Federal Reserve Bank of New York
Staff Reports

Time-Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum

Marco Del Negro
Giorgio Primiceri

Staff Report No. 619
May 2013
Revised October 2014

This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

Time-Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum
Marco Del Negro and Giorgio Primiceri
Federal Reserve Bank of New York Staff Reports, no. 619
May 2013; revised October 2014
JEL classification: C11, C15

Abstract

This note corrects a mistake in the estimation algorithm of the time-varying structural vector autoregression model of Primiceri (2005) and shows how to correctly apply the procedure of Kim, Shephard, and Chib (1998) to the estimation of VAR, DSGE, factor, and unobserved components models with stochastic volatility. Relative to Primiceri (2005), the main difference in the new algorithm is the ordering of the various Markov Chain Monte Carlo steps, with each individual step remaining the same.

Key words: Bayesian methods, time-varying volatility

Del Negro: Federal Reserve Bank of New York (e-mail: marco.delnegro@ny.frb.org). Primiceri: Northwestern University (e-mail: g-primiceri@northwestern.edu). The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

1 The Model in Short

This note is a corrigendum of Primiceri (2005), but its lesson applies more broadly to several empirical macro models with stochastic volatility that are estimated using the approach of Kim, Shephard, and Chib (1998, KSC hereafter). Consider the time-varying VAR model of Primiceri (2005)

$$y_t = c_t + B_{1,t}y_{t-1} + ... + B_{k,t}y_{t-k} + A_t^{-1}\Sigma_t\varepsilon_t, \tag{1}$$

where y_t is an $n \times 1$ vector of observed endogenous variables; c_t is a vector of time-varying intercepts; $B_{i,t}$, $i = 1, ..., k$, are matrices of time-varying coefficients; A_t is a lower triangular matrix with ones on the main diagonal and time-varying coefficients below it; Σ_t is a diagonal matrix of time-varying standard deviations; ε_t is an $n \times 1$ vector of unobservable shocks with variance equal to the identity matrix. All the time-varying coefficients evolve as random walks, except for the diagonal elements of Σ_t, which behave as geometric random walks. All the innovations in the model (shocks to coefficients, log-volatilities and ε_t) are jointly normally distributed, with mean equal to zero and covariance matrix equal to V. The matrix V is block diagonal, with blocks corresponding to the time-varying elements of the B's, A, Σ and ε. The block structure of the matrix V is described in detail in Primiceri (2005).

2 The Original Algorithm of Primiceri (2005) and Why It is Wrong

The unknown objects of the model are the history of the volatilities (Σ^T), the history of the coefficients (B^T and A^T), and the covariance matrix of the innovations (V). To simplify the notation, define $\theta \equiv \left[B^T, A^T, V\right]$. Primiceri (2005) proposed to simulate the posterior distribution of the model coefficients by Gibbs sampling, drawing the history of volatilities with the multi-move algorithm of KSC.

The difficulty with drawing Σ^T is that they enter the model multiplicatively. For given θ, however, simple algebraic manipulations of (1) yield a linear system in the log volatilities. A consequence of applying these transformations is that they also convert ε_t into $\log \varepsilon_t^2$, which is a vector of $\log \chi^2(1)$ random variables. The method of KSC relies on approximating each element of $\log \varepsilon_t^2$ with a mixture of normals. Conditioning on the mixture indicators makes it possible to use standard Gaussian state-space methods to conduct inference on the volatilities. As a consequence, the Gibbs sampler is augmented to include the mixture indicators $s^T \equiv \{s_t\}_{t=1}^T$ that select the component of the mixture for each variable at each date.

Primiceri (2005) adopts the following algorithm to obtain posterior draws for Σ^T, s^T and θ:

Algorithm 1 (wrong algorithm)

1. Draw Σ^T from $\tilde{p}\left(\Sigma^T | y^T, \theta, s^T\right)$
2. Draw s^T from $\tilde{p}\left(s^T | y^T, \Sigma^T, \theta\right)$
3. Draw θ from $p\left(\theta | y^T, \Sigma^T\right)$,

where the "~" in step 1 and 2 indicates that the conditional posteriors of Σ^T and s^T correspond to the product of their conditional priors by $\tilde{p}(y^T | \Sigma^T, \theta, s^T)$, i.e. the likelihood of the data conditional on the components of the mixture-of-normals approximation of the $\log \chi^2(1)$ distribution for each date and variable. The conditional posterior of θ in step 3 is instead obtained using the true likelihood implied by model (1), i.e. $p\left(y^T | \Sigma^T, \theta\right)$.

There are two reasons why this algorithm does not yield draws from the correct posterior distribution of the model parameters. First of all, the algorithm alternates between the use of two different likelihood functions: steps 1 and 2 of the sampler make use of the mixture-of-normals approximation, to facilitate the draw of Σ^T; step 3, instead, uses the correct likelihood.

More important, the second problem with Algorithm 1 is related to the fact that it was conceived as a Gibbs sampler with "blocks" Σ^T, θ , and s^T. In a Gibbs sampler, one has to draw from each block conditional on all the others. However, the draw of θ in step 3 is not conditional on s^T. Primiceri (2005) erroneously assumed that conditioning on s^T in step 3 does not make a difference, but instead it does: the knowledge of which components of the mixture have been selected for each date and variable changes the likelihood of the data, thus affecting the conditional posterior of θ. This simple observation invalidates Algorithm 1, even abstracting from the approximation error. In other words, Algorithm 1 would not yield draws from the correct posterior even if we used an arbitrarily large number of mixture components to make the approximation arbitrarily accurate.

3 A Gibbs Sampler with Different Blocking

Fixing this problem of Algorithm 1 by simply replacing step 3 with "Draw from $\tilde{p}\left(\theta|y^T, \Sigma^T, s^T\right)$" is not a viable option because $\varepsilon_t|s_t$ in not Gaussian, which precludes the possibility of drawing easily from $\tilde{p}\left(\theta|y^T, \Sigma^T, s^T\right)$. An alternative strategy is to use a Gibbs sampler with different blocking. Instead of using three blocks, Σ^T, θ , and s^T, one can use two blocks, i.e. Σ^T and (θ, s^T). The first step of the new sampler is to draw Σ^T conditional on (θ, s^T) and the data y^T. The second step is to draw from the joint distribution of (θ, s^T) conditional on Σ^T and the data. Of course, drawing from the joint of $\left(\theta, s^T\right)$ can be accomplished by drawing first from the marginal of θ and then from the conditional of s^T given θ. This yields the following algorithm (of which the online appendix presents a more formal treatment):

Algorithm 2 (correct algorithm under no approximation error)

1. Draw Σ^T from $\tilde{p}\left(\Sigma^T|y^T, \theta, s^T\right)$
2. Draw (θ, s^T) from $\tilde{p}\left(\theta, s^T|y^T, \Sigma^T\right)$, which is accomplished by
   (a) Drawing θ from $p\left(\theta|y^T, \Sigma^T\right)$
   (b) Drawing s^T from $\tilde{p}\left(s^T|y^T, \Sigma^T, \theta\right)$,

where the "$\tilde{}$" notation in steps 1 and 2b continues to indicate the use of the auxiliary approximating model—as opposed to the true likelihood—to facilitate the draw of the history of volatilities.

Like Algorithm 1, also Algorithm 2 alternates between the use of the correct and the approximate likelihood. However, unlike Algorithm 1, Algorithm 2 has the property that it would yield draws from the correct posterior in the hypothetical case in which the mixture of normals represented a perfect approximation for the $\log \chi^2(1)$ distribution, as we formally show in the online appendix. As we stress in the next section, in practice, the mixture of normals is of course only an approximation of the $\log \chi^2(1)$ distribution. We therefore think of Algorithm 2 as a sampler from an approximate posterior.

Finally, notice that the individual steps in Algorithms 1 and 2 are the same, but the order is different: in Algorithm 2 the indicators s^T are sampled *after* θ and *before* Σ^T. Since the individual steps remain the same, they can all be implemented as in Primiceri (2005).[1] Algorithm 2 is therefore equivalent to switching steps (d) and (e) in the algorithm summarized in Appendix A.5 of Primiceri (2005).[2] This order is key to derive Algorithm 2 as a Gibbs sampler based on the two blocks Σ^T and (θ, s^T), and thus to justify the draw of θ from a posterior that does not conditions on s^T.

4 Addressing the Approximation Problem

In this section we explicitly deal with the issue of the approximation error, recognizing the fact that the finite mixture of normals is only used as an approximation of the $\log \chi^2(1)$ distribution. Stroud et al. (2003) show how to address this problem by turning step 1 of Algorithm 2 into a Metropolis-Hastings step, where the distribution $\tilde{p}\left(\Sigma^T|y^T, \theta, s^T\right)$ is used as a proposal density. Specifically, we set up another algorithm, which we denote by *Algorithm 3 (correct algorithm)*. Steps 2a and 2b of Algorithm 3 are the same as in Algorithm 2. Step 1 is instead replaced with a candidate draw from the proposal density $\tilde{p}\left(\Sigma^T|y^T, \theta, s^T\right)$. This draw is then accepted with probability proportional to the ratio between the conditional density of the new and previous draw, re-weighted by the ratio between

[1] In particular, step (2) can be implemented by drawing from $p\left(B^T|y^T, A^T, V, \Sigma^T\right)$, $p\left(A^T|y^T, B^T, V, \Sigma^T\right)$ and $p\left(V|y^T, A^T, B^T, \Sigma^T\right)$.

[2] Section A.5 in Primiceri (2005) actually contains a typo: step (d) of the algorithm should be $p(s^T|y^T, \mathbf{B^T}, A^T, \Sigma^T, V)$ as opposed to $p(s^T|y^T, A^T, \Sigma^T, V)$. Unlike the conceptual mistake outlined in the previous section, this typo was inconsequential given that it is mechanically not possible to draw s^T without conditioning on B^T.

the proposal density of the previous and the new draw, as standard in each Metropolis-Hastings algorithm. If the candidate draw of Σ^T is not accepted, the draw of Σ^T is set equal to the previous draw. The functional form of the acceptance probability is shown in equation (11) of Stroud et al. (2003), and re-derived in our online appendix for the specific case of our model.

A formal illustration of Algorithm 3 requires some investment in notation and is therefore relegated to the online appendix. We stress that this sampler is correct (i.e. eventually yields the right posterior density of Σ^T and θ) regardless of the quality of the approximation, which matters only for its efficiency. We also emphasize that a key step in Algorithm 3, as in Algorithm 2, consists in integrating out the mixture components when drawing θ, which implies inverting the order of the draws of Σ^T and s^T relative to the original Gibbs sampler. This is the main difference relative to Primiceri (2005). Our correction implies that researchers using the KSC approach to estimate VARs, DSGEs, or factor models with time-varying volatility need to make sure they sample the indicators s^T right before the history of volatilities. Examples of such papers are numerous in the past decade, e.g. Justiniano and Primiceri (2008).[3] This lesson also applies to unobserved components models with stochastic volatility (e.g., Stock and Watson, 2007).

5 Consequences for the Results

In the online appendix, we have applied Geweke's (2004) "Joint Distribution Tests of Posterior Simulators" to further confirm that Algorithm 1 is incorrect, Algorithm 2 is approximately correct, and Algorithm 3 is fully correct. In addition, we have re-estimated the model of Primiceri (2005) using Algorithm 2 and 3, and compared the results to the original ones obtained with Algorithm 1.

Algorithm 2 generates results that are indistinguishable from those obtained with Algorithm 3, suggesting that the mixture-of-normals approximation error involved in the pro-

[3]The estimation algorithm of the DSGE model with stochastic volatility of Justiniano and Primiceri (2008) is correct, although their appendix describes an algorithm with the wrong order.

cedure of KSC is negligible in our application (as it was in theirs). The results based on Algorithm 2 and 3 are instead not the same as those obtained with Algorithm 1, albeit qualitatively similar. The main difference is that some estimates of the time-varying objects are now smoother. The full set of new results can be found in the online appendix.

REFERENCES

GEWEKE, J. (2004), "Getting It Right: Joint Distribution Tests of Posterior Simulators," *Journal of the American Statistical Association*, 99, 799–804.

JUSTINIANO, A. and PRIMICERI, G. E. (2008), "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98, 604–641.

KIM, S., SHEPHARD, N. and CHIB, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361–393.

PRIMICERI, G. E. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852.

STOCK, J.H. and WATSON, M.W. (2007), "Why Has U.S. Inflation Become Harder to Forecast?" Journal of Money, Credit, and Banking, 39(1), 3–33.

STROUD, J.R., MÜLLER, P. and POLSON, N.G. (2003), "Nonlinear State-Space Models With State-Dependent Variances," *Journal of the American Statistical Association*, 98, 377–386.

In this appendix, we (i) re-estimate the model of Primiceri (2005) using Algorithm 2 (the sampler from the approximate posterior) and Algorithm 3 (the sampler from the true posterior), and compare these results with those obtained with Algorithm 1 (the original, incorrect algorithm of Primiceri, 2005); (ii) present a more formal treatment of Algorithm 2 and Algorithm 3; (iii) formally explain why Algorithm 1 is incorrect; and (iv) apply Geweke's (2004) "Joint Distribution Tests of Posterior Simulators" to Algorithm 1, 2 and 3, and present the results of these tests.

A New Results Based on Algorithm 2 and 3

In this section, we reproduce the figures of Primiceri (2005) using Algorithm 2 (the sampler from the approximate posterior) and Algorithm 3 (the sampler from the true posterior), and compare these results with those obtained with Algorithm 1 (the original, incorrect algorithm of Primiceri, 2005). The new results are based on 70,000 draws of the Gibbs sampler, discarding the first 20,000 to allow for convergence to the ergodic distribution.

