varepsilon_{t-i}$, where $q_{i,k} = \sum_{j=1}^{k}\phi^{k-j}b_{i+j}$. Now we will show that $q_{i,k}$ are absolutely summable. $\sum_{i=0}^{\infty}|q_{i,k}| = \sum_{i=0}^{\infty}\left|\sum_{j=1}^{k}\phi^{k-j}b_{i+j}\right| \leq \sum_{i=0}^{\infty}\sum_{j=1}^{k}\left|\phi^{k-j}\right||b_{i+j}| \leq \sum_{i=0}^{\infty}\sum_{j=1}^{k}e^{|c|}|b_{i+j}|$

$\leq e^{|c|}\sum_{i=0}^{\infty}\sum_{j=1}^{\infty}|b_{i+j}| = e^{|c|}\sum_{i=0}^{\infty} i\,|b_{i+j}| < \infty$. Therefore $\sum_{i=0}^{\infty}q_{i,k}\varepsilon_{t-i} = o_p(T^{1/2})$. Using the parameterization $k = [\pi T]$, $\pi \in [0,1]$, $T^{-1/2}E_t\left(y_{1,t+k}\right) = T^{-1/2}\phi^{[\pi T]}y_{1,t} + o_p(1) \Rightarrow e^{c\pi}\omega^{1/2}J_c(s)$, where $t = [sT]$ and ω is the long run variance of u_{1t}. Since $y_{n,t} = \alpha_n + \frac{1}{n}\sum_{i=0}^{n-1}E_t\left[y_{1,t+i}\right]$, using the above result and $n = [\pi T]$, we obtain: $T^{-1/2}y_{n,t} = \frac{1}{\left[\pi T^{3/2}\right]}\sum_{i=0}^{[\pi T]}E_t\left[y_{1,t+i}\right] + op(1) \Rightarrow \left(\frac{e^{c\pi}-1}{c\pi}\right)\omega^{1/2}J_c(s)$. Parts 2 and part 3 are obtained with similar calculations. For part 4, note that $s_{nt}^{*} = \frac{1}{n}\sum_{i=0}^{n}y_{1,t+i} - y_{1,t} - \frac{1}{n}y_{1,t+n}$, $T^{-1/2}s_{nt}^{*} = \frac{1}{\left[\pi T^{3/2}\right]}\sum_{i=0}^{[\pi T]}y_{1,t+i} - \frac{1}{T^{1/2}}y_{1,t} - \frac{1}{\left[\pi T^{3/2}\right]}y_{1,t+n} \Rightarrow \frac{1}{\pi}\omega^{1/2}\int_s^{\pi+s}J_c(\tau)d\tau - \omega^{1/2}J_c(s)$. ■

Theorem 1 Proof. Follows from Lemma 1. For example, for the first regression (omitting the intercept for clarity of exposition,) $\widehat{\beta}_n = \frac{\sum(y_{n-1,t+1}-y_{n,t})(s_{nt}/(n-1))}{\sum(s_{nt}/(n-1))^2}$.

After noting that $y_{n-1,t+1} - y_{n,t} = \frac{e^{c\pi}-1}{c\pi}\left\{\frac{c}{T}y_{1t} + u_{1,t+1}\right\} + op(1)$, the numerator simplifies to: $\frac{(e^{c\pi}-1)(e^{c\pi}-1-c\pi)}{c\pi^3}\frac{1}{T^2}\sum y_{1,t} + \frac{(e^{c\pi}-1)(e^{c\pi}-1-c\pi)}{c^2\pi^3}\frac{1}{T}\sum y_{1,t}u_{1,t+1} + o_p(1) \Rightarrow \frac{(e^{c\pi}-1)(e^{c\pi}-1-c\pi)}{c\pi^3}\omega\int J_c^2 + \frac{(e^{c\pi}-1)(e^{c\pi}-1-c\pi)}{c^2\pi^3}\left[\omega\int J_c dW + \omega_1\right]$. Similarly, the denominator converges to: $\frac{(e^{c\pi}-1-c\pi)^2}{(c\pi)^2}\frac{1}{\pi^2}\omega\int J_c^2$. Therefore, $\widehat{\beta} \Rightarrow \frac{(e^{c\pi}-1)c\pi}{(e^{c\pi}-1-c\pi)} + \frac{(e^{c\pi}-1)\pi}{(e^{c\pi}-1-c\pi)}\frac{\int J_c dW + \omega_1/\omega}{\int J_c^2}$. If we let $z_t = y_{n-1,t+1} - y_{n,t}, x_t = s_{nt}/(n-1)$, $\overline{z} = \frac{1}{T}\sum_{t=1}^T z_t$, and $\overline{x} = \frac{1}{T}\sum_{t=1}^T x_t$ then $s^2 = \frac{1}{T}\sum(z_t - \overline{z})^2 - \hat{\beta}_n^2\frac{1}{T}\sum(x_t - \overline{x})^2 = \frac{1}{T}\sum z_t^2 - \overline{z}^2 - \hat{\beta}_n^2\frac{1}{T}\sum x_t^2 + \hat{\beta}_n^2\overline{x}^2 = A - B - \hat{\beta}_n^2 C + \hat{\beta}_n^2 D$. It is easy to see that $A \xrightarrow{p} \left(\frac{e^{c\pi}-1}{c\pi}\right)^2\lambda_{1,0}$, and $B = C = D = o_p(1)$. Therefore, $s^2 \xrightarrow{p} \left(\frac{e^{c\pi}-1}{c\pi}\right)^2\lambda_{1,0}$. Also, $T^{-1}\sum(x_t - \overline{x})^2 = T^{-1}\sum x_t^2 + \overline{x}^2 \Rightarrow \omega\left(\frac{e^{c\pi}-1-c\pi}{c\pi^3}\right)\int J_c^2$, because $\overline{x} = o_p(1)$. Then, $t_{\hat{\beta}} = \frac{\hat{\beta}_n - 1}{s\left[\sum(x_t-\overline{x})^2\right]^{-1/2}} = O_p(1)$. (We can put the above pieces together to find the expression of the asymptotic distribution, but no useful insights are obtained).

For the second regression, $\hat{\psi}_n = \frac{\sum(s_{nt}^* - \overline{s}_n^*)(s_{nt} - \overline{s}_n)}{\sum(s_{nt} - \overline{s}_n)^2}$, where $\overline{x} = \frac{1}{T}\sum_{t=1}^T x_t$. After some tedious calculations, the numerator is: $T^{-2}\sum\left(s_{nt}^* - \overline{s}_n^*\right)\left(s_{nt} - \overline{s}_n\right) \Rightarrow$

$$\frac{1}{\pi}\left(\frac{e^{c\pi}-1-c\pi}{c\pi}\right)\left[\left\{\int_0^1 J_c(s)\int_s^{s+\pi} J_c(\tau)d\tau ds\right\} - \left\{\int_0^1 J_c(s)ds\right\}\left\{\int_0^1\int_s^{s+\pi} J_c(\tau)d\tau\right\}\right] - \left(\frac{e^{c\pi}-1-c\pi}{c\pi}\right)\left[\int_0^1 J_c^2(s)ds - \left(\int_0^1 J_c(s)ds\right)^2\right] \stackrel{def}{=} \frac{1}{\pi}\left(\frac{e^{c\pi}-1-c\pi}{c\pi}\right)\int_0^1 J_c{}^{\mu}(s)\int_s^{s+\pi} J_c{}^{\mu}(\tau)d\tau ds - \left(\frac{e^{c\pi}-1-c\pi}{c\pi}\right)\int_0^1\left(J_c{}^{\mu}(s)\right)^2 ds.$$

Since the denominator is $T^{-2}\sum\left(s_{nt} - \overline{s}_n\right)^2 \Rightarrow \left(\frac{e^{c\pi}-1-c\pi}{c\pi}\right)^2\int_0^1\left(J_c{}^{\mu}(s)\right)^2 ds$, we obtain the result in the theorem. Similar calculations yield the rest of the results.■

Theorem 2 Proof. Let $z_t^1 = (\widetilde{\triangle}y_{1,t+k}, ..., \widetilde{\triangle}y_{1,t-k})'$, $z_t^2 = y_{1,t}$. Note that z_t^1 is I(0) and z_t^2 is I(1). Let $z_t = \left(z_t^{1'}, z_t^2\right)'$ be a $(2k+2)$ dimensional vector of canonical regressors (Sims et al. (1990)). Define $A = (a_1, ..., a_{q-1})$ where $a_j = (d_{-k,j}, ..., d_{k,j}, \gamma(c_{j+1}))'$ and $d_{-k,j}$ is the jth coefficient of d_{-k}. In other words, the jth equation in (7) is: $y_{j+1,t} = a_j^{'}z_t + v_{j+1,t}$. Let $Y = (Y_{2,1}, ..., Y_{2,T})'$, $Z = (z_1, ..., z_T)'$, and $V = (v_1, ..., v_T)'$ be $T \times (q-1)$, $T \times (2k+2)$ and $T \times (q-1)$ matrices. If we stack the equations in (7) observation by observation, then $Y = ZA + V$. If $\widetilde{y} = vec(Y')$, $\widetilde{a} = vec(A')$ and $\widetilde{v} =$

$vec\,(V')$, then we can rewrite the stacked regression as $\widetilde{y} = (Z \bigotimes I_{n-1})\widetilde{a} + \widetilde{v}$. The least squares estimator of $\widetilde{a}$ is

$$\widehat{a} = \left[\sum z_t z_t' \bigotimes I_{n-1}\right]^{-1} \left[\sum \left(z_t \bigotimes I_{n-1}\right) Y_{2,t}\right]$$

Define $\Upsilon = diag\left(T^{1/2} I_{2k+1}, T\right)$. Follow arguments identical to Sims et al. (1990) or Stock and Watson (1993), we can show that the matrix $(\Upsilon \bigotimes I_{n-1}) \left[\sum z_t z_t' \bigotimes I_{n-1}\right]^{-1} (\Upsilon \bigotimes I_{n-1})$ is asymptotically block diagonal, conformable with the partition of $(\Upsilon \bigotimes I_{n-1})$ (i.e. the off-diagonal elements of the matrix are $op(1)$). To see this, note that $\Upsilon^{-1} \left(\sum z_t z_t'\right) \Upsilon^{-1} = \begin{bmatrix} T^{-1} \sum z_t^1 z_t^{1\prime} & T^{-3/2} \sum z_t^1 z_t^{2\prime} \\ T^{-3/2} \sum z_t^2 z_t^{1\prime} & T^{-2} \sum z_t^2 z_t^{2\prime} \end{bmatrix}$, and $T^{-3/2} \sum z_t^1 z_t^{2\prime} = T^{-3/2} \left[\sum z_t^{1\prime} z_t^2\right]' = op(1)$. Therefore, taking the last $(q-1)$ elements of $\widetilde{a}$, appropriately scaled,

$$T\left(\widehat{\Gamma(\underline{c})} - \Gamma(\underline{c})\right) = \left[T^{-2} \sum (y_{1,t})^2\right]^{-1} \left[T^{-1} \sum (y_{1,t})\, v_t\right] + op(1)$$

and noting that $T^{-1} \sum_{t=1}^{[sT]} v_t \Rightarrow \Omega_{2.1} W_{2.1}(s)$, where $\Omega_{2.1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$, then

$$T\left(\widehat{\Gamma(\underline{c})} - \Gamma(\underline{c})\right) \Rightarrow \Omega_{11}^{-1/2}\Omega_{2.1}^{1/2} \int J_c^\mu dW_{2.1} \left(\int (J_c^\mu)^2\right)^{-1}$$

■

Theorem 3 Proof. Under assumptions equivalent to assumptions B and C, Elliott (1994,1998) shows that

$$T\left(\widehat{\widehat{\Gamma(\underline{c})}} - \Gamma(\underline{c})\right) \Rightarrow \left(\Omega_{11} \int (J_c^\mu)^2\right)^{-1} \left(\Omega_{11}^{1/2}\Omega_{2.1}^{1/2} \int J_c^\mu dW_{2.1}\right) - \Omega_{11}^{-1}\Omega_{21} c$$

Since $G(.)$ and $D(.)$ are both continuous functions of $\underline{c}$, use the Mean Value Theorem to write $G(\widehat{\widehat{\Gamma(\underline{c})}}) - G(\Gamma(\underline{c_0})) = \left(\widehat{\widehat{\underline{c}}} - \underline{c_0}\right) = D(c^*)\left(\widehat{\widehat{\Gamma(\underline{c})}} - \Gamma(\underline{c_0})\right)$ where c^* is a vector whose elements are between the corresponding elements of $\widehat{\widehat{c}}$ and $\underline{c_0}$, and then delta-method arguments to obtain

$(T_{n-1})\left(\widehat{\underline{c}}-\underline{c}_0\right) \Rightarrow \left(D_0^{-1}\Omega_{11}\int (J_c^{\mu})^2 D_0^{-1}\right)^{-1} D_0^{-1}\left(\Omega_{11}^{1/2}\Omega_{2.1}^{1/2}\int J_c^{\mu}dW_{2.1}\right)-$ $\Omega_{11}^{-1}D_0\Omega_{21}c$. ■

Corollary 5 Proof. Note that $\hat{a}'\underline{c}_0 = c$ and $\hat{a}'\widehat{\underline{c}}=\widetilde{c}$. Then, $T\left(\widetilde{c}-c\right) = T\left(\hat{a}'\left(\widehat{\underline{c}}-\underline{c}_0\right)\right) \Rightarrow \hat{a}'\left(D_0^{-1}\Omega_{11}\int (J_c^{\mu})^2 D_0^{-1}\right)^{-1} D_0^{-1}\left(\Omega_{11}^{1/2}\Omega_{2.1}^{1/2}\int J_c^{\mu}dW_{2.1}\right)$ $-\Omega_{11}^{-1}\hat{a}'D_0\Omega_{21}c$ using the results from Theorem **3.**■

Theorem 6 Proof. Using $\widehat{\Gamma}^1(c) = \widehat{\widehat{\Gamma(c)}} - \frac{\widehat{B}}{T}$, $T\left(\widehat{\Gamma}^1(\underline{c})-\Gamma(\underline{c}_0)\right) = T\left(\widehat{\widehat{\Gamma(c)}}-\Gamma(\underline{c}_0)\right)-\widehat{B}$. Since $T\left(\widehat{\widehat{\Gamma(\underline{c})-}}\Gamma(\underline{c}_0)\right) \Rightarrow \left(\Omega_{11}\int (J_c^{\mu})^2\right)^{-1}\left(\Omega_{11}^{1/2}\Omega_{2.1}^{1/2}\int J_c^{\mu}dW_{2.1}\right)+$ B, the first result is obtained immediately, because $\left(B-\widehat{B}\right) = op(1)$. The second result is obtained by applying the delta-method as in Theorem 3.■

Theorem 7 Proof. From (9), $Y_{2,t} = \Gamma(\underline{c})y_{1,t} + \sum_{i=-k}^{k} d_i\widehat{\widetilde{\triangle}}y_{1,t-i} + v_t$. Adding and subtracting $\sum_{i=-k}^{k} d_i\widetilde{\triangle}y_{1,t-i}$ to obtain $Y_{2,t} = \Gamma(\underline{c})y_{1,t} + \sum_{i=-k}^{k} d_i\widetilde{\triangle}y_{1,t-i} + \sum_{i=-k}^{k} d_i\widehat{\widetilde{\triangle}}y_{1,t-i} - \sum_{i=-k}^{k} d_i\widetilde{\triangle}y_{1,t-i} + v_t$

$= \Gamma(\underline{c})y_{1,t} + \sum_{i=-k}^{k} d_i\widetilde{\triangle}y_{1,t-i} + \frac{c-\widehat{c}}{T}\sum_{i=-k}^{k} d_iy_{1,t-i-1} + v_t = \Gamma(\underline{c})y_{1,t} + \sum_{i=-k}^{k} d_i\widetilde{\triangle}y_{1,t-i} + r_t$, where $r_t = \frac{c-\tilde{c}}{T}\sum_{i=-k}^{k} d_iy_{1,t-i-1}+v_t$. Following closely the steps of the Theorem 2 proof above, let $z_t^1 = (\widetilde{\triangle}y_{1,t+k},...,\widetilde{\triangle}y_{1,t-k})'$, $z_t^2 = y_{1,t}$. Let $z_t = \left(z_t^{1'}, z_t^2\right)'$ be a $(2k+2)$ dimensional vector of canonical regressors. Define $A = (a_1,...,a_{q-1})$ where $a_j = (d_{-k,j},...,d_{k,j},\gamma(c_{j+1}))'$ and $d_{-k,j}$ is the jth coefficient of d_{-k}. In other words, the jth equation in (7) is: $y_{j+1,t} = a_j'z_t + r_{j+1,t}$. If $Y = (Y_{2,1},...,Y_{2,T})'$, $Z = (z_1,...,z_T)'$, and $R = (r_1,...,r_T)'$ be $T\times(n-1)$, $T\times(2k+2)$ and $T\times(n-1)$ matrices, then $Y = ZA + V$. Let $\widetilde{y} = vec(Y')$, $\widetilde{a} = vec(A')$ and $\widetilde{r} = vec\left(R'\right)$, then we can rewrite the stacked regression as $\widetilde{y} = (Z\bigotimes I_{n-1})\widetilde{a} + \widetilde{r}$. The least squares estimator of $\widetilde{a}$ is $\widehat{a} = \left[\sum z_tz_t'\bigotimes I_{n-1}\right]^{-1}\left[\sum\left(z_t\bigotimes I_{n-1}\right)Y_{2,t}\right]$. Define $\Upsilon = diag\left(T^{1/2}I_{2k+1}, T\right)$. Again, the matrix $\left[(\Upsilon\bigotimes I_{n-1})^{-1}\left[\sum z_tz_t'\bigotimes I_{n-1}\right](\Upsilon\bigotimes I_{n-1})^{-1}\right]^{-1}$ is asymptotically block diagonal, conformable with the partition of $(\Upsilon\bigotimes I_{n-1})$. Therefore, we can write $T\left(\widehat{\Gamma}^2(\underline{c})-\Gamma(\underline{c}_0)\right) = \left[T^{-2}\left(\sum y_{1,t}^2\right)\right]^{-1}\left[T^{-1}\sum y_{1,t}r_t\right]$ $+op(1)$. Using local-to-unity asymptotics, $\left[T^{-2}\left(\sum y_{1,t}^2\right)\right]^{-1} \Rightarrow \left[\Omega_{11}\int (J_c^{\mu})^2\right]^{-1}$.

