Transmission of Government Spending Shocks in the Euro Area: Time Variation and Driving Forces

Technical Annex – Not for Publication

Markus Kirchner Jacopo Cimadomo Sebastian Hauptmeier

June 2010

A Details of the Gibbs Sampler

This annex outlines the details of the Gibbs sampling algorithm used for estimation of the TVP-VAR. The algorithm generates a Markov chain which is a sample the joint posterior distribution of the VAR parameters (i.e. coefficient states, covariance states, volatility states and hyperparameters). It combines elements of Cogley and Sargent (2005), Primiceri (2005) and Benati and Mumtaz (2007), with a few additional restrictions on the structure of the hyperparameters. In what follows, x^t denotes the history of x up to time t, i.e. $x^t = [x_1', x_2', \ldots, x_t']'$ and T denotes the sample length. Furthermore, rewrite the observation equation (2) in the main text conveniently as

$$y_t = X_t'\beta_t + u_t \tag{A.1}$$

where $X_t' = I \otimes [y_{t-1}', \ldots, y_{t-p}', z_t']$. The estimation proceeds in four steps.

Drawing coefficient states β^T. Conditional on A^T and H^T one obtains a history R^T. Then, conditional on y^T, R^T and Q, the observation equation (2) is linear with Gaussian innovations and a known covariance matrix. The posterior density of the coefficients can be factored as[1]

$$f(\beta^T|y^T, R^T, Q) = f(\beta_T|y^T, R^T, Q) \prod_{t=1}^{T-1} f(\beta_t|\beta_{t+1}, y^t, R_t, Q) \tag{A.2}$$

where

$$\begin{aligned} \beta_t|\beta_{t+1,}y^t, R^T, Q &\sim N(\beta_{t|t+1,}P_{t|t+1}) \\ \beta_{t|t+1} &\sim \mathrm{E}[\beta_t|\beta_{t+1}, y^t, R^T, Q] \\ P_{t|t+1} &\sim \mathrm{E}[P_t|P_{t+1}, y^t, R^T, Q] \end{aligned}$$

[1] We omit conditioning factors which are redundant in the respective step.

The conditional means and variances can be computed using the Kalman filter and a backward recursion (see Carter and Kohn, 1994). The Kalman filter delivers

$$\begin{aligned}
P_{t|t-1} &= P_{t-1|t-1} + Q \\
K_t &= P_{t|t-1} X_t (X_t' P_{t|t-1} X_t + R_t)^{-1} \\
\beta_{t|t} &= \beta_{t-1|t-1} + K_t (y_t - X_t' \beta_{t-1|t-1}) \\
P_{t|t} &= P_{t|t-1} - K_t X_t' P_{t|t-1}
\end{aligned}$$

The initial values $\beta_{0|0}$ for this recursion are the OLS point estimates from the initial sample, and the initial value $P_{0|0}$ is their covariance matrix. The initial R_t is the OLS covariance matrix of the reduced-form VAR. The covariance matrix Q is a scaled version of the variance-covariance matrix of the coefficients.

The Kalman filter delivers as its last points $\beta_{T|T}$ and $P_{T|T}$. Draws from (A.2) are then obtained by a backward recursion. The first point in the backward recursion is a draw from $N(\beta_{T|T}, P_{T|T})$. The remaining draws are from $N(\beta_{t|t+1}, P_{t|t+1})$ where the means and variances are derived as follows:

$$\begin{aligned}
\beta_{t|t+1} &= \beta_{t|t} + P_{t|t} P_{t+1|t}^{-1} (\beta_{t+1} - \beta_{t|t}) \\
P_{t|t+1} &= P_{t|t} - P_{t|t} P_{t+1|t}^{-1} P_{t|t}
\end{aligned}$$

Drawing covariance states A^T. Conditional on y^T, β^T and H^T, write the system of equations (A.1) as

$$A_t (y_t - X_t' \beta_t) = A_t \widehat{y}_t = H_t^{1/2} v_t \tag{A.3}$$

Moreover, A_t is lower diagonal (with ones on the main diagonal) such that (A.3) can be rewritten as

$$\widehat{y}_t = Z_t \alpha_t + H_t^{1/2} v_t \tag{A.4}$$

where α_t is defined as in the main text and Z_t has the structure

$$Z_t = \begin{bmatrix}
0 & \cdots & \cdots & 0 \\
-\widehat{y}_{1,t} & 0 & \cdots & \vdots \\
0 & (-\widehat{y}_{1,t}, -\widehat{y}_{2,t}) & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & (-\widehat{y}_{1,t}, \cdots, -\widehat{y}_{n-1,t})
\end{bmatrix}$$

where n denotes the number of variables in the VAR. The system of equations (A.4) has a Gaussian but non-linear state-space form. However, under the assumption of (block) diagonality of S the problem becomes linear (see Primiceri, 2005). The forward (Kalman filter) and backward recursions of the previous step can then be applied equation by equation. Hence, the procedure allows to recover α^T by

$$\begin{aligned}\alpha_{i,t|t+1} &= \mathrm{E}[\alpha_{i,t}|\alpha_{i,t+1}, y^t, \beta^T, H^T, S_i] \\ \Lambda_{i,t|t+1} &= \mathrm{var}[\alpha_{i,t}|\alpha_{i,t+1}, y^t, \beta^T, H^T, S_i]\end{aligned}$$

where $\alpha_{i,t}$ is the block of α_t corresponding to the i-th equation and S_i is the associated i-th block of S. The initial values for the Kalman filter are obtained from a decomposition of the OLS covariance matrix, using the prior mean and the prior variance of α_0 as described in Annex B.

Drawing volatility states H^T. To sample the stochastic volatilities the univariate algorithm of Jacquier, Polson, and Rossi (1994) is applied each element of H_t. The orthogonalized residuals $v_t = A_t u_t$ are observable conditional on y^T, β^T and A^T. We can use the univariate setting because the stochastic volatilities are assumed to be independent, following Cogley and Sargent (2005). Jacquier, Polson, and Rossi (1994) show that the conditional kernel is

$$f(h_{i,t}|h_{-i,t}, v_i^T, w_i) \propto f(h_{i,t}|h_{i,t-1}, h_{i,t+1}, v_i^T, w_i)$$

where w_i is the i-th diagonal element of W and $h_{-i,t}$ represents the vector of h's at all other dates. Using Bayes' theorem the above conditional kernel can be expressed as

$$\begin{aligned}f(h_{i,t}|h_{i,t-1}, h_{i,t+1}, v_i^T, w_i) &\propto f(u_{i,t}|h_{i,t}) f(h_{i,t}|h_{i,t-1}) f(h_{i,t+1}|h_{i,t}) \\ &\propto h_{i,t}^{-1.5} \exp\left(-\frac{v_{i,t}^2}{2h_{i,t}}\right) \exp\left(-\frac{(\ln h_{i,t} - \mu_{i,t})^2}{2\sigma_{ic}^2}\right)\end{aligned} \tag{A.5}$$

where $\mu_{i,t}$ and σ_{ic}^2 are the conditional mean and variance of $h_{i,t}$ implied by equation (5) in the main text and knowledge of $h_{i,t-1}$ and $h_{i,t+1}$. For a geometric random walk these parameters are

$$\mu_{i,t} = 0.5(\log h_{i,t-1} + \log h_{i,t+1}) \quad \text{and} \quad \sigma_{ic}^2 = 0.5 w_i$$

In practice $h_{i,t+1}$ is taken from the previous Gibbs iteration.[2] Jacquier, Polson, and Rossi

[2] In the first iteration, we use squared orthogonalized residuals $v_{i,t}^2$ in order to initialize the volatilities, which

(1994) propose a Metropolis step instead of a Gibbs step, because the normalizing constant is expensive to calculate in (A.5). Hence, one draws from a stand-in density and then uses the conditional likelihood $f(u_{i,t}|h_{i,t})$ to calculate the acceptance probability for that draw. Cogley and Sargent (2005) suggest to use the log-normal implied by equation (5) in the main text as the stand-in density:

$$g(h_{i,t}) \propto h_{i,t}^{-1} \exp\left(-\frac{(\log h_{i,t} - \mu_{i,t})^2}{2\sigma_{ic}^2}\right)$$

