Policy Risk and the Business Cycle: Online Appendix

Benjamin Born* Johannes Pfeifer†

June 12, 2014

*University of Mannheim and CESifo, L7, 3-5, 68131 Mannheim, Germany, Tel.: +49-(0)621-181 1806, born@uni-mannheim.de

†University of Mannheim, L7, 3-5, 68131 Mannheim, Germany, Tel.: +49-(0)621-181 3430, pfeifer@uni-mannheim.de.

A Data construction

Unless otherwise noted, all data are from the Bureau of Economic Analysis (BEA)'s NIPA Tables and available in quarterly frequency from 1970Q1 until 2012Q2.

A.1 Data for the exogenous processes

Capital and labor tax rates. Our approach to calculate average tax rates closely follows Mendoza et al. (1994), Jones (2002), and Leeper et al. (2010). We first compute the average personal income tax rate

$$\tau^p = \frac{IT}{W + PRI/2 + CI} \ ,$$

where IT is personal current tax revenues (Table 3.1 line 3), W is wage and salary accruals (Table 1.12 line 3), PRI is proprietor's income (Table 1.12 line 9), and $CI \equiv PRI/2 + RI + CP + NI$ is capital income. Here, RI is rental income (Table 1.12 line 12), CP is corporate profits (Table 1.12 line 13), and NI denotes the net interest income (Table 1.12 line 18).

The average labor and capital income tax rates can then be computed as

$$\tau^n = \frac{\tau^p(W + PRI/2) + CSI}{EC + PRI/2} \ ,$$

where CSI denotes contributions to government social insurance (Table 3.1 line 7), and EC is compensation of employees (Table 1.12 line 2), and

$$\tau^k = \frac{\tau^p CI + CT + PT}{CI + PT} \ ,$$

where CT is taxes on corporate income (Table 3.1 line 5), and PT is property taxes (Table 3.3 line 8).

Government spending. Government spending is the sum of government consumption (Table 3.1 line 16) and government investment (Table 3.1 line 35) divided by the GDP deflator (Table 1.1.4 line 1) and the civilian noninstitutional population (BLS, Series LNU00000000Q).

Debt. Federal debt held by the public (St.Louis FED - FRED Database, Series FYGFDPUN).

Monetary policy shock. Computed as the residual from a Clarida et al. (2000)-type Taylor rule (see Appendix B.8).

Total factor productivity (TFP). The TFP series is taken from Fernald (2012), who closely follows Basu et al. (2006) and provides a quarterly series that is adjusted for capital and labor utilization.

Relative price of investment. The relative price of investment is taken from Schmitt-

Grohé and Uribe (2011) and only available until 2006Q4. They base their calculations on Fisher (2006).

The different sample lengths are not an issue as we estimate each exogenous process separately. Using the longest available sample assures that we make optimal use of the available information for each series.

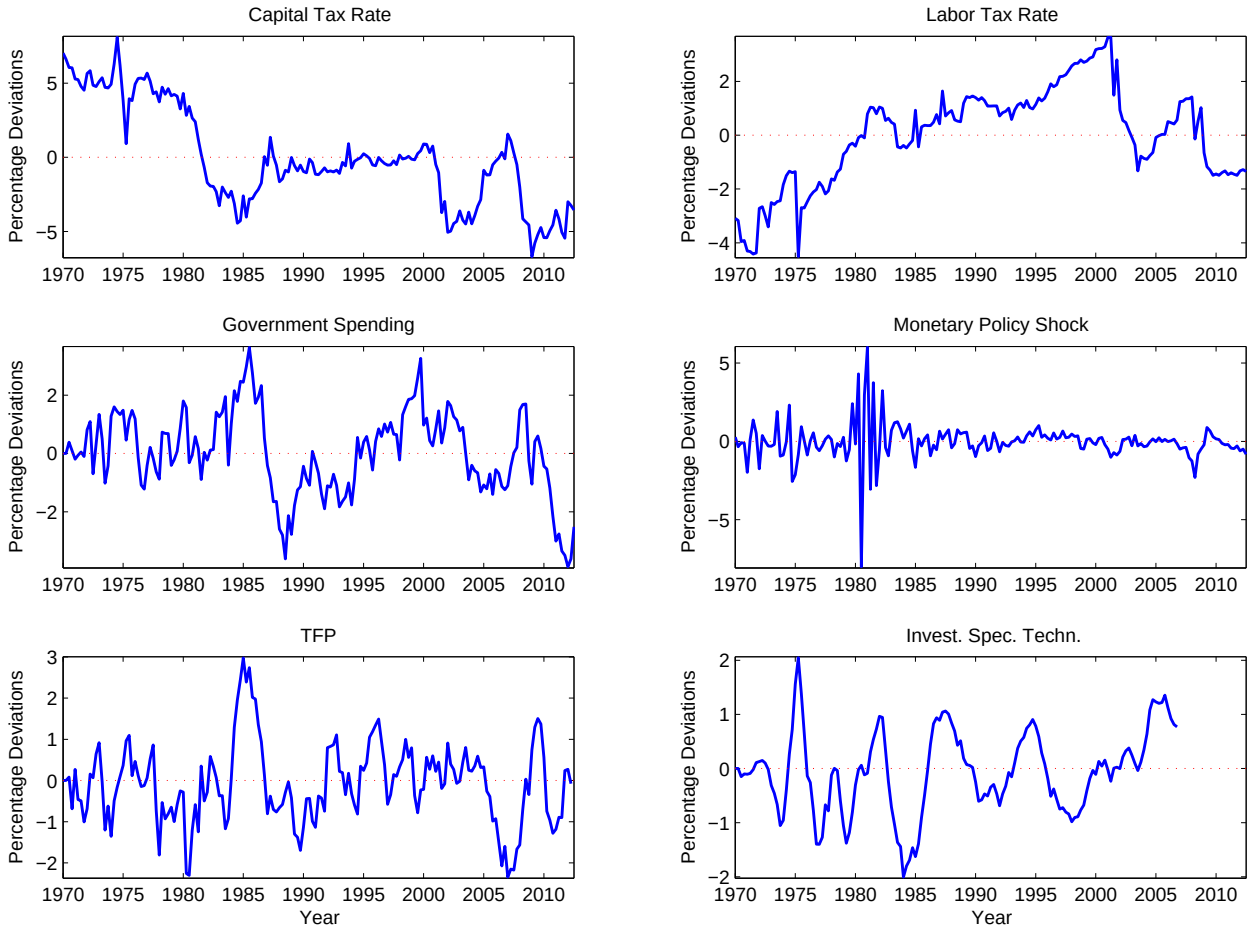


Figure A.1: Time series of exogenous driving processes.
Notes: Tax rates are demeaned; government spending and technology processes are detrended using the one-sided HP-filter.

A.2 Data for SMM

Output. Nominal GDP (Table 1.1.5 line 1) divided by the GDP deflator (Table 1.1.4 line 1) and the civilian noninstitutional population (BLS, Series LNU00000000Q).

Investment. Sum of Residential fixed investment (Table 1.1.5 line 12) and nonresidential fixed investment (Table 1.1.5 line 9) divided by the GDP deflator (Table 1.1.4 line 1) and the civilian noninstitutional population (BLS, Series LNU00000000Q).

Consumption. Sum of personal consumption expenditures for nondurable goods (Table 1.1.5 line 5) and services (Table 1.1.5 line 6) divided by the GDP deflator (Table 1.1.4 line 1) and the civilian noninstitutional population (BLS, Series LNU00000000Q).
Real wage. Hourly compensation in the nonfarm business sector (BLS, Series PRS85006103) divided by the GDP deflator (Table 1.1.4 line 1).
Inflation. Computed as the log-difference of the GDP deflator (Table 1.1.4 line 1).
Nominal interest rate. Geometric mean of the effective Federal Funds Rate (St.Louis FED - FRED Database, Series FEDFUNDS).
Hours. Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing (St.Louis FED - FRED Database, Series AWHMAN) normalized by 5 days times 24 hours (providing a mean of about 1/3).

A.3 Additional data for GMM

Interest term spread. We use the difference of the quarterly geometric mean of the 10-Year Treasury Constant Maturity Rate (FRED Database, Series GS10) and the quarterly geometric mean of the 3-Month Treasury Bill: Secondary Market Rate (FRED Database, Series TB3MS).
Money growth rate. Growth rate of the M2 Money Stock (FRED Database, Series M2SL).
Commodity inflation. Commodity inflation is computed as the growth rate of the X12-seasonally adjusted Producer Price Index: All Commodities (FRED Database, Series PPI-ACO).
Output gap. The output gap is constructed as the deviation of real GDP (FRED Database, Series GDPC96) from its HP-trend ($\lambda = 1600$).

B Econometric Methods

B.1 Modeling Time-Varying Volatility

There are two major competing approaches to model time-varying standard deviations: GARCH models and stochastic volatility (SV) models (Fernández-Villaverde and Rubio-Ramírez, 2010). In the standard GARCH model, σ_t^2 is a function of the squared scaled lagged innovation in the level equation ν_{t-1}^2 and its own lagged value: $\sigma_t^2 = \omega + \alpha(\sigma_{t-1}\nu_{t-1})^2 + \beta\sigma_{t-1}^2$. The GARCH model has one important drawback: there are no distinct volatility shocks. The only innovations to the volatility equation are past level shocks, meaning that they cannot be separated from volatility shocks. As we are especially interested in the effects of shocks to the volatility, we cannot use a GARCH model but instead employ a stochastic volatility model.

Note that the SV-framework used in the present study does not imply a mechanical link between the level shocks and the volatility shocks as a GARCH-model would do. Of course, as a comparison of Figures **??** and A.1 shows, a large level shock tends to coincide with an increase in the conditional variance. However, the reason for this increase in the estimated conditional variance is not a mechanical effect of this level shock subsequently entering the volatility equation. Rather, the Bayesian estimation of the SV-model weighs the likelihood of observing such a large shock being drawn from a narrow distribution, i.e., without observing a simultaneous/previous volatility shock, against the likelihood of observing a shock of this size that is drawn from a wider distribution due to the occurrence of a variance shock.

B.2 The Particle Filter

For ease of exposition, let x_t be a generic observable AR(1) process

$$x_t = \rho x_{t-1} + e^{\sigma_t}\nu_t \ , \quad \nu_t \overset{iid}{\sim} \mathcal{N}(0,1) \tag{B.1}$$

where the unobserved/latent state σ_t follows a stochastic volatility process

$$\sigma_t = (1-\rho^\sigma)\,\bar{\sigma} + \rho^\sigma \sigma_{t-1} + \eta\varepsilon_t \ , \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1), \tag{B.2}$$

where $\bar{\sigma}$ is is the unconditional mean of σ_t. The shock to the volatility ε_t is assumed to be independent from the level shock ν_t.

Hence, a filter is required to obtain the so-called filtering density $p\left(\sigma_t | \, x^t; \Theta\right)$. Due to the nonlinearity embedded in the stochastic volatility setup of the shocks, we cannot simply employ the Kalman filter as in the case of linearity and normally distributed shocks. Instead, we employ the *Sequential Importance Resampling* (SIR) particle filter, a special application of the more general class of *Sequential Monte Carlo* methods, to evaluate the likelihood (Fernández-Villaverde and Rubio-Ramírez, 2007; Fernández-Villaverde et al., 2011). Given the structure in (B.1) and (B.2) and some initial value x_0, the factorized likelihood of observing x^T can be written as

$$\begin{aligned}
p\left(x^T;\Theta\right) &= \prod_{t=1}^{T} p\left(x_t | x^{t-1};\Theta\right) \\
&= \int p\left(x_1|x_0,\sigma_0;\Theta\right) d\sigma_0 \prod_{t=2}^{T} \int p\left(x_t|x_{t-1},\sigma_t;\Theta\right) p\left(\sigma_t|x^{t-1};\Theta\right) d\sigma_t \\
&= \int \frac{1}{e^{\sigma_0}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_1-\rho x_0}{e^{\sigma_0}}\right)^2\right] d\sigma_0 \\
&\times \prod_{t=2}^{T} \int \frac{1}{e^{\sigma_t}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_t-\rho x_{t-1}}{e^{\sigma_t}}\right)^2\right] p\left(\sigma_t|x^{t-1};\Theta\right) d\sigma_t \ ,
\end{aligned} \tag{B.3}$$

where x^t is a $(t \times 1)$ vector that stacks the observations on x up to time t, Θ stacks the parameters, and the last equality follows from the assumption of normally distributed shocks. Although we do not have an analytical expression for $p\left(\sigma_t|x^{t-1};\Theta\right), t = 1,\ldots,T$, and can therefore not compute it directly, we can employ the particle filter to estimate the likelihood by iteratively drawing from $p\left(\sigma_t|x^{t-1};\Theta\right)$.

The underlying idea of the particle filter is to use an approximation of the filtering density $p\left(\sigma_t|\, x^t;\Theta\right)$ with a simulated distribution generated from empirical data. This distribution can be formed from mass points, or particles,

$$p\left(\sigma_t|\, x^t;\Theta\right) \simeq \sum_{i=0}^{N} \omega_t^i \delta_{\sigma_t^i}\left(\sigma_t\right), \quad \sum_{i=0}^{N} \omega_t^i = 1, \quad \omega_t^i \geqslant 0 \tag{B.4}$$

where δ is the Dirac delta function and ω_t^i is the weight attached to the respective draw/particle σ_t^i (Godsill et al., 2004). We can then use a *Sequential Importance Resampling* (SIR)-approach to update particles from time t to $t+1$ and obtain the new filtering distribution at $t+1$ (see, e.g., Fernández-Villaverde et al., 2011). A convenient by-product of this filtering approach is that we also approximate $p\left(\sigma_t|x^{t-1};\Theta\right)$, the distribution we need to build the likelihood.

