No Arbitrage Priors, Drifting Volatilities, and the Term Structure of Interest Rates[1]

Andrea Carriero[2] *Todd E. Clark*[3] *Massimiliano Marcellino*[4]

June 2014

[1] The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

[2] Queen Mary, University of London
[3] Federal Reserve Bank of Cleveland
[4] Bocconi University and CEPR

Introduction

- Producing accurate forecasts of the term structure of interest rates is crucial for bond portfolio management, derivatives pricing, and risk management.
- Most contributions have focused on point forecasts of the yield curve. However, assessing the whole predictive distribution of the yield curve is more important for the success of portfolio and risk management strategies.
- Which ingredients for a good density forecast?
  - The time series of interest rates typically feature comovement and heteroskedasticity. Having a joint dynamic model featuring time variation in volatility is key
  - A good model for point forecasts

Introduction

- Gaussian Affine Term Structure Models (GATSM): widely used and successful for in-sample analysis
  - Vasiceck (1977), Duffie and Kan (1996), Dai and Singleton (2000), Duffee (2002), Ang and Piazzesi (2003), Christensen, Diebold and Rudebush (2011).
- How do they forecast yields out of sample? Duffee (2002) and Ang and Piazzesi (2003): beating a random walk with a traditional no arbitrage GATSM is difficult.
- The assumption of absence of arbitrage - which is per se reasonable in well developed markets - needs to be translated into a set of restrictions on a particular model. These specification assumptions are not necessarily holding in the data.
- Using a GATSM as a prior rather than as a set of sharp restrictions allows to deal with the potential misspecification (Del Negro and Schorfheide 2004).

Summary of the paper

- In this paper we propose a model that:
  - Shrinks the point (and density) forecasts towards a no arbitrage model.
  - Allows for time variation in the volatilities.
- We impose the GATSM as a prior rather than as sharp restrictions to account for its possible misspecification.
- As the volatilities of a panel of yields move closely together, we impose on them a factor structure.
- We derive the conditional posterior kernels of this model and use a MCMC sampler for posterior simulation.
- Such modelling choices result in a clear improvement in point and density forecasting performance.

Relation to (some) literature

- The method can be applied for a wide range of alternative models, and can be considered as an extension of the method of Del Negro and Schorfheide (2004) to VARs featuring drifting volatilities.
  - Can be applied to DSGE models
- The model generalizes the approach of Giannone, Lenza and Primiceri (2012)
  - *Both* the prior variance *and* the prior mean of the VAR coefficients are specified hierarchically, and errors are heteroskedastic.
- With respect to Carriero (2011) we introduce drifting volatilities and consider the new canonical form of Joslin, Singleton, and Zhu (2011).
  - This representation presents very important advantages in the computation of the likelihood
- Density forecasts (Shin and Zhong 2013, Chib and Kang 2013)

Joslin, Singleton and Zhu (2011) model

- The yields of maturity $\tau = 1, ..., N$ are collected in the $N \times 1$ vector y_t, and are a funcion of a set of n factors P_t :

$$\Delta P_t = K^P_{0P} + K^P_{1P} P_{t-1} + \Sigma_P \varepsilon^P_t \tag{1}$$

$$y_t = A_p + B_p P_t + \Sigma_y \varepsilon^y_t \tag{2}$$

  where P_t are factors, and Σ_P is the Cholesky factor of their conditional variance.

- No arbitrage implies that the coefficients appearing in A_p, B_p are a function of some deep coefficients. We collect these deep coefficients and the variance matrices in the vector θ:

$$\theta = (\lambda^Q, k^Q_\infty, \Sigma_P, \Sigma_y). \tag{3}$$

  (K^P_{0P} and K^P_{1P} can be concentrated out of the likelihood).