The acceptance probability for the m-th draw is

$$\begin{aligned} q_m &= \frac{f(v_{i,t}|h_{i,t}^m)g(h_{i,t}^m)}{g(h_{i,t}^m)} \frac{g(h_{i,t}^{m-1})}{f(v_{i,t}|h_{i,t}^{m-1})g(h_{i,t}^{m-1})} \\ &= \frac{(h_{i,t}^m)^{-1/2} \exp\left(-v_{i,t}^2/2h_{i,t}^m\right)}{(h_{i,t}^{m-1})^{-1/2} \exp\left(-v_{i,t}^2/2h_{i,t}^{m-1}\right)} \end{aligned}$$

where $h_{i,t}^m = h_{i,t}^{m-1}$ if the draw is rejected. This algorithm is applied on a date by date basis to each element of u_t. The formulas are slightly different for the first and last element. For the first element we have

$$\mu_{i1} = \sigma_{ic}^2 \left(\frac{\mu_{i0}}{\sigma_{hi0}^2} + \frac{\log h_{i,t+1}}{w_i}\right) \quad \text{and} \quad \sigma_{ic}^2 = \frac{\sigma_{hi0}^2 w_i}{\sigma_{hi0}^2 + w_i}$$

and the acceptance probability is 1 since there is no previous draw. For the last element we have

$$\mu_{iT} = \log h_{i,t-1} \quad \text{and} \quad \sigma_{ic}^2 = w_i$$

where the prior on the distribution of $\log h_0$, providing values for the mean μ_{i0} and the variance σ_{hi0}^2, is described in Annex B.

Drawing hyperparameters. The hyperparameters of the model are the covariance matrices of the innovations, i.e. Q (coefficient states), S (covariance states), and W (volatility states). Conditional on y^T, β^T, A^T and H^T, these state innovations are observable. Since the hyperparameters are assumed to be independent, each covariance matrix can be drawn from its respective distribution.

Since we have restricted the hyperparameter matrix Q to be diagonal, its diagonal elements q_i have univariate inverse Gamma distributions with scale parameter $\gamma_{i,1}^q$ and degrees of freedom

are obtained from the application of the OLS estimates from the initial sample on the actual sample.

δ_1^q:

$$f(q_i|y^T, \beta^T) = \text{IG}\left(\frac{\gamma_{i,1}^q}{2}, \frac{\delta_1^q}{2}\right)$$

where $\delta_1^q = \delta_0^q + T$ and $\gamma_{i,1}^q = \gamma_{i,0}^q + \sum_{t=1}^T \varepsilon_{i,t}^2$ (see e.g. Kim and Nelson, 1999).

Similarly, restricting S to be diagonal, each of its diagonal elements s_i has an inverse Gamma distribution with scale parameter $\gamma_{i,1}^s$ and degrees of freedom δ_1^s:

$$f(s_i|y^T, A^T) = \text{IG}\left(\frac{\gamma_{i,1}^s}{2}, \frac{\delta_1^s}{2}\right)$$

where $\delta_1^s = \delta_0^s + T$ and $\gamma_{i,1}^s = \gamma_{i,0}^s + \sum_{t=1}^T \nu_{i,t}^2$.

Finally, the diagonal elements w_i of W have univariate inverse Gamma distributions with scale parameter $\gamma_{i,1}^w$ and degrees of freedom δ_1^w:

$$f(w_i|y^T, H^T) = \text{IG}\left(\frac{\gamma_{i,1}^w}{2}, \frac{\delta_1^w}{2}\right)$$

where $\delta_1^w = \delta_0^w + T$ and $\gamma_{i,1}^w = \gamma_{i,0}^w + \sum_{t=1}^T \omega_{i,t}^2$.

Summary. The Gibbs sampling algorithm is summarized as follows:

1. Initialize R^T, Q, S and W.
2. Draw coefficients β^T from $f(\beta^T|y^T, R^T, Q)$.
3. Draw covariances A^T from $f(A^T|y^T, H^T, S)$.
4. Draw volatilities H^T from $f(H^T|y^T, \beta^T, A^T, W)$.
5. Draw hyperparameters from $f(q_i|y^T, \beta^T)$, $f(s_i|y^T, A^T)$ and $f(w_i|y^T, H^T)$.
6. Go to step 2.

B Calibration of the Priors

This annex discusses the choice of our priors. We closely follow common choices in the TVP-VAR literature and impose relatively conservative priors, particularly on the hyperparameters (see e.g. Cogley and Sargent, 2001; Cogley and Sargent, 2005; Primiceri, 2005; Benati and Mumtaz, 2007). However, unlike most previous studies those priors are not calibrated based on OLS estimates from an initial "training sample" which is then discarded. This would mean sacrificing part of our already relatively short sample. Instead, we calibrate our priors based on

OLS estimates from the full sample. Such a strategy is suggested by Canova (2007) and Canova and Ciccarelli (2009) for cases where a training sample is not available. A fixed-coefficient VAR model is thus estimated by OLS (equation by equation) on the full sample from 1980Q1-2008Q4.

VAR coefficients. Let $\hat{\beta}$ denote the OLS estimate of the VAR coefficients, and $\widehat{\Xi}$ their covariance matrix. We set

$$\beta_0 \sim \mathrm{N}(\hat{\beta}, 4 \times \widehat{\Xi})$$

where the variance scaling factor increases the uncertainty about the size of the VAR coefficients in the initial sample versus the actual sample.

Elements of H_t. Denote the OLS estimate of the VAR covariance matrix as $\widehat{\Sigma}$. We apply a triangular decomposition of this matrix similar to equation (4) in the main text, $\widehat{\Sigma} = \widehat{\Psi}^{-1}\widehat{\Phi}(\widehat{\Psi}^{-1})'$, and denote the vector of diagonal elements of $\widehat{\Phi}$ as ϕ_0. Our prior for the diagonal elements of the matrix H_t is

$$h_0 \sim \mathrm{N}(\phi_0, 10 \times I)$$

The variance scaling factor 10 is arbitrary but large relative to the mean ϕ_0.

Elements of A_t. Denote the vector of non-zero off-diagonal elements of $\widehat{\Psi}$ as ψ_0, ordered by rows. The prior for the elements of A_t is

$$\alpha_0 \sim \mathrm{N}\left(\psi_0, 10 \times \mathrm{diag}(\psi_0)\right)$$

where the variance of α_0 is scaled up taking into the magnitude of the respective elements of the mean ψ_0, as in Benati and Mumtaz (2007).

Hyperparameters. The prior on the diagonal elements of the coefficient state error variance Q is also inverse Gamma:

$$q_i \sim \mathrm{IG}\left(\frac{\gamma_{i,0}^q}{2}, \frac{\delta_0^q}{2}\right)$$

where $\gamma_{i,0}^q = k_Q \times \hat{\xi}_i$, where $\hat{\xi}_i$ denotes the i-th diagonal element of the OLS covariance matrix $\widehat{\Xi}$ and $k_Q = 10^{-4}$. Hence, our prior attributes only 0.01% of the uncertainty surrounding the OLS estimates to time variation following Cogley and Sargent (2001). The degrees of freedom δ_0^q are set to 1, which is the minimum for the prior to be proper. We thus put as little weight on the prior as possible.

The prior on the diagonal elements of the hyperparameter matrix S for the covariance states

is also inverse Gamma:

$$s_i \sim \text{IG}\left(\frac{\gamma_{i,0}^s}{2}, \frac{\delta_0^s}{2}\right)$$

where $\gamma_{i,0}^s = k_S \times \hat{\psi}_i$, where $\hat{\psi}_i$ denotes the i-th diagonal element of the OLS covariance matrix $\widehat{\Psi}$ and $k_S = 10^{-2}$. Here we follow Primiceri (2005), who makes similar choices for a block diagonal structure of S. The degrees of freedom δ_0^s are again set to the minimum value of 1.