The SIR is a two-step procedure that, by using a prediction and a resampling/filtering step for each time period, ultimately allows to iteratively draw from $p\left(\sigma_t|x^{t-1};\Theta\right)$. Starting with $p\left(\sigma_0|\, x^0;\Theta\right) = p\left(\sigma_0;\Theta\right)$, the prediction step uses the law of motion for the states $f(\sigma_{t+1}|\sigma_t)$, equation (B.2), to obtain the conditional density $p\left(\sigma_1|\, x^0;\Theta\right) = p\left(\varepsilon_1\right) p\left(\sigma_0|\, x^0;\Theta\right)$. That is, given N draws $\left\{\sigma_{t|t}^i\right\}_{i=1}^N$ from $p\left(\sigma_t|\, x^t;\Theta\right)$, (here $p\left(\sigma_0|\, x^0;\Theta\right)$) and a draw of exogenous shocks $\varepsilon_t^i \sim \mathcal{N}(0,1)$, we can use equation (B.2) to compute $\left\{\sigma_{t+1|t}^i\right\}_{i=1}^N$.[1]

Next, the resampling/filtering step uses importance resampling to update the conditional probability from $p\left(\sigma_t|\, x^{t-1};\Theta\right)$ to $p\left(\sigma_t|\, x^t;\Theta\right)$. The crucial idea is that if $\left\{\sigma_{t|t-1}^i\right\}_{i=1}^N$ is a draw from $p\left(\sigma_t|x^{t-1};\Theta\right)$ and $\{\widetilde{\sigma}_t^i\}_{i=1}^N$ is a draw with replacement from $\left\{\sigma_{t|t-1}^i\right\}_{i=1}^N$ using the resampling probabilities

$$\omega_t^i = \frac{p\left(x_t|x^{t-1},\sigma_{t|t-1}^i;\Theta\right)}{\sum_{i=1}^N p\left(x_t|x^{t-1},\sigma_{t|t-1}^i;\Theta\right)}\,, \tag{B.5}$$

then $\left\{\sigma_{t|t}\right\}_{i=1}^N = \{\widetilde{\sigma}_t^i\}_{i=1}^N$ is a draw from $p\left(\sigma_t|x^t;\Theta\right)$. The resampling with probabilities given in (B.5) serves two purposes. First, the reweighting implements an importance sampling approach, i.e., draws are obtained from a proposal density that is easy to draw from and are then subsequently reweighted to reflect the density to be approximated (see Arulampalam et al., 2002, for a derivation).[2] Second, without the resampling step, there would be an

[1] The notation $t+1|t$ indicates a draw at time $t+1$ conditioned on the information available at time t.

[2] In our case, we use the prior density $p\left(\sigma_t|\,\sigma^{t-1};\Theta\right)$ as the importance density.

increase in the unconditional variance of ω_t over time, yielding only one particle with non-zero weight (known as degeneracy or sample impoverishment, see Arulampalam et al. (2002)). By resampling, we keep only those particles with high ω_t^i (i.e., those that are closer to the true state vector). Having now obtained draws from $p\left(\sigma_t \middle| x^t; \Theta\right)$, we can again start with the prediction step to obtain draws for time period $t+1$.

After T iterations, we get an estimate of our likelihood as[3]

$$\begin{aligned} p\left(x^T;\Theta\right) \simeq \frac{1}{N}\sum_{i=1}^{N}\frac{1}{e^{\sigma^i_{0|0}}\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x_1-\rho x_0}{e^{\sigma^i_{0|0}}}\right)^2\right] \\ \times\prod_{t=2}^{T}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{e^{\sigma^i_{t|t-1}}\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x_t-\rho x_{t-1}}{e^{\sigma^i_{t|t-1}}}\right)^2\right] \ . \end{aligned} \tag{B.6}$$

For the autoregressive parameters of the AR(2) level equations, ρ_1^i and ρ_2^i, we impose a uniform prior for each of the corresponding autoregressive roots over the stability region $(-1,+1)$. Let ξ_1 and ξ_2 be the roots of such an $AR(2)$-process. The autoregressive parameters corresponding to these roots can be recovered from: $\rho_1 = \xi_1 + \xi_2$ and $\rho_2 = -\xi_1\xi_2$. The posterior distribution was computed from a 20,500 draw Monte Carlo Markov Chain using 10,000 particles, where the first 5,500 draws were discarded as burn-in draws. Acceptance rates were generally between 20% and 45%. We also checked identifiability of the SV-process by simulating data from the process and trying to recover the true parameters from this artificial data.

B.3 Particle Smoother

We employ the *backward-smoothing routine* suggested by Godsill et al. (2004) to draw from the smoothing density $p(\sigma^T|x^T;\Theta)$ to get a historical distribution of the volatilities. Specifically, we start with the factorization

$$p(\sigma^T|x^T;\Theta) = p(\sigma_T|x^T;\Theta)\prod_{t=1}^{T-1}p(\sigma_t|\sigma_{t+1:T},x^T;\Theta) \ . \tag{B.7}$$

[3]See Fernández-Villaverde and Rubio-Ramírez (2007) and Doucet and Johansen (2009) and the references contained therein for the conditions required for a central limit theorem to apply, yielding a consistent estimator of $p\left(x^T;\Theta\right)$.

The second factor can be further simplified

$$\begin{aligned} p(\sigma_t|\sigma_{t+1:T}, x^T; \Theta) &= p(\sigma_t|\sigma_{t+1}, x^t; \Theta) \\ &= \frac{p(\sigma_t|x^t; \Theta) f(\sigma_{t+1}|\sigma_t)}{p(\sigma_{t+1}|x^t)} \\ &\propto p(\sigma_t|x^t; \Theta) f(\sigma_{t+1}|\sigma_t) \ , \end{aligned} \tag{B.8}$$

where the first equality results from the Markovian properties of the model and f denotes the state transition density following from equation (B.2). Equation (B.4) describes how to approximate $p(\sigma_t|x^t; \Theta)$ by forward filtering. Therefore, we can approximate $p(\sigma_t|\sigma_{t+1:T}, x^T; \Theta) \propto p(\sigma_t|x^t; \Theta) f(\sigma_{t+1}|\sigma_t)$ by

$$p(\sigma_t|\sigma_{t+1}, x^t; \Theta) \simeq \sum_{i=1}^{N} \omega_{t|t+1}^i \delta_{\sigma_t^i}(\sigma_t) \ , \tag{B.9}$$

where the new weights $\omega_{t|t+1}^i$ are given by

$$\omega_{t|t+1}^i = \frac{\omega_t^i f(\sigma_{t+1}|\sigma_t^i)}{\sum_{j=1}^{N} \omega_t^j f(\sigma_{t+1}|\sigma_t^j)} \ . \tag{B.10}$$

and the ω_t^i are the weights obtained in the filtering step. Denote with $\tilde{\sigma}_t^i$ the i^{th} draw from the smoothing density at time t. At time T, we can obtain draws $\tilde{\sigma}_T^i$ by drawing from $p(\sigma_T|x^T)$ with the weights ω_T^i. Then, going backwards in time, we can use the above recursions to iteratively obtain draws $\tilde{\sigma}_t^i$ by resampling using the weights given in (B.10). In order to get the smoothing distribution, we can use the above recursion repeatedly to simulate different independent smoothing trajectories. Moreover, given the sequence of smoothed states, we can also extract the smoothed residuals for both the level and the volatility equation. The smoothed values were computed at the mean of the posterior distribution using 10,000 trajectories with 10,000 particles each.

B.4 Tailored Randomized Block Metropolis Hastings Algorithm

Let Θ, $p\left(x^T|\Theta\right)$, and $\pi(\Theta)$ denote the vector of parameters to be estimated, the likelihood function, and the prior distribution of the parameters, respectively. The posterior distribution $\pi(\Theta|x^T)$ can be computed as

$$\pi\left(\Theta|x^T\right) \propto p\left(x^T|\Theta\right) \pi\left(\Theta\right) \ . \tag{B.11}$$

Given this usually analytically intractable posterior, most macroeconomic applications employ a Random Walk Metropolis-Hastings (RW-MH) algorithm to generate draws from the posterior distribution. However, the standard RW-MH algorithm often has poor mixing properties,

leading to highly autocorrelated draws, and is therefore often very inefficient. Hence, to increase the efficiency, we use the Tailored Randomized Block Metropolis Hastings (TaRB-MH) algorithm proposed by Chib and Ramamurthy (2010).[4] Instead of in each iteration step simultaneously drawing an entire new parameter vector from a proposal density, the parameter vector is randomly split up into several blocks. Each block is then subsequently updated by a separate MH run, conditional on the previous step's values of the parameters in the other blocks. Ideally, the blocks should be formed according to the correlation between parameters, with highly correlated parameters belonging to the same block. However, we have no a priori knowledge about the correlation between parameters and resort to a blocking scheme where both the number of blocks and its composition are randomized in each step. This algorithm provides a good compromise between the standard RW-MH and tailored multiple block MH algorithms that use multiple blocks, which are particularly designed for the problem at hand. The second feature that improves on the standard RW-MH is that in each step the proposal density is "tailored" to the location and the curvature of the posterior density in that block by using a non-derivative based global optimizer. We deviate from Chib and Ramamurthy (2010) by using the CMAES algorithm (Hansen et al., 2003) instead of a simulated annealing as the former has been shown to be more efficient (Andreasen, 2010).[5] Moreover, it requires considerably less tuning than a simulated annealing. The TaRB-MH algorithm proceeds as follows.

1. At each iteration step n, $n = 1, \ldots, N$, the elements of the parameter vector θ are separated into random blocks $(\theta_{n,1}, \theta_{n,2}, \ldots, \theta_{n,p_n})$ by perturbing their initial ordering and assigning the first parameter in the perturbed vector to the first block and each following parameter with probability $p = 0.5$ to a new block, leaving us with 2.5 blocks on average as we estimate 5 parameters.

2. At each iteration step n, each block $\theta_{n,l}, l = 1, \ldots, p_n$ is sampled by a Metropolis-Hastings step using a proposal density adapted to the posterior in the following way. Denote with $\theta_{n,-l}$ the most current value of all blocks except for the lth one, i.e., their value at the end of step $n-1$. To generate a new draw for $\theta_{n,l}$, the CMAES-algorithm is used to find
$$\hat{\theta}_{n,l} = \arg\max_{\theta_{n,l}} \log \left[p\left(x^T | \theta_{n,l}, \theta_{n,-l}\right) \pi\left(\Theta\right) \right]. \tag{B.12}$$
That is, we use a global optimizer to maximize the posterior over the current block l, given the value of all other parameters at the end of step $n-1$. Having found the "conditional mode" $\hat{\theta}_{n,l}$, we compute the curvature of the target posterior distribution in

[4] Using the TaRB-MH decreased the inefficiency factors from values around 10 to below 2.

[5] For an intuitive introduction to the working of the CMAES algorithm, see Binsbergen et al. (2012).

the standard way as the negative inverse of the Hessian[6] at the "conditional mode"

$$V_{n,l} = \left(-\frac{\partial \log \left[p\left(x^T | \theta_{n,l}, \theta_{n,-l}\right) \pi\left(\Theta\right) \right]}{\partial \theta_{n,l} \theta'_{n,l}} \right)^{-1} \Bigg|_{\theta_{n,l}=\hat{\theta}_{n,l}} . \tag{B.13}$$

Following Chib and Ramamurthy (2010), we use a multivariate t-distribution with ν degrees of freedom as proposal density for $\theta_{n,l}$, $q_l\left(\theta_{n,l} | \, \theta_{n,-l}, x^T\right)$. Mean and variance are set to the "conditional mode" and the negative inverse of the Hessian at this point:

$$q_l\left(\theta_{n,l} | \, \theta_{n,-l}, x^T\right) = t\left(\theta_{n,l} | \, \hat{\theta}_{n,l}, V_{n,l}, \nu\right). \tag{B.14}$$

In the Metropolis-Hastings-step, a proposed value $\theta^*_{n,l}$ is accepted as the new value of the block with probability

$$\alpha_l\left(\theta_{n,l}, \theta^*_{n,l} \middle| \, \theta_{n,-l}, x^T\right) = \min\left[\frac{p\left(x^T | \theta^*_{n,l}, \theta_{n,-l}\right) \pi\left(\theta^*_{n,l}\right)}{p\left(x^T | \theta_{n,l}, \theta_{n,-l}\right) \pi\left(\theta_{n,l}\right)} \frac{t\left(\theta_{n,l} | \, \hat{\theta}_{n,l}, V_{n,l}, \nu\right)}{t\left(\theta^*_{n,l} \middle| \, \hat{\theta}_{n,l}, V_{n,l}, \nu\right)}, 1 \right]. \tag{B.15}$$

If the proposed value $\theta^*_{n,l}$ is rejected, we set $\theta_{n+1,l} = \theta_{n,l}$. This step is repeated for all p_n blocks before the algorithm starts over with step 1.

Setting $\nu = 5$ and iterating over steps 1 and 2, we can - after a suitable burn-in-period - obtain samples from the desired posterior distribution, which is the invariant distribution of the resulting Markov Chain. In our case, a burn-in of 2500 proved sufficient.

