Online Appendix to "Correcting Estimation Bias in Dynamic Term Structure Models"

Michael D. Bauer[*], Glenn D. Rudebusch[†], Jing Cynthia Wu[‡]

May 4, 2012

A   Bootstrap bias correction

The bootstrap has become a common method for correcting small-sample mean bias.[1] Here we detail the steps for bias-corrected estimation of the VAR specified in the paper.

Denote the demeaned observations by $\tilde{X}_t = X_t - T^{-1} \sum_{i=1}^{T} X_i$, and let B denote the number of bootstrap samples. The algorithm for mean bias correction using the bootstrap is as follows:

1. Estimate the model by OLS and save the OLS estimates $\hat{\theta} = \text{vec}(\hat{\Phi})$ and the residuals. Set $b = 1$.

2. Generate bootstrap sample b using the residual bootstrap: Resample the OLS residuals, denoting the bootstrap residuals by u_t^*. Randomly choose a starting value among the T observations. For $t > 1$, construct the bootstrap sample using $\tilde{X}_t^* = \hat{\Phi} \tilde{X}_{t-1}^* + u_t^*$.

3. Calculate the OLS estimates on bootstrap sample b and denote it by $\hat{\theta}_b^*$.

4. If $b < B$ then increase b by one and return to step two.

5. Calculate the average over all samples as $\bar{\theta}^* = B^{-1} \sum_{b=1}^{B} \hat{\theta}_b^*$.

6. Calculate the bootstrap bias-corrected estimate as

$$\tilde{\theta}^B = \hat{\theta} - \left[\bar{\theta}^* - \hat{\theta} \right] = 2\hat{\theta} - \bar{\theta}^*.$$

For large B, the estimated bias $\bar{\theta}^* - \hat{\theta}$ will be close to $b_T(\hat{\theta})$. The motivation for this approach comes from the fact that $E(b_T(\hat{\theta}_T)) = b_T(\theta_0) + O(T^{-2})$, thus we can reduce the bias to order T^{-2} by using this bias correction (Horowitz, 2001).

---

[*]Federal Reserve Bank of San Francisco, michael.bauer@sf.frb.org
[†]Federal Reserve Bank of San Francisco, glenn.rudebusch@sf.frb.org
[‡]Booth School of Business, University of Chicago, cynthia.wu@chicagobooth.edu
[1]For a detailed exposition see Hall (1992) or Efron and Tibshirani (1993, Chapter 10); for a review of the bootstrap including its application to bias correction, refer to Horowitz (2001).

If the bias were constant in a neighborhood around θ_0 that contains $\hat{\theta}$, this procedure would eliminate the bias (up to simulation error), which prompted MacKinnon and Smith (1998) to call this a "constant-bias-correcting" (CBC) estimator. In general, however, the bias function is not constant, thus the bootstrap will systematically get the bias estimate wrong. Differently put, this method only removes first order bias. To obtain higher accuracy and remove high-order bias, one can use the iterated bootstrap (Hall, 1992), but the computational burden quickly becomes prohibitively costly.

B   Indirect inference bias correction

The idea of bias-correcting indirect inference estimation of θ_0 is to choose that parameter value which yields a distribution of the OLS estimator with a mean equal to the OLS estimate in the actual data (Gourieroux et al., 2000). Define $g_T(\theta) = E_\theta(\hat{\theta}_T)$, the mean of the OLS estimator if the data are generated under θ. The bias-corrected estimator of θ_0 is the value of θ that solves

$$g_T(\theta) = \hat{\theta}_T. \tag{1}$$

We denote this bias-corrected estimator by $\tilde{\theta}_T$. MacKinnon and Smith (1998) call this estimator a "nonlinear-bias-correcting" (NBC) estimator, and point out that it is not unbiased. We have

$$E_{\theta_0}(\tilde{\theta}_T) = E_{\theta_0}(g_T^{-1}(\hat{\theta}_T)) \neq g_T^{-1}(E_{\theta_0}(\hat{\theta}_T)) = g_T^{-1}(g_T(\theta_0)) = \theta_0,$$

except for the unlikely special case that the bias function is linear, since the expectation operator does not go through nonlinear functions.

For identifiability, i.e., for a unique solution of 1 to exist, we need $g_T(\cdot)$ to be uniformly continuous and one-to-one (injective) in a neighborhood around θ_0 that includes $\hat{\theta}_T$. Since the function $g_T(\theta)$ is not known analytically, a proof that these conditions are fulfilled is not possible. However, intuition and simulation exercises (not shown) suggest that these conditions are likely to be satisfied in the context of a stationary VAR.