The prior on the diagonal elements of the variance W for the volatility states is inverse Gamma:

$$w_i \sim \text{IG}\left(\frac{\gamma_{i,0}^w}{2}, \frac{\delta_0^w}{2}\right)$$

where $\gamma_{i,0}^w = k_W$. We set $k_W = 10^{-4}$ and $\delta_0^w = 1$. The parameters of the distribution are the same as in Cogley and Sargent (2005) and Benati and Mumtaz (2007).

C Convergence of the Markov Chain

This annex assesses the convergence of the Markov chain produced by the Gibbs sampler. We apply three types of convergence checks to the VAR coefficients, the covariances and the volatilities.[3] We omit the hyperparameters, since these are not the direct objects of interest.

The first convergence check is the diagnostics due to Raftery and Lewis (1992), which is used to assess the total number of iterations required to achieve a certain precision, and the minimum burn-in period and thinning factor. The parameters for the diagnostic are specified as follows: quantile = 0.025; desired accuracy = 0.025; required probability of attaining the required accuracy = 0.95. We generate a Markov chain with 5,000 draws as suggested by Raftery and Lewis (1992) which is used as an input for the diagnostics. Table C.1 reports the diagnostics. For all three state vectors, the required number of runs is far below the total number of iterations actually applied. The same holds for the number of burn-in replications and the thinning factor. The choices made to generate the Markov chain therefore seem appropriate.

Our second convergence diagnostic are the inefficiency factors (IFs) for the posterior estimates of the parameters. The IF is the inverse of Geweke's (1989) relative numerical efficiency measure, i.e. $IF = 1 + 2\sum_{k=1}^{\infty} \rho_k$, where ρ_k is the k-th order autocorrelation of the chain. This diagnostic therefore serves to judge how well the chain mixes. Low autocorrelations suggest that the draws are close to independent, which increases the efficiency of the algorithm (Primiceri, 2005). We use a 4% tapered window for the estimation of the spectral density at frequency zero. Values of the IFs below or around 20 are regarded as satisfactory, according to Primiceri (2005). The left panels of Figure C.1 report the IFs for the state vectors. The IFs are

[3] See Koop (2003), chapter 4, for a review of convergence diagnostics.

far below 20 for the coefficients and the covariances, but around 30-35 for the volatilities. Compared to the results reported e.g. in Primiceri (2005) and considering the higher dimensionality of our problem, however, these results still seem satisfactory.

The final convergence test applied is the convergence diagnostic (CD) due to Geweke (1992). This diagnostic is based on the idea that, if a sufficiently large number of draws have been taken, the posterior estimates based on the first half of draws should be essentially the same as the estimates based on the second half of draws. If they are very different, either too few draws have been taken and estimates are inaccurate or the effects of the initial values of the chain have not worn off (Koop, 2003). We therefore divide the 1,000 draws from the posterior distribution into a first set of $N_1 = 100$ draws, a middle set of 500 draws and a last set of $N_2 = 400$ draws as suggested by Koop (2003). We drop the middle set of draws and therefore make it likely that the first and last set are independent of each other. The convergence diagnostic is given by

$$CD = \frac{\widehat{\theta}_1 - \widehat{\theta}_2}{\widehat{\sigma}_1/\sqrt{N_1} + \widehat{\sigma}_2/\sqrt{N_2}} \to \mathrm{N}(0,1)$$

by a central limit theorem, where $\widehat{\theta}_i$ and $\widehat{\sigma}_i/\sqrt{N_i}$ denote the posterior means of the parameters and their numerical standard errors based on the i-th set of draws, for $i = 1, 2$. We plot the p-values for the null hypothesis that the set of draws are the same in the right panels of Figure C.1. The p-values are typically larger than conventional significance levels for the VAR coefficients and the covariances, indicating that a sufficiently large number of draws has been taken for these parameters. However, the null hypothesis is often rejected for the volatilities.

To summarize, the coefficients and covariances have in general better convergence properties than the volatilities. Since the focus of our analysis is on impulse responses which are determined by the contemporaneous relations among variables and the propagation mechanism rather than the size of stochastic shocks we conclude that the convergence properties of the Markov chain are satisfactory.

D Additional Robustness Checks

Sign restrictions. We now compare the benchmark TVP-VAR results to results obtained using an alternative identification approach, based on imposing sign restrictions on the estimated impulse responses as in Uhlig (2005) and Mountford and Uhlig (2009). In fact, one might deem the causal ordering of variables implied by the recursive approach as too restrictive. The identification via sign restrictions has the advantage that zero restrictions need not be imposed. Following Mountford and Uhlig (2009), the government spending shock is identified by the

assumption that it is orthogonal to a business cycle shock and that it raises government spending during at least four quarters. The business cycle shock is identified by the requirement that it raises output and consumption during at least four quarters.

Formally, decompose the matrix C_t in $u_t = C_t e_t$ as $C_t = K_t \Lambda_t$, where K_t is the lower triangular Cholesky factor of the reduced-form covariance matrix R_t and Λ_t is an orthonormal matrix satisfying $\Lambda_t \Lambda_t' = I$. This matrix collects the identifying weights on the Cholesky factor in its columns λ_t which can be used to calculate impulse vectors to unit shocks to the elements of e_t, i.e. the columns of the matrix C_t. Uhlig (2005) shows that the responses to these impulse vectors can be written as linear combinations of the impulse responses to the Cholesky decomposition of R_t. In what follows, define $r_{ji}(k)$ as the impulse response of the j-th variable to the i-th column of K_t, and the 4-dimensional column vector $r_i(k)$ as $[r_{1i}(k), \ldots, r_{4i}(k)]'$. Since we are only interested in two structural shocks we only derive two impulse vectors satisfying the identifying restrictions, applying the following filter:

1. Draw β_t and R_t from their posterior distribution, and compute the lower triangular matrix K_t such that $K_t K_t' = R_t$.

2. Draw $\xi_{1,t}$ from a 4-variate standard normal distribution and calculate the vector $\lambda_{1,t} = \xi_{1,t} / \|\xi_{1,t}\|$.

3. Compute $r_{\lambda_{1,t}}(k) = \sum_{i=1}^{4} \lambda_{1i,t} r_i(k)$, where $\lambda_{1i,t}$ denotes the i-th element of $\lambda_{1i,t}$. Check whether $r_{\lambda_{1,t}}(k)$ satisfies the sign restrictions for the business cycle shock for $k = 1, \ldots, 4$. If so, go to step 4. Otherwise, go back to step 2.

4. Calculate $\lambda_{2,t}$ satisfying $\lambda_{2,t}' \lambda_{2,t} = 1$ and $\lambda_{2,t}' \lambda_{1,t} = 0$ (normality and orthogonality to the business cycle shock). Compute $r_{\lambda_{2,t}}(k) = \sum_{i=1}^{4} \lambda_{2i,t} r_i(k)$ and check whether it satisfies the sign restrictions for the spending shock. If so, save the impulse responses and go to step 5. Otherwise, repeat step 4.

5. Repeat steps 1-4 until 1,000 draws of impulse responses satisfying all sign restrictions are obtained for each $t = 1, \ldots, T$.[4]

In practice, steps 2-3 are repeated a maximum of 100 times before taking a new draw from the posterior if the sign restrictions are never satisfied for the current draw. Step 4 is implemented using a Gauss-Newton minimization routine.[5]

[4] Notice that, similarly as in the benchmark case, we apply a local approximation to the impulse responses at time t, using the reduced-form covariance matrix and the VAR coefficients at time t to calculate the implied response at horizon $t + k$, for $k = 0, \ldots, 20$ quarters.

[5] The MATLAB function `fsolve` is applied.

The results of this exercise are reported in Figure D.1, in terms of impulse responses at selected horizons. The figure shows that there is more high-frequency time variation in the estimated responses, compared to the results reported in the main text. However, it should be noted (not reported) that the impulse responses are estimated with significantly less precision than in the benchmark case, so this result should be taken with caution. The overall time evolution of the impulse responses, in terms of the order of magnitude of the estimated responses, is however very similar as in the benchmark case. We therefore conclude that the TVP-VAR results are broadly robust to using this alternative identification approach.