B.5 Model Solution

Let s_t denote the $n_s \times 1$ vector of state variables in deviations from steady state, including the exogenous shocks and the perturbation parameter Λ, and let s^i_t denote its ith entry. The policy function/law of motion for an arbitrary model variable $\widehat{X_t}$ then has the form

$$\widehat{X_t} = \sum_{i=1}^{n_s} \xi^X_i s^i_t + \frac{1}{2} \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \xi^X_{i,j} s^i_t s^j_t + \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \sum_{l=1}^{n_s} \xi^X_{i,j,l} s^i_t s^j_t s^l_t \, , \tag{B.16}$$

[6] While we use a derivative-free method to find the mode, we use standard numerical differentiation to compute the Hessian (see Abramowitz and Stegun, 1965, formulas 25.3.24 and 25.3.27) after the actual mode finding. Hessians resulting from such an approach in larger models are often not positive definite. However, the dimensionality of our estimation problem is small and the mode quickly discovered, leading to generally decent numerical behavior. This allows us to use a rather pragmatic approach. In the rare cases where we encounter an `NaN` in the computation of the inverse Hessian, we proceed with a 10^{-4} identity matrix. If the inverse Hessian is simply not positive definite, we perform a Jordan decomposition of the inverse Hessian, set the eigenvalues smaller than 1e-8 to 1e-8, and recompose the matrix. But, as mentioned, this rarely ever happens. As the likelihood function from the particle filter is only asymptotically differentiable, this involves some numerical error (see e.g. Pitt, 2002). However, as any positive definite matrix allows drawing from the posterior, this only entails some loss of efficiency (see e.g. Chib and Greenberg, 1995).

where the ξ's are scalars that depend on the deep parameters of the model and hats denote percentage deviations from steady state. Equation (B.16) shows why lower-order approximations would not be sufficient for our purpose.

As is well known, a first-order approximation exhibits certainty equivalence. This implies $\xi_v^X = 0$, where v denotes the position of a volatility shock in the state vector s. That is, up to first order, uncertainty shocks do not enter the policy function at all.

For a second-order approximation, it is well known from Schmitt-Grohé and Uribe (2004) for the homoskedastic case that uncertainty only enters the policy function through a constant term via the second derivative with respect to the perturbation parameter, i.e., through $\xi_{\Lambda,\Lambda} \neq 0$. However, things are more complicated in the heteroskedastic case where shocks to the variance occur, leading to an additional effect. Fernández-Villaverde et al. (2010) prove that in this case, the volatility shocks additionally only enter the policy function with non-zero coefficients in their interaction term with the respective level shock. Algebraically, only the cross-product of $\hat{\sigma}^i \times \hat{\nu}^i$ is different from 0. In contrast, all other cross-terms with the uncertainty shocks are zero, i.e $\xi_{v,j\neq u}^X = 0$, where v and u denote the positions of a volatility and its corresponding level shock in the state vector s, respectively. Hence, the effect of uncertainty is always mediated through level shocks. It is not possible to shock the variance of the level shocks independently from the level shock as its effect would be 0 by construction.

Only in the third-order approximation do the volatility shocks enter the policy function separately from the level shocks in a non-constant form. Most importantly, the term $\xi_{i,\Lambda,\Lambda}$ is in general different from 0 for all volatility shocks.

B.6 Simulated Method of Moments

The idea of the Simulated Method of Moments (SMM) is the following. Let x_t be a time t vector of observables from a stationary and ergodic distribution and let $\{x_t\}_{t=1}^T$ be the corresponding sequence. Furthermore, let $m(x_t)$ denote a $k \times 1$ vector of empirical moments computed from this data. Denote with $\{x_t^{sim}(\theta)\}_{t=1}^{aT}$ the corresponding time series of length aT generated from simulating the model using the $p \times 1$ parameter vector $\theta \in \Theta$, with $\Theta \subset R^p$. Let $m(x_t^{sim}(\theta))$ be the vector of simulated moments computed from the artificial data. The SMM estimator is the value of θ that satisfies

$$\widehat{\theta} = \underset{\theta\in\Theta}{\arg\min} \left[m(x_t) - m\left(x_t^{sim}(\theta)\right)\right]' W \left[m(x_t) - m\left(x_t^{sim}(\theta)\right)\right] , \tag{B.17}$$

where W is a $p \times p$ positive definite weighting matrix. Under the assumption that the model with $\theta = \theta_0$ is a correct representation of the true process that generated $m(x_t)$ and the regularity conditions spelled out in Duffie and Singleton (1993), $\hat{\theta}$ is a consistent estimator of

θ_0 with asymptotic distribution

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right) \xrightarrow{d} \mathcal{N}\left(0,\left(1+1/a\right)\left(J'WJ\right)^{-1}J'WSWJ\left(J'WJ\right)^{-1}\right), \tag{B.18}$$

where

$$S=\lim_{T\to\infty} Var\left(\left(1/\sqrt{T}\right)\sum_{t=1}^{T} m(x_t)\right), \tag{B.19}$$

and $J=E(\partial m(x_t^{sim})/\partial\theta)$ (see Ruge-Murcia, 2012).

This estimator is asymptotically efficient when using the weighting matrix

$$W=\left(V^{longrun}\right)^{-1}=\left[\lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m\left(x_t\right)\right)\right]^{-1}. \tag{B.20}$$

The ideal weighting matrix places the most weight on the linear combination of moments that are the most precisely measured in the data. However, for two reasons, we use only the diagonal of the optimal weighting matrix:

$$W^{diag}=\operatorname{diag}\left(V^{longrun}\right)^{-1}. \tag{B.21}$$

First, we would like to put more weight on moments that are actually observed in the data and that are economically meaningful, rather than on a linear combination of moments (see also Cochrane, 2005).[7] Second, in practice, fully specified weighting matrices often lead to diverging parameter estimates. As shown in Ruge-Murcia (2012), using only the main diagonal of the optimal weighting matrix leads to a loss in efficiency but nevertheless delivers good results in most cases.

When testing the starting values and the convergence of the SMM estimator, we often found the global minimum being located in regions of the parameter space typically considered to be extremely unlikely based on estimates using micro data (see e.g. An and Schorfheide, 2007, on the "dilemma of absurd parameter estimates"). At the same time, the local minima in more plausible regions of the parameter space were characterized by only slightly worse values of the target function. Thus, we formally include our prior knowledge about plausible parameter ranges into our SMM estimation as suggested in Ruge-Murcia (2010). Similar to adding the log prior to the log likelihood in full information Bayesian estimation, we augment the traditional SMM by using the prior distribution of the parameters as an additional moment restriction. Denote the mean of the prior distribution with θ^{prior} and its covariance matrix

[7] We additionally increase the weight on output volatility by multiplying the associated entry of the weighting matrix by 0.05.

with $\Omega(\theta^{prior})$. Instead of (B.17), we minimize the distance function

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} \left[\begin{array}{c} m\left(x_t\right) - m\left(x_t^{sim}\left(\theta\right)\right) \\ \theta^{prior} - \theta \end{array} \right]' \left[\begin{array}{cc} W & 0 \\ 0 & \Omega\left(\theta^{prior}\right)^{-1} \end{array} \right] \left[\begin{array}{c} m\left(x_t\right) - m\left(x_t^{sim}\left(\theta\right)\right) \\ \theta^{prior} - \theta \end{array} \right]. \tag{B.22}$$

This approach only requires us to specify the first two moments of the prior distribution of parameters, but not the whole distribution. Thus, given the natural bounds on some parameters, it can be interpreted as using a (truncated) normal distribution. The natural parameter bounds are imposed during estimation by using a logistic/log transformation to transform bounded parameters into unbounded ones.

Table B.1: Prior Distributions of Estimated Structural Parameters

Parameter	Description	Mean	Var.
ϕ_c	Consumption habits	0.70	2
σ_l	Wealth effect labor supply	1.00	1
θ_w	Calvo parameter wages	0.50	0.4
χ_w	Wage indexing	0.50	2
θ_p	Calvo parameter prices	0.5	0.4
χ_p	Price indexing	0.50	2
ρ_R	Interest smoothing	0.5	0.5
ϕ_{Ry}	Output feedback Taylor rule	0.50	0.5
$\phi_{R\pi}$	Inflation feedback Taylor rule	1.50	6
γ	Labor disutility parameter	9.00	∞
δ_2/δ_1	Elasticity of capital utilization	0.15	4
κ	Capital adjustments costs	7.00	40
σ_{me}	Measurement error on wages	-7.13	1

Notes: The table shows the prior mean and variance for the respective parameters used in the SMM estimation. A variance of infinity indicates the use of a flat prior.

We generally use relatively uninformative priors as shown in Table B.1. To get a feeling about the shape of the prior distributions, Figure B.1 compares the prior for the Calvo parameter θ_p used in the present study to the one in Smets and Wouters (2007). While they use a beta prior with mean 0.5 and standard deviation of 0.15, we use a prior mean of 0.5 and a prior variance 0.4. Given that the parameter is bounded to the range $[0, 1]$, this corresponds to a truncated normal distribution with mean 0.5 and standard deviation 0.2768.[8] As can be seen, the priors used in our study are considerably flatter and less informative than the ones typically used in full information estimation.

[8]The mean and the variance of a normal distribution with mean μ and standard deviation σ truncated to

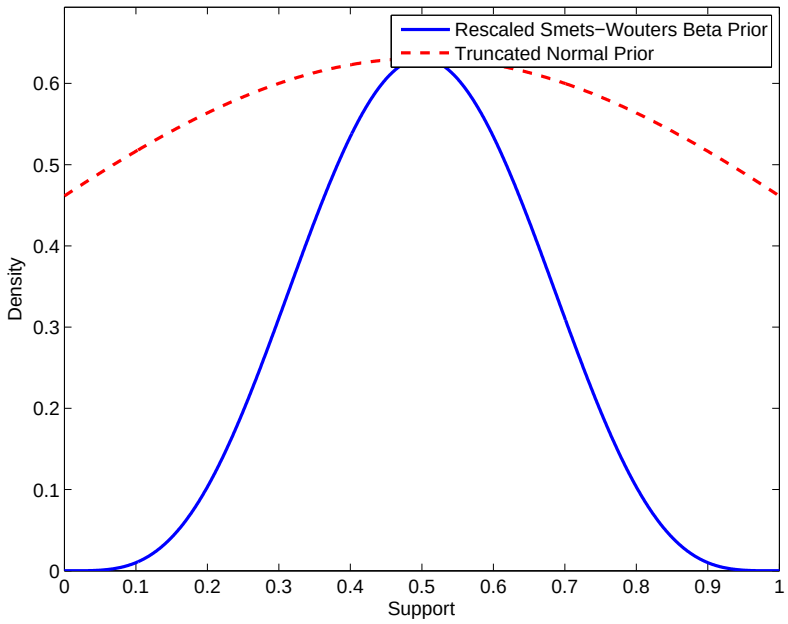


Figure B.1: Exemplary comparison between the prior used for the Calvo parameter θ_p in this study (truncated normal) and the beta prior used in Smets and Wouters (2007). The beta prior has been rescaled to have the same mode as the truncated normal.

The simulation proceeds as follows. Starting at the deterministic steady state, we simulate the model for 3200 quarters using shocks drawn from the estimated shock distributions. To assure non-explosive behavior of the simulations, we use the pruning algorithm of Andreasen et al. (2013). We discard the first 1500 quarters as a burn-in in order to reach the ergodic distribution. We then use the remaining 1700 quarters to compute the respective moments. The results are robust to using a longer burn-in period. The choice of using ten times the length of the original data sample (i.e., $a = 10$) to compute the moments is motivated by the simulations in Ruge-Murcia (2012), who finds this choice to deliver a good balance between the precision of the estimates and computation time.

---

the interval $[a, b]$ is given by

$$E\left(X|a<X<b\right)=\mu+\frac{\phi\left(\frac{a-\mu}{\sigma}\right)-\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}\sigma$$

$$Var\left(X|a<X<b\right)=\sigma^2\left[1+\frac{\frac{a-\mu}{\sigma}\phi\left(\frac{a-\mu}{\sigma}\right)-\frac{b-\mu}{\sigma}\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}-\left(\frac{\phi\left(\frac{a-\mu}{\sigma}\right)-\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}\right)^2\right]$$

B.7 Impulse Responses

The nonlinearity of our model complicates the computation of impulse responses compared to linear models. We follow Fernández-Villaverde et al. (2011) and generate impulse responses as the response to a two standard deviation shock to uncertainty at the ergodic mean in the absence of shocks/stochastic steady state.[9] First, we simulate the model for 25,000 quarters without shocks starting from the steady state. We then check for convergence, i.e., whether the maximum change during the last 500 periods was bigger than 1e-12. If yes, we iterate another 5000 periods. We repeat this until convergence. To assure non-explosive behavior of the simulations, we use the pruning algorithm of Andreasen et al. (2013). Starting at the ergodic mean in the absence of shocks, we compute the IRFs as the percentage difference of the respective variables between the system shocked with the respective shock and the baseline model response, i.e., the model response without shocks.

B.8 GMM

We construct the monetary policy shocks by specifying the Federal Reserve's policy reaction function and estimating it by the generalized method of moments (GMM). Our approach is similar to the one used in Clarida et al. (2000), but slightly modified to stay consistent with our DSGE-model. Specifically, the policy reaction function to be estimated is given by

$$r_t = \rho r_{t-1} + (1-\rho)\left[\bar{r} + \phi_\pi\left(\pi_t - \bar{\pi}\right) + \phi_y y_t^{gap}\right] + \varepsilon_t \ , \tag{B.23}$$

where π_t is inflation with target rate $\bar{\pi}$, y_t^{gap} is the output gap, r_{t-1} allows for interest smoothing, $\bar{r}$ ist the target nominal interest rate, and ε_t is an error term. Using the vector of instruments z_t, the set of moment conditions for our GMM estimation procedure can be written as

$$E\left[\left\{r_t - \rho r_{t-1} - \alpha - \beta\pi_t - \gamma y_t^{gap}\right\} z_t\right] = 0 \tag{B.24}$$

where $\alpha = (1-\rho)\left(\bar{r} + \phi_\pi\bar{\pi}\right)$ collects all constant terms, $\beta = (1-\rho)\phi_\pi$, and $\gamma = (1-\rho)\phi_y$.

Hence, we regress the average effective Federal Funds Rate in the first month of the quarter on the lagged FFR, the inflation rate, and the output gap, where all rates are annualized. The set of instruments includes four lags of the FFR, the inflation rate, the output gap, commodity price inflation, money growth, and the interest term spread. Because we are only interested in the residuals of the policy reaction function $\widehat{\varepsilon}_t$, we do not need to separately identify the target nominal rate $\bar{r}$ and target inflation $\bar{\pi}$.