Since the function $g_T(\theta)$ is not known analytically, it is evaluated by means of simulation. We use a residual bootstrap for this purpose, so that we do not have to make distributional assumptions about the error term of the VAR. Define $R(\theta) = \hat{\theta} - g_T(\theta)$. The bias-corrected estimates are given by the root of this function, i.e., by $R(\tilde{\theta}) = 0$.[2] Because measurements of this function are obtained by simulation and are contaminated with noise, classical root-finding methods are not applicable.

We now detail our algorithm, which provides a fast and reliable way to calculate the indirect inference estimator with low computational cost even if $dim(\theta)$ is large. Finding the root of a function that is measured with error is a problem in the area of "stochastic approximation" (SA), pioneered by Robbins and Monro (1951). Their crucial insight was that for each attempted value of θ we do not need a very precise measurement of $R(\theta)$, because it is only used to lead us in the right direction. For our application that means that a small number of bootstrap replications is sufficient in each iteration, which greatly lowers our computational

---

[2]The conditions mentioned above for $g_T(\cdot)$ imply that $R(\theta) = 0$ has a unique solution.

cost. The basic stochastic approximation algorithm is to construct a sequence according to

$$\theta^{(j+1)} = \theta^{(j)} + \alpha^{(j)} Y^{(j)}, \tag{2}$$

where $\alpha^{(j)}$ is a deterministic scalar sequence and $Y^{(j)}$ is a noisy measurement of $R(\theta^{(j)})$. Under some specific conditions about $\alpha^{(j)}$, the sequence will converge to $\tilde{\theta}$. However, the sequence of averages, $\bar{\theta}^{(j)} = j^{-1} \sum_{i=1}^{j} \theta^{(i)}$, converges even if $\alpha^{(j)}$ is taken to be a constant (between zero and one) and it does so at an optimal rate (Polyak and Juditsky, 1992). Under some rather weak conditions on $R(\cdot)$, $\alpha^{(j)}$ and the measurement error, we have $\bar{\theta}^{(j)} \to \tilde{\theta}$ almost surely, $\sqrt{j}$-asymptotic normality, as well as optimality in the sense of a maximum rate of convergence.[3]

Motivated by these results, we use the following algorithm:

1. Choose as a starting value $\theta^{(1)} = \hat{\theta}$. Set $j = 1$.

2. Using $\theta^{(j)}$, obtain a measurement $Y^{(j)}$: estimate $g_T(\theta^{(j)})$ using a residual bootstrap with B replications (for details, see below) and set $Y^{(j)}$ equal to the difference between $\hat{\theta}$ and this estimate.

3. Calculate $\theta^{(j+1)}$ using equation (2).

4. If $j < N_0 + N_1$ increase j by one and return to step 2.

5. Calculate the bias-corrected estimate as

$$\hat{\hat{\theta}} = N_1^{-1} \sum_{i=N_0+1}^{N_0+N_1} \theta^{(i)}.$$

In step two the approximate mean of the OLS estimator for a given $\theta^{(j)}$, i.e., an estimate of $g_T(\theta^{(j)})$, is obtained using a residual bootstrap with B replications. We randomly choose the starting values among the T observations. For $t > 1$ the bootstrapped series is obtained using $\tilde{X}_t^* = \Phi^{(j)} \tilde{X}_{t-1}^* + u_t^*$, where u_t^* are the bootstrap residuals, and $\Phi^{(j)}$ denotes the $N \times N$ matrix containing the elements of of $\theta^{(j)}$. Importantly, the bootstrap residuals have to be obtained for a given $\theta^{(j)}$: One cannot resample the original VAR residuals since these do not, together with $\theta^{(j)}$, generate the original data. Instead one has to first obtain a series of residuals $\hat{u}_t = \tilde{X}_t - \Phi^{(j)} \tilde{X}_{t-1}$, for $t > 1$, which then can be resampled in the usual way to create the bootstrap residuals u_t^*.[4] In other words, the bootstrap residuals are draws not from the empirical distribution of the original VAR residuals, but from the empirical distribution of the VAR residuals that are obtained given $\Phi^{(j)}$.