The first thing to notice is that the results based on Algorithm 3 are qualitatively similar to the original ones obtained with Algorithm 1, but they are not the same (figures 1-8). The main difference is that some estimates of the time-varying objects are now smoother. For example, the standard deviation of monetary policy shocks (figure 1c) exhibits substantial time variation, but not as much as in the original results. A similar comment applies to the interest rate response to a permanent increase in inflation and unemployment (figures 5 and 7).

The results obtained using Algorithm 2 and 3 are instead indistinguishable from each other (figures 9-16), suggesting that the mixture-of-normals approximation error involved in the procedure of Kim, Shephard and Chib (1998, KSC hereafter) is negligible in our application (as it was in theirs).

B A Formal Treatment of Algorithm 2 and 3

In this section, we present a formal derivation of Algorithm 2 and 3.

B.1 Algorithm 2

The joint posterior distribution of Σ^T and θ is given by

$$p(\Sigma^T, \theta | y^T) \propto p(y^T | \Sigma^T, \theta) \cdot p(\Sigma^T, \theta) \tag{2}$$

where $p(y^T|\Sigma^T, \theta)$ is the likelihood function implied by equation (1.1) of the corrigendum, and $p(\Sigma^T, \theta)$ is the prior density of Σ^T and θ. In principle, one could use a two-block Gibbs sampler in Σ^T and θ with steps: i) draw Σ^T from $p(\Sigma^T|y^T, \theta) \propto p(y^T|\Sigma^T, \theta) \cdot p(\Sigma^T|\theta)$, and ii) draw θ from $p(\theta|y^T, \Sigma^T) \propto p(y^T|\Sigma^T, \theta) \cdot p(\theta|\Sigma^T)$. While step (ii) is straightforward, step (i) is not: the time-varying volatilities Σ^T enter the model multiplicatively, making it impossible to use linear and Gaussian state-space methods. KSC's idea consists of approximating the likelihood $p(y^T|\theta, \Sigma^T)$ using the mixture-of-normals $\int \tilde{p}(y^T|\Sigma^T, \theta, s^T)\pi(s^T)ds^T$, where s^T represents the components of the mixture for each date and variable, $\pi(s^T)$ are the corresponding mixture weights, and $\tilde{p}(y^T|\Sigma^T, \theta, s^T)$ is the likelihood of the data conditional on the mixture components s^T.[4]

Let $\tilde{p}(\Sigma^T, \theta, s^T|y^T)$ denote the product of $\tilde{p}(y^T|\Sigma^T, \theta, s^T)$ and the prior of θ, Σ^T and s^T, that is

$$\tilde{p}(\Sigma^T, \theta, s^T|y^T) = \tilde{p}(y^T|\Sigma^T, \theta, s^T) \cdot p(\Sigma^T, \theta) \cdot \pi(s^T). \tag{3}$$

In addition, for the sake of argument, suppose that the mixture-of-normals provides a perfect approximation of the likelihood, i.e. that $p(y^T|\theta, \Sigma^T) = \int \tilde{p}(y^T|\Sigma^T, \theta, s^T)\pi(s^T)ds^T$. Integrating out the mixture components s^T from $\tilde{p}(\Sigma^T, \theta, s^T|y^T)$ in (3), we obtain $p(y^T|\theta, \Sigma^T) \cdot p(\Sigma^T, \theta)$, which is proportional to the the posterior of interest $p(\theta, \Sigma^T|y^T)$. This implies that, if we device an algorithm for drawing from $\tilde{p}(\Sigma^T, \theta, s^T|y^T)$, after discarding the draws of s^T,

[4] Note that there is a clear abuse of notation in writing $\int \tilde{p}(y^T|\theta, \Sigma^T, s^T)\pi(s^T)ds^T$ given that the s^T are discrete indicators, but this shortcut simplifies the notation considerably.

we are left with draws of θ and Σ^T from the desired distribution. *Algorithm 2*, which we rewrite below, represents such an algorithm:

1. Draw Σ^T from $\tilde{p}\left(\Sigma^T|y^T,\theta,s^T\right) \propto \tilde{p}(y^T|\Sigma^T,\theta,s^T)\cdot p(\Sigma^T|\theta)$
2. Draw (θ, s^T) from $\tilde{p}\left(\theta,s^T|y^T,\Sigma^T\right)$, which is accomplished by
   (a) Drawing θ from $p\left(\theta|y^T,\Sigma^T\right) \propto p(y^T|\Sigma^T,\theta)\cdot p(\theta|\Sigma^T)$.
   (b) Drawing s^T from $\tilde{p}\left(s^T|y^T,\Sigma^T,\theta\right) \propto \tilde{p}(y^T|\Sigma^T,\theta,s^T)\cdot\pi(s^T)$.

As emphasized in the note, Algorithm 2 is conceived as a two-blocks sampler, with blocks Σ^T and (θ, s^T). We draw from the joint of (θ, s^T) given Σ^T and y^T by first drawing from the marginal $p\left(\theta|y^T,\Sigma^T\right)$ and then from the conditional $\tilde{p}\left(s^T|y^T,\Sigma^T,\theta\right)$. It is precisely the fact that we draw from the marginal of θ that allows us to use the original likelihood $p(y^T|\Sigma^T,\theta)$ in step 2a: under the assumption that there is no approximation error, integrating out the s^T from the joint distribution (3) yields

$$p(\Sigma^T,\theta)\cdot\int\tilde{p}(y^T|\Sigma^T,\theta,s^T)\pi(s^T)ds^T = p(y^T|\Sigma^T,\theta)\cdot p(\Sigma^T,\theta) \propto p(y^T|\Sigma^T,\theta)\cdot p(\theta|\Sigma^T).$$

Furthermore, step 1 is also simple: as discussed in the paper, conditional on s^T, the model is linear and Gaussian in the log-volatilities, making the distribution $\tilde{p}(y^T|\Sigma^T,\theta,s^T)$ amenable to the use of linear and Gaussian state-space methods.

B.2 Algorithm 3

In the previous section we have provided a justification for Algorithm 2 under the assumption that $p(y^T|\theta,\Sigma^T) = \int\tilde{p}(y^T|\Sigma^T,\theta,s^T)\pi(s^T)ds^T$. Of course, in practice, this is not correct: the mixture of normals is only an approximation of the true likelihood. In this subsection we present a formal treatment of Algorithm 3, which addresses this issue.

Construct a joint posterior of Σ^T, θ and s^T as follows:

$$\begin{aligned} p(\theta,\Sigma^T,s^T|y^T) &= p(\theta,\Sigma^T|y^T)\cdot\tilde{p}(s^T|\Sigma^T,\theta,y^T) \\ &\propto p(y^T|\theta,\Sigma^T)\cdot p(\Sigma^T,\theta)\cdot\tilde{p}(s^T|\Sigma^T,\theta,y^T), \end{aligned} \tag{4}$$

with

$$\tilde{p}(s^T|\Sigma^T,\theta,y^T) = \frac{\tilde{p}(y^T|\Sigma^T,\theta,s^T)\cdot\pi(s^T)}{c(\Sigma^T,\theta,y^T)}, \tag{5}$$

where $c(\Sigma^T,\theta,y^T) \equiv \int \tilde{p}(y^T|\Sigma^T,\theta,s^T)\pi(s^T)ds^T$ guarantees that the density in (5) integrates to one.

As discussed above, a perfectly fine approach for obtaining draws from the posterior of interest, $p(\theta,\Sigma^T|y^T)$, is to sample from $p(\theta,\Sigma^T,s^T|y^T)$, and then discard the draws of s^T. This is precisely what Algorithm 3 does. Like Algorithm 2, Algorithm 3 has the structure of a two-block sampler, with blocks Σ^T and $\left(\theta,s^T\right)$. However, Algorithm 3 follows Stroud et al. (2003) in using a Metropolis-Hastings step for drawing Σ^T conditional on (θ,s^T), where the proposal pdf is given by

$$\tilde{p}\left(\Sigma^T|y^T,\theta,s^T\right) \propto \tilde{p}(y^T|\Sigma^T,\theta,s^T)\cdot p(\Sigma^T|\theta), \tag{6}$$

which is the density used in step 1 of Algorithm 2.[5]

Specifically, *Algorithm 3* consists of the following steps:

1. Draw Σ^T from $p\left(\Sigma^T|y^T,\theta,s^T\right)$ as follows: Draw a candidate $\tilde{\Sigma}^T$ from the proposal density $\tilde{p}\left(\Sigma^T|y^T,\theta,s^T\right)$ of Algorithm 2, and set

$$\Sigma^{(j)\ T} = \begin{cases} \tilde{\Sigma}^T & \text{with probability } \alpha \\ \Sigma^{(j-1)\ T} & \text{with probability } 1-\alpha \end{cases},$$

   where the superscript (j) denotes the iteration of the sampler, and where

$$\alpha = \frac{p\left(\tilde{\Sigma}^T|y^T,\theta,s^T\right)}{p\left(\Sigma^{(j-1)\ T}|y^T,\theta,s^T\right)}\frac{\tilde{p}\left(\Sigma^{(j-1)\ T}|y^T,\theta,s^T\right)}{\tilde{p}\left(\tilde{\Sigma}^T|y^T,\theta,s^T\right)}.$$

2. Draw (θ,s^T) from $p\left(\theta,s^T|y^T,\Sigma^T\right)$, which is accomplished by

   (a) Drawing θ from

$$\begin{aligned} p\left(\theta|y^T,\Sigma^T\right) &= \int p(\theta,s^T|y^T,\Sigma^T)ds^T \\ &\propto p(y^T|\theta,\Sigma^T)\cdot p(\theta|\Sigma^T)\cdot\int \tilde{p}(s^T|\Sigma^T,\theta,y^T)ds^T = p(y^T|\Sigma^T,\theta)\cdot p(\theta|\Sigma^T). \end{aligned}$$

[5] Stroud et al. (2003) study the use of mixture approximations in Gibbs samplers, and thus generalize the results of KSC.

(b) Drawing s^T from $\tilde{p}\left(s^T|y^T,\Sigma^T,\theta\right) \propto \tilde{p}(y^T|\Sigma^T,\theta,s^T)\cdot\pi(s^T)$.

Observe that, since step 1 takes θ and s^T as given, the acceptance probability can be rewritten as

$$\alpha = \frac{p\left(\tilde{\Sigma}^T,\theta,s^T|y^T\right)}{p\left(\Sigma^{(j-1)\ T},\theta,s^T|y^T\right)} \frac{\tilde{p}\left(\Sigma^{(j-1)\ T}|y^T,\theta,s^T\right)}{\tilde{p}\left(\tilde{\Sigma}^T|y^T,\theta,s^T\right)}.$$

Using (4), (5) and (6), we then obtain

$$\alpha = \frac{p(y^T|\theta,\tilde{\Sigma}^T)}{p(y^T|\theta,\Sigma^{(j-1)\ T})} \frac{c(\Sigma^{(j-1)\ T},\theta,y^T)}{c(\tilde{\Sigma}^T,\theta,y^T)}.$$

Finally, notice that $c(\Sigma^T,\theta,y^T)$ coincides with the mixture-of-normals approximation of the original likelihood $p(y^T|\Sigma^T,\theta)$, hence

$$\alpha = \frac{\left(\prod_t \phi\left(y_t^*|0_{n\times1},\tilde{\Sigma}_t\tilde{\Sigma}_t'\right)\right)\left(\prod_t\prod_i mn_{KSC}\left(y_{i,t}^{**}-2\log\sigma_{i,t}^{(j-1)}\right)\right)}{\left(\prod_t \phi\left(y_t^*|0_{n\times1},\Sigma_t^{(j-1)}\Sigma_t^{(j-1)\ \prime}\right)\right)\left(\prod_t\prod_i mn_{KSC}\left(y_{i,t}^{**}-2\log\tilde{\sigma}_{i,t}\right)\right)}, \tag{7}$$

where $y_t^* = A_t\left(y_t - c_t - B_{1,t}y_{t-1} - ... - B_{k,t}y_{t-k}\right)$, $y_{i,t}^{**} = \log\left(y_{i,t}^{*2}+0.001\right)$, $\sigma_{i,t}$ is the i-th element of the diagonal of Σ_t, $\phi\left(\cdot|0_{n\times1},\tilde{\Sigma}_t\tilde{\Sigma}_t'\right)$ is the pdf of an $n-$variate Gaussian distribution with mean zero and variance $\tilde{\Sigma}_t\tilde{\Sigma}_t'$, and $mn_{KSC}(\cdot)$ denotes the pdf of the mixture-of-normals distribution with means, variances and mixing proportions specified in KSC.

C The Fixed-Point Integral Equation

In this section, we formally explain why the original algorithm of Primiceri (2005) is incorrect. The reason why Algorithm 1 is not a proper Gibbs sampler can be understood from inspecting the key equation showing why the Markov chain converges (we omit the conditioning on y^T to simplify notation):

$$p(\theta,\Sigma^T) = \int h\left(\theta,\Sigma^T|\theta',\Sigma^{T\prime}\right) p\left(\theta',\Sigma^{T\prime}\right) \mathrm{d}\left(\theta',\Sigma^{T\prime}\right) \tag{8}$$

where $h\left(\theta,\Sigma^T|\theta',\Sigma^{T\prime}\right)$ is a Markov transition kernel defined by

$$h\left(\theta,\Sigma^T|\theta',\Sigma^{T\prime}\right) = \int p(\theta,\Sigma^T|s^T)p(s^T|\theta',\Sigma^{T\prime})\mathrm{d}s^T. \tag{9}$$

Equation (8) defines a fixed point integral equation for which the true marginal $p(\theta, \Sigma^T)$ is a solution, which is readily seen by plugging (9) into (8) and changing the order of integration, as shown below (Chib and Greenberg, 1996 and references therein discuss why i) it is the unique solution, and ii) there is convergence from any initial $p\left(\theta', \Sigma^{T\prime}\right)$ under general conditions). Omitting to condition on s^T when drawing from $p(\theta, \Sigma^T|s^T)$ (as done in step 3) implies using the wrong kernel, hence the fixed point argument breaks down: even if one were to draw $(\theta', \Sigma^{T\prime})$ from the correct joint distribution, the resulting (θ, Σ^T) in the next iteration would not be from $p(\theta, \Sigma^T)$.