Table B.2 presents the estimation results, which are all in the range typically reported in

[9] See Born and Pfeifer (2014) for details.

Table B.2: GMM Estimation of Taylor Rule

Coefficient	Mean	Std. Error	t-Statistic	Prob.
ρ	0.902	0.019	47.728	0.000
α	0.000	0.000	0.837	0.404
β	0.127	0.030	4.218	0.000
γ	0.071	0.008	9.311	0.000
R-squared	0.898	Mean dependent var		0.0150
Adjusted R-squared	0.897	Sum squared resid		0.002
S.E. of regression	0.003	J-statistic		15.730
Durbin-Watson stat	2.508	pval(J-statistic)		0.785

Note: GMM estimation using EViews; estimation weighting matrix: HAC (Bartlett kernel, Newey-West fixed bandwidth = 5.0000); standard errors & covariance computed using estimation weighting matrix.

the literature. There is evidence of interest smoothing with $\rho = 0.902$. The point estimates of the feedback parameters are $\phi_\pi = 1.296$ and $\phi_y = 0.725$. The test of overidentifying restrictions shows that the model cannot be rejected at conventional significance levels.

C Diagnostics

C.1 Testing for Heteroskedasticity

Table C.1 presents evidence of the need to model time-varying volatility. Despite our relatively short sample size and the low power of tests for heteroskedasticity, the null hypothesis of homoskedastic shocks can be rejected at the 10% level for all series except TFP. This result is consistent with evidence that the standard deviation of structural shocks has changed over time (see, e.g., Justiniano and Primiceri, 2008; Primiceri, 2005).

Table C.1: Tests for Heteroskedasticity

	τ^k	τ^n	z	z_I	g	m
White	0.000*	0.000*	0.958	0.023*	0.383	0.000*
White/Wooldridge	0.883	0.206	0.958	0.003*	0.050*	0.000*
Breusch/Pagan/Koenker	0.001*	0.003*	0.779	0.553	0.140	0.001*

Notes: Asterisks indicate significance at the 10% level. White refers to the standard White (1980)-test, WW refers to the Wooldridge (1990)-version of this test, and BPK refers to the Breusch and Pagan (1979)/Koenker (1981)-test.

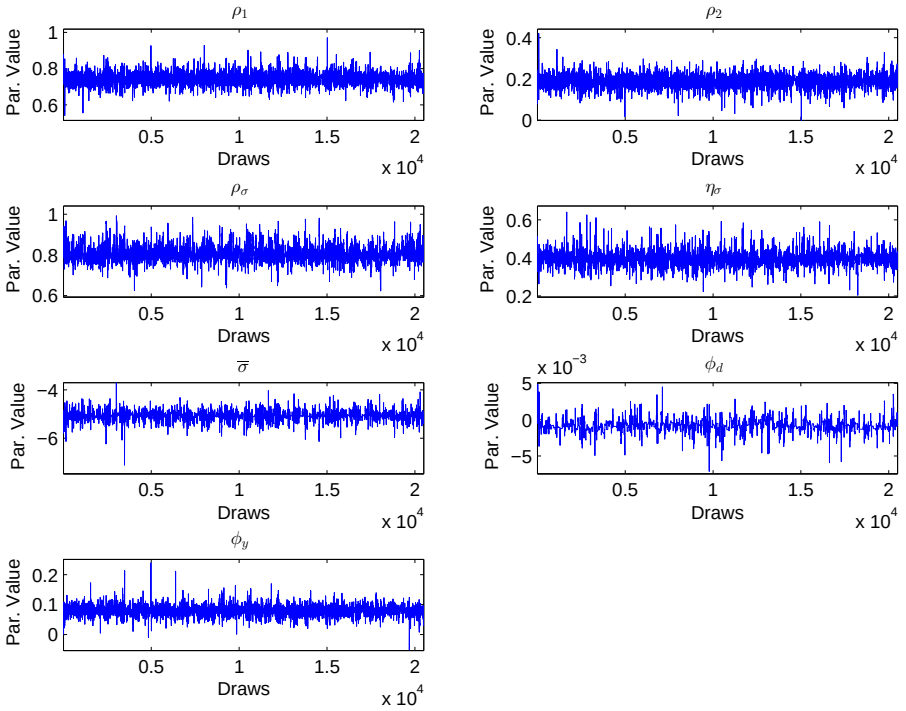
C.2 Convergence Diagnostics

Table (C.2) shows the results from the Geweke (1992)-convergence diagnostics that compares the means of the first 20% of draws with that of the last 50% of the draws. In general, all MCMC chains have converged to their stationary distribution as indicated by the p-values of the χ^2-test for equal means. Figures C.1 to C.6 show the corresponding mean plots.

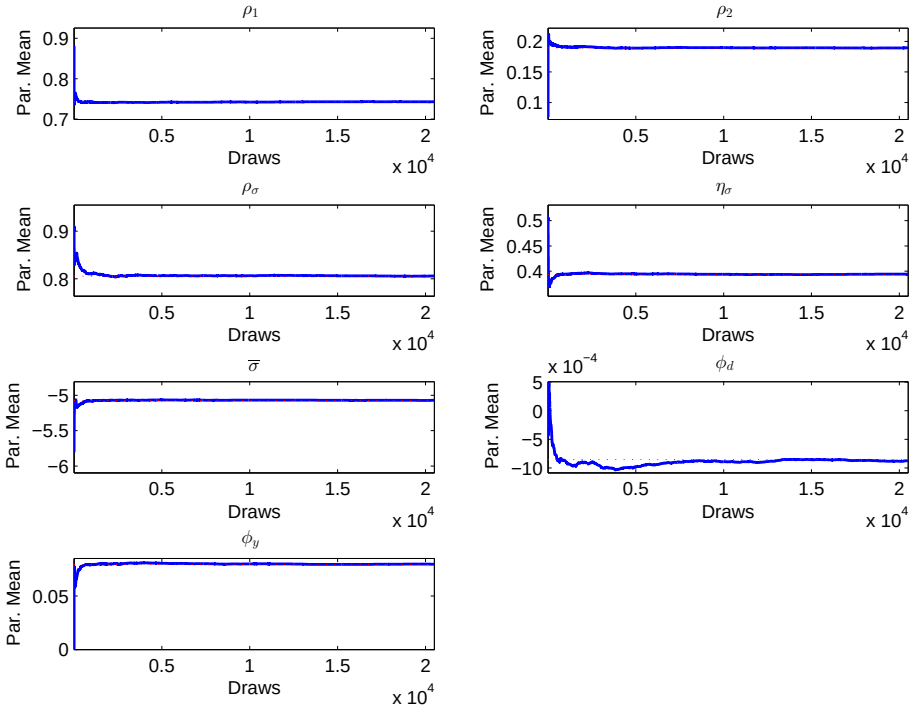
Table C.2: Geweke (1992) Convergence Diagnostics

Parameter	4% taper	8% taper	15% taper	4% taper	8% taper	15% taper
	Capital Tax Rates			Labor Tax Rates		
ρ_1	0.256	0.287	0.331	0.622	0.624	0.621
ρ_2	0.760	0.760	0.773	0.622	0.624	0.621
ρ_σ	0.896	0.904	0.906	0.387	0.403	0.382
η_σ	0.127	0.121	0.091	0.943	0.943	0.934
$\bar{\sigma}$	0.499	0.529	0.532	0.608	0.611	0.616
ϕ_d	0.129	0.085	0.070	0.369	0.379	0.326
ϕ_y	0.414	0.434	0.482	0.882	0.878	0.872
	Investment Specific Technology			Government Spending		
ρ_1	0.078	0.109	0.113	0.179	0.182	0.183
ρ_2	0.100	0.140	0.148	0.148	0.163	0.131
ρ_σ	0.627	0.614	0.598	0.878	0.889	0.897
η_σ	0.479	0.491	0.491	0.757	0.755	0.722
$\bar{\sigma}$	0.721	0.712	0.674	0.969	0.969	0.968
ϕ_d				0.843	0.843	0.828
ϕ_y				0.303	0.297	0.222
	Total Factor Productivity			Monetary Policy Shock		
ρ_1	0.681	0.670	0.653	0.778	0.753	0.723
ρ_σ	0.217	0.243	0.258	0.785	0.781	0.759
η_σ	0.047	0.029	0.011	0.130	0.101	0.048
$\bar{\sigma}$	0.749	0.752	0.770	0.874	0.871	0.854

Notes: Numbers are p-values of the χ^2-test for equal means of the first 20% of draws and the last 50% of the draws (after the first 6500 draws are discarded as burn-in). ρ denotes the autocorrelation coefficient(s) of the level equation, ρ_σ the one of the volatility equation, $\bar{\sigma}$ is the steady state volatility of the level shocks, η_σ denotes the standard deviation of the volatility shocks, ϕ_d is the debt feedback coefficient, and ϕ_y the output feedback coefficient.

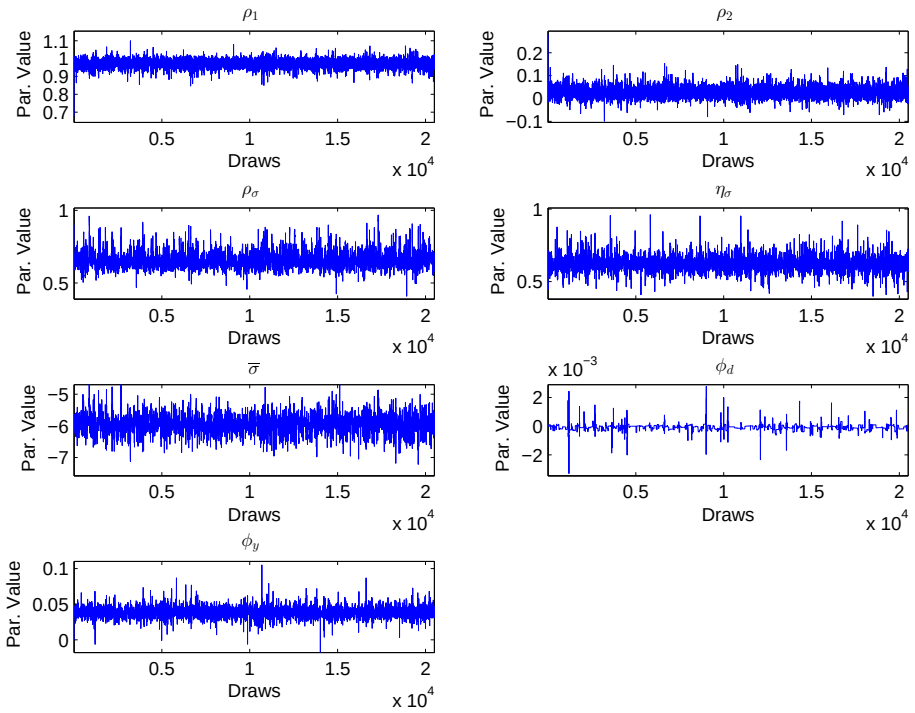


(a) MCMC draws.

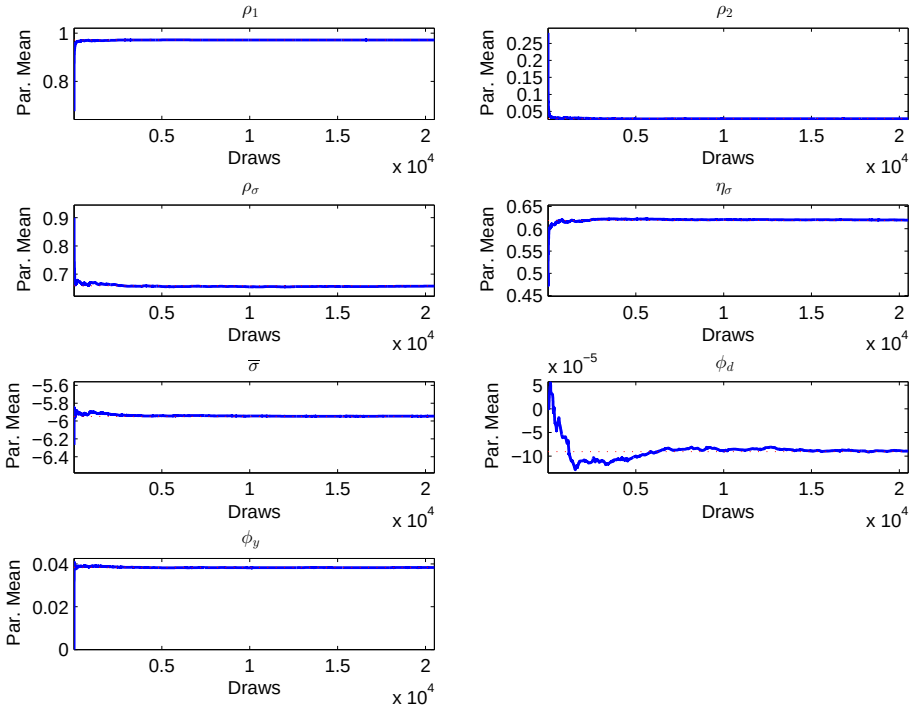


(b) Mean of the parameters over time.

Figure C.1: Evolution of MCMC sampler over time for τ^k.