Also $T^{-1}\sum y_{1,t}r_t = T^{-2}(c-\tilde{c})\, d_{-k}\sum y_{1,t}y_{1,t+k-1}+...+T^{-2}(c-\tilde{c})\, d_k \sum y_{1,t}y_{1,t-k-1}+$ $T^{-1}\sum y_{1,t}v_t = op(1)+...+op(1)+ Op(1) \Rightarrow \Omega_{11}^{1/2}\Omega_{2.1}^{1/2}\int J_c{}^{\mu}dW_{2.1}$. Note that, since $\sum y_{1,t}y_{1,t-j} = Op(T^2)$ for all j integer, the fact that $(c-\tilde{c}) = op(1)$ was critical in the last step. Therefore, $T\left(\widehat{\Gamma}^2(\underline{c}) - \Gamma(\underline{c}_0)\right) \Rightarrow \Omega_{11}^{-1/2}\Omega_{2.1}^{1/2}\left[\int (J_c{}^{\mu})^2\right]^{-1}\int J_c{}^{\mu}dW_{2.1}$, and since $J_c{}^{\mu}$ and $W_{2.1}$ are independent, we obtain the first result. The second result is obtained by applying delta-method arguments.■

Corollary 8 Proof. Follows directly by using the results in Theorems 6 and 7, the independence of $J_c{}^{\mu}$ and $W_{2.1}$, and Corollary 5.3 of Park and Phillips (1988).■

References

[1] Andrews, D. (1993): "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," Econometrica, 61:1,139-165.

[2] Bekaert, G., R. J. Hodrick, and D. A. Marshall (1997): "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," Journal of Financial Economics, 44, 309-348.

[3] Burnside, C. and M. Eichenbaum (1996): "Small-Sample Properties of GMM-Based Wald Tests," Journal of Business and Economic Statistics, 14(3), 294-308.

[4] Chan, N. H. (1988): "The Parameter Inference for Nearly Nonstationary Time Series," Journal of the American Statistical Association, 83, 857-862.

[5] Chan, N. H., and C. Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," Annals of Statistics, 15:3, 1050-1063.

[6] Campbell, J. Y. and R. J. Shiller (1987): "Cointegration and Tests of Present Value Models," Journal of Political Economy, 95:9, 1062-1088.

[7] Campbell, J. Y. and R. J. Shiller (1991): "Yield Spreads and Interest Rate Movements: A Bird's Eye View," Review of Economic Studies, 58:3, 495-514.

[8] Cavanagh, C. L., G. Elliott, and J. H. Stock (1995): "Inference in Models with Nearly Integrated Regressors," Econometric Theory, 11, 1131-1147.

[9] Cox, J. C, J. E. Ingersoll, and S. A. Ross (1985): "A Theory of the Term Structure of Interest Rates," Econometrica, 53, 385-407.

[10] Dufour, J.-M. (1990), "Exact Tests and Confidence sets in Linear Regressions with Autocorrelated Errors," Econometrica, 58:2,475-494.

[11] Elliott, G. (1994): "Application of Local to Unity Asymptotic Theory to Time Series Regression," Unpublished Ph.D. Dissertation, Department of Economics, Harvard University.

[12] Elliott, G. (1998): "On the Robustness of Cointegration Methods when Regressors Almost Have Unit Root," Econometrica, 66:1, 149-158.

[13] Engle, R. F., and C. W. J. Granger (1987): "Co-Integration and Error Correction: Representation, Estimation, and Testing," Econometrica, 55:2, 251-276.

[14] Hansen, B.E.(1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," Econometric Theory, 8, 485-500.

[15] McCulloch, J. and H. Kwon (1993): "U.S. Term Structure Data, 1947-1991," Working Paper, 93-6, Ohio State University.

[16] Ng, S., and P. Perron (1995): "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation," Journal of the American Statistical Association, 90, 268-281.

[17] Ng, S., and P. Perron (1998): "Lag Length Selection and the Construction of Unit Root Tests with Good size and Power," manuscript.

[18] Park, J. Y. and P. C. B. Phillips (1988): "Statistical Inference in Regressions with Integrated Processes:Part 1," Econometric Theory, 4, 468-497.

[19] Park, J. Y. (1992): "Canonical Cointegrating Regressions," Econometrica, 60, 119-143.

[20] Phillips, P. C. B. (1986): "Understanding Spurious Regressions in Econometrics," Journal of Econometrics, 33, 311-340.

[21] Phillips, P. C. B. (1987): "Toward a Unified Asymptotic Theory for Autoregression." Biometrika, 74:3, 535-547.

[22] Phillips, P. C. B. (1988): "Regression Theory for Near-Integrated Time Series," Econometrica, 56:5, 1021-1043.

[23] Phillips, P. C. B. (1991): "Optimal Inference in Cointegrated Systems," Econometrica, 59:2,283-30.

[24] Phillips, P. C. B. and M. Loretan (1991): "Estimating Long-Run Economic Equilibria," Review of Economic Studies, 57, 99-125.

[25] Saikkonen, P. (1991): "Asymptotically efficient estimation of cointegration Regressions," Econometric Theory, 8: 1-27.

[26] Shiller, R. J. (1990): "The Term Structure of Interest Rates," in Friedman, B. and Hahn, F. (eds.), The Handbook of Monetary Economics, Amsterdam: North Holland, Chapter 13.

[27] Shiller, R. J., J. Y. Campbell, and K. L. Schoenholtz (1983): "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates," Brookings Papers of Economic Activity, 1,173-217.

[28] Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in Linear Time Series Models with Some Unit Roots," Econometrica, 58:1, 113-144.

[29] Stock, J. H. (1996): "VAR, Error Correction and Pretest Forecasts at Long Horizons," Oxford Bulletin of Economics and Statistics, 58:4, 685-701.

[30] Stock, J. H. (1987): "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," Econometrica, 55:5, 1035-1056.

[31] Stock, J. H. (1991): "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," Journal of Monetary Economics, 28, 435-459.

[32] Stock, J. H. (1994): "Unit roots, Structural Breaks and Trends" in Engle, R. F. and McFadden, D. (eds.), Handbook of Econometrics, Vol IV, Amsterdam: Elsevier, Chapter 46.

[33] Stock, J. H. and M.W. Watson (1993): "A simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," Econometrica, 61:4, 783-820.

[34] Vasicek, O. A. (1977): :An Equilibrium Characterization of the Term Structure," Journal of Financial Economics, 5, 177-88.

[35] Watson, M. W. (1994): "Vector Autoregressions and Cointegration." In Engle, R. F. and McFadden, D. (eds.), Handbook of Econometrics, Vol. IV, Amsterdam:Elsevier, Chapter 47.

Empirical distribution of the OLS esimator of β (regression 2.3)									
T	c	π	.1	.2	.5	.8	.9	mean	s.d.
100	-10	0.05	1.104	1.328	1.943	2.724	3.222	2.067	0.865
100	-10	0.10	1.142	1.401	1.998	2.778	3.326	2.150	0.883
100	-10	0.25	1.055	1.292	1.827	2.633	3.103	1.989	0.823
100	-5	0.05	1.228	1.690	2.719	4.056	5.132	2.967	1.578
100	-5	0.10	1.352	1.721	2.766	4.371	5.295	3.130	1.677
100	-5	0.25	1.282	1.687	2.716	4.311	5.455	3.069	1.663
100	-3	0.05	1.550	2.095	3.550	5.769	7.156	4.034	2.381
100	-3	0.10	1.504	2.265	4.016	6.517	8.418	4.532	2.643
100	-3	0.25	1.568	2.134	3.941	6.494	8.251	4.490	2.687
100	-1	0.05	2.554	4.098	8.270	14.584	19.229	9.721	7.011
100	-1	0.10	2.959	4.478	9.088	17.184	21.424	11.054	7.697
100	-1	0.25	2.972	4.954	9.548	17.635	22.936	11.597	8.473
100	1	0.05	-14.483	-10.008	-4.079	-0.563	0.549	-5.687	6.369
100	1	0.10	-16.986	-11.877	-4.528	-0.557	0.947	-6.622	7.868
100	1	0.25	-16.927	-12.257	-4.788	-0.453	1.132	-6.868	8.324
100	3	0.05	-1.781	0.197	1.305	1.564	1.683	0.579	1.888
100	3	0.10	-1.319	0.473	1.515	1.796	1.916	0.778	1.996
100	3	0.25	-2.073	0.416	1.701	2.076	2.241	0.800	2.483
500	-10	0.05	1.335	1.630	2.362	3.345	3.995	2.555	1.117
500	-10	0.10	1.361	1.615	2.276	3.304	3.895	2.499	1.029
500	-10	0.25	1.122	1.343	1.853	2.712	3.126	2.056	0.898
500	-5	0.05	1.451	1.882	3.232	5.301	6.463	3.685	2.085
500	-5	0.10	1.395	1.909	3.196	5.014	6.196	3.575	2.007
500	-5	0.25	1.359	1.795	2.980	4.446	5.528	3.265	1.774
500	-3	0.05	1.735	2.421	4.312	7.299	9.474	5.041	3.235
500	-3	0.10	1.788	2.506	4.419	7.435	9.421	5.142	3.173
500	-3	0.25	1.784	2.437	4.142	7.218	9.014	4.847	2.981
500	-1	0.05	3.242	5.490	10.606	17.927	23.713	12.406	8.626
500	-1	0.10	2.904	4.800	9.962	18.934	23.913	11.999	8.714
500	-1	0.25	3.436	5.558	10.612	17.871	24.738	12.620	9.182
500	1	0.05	-18.878	-12.880	-4.965	-0.854	0.517	-7.314	8.246
500	1	0.10	-18.948	-13.395	-4.975	-0.547	0.824	-7.389	8.625
500	1	0.25	-19.404	-12.950	-4.604	-0.412	1.112	-7.427	9.777
500	3	0.05	-1.600	0.543	1.594	1.876	2.007	0.790	2.220
500	3	0.10	-1.785	0.389	1.661	1.960	2.093	0.802	2.270
500	3	0.25	-1.302	0.860	1.873	2.184	2.348	1.091	2.278

Table 1a: Empirical distribution of the OLS estimator of β for $c = (-10, -5, -3, -1, 1, 3)$, $T = (100, 500)$, $\pi = (0.05, 0.10, 0.25)$ and $u_t = \varepsilon_t$, $\varepsilon_t \sim N(0, I_2)$. The 10th, 20th, 50th, 80th, and 90th quantiles are reported as well as the mean and the standard deviation. The parameter c has a big impact on the distributions.

Simulated limiting distribution of β (regression 2.3)									
T	c	π	.1	.2	.5	.8	.9	mean	s.d.
100	-10	0.05	1.380	1.659	2.428	3.405	4.027	2.584	1.081
100	-10	0.10	1.269	1.556	2.220	3.087	3.696	2.388	0.981
100	-10	0.25	1.099	1.346	1.904	2.742	3.232	2.072	0.857
100	-5	0.05	1.535	2.113	3.398	5.070	6.415	3.709	1.972
100	-5	0.10	1.503	1.912	3.073	4.856	5.883	3.477	1.863
100	-5	0.25	1.335	1.758	2.829	4.490	5.682	3.197	1.733
100	-3	0.05	1.937	2.619	4.437	7.211	8.945	5.042	2.976
100	-3	0.10	1.671	2.517	4.463	7.241	9.354	5.035	2.937
100	-3	0.25	1.633	2.223	4.106	6.764	8.595	4.677	2.799
100	-1	0.05	3.193	5.123	10.337	18.230	24.036	12.151	8.764
100	-1	0.10	3.288	4.975	10.098	19.093	23.805	12.282	8.552
100	-1	0.25	3.095	5.160	9.946	18.370	23.892	12.080	8.826
100	1	0.05	-18.103	-12.510	-5.098	-0.703	0.686	-7.109	7.961
100	1	0.10	-18.873	-13.197	-5.032	-0.619	1.052	-7.358	8.742
100	1	0.25	-17.632	-12.768	-4.987	-0.472	1.180	-7.155	8.671
100	3	0.05	-2.226	0.246	1.631	1.955	2.104	0.724	2.360
100	3	0.10	-1.465	0.525	1.683	1.995	2.129	0.865	2.217
100	3	0.25	-2.159	0.433	1.772	2.163	2.334	0.834	2.587
500	-10	0.05	1.391	1.698	2.460	3.485	4.162	2.662	1.164
500	-10	0.10	1.389	1.648	2.322	3.371	3.974	2.550	1.050
500	-10	0.25	1.131	1.354	1.867	2.734	3.151	2.073	0.905
500	-5	0.05	1.511	1.961	3.367	5.522	6.733	3.838	2.172
500	-5	0.10	1.423	1.948	3.261	5.116	6.322	3.648	2.048
500	-5	0.25	1.370	1.810	3.004	4.482	5.573	3.292	1.788
500	-3	0.05	1.807	2.521	4.492	7.603	9.868	5.251	3.370
500	-3	0.10	1.825	2.557	4.509	7.587	9.613	5.247	3.237
500	-3	0.25	1.798	2.457	4.175	7.277	9.087	4.886	3.006
500	-1	0.05	3.377	5.719	11.048	18.674	24.701	12.923	8.986
500	-1	0.10	2.963	4.898	10.166	19.320	24.401	12.244	8.892
500	-1	0.25	3.463	5.603	10.698	18.016	24.938	12.721	9.256
500	1	0.05	-19.665	-13.417	-5.172	-0.889	0.538	-7.619	8.589
500	1	0.10	-19.335	-13.669	-5.077	-0.558	0.841	-7.540	8.801
500	1	0.25	-19.560	-13.055	-4.641	-0.415	1.121	-7.487	9.856
500	3	0.05	-1.667	0.565	1.661	1.954	2.090	0.823	2.313
500	3	0.10	-1.821	0.397	1.695	2.000	2.136	0.819	2.316
500	3	0.25	-1.312	0.867	1.888	2.202	2.366	1.100	2.296

Table 1b: Simulated limiting distribution of the OLS estimator of β for $c = (-10, -5, -3, -1, 1, 3)$, $T = (100, 500)$, $\pi = (0.05, 0.10, 0.25)$ and $u_t = \varepsilon_t$, $\varepsilon_t \sim N(0, I_2)$. The stochastic integrals are simulated by rescaled partial sums of $y_{1,t}$. The 10th, 20th, 50th, 80th, and 90th quantiles are reported as well as the mean and the standard deviation. The parameter c has a big impact on the distributions.