Imposing stationarity conditions. Cogley and Sargent (2001) propose to impose a prior restriction on the VAR coefficients saying that draws from the Gibbs sampler which do not satisfy stationarity conditions are discarded. We have argued in the main text that such a restriction is difficult to defend for aggregate euro area fiscal data since there may have been instability in the effects of fiscal policy in some countries in the past. The potential downside of not imposing the stationarity conditions is that this may exaggerate time variation due to a potentially large amount of unstable draws. We therefore check the robustness of the TVP-VAR results to imposing the stationarity conditions.

Formally, the random walk process (3) in the main text for the VAR coefficients β_t, $t = 1, \ldots, T$, characterizes the conditional density $f(\beta_t|\beta_{t-1}, Q)$. Following Cogley and Sargent (2001), introduce an indicator function $I(\beta_t)$ which rejects unstable draws not satisfying standard eigenvalue stability conditions and which thus enforces stationarity of the estimated TVP-VAR at each point of time. The VAR coefficients are thus postulated to evolve according to

$$p(\beta_t|\beta_{t-1}, Q) = I(\beta_t) f(\beta_t|\beta_{t-1}, Q)$$

Figure D.2 shows state-dependent spending multipliers, at selected horizons, with the stationarity conditions imposed. A comparison with the results in the main text indicates no significant differences to the benchmark case. The multipliers show somewhat less high-frequency variation, but the broad patterns are similar.

Training sample. In the benchmark specification, the priors are calibrated based on OLS estimates on the full sample. Alternatively, we check the robustness of the TVP-VAR results to the use of a shorter sample closer to the initial date for the calibration of the prior means of β_0, h_0 and α_0 as well as the priors on the hyperparameters. In particular, the prior is now calibrated using the first half of the sample, i.e. 1980Q1-1994Q2. The results are reported in Figure

D.3. In comparison to the results in the main text, the short-term multipliers on output and consumption show less time variation, but the long-term multipliers are again clearly declining. On the other hand, the interest rate response shows significantly less changes over time.

Scaling factors. We have elicited relative conservative priors in terms of time variation, in particular the scaling factors k_Q, k_S and k_W which parameterize the priors on the covariance matrices of the shocks in the state equations. The benchmark specification has $k_Q = k_W = 10^{-4}$ and $k_S = 10^{-2}$, following the literature. Instead, here we reduce the scaling factors one at a time to $k_Q = 5^{-5}$, $k_S = 5^{-3}$ and $k_W = 5^{-5}$, keeping the remaining two factors fixed at their benchmark values. The results are shown in Figures D.4, D.5 and D.6, respectively. It turns out that the reduction of the coefficients scaling factor k_Q actually increases time variation e.g. in the short-term multiplier on output and consumption. It is possible that some of the time variation in the coefficients thus goes into the covariance states instead. On the other hand, reduction of k_S and k_W leads to only small changes in the amount and direction of time variation.

No standard error adjustment. Since the dependent variables in the second-stage regression step are estimated parameters, in the main text we have taken into account the uncertainty on those variables through an appropriate standard error adjustment. That is, we have generated draws from the posterior distribution of regression coefficients conditional on the full posterior distribution of estimated multipliers. One may however ask how the results would look like without this adjustment. Therefore, in Tables C.2 and C.3 we reproduce the regression results in the main text, using as dependent variables the posterior means of the estimated multipliers instead of their full posterior distribution.

In comparison to the benchmark results in the main text, there are some (small) differences in the size of the estimated coefficients, but their signs are identical. The standard errors are smaller for most variables and the effects of some variables on the estimated multipliers therefore turn statistically significant, such as the effect of the debt ratio on contemporaneous multipliers and the effect of the import ratio (and some other variables) on long-term multipliers. These results are a natural consequence of the fact that the uncertainty on the dependent variables is not taken into account, which may give a biased view on the importance of the restrictions implied by the explanatory variables. Therefore, it seems more cautious to assign higher importance to the results which were generated with the standard error adjustment.

References

Benati, L., and H. Mumtaz (2007): "U.S. Evolving Macroeconomic Dynamics: A Structural Investigation," ECB Working Paper No. 746.

Canova, F. (2007): *Methods for Applied Macroeconomic Research*. Princeton University Press, Princeton.

Canova, F., and M. Ciccarelli (2009): "Estimating Multi-Country VAR Models," *International Economic Review*, 50, 929–959.

Carter, C. K., and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553.

Cogley, T., and T. J. Sargent (2001): "Evolving Post-World War II U.S. Inflation Dynamics," in *NBER Macroeconomics Annual*, vol. 16. MIT Press, Cambridge, MA.

——— (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.," *Review of Economic Studies*, 8, 262–302.

Geweke, J. (1989): "Bayesian Inference in Econometric Models using Monte Carlo Integration," *Econometrica*, 57, 317–339.

——— (1992): "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics*, ed. by J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith. Oxford University Press, Oxford.

——— (1993): "Bayesian Treatment of the Independent Student-t Linear Model," *Journal of Applied Econometrics*, 8, S19–S40.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371–389.

Kim, C., and C. R. Nelson (1999): *State-Space Models with Regime Switching*. MIT Press, Cambridge, MA.

Koop, G. (2003): *Bayesian Econometrics*. Wiley & Sons.

Mountford, A., and H. Uhlig (2009): "What are the Effects of Fiscal Policy Shocks?," *Journal of Applied Econometrics*, 24, 960–992.

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

Raftery, A. E., and S. Lewis (1992): "How Many Iterations in the Gibbs Sampler?," in *Bayesian Statistics*, ed. by J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith. Oxford University Press, Oxford.

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.

Table C.1: Raftery and Lewis (1992) diagnostics.[a,b]

	Estim. Parameters	Thinning Factor	Burn-in Replic.	Total Runs
Coefficients	4068	1	2	150
Covariances	452	1	10	429
Volatilities	678	1	4	208

[a] Parameters for Raftery and Lewis (1992) diagnostics are quantile = 0.025; desired accuracy = 0.025; required probability of attaining the required accuracy = 0.95.

[b] Results are based on 5,000 iterations of the Gibbs sampler with zero burn-in replications and thinning factor 1.