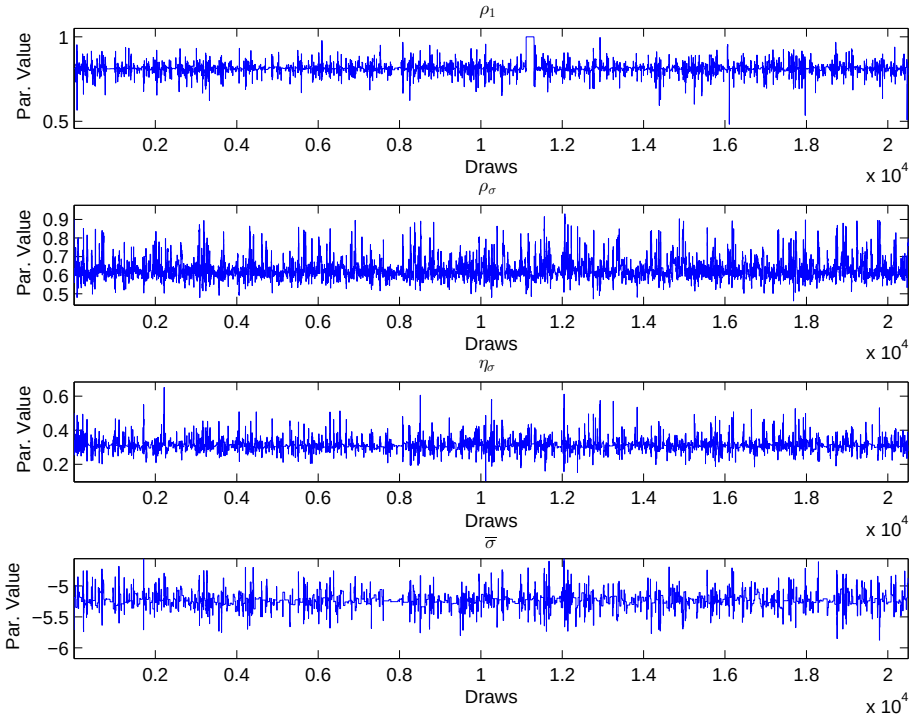


(a) MCMC draws.

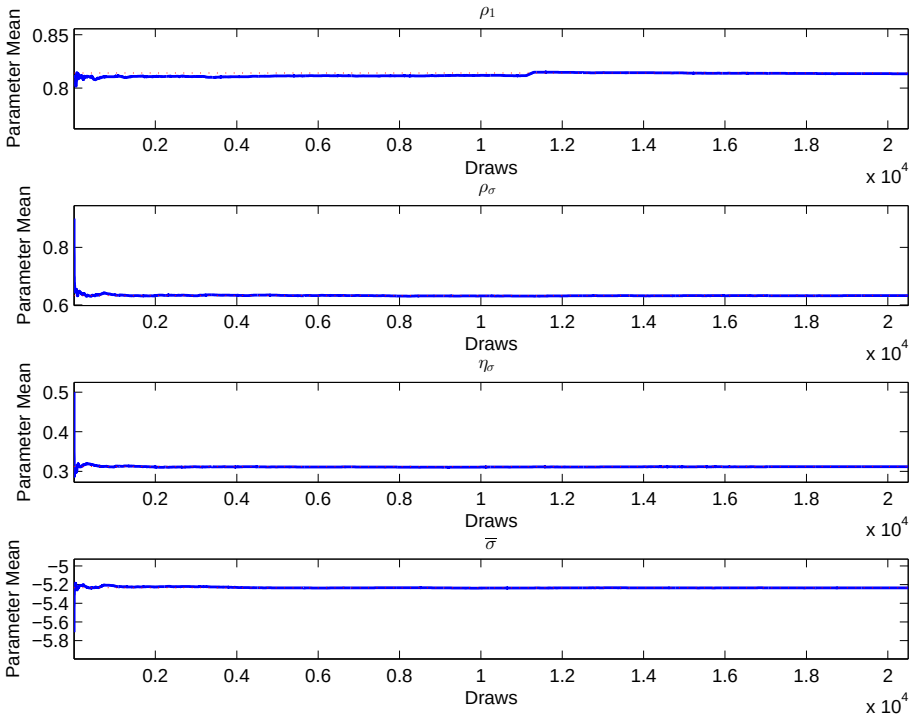


(b) Mean of the parameters over time.

Figure C.2: Evolution of MCMC sampler over time for τ^n.

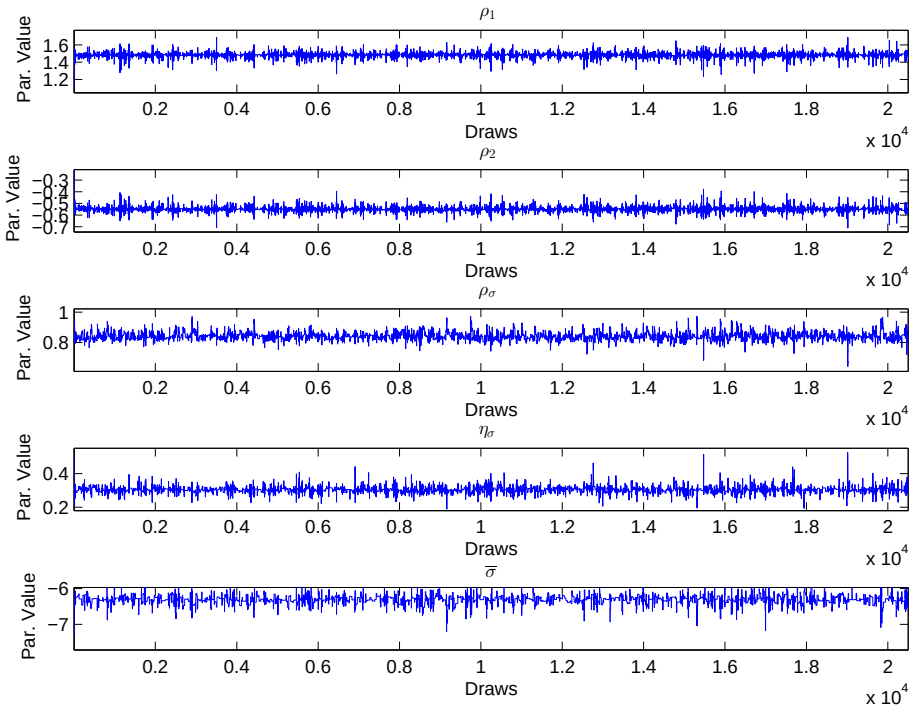


(a) MCMC draws.

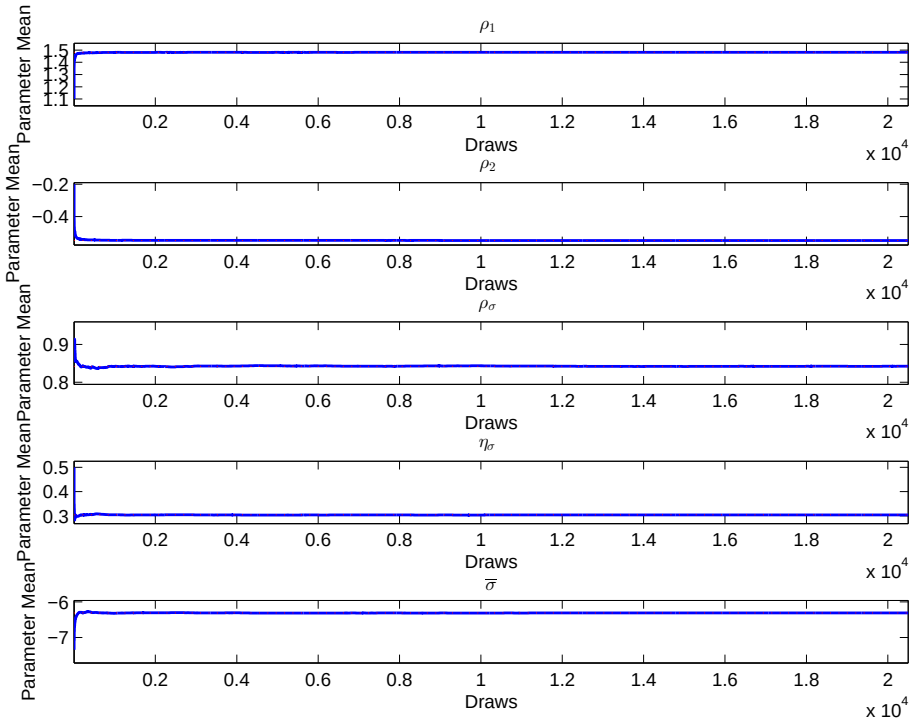


(b) Mean of the parameters over time.

Figure C.3: Evolution of MCMC sampler over time for z.

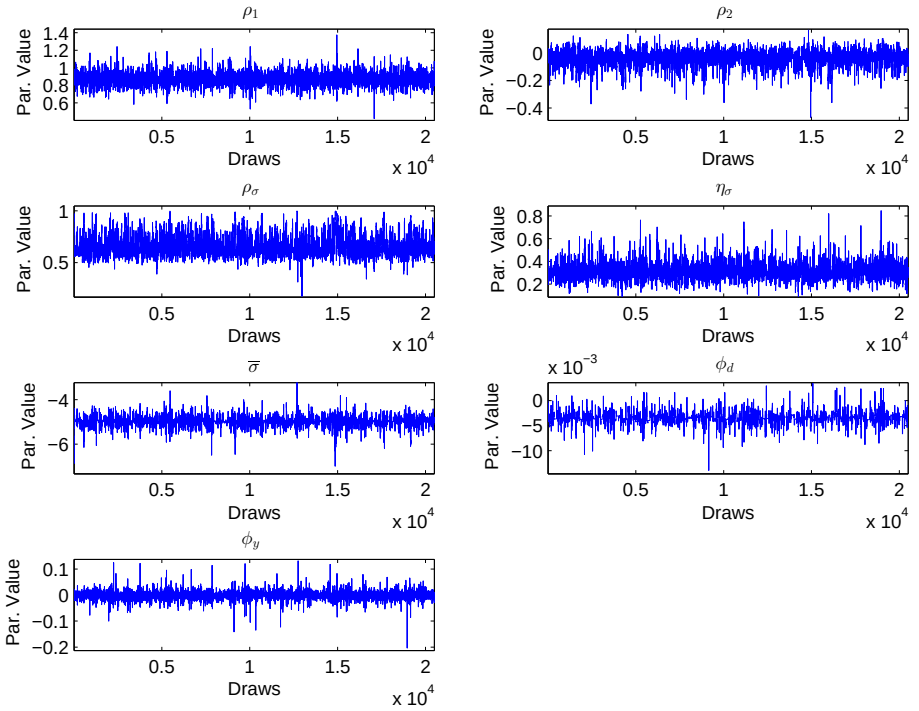


(a) MCMC draws.

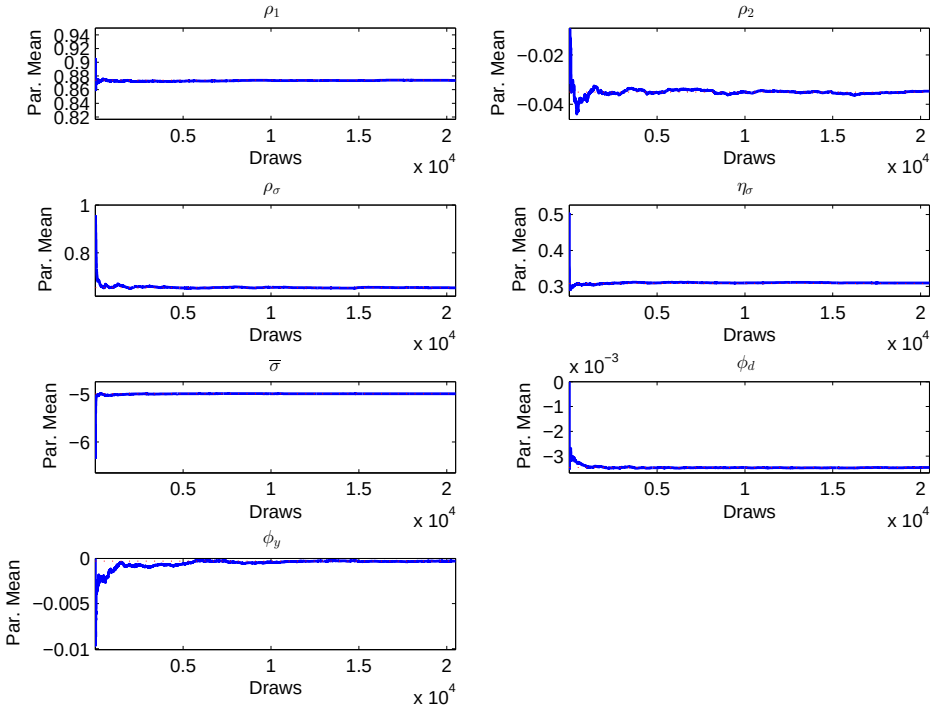


(b) Mean of the parameters over time.

Figure C.4: Evolution of MCMC sampler over time for z^I.

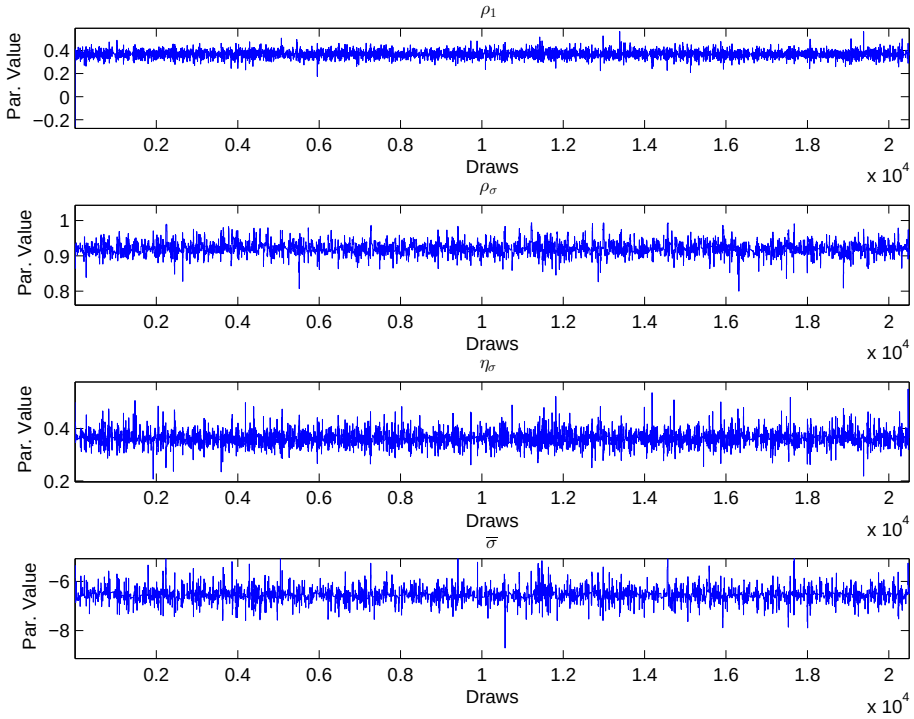


(a) MCMC draws.