Empirical distribution of the OLS estimator of γ (regression 2.4)									
T	c	π	.1	.2	.5	.8	.9	mean	s.d.
100	-10	0.05	1.173	1.389	1.932	2.520	2.906	1.986	0.671
100	-10	0.10	0.976	1.179	1.552	1.973	2.198	1.571	0.456
100	-10	0.25	0.985	1.115	1.345	1.530	1.599	1.316	0.245
100	-5	0.05	1.267	1.646	2.535	3.709	4.401	2.739	1.292
100	-5	0.10	1.053	1.313	2.031	2.823	3.257	2.082	0.836
100	-5	0.25	0.979	1.284	1.686	2.132	2.278	1.670	0.503
100	-3	0.05	1.404	1.984	3.386	5.259	6.428	3.707	2.036
100	-3	0.10	1.243	1.741	2.694	4.101	4.821	2.914	1.395
100	-3	0.25	1.197	1.499	2.143	2.866	3.210	2.165	0.774
100	-1	0.05	2.497	4.099	7.722	12.749	15.524	8.473	5.240
100	-1	0.10	2.276	3.354	6.335	10.009	12.065	6.770	3.923
100	-1	0.25	1.995	2.928	5.112	6.852	7.725	4.963	2.193
100	1	0.05	-13.046	-9.126	-3.292	-0.375	0.568	-4.969	5.838
100	1	0.10	-9.045	-6.497	-2.450	-0.207	0.627	-3.440	3.851
100	1	0.25	-5.904	-4.684	-1.996	-0.117	0.629	-2.348	2.468
100	3	0.05	-0.873	0.382	1.218	1.447	1.542	0.708	1.370
100	3	0.10	-0.560	0.457	1.019	1.183	1.277	0.656	0.991
100	3	0.25	0.056	0.577	0.941	1.078	1.141	0.725	0.645
500	-10	0.05	0.847	0.994	1.401	1.918	2.157	1.467	0.533
500	-10	0.10	0.873	1.034	1.367	1.754	1.947	1.389	0.412
500	-10	0.25	0.940	1.067	1.286	1.481	1.568	1.267	0.245
500	-5	0.05	0.898	1.181	1.884	2.798	3.399	2.031	0.988
500	-5	0.10	0.963	1.193	1.807	2.489	2.952	1.875	0.763
500	-5	0.25	0.981	1.180	1.626	2.024	2.220	1.599	0.480
500	-3	0.05	1.062	1.504	2.612	4.072	5.092	2.835	1.554
500	-3	0.10	1.002	1.388	2.305	3.393	4.053	2.429	1.186
500	-3	0.25	1.153	1.492	2.128	2.783	3.131	2.120	0.757
500	-1	0.05	1.921	3.107	5.720	9.414	11.790	6.390	4.055
500	-1	0.10	1.809	3.088	5.569	8.583	10.505	5.921	3.443
500	-1	0.25	1.670	2.651	4.649	6.589	7.414	4.617	2.176
500	1	0.05	-8.809	-6.431	-2.546	-0.245	0.461	-3.422	3.799
500	1	0.10	-7.839	-5.712	-2.365	-0.161	0.545	-2.986	3.222
500	1	0.25	-5.292	-4.061	-1.502	-0.002	0.533	-1.973	2.226
500	3	0.05	-0.668	0.328	0.889	1.035	1.102	0.511	1.038
500	3	0.10	-0.314	0.490	0.863	1.010	1.077	0.588	0.814
500	3	0.25	-0.199	0.530	0.881	1.011	1.077	0.655	0.640

Table 2a: Empirical distribution of the OLS estimator of ψ for $c = (-10, -5, -3, -1, 1, 3)$, $T = (100, 500)$, $\pi = (0.05, 0.10, 0.25)$ and $u_t = \varepsilon_t$, $\varepsilon_t \sim N(0, I_2)$. The 10th, 20th, 50th, 80th, and 90th quantiles are reported as well as the mean and the standard deviation. The parameter c has a big impact on the distributions.

Simulated limiting distribution of γ (regression 2.4)									
T	c	π	.1	.2	.5	.8	.9	mean	s.d.
100	-10	0.05	0.997	1.180	1.641	2.141	2.468	1.687	0.570
100	-10	0.10	0.925	1.117	1.471	1.870	2.084	1.489	0.432
100	-10	0.25	0.979	1.108	1.336	1.520	1.589	1.307	0.243
100	-5	0.05	1.046	1.359	2.093	3.063	3.634	2.262	1.067
100	-5	0.10	0.976	1.217	1.883	2.616	3.018	1.929	0.775
100	-5	0.25	0.962	1.262	1.657	2.096	2.239	1.642	0.494
100	-3	0.05	1.146	1.618	2.762	4.290	5.243	3.023	1.661
100	-3	0.10	1.140	1.596	2.470	3.760	4.420	2.672	1.279
100	-3	0.25	1.168	1.463	2.091	2.796	3.132	2.113	0.755
100	-1	0.05	2.010	3.301	6.218	10.267	12.502	6.824	4.220
100	-1	0.10	2.062	3.039	5.739	9.067	10.929	6.133	3.554
100	-1	0.25	1.928	2.829	4.938	6.619	7.462	4.794	2.118
100	1	0.05	-10.366	-7.252	-2.616	-0.298	0.452	-3.948	4.639
100	1	0.10	-8.085	-5.807	-2.190	-0.185	0.560	-3.075	3.442
100	1	0.25	-5.628	-4.464	-1.903	-0.111	0.600	-2.238	2.353
100	3	0.05	-0.684	0.299	0.955	1.134	1.208	0.555	1.074
100	3	0.10	-0.494	0.402	0.897	1.042	1.125	0.578	0.873
100	3	0.25	0.053	0.541	0.882	1.009	1.069	0.679	0.604
500	-10	0.05	0.822	0.966	1.361	1.863	2.095	1.425	0.518
500	-10	0.10	0.864	1.024	1.353	1.736	1.927	1.375	0.407
500	-10	0.25	0.939	1.066	1.284	1.479	1.566	1.266	0.244
500	-5	0.05	0.867	1.141	1.820	2.702	3.283	1.962	0.954
500	-5	0.10	0.950	1.176	1.781	2.453	2.909	1.849	0.752
500	-5	0.25	0.978	1.176	1.620	2.017	2.212	1.594	0.478
500	-3	0.05	1.023	1.450	2.517	3.924	4.907	2.732	1.498
500	-3	0.10	0.986	1.365	2.267	3.337	3.987	2.389	1.166
500	-3	0.25	1.148	1.485	2.118	2.769	3.116	2.109	0.753
500	-1	0.05	1.847	2.986	5.498	9.050	11.333	6.142	3.898
500	-1	0.10	1.775	3.030	5.465	8.423	10.309	5.810	3.378
500	-1	0.25	1.658	2.633	4.618	6.544	7.364	4.586	2.161
500	1	0.05	-8.445	-6.166	-2.441	-0.235	0.442	-3.280	3.642
500	1	0.10	-7.671	-5.590	-2.315	-0.158	0.533	-2.923	3.153
500	1	0.25	-5.242	-4.022	-1.488	-0.002	0.528	-1.954	2.205
500	3	0.05	-0.639	0.313	0.850	0.990	1.054	0.488	0.992
500	3	0.10	-0.307	0.478	0.842	0.986	1.051	0.573	0.794
500	3	0.25	-0.197	0.523	0.870	0.998	1.063	0.647	0.632

Table 2b: Simulated limiting distribution of the OLS estimator of ψ for $c = (-10, -5, -3, -1, 1, 3)$, $T = (100, 500)$, $\pi = (0.05, 0.10, 0.25)$ and $u_t = \varepsilon_t$, $\varepsilon_t \sim N(0, I_2)$. The stochastic integrals are simulated by rescaled partial sums of $y_{1,t}$. The 10th, 20th, 50th, 80th, and 90th quantiles are reported as well as the mean and the standard deviation. The parameter c has a big impact on the distributions.

OLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	-9.228	-6.943	-6.007	-4.353	-2.739	-1.848	0.333	0.699	2.902	2.934
0.10	-7.590	-6.145	-4.500	-3.780	-1.918	-1.457	0.689	0.862	2.941	2.960
0.20	-6.895	-5.839	-4.015	-3.471	-1.492	-1.221	0.856	0.932	2.970	2.986
0.30	-6.820	-5.740	-3.810	-3.385	-1.341	-1.160	0.909	0.957	2.979	2.990

DOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	-6.333	-5.597	-3.956	-3.294	-1.356	-1.107	1.130	1.130	3.567	3.280
0.10	-5.910	-5.357	-3.332	-3.211	-1.198	-1.095	1.070	1.057	3.228	3.117
0.20	-5.693	-5.283	-3.303	-3.129	-1.094	-1.023	1.044	1.029	3.097	3.048
0.30	-5.715	-5.258	-3.231	-3.094	-1.057	-1.016	1.029	1.018	3.048	3.025

QDOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	-4.798	-4.937	-3.061	-2.951	-1.078	-1.031	0.946	1.064	2.919	2.991
0.10	-5.043	-4.994	-2.901	-3.013	-1.071	-1.046	0.973	1.015	2.962	2.989
0.20	-5.097	-5.026	-3.040	-3.011	-1.028	-0.995	1.000	1.007	2.995	2.999
0.30	-5.178	-5.030	-3.032	-3.000	-1.012	-0.995	1.001	1.004	2.992	2.998

Iterated QDOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	-5.279	-5.036	-3.329	-3.005	-1.164	-1.051	0.988	1.081	3.064	3.024
0.10	-5.206	-5.025	-2.970	-3.029	-1.089	-1.051	0.985	1.018	2.989	2.996
0.20	-5.174	-5.041	-3.066	-3.016	-1.033	-0.996	1.003	1.008	3.000	3.000
0.30	-5.244	-5.042	-3.047	-3.004	-1.015	-0.995	1.002	1.004	2.994	2.998

Nonparametric QDOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	-5.709	-5.366	-3.473	-3.130	-1.138	-1.032	0.972	1.054	2.938	3.000
0.10	-5.524	-5.230	-3.102	-3.121	-1.122	-1.057	0.999	1.019	2.971	2.998
0.20	-5.468	-5.183	-3.157	-3.074	-1.054	-1.005	1.010	1.013	3.003	3.001
0.30	-5.497	-5.160	-3.125	-3.051	-1.027	-1.004	1.007	1.008	2.998	3.001

Table 3a: Estimation of c with various estimators. The data is simulated 5000 times using equations $(3.2a - 3.2b)$ for two yields, $c = (-5, -3, -1, 1, 3), T = (200, 400)$, $\pi = (0.05, 0.1, 0.2, 0.3)$, and $u_t = \varepsilon_t$, $\varepsilon_t \sim N(0, \Sigma)$, $\Sigma = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 5 \end{bmatrix}$.

OLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	28.278	23.975	27.411	22.940	25.952	21.453	21.275	16.814	10.898	8.214
0.10	13.259	11.111	13.130	10.938	12.825	10.310	10.770	8.600	6.845	5.396
0.20	5.979	4.738	6.115	4.809	6.281	4.975	5.779	4.670	4.548	4.154
0.30	3.662	2.741	3.776	2.986	3.973	3.199	3.892	3.282	3.944	3.621

DOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	-5.797	-5.683	-3.389	-3.369	-0.909	-1.114	1.379	1.142	3.599	3.319
0.10	-5.531	-5.416	-3.138	-3.203	-0.951	-1.070	1.197	1.058	3.271	3.149
0.20	-5.341	-5.293	-3.108	-3.123	-0.979	-1.028	1.096	1.031	3.113	3.061
0.30	-5.355	-5.250	-3.085	-3.105	-0.977	-1.020	1.061	1.019	3.061	3.032

QDOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	-4.429	-4.873	-2.649	-2.916	-0.745	-1.065	1.139	1.088	2.975	2.961
0.10	-4.771	-4.965	-2.762	-2.968	-0.855	-1.003	1.091	1.000	3.015	2.994
0.20	-4.835	-4.987	-2.880	-2.983	-0.924	-0.993	1.050	1.004	3.013	3.001
0.30	-4.900	-4.979	-2.906	-2.993	-0.938	-0.995	1.033	1.002	3.008	3.000

Iterated QDOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	-4.763	-4.990	-2.818	-2.979	-0.799	-1.091	1.196	1.116	3.085	3.000
0.10	-4.879	-5.003	-2.809	-2.985	-0.867	-1.008	1.100	1.003	3.035	3.002
0.20	-4.884	-5.005	-2.898	-2.989	-0.928	-0.994	1.052	1.004	3.017	3.002
0.30	-4.939	-4.993	-2.916	-2.997	-0.940	-0.996	1.034	1.002	3.009	3.000

Nonparametric QDOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	-4.394	-5.013	-2.625	-2.999	-0.714	-1.000	1.104	1.033	2.849	3.016
0.10	-4.741	-5.040	-2.752	-3.009	-0.848	-1.013	1.080	1.009	2.958	3.020
0.20	-4.816	-5.040	-2.875	-3.008	-0.922	-0.999	1.046	1.008	2.991	3.010
0.30	-4.880	-5.025	-2.900	-3.013	-0.937	-1.000	1.031	1.005	2.997	3.005

Table 3b: Estimation of c with various estimators. The data is simulated 5000 times using equations $(3.2a - 3.2b)$ for two yields, $c = (-5, -3, -1, 1, 3), T = (200, 400)$, $\pi = (0.05, 0.1, 0.2, 0.3)$, and $u_t = Au_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, \Sigma)$, $\Sigma = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 5 \end{bmatrix}$, $A = \begin{bmatrix} -0.5 & -0.5 \\ 0.05 & -1.0 \end{bmatrix}$.

Ratio of variances of various estimators, for T= 200 and 400																				
c	**-5**				**-3**				**-1**				**1**				**3**			
π	**0.05**	**0.10**	**0.20**	**0.30**	**0.05**	**0.10**	**0.20**	**0.30**	**0.05**	**0.10**	**0.20**	**0.30**	**0.05**	**0.10**	**0.20**	**0.30**	**0.05**	**0.10**	**0.20**	**0.30**
OLS	6.2	6.0	6.8	7.6	4.2	4.5	5.4	4.6	4.0	5.2	4.9	4.6	3.9	4.3	4.6	4.1	3.0	3.3	5.0	4.1
DOLS	5.6	5.1	5.6	6.6	4.1	4.3	4.8	4.0	3.7	4.8	4.5	4.2	3.9	3.9	4.5	4.8	3.5	3.7	4.6	4.4
QDOLS	4.5	3.8	4.7	4.9	3.2	3.6	4.3	3.7	2.9	4.3	4.2	4.0	3.1	3.6	4.3	4.6	3.3	3.4	4.6	4.4
IQDOLS	5.1	4.2	4.9	5.3	3.6	3.8	4.4	3.8	3.2	4.4	4.2	4.1	3.4	3.7	4.3	4.6	3.3	3.4	4.6	4.4
NQDOLS	6.2	4.9	5.7	6.7	3.6	4.0	4.4	4.0	3.2	4.7	4.3	4.2	3.2	3.7	4.4	4.7	3.1	3.1	4.4	4.3

Ratio of variances of various estimators, for T= 200 and 400																				
c	**-5**				**-3**				**-1**				**1**				**3**			
π	**0.05**	**0.10**	**0.20**	**0.30**	**0.05**	**0.10**	**0.20**	**0.30**	**0.05**	**0.10**	**0.20**	**0.30**	**0.05**	**0.10**	**0.20**	**0.30**	**0.05**	**0.10**	**0.20**	**0.30**
OLS	0.8	0.8	0.8	0.7	0.7	0.8	0.8	0.8	0.9	0.9	0.8	0.9	1.1	1.1	1.0	1.1	1.5	1.7	1.4	1.7
DOLS	5.2	4.1	4.9	4.8	4.2	4.9	4.3	4.2	3.7	5.1	4.1	4.4	5.0	4.6	4.9	4.6	7.7	4.0	4.9	5.5
QDOLS	4.0	3.6	4.4	4.4	3.3	4.3	4.0	4.1	3.2	4.7	3.9	4.3	4.6	4.4	4.7	4.5	6.9	3.9	4.8	5.7
IQDOLS	4.4	3.7	4.5	4.4	3.7	4.4	4.0	4.1	3.2	4.8	3.9	4.3	4.5	4.5	4.7	4.5	7.1	3.9	4.8	5.7
NQDOLS	3.7	3.3	4.1	4.0	3.1	4.2	3.8	3.8	2.9	4.5	3.8	4.2	4.1	4.3	4.8	4.5	7.9	5.0	5.9	6.4

Table 4: Empirical simulation of the rate of convergence of $\widetilde{c}$. The data in the first table is simulated 5000 times using equations $(3.2a-3.2b)$ for two yields, using $c=(-5,-3,-1,1,3), T=(200,400)$, $\pi=(0.05,0.1,0.2,0.3)$, and $u_t=\varepsilon_t$, $\varepsilon_t\sim N(0,\Sigma)$, $\Sigma=\begin{bmatrix}0.5 & -0.5\\ -0.5 & 5\end{bmatrix}$. In the second table, the errors are: $u_t=Au_{t-1}+\varepsilon_t$, $\varepsilon_t\sim N(0,\Sigma)$, $\Sigma=\begin{bmatrix}0.5 & -0.5\\ -0.5 & 5\end{bmatrix}$, $A=\begin{bmatrix}-0.5 & -0.5\\ 0.05 & -1.0\end{bmatrix}$. The entries are ratios of the variances of $\widetilde{c}$ for $T=200$ versus $T=400$. If $\widetilde{c}$ converges at rate T, the ratios should be close to 4.