In the three-block Gibbs sampler, equation (3.1)—the fixed point integral equation—becomes

$$p(\theta, \Sigma^T, s^T) = \int .. \int h\left(\theta, \Sigma^T, s^T|\theta', \Sigma^{T'}, s^{T'}\right) p\left(\theta', \Sigma^{T'}, s^{T'}\right) \mathrm{d}\theta' \mathrm{d}\Sigma^{T'} \mathrm{d}s^{T'} \tag{10}$$

where $h\left(\theta, \Sigma^T, s^T|\theta', \Sigma^{T'}, s^{T'}\right)$ is a Markov transition kernel defined by

$$h\left(\theta, \Sigma^T, s^T|\theta', \Sigma^{T'}, s^{T'}\right) = p(\theta|\Sigma^T, s^T)p(\Sigma^T|\theta', s^T)p(s^T|\theta', \Sigma^{T'}). \tag{11}$$

Here we follow Chib and Greenberg (1996) and show that $p(\theta, \Sigma^T, s^T)$ is indeed the solution to (10). In fact, one can write the right hand side of expression (10), after substituting in the definition of the transition kernel (11), as:

$$\begin{aligned}
&\int .. \int p(\theta|\Sigma^T, s^T)p(\Sigma^T|\theta', s^T)p(s^T|\theta', \Sigma^{T'})p\left(\theta', \Sigma^{T'}, s^{T'}\right) \mathrm{d}\theta' \mathrm{d}\Sigma^{T'} \mathrm{d}s^{T'} = \\
&\quad \int .. \int p(\theta|\Sigma^T, s^T)\frac{p(\Sigma^T|s^T)p(\theta'|\Sigma^T, s^T)}{p(\theta'|s^T)} \frac{p(s^T)p(\theta', \Sigma^{T'}|s^T)}{p(\theta', \Sigma^{T'})} p\left(\theta', \Sigma^{T'}, s^{T'}\right) \mathrm{d}\theta' \mathrm{d}\Sigma^{T'} \mathrm{d}s^{T'}
\end{aligned}$$

where we used Bayes law to express $p(\Sigma^T|\theta', s^T)$ and $p(s^T|\theta', \Sigma^{T'})$. Note that the terms

$$p(\theta|\Sigma^T, s^T)p(\Sigma^T|s^T)p(s^T) = p(\theta, \Sigma^T, s^T)$$

can be taken out of the integral as they do not depend on the $'$ variables, and their product is precisely $p(\theta, \Sigma^T, s^T)$. Therefore we just have to show that

$$\int .. \int \frac{p(\theta'|\Sigma^T, s^T)}{p(\theta'|s^T)} \frac{p(\theta', \Sigma^{T'}|s^T)}{p(\theta', \Sigma^{T'})} p\left(\theta', \Sigma^{T'}, s^{T'}\right) \mathrm{d}\theta' \mathrm{d}\Sigma^{T'} \mathrm{d}s^{T'} = 1.$$

This is the case because

$$\int .. \int \frac{p(\theta'|\Sigma^T, s^T)}{p(\theta'|s^T)} \frac{p(\theta', \Sigma^{T'}|s^T)}{p(\theta', \Sigma^{T'})} p(\theta', \Sigma^{T'}, s^{T'}) \mathrm{d}\theta' \mathrm{d}\Sigma^{T'} \mathrm{d}s^{T'} =$$

$$\int .. \int \frac{p(\theta'|\Sigma^T, s^T)}{p(\theta'|s^T)} \frac{p(\theta'|s^T) p(\Sigma^{T'}|\theta', s^T)}{p(\theta', \Sigma^{T'})} p(\theta', \Sigma^{T'}) p(s^{T'}|\theta', \Sigma^{T'}) \mathrm{d}\theta' \mathrm{d}\Sigma^{T'} \mathrm{d}s^{T'} =$$

$$\int p(\theta'|\Sigma^T, s^T) \left(\int p(\Sigma^{T'}|\theta', s^T) \left(\int p(s^{T'}|\Sigma^{T'}, s^{T'}) \mathrm{d}s^{T'} \right) \mathrm{d}\Sigma^{T'} \right) \mathrm{d}\theta' = 1,$$

where in the second line we again used Bayes law and in the fourth line we realized that we are left with three conditional distributions, all integrating to one. Clearly, omitting to condition on s^T when drawing from $p\left(\theta|\Sigma^T, s^T\right)$ implies using the wrong kernel, and the fixed-point arguments breaks down.

D Geweke's (2004) "Getting It Right"

In this section, we apply Geweke's (2004) "Joint Distribution Tests of Posterior Simulators" to the three algorithms discussed in the note, and present further evidence that Algorithm 1 is wrong, Algorithm 2 is approximately correct, and Algorithm 3 is correct. Geweke's idea is to compare two ways of obtaining draws from the joint distribution of the data and the model parameters, $p\left(y^T, \theta, \Sigma^T, s^T\right)$:

a. Draw the parameters from the prior, and then the data from the data-generating process (that is, draw sequentially from $p\left(\theta, \Sigma^T, s^T\right)$ and $p\left(y^T|\theta, \Sigma^T, s^T\right)$).

b. Draw from the posterior using the MCMC algorithm given a draw of the data, and then use this draw to generate another draw from data, and so on (that is, draw sequentially from $p\left(\theta, \Sigma^T, s^T|y^T\right)$ and then from $p\left(y^T|\theta, \Sigma^T, s^T\right)$).

If the MCMC algorithm is correct, (a) and (b) should yield the same distribution, and in particular the same marginal for the model parameters (which, in the case of (a), is of course the prior). Therefore, if the MCMC algorithm is correct, P-P plots constructed using the draws from (a) and (b) should lie on the 45-degree line.

We now present the results obtained by applying this procedure to the various algorithms that we have discussed so far. Note that, for computational reasons, we use $T = 10$ in running these tests, which is smaller than the actual sample size. For a T as large as that in the sample, it simply takes so many draws for (b) to converge (even if the MCMC algorithm is right) that the test is computationally not feasible. Since Geweke's approach applies to any T, we are justified in using a smaller T that makes the comparison feasible.

We concentrate on the P-P plots for the distribution of the log-volatilities at a particular point in time ($t = 7$), because the differences are smaller for the other coefficients. Figure 17 shows the results related to the original algorithm (Algorithm 1). It is evident that the P-P plots are very far from the 45-degree line, indicating that the draws generated with (a) and (b) belong to different distributions. This suggests the presence of a mistake in Algorithm 1, as we have argued above.

Figure 18 plots the results obtained using Algorithm 2. The fact that the P-P plots in figure 18 are now much closer to the 45-degree line is a sign of dramatic improvement in the accuracy of the algorithm. The natural question is of course why these P-P plots do not lie exactly on top of the 45-degree line, but just close to it. This is due to the minor error involved in the mixture-of-normals approximation proposed by KSC. A property of the Geweke (2004) approach is that it amplifies subtle discrepancies in the sampler, such as these small approximation errors. Figure 19 confirms this conjecture by presenting the P-P plots obtained by running the Geweke procedure using Algorithm 3. In this case, the P-P plots essentially coincide with the 45-degree lines, which verifies that there is no problem with Algorithm 2, other than the fact that it uses the mixture approximation to increase efficiency and speed of convergence. Recall from section 1 that this approximation is absolutely inconsequential for the estimation results, i.e. for the construction the posterior distribution given the observed data. Conversely, applying the same correction for the mixture-of-normals approximation error in step 1 of the original algorithm does not improve the P-P plots at all, as shown in figure 20.

Figure 1: Posterior mean, 16th and 84th percentiles of the standard deviation of (a) the residuals of the inflation equation, (b) the residuals of the unemployment equation and (c) the residuals of the interest rate equation or monetary policy shocks.

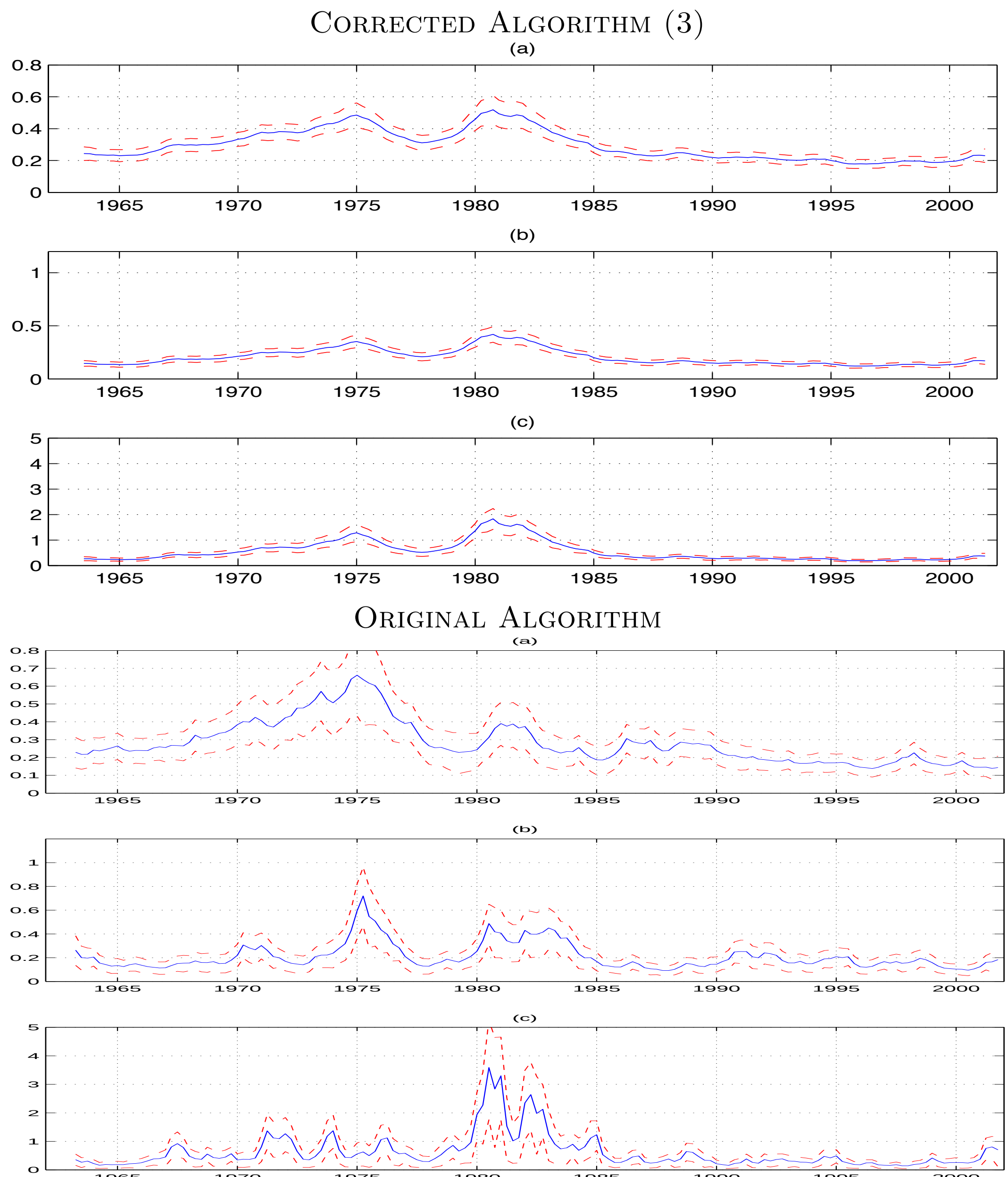
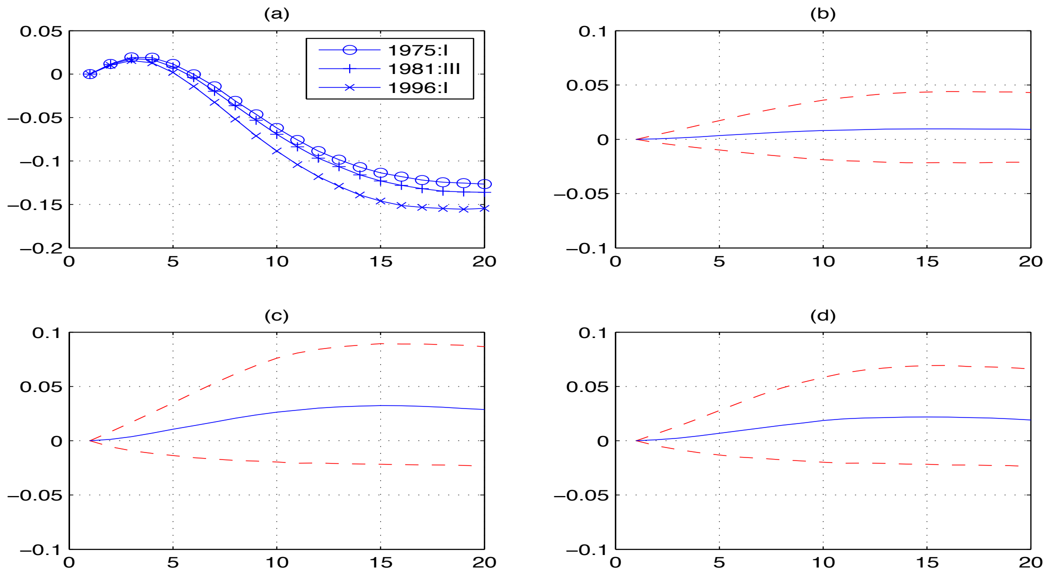


Figure 2: (a) impulse responses of inflation to monetary policy shocks in 1975:I, 1981:III and 1996:I, (b) difference between the responses in 1975:I and 1981:III with 16th and 84th percentiles, (c) difference between the responses in 1975:I and 1996:I with 16th and 84th percentiles, (d) difference between the responses in 1981:III and 1996:I with 16th and 84th percentiles.