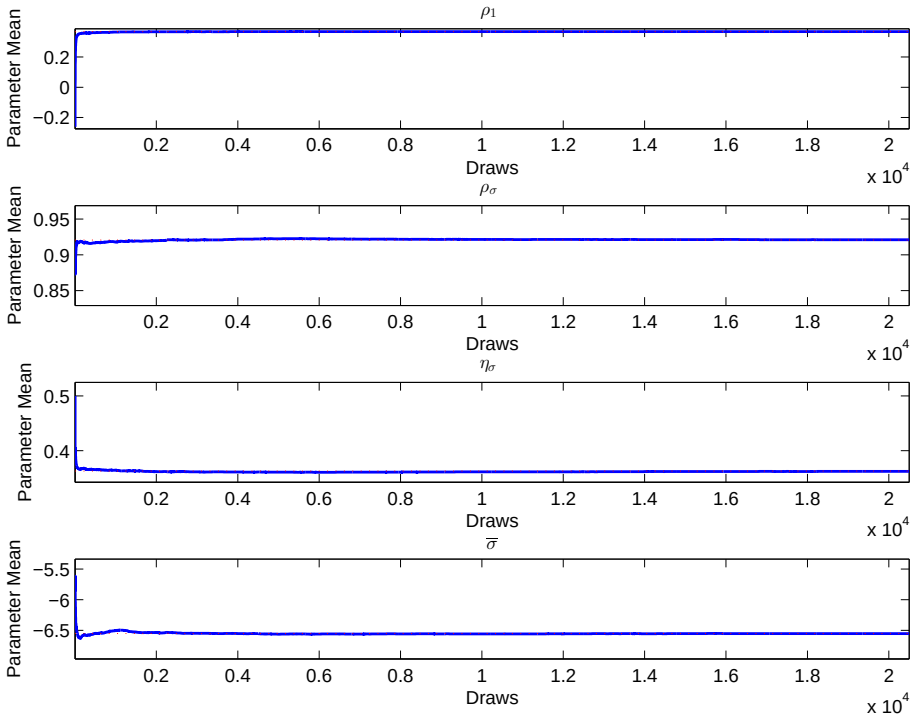


(b) Mean of the parameters over time.

Figure C.5: Evolution of MCMC sampler over time for g.



(a) MCMC draws.



(b) Mean of the parameters over time.

Figure C.6: Evolution of MCMC sampler over time for m.

C.3 Model Misspecification Diagnostics

Following Kim et al. (1998), we can test the specification of our SV-model. Using N draws from the prediction density $p\left(x_t|x^{t-1};\Theta\right)$, we can compute the probability that x_{t+1}^2 will be less or equal than the actually observed value of $\left(x_{t+1}^{obs}\right)^2$:

$$\Pr\left(x_{t+1}^2 \leqslant \left(x_{t+1}^{obs}\right)^2 \left| x^t\,;\Theta\right.\right) \simeq u_{t+1} = \frac{1}{N}\Pr\left(x_{t+1}^2 \leqslant \left(x_{t+1}^{obs}\right)^2 \left| x^t, \sigma_{t+1|t}\,;\Theta\right.\right)\,, \tag{C.1}$$

$\forall t = 1,\ldots T-1$. If the SV-model is correctly specified, the sequence of u_t converges in distribution to *i.i.d.* uniform variables as the number of particles N goes to infinity (Rosenblatt, 1952). Under the null hypothesis of a correctly specified model, the u_t can be transformed to *i.i.d.* standard normal variables using the inverse normal CDF. Hence, we can perform a simple test for misspecification by testing the resulting series for their normality. Figure C.7 shows the corresponding QQ-plots.

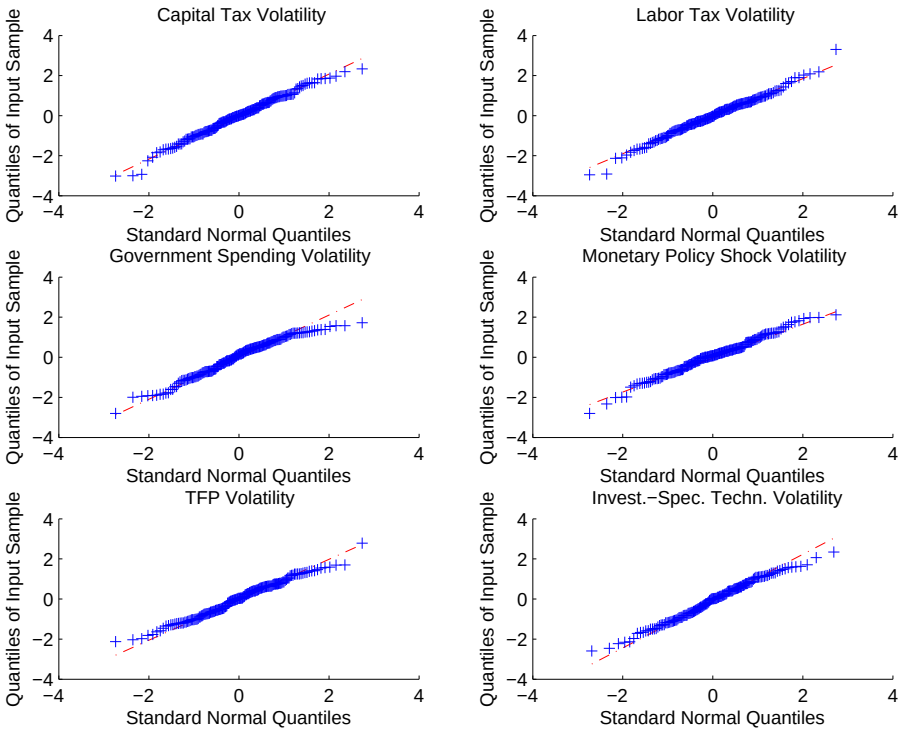


Figure C.7: QQ-plots. From left to right and top to bottom: capital taxes, labor taxes, TFP, investment-specific technology, monetary policy shocks, and government spending.

Table C.3 presents the results from three commonly used normality tests. In general, a correct specification of the model tends to not be rejected. Only for g, the Jarque-Bera and the Shapiro-Wilk tests reject normality. However, this effect is driven by the outliers visible in the center left panel of Figure C.7. In contrast, when shutting off the time-varying volatility

and setting the volatility to its unconditional mean, the specification is generally rejected (results are not shown here).

Table C.3: Tests for Model Misspecification

	JB	KS	SW
Capital Tax Rate τ_k	0.340	0.900	0.385
Labor Tax Rate τ_n	0.164	0.802	0.266
Government Spending g	0.041*	0.420	0.003*
Monetary Policy m	0.500	0.214	0.365
TFP z	0.444	0.789	0.179
Invest.-Spec. Technology z_I	0.218	0.345	0.399

Note: The table displays p-values of model-misspecification tests for the exogenous processes estimated in the first stage. Asterisks indicate significance at the 5% level. JB refers to the Jarque and Bera (1987)-test, KS refers to the Kolmogorov (1933)/Smirnov (1948)-test, and SW refers to the Shapiro and Wilk (1965)-test.

D Supplementary Tables and Figures

D.1 Baseline Model Without Stochastic Volatility

Table D.1: Simulated and empirical moments for the model without time-varying volatility

	Model	Data	Model	Data	Model	Data	Model	Data	Model	Data
	$\sigma(x_t)$		$\sigma_{x_t}/\sigma_{y_t}$		$\rho(x_t, y_t)$		$\rho(x_t, x_{t-1})$		$\rho(x_t, x_{t-2})$	
Δy	0.50%	0.85%	1.00	1.00	1.00	1.00	0.60	0.29	0.30	0.24
Δc	0.43%	0.55%	0.86	0.64	0.77	0.58	0.48	0.43	0.11	0.22
Δi	1.65%	2.44%	3.29	2.86	0.80	0.69	0.74	0.62	0.45	0.46
π	0.53%	0.62%	1.06	0.72	-0.26	-0.20	0.83	0.88	0.58	0.84
Δw	0.17%	0.64%	0.33	0.75	0.28	-0.00	0.43	0.03	0.22	0.03
r	0.48%	0.94%	0.96	1.10	-0.66	-0.11	0.92	0.94	0.78	0.89
l	1.74%	1.64%	3.47	1.91	0.06	0.33	0.83	0.93	0.65	0.82

Notes: Time Series x_t are output (y_t), consumption (c_t), investment (i_t), inflation (π_t), the real wage (w_t), the nominal interest rate (R_t), and hours worked l_t. Lowercase letters denotes variables that are logged. The columns show the standard deviation $\sigma(x_t)$, the relative standard deviation compared to output volatility $\sigma_{x_t}/\sigma_{y_t}$, the correlation with output $\rho(x_t, y_t)$, and the first two autocorrelations $\rho(x_t, x_{t-1})$ and $\rho(x_t, x_{t-2})$. Some of the target moments are transformed to correlations for better interpretation. The relative standard deviations with respect to the standard deviation of output are only implicitly targeted through the standard deviations of the respective series.

D.2 Overview about the Individual Shocks

D.2.1 Level Shocks

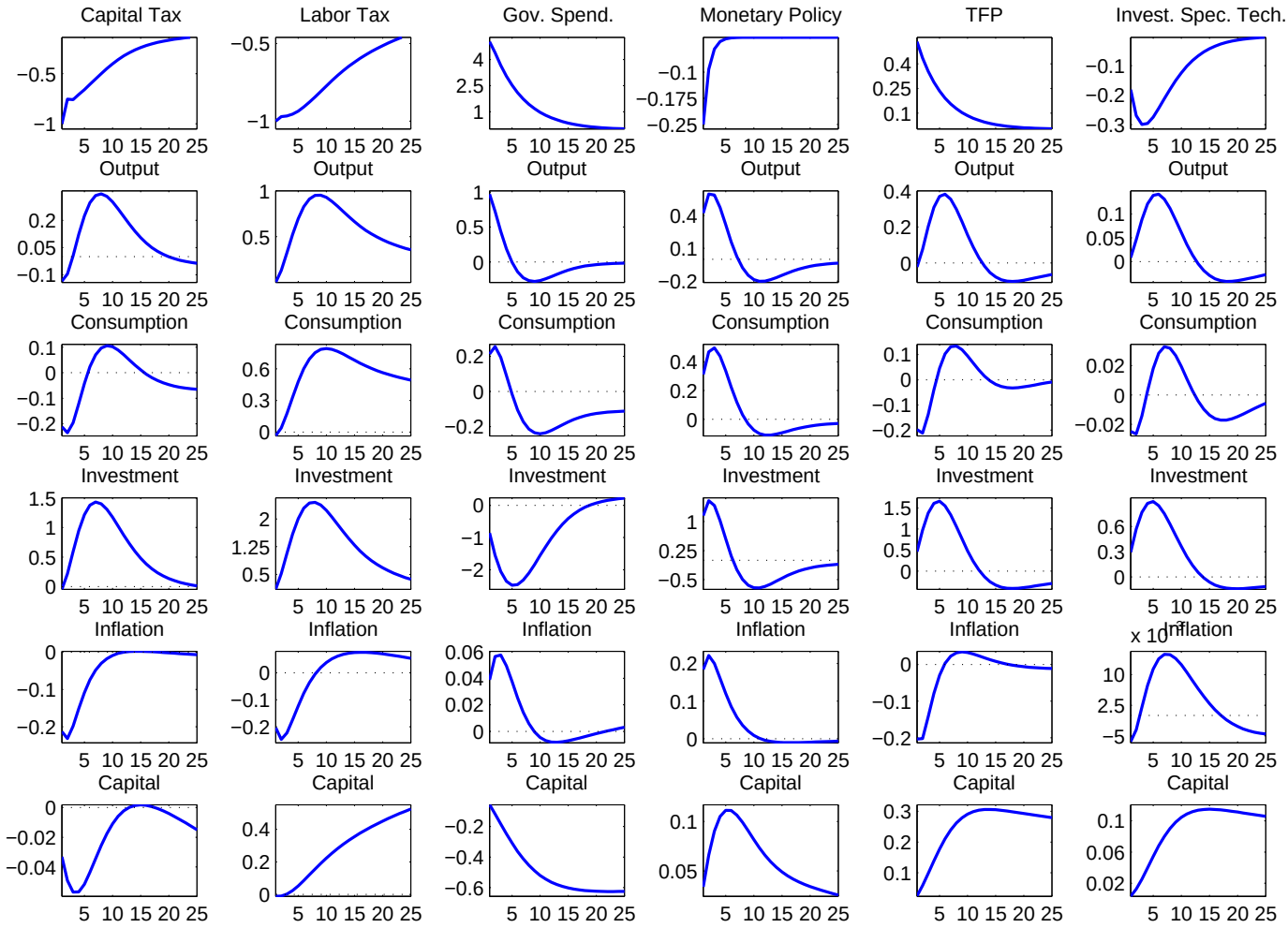


Figure D.1: The columns show the IRFs for output, consumption, investment, and inflation to a shock to the variable displayed in the first row. The capital and labor shocks are 1 percentage point cuts to the respective rates, while the government spending shock amounts to 1% of GDP. The monetary policy shocks of 25 basis points corresponds to a 1% exogenous cut in the Federal Funds Rate. The technology shocks correspond to one standard deviation shocks.