Real Coverage between the ± 1.96 std. dev., QDOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	0.956	0.940	0.947	0.947	0.941	0.941	0.939	0.933	0.947	0.944
0.10	0.938	0.946	0.952	0.947	0.942	0.945	0.938	0.943	0.951	0.944
0.20	0.941	0.947	0.964	0.946	0.944	0.944	0.948	0.938	0.935	0.942
0.30	0.955	0.949	0.952	0.948	0.938	0.939	0.940	0.936	0.947	0.947

Real Coverage between the ± 1.96 std. dev., NQDOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	0.955	0.932	0.940	0.947	0.941	0.938	0.936	0.938	0.950	0.949
0.10	0.945	0.933	0.948	0.943	0.950	0.943	0.939	0.939	0.953	0.947
0.20	0.939	0.942	0.956	0.944	0.945	0.943	0.951	0.932	0.936	0.949
0.30	0.958	0.946	0.950	0.942	0.935	0.935	0.941	0.936	0.951	0.946

Table 5a: Real versus Nominal Coverage. The data is simulated 5000 times using equations $(3.2a - 3.2b)$ for two yields, using $c = (-5, -3, -1, 1, 3), T = (200, 400)$, $\pi = (0.05, 0.1, 0.2, 0.3)$, and $u_t = \varepsilon_t$, $\varepsilon_t \sim N(0, \Sigma)$, $\Sigma = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 5 \end{bmatrix}$. The entries are the fraction of Monte Carlo replications for which the true value of c was contained within 1.96 standard deviations from the estimated c.

Real Coverage between the ± 1.96 std. dev., QDOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	0.920	0.941	0.923	0.953	0.917	0.939	0.946	0.934	0.948	0.936
0.10	0.936	0.947	0.921	0.947	0.934	0.941	0.938	0.941	0.929	0.943
0.20	0.934	0.942	0.924	0.938	0.924	0.950	0.926	0.941	0.933	0.930
0.30	0.938	0.940	0.932	0.953	0.931	0.954	0.919	0.932	0.931	0.935

Real Coverage between the ± 1.96 std. dev., NQDOLS										
	c=-5		**c=-3**		**c=-1**		**c= 1**		**c= 3**	
$\pi \mid T$	200	400	200	400	200	400	200	400	200	400
0.05	0.915	0.948	0.920	0.955	0.925	0.939	0.948	0.939	0.955	0.942
0.10	0.925	0.947	0.919	0.953	0.932	0.940	0.938	0.941	0.939	0.942
0.20	0.930	0.946	0.919	0.936	0.923	0.951	0.925	0.938	0.950	0.926
0.30	0.936	0.943	0.931	0.953	0.930	0.954	0.923	0.936	0.951	0.937

Table 5b: Real versus Nominal Coverage. The data is simulated 5000 times using equations $(3.2a - 3.2b)$ for two yields, using $c = (-5, -3, -1, 1, 3), T = (200, 400)$, $\pi = (0.05, 0.1, 0.2, 0.3)$, and $u_t = Au_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, \Sigma)$, $\Sigma = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 5 \end{bmatrix}$, $A = \begin{bmatrix} -0.5 & -0.5 \\ 0.05 & -1.0 \end{bmatrix}$. The entries are the fraction of Monte Carlo replications for which the true value of c was contained within 1.96 standard deviations from the estimated c.

QDOLS: Ratio of variances of c for various π										
	T=200					**T=400**				
Ratio	c=-5	c=-3	c=-1	c=1	c=3	c=-5	c=-3	c=-1	c=1	c=3
0.05/0.10	3.022	3.062	2.753	3.412	4.305	2.567	3.474	4.084	3.919	4.452
	2.897	*3.287*	*3.744*	*4.278*	*4.904*	*2.897*	*3.287*	*3.744*	*4.278*	*4.904*
0.05/0.20	5.053	5.968	9.121	13.512	19.390	5.271	8.015	13.113	18.474	26.787
	6.215	*8.952*	*13.127*	*19.583*	*29.690*	*6.215*	*8.952*	*13.127*	*19.583*	*29.690*
0.05/0.30	5.804	10.800	16.480	31.047	57.536	6.336	12.682	22.934	45.778	75.938
	7.734	*13.859*	*25.923*	*50.483*	*102.043*	*7.734*	*13.859*	*25.923*	*50.483*	*102.043*

Nonparametric QDOLS: Ratio of variances of c for various π										
	T=200					**T=400**				
Ratio	c=-5	c=-3	c=-1	c=1	c=3	c=-5	c=-3	c=-1	c=1	c=3
0.05/0.10	3.566	3.312	2.892	3.708	4.561	2.826	3.646	4.242	4.236	4.570
	2.897	*3.287*	*3.744*	*4.278*	*4.904*	*2.897*	*3.287*	*3.744*	*4.278*	*4.904*
0.05/0.20	6.233	7.148	10.441	15.184	21.758	5.748	8.790	13.984	20.624	30.660
	6.215	*8.952*	*13.127*	*19.583*	*29.690*	*6.215*	*8.952*	*13.127*	*19.583*	*29.690*
0.05/0.30	6.464	12.716	19.071	35.270	65.867	6.939	13.984	24.721	51.365	92.237
	7.734	*13.859*	*25.923*	*50.483*	*102.043*	*7.734*	*13.859*	*25.923*	*50.483*	*102.043*

QDOLS: Ratio of variances of c for various π										
	T=200					**T=400**				
Ratio	c=-5	c=-3	c=-1	c=1	c=3	c=-5	c=-3	c=-1	c=1	c=3
0.05/0.10	2.567	2.683	2.977	3.742	5.919	2.343	3.488	4.435	3.548	3.316
	2.897	*3.287*	*3.744*	*4.278*	*4.904*	*2.897*	*3.287*	*3.744*	*4.278*	*4.904*
0.05/0.20	4.935	6.533	10.181	14.844	25.848	5.441	7.801	12.532	15.153	17.798
	6.215	*8.952*	*13.127*	*19.583*	*29.690*	*6.215*	*8.952*	*13.127*	*19.583*	*29.690*
0.05/0.30	6.052	10.175	18.122	41.996	68.979	6.667	12.549	24.237	40.642	56.969
	7.734	*13.859*	*25.923*	*50.483*	*102.043*	*7.734*	*13.859*	*25.923*	*50.483*	*102.043*

Nonparametric QDOLS: Ratio of variances of c for various π										
	T=200					**T=400**				
Ratio	c=-5	c=-3	c=-1	c=1	c=3	c=-5	c=-3	c=-1	c=1	c=3
0.05/0.10	2.642	2.620	2.873	3.825	5.680	2.359	3.588	4.424	4.013	3.576
	2.897	*3.287*	*3.744*	*4.278*	*4.904*	*2.897*	*3.287*	*3.744*	*4.278*	*4.904*
0.05/0.20	4.788	6.479	9.856	15.167	26.474	5.347	8.000	13.006	17.693	19.807
	6.215	*8.952*	*13.127*	*19.583*	*29.690*	*6.215*	*8.952*	*13.127*	*19.583*	*29.690*
0.05/0.30	6.212	10.238	17.566	42.639	77.325	6.629	12.765	25.167	47.199	62.852
	7.734	*13.859*	*25.923*	*50.483*	*102.043*	*7.734*	*13.859*	*25.923*	*50.483*	*102.043*

Table 6: Dependence of the variance of $\widetilde{c}$ on π. The data in the rst tableau are simulated 5000 times using equations $(3.2a - 3.2b)$ for two yields, $c = (-5, -3, -1, 1, 3), T = (200, 400)$, $\pi = (0.05, 0.1, 0.2, 0.3)$, and $u_t = \varepsilon_t$, $\varepsilon_t \sim N(0, \Sigma)$, $\Sigma = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 5 \end{bmatrix}$. In the second tableau, the errors are: $u_t = Au_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, \Sigma)$, $\Sigma = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 5 \end{bmatrix}$, $A = \begin{bmatrix} -0.5 & -0.5 \\ 0.05 & -1.0 \end{bmatrix}$. The entries are ratios of the variances for $\pi = 0.05$ versus $\pi = 0.1$, 0.2, 0.3. The numbers in italics are the theoretical values of the ratio.

	System			
Hypothesis	5mo-13yr	13mo-13yr	17mo-13yr	3yr-13yr
-0.5	1244	1053	993	661
0.00	912	1177	1039	772
1.00	9986	21044	16452	14306
unspeci ed	703	947	869	593

Table 7: Wald statistics for various null hypotheses and various speci cations of the system in (3.2a-3.2b), using the US yield data from McCulloch and Kwon (1993). For example, the value of the test for Ho : $c_{k1} = c_{k2} = ... = c_{kn} = -0.5$, where $k1 = 36$, $k2 = 42$, ... , $kn = 156$ (months), is 661. The statistics in the last rows test for $Ho : c_{k1} = c_{k2} = ... = c_{kn}$. The REHTS is rejected.

Real coverage					
c	π	0.90	0.80	0.70	median
-5	0.00	0.89	0.79	0.70	0.45
-3	0.00	0.89	0.79	0.67	0.50
-3	0.05	0.83	0.74	0.65	0.65
-3	0.10	0.83	0.71	0.61	0.65
-1	0.00	0.91	0.81	0.72	0.50
-1	0.05	0.86	0.75	0.65	0.60
-1	0.10	0.86	0.74	0.66	0.61
-1	0.20	0.87	0.75	0.66	0.65

Real coverage					
c	π	0.90	0.80	0.70	median
-5	0.00	0.88	0.77	0.66	0.47
-5	0.05	0.88	0.77	0.68	0.46
-5	0.10	0.85	0.76	0.65	0.49
-5	0.20	0.87	0.75	0.65	0.47
-3	0.00	0.90	0.81	0.70	0.45
-3	0.05	0.88	0.77	0.66	0.44
-3	0.10	0.89	0.77	0.67	0.46
-3	0.20	0.89	0.80	0.71	0.46
-1	0.00	0.89	0.80	0.71	0.48
-1	0.05	0.90	0.79	0.69	0.50
-1	0.10	0.88	0.77	0.67	0.47
-1	0.20	0.89	0.79	0.72	0.47

Table 8: The fraction of Monte Carlo simulations for which the true value of c was contained within the respective centered con dence intervals or below the median unbiased estimate. The data in the rst tableau is simulated 5000 times using equations $(3.2a - 3.2b)$ for two yields, $c = (-5, -3, -1, 1, 3), T = (200, 400)$, $\pi = (0.05, 0.1, 0.2, 0.3)$, and $u_t = \varepsilon_t$, $\varepsilon_t \sim N(0, \Sigma)$, $\Sigma = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 5 \end{bmatrix}$. In the second tableau, the errors are: $u_t = Au_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, \Sigma)$, $\Sigma = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 5 \end{bmatrix}$, $A = \begin{bmatrix} -0.5 & -0.5 \\ 0.05 & -1.0 \end{bmatrix}$. For a given experiment (c, π, T), if more than one percent of the test values could not be inverted (outside of Stock s tables), the experiment was omitted from the table.

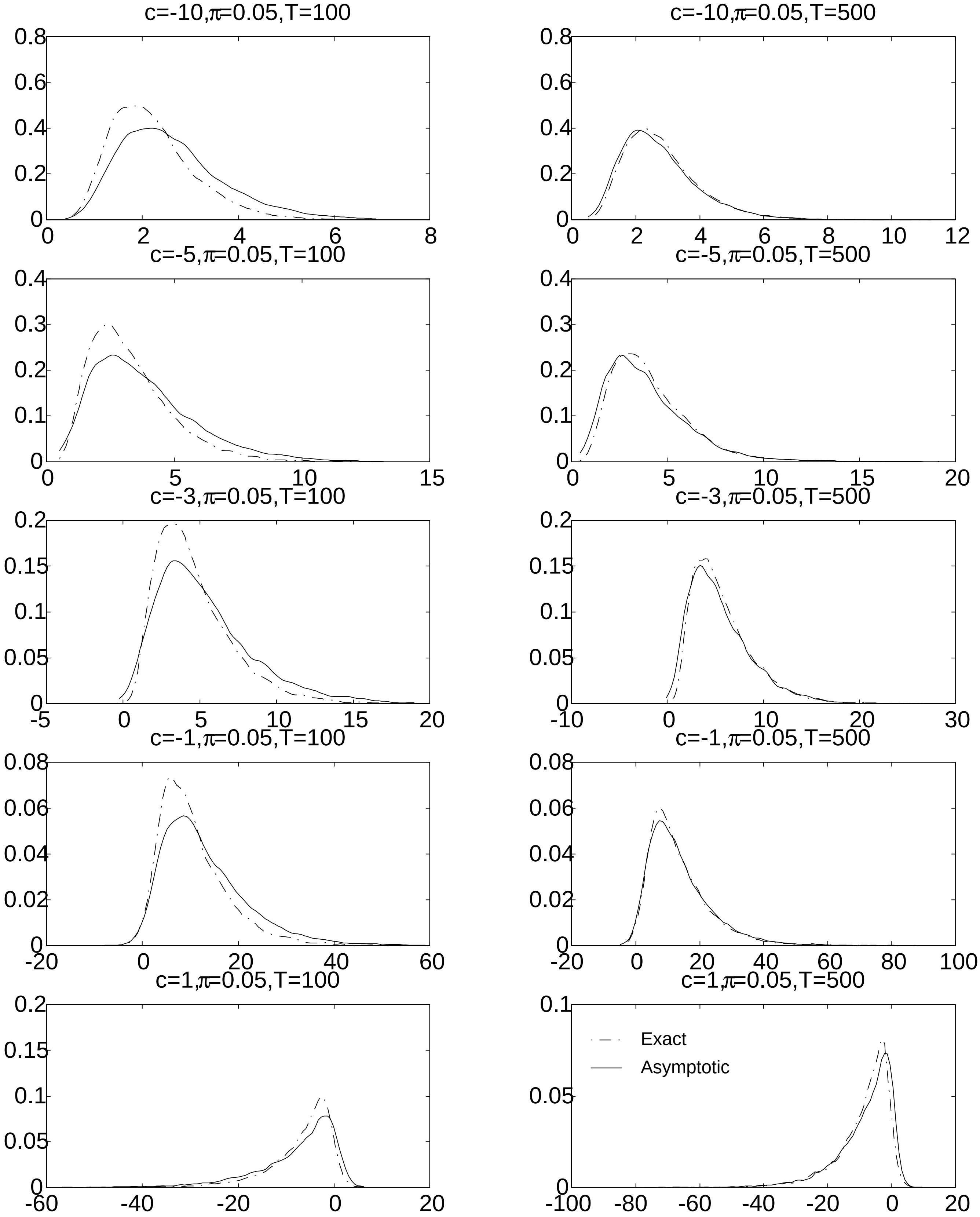


Figure 1a: Comparison of the empirical distributions of $\hat{\beta}$ and the simulated asymptotic distributions for various (c, π, T)