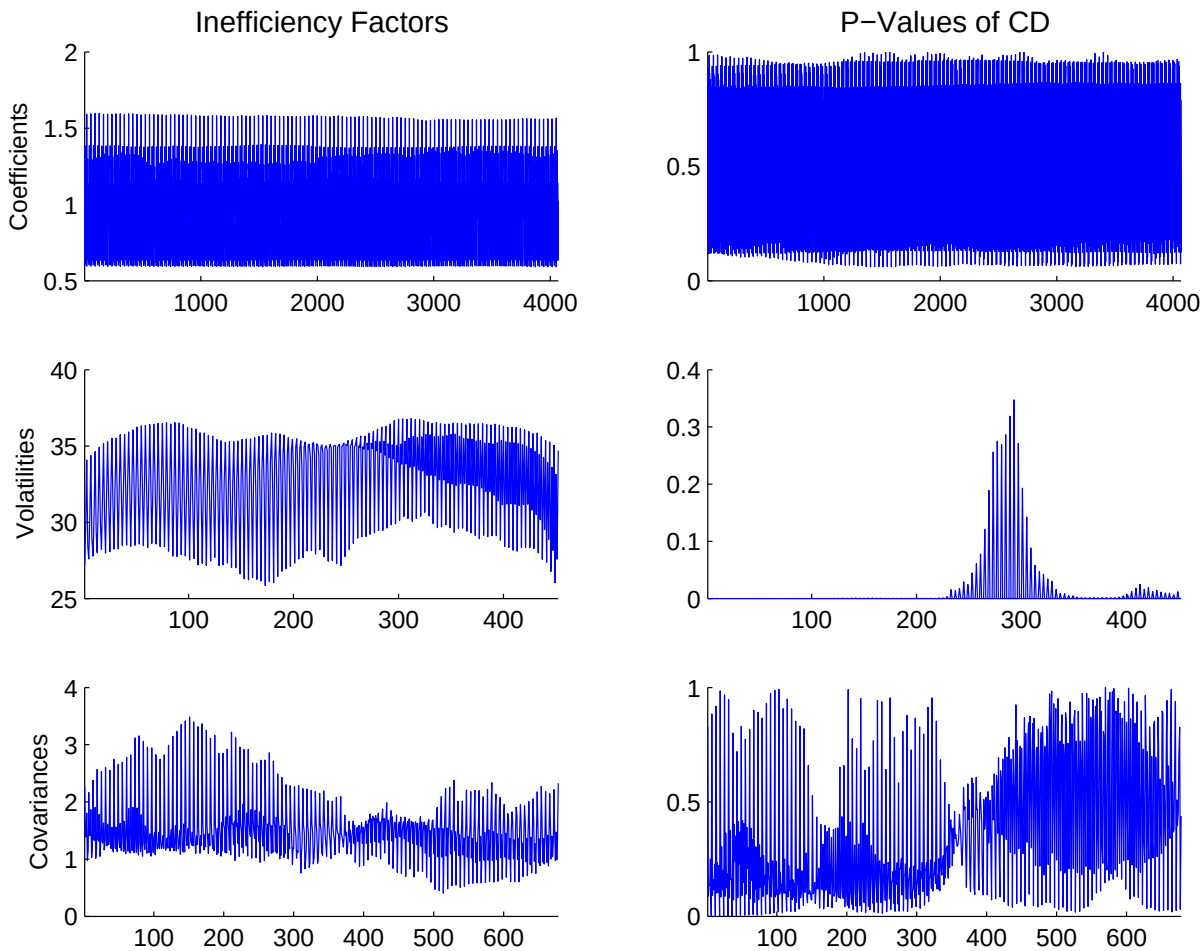

Figure C.1: Convergence diagnostics for state vectors. *Notes.* Horizontal axes refer to vectors of time-varying parameters with one point representing one parameter at a given time (e.g. volatilities $h_{i,t}$); left panels: inefficiency factors, i.e. inverse of Geweke's (1992) relative numerical efficiency measure; computed as $IF = 1+2\sum_{k=1}^{\infty}\rho_k$, where ρ_k is the k-th order autocorrelation of the Markov chain; right panels: P-values of Geweke's (1992) convergence diagnostic; computed as $CD = (\widehat{\theta}_1 - \widehat{\theta}_2)/(\widehat{\sigma}_1/\sqrt{N_1} + \widehat{\sigma}_2/\sqrt{N_2}) \to \mathrm{N}(0,1)$, where $N_1 = 100$, $N_2 = 400$, middle 500 draws dropped.

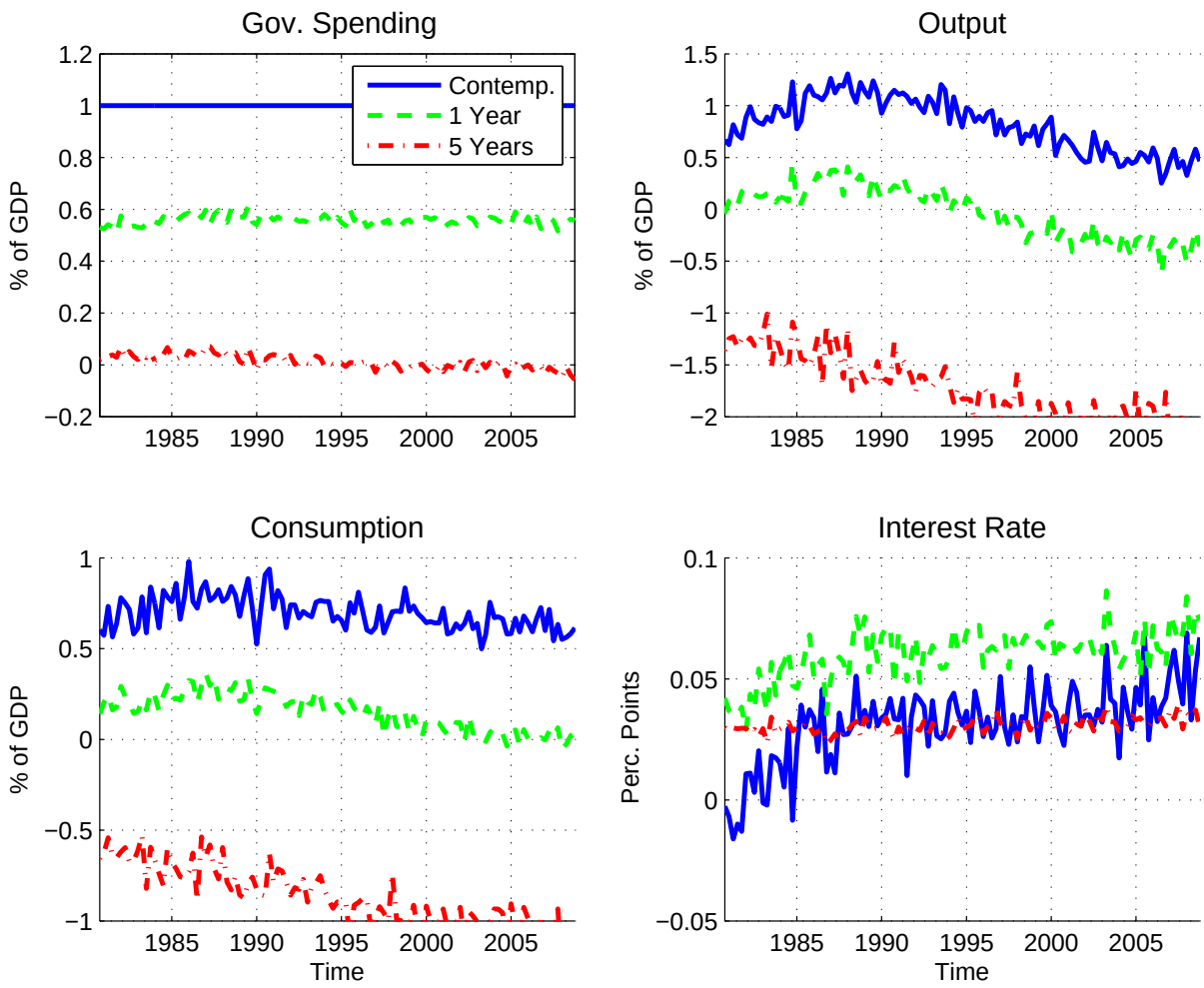

Figure D.1: Impulse responses to a spending shock at selected horizons, TVP-VAR with identification by sign restrictions. *Notes.* Median impulse responses are reported; the responses of output and consumption are measured in % of GDP to 1% of GDP spending shocks, i.e. multiplier at time t and horizon k = responding variable's response at time t and horizon k/(spending response at time t and horizon 0 $\times$ ratio of spending to responding variable at time t); the response of the interest rate is measured in percentage points to 1% spending shocks.