Corrected Algorithm (3)



Original Algorithm

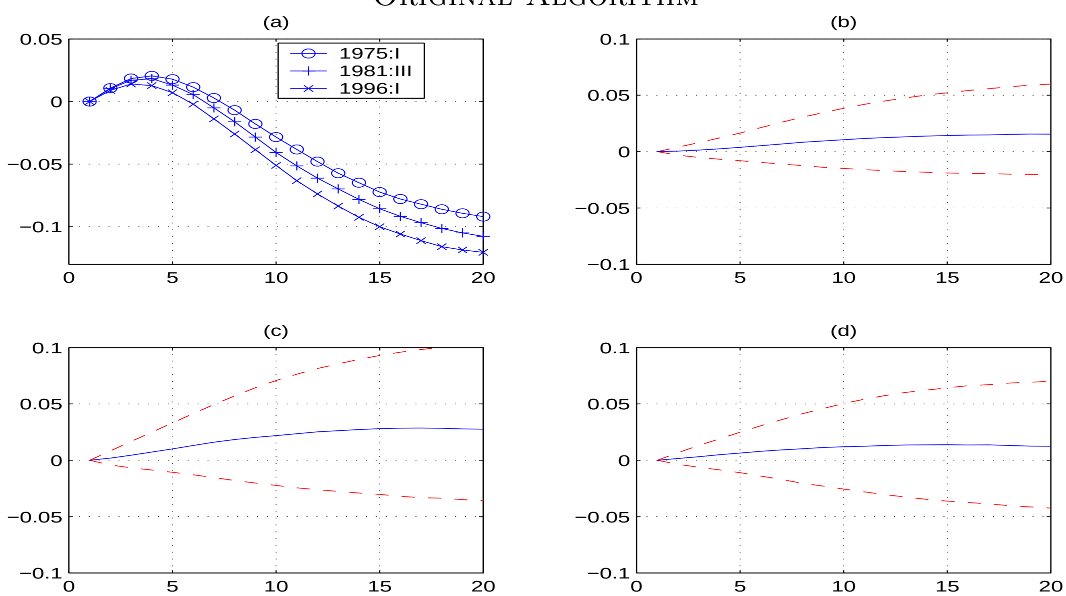
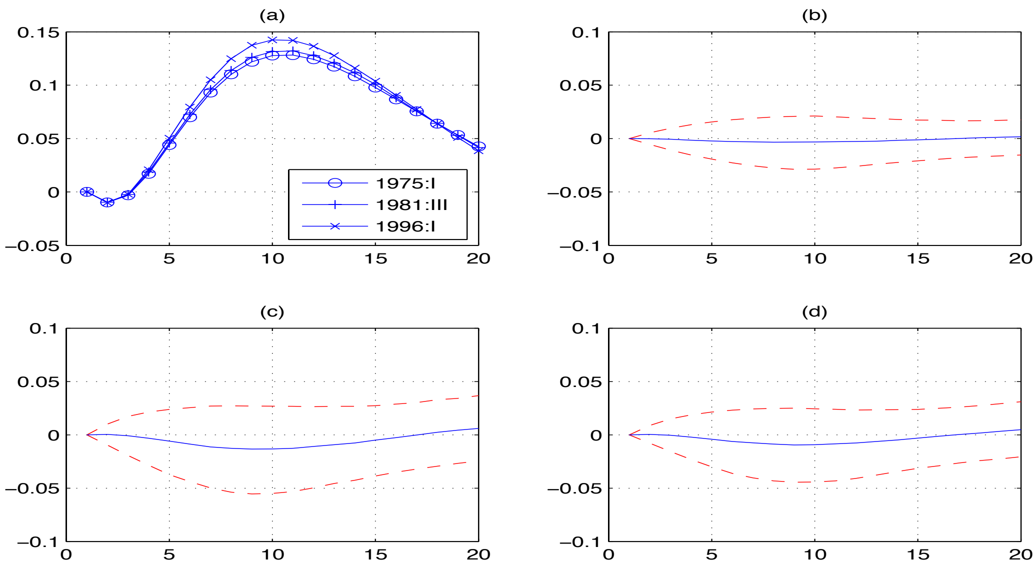


Figure 3: (a) impulse responses of unemployment to monetary policy shocks in 1975:I, 1981:III and 1996:I, (b) difference between the responses in 1975:I and 1981:III with 16th and 84th percentiles, (c) difference between the responses in 1975:I and 1996:I with 16th and 84th percentiles, (d) difference between the responses in 1981:III and 1996:I with 16th and 84th percentiles.

Corrected Algorithm (3)



Original Algorithm

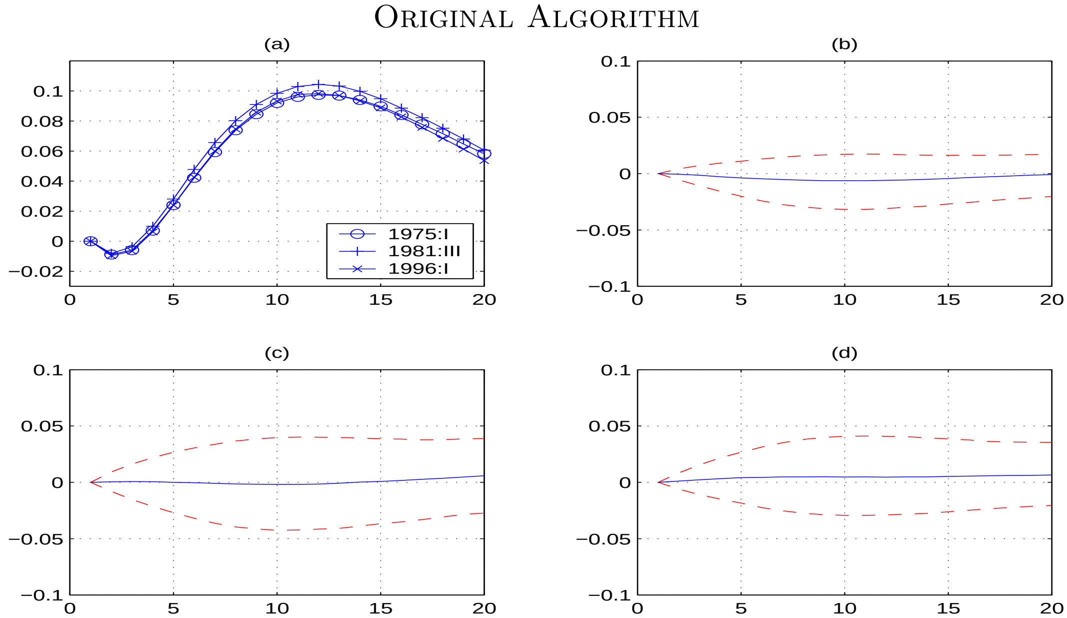
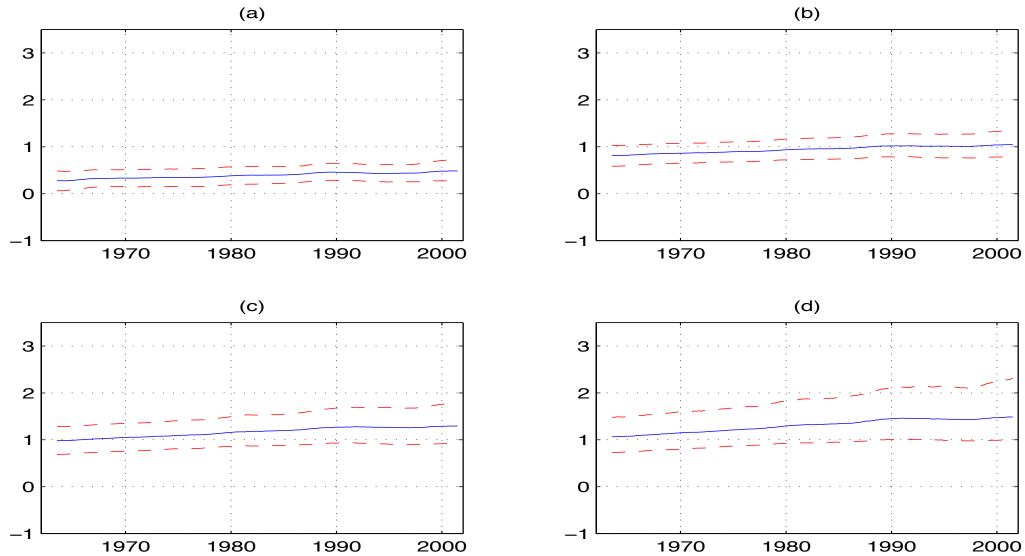


Figure 4: Interest rate response to a 1% permanent increase of inflation with 16th and 84th percentiles. (a) Simultaneous response, (b) response after 10 quarters, (c) response after 20 quarters, (d) response after 60 quarters.

Corrected Algorithm (3)



Original Algorithm

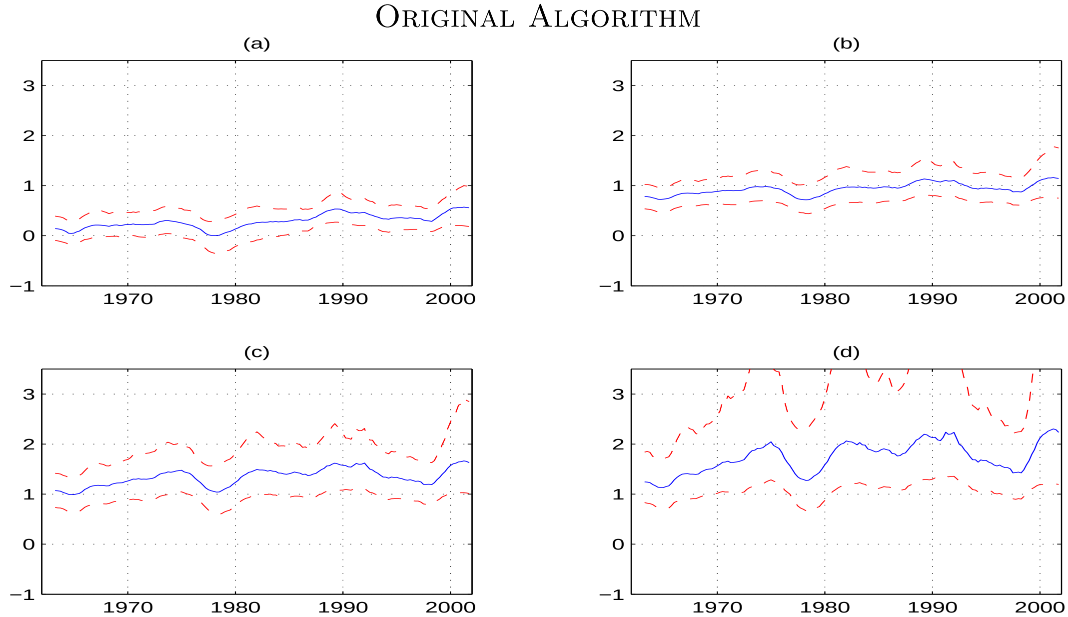
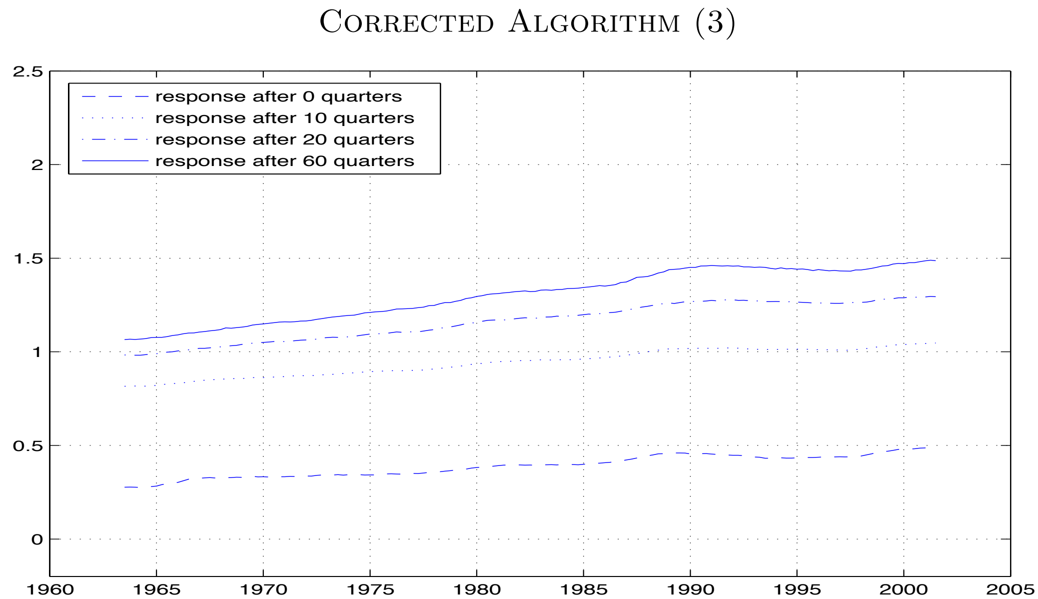


Figure 5: Interest rate response to a 1% permanent increase of unemployment with 16th and 84th percentiles. (a) Simultaneous response, (b) response after 10 quarters, (c) response after 20 quarters, (d) response after 60 quarters.