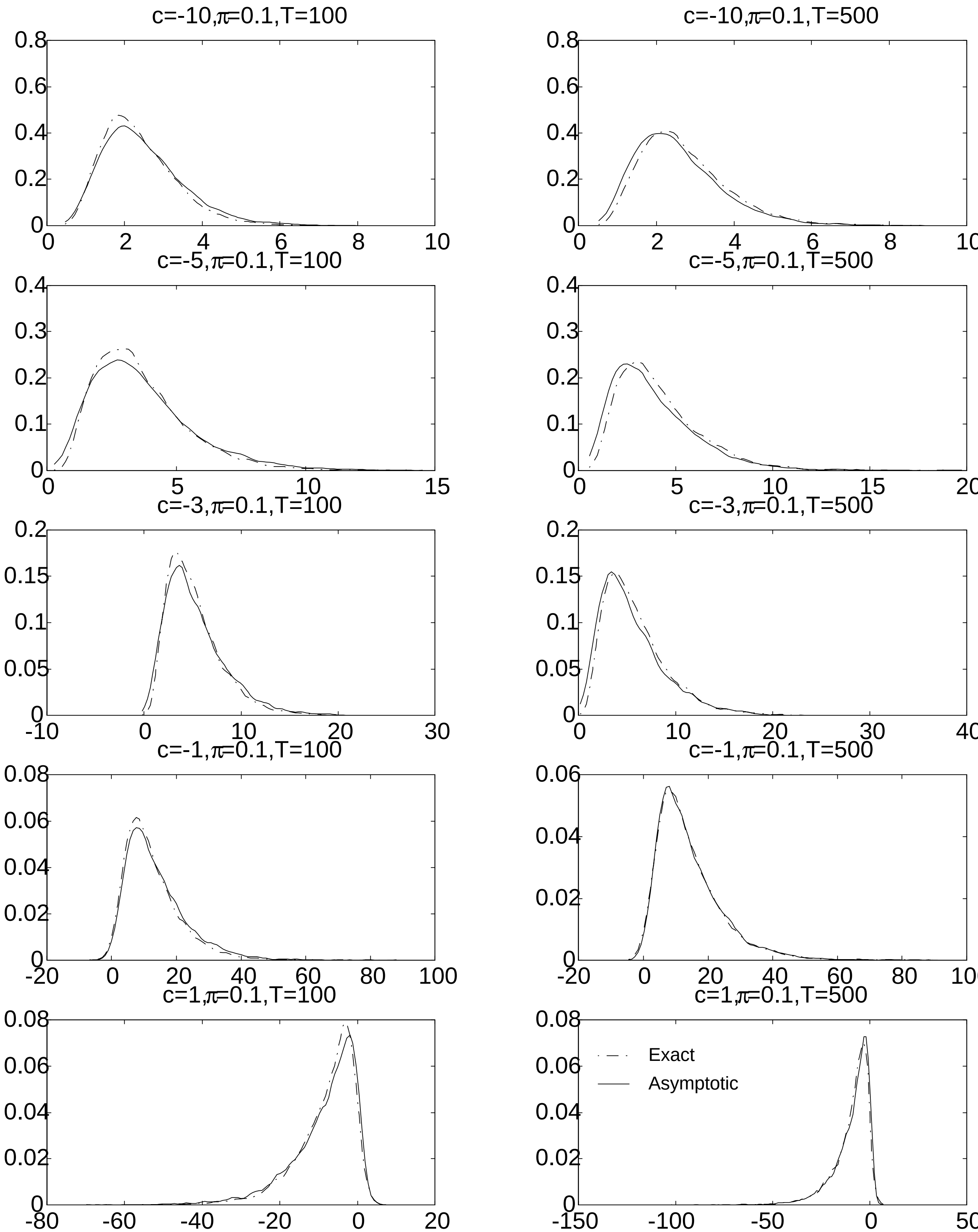


Figure 1b: Comparison of the empirical distributions of $\hat{\beta}$ and the simulated asymptotic distributions for various (c, π, T)

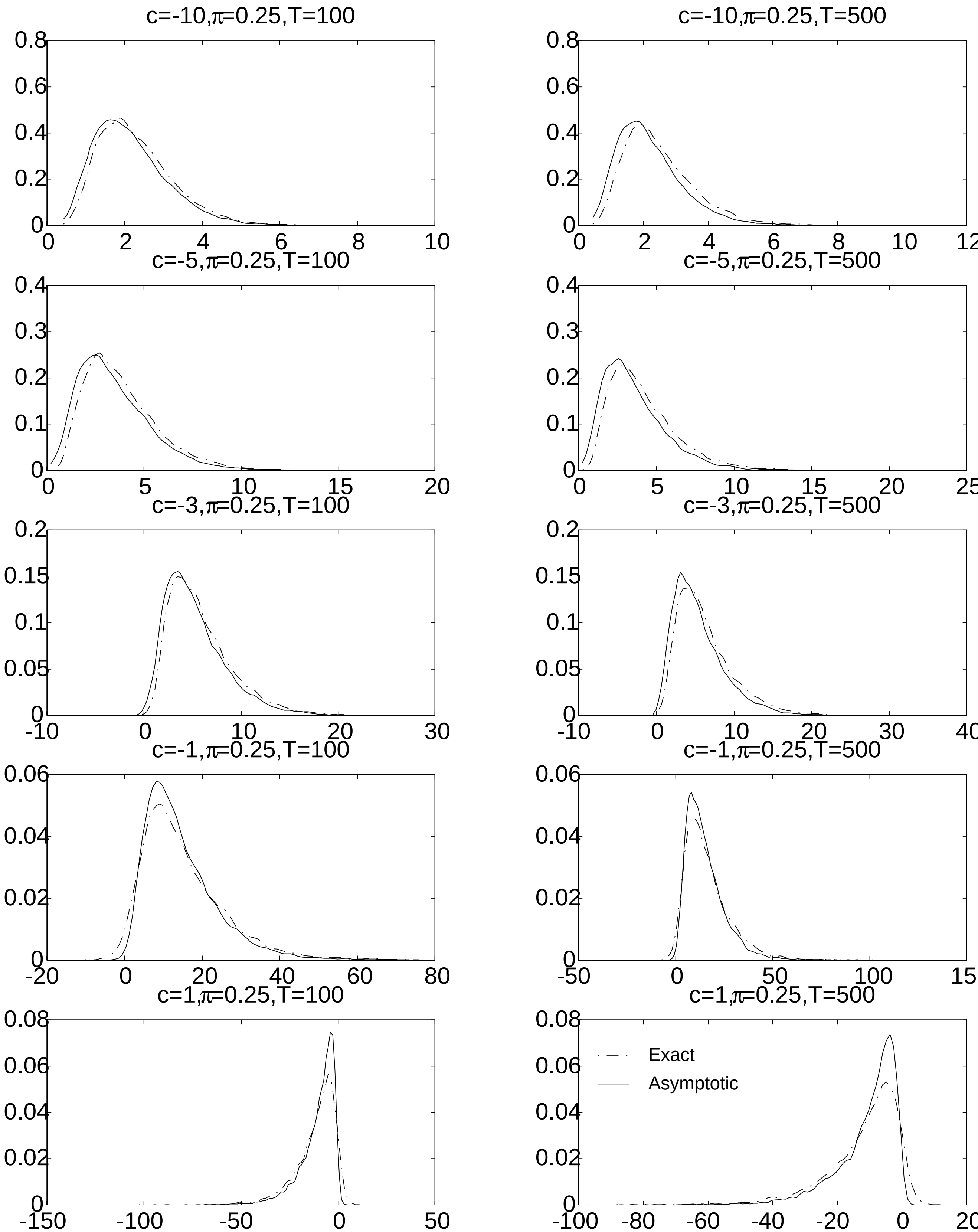
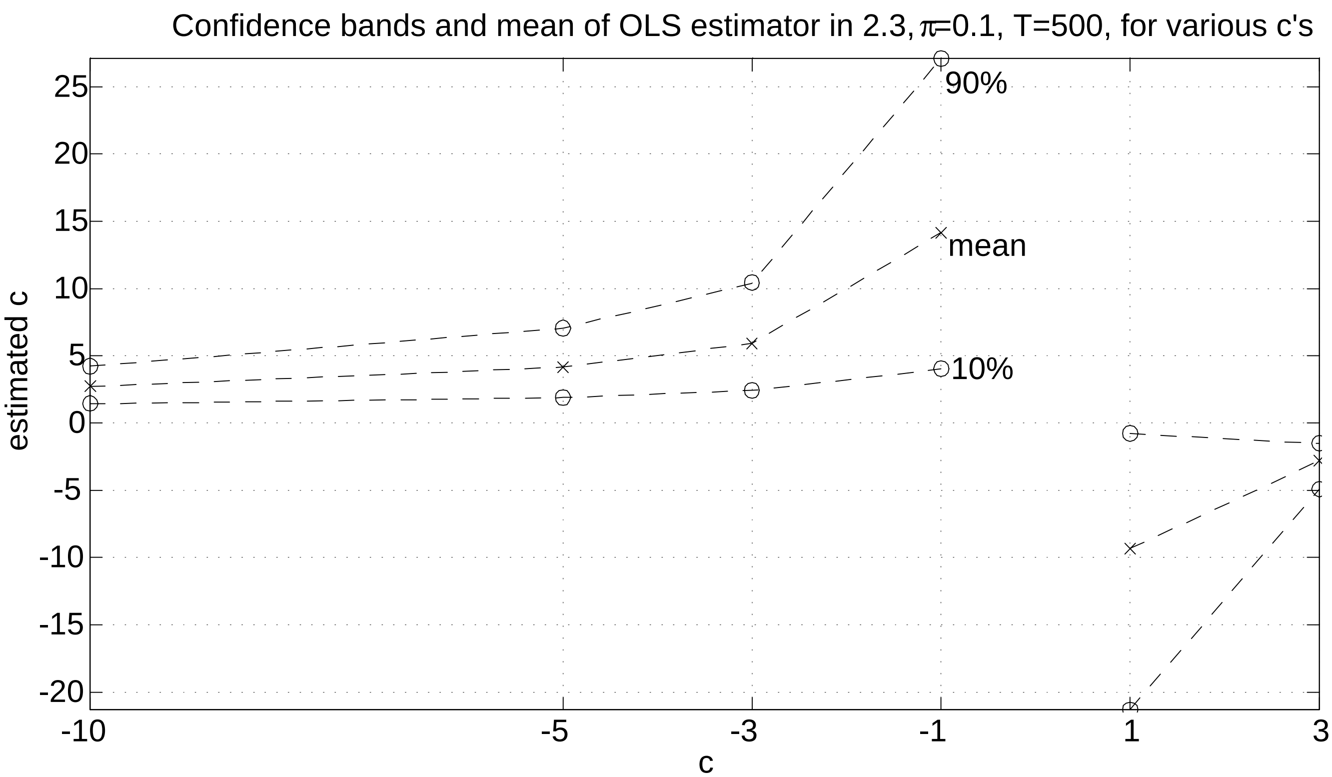


Figure 1c: Comparison of the empirical distributions of $\hat{\beta}$ and the simulated asymptotic distributions for various (c, π, T)



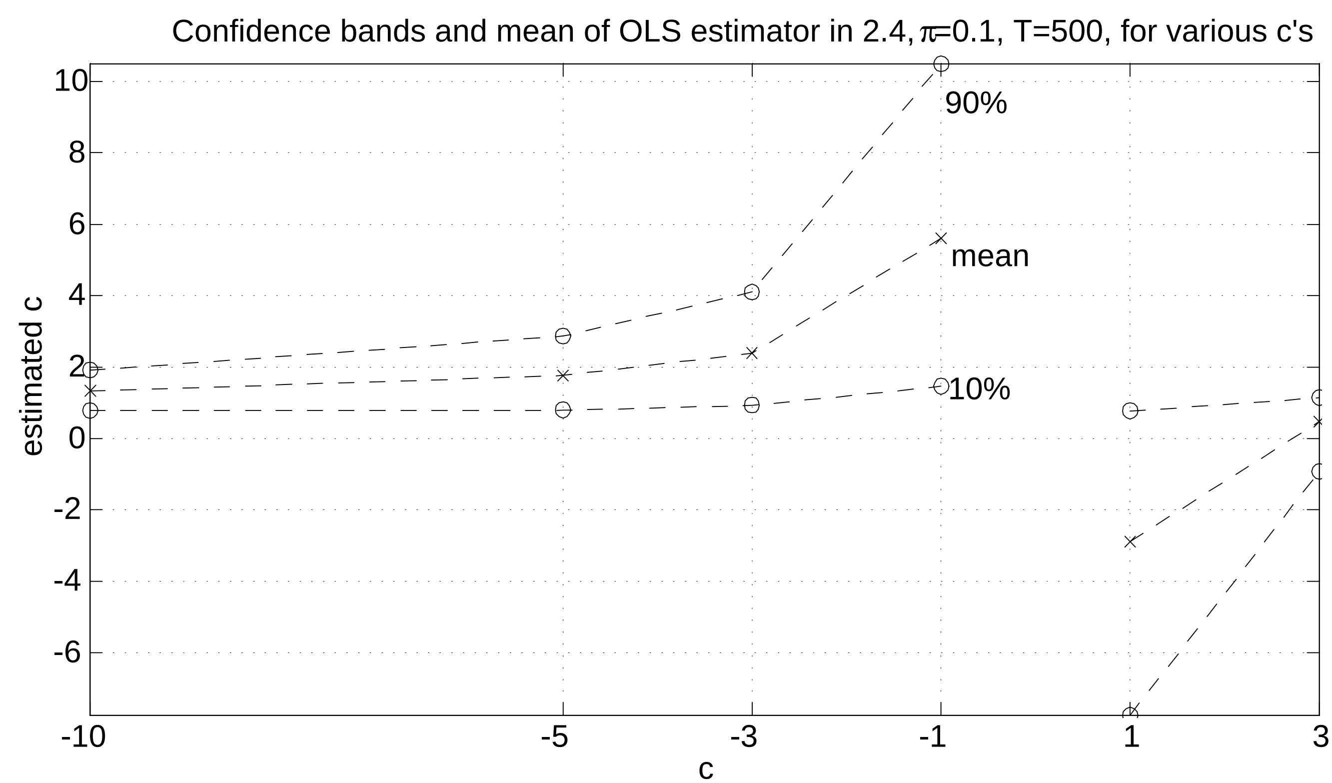


Figure 2: Mean, 10th, and 90th percentiles of the asymptotic distributions of $\hat{\beta}$ and $\hat{\psi}$ for $\pi = 0.1$, $T = 500$, and $c = (-5, -3, -1, 1, 3)$. The intervals are simulated 5000 times from rescaled partial sums of $y_{1,t}$.