- For a given choice of θ it is possible to compute the moments $E[y_t y_{t-h}]$ under the state space model in (1)-(2):
- These moments will be used to form a prior for a VAR and are ultimately a function of the deep coefficients θ

VAR with common stochastic volatility (CSV)

- We use the specification proposed by Carriero, Clark and Marcellino (2012):

$$
\begin{aligned}
y_t &= \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t. &(4)\\
u_t &= \lambda_t^{0.5} \epsilon_t, \ \epsilon_t \sim N(0, V), &(5)\\
\log(\lambda_t) &= \phi_0 + \phi_1 \log(\lambda_{t-1}) + \nu_t, \ \nu_t \sim \text{iid } N(0, \phi_2). &(6)
\end{aligned}
$$

- The assumption of common stochastic volatility is predicated on the fact that the volatilities of yields feature a strong factor structure
  - The first principal component explains most of the variation in the panel (e.g. 89% in our data set)
- Modelling volatility as common produces a likelihood featuring a variance matrix with Kronecker structure, which allows to use a conjugate N-IW prior (cond. on volatilities)
- The priors on Φ and V are set up based on the moments of the GATSM for a given hyperparameter θ, with tightness γ.

Algorithm

- Let Λ denote the history of volatility and Y collect all the data. Draws from the joint posterior:

$$p(\Phi, V, \theta, \gamma, \Lambda, \phi | Y) \tag{7}$$

  are obtained by drawing in turn from:

$$p(\Phi, V, \theta, \gamma | Y, \Lambda, \phi) \tag{8}$$

  and:

$$p(\Lambda, \phi | Y, \Phi, V, \theta, \gamma) \tag{9}$$

- Draws from (8) are obtained as in Del Negro and Schorfheide (2004), plus an additional step to draw the tightness hyperparameter γ.
- Draws from (9) are obtained as in Carriero, Clark, Marcellino (2012) using a modification of Cogley and Sargent (2005) algorithm (to account for common volatility)

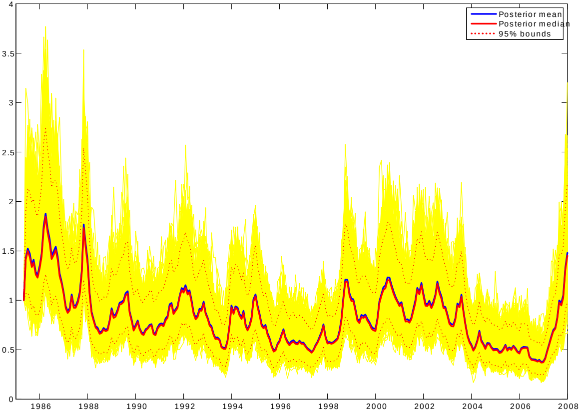
Priors

- The priors on the VAR coefficients V and Φ are set up hierarchically, using equations (19) and (20). We do not need a prior on the first observation of the volatility process.
- Therefore, we only need to specify priors for γ, θ, ϕ. For these parameters, we use weakly informative priors (and also check with diffuse priors for robustness)
  - For θ the prior reflects the belief that the first factor in the GATSM is (close to) a random walk, the second is stationary but very persistent, and the third is moderately persistent.
  - The prior mean for γ is centered in 1, which corresponds to giving a-priori the same weight to the GATSM and the unrestricted VAR. We implement the restriction $\gamma > (k + N)/T$, necessary for the priors on V and Φ to be proper, by truncating the posterior draws.
  - The prior for ϕ reflects the belief that volatility is persistent. $\phi_0 \sim N(0, 0.025)$, $\phi_1 \sim N(0.96, 0.025)$, and $\phi_2 \sim IG(3 \cdot 0.05, 3)$.

Data

Zero-coupon Fama-Bliss yields, at monthly frequency, and ranging from January 1985 to December 2007.

In sample results: CSV factor



Forecasting exercise

- Out-of-sample forecasting exercise.
- We estimate the model recursively and produce forecasts up to 12 steps ahead.
  - Initial estimation window : January 1985 to December 1994
  - Final estimation window : January 1985 to November 2007
  - Forecast evaluation period: January 1995 to December 2007.
- We obtain forecast distributions by sampling as appropriate from the posterior distribution of the considered models.

Forecast evaluation

- Point forecasts are evaluated using RMSE:

$$RMSFE_{i,h}^{M} = \sqrt{\frac{1}{P}\sum\left(\hat{y}_{t+h}^{(i)}(M) - y_{t+h}^{(i)}\right)^2}, \tag{10}$$

where $\hat{y}_{t+h}^{(i)}(M)$ denote the forecast of the i-th yield $y_{t+h}^{(i)}$ made by model M.