---

[3]The only assumption that needs mentioning here is that the Jacobian at the solution point needs to be a Hurwitz matrix, i.e., the real parts of the eigenvalues of $R'(\tilde{\theta})$ need to be strictly negative. Only if $R(\cdot)$ is decreasing in this sense does it make sense to increase the value of $\theta^{(j)}$ when we have positive measurements (equation 2). We check this condition by estimating the Jacobian at $\hat{\theta}$, verifying that it is Hurwitz, and relying on the assumption that this does not change between $\hat{\theta}$ and $\tilde{\theta}$. Details on how we estimate the Jacobian in this particular setting are available upon request.

[4]This notation suppresses the dependence on the bootstrap replication b and on the iteration j.

We choose $\alpha^{(j)} = 0.5$ and $B = 50$, unless otherwise specified. Instead of devising a specific exit condition which might be computationally costly to check, we simply run the algorithm for a fixed number of iterations. We do not calculate $\bar{\theta}^{(j)}$ using all iterations but instead discard the first part of the sample, corresponding to the idea of a burn-in sample in the Markov chain Monte Carlo literature. Unless otherwise specified, we use $N_0 = 1000$ iterations as a burn-in sample and then take as our estimate the average of the next $N_1 = 5000$ iterations.

To verify the convergence of the algorithm, we then check how close $\hat{\tilde{\theta}}$ is to $\tilde{\theta}$. This is feasible despite $\tilde{\theta}$ being unknown, since we can obtain a measurement of $R(\hat{\tilde{\theta}})$ with arbitrarily small noise by using a large B, and check how close it is to zero. As the distance measure, we take root mean-square distance, that is $d(a, b) = (l^{-1}(a-b)'(a-b))^{1/2}$ for two vectors a, b of equal length l. We use this distance metric because it is invariant to the dimensionality of θ. We calculate $d(R(\hat{\tilde{\theta}}), 0)$, using a precision for the measurement of $B = 100,000$, and verify that this distance is small, e.g., on the order of less than 10^{-3}. In short, this additional step tells us whether we have really found a value of θ that is very close to $\tilde{\theta}$, and yields a mean for the OLS estimator close to $\hat{\theta}$. Should we have been worried about some of the required conditions for equation (1) to have a solution and for our SA algorithm to work reliably in finding it, this step verifies that we have found a solution.

While the structure of our algorithm has solid theoretical foundations, our specific configuration ($\alpha^{(j)}$, B, number of burn-in/actual iterations) is admittedly arbitrary. We chose it based on our own experience with the algorithm. The specifics of the problem likely would allow us to reduce the computational cost further by choosing the configuration in some optimal way. We leave this for future research. With our configuration, the computational costs are very manageable.

C   VAR Monte Carlo study

To assess the performance of our bias correction method, we present the results of a simulation study, which considers a bivariate VAR model. To create a setting that is comparable to the reality faced by researchers analyzing interest rates, we first estimate such a model on actual interest rates. We use the same monthly data set as in Section 4. We extract the first two principal components from the cross section of yields and estimate the VAR. Then we take the OLS estimates, rounded to two decimals, as our DGP parameters:

$$X_{t+1} = \begin{pmatrix} .98 & .01 \\ 0 & .97 \end{pmatrix} X_t + \varepsilon_{t+1}, \quad \varepsilon_t \overset{iid}{\sim} N(0, I_2).$$

We generate $M = 2000$ samples and calculate for each replication six alternative estimates: OLS, analytical bias correction as in Pope (1990), bootstrap and indirect inference bias correction. For bootstrap bias correction, we use 1,000 replications. For calculating the indirect inference estimator, we use 1500 iterations in our algorithm, discarding the first 500, with 5 bootstrap samples in each iteration. Estimates that imply explosive VAR dynamics are stationarity-adjusted in the same way as in the DTSM Monte Carlo study in Section 4.4.

Table C.1: VAR Monte Carlo simulation results

	DGP	OLS	analytical	bootstrap	ind. inf.
Φ_{11}	0.98	-0.0281	-0.0087	-0.0068	-0.0049
Φ_{12}	0.01	0.0015	-0.0007	-0.0004	-0.0007
Φ_{21}	0.00	-0.0028	-0.0010	-0.0011	-0.0014
Φ_{22}	0.97	-0.0305	-0.0102	-0.0086	-0.0056
RMSB		0.0208	0.0067	0.0055	0.0038
TAB		0.0629	0.0206	0.0168	0.0126
max. eig.	0.98	-0.0154	0.0058	0.0081	0.0090
half-life	34	-14.1930	13.7945	14.3122	23.0127
IRF at $h = 60$	0.2976	-0.1726	0.0978	0.1433	0.1644
freq. expl.		0.25%	23.90%	35.00%	30.80%

Notes: True values and bias for parameters and persistence measures, and summary statistics capturing total bias for the VAR Monte Carlo study. For details refer to the text.