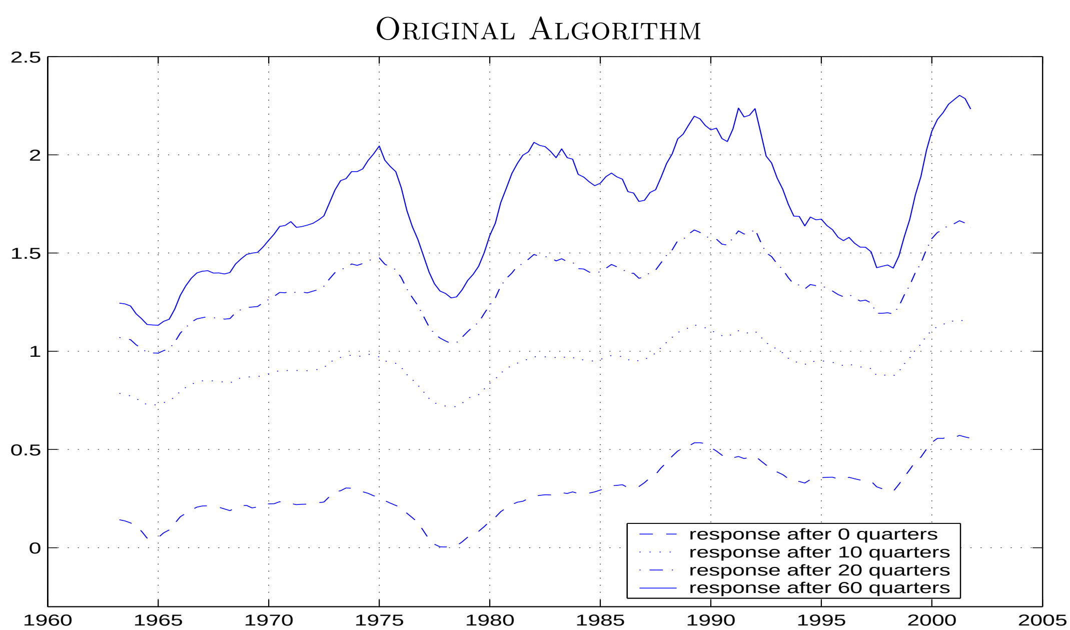
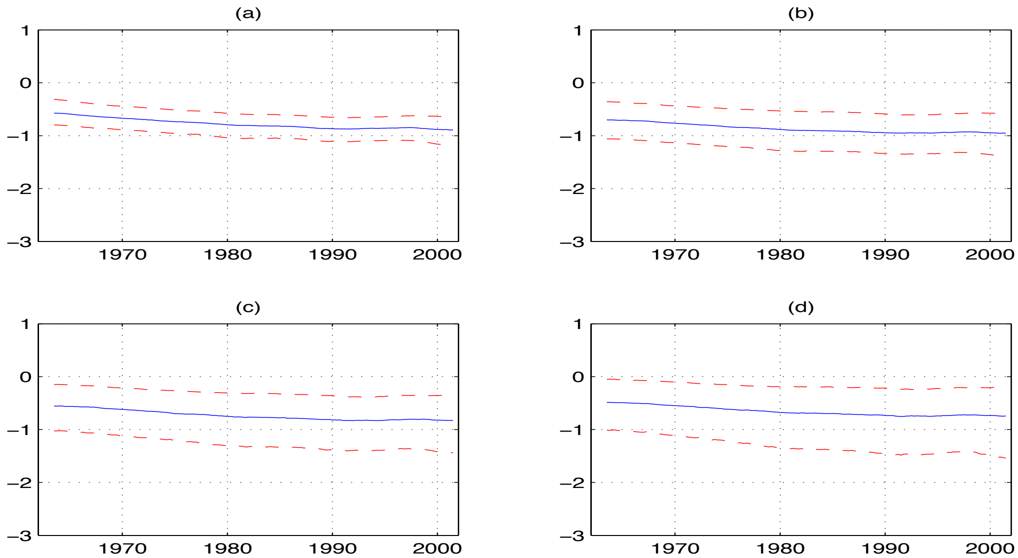


Figure 6: Interest rate response to a 1% permanent increase of unemployment with 16th and 84th percentiles. (a) Simultaneous response, (b) response after 10 quarters, (c) response after 20 quarters, (d) response after 60 quarters.

Corrected Algorithm (3)



Original Algorithm

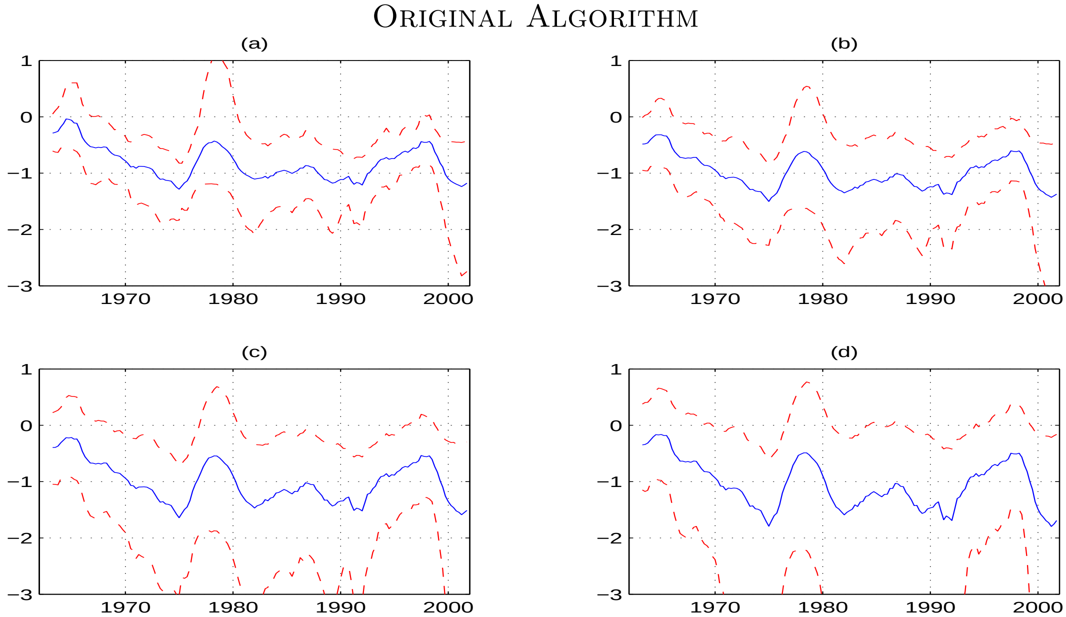


Figure 7: Interest rate response to a 1% permanent increase of unemployment.

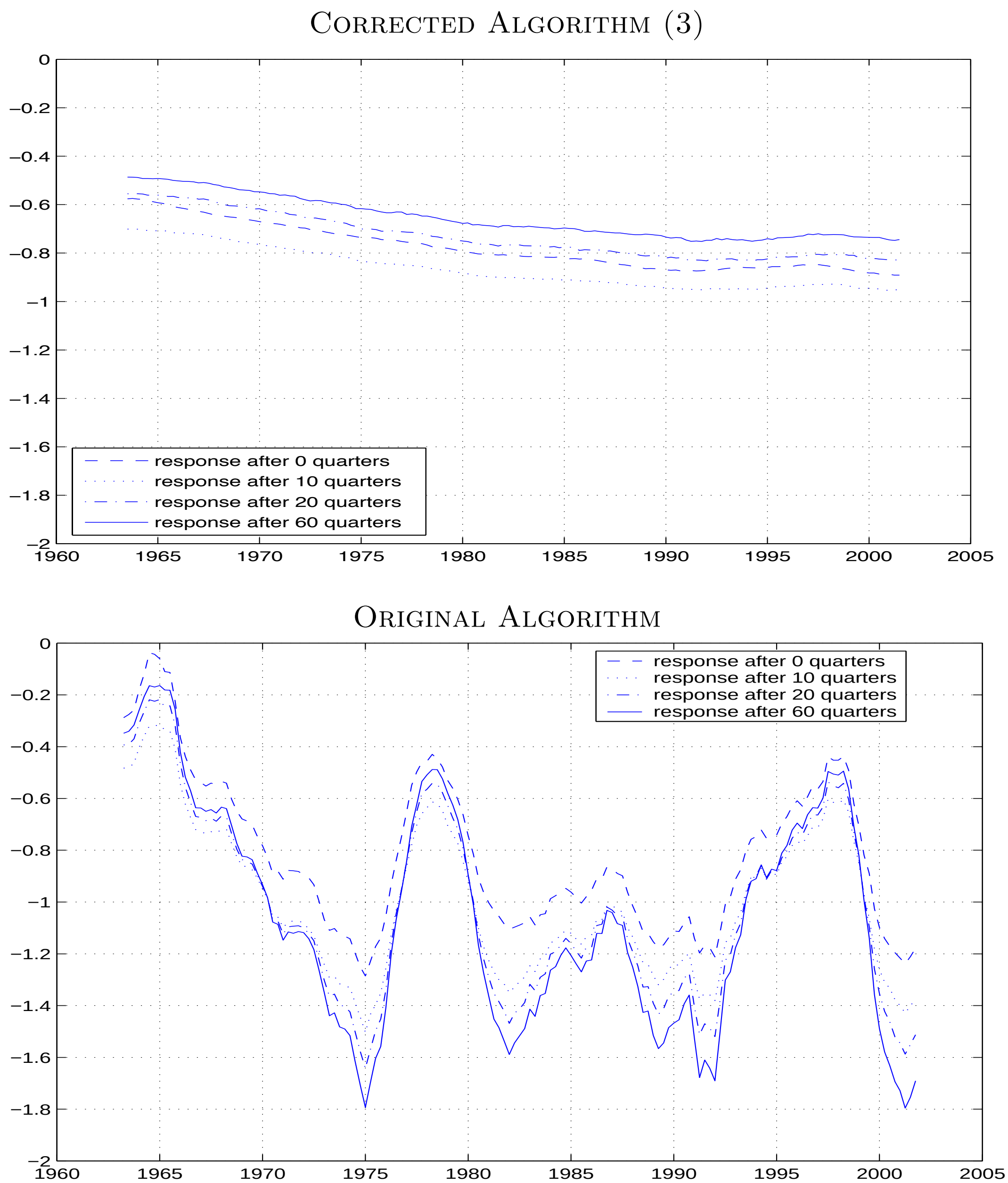
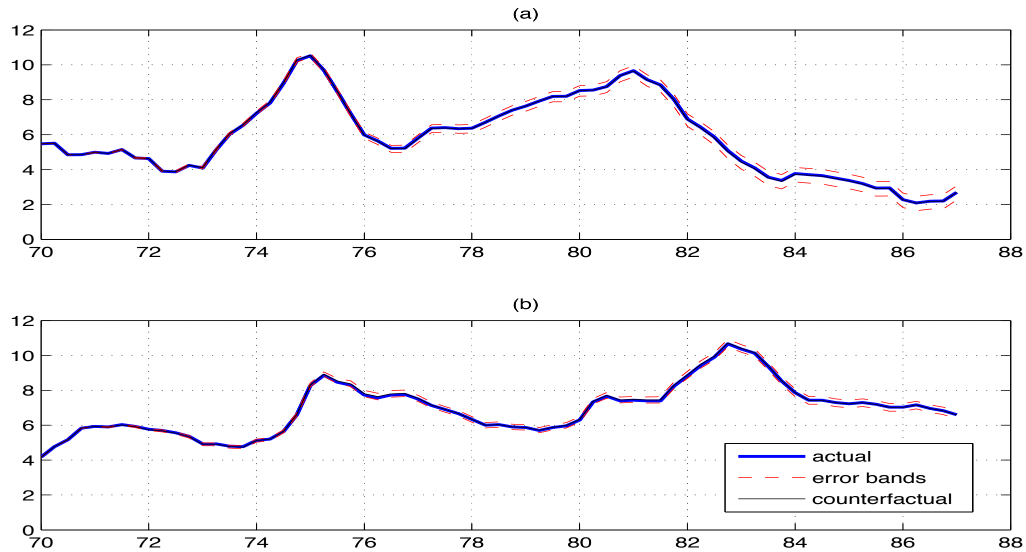


Figure 8: Counterfactual historical simulation drawing the parameters of the monetary policy rule from their 1991-1992 posterior. (a) Inflation, (b) unemployment.