  - Significance via Diebold and Mariano (1995) t-statistic, rough guide (conservative in small sample - Clark and McCracken 2011a,b)

- Density forecasts evaluated with log predictive density scores:

$$SCORE_{i,h}^{M} = \frac{1}{P}\sum \log p(y_{t+h}|y^{(t)}, M), \tag{11}$$

where the predictive density $p(\cdot)$ is obtained by univariate kernel estimation based on the MCMC output.

  - Significance via Amisano and Giacomini (2007) t-statistic, rough guide (asymptotic validity would require rolling windows)

Table 3. Evaluation of Point Forecasts. Sample 1995:2007

Maturity→ / step-ahead ↓	0.25-yrs	1-yrs	2-yrs	3-yrs	5-yrs	7-yrs	10-yrs
RW point forecasting performance							
1	20.80	23.34	27.30	29.02	28.75	27.44	26.32
2	33.71	36.73	42.08	43.66	42.61	40.49	38.31
3	45.22	48.67	53.26	53.52	51.17	48.04	44.67
6	77.98	79.20	80.21	77.21	72.45	66.48	60.29
12	135.78	132.78	122.24	111.25	97.55	86.92	77.52
JSZ-VAR vs Random Walk							
1	0.86 ***	0.97	1.03	1.01	1.02	1.03	**1.00**
2	**0.80** **	0.95	1.03	1.03	1.03	1.03	**0.99**
3	**0.78** **	0.92	**1.01**	**1.01**	**1.02**	**1.02**	**0.98**
6	**0.78** *	**0.90**	**0.96**	**0.96**	**0.96**	**0.96**	**0.93**
12	**0.80**	**0.85**	**0.86**	**0.86**	**0.86** *	**0.88**	**0.86** *
JSZ-VAR-CSV vs Random Walk							
1	**0.85** ***	**0.95**	1.02	**1.00**	**1.01**	1.02	**1.00**
2	**0.79** **	**0.93**	1.03	**1.02**	**1.02**	1.03	1.00
3	**0.77** **	**0.91**	**1.01**	**1.01**	**1.02**	**1.02**	1.00
6	**0.78** *	0.91	0.98	0.99	1.00	1.01	0.99
12	0.81 *	0.88	0.91	0.91	0.92	0.94	0.95
BVAR-CSV vs Random Walk							
1	0.93 ***	0.98	**1.00**	1.01	1.01	**1.01**	1.01
2	0.92 ***	0.99	**1.02**	1.02	1.02	**1.02**	1.02
3	0.92 ***	0.99	1.03	1.03	1.03	**1.02**	1.02
6	0.95 *	1.02	1.06	1.06	1.06	1.04	1.03
12	0.95	1.01	1.06	1.08	1.09	1.08	1.05

Table 4. Evaluation of Density Forecasts. Sample 1995:2007

Maturity→ / step-ahead ↓	0.25-yrs	1-yrs	2-yrs	3-yrs	5-yrs	7-yrs	10-yrs
RW density forecasting performance							
1	-4.54	-4.66	-4.78	-4.82	-4.81	-4.77	-4.72
2	-4.96	-5.06	-5.17	-5.21	-5.19	-5.15	-5.10
3	-5.26	-5.34	-5.41	-5.41	-5.38	-5.33	-5.26
6	-5.86	-5.81	-5.81	-5.77	-5.73	-5.66	-5.58
12	-7.34	-6.42	-6.24	-6.14	-6.05	-5.97	-5.89
JSZ-VAR vs Random Walk							
1	0.13 ***	0.05 **	0.00	0.00	0.00	-0.02	-0.02
2	0.16 **	0.03	-0.04	-0.04	-0.03	-0.03	**0.00**
3	0.20 *	0.06	-0.02	-0.03	-0.01	-0.02	**0.01**
6	0.28	0.08	0.03	0.02	0.04	0.02	0.03
12	1.19	0.23	0.10	0.09	**0.10** *	**0.08** *	**0.07** *
JSZ-VAR-CSV vs Random Walk							
1	**0.30** ***	**0.16** ***	**0.04**	**0.03**	**0.01**	**0.01**	**0.01**
2	**0.29** ***	**0.11** **	**0.00**	**-0.01**	**-0.02**	-0.03	-0.01
3	**0.31** ***	**0.13** **	**0.01**	**0.01**	**0.01**	**-0.01**	0.00
6	**0.37** *	**0.15** *	**0.07**	**0.05**	**0.06**	**0.04**	**0.04**
12	**1.26**	**0.29**	**0.12** *	**0.09**	0.09 *	**0.08** *	**0.07** *
BVAR-CSV vs Random Walk							
1	0.19 ***	0.12 ***	**0.04**	0.02	0.00	**0.01**	-0.01
2	0.16 ***	0.07 **	**0.00**	-0.02	-0.02	**-0.01**	-0.01
3	0.14 ***	0.07 *	-0.02	-0.03	-0.01	-0.01	-0.01
6	0.12	0.04	0.00	-0.02	0.01	0.01	0.01
12	1.02	0.14	0.00	-0.03	-0.02	0.00	0.02