D.2.2 Risk Shocks

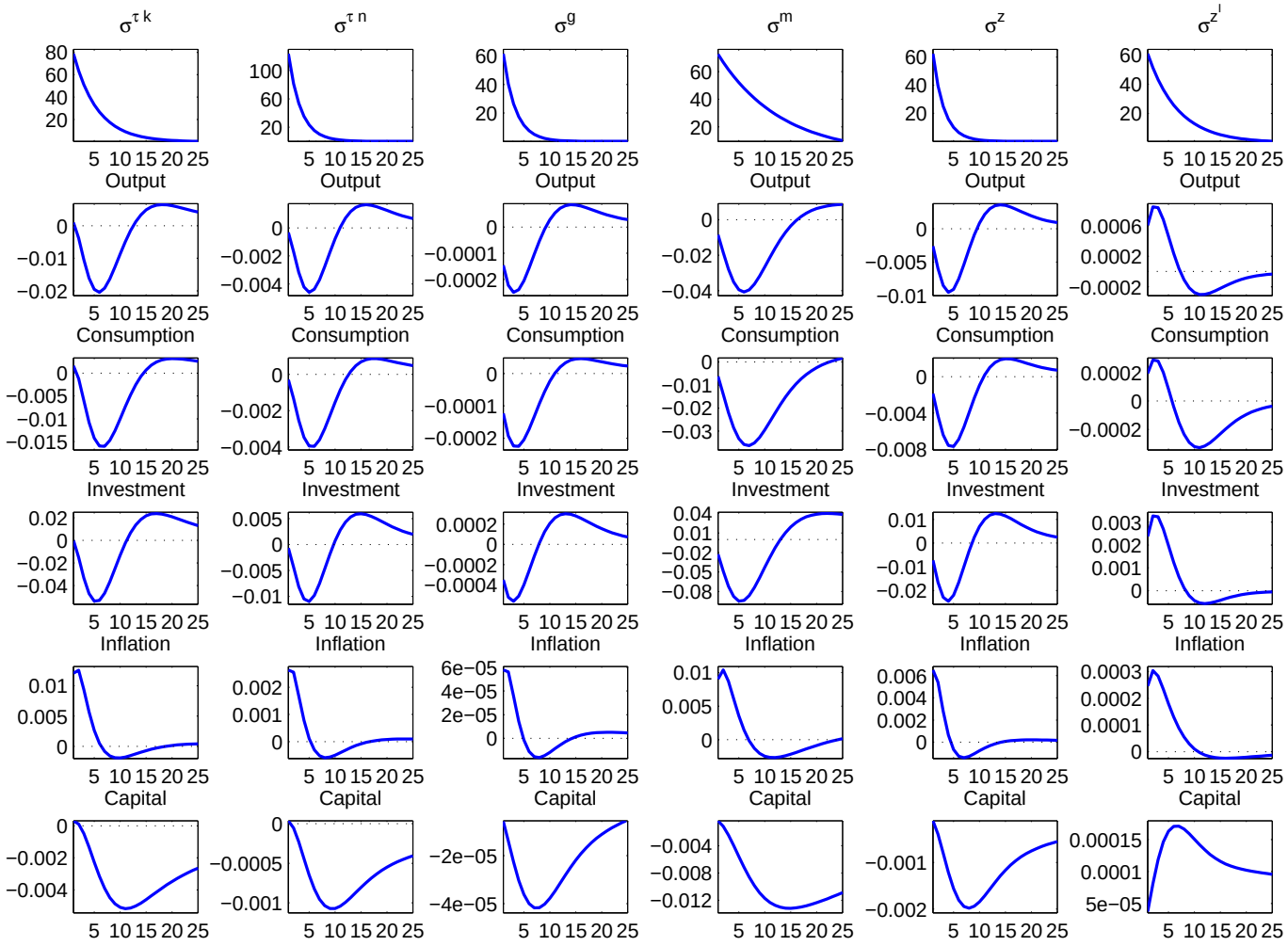


Figure D.2: The columns show the IRFs for output, consumption, investment, and inflation to a two-standard deviation shock to the volatility displayed in the first row.

D.3 Parameter Perturbations

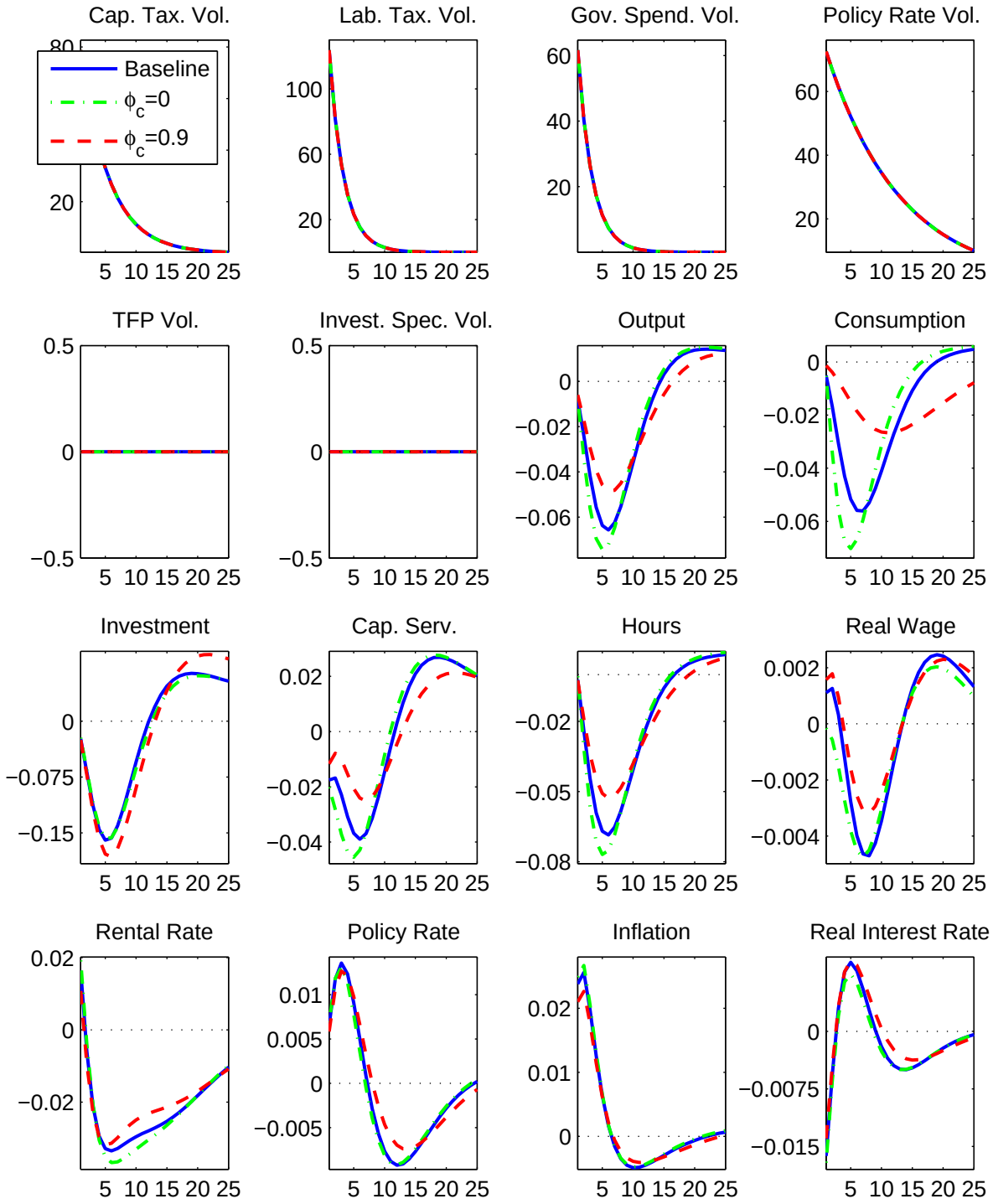


Figure D.3: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed and the green dashed dotted line indicate a higher and lower habit parameter ϕ_c.

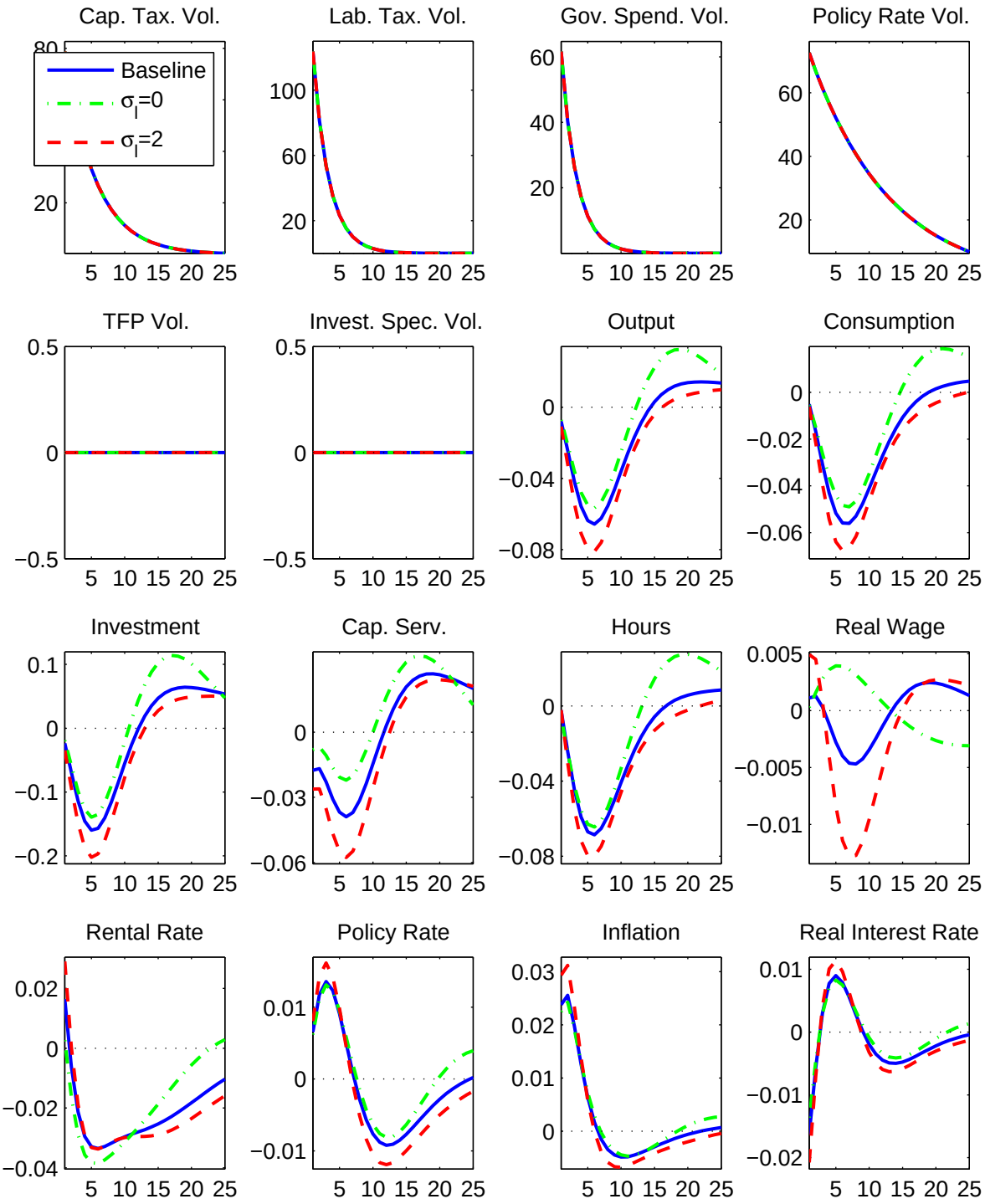


Figure D.4: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed and the green dashed dotted line indicate a higher and lower inverse Frisch elasticity parameter σ_l.