CORRECTED ALGORITHM (3)



ORIGINAL ALGORITHM

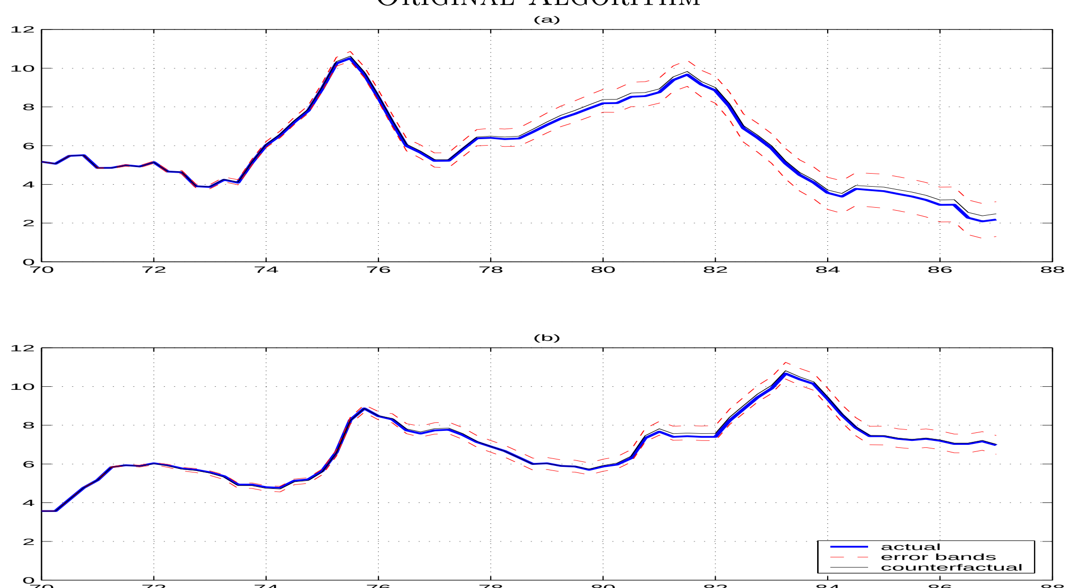
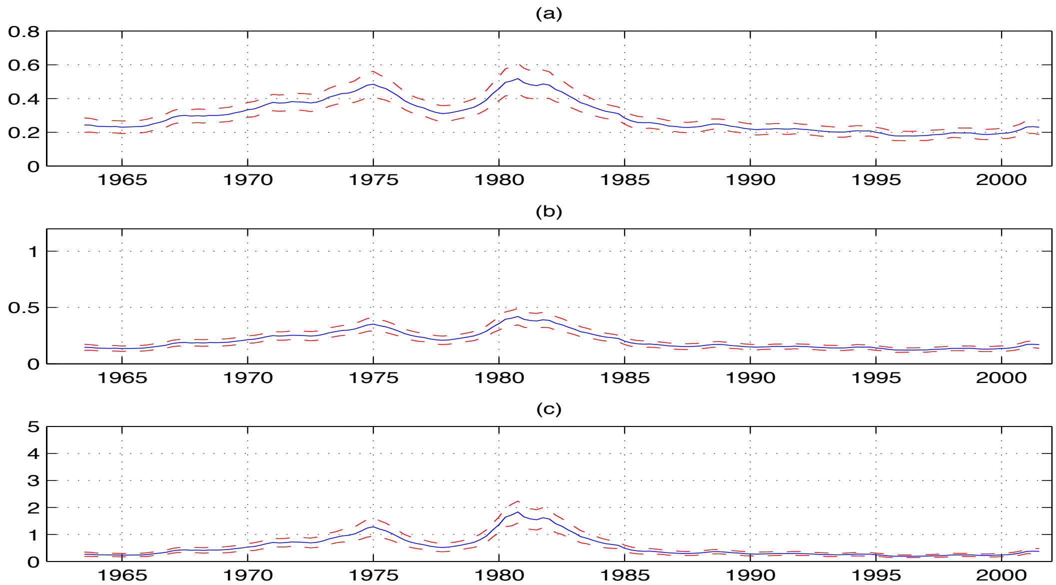


Figure 9: Posterior mean, 16th and 84th percentiles of the standard deviation of (a) the residuals of the inflation equation, (b) the residuals of the unemployment equation and (c) the residuals of the interest rate equation or monetary policy shocks.

Algorithm 3



Algorithm 2

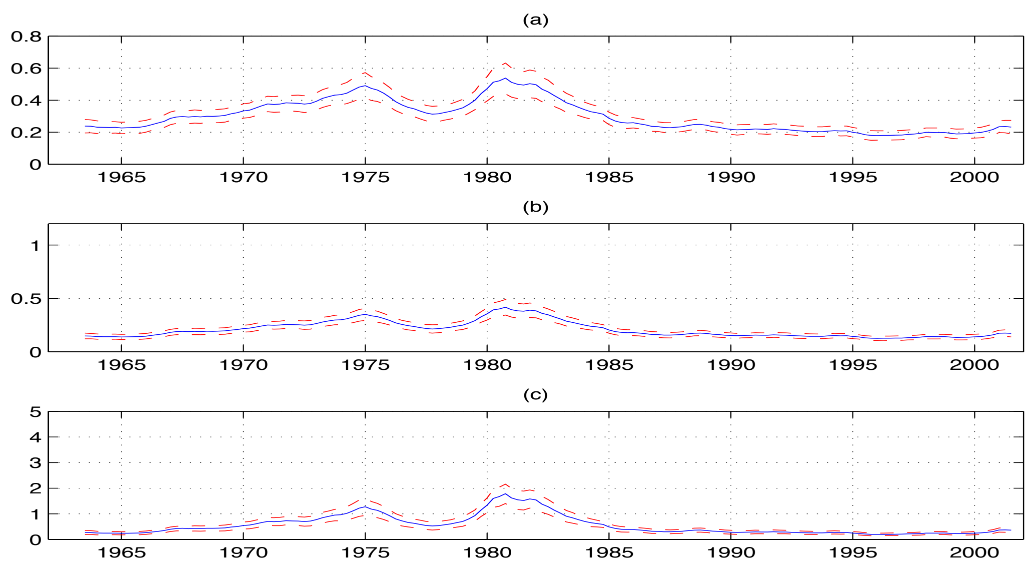
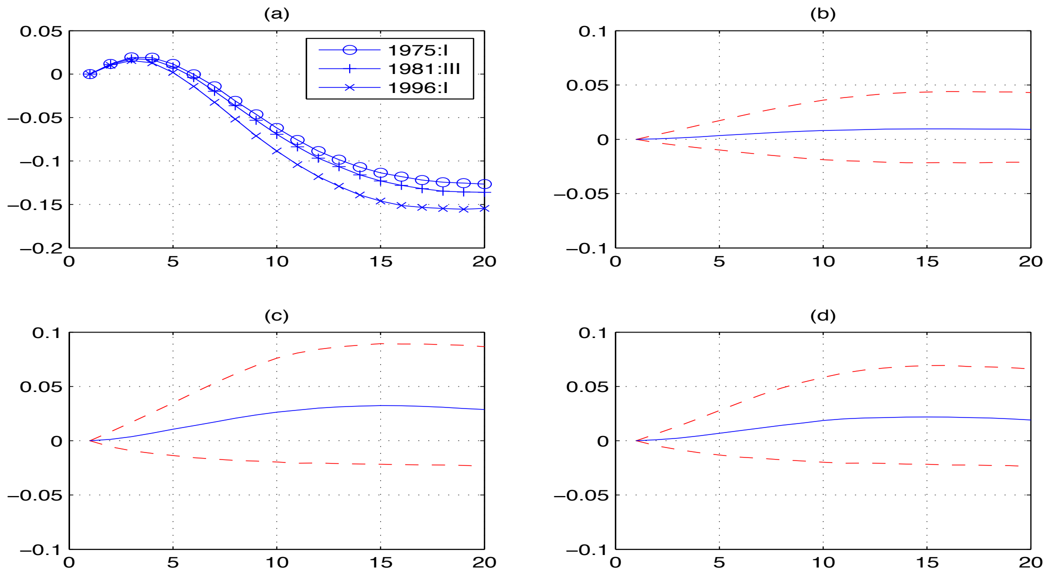


Figure 10: (a) impulse responses of inflation to monetary policy shocks in 1975:I, 1981:III and 1996:I, (b) difference between the responses in 1975:I and 1981:III with 16th and 84th percentiles, (c) difference between the responses in 1975:I and 1996:I with 16th and 84th percentiles, (d) difference between the responses in 1981:III and 1996:I with 16th and 84th percentiles.

Algorithm 3



Algorithm 2

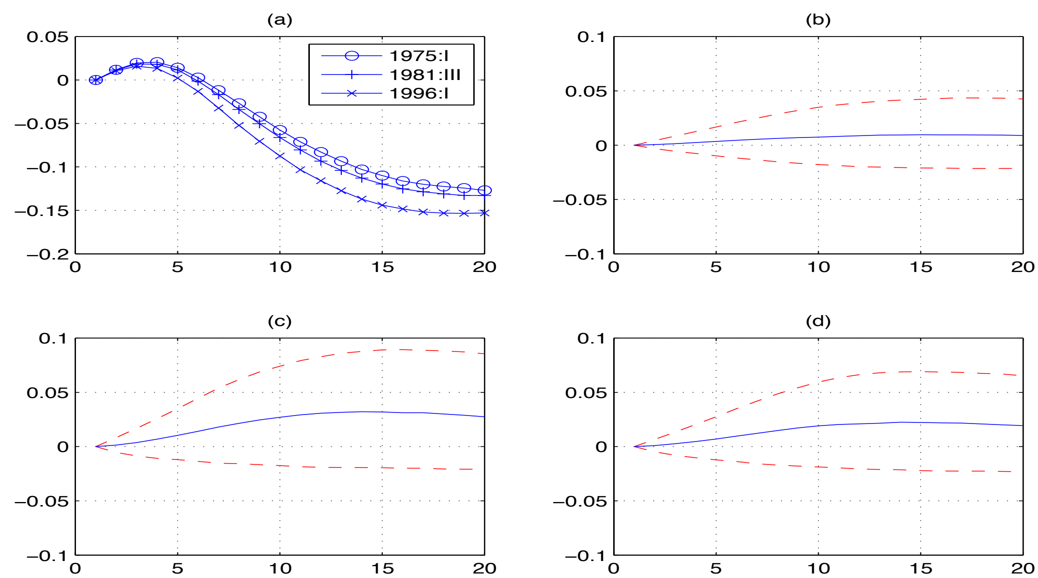
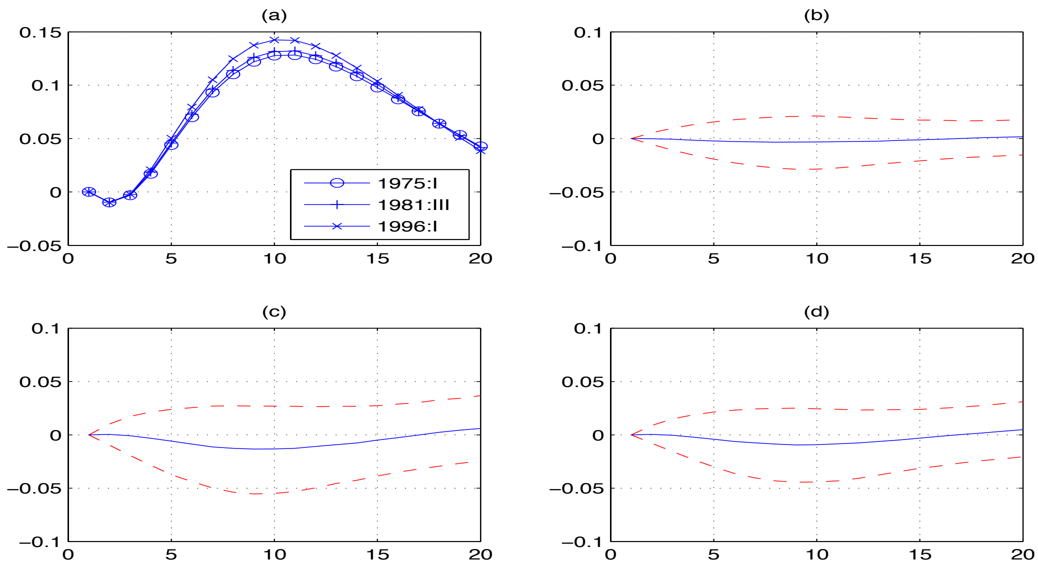


Figure 11: (a) impulse responses of unemployment to monetary policy shocks in 1975:I, 1981:III and 1996:I, (b) difference between the responses in 1975:I and 1981:III with 16th and 84th percentiles, (c) difference between the responses in 1975:I and 1996:I with 16th and 84th percentiles, (d) difference between the responses in 1981:III and 1996:I with 16th and 84th percentiles.