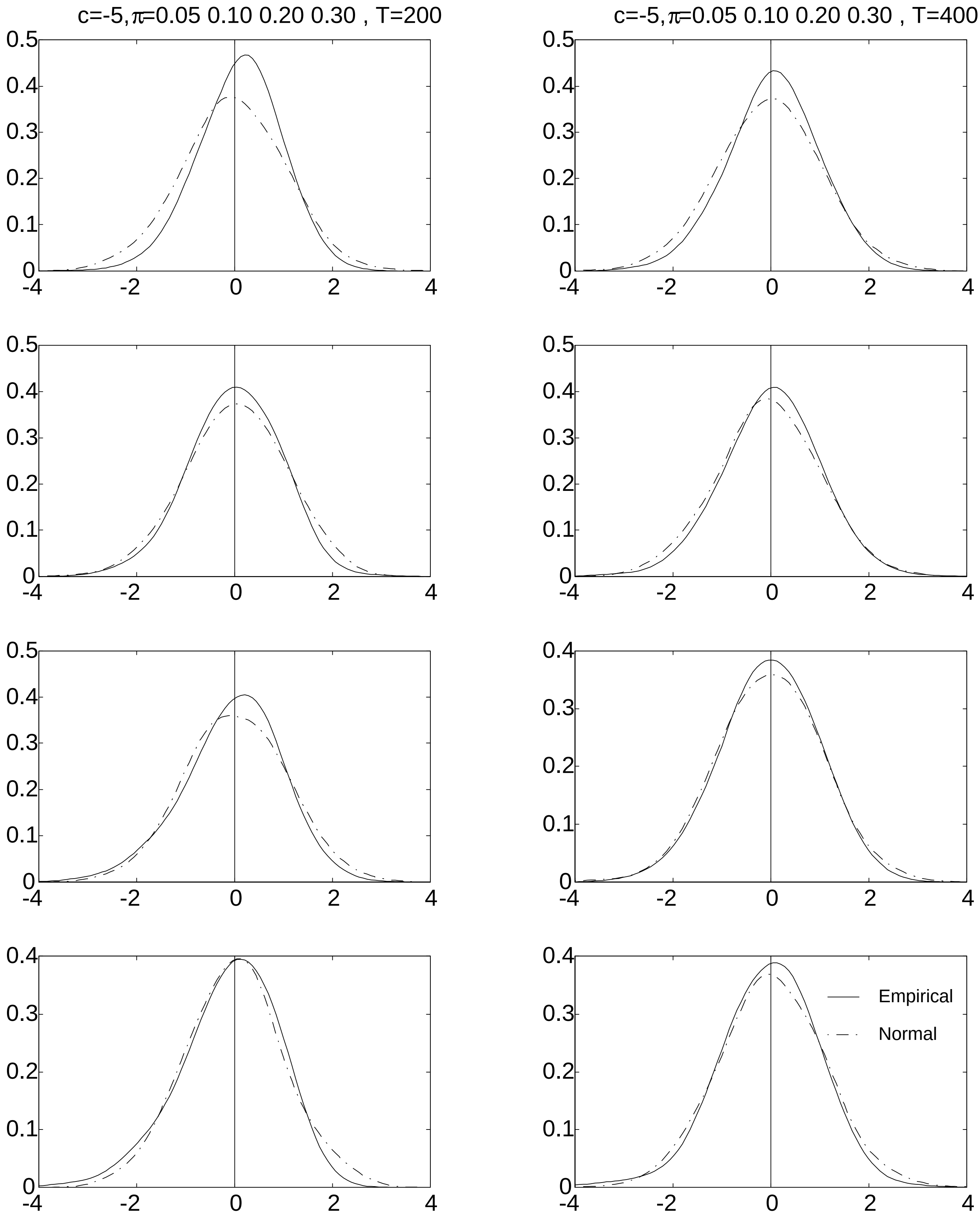


Figure 3a: Empirical distributions of $\widetilde{c}$ computed by QDOLS. The data are simulated from equations $(3.2a - 3.2b)$ for two yields, using $c = -5$, T=$(200, 400)$, $\pi = (0.05, 0.1, 0.2, 0.3)$, $u_t = \varepsilon_{t,}$ where $\varepsilon_t \sim N(0, \Sigma)$, and $\Sigma = \begin{bmatrix} 0.4 & 0.5 \\ 0.3 & 0.4 \end{bmatrix}$. The empirical distributions of $\widetilde{c}$, standardized by their mean and standard deviation are compared to N(0,1).

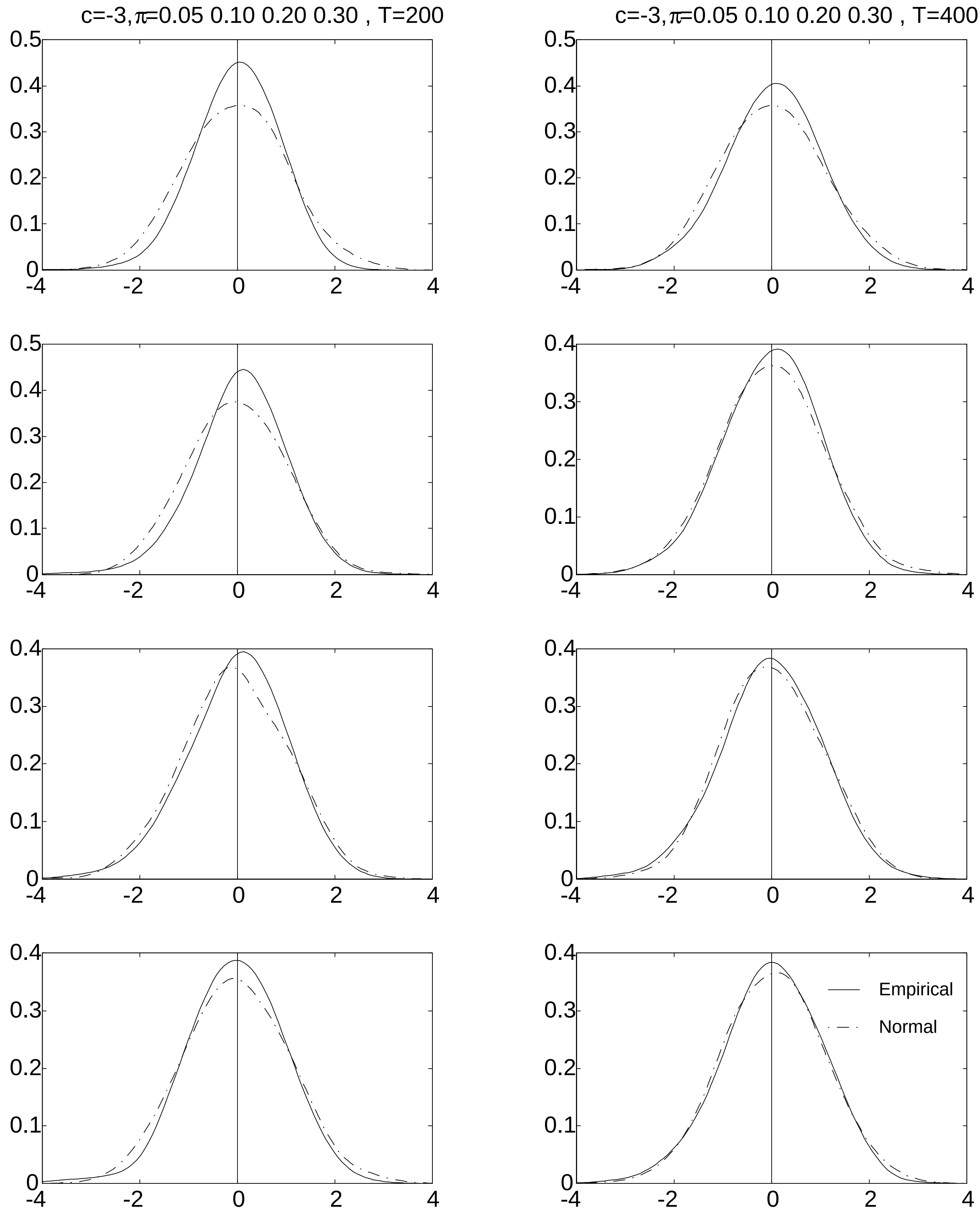


Figure 3b: Empirical distributions of $\widetilde{c}$ computed by QDOLS. The data are simulated from equations $(3.2a - 3.2b)$ for two yields, using $c = -3$, T=$(200, 400)$, $\pi = (0.05, 0.1, 0.2, 0.3)$, $u_t = \varepsilon_t,$ where $\varepsilon_t \sim N(0, \Sigma)$, and $\Sigma = \begin{bmatrix} 0.4 & 0.5 \\ 0.3 & 0.4 \end{bmatrix}$. The empirical distributions of $\widetilde{c}$, standardized by their mean and standard deviation are compared to N(0,1).

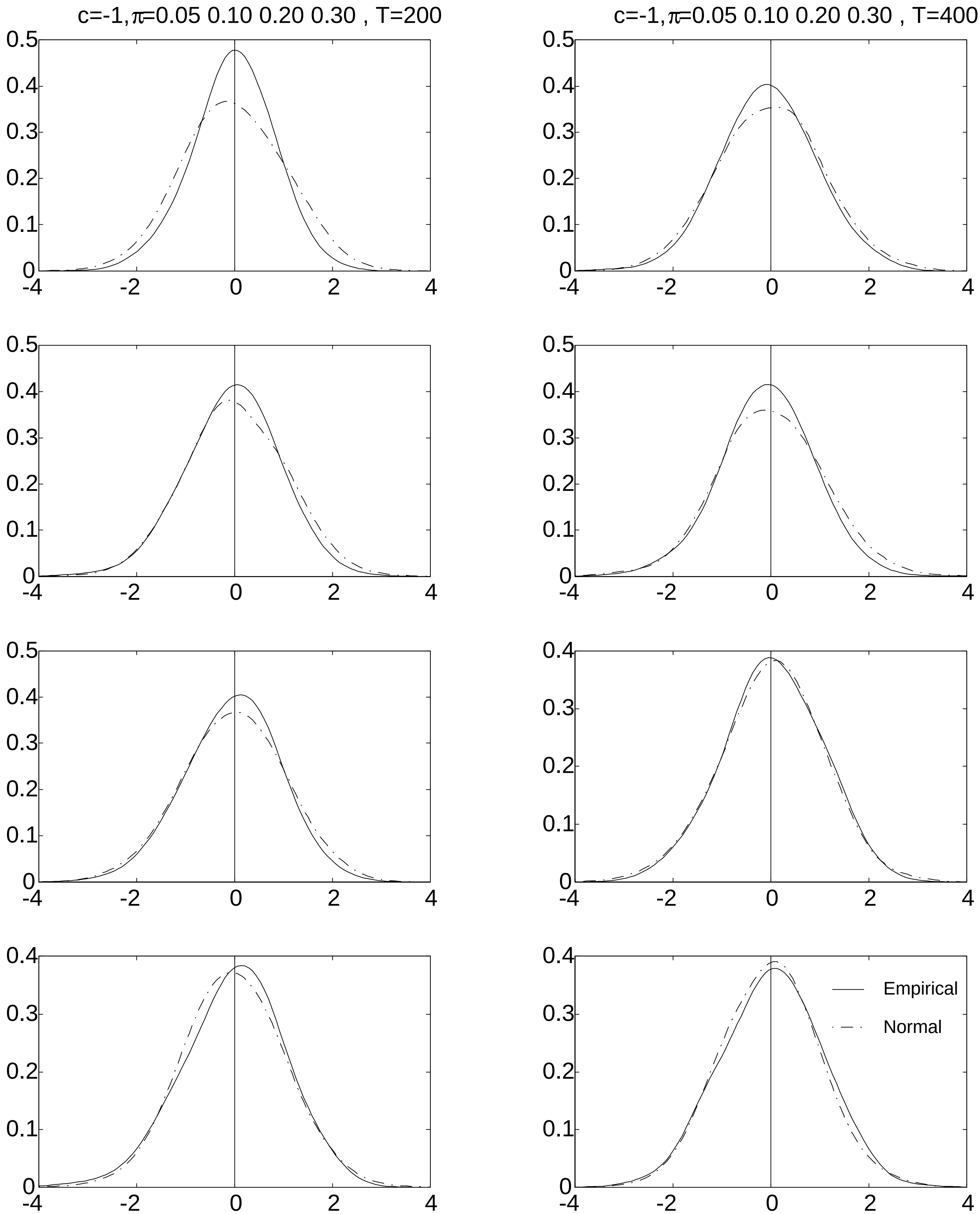


Figure 3c: Empirical distributions of $\widetilde{c}$ computed by QDOLS. The data are simulated from equations $(3.2a - 3.2b)$ for two yields, using $c = -1$, T=$(200, 400)$, $\pi = (0.05, 0.1, 0.2, 0.3)$, $u_t = \varepsilon_{t,}$ where $\varepsilon_t \sim N(0, \Sigma)$, and $\Sigma = \begin{bmatrix} 0.4 & 0.5 \\ 0.3 & 0.4 \end{bmatrix}$. The empirical distributions of $\widetilde{c}$, standardized by their mean and standard deviation are compared to N(0,1).

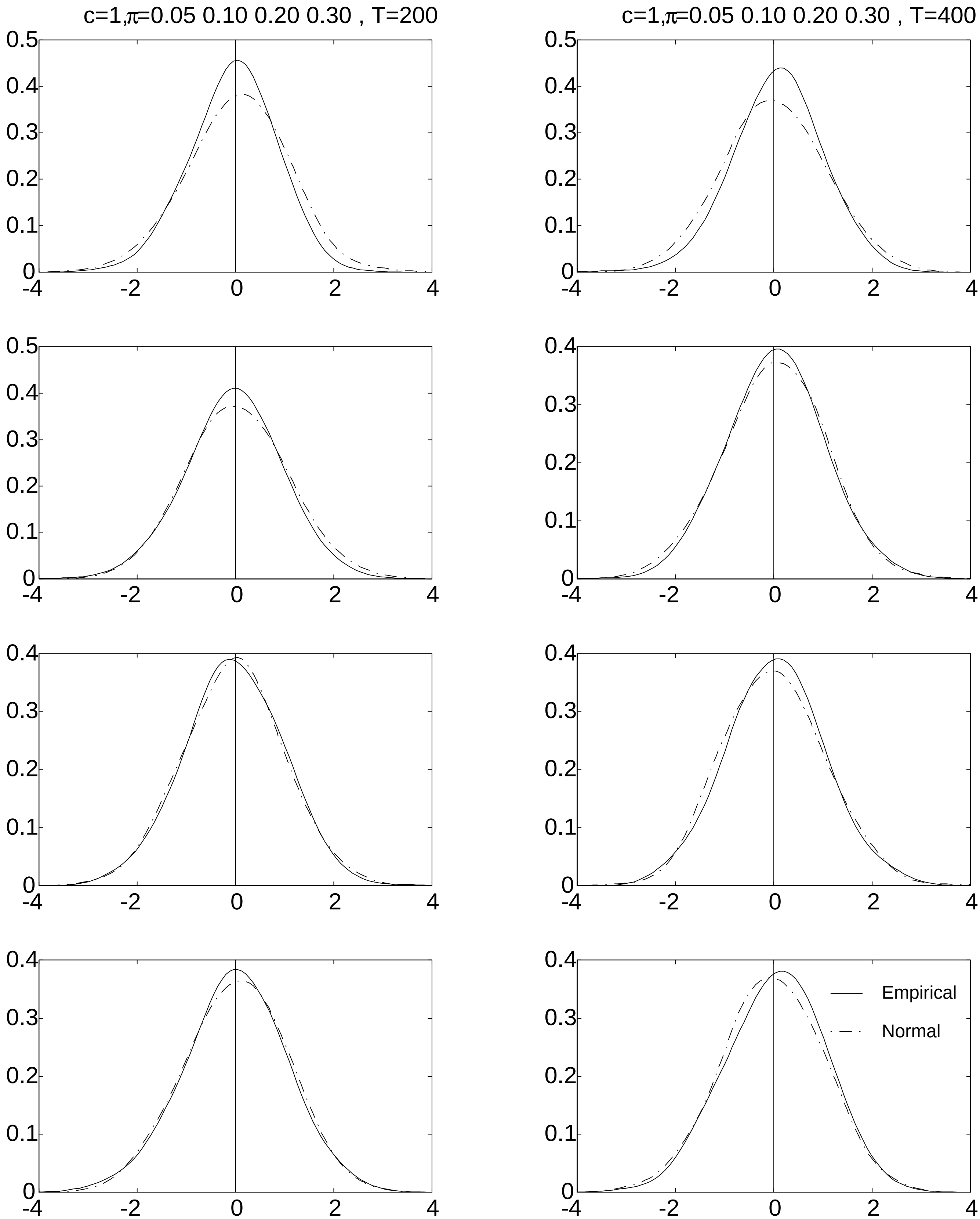


Figure 3d: Empirical distributions of $\widetilde{c}$ computed by QDOLS. The data are simulated from equations $(3.2a-3.2b)$ for two yields, using $c=1$, T=$(200,400)$, $\pi=(0.05,0.1,0.2,0.3)$, $u_t=\varepsilon_t$, where $\varepsilon_t \sim N(0,\Sigma)$, and $\Sigma = \begin{bmatrix} 0.4 & 0.5 \\ 0.3 & 0.4 \end{bmatrix}$. The empirical distributions of $\widetilde{c}$, standardized by their mean and standard deviation are compared to N(0,1).