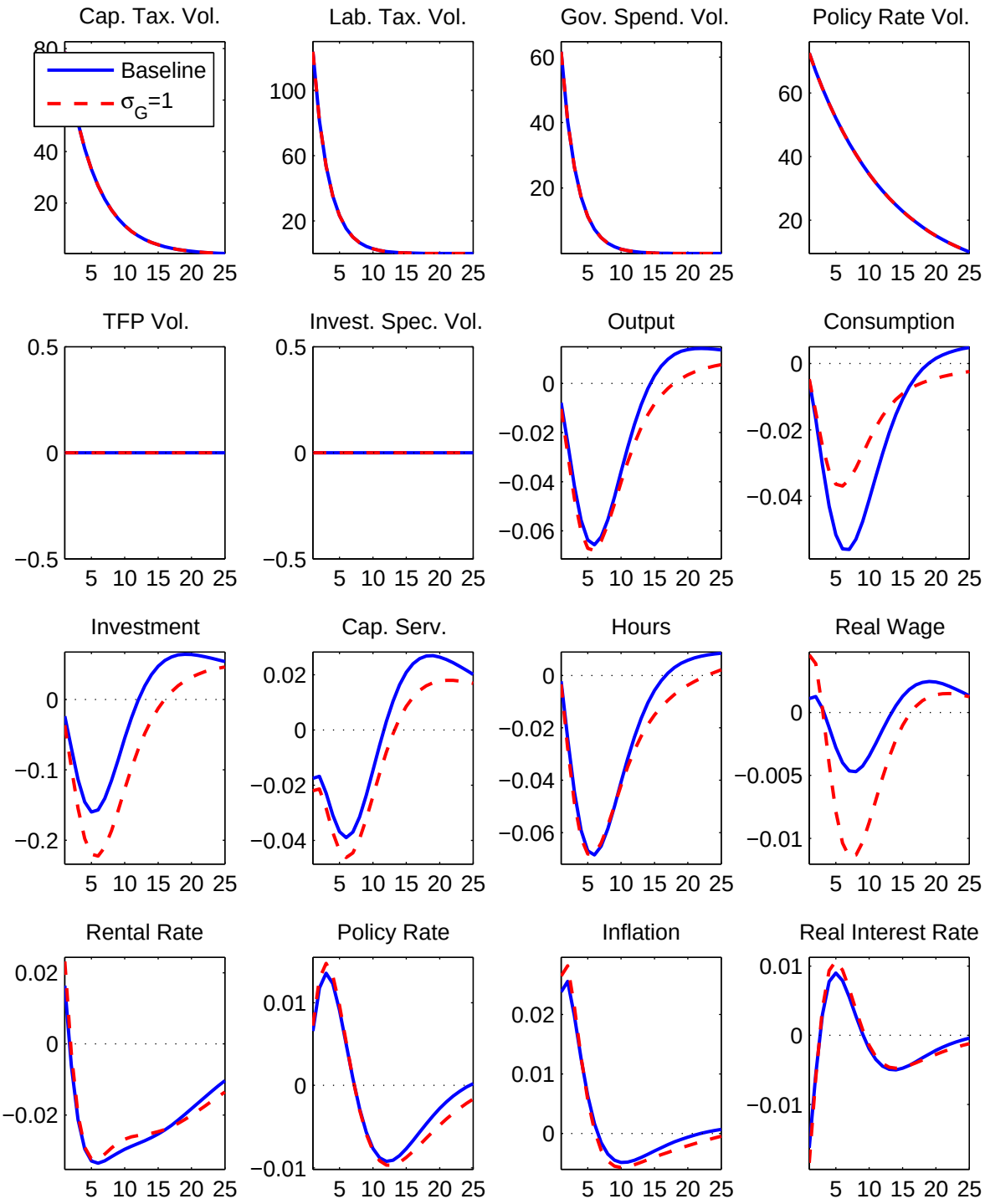


Figure D.5: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration with preferences close to the GHH case, while the red dashed line indicates King-Plosser-Rebelo preferences.

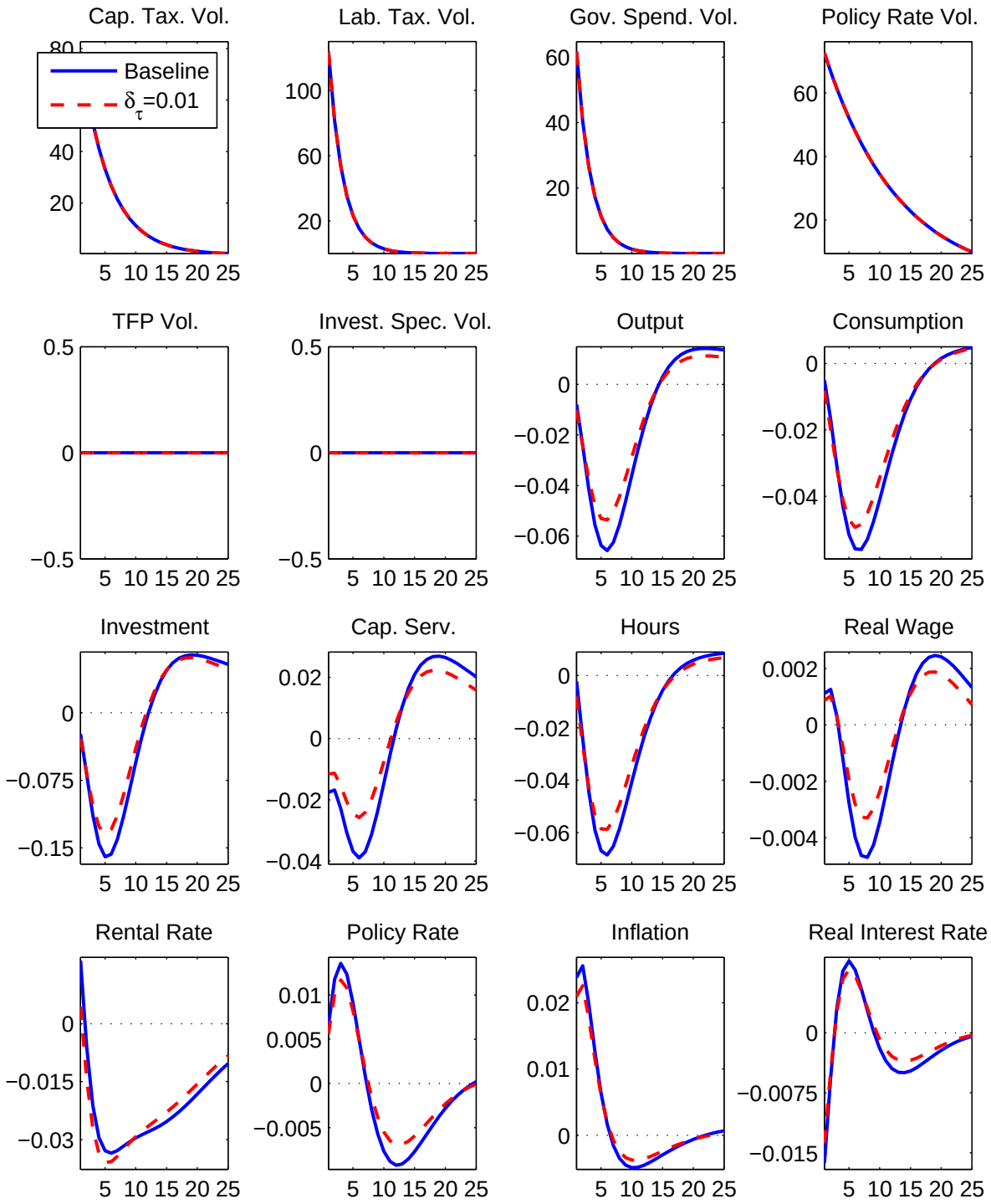


Figure D.6: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed line indicates lower depreciation rate for tax purposes δ_τ.

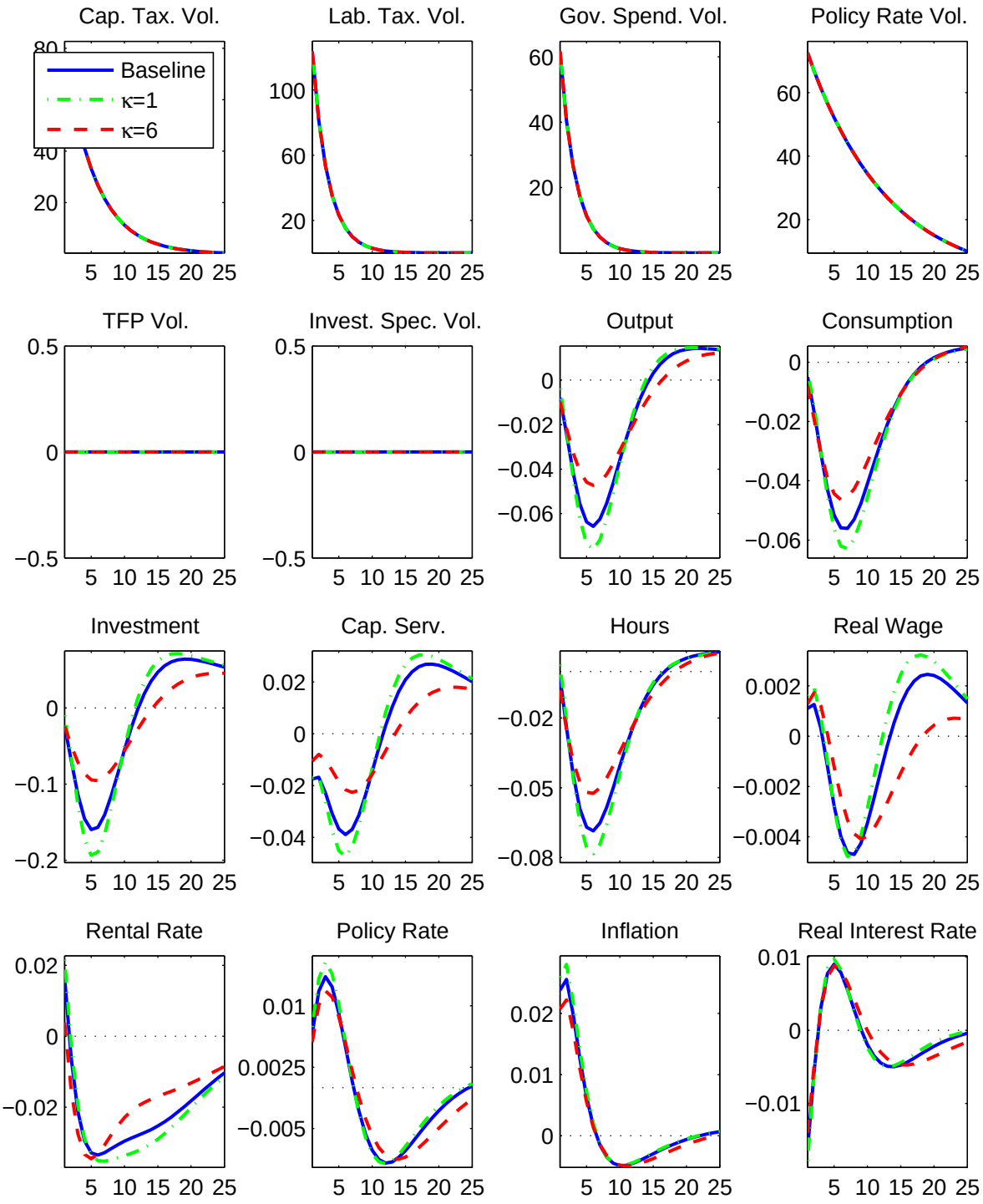


Figure D.7: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed and the green dashed dotted line indicate higher and lower investment adjustment costs.

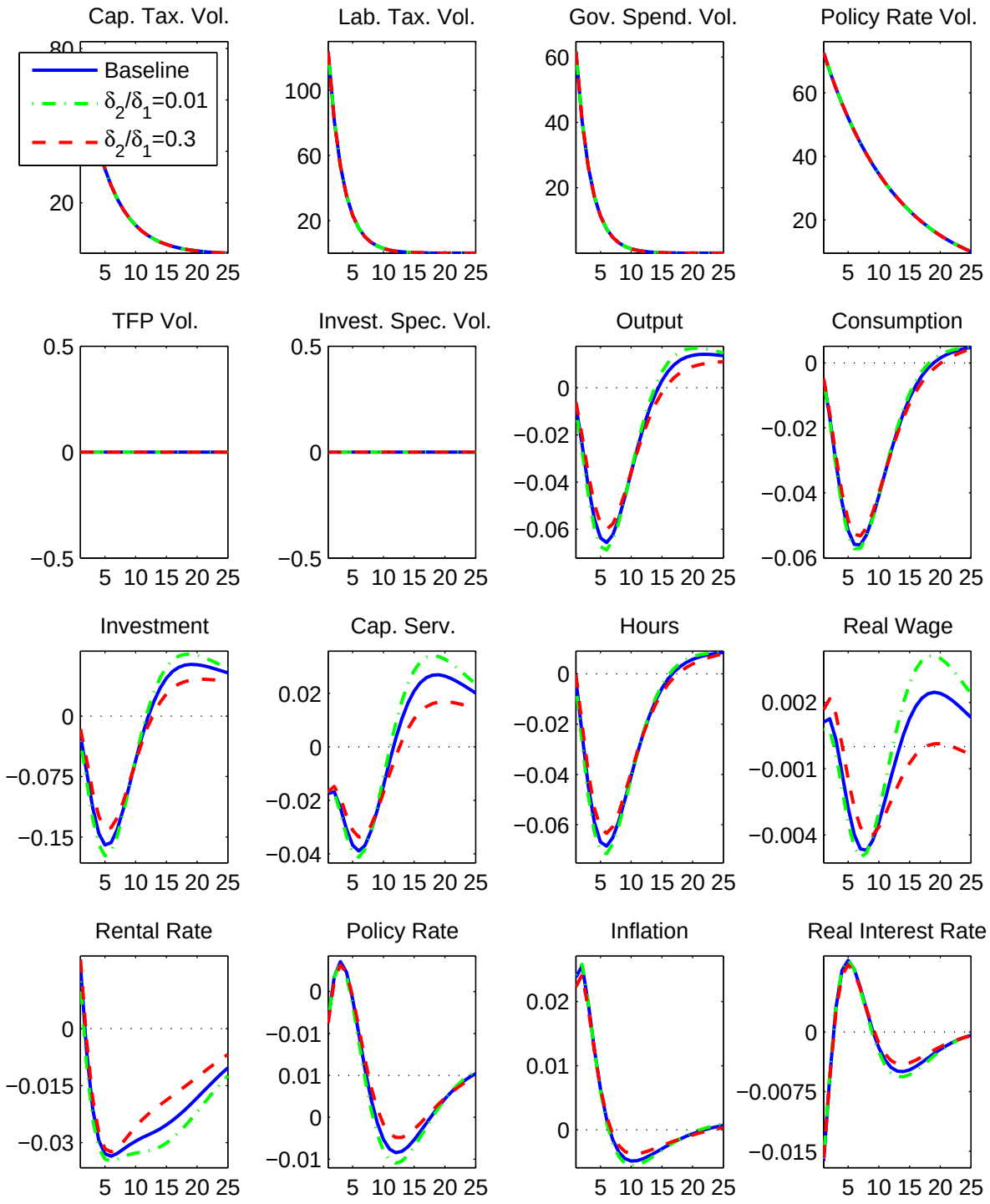


Figure D.8: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed and the green dashed dotted line indicate higher and lower capital utilization costs.