Algorithm 3



Algorithm 2

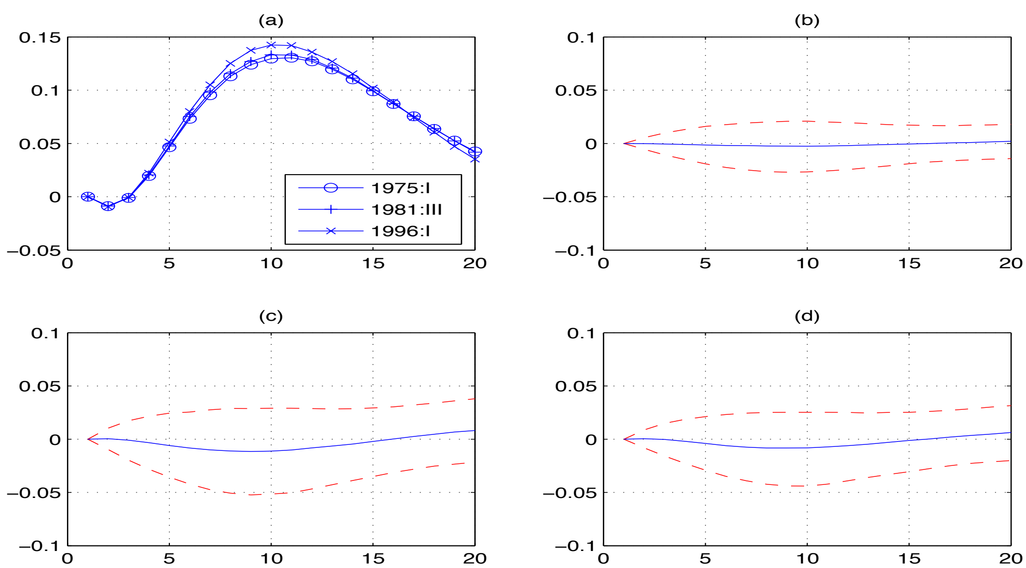
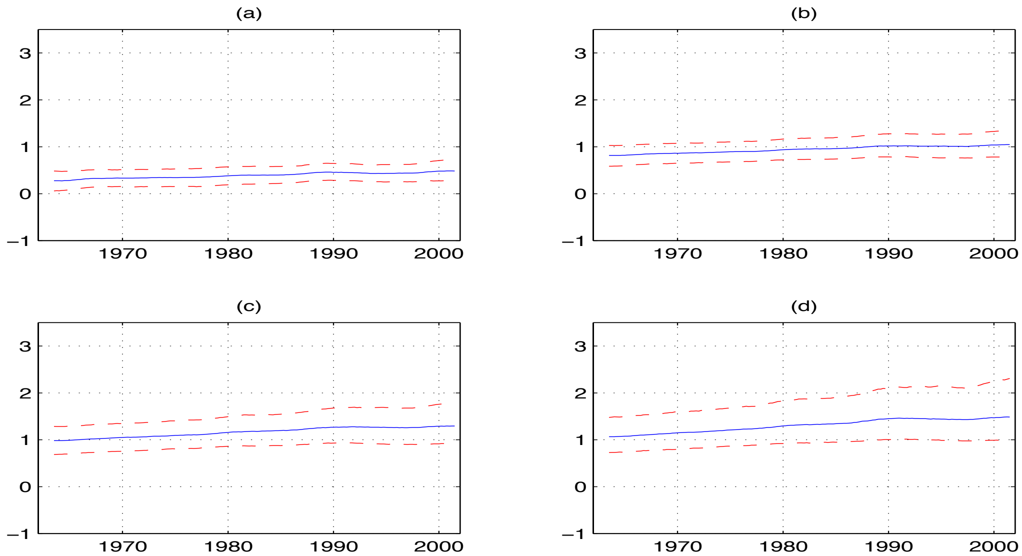


Figure 12: Interest rate response to a 1% permanent increase of inflation with 16th and 84th percentiles. (a) Simultaneous response, (b) response after 10 quarters, (c) response after 20 quarters, (d) response after 60 quarters.

Algorithm 3



Algorithm 2

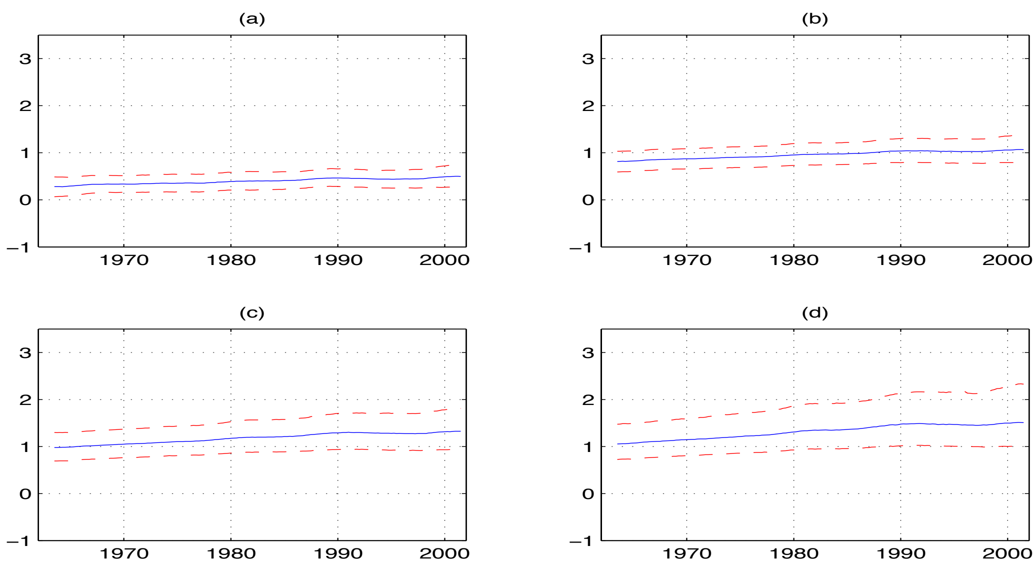
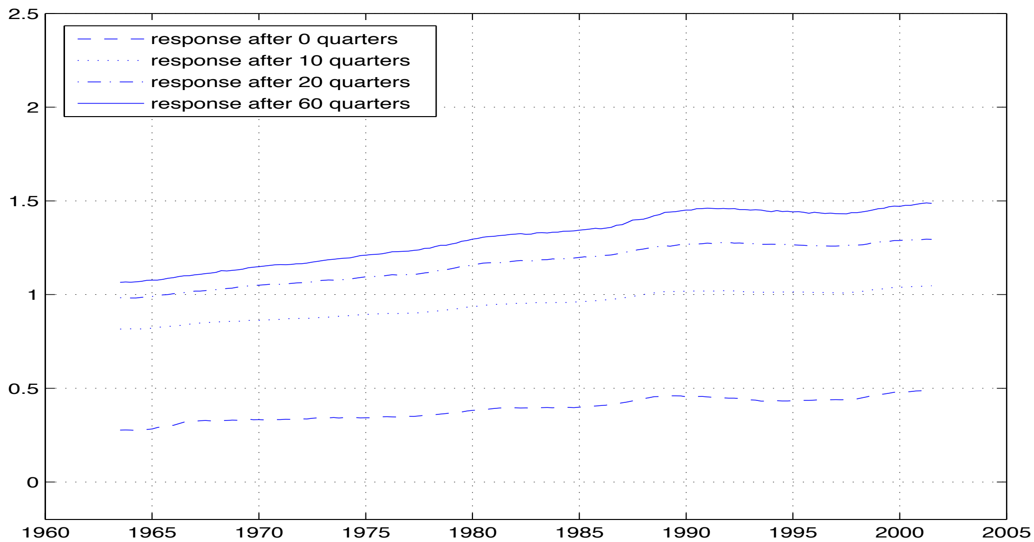


Figure 13: Interest rate response to a 1% permanent increase of unemployment with 16th and 84th percentiles. (a) Simultaneous response, (b) response after 10 quarters, (c) response after 20 quarters, (d) response after 60 quarters.

ALGORITHM 3



ALGORITHM 2

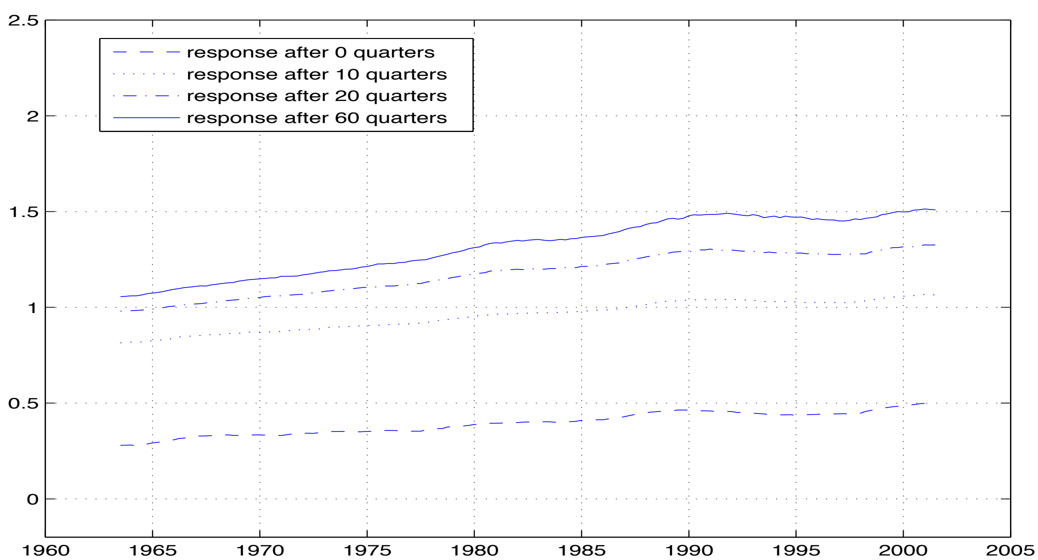
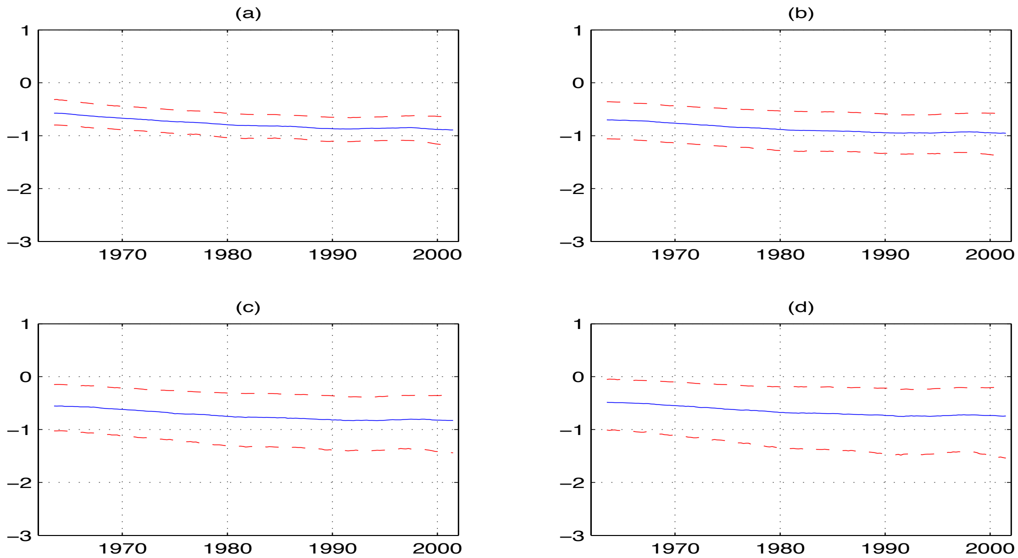


Figure 14: Interest rate response to a 1% permanent increase of unemployment with 16th and 84th percentiles. (a) Simultaneous response, (b) response after 10 quarters, (c) response after 20 quarters, (d) response after 60 quarters.

Algorithm 3



Algorithm 2

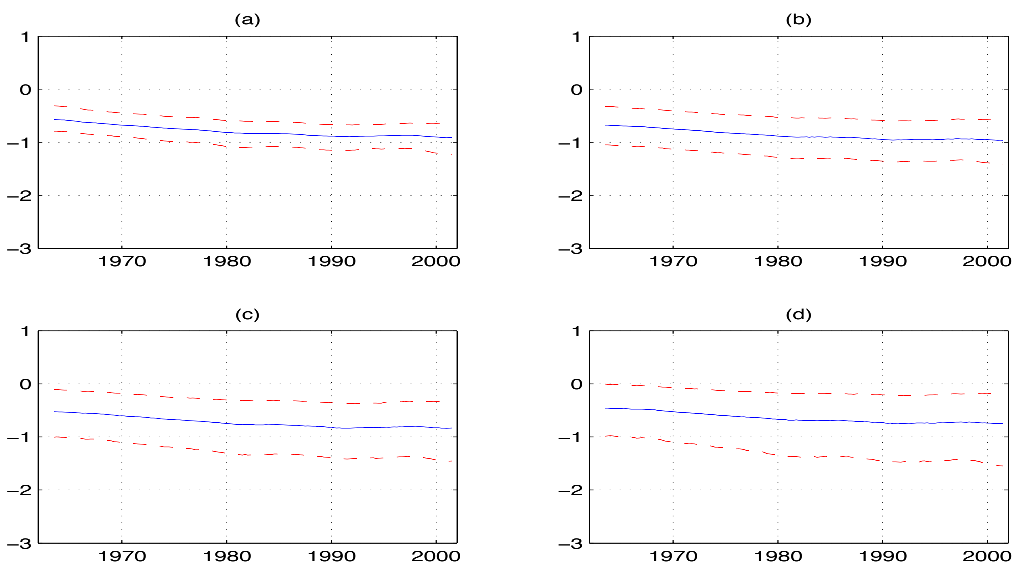
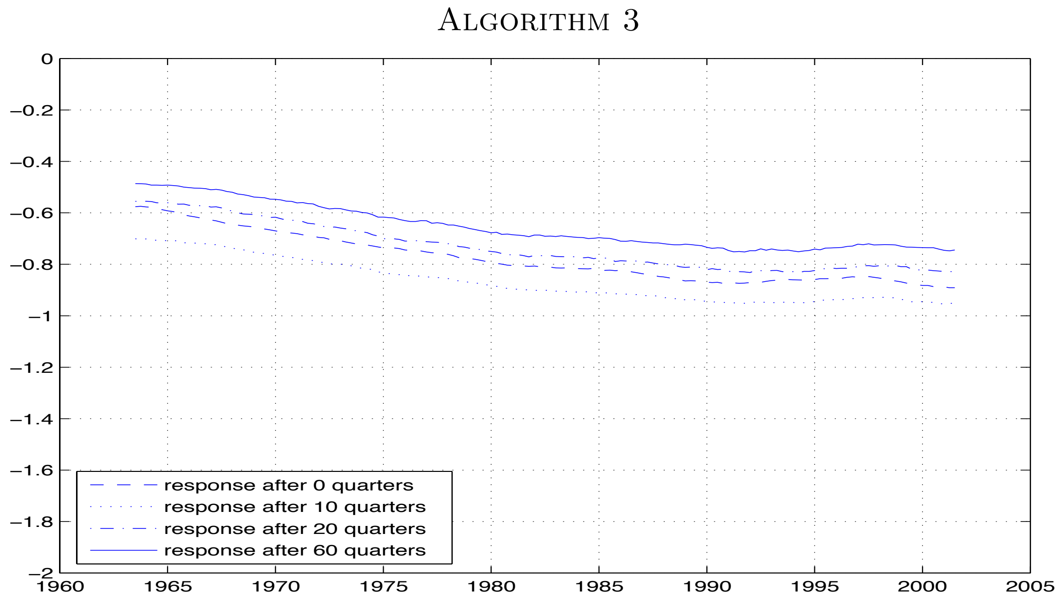


Figure 15: Interest rate response to a 1% permanent increase of unemployment.