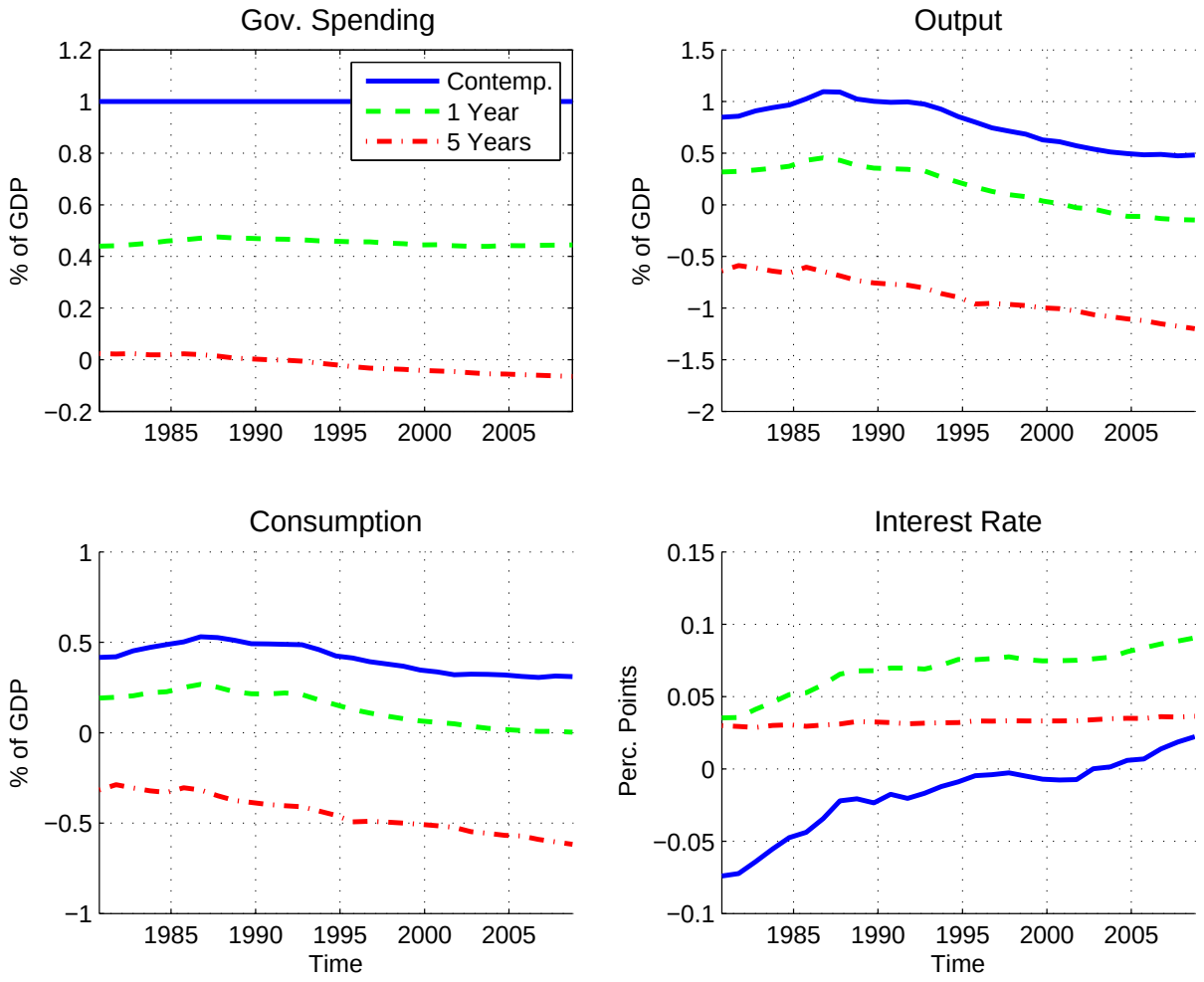

Figure D.2: Impulse responses to a spending shock at selected horizons, TVP-VAR with stationarity conditions imposed. *Notes.* See Figure D.1.

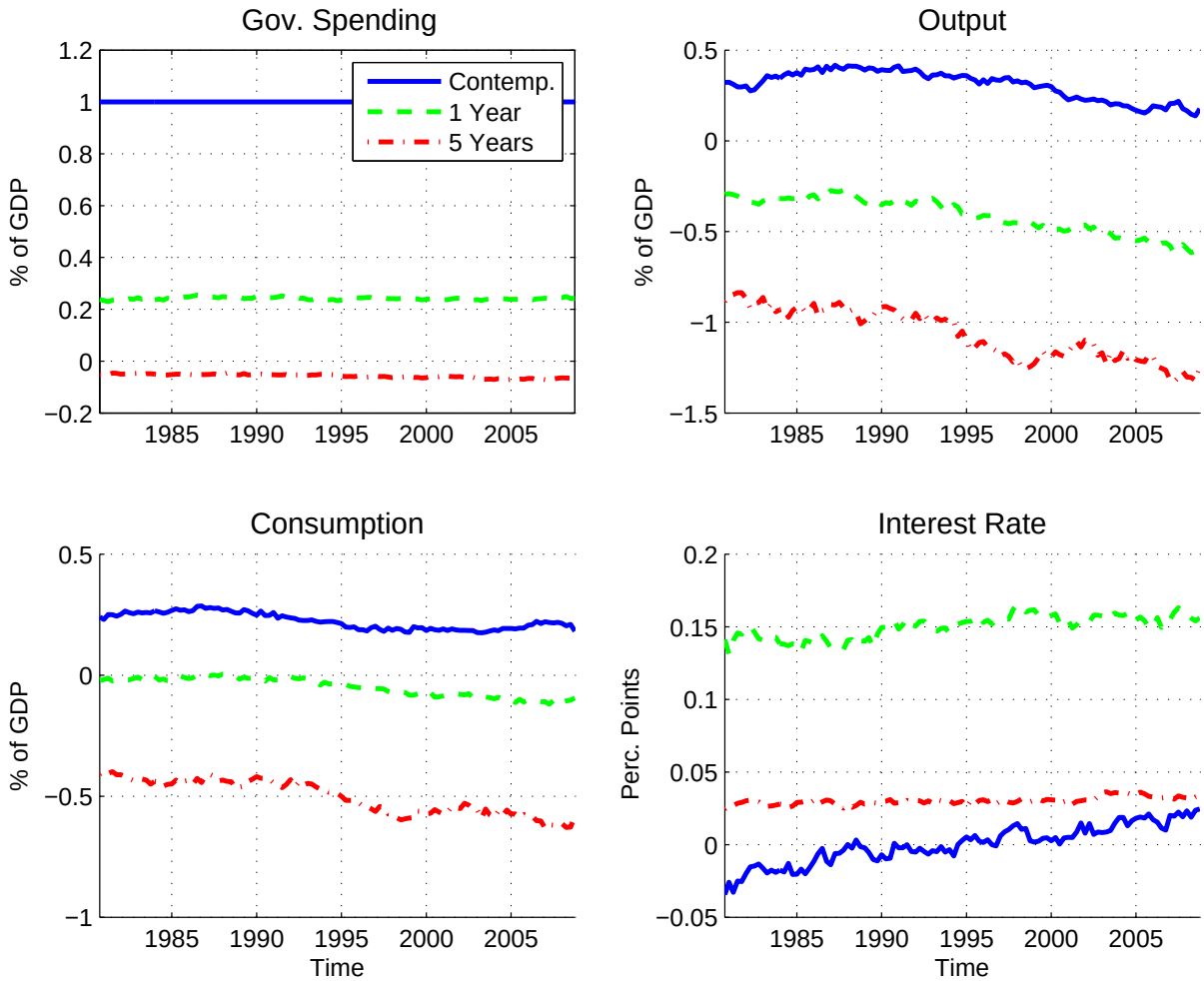

Figure D.3: Impulse responses to a spending shock at selected horizons, TVP-VAR with shorter training sample. *Notes.* See Figure D.1.