Table 5: JSZ-VAR vs GATSM

Maturity→ step-ahead ↓	0.25-yrs	1-yrs	2-yrs	3-yrs	5-yrs	7-yrs	10-yrs
Relative RMSFE (point forecasting performance)							
1	0.95	0.87**	0.97	1.01	0.98	0.97	0.93**
2	0.92	0.89	0.98	1.01	0.98	0.98	0.95
3	0.88	0.86*	0.94	0.97	0.94	0.93	0.90*
6	0.85	0.84*	0.88	0.90	0.87	0.86	0.81**
12	0.83	0.81*	0.81*	0.80*	0.77*	0.77*	0.72**
Average Difference in SCORE (density forecasting performance) *							
1	0.640	0.581	0.500	0.493	0.528	0.549	0.602
2	0.588	0.499	0.387	0.385	0.440	0.485	0.553
3	0.536	0.459	0.388	0.404	0.471	0.515	0.596
6	0.408	0.395	0.385	0.422	0.485	0.530	0.603
12	0.245	0.285	0.338	0.405	0.494	0.536	0.592

***** All differences in density forecasts are significant at the 1% level**

Table 6: JSZ-VAR-CSV vs VAR-CSV with factor structure only

Maturity→ step-ahead ↓	0.25-yrs	1-yrs	2-yrs	3-yrs	5-yrs	7-yrs	10-yrs
Relative RMSFE (point forecasting performance)							
1	1.00	0.99	1.00	1.00	0.99	1.00	1.00
2	1.00	1.00	0.99	0.99	0.99	1.00	1.00
3	0.99	0.99	0.99	0.99	0.99	1.00	1.00
6	0.99	0.99	0.99	0.99	0.99	1.00	1.00
12	0.99	1.00	0.99	0.99	0.99	1.00	1.00
Average Difference in SCORE (density forecasting performance)							
1	0.0664*	0.0166**	0.007	0.005	0.005	0.003	0.009
2	0.0233*	0.007	0.011	0.012	0.005	0.002	0.001
3	0.017	0.003	0.007	0.012	0.013	0.008	0.004
6	0.024	0.007	0.004	0.003	0.001	0.000	0.004
12	0.031	0.0247*	0.022	0.013	0.012	0.009	0.010

Summary of Results

- Both the $JSZ - VAR - CSV$ and the $JSZ - VAR$ models produce competitive point and density forecasts, systematically outperforming the RW benchmark.
- The gains against the random walk increase with the forecast horizon.
- The $JSZ - VAR - CSV$ specification produced the best density forecasts throughout the sample.
- The gains in using a specification with time varying volatility tend to die out as the forecast horizon increases.
- Although the differences are statistically insignificant, long-term point forecasts of the $JSZ - VAR$ model are slightly superior to those of the $JSZ - VAR - CSV$.
- We suggest to use the heteroschedastic version of our model for short term forecasting, while at long horizons it is probably safer to complement it with the homoschedastic version of the model.

Conclusions

- We propose a way to impose a no arbitrage affine term structure model as a prior on a VAR, while allowing also for time variation in the error volatilities.
- We provide the conditional posterior distribution kernels of the model and we propose a MCMC algorithm to perform estimation.
- The method can be applied to several models, including DSGE (work in progress), and is an extension of the method of Del Negro and Schorfheide (2004) to VARs featuring drifting volatilities.
- Our model generalizes the one of Giannone, Lenza and Primiceri (2012), by introducing heteroskedastic errors and specifying hierarchically *both* the prior variances *and* the prior means of the VAR coefficients.
- Based on U.S. data we provide evidence that both the shrinkage towards the term structure model and the time variation in volatilities can produce substantial gains in forecasting the yield curve.