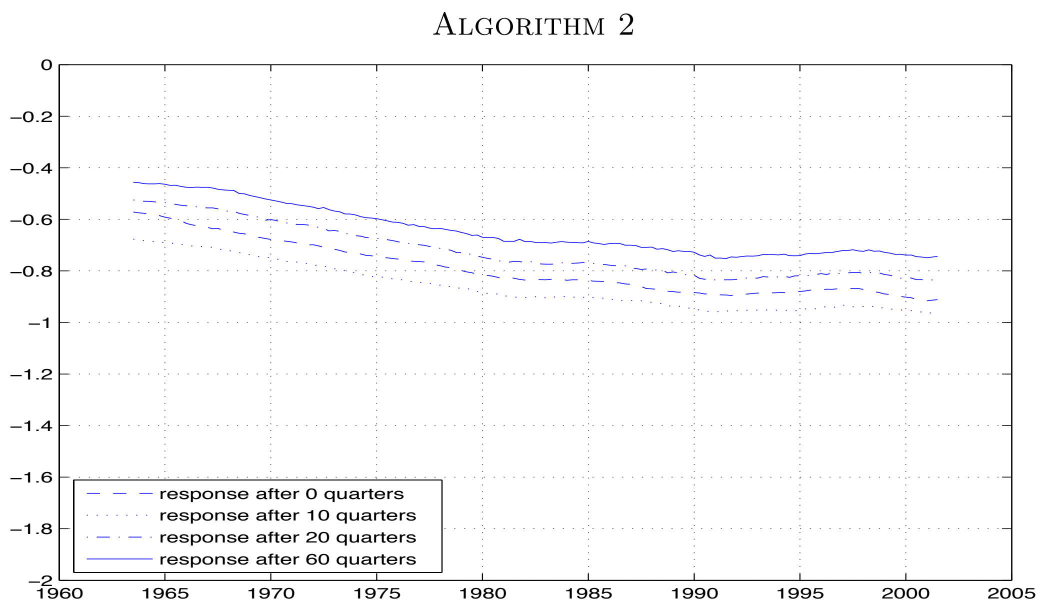
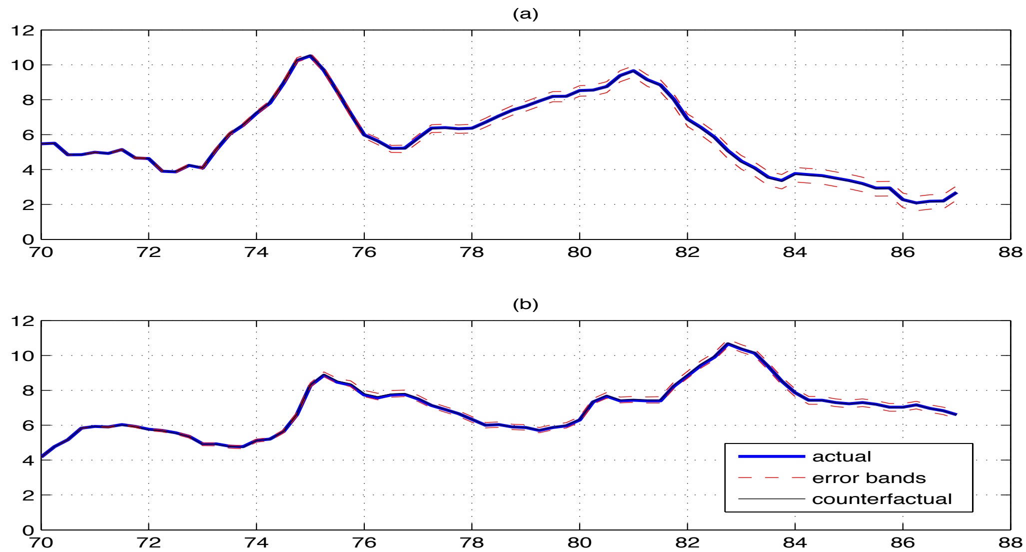


Figure 16: Counterfactual historical simulation drawing the parameters of the monetary policy rule from their 1991-1992 posterior. (a) Inflation, (b) unemployment.

Algorithm 3



Algorithm 2

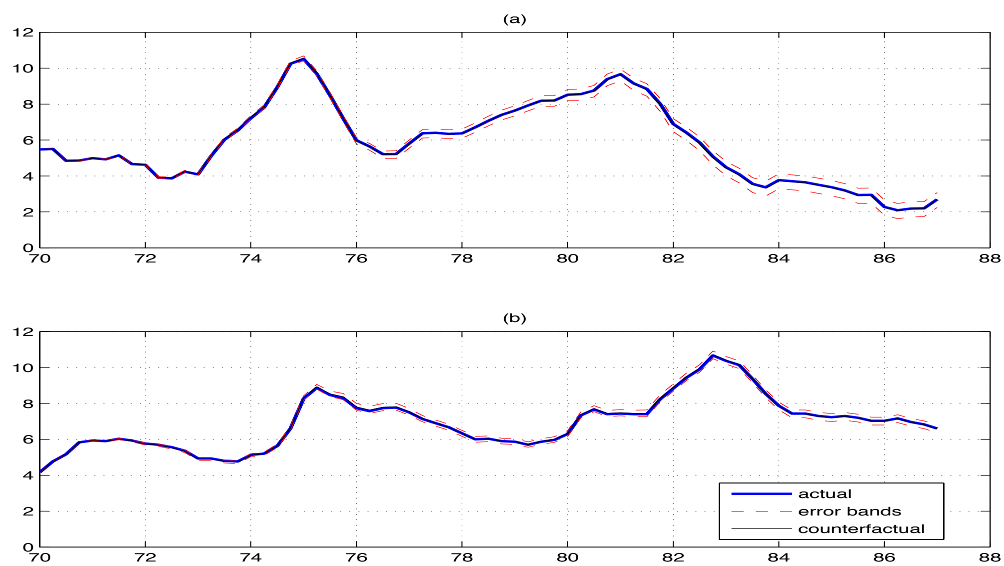


Figure 17: P-P plots obtained by applying the Geweke's (2004) procedure to Algorithm 1. The plots refer to the distribution of $\log \sigma_{i,t}$, with $t = 7$, and $i = 1$ in panel (a), $i = 2$ in panel (b), and $i = 3$ in panel (c).

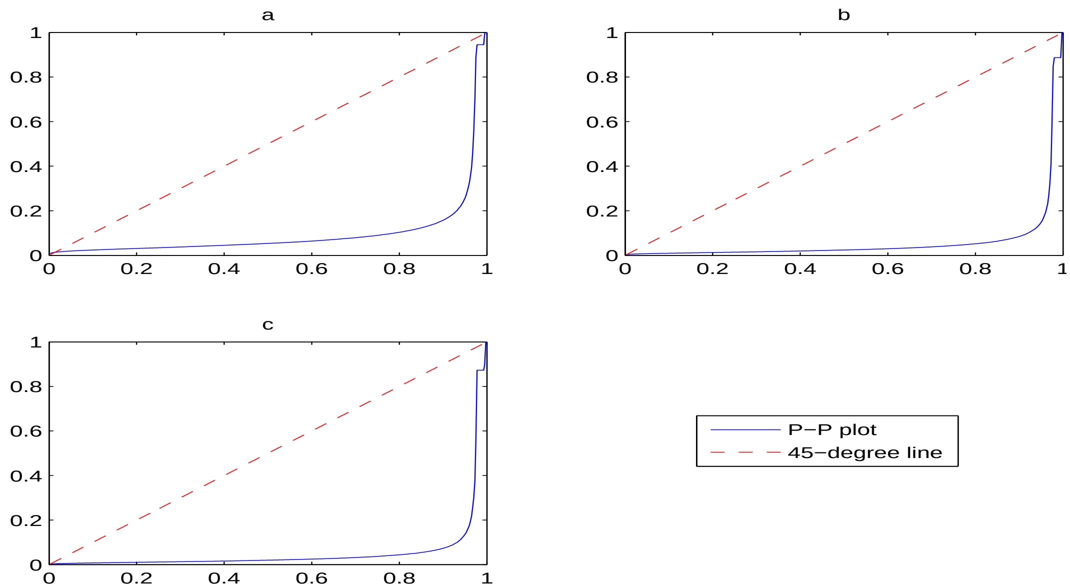


Figure 18: P-P plots obtained by applying the Geweke's (2004) procedure to Algorithm 2. The plots refer to the distribution of $\log \sigma_{i,t}$, with $t = 7$, and $i = 1$ in panel (a), $i = 2$ in panel (b), and $i = 3$ in panel (c).

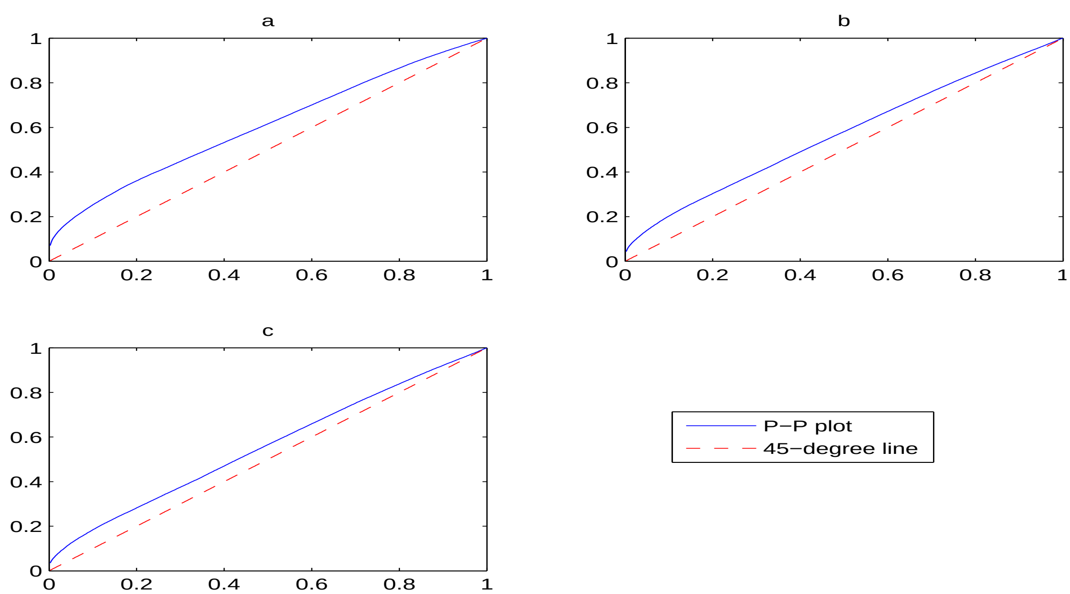


Figure 19: P-P plots obtained by applying the Geweke's (2004) procedure to Algorithm 3. The plots refer to the distribution of $\log \sigma_{i,t}$, with $t = 7$, and $i = 1$ in panel (a), $i = 2$ in panel (b), and $i = 3$ in panel (c).

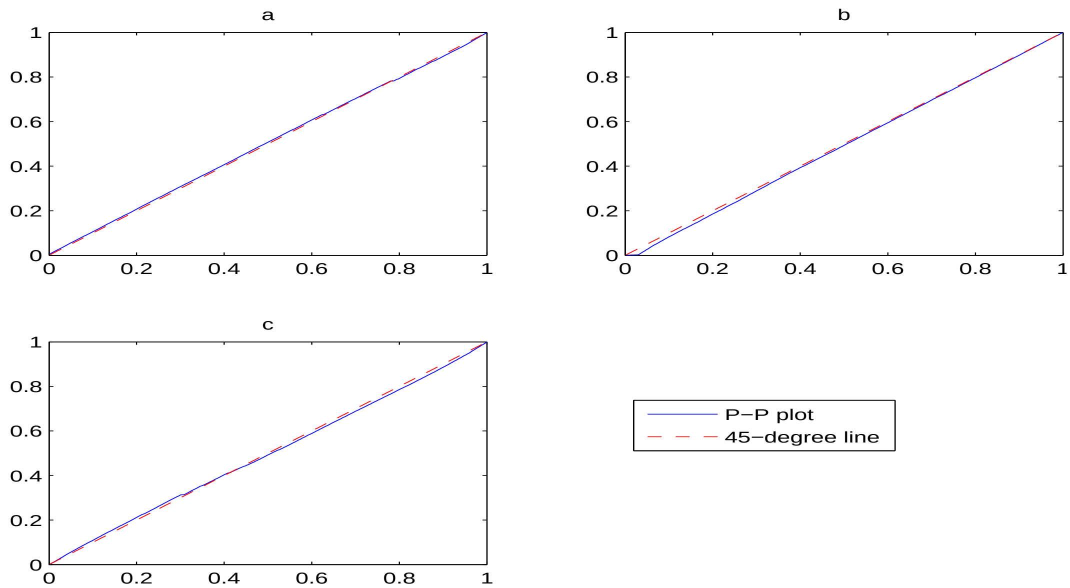


Figure 20: P-P plots obtained by applying the Geweke's (2004) procedure to Algorithm 1 augmented with a Metropolis-Hastings step to correct for the mixture-of-normals approximation error. The plots refer to the distribution of $\log \sigma_{i,t}$, with $t = 7$, and $i = 1$ in panel (a), $i = 2$ in panel (b), and $i = 3$ in panel (c).