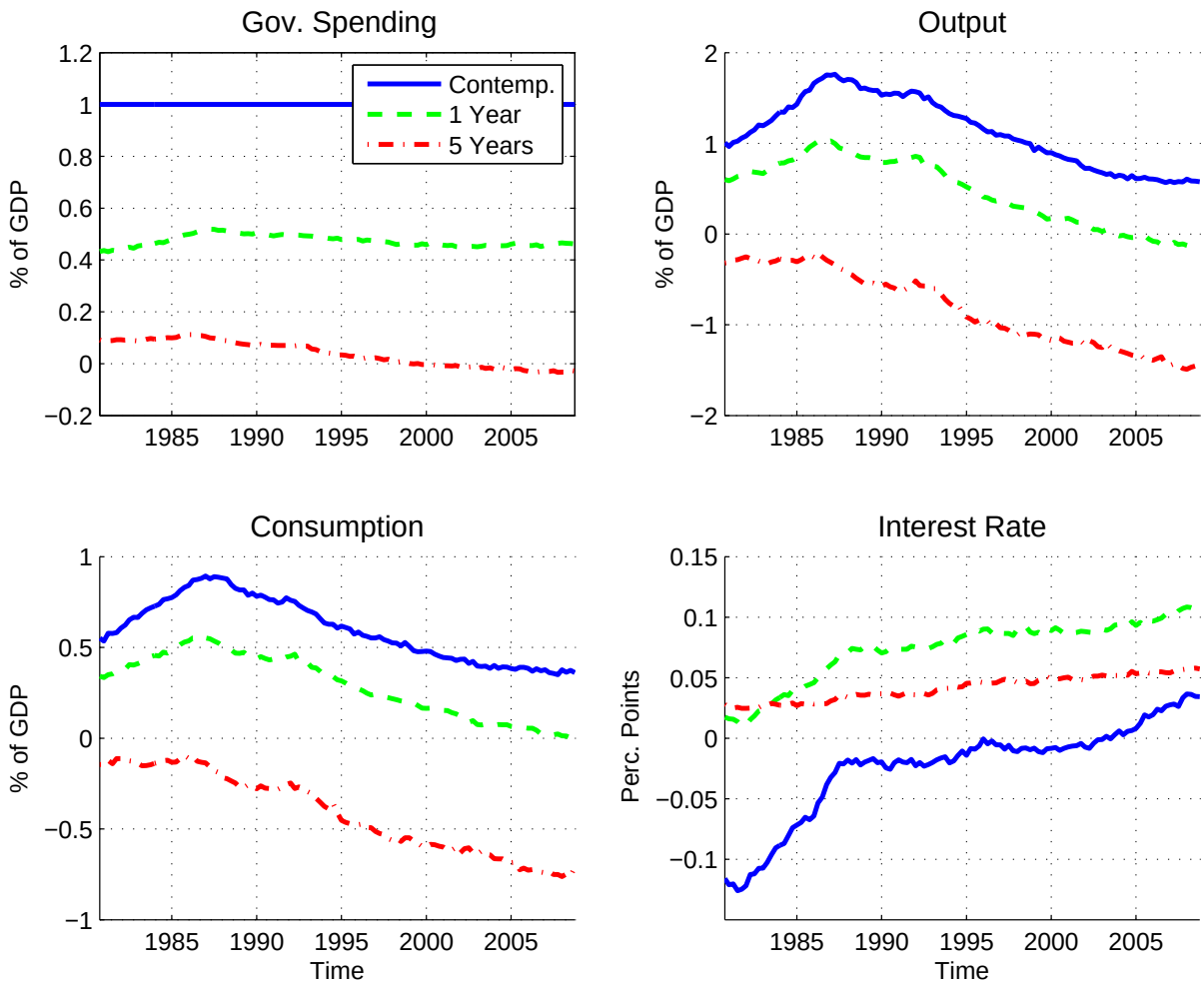

Figure D.4: Impulse responses to a spending shock at selected horizons, TVP-VAR with smaller k_Q. *Notes.* See Figure D.1.

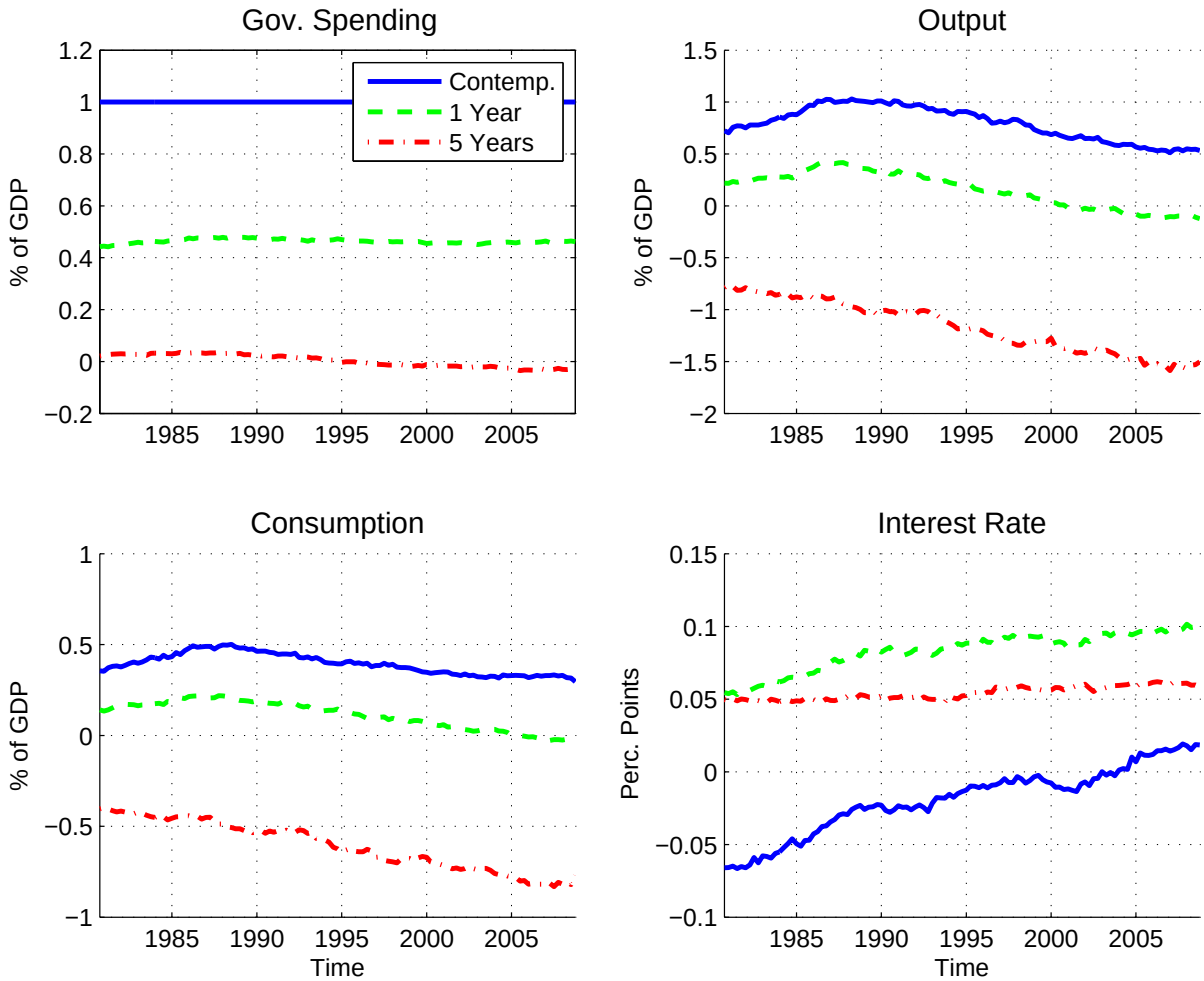

Figure D.5: Impulse responses to a spending shock at selected horizons, TVP-VAR with smaller k_S. *Notes.* See Figure D.1.