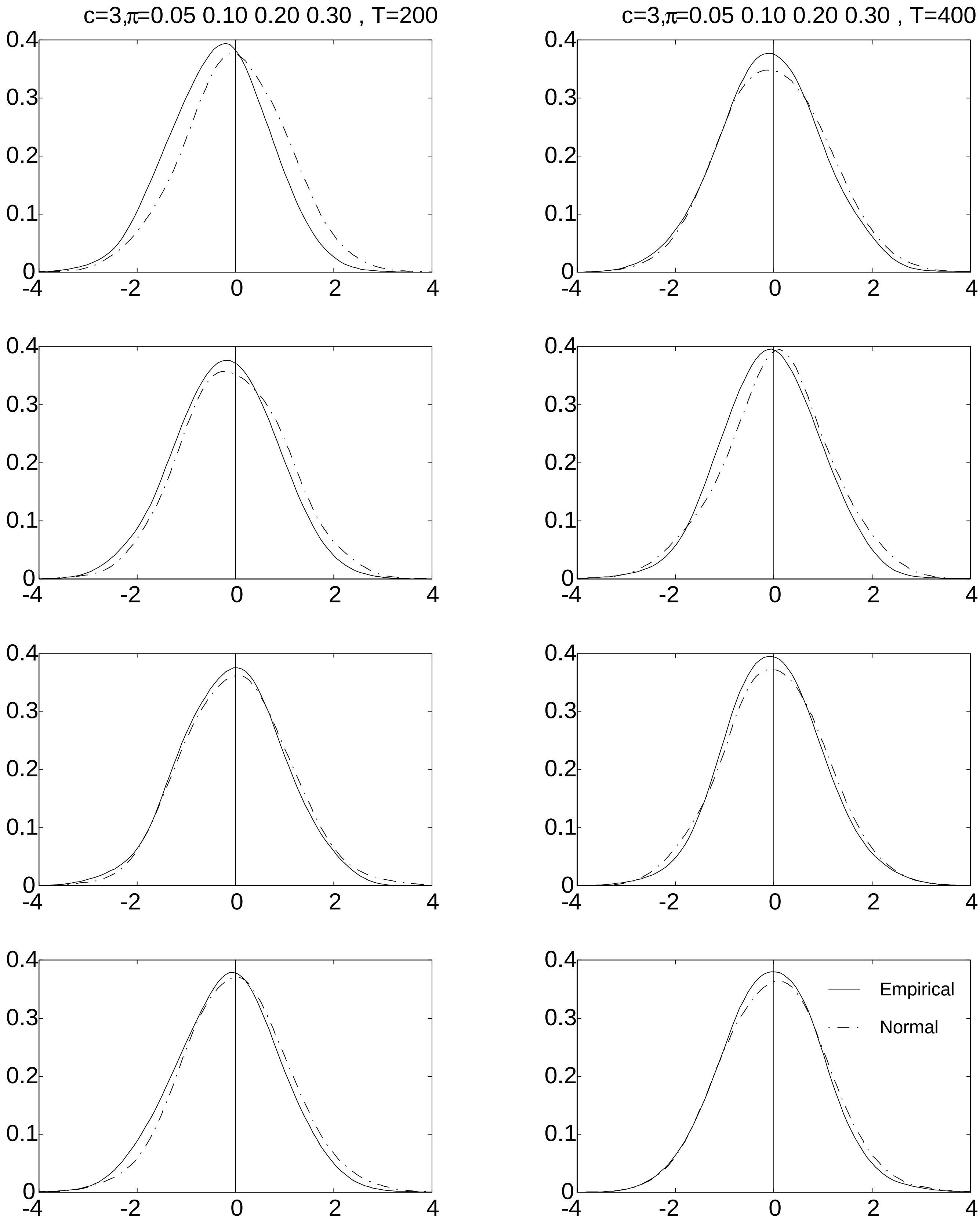


Figure 3e: Empirical distributions of $\widetilde{c}$ computed by QDOLS. The data are simulated from equations $(3.2a - 3.2b)$ for two yields, using $c = 3$, T=$(200, 400)$, $\pi = (0.05, 0.1, 0.2, 0.3)$, $u_t = \varepsilon_{t,}$ where $\varepsilon_t \sim N(0, \Sigma)$, and $\Sigma = \begin{bmatrix} 0.4 & 0.5 \\ 0.3 & 0.4 \end{bmatrix}$. The empirical distributions of $\widetilde{c}$, standardized by their mean and standard deviation are compared to N(0,1).

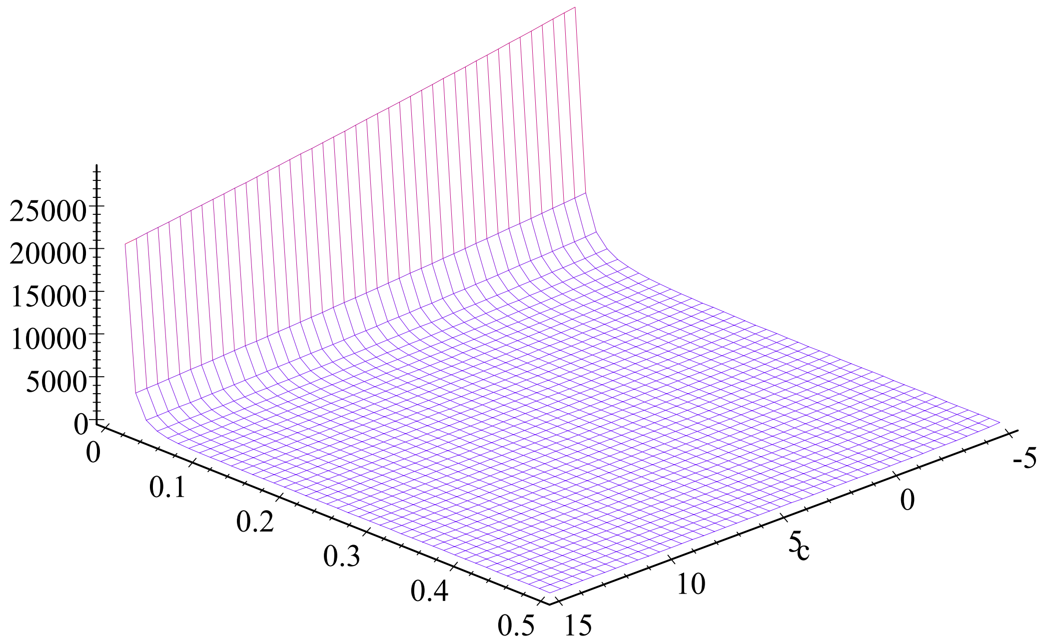


Figure 4: Plot of $D^2(c) = \left(\frac{c^2\pi}{e^{c\pi}c\pi - e^{c\pi}+1}\right)^2 \approx \frac{4}{\pi^2}$ for $c \in (-5, 15) \setminus \{0\}$ and $\pi \in (0, 0.5)$. If the system $(3.2a - 3.2b)$ is 2-dimensional, then $D^2(c)$ is a scalar. As π approaches 0, $D^2(c)$ diverges, implying that the variance of estimates of c in the short end of the term structure will have higher variance than estimates in the long end.

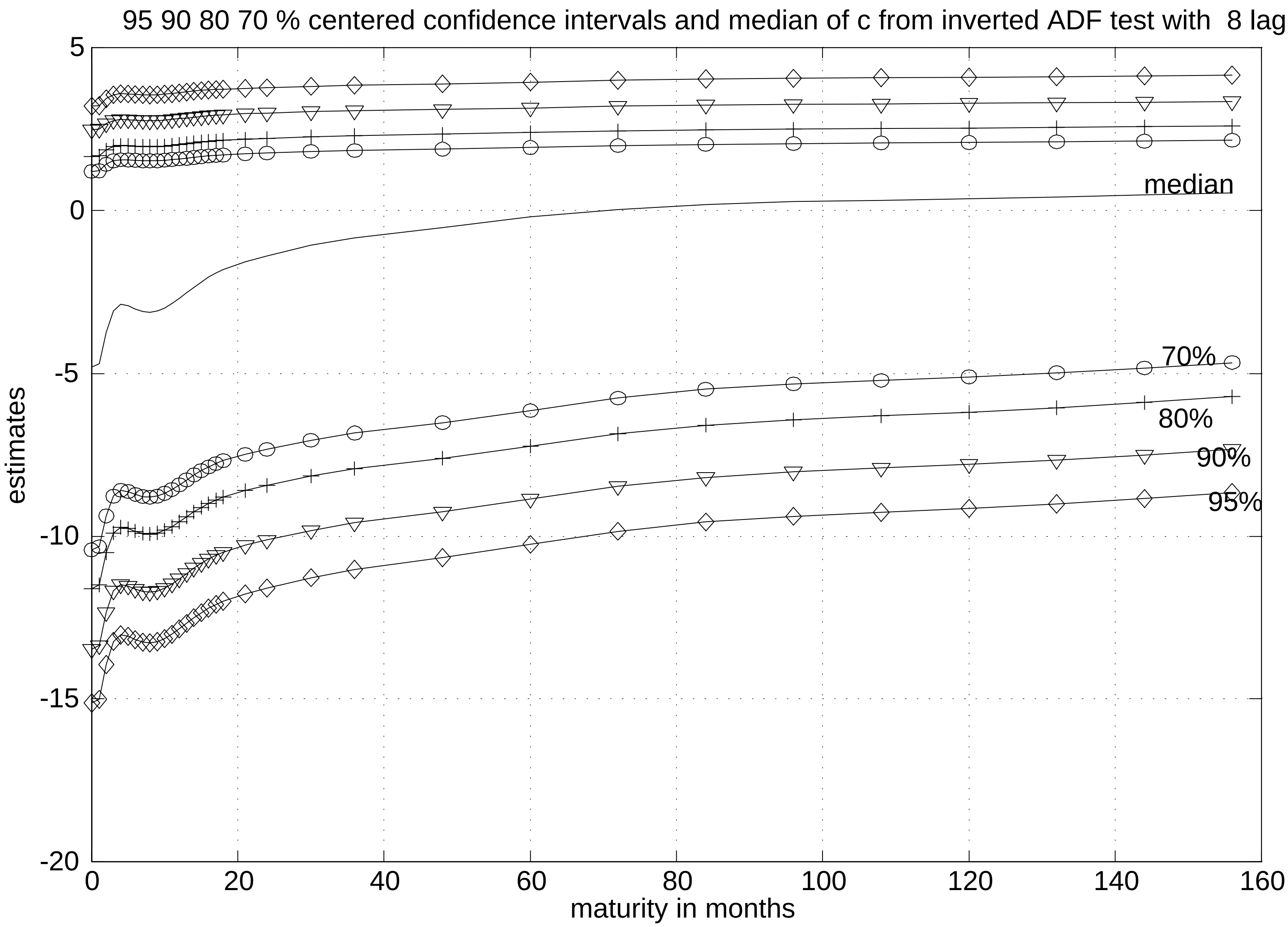


Figure 5a: Median unbiased estimates of c and centered confidence intervals found by inverting an ADF statistic, using the yield data from McCulloch and Kwon (1993) for zero-coupon bonds of maturities from 1 month to 13 years. Monthly data for the period 1952:1–1991:2. The inversion of the statistic was performed by linear interpolation from the tables in Stock (1991). The ADF test is specified with 8 lags, chosen with sequential t-tests.

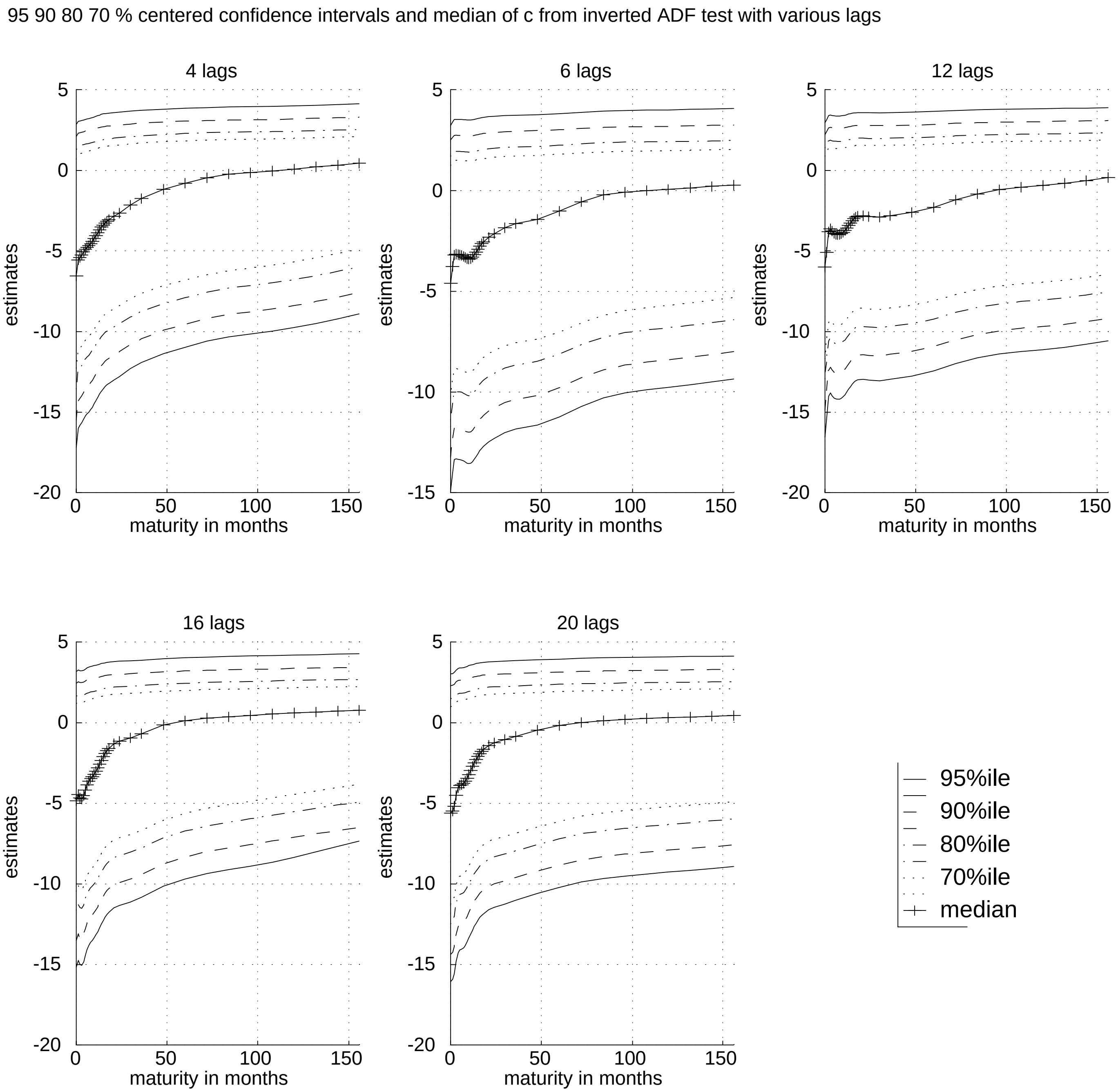


Figure 5b: Robustness analysis of the median unbiased estimation. The ADF test is specified with various lag structures, from 4 to 20 lags.