VAR with common stochastic volatility (CSV)

- A theoretical sample of dimension T^* obeying the GATSM has an approximate VAR representation:

$$
\begin{aligned}
y_t^* &= \Phi_0 + \Phi_1 y_{t-1}^* + ... + \Phi_p y_{t-p}^* + u_t^*. && (12) \\
u_t^* &= \lambda_t^{*0.5} \epsilon_t, \ \epsilon_t \sim N(0, V), && (13) \\
\lambda_t^* &= 1 \text{ for all } t. && (14)
\end{aligned}
$$

- Under the GATSM, the common stochastic volatility factor stays constant at its initial value of 1
- The yields y_t^* have the moment matrices implied by the GATSM, for any given choice of the hyperparameters θ.
- Conditional on λ_t these moment matrices can be used as a prior for the VAR coefficients Φ and V, as in Del Negro and Schorfheide (2004).
- The prior is conditional on θ and has tightness $\gamma = T^*/T$.

VAR with CSV: JSZ prior

- We derive the prior as Del Negro and Schorfheide (2004), i.e. by using the moments of the underlying state space system:

$$\Phi|V,\theta,\gamma,\Lambda \sim N(\hat{\Phi}^*(\theta), V \otimes (\gamma T \Gamma_{\tilde{X}^{*\prime}\tilde{X}^*}(\theta))^{-1}), \quad (19)$$

$$V|\theta,\gamma,\Lambda \sim IW(\hat{S}^*(\theta), \gamma T - k), \quad (20)$$

where:

$$\begin{aligned}
\hat{\Phi}^*(\theta) &= \Gamma^{-1}_{\tilde{X}^{*\prime}\tilde{X}^*}(\theta)\Gamma_{\tilde{X}^{*\prime}\tilde{Y}^*}(\theta), \\
\hat{S}^*(\theta) &= \gamma T(\Gamma_{\tilde{Y}^{*\prime}\tilde{Y}^*}(\theta) - \Gamma_{\tilde{Y}^{*\prime}\tilde{X}^*}(\theta)\Gamma^{-1}_{\tilde{X}^{*\prime}\tilde{X}^*}(\theta)\Gamma_{\tilde{X}^{*\prime}\tilde{Y}^*}(\theta)),
\end{aligned}$$

- Here $\Gamma_{\tilde{Y}^{*\prime}\tilde{Y}^*}(\theta), \Gamma_{\tilde{Y}^{*\prime}\tilde{X}^*}(\theta), \Gamma_{\tilde{X}^{*\prime}\tilde{X}^*}(\theta)$ are the moments of the (rescaled) yields under the GATSM model.
- Therefore the prior is conditional on θ.
- It is also conditional on the γ (the overall tightness of the prior). We will draw γ as Giannone et al. (2012).

VAR with CSV: Conditional posteriors

- The posterior distributions of Φ and V will be proportional to the likelihood times the prior, and are conjugate (Zellner, 1973):

$$\Phi | Y, V, \theta, \gamma, \Lambda \sim N(\tilde{\Phi}(\theta), V \otimes (\gamma T \Gamma_{\tilde{X}^{*\prime} \tilde{X}^{*}}(\theta) + \tilde{X}' \tilde{X})^{-1}), \quad (21)$$

$$V | Y, \theta, \gamma, \Lambda \sim IW(\tilde{S}(\theta), (\gamma + 1) T - k), \quad (22)$$

where:

$$\tilde{\Phi}(\theta) = (\gamma T \Gamma_{\tilde{X}^{*\prime} \tilde{X}^{*}}(\theta) + \tilde{X}' \tilde{X})^{-1} (\gamma T \Gamma_{\tilde{X}^{*\prime} \tilde{Y}^{*}}(\theta) + \tilde{X}' \tilde{Y}),$$

$$\tilde{S}(\theta) = [(\gamma T \Gamma_{\tilde{Y}^{*\prime} \tilde{Y}^{*}}(\theta) + \tilde{Y}' \tilde{Y}) - (\gamma T \Gamma_{\tilde{Y}^{*\prime} \tilde{X}^{*}}(\theta) + \tilde{Y}' \tilde{X})(\gamma T \Gamma_{\tilde{X}^{*\prime} \tilde{X}^{*}}(\theta) + \tilde{X}' \tilde{X})^{-1} (\gamma T \Gamma_{\tilde{X}^{*\prime} \tilde{Y}^{*}}(\theta) + \tilde{X}' \tilde{Y})].$$