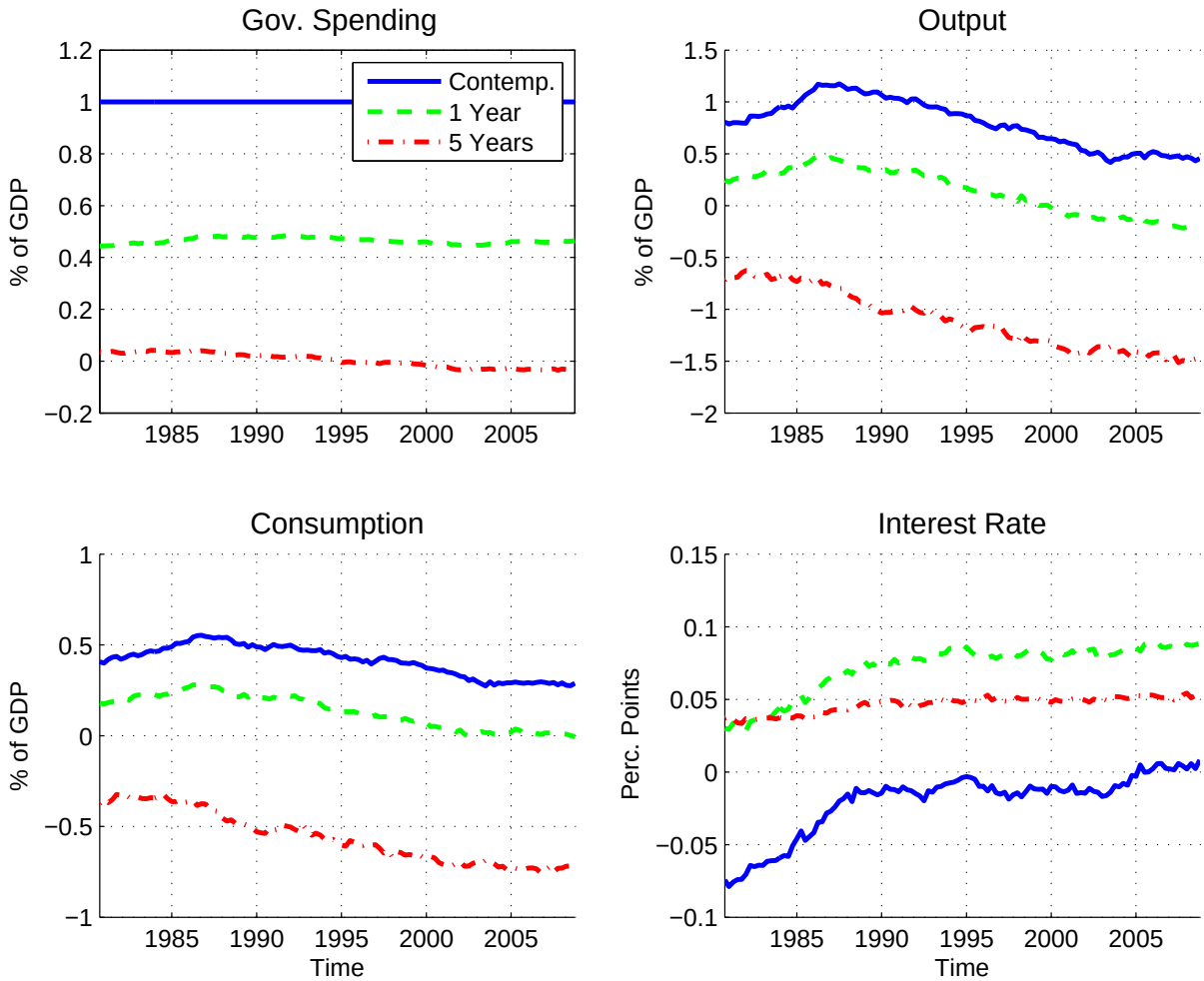

Figure D.6: Impulse responses to a spending shock at selected horizons, TVP-VAR with smaller k_W. *Notes.* See Figure D.1.

Table C.2: Bayesian linear regressions, dependent variables are contemporaneous multipliers.[a,b,c,d]

	Multiplier on Output						Multiplier on Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Gov. Debt over GDP (-1)	0.01***	-0.01***	-0.01***	-0.01***	-0.01**	-0.01*	0.00***	-0.00***	-0.00***	-0.00***	-0.00*	-0.00*
	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Credit over GDP (-1)		-0.06***	-0.06***	-0.04***	-0.04***	-0.06***		-0.02***	-0.02***	-0.01***	-0.01***	-0.02***
		(0.00)	(0.00)	(0.00)	(0.01)	(0.01)		(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Output Gap (-1)			0.02**	0.04***	0.03***	0.02*			0.01***	0.02***	0.01***	0.01**
			(0.01)	(0.01)	(0.01)	(0.01)			(0.00)	(0.01)	(0.01)	(0.01)
Imports over GDP (-1)				-0.01**	-0.01*	-0.01*				-0.01**	-0.01**	-0.00
				(0.01)	(0.01)	(0.01)				(0.00)	(0.01)	(0.01)
Investment Share					0.03*	0.07***					0.00	0.01**
					(0.02)	(0.02)					(0.01)	(0.01)
Wage Share						-0.05***						-0.01***
						(0.01)						(0.01)
Constant	0.82***	3.04***	2.94***	2.90***	2.10***	4.91***	0.46***	1.08***	1.04***	1.02***	0.93***	1.83***
	(0.12)	(0.17)	(0.18)	(0.17)	(0.52)	(0.88)	(0.03)	(0.06)	(0.06)	(0.06)	(0.20)	(0.37)
Trend	-0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	-0.00***	0.00***	0.00***	0.00**	0.00*	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	112	112	112	112	112	112	112	112	112	112	112	112

[a] Bayesian regressions allow for heteroskedastic errors following Geweke (1993). Dependent variables are posterior means of the posterior distribution of impulse responses from the identified TVP-VAR. All regressions are estimated using a Gibbs sampling algorithm with 1,100 draws and 100 omitted draws. This leaves us with 1,000 posterior draws of regression coefficients.

[b] Multiplier at time t and horizon k = responding variable's response at time t and horizon k/(spending response at time t and horizon 0 $\times$ ratio of spending to responding variable at time t).

[c] Point estimates are posterior means of the posterior distribution. Standard deviations are reported in parentheses. Asterisks indicate posterior probabilities that the regression coefficients are non-positive if the point estimates are positive or non-negative if the point estimates are negative (*less than 10%, **less than 5%, ***less than 1%.)

[d] Explanatory variables are measured in percent.

Table C.3: Bayesian linear regressions, dependent variables are multipliers after five years.[a,b,c,d]

	Multiplier on Output						Multiplier on Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Gov. Debt over GDP (-1)	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Credit over GDP (-1)		-0.01***	-0.01***	-0.01***	-0.01***	-0.01***		-0.00**	-0.00**	-0.01***	-0.01***	-0.00**
		(0.00)	(0.00)	(0.01)	(0.01)	(0.01)		(0.00)	(0.01)	(0.01)	(0.00)	(0.00)
Output Gap (-1)			-0.02***	-0.02***	-0.02***	-0.02***			-0.00	-0.01*	-0.01**	-0.01***
			(0.01)	(0.01)	(0.01)	(0.01)			(0.01)	(0.01)	(0.00)	(0.00)
Imports over GDP (-1)				0.01*	0.01	0.01*				0.00**	0.00**	0.00***
				(0.00)	(0.00)	(0.00)				(0.00)	(0.00)	(0.00)
Investment Share					0.00	0.00					0.01**	0.00
					(0.01)	(0.01)					(0.01)	(0.01)
Wage Share						0.00						0.01**
						(0.00)						(0.00)
Constant	-0.40***	-0.07	-0.02	-0.02	-0.07	-0.14	-0.25***	-0.13***	-0.13***	-0.12***	-0.30***	-0.60***
	(0.03)	(0.12)	(0.12)	(0.12)	(0.29)	(0.44)	(0.01)	(0.05)	(0.05)	(0.05)	(0.13)	(0.19)
Trend	-0.01***	-0.00***	-0.00**	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	112	112	112	112	112	112	112	112	112	112	112	112

[a] Bayesian regressions allow for heteroskedastic errors following Geweke (1993). Dependent variables are posterior means of the posterior distribution of impulse responses from the identified TVP-VAR. All regressions are estimated using a Gibbs sampling algorithm with 1,100 draws and 100 omitted draws. This leaves us with 1,000 posterior draws of regression coefficients.

[b] Multiplier at time t and horizon k = responding variable's response at time t and horizon k/(spending response at time t and horizon 0 $\times$ ratio of spending to responding variable at time t).

[c] Point estimates are posterior means of the posterior distribution. Standard deviations are reported in parentheses. Asterisks indicate posterior probabilities that the regression coefficients are non-positive if the point estimates are positive or non-negative if the point estimates are negative (*less than 10%, **less than 5%, ***less than 1%.)

[d] Explanatory variables are measured in percent.