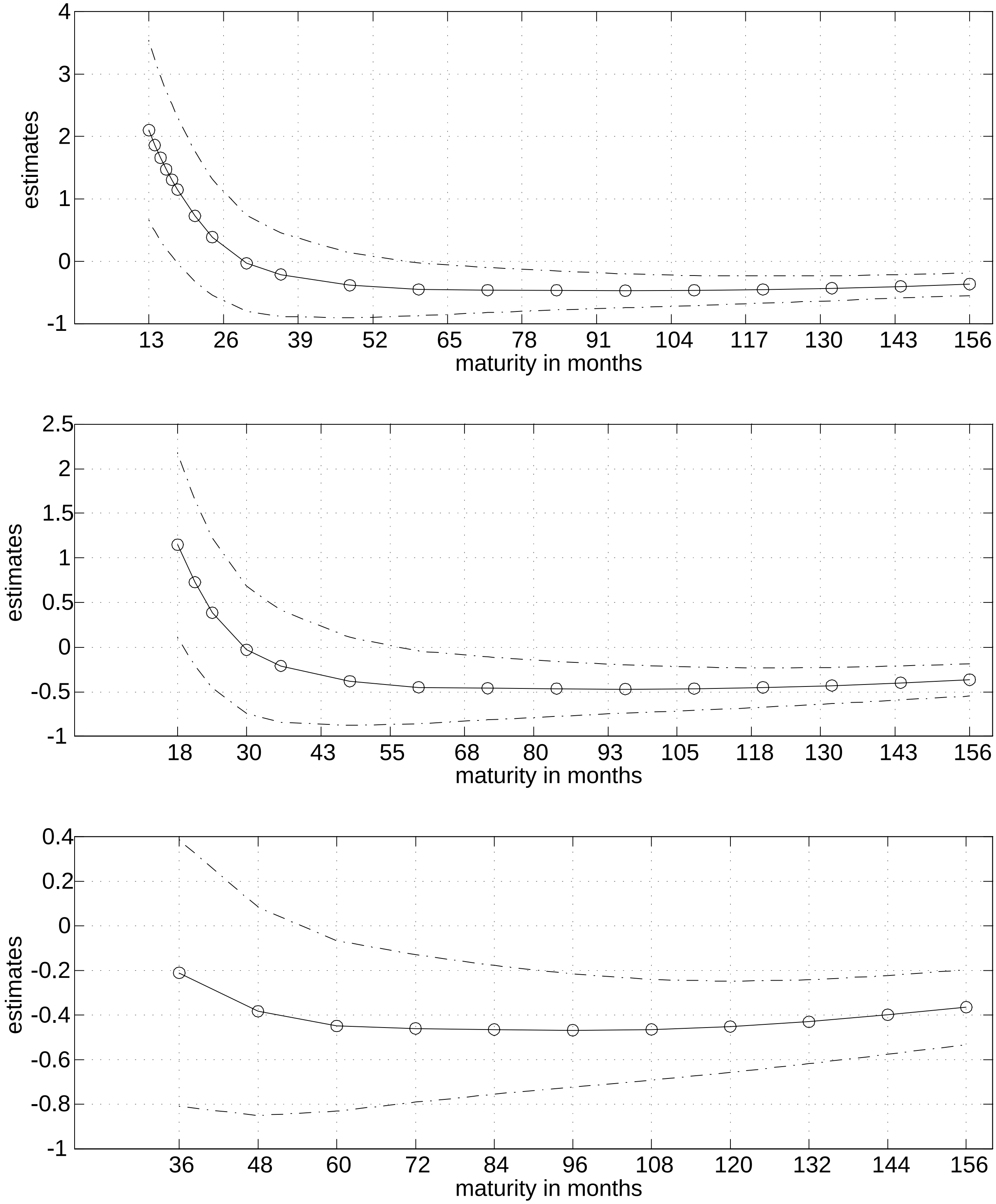


Figure 6a: Consistent and asymptotically unbiased estimate of $\underline{c}$ using QDOLS, plotted for the respective maturities, using the yield data from McCulloch and Kwon (1993). Error bands are computed using ± 2 standard deviations. The first plot uses bonds with maturity higher than 1 year, the second with maturity of 18 months or higher, and the third with maturity of 3 or more years. The QDOLS uses 3 leads and 3 lags, chosen by sequential t-tests.

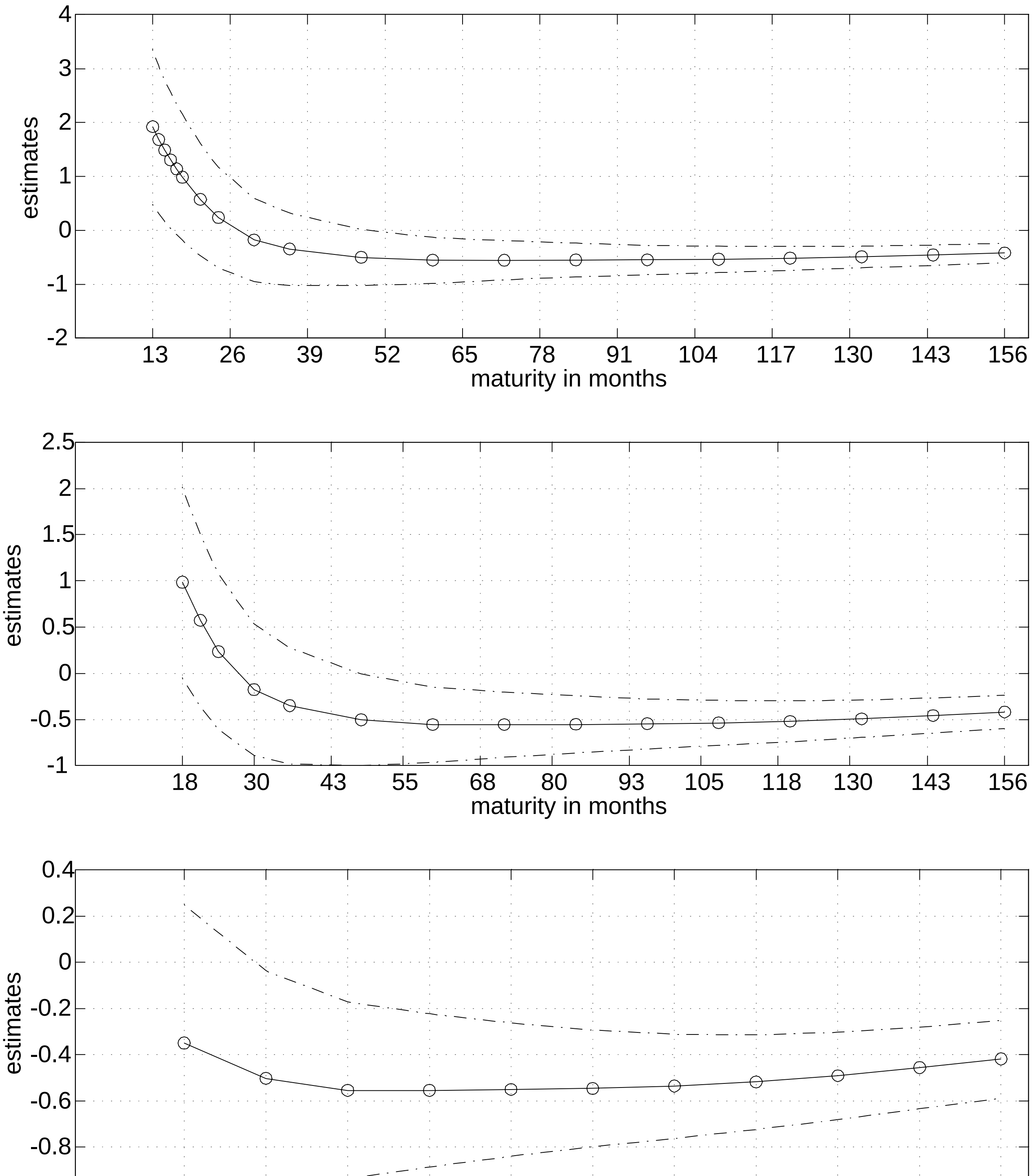


Figure 6b: Robustness analysis for the consistent and asymptotically unbiased estimate of $\underline{c}$ from figure 6a. The results do not change significantly, when the QDOLS is estimated using different lead/lag structures. The first plot uses bonds with maturity higher than 1 year, the second with maturity of 18 months or higher, and the third with maturities of 3 or more years. The QDOLS is estimated with 2 leads and 2 lags.

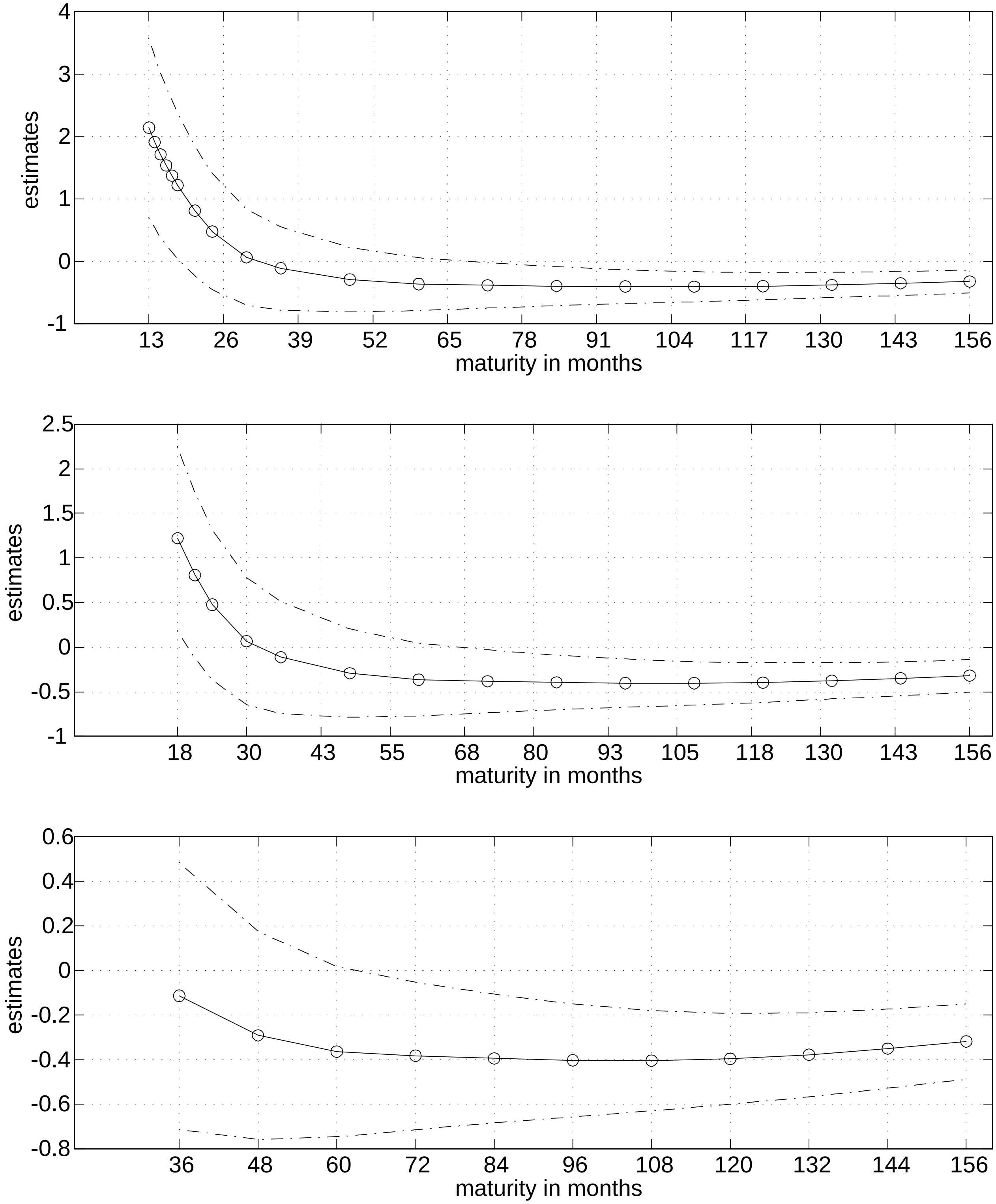


Figure 6c: Robustness analysis for the consistent and asymptotically unbiased estimate of $\underline{c}$ from figure 6a. The results do not change significantly, when the QDOLS is estimated using different lead/lag structures. The first plot uses bonds with maturity higher than 1 year, the second with maturity of 18 months or higher, and the third with maturities of 3 or more years. The QDOLS is estimated with 4 leads and 4 lags.

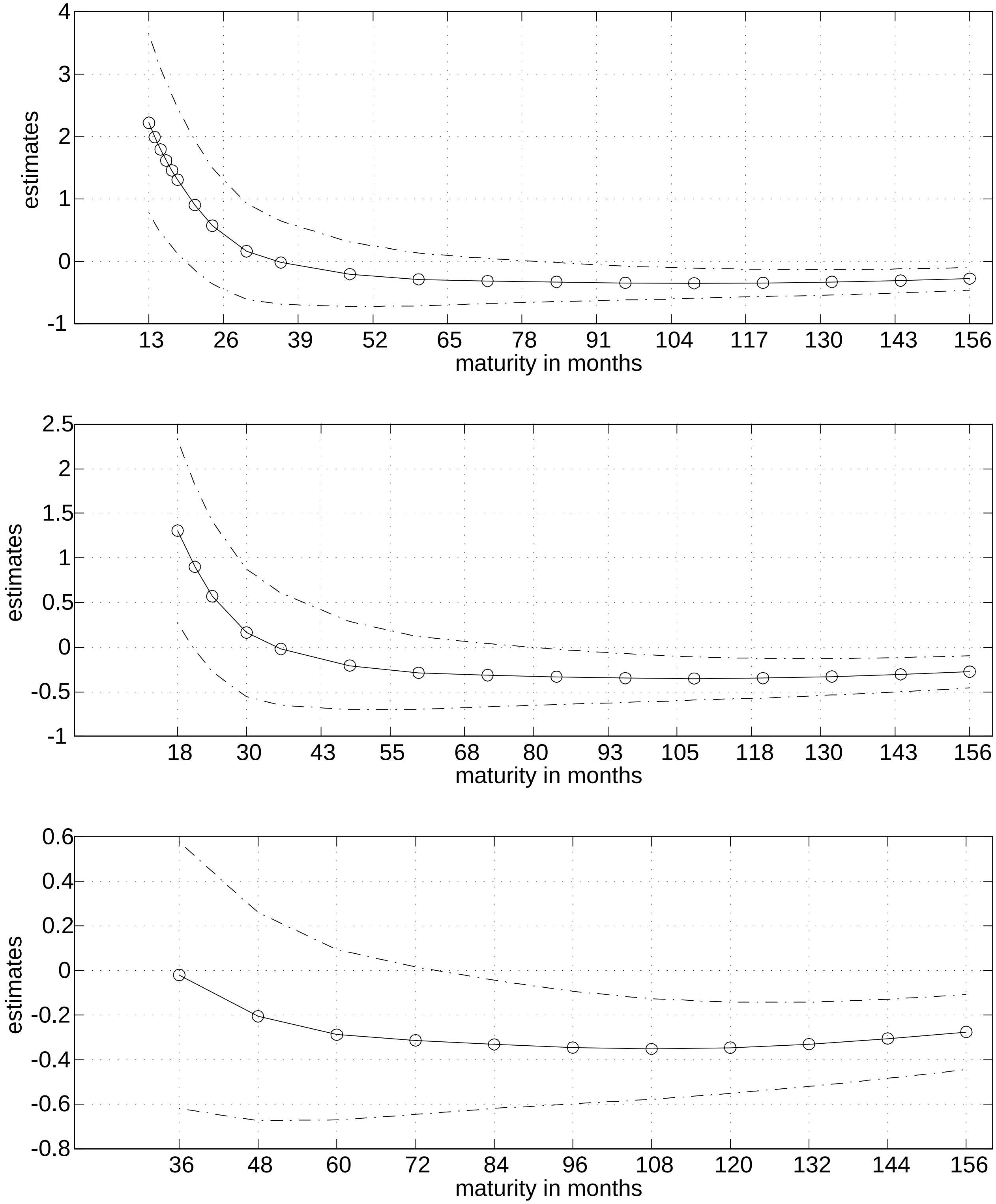


Figure 6d: Robustness analysis for the consistent and asymptotically unbiased estimate of $\underline{c}$ from figure 6a. The results do not change significantly, when the QDOLS is estimated using different lead/lag structures. The first plot uses bonds with maturity higher than 1 year, the second with maturity of 18 months or higher, and the third with maturities of 3 or more years. The QDOLS is estimated with 5 leads and 5 lags.