Volatility

- We use a modification of Cogley and Sargent (2005) algorithm.
- Defining the orthogonalized residuals $w_t = (w_{1t}, ..., w_{nt}) = V^{-1/2} u_t$ the kernel of $p(\lambda_t | Y, \Phi, V, \theta, \gamma, \phi)$ is given by:

$$\lambda_t^{-n \times 1.5} \prod_{i=1}^{n} \exp(-0.5 w_{it}^2 / \lambda_t) \cdot \exp(-0.5 (\ln \lambda_t - \mu_t)^2 / \sigma_c^2), \tag{23}$$

  where μ_t and σ_c^2 are the conditional mean and variance of $\ln \lambda_t$.
- By choosing an appropriate proposal density, this kernel can be used as a basis for a Metropolis step with acceptance probability:

$$a = \min \left(\frac{\lambda_t^{* - n \times 0.5} \prod_{i=1}^{n} \exp(-0.5 w_{it}^2 / \lambda_t^*)}{\lambda_t^{-n \times 0.5} \prod_{i=1}^{n} \exp(-0.5 w_{it}^2 / \lambda_t)}, 1 \right). \tag{24}$$

- Note (23) and (24) differ from Cogley and Sargent (2005), as in their case each volatility process λ_{it} is drawn separately conditional on the remaining $n - 1$
- Draws from $p(\phi | Y, \Lambda)$ using standard results for univariate linear regressions

MCMC sampler

To sum up, the algorithm will draw in turn from the distributions, as follows:

1. Draw from the conditional posterior distribution of θ, $p(\theta|Y, \gamma, \Lambda)$;
2. Draw from the conditional posterior distribution of γ, $p(\gamma|Y, \theta, \Lambda)$;
3. Draw from the conditional posterior distribution of V, $p(V|Y, \theta, \gamma, \Lambda)$;
4. Draw from the conditional posterior distribution of Φ, $p(\Phi|Y, V, \theta, \gamma, \Lambda)$;
5. Draw from the conditional posterior distribution of Λ, $p(\Lambda|Y, \Phi, V, \phi)$; and
6. Draw from the conditional posterior distribution of ϕ, $p(\phi|Y, \Lambda)$.

Algorithm

- The joint p.d.f. (8) can be factorized as follows:

$$p(\Phi, V, \theta, \gamma | Y, \Lambda) \propto p(\Phi | Y, \Lambda, V, \theta, \gamma) p(V | Y, \Lambda, \theta, \gamma) p(\theta, \gamma | Y, \Lambda, \gamma). \quad (25)$$

  where we have omitted conditioning on ϕ because they are redundant under knowledge of Λ.
- Draws from $\theta, \gamma | Y, \Lambda$ can be obtained using Metropolis steps using the kernel of the p.d.f. of this distribution, which is available
- Draws from $V | Y, \Lambda, \theta, \gamma$, and $\Phi | Y, \Lambda, V, \theta, \gamma$ can be obtained via MC steps.
- This step is the same as DS (2004), except we also have an additional block to draw the tightness hyperparameter γ.

Algorithm

- The second step involves simulating from the joint posterior of the volatility process λ_t and its law of motion parameters ϕ, conditional on the VAR coefficients:

$$p(\Lambda, \phi | Y, \Phi, V) \tag{26}$$

where we have omitted conditioning on θ and γ because they are redundant under knowledge of Φ and V.

- Draws are obtained by drawing in turn from $\phi | Y, \Lambda$ and $\Lambda | Y, \Phi, V, \phi$.
- Following Cogley and Sargent (2005) we specify conjugate priors on the parameters in ϕ, so that $\phi | Y, \Lambda$ can be obtained via a MC step.
- To draw from $\Lambda, \phi | Y, \Phi, V$ we use the method proposed in Carriero, Clark, and Marcellino (2012). Such method is a modification of Cogley and Sargent (2005) to allow for a single stochastic volatility factor.