Table C.1 shows the results of our simulations. The first four rows show for each parameter the true value, the mean bias of OLS and the three bias-correcting estimators. The fifth and sixth row show the "root-mean-squared bias" (RMSB), which is the square root of the mean-squared bias across the four parameters, and the "total absolute bias" (TAB), which is the sum of the absolute values of the bias in each parameter. The results show that bias correction significantly reduces the small-sample parameter bias. This holds for all three bias correction methods. The total absolute bias is reduced by about 75% using the bootstrap. The indirect inference estimator is able to further reduce bias. It estimates most elements of Φ with less bias, and reduces the measures of total bias.

The seventh to ninth rows show true values and the bias for three measures of persistence—the largest eigenvalue of Φ as well as the half-life and value of the IRF at a horizon of 60 periods for the response of the first variable to own shocks. To calculate the half-life we use the same approach as in Kilian and Zha (2002), with a cutoff of 500 periods—the mean of the half-life is calculated across those replications for which it is available, which consequently excludes values for which the half-life would be larger than 500.

The downward bias in estimated persistence for the OLS estimator is sizable. The half-life and long-horizon IRF are on average half as large as for the true DGP. Bias correction increases the estimated persistence significantly, and the dramatic downward bias in the largest eigenvalue disappears. Notably, in this setting the bias corrected estimates tend to be even more persistent than the DGP. Bootstrap and indirect inference bias correction display similar performance in terms of the implied persistence of the estimates.

The last row shows the frequency with which explosive eigenvalues occur. Evidently, although the DGP is stationary and the estimated model is correctly specified, there is a sizable probability that the realized value of the OLS estimator is such that bias correction leads to explosive estimates. Consequently, in practice one will often have to perform some type of stationarity adjustment to ensure that estimated dynamics are not explosive.

We draw two conclusions from this simulation study: First, both the bootstrap and in-

Table D.1: Maximally-flexible DTSM – summary statistics

	OLS	analytical BC	bootstrap BC	ind. inf. BC
$max(eig(\Phi))$	0.9678	0.9961	0.9999	0.9991
half-life	24	147	n.a.	265
IRF at 5y	0.16	0.75	0.95	0.93
$\sigma(f_t^{47,48})$	1.392	1.392	1.392	1.392
$\sigma(\tilde{f}_t^{47,48})$	0.388	1.206	1.431	1.635
$\sigma(ftp_t^{47,48})$	1.301	1.216	1.322	1.656

Notes: Summary statistics for OLS and bias-corrected estimates of the DTSM in Joslin et al. (2011). First row: maximum eigenvalue of the estimated Φ. Second and third row: half-life and value of the impulse response function at the five-year horizon for the response of the level factor to a level shock. Rows six to eight show sample standard deviations of the fitted 47-to-48-month forward rates and of the corresponding risk-neutral forward rates and forward term premia.

direct inference are useful and reliable methods to reduce the bias in OLS estimates of VAR parameters. Second, indirect inference is a superior bias correction method compared to analytical or bootstrap correction in a setting like ours, because it reduces higher-order bias and hence further improves the accuracy of the parameter estimates.

D   Maximally-flexible DTSM: alternative bias correction methods

In the main text we present bias-corrected DTSM estimates that are obtained by applying indirect inference bias correction. Here we compare the results for the maximally-flexible model of Section 4 to those obtained using alternative bias correction methods.

Table D.1 shows the summary statistics for six alternative sets of estimates: OLS, analytical bias correction, bootstrap bias correction, and indirect inference bias correction. Correcting for bias using the bootstrap leads to explosive VAR dynamics, so we apply Kilian's stationarity adjustment in this case—for the resulting Φ matrix the IRF does not fall below 0.5 within 40 years, our cutoff for half-life calculation as in Kilian and Zha (2002).

There are some differences in results across bias correction methods. Analytical bias correction leads to a less persistent VAR system than indirect inference bias correction. But the key result is robust to using these alternative approaches: The persistence is substantially increased by bias correction, so that short rate forecasts and risk-neutral rates are much more volatile than for OLS. In practice, a researcher will likely use that method for bias correction that (s)he is most comfortable with on practical or theoretical grounds. For us, this is the indirect inference method.

Table E.1: DTSM Monte Carlo study – parameter bias

	DGP			Bias OLS			Bias BC		
1200μ	-0.328	-0.232	0.074	-0.392	0.143	-0.019	-0.207	0.069	-0.016
Φ	0.999	0.000	0.472	-0.024	0.018	-0.021	-0.013	0.008	0.000
	-0.002	0.993	0.347	0.006	-0.018	0.005	0.003	-0.009	0.001
	0.000	0.002	0.881	-0.003	-0.003	-0.026	-0.002	-0.001	-0.009
$1200 r_\infty^{\mathbb{Q}}$	8.657			-1.126			-1.092		
$\lambda^{\mathbb{Q}}$	0.998	0.952	0.929	-0.001	-0.001	-0.002	-0.001	-0.001	-0.002
1200Σ	0.644	0	0	-0.018	0	0	-0.009	0	0
	-0.147	0.211	0	0.003	0.014	0	0.003	0.017	0
	0.063	-0.011	0.087	-0.001	0.005	0.027	-0.001	0.005	0.028

Notes: DGP parameter values and bias of OLS and BC estimates. For details refer to text.

E   Parameter bias in DTSM Monte Carlo study

In Table E.1 we show the DGP parameters that are used to simulate interest rate data from the DTSM, as well as parameter bias of the OLS and BC estimates. The bias in the estimates of VAR parameters μ and Φ is sizable for the OLS estimates. As expected, the bias-corrected estimates generally display reduced bias. With regard to the $\mathbb{Q}$-measure parameters, the values of $\lambda^{\mathbb{Q}}$ are estimated very accurately for all three estimators. This confirms the intuition that in DTSM estimation, cross-sectional information helps to pin down the parameters determining cross-sectional loadings very precisely. The risk-neutral long-run mean $r_\infty^{\mathbb{Q}}$ is estimated with a slight downward bias by all estimators because the largest root under $\mathbb{Q}$ is very close to one, which naturally makes inference about the long-run mean under $\mathbb{Q}$ difficult. This has motivated some researchers to restrict the largest root under $\mathbb{Q}$ to one and $r_\infty^{\mathbb{Q}}$ to zero, in which case the long end of the yield curve is determined by a level factor (Bauer, 2011; Christensen et al., 2011).

In sum, bias-corrected estimation of the maximally-flexible DTSM that we consider here reduces bias in estimates of the VAR parameters in comparison to MLE, whereas the remaining parameters are estimated with similar accuracy.

References

Bauer, Michael D., "Nominal Rates and the News," Working Paper 2011-20, Federal Reserve Bank of San Francisco August 2011.

Christensen, Jens H.E., Francis X. Diebold, and Glenn D. Rudebusch, "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models," *Journal of Econometrics*, September 2011, *164* (1), 4–20.

Efron, Bradley and Robert J. Tibshirani, *An introduction to the bootstrap*, Chapman & Hall/CRC, 1993.

Gourieroux, Christian, Eric Renault, and Nizar Touzi, "Calibration by simulation for small sample bias correction," in Robert Mariano, Til Shcuermann, and Melvyn J. Weeks, eds., *Simulation-based Inference in Econometrics: Methods and Applications*, Cambridge University Press, 2000, chapter 13, pp. 328–358.

Hall, Peter, *The bootstrap and Edgeworth expansion*, Springer Verlag, 1992.

Horowitz, Joel L., "The Bootstrap," in J.J. Heckman and E.E. Leamer, eds., *Handbook of Econometrics*, Vol. 5, Elsevier, 2001, chapter 52, pp. 3159–3228.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu, "A New Perspective on Gaussian Dynamic Term Structure Models," *Review of Financial Studies*, 2011, *24* (3), 926–970.

Kilian, Lutz and Tao Zha, "Quantifying the Uncertainty About the Half-life of Deviations from PPP," *Journal of Applied Econometrics*, 2002, *17* (2), 107–125.

MacKinnon, James G. and Anthony A. Jr. Smith, "Approximate bias correction in econometrics," *Journal of Econometrics*, 1998, *85* (2), 205–230.

Polyak, B.T. and A.B. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM Journal on Control and Optimization*, 1992, *30*, 838.

Pope, Alun L., "Biases of Estimators in Multivariate Non-Gaussian Autoregressions," *Journal of Time Series Analysis*, 1990, *11* (3), 249–258.

Robbins, H. and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, 1951, *22* (3), 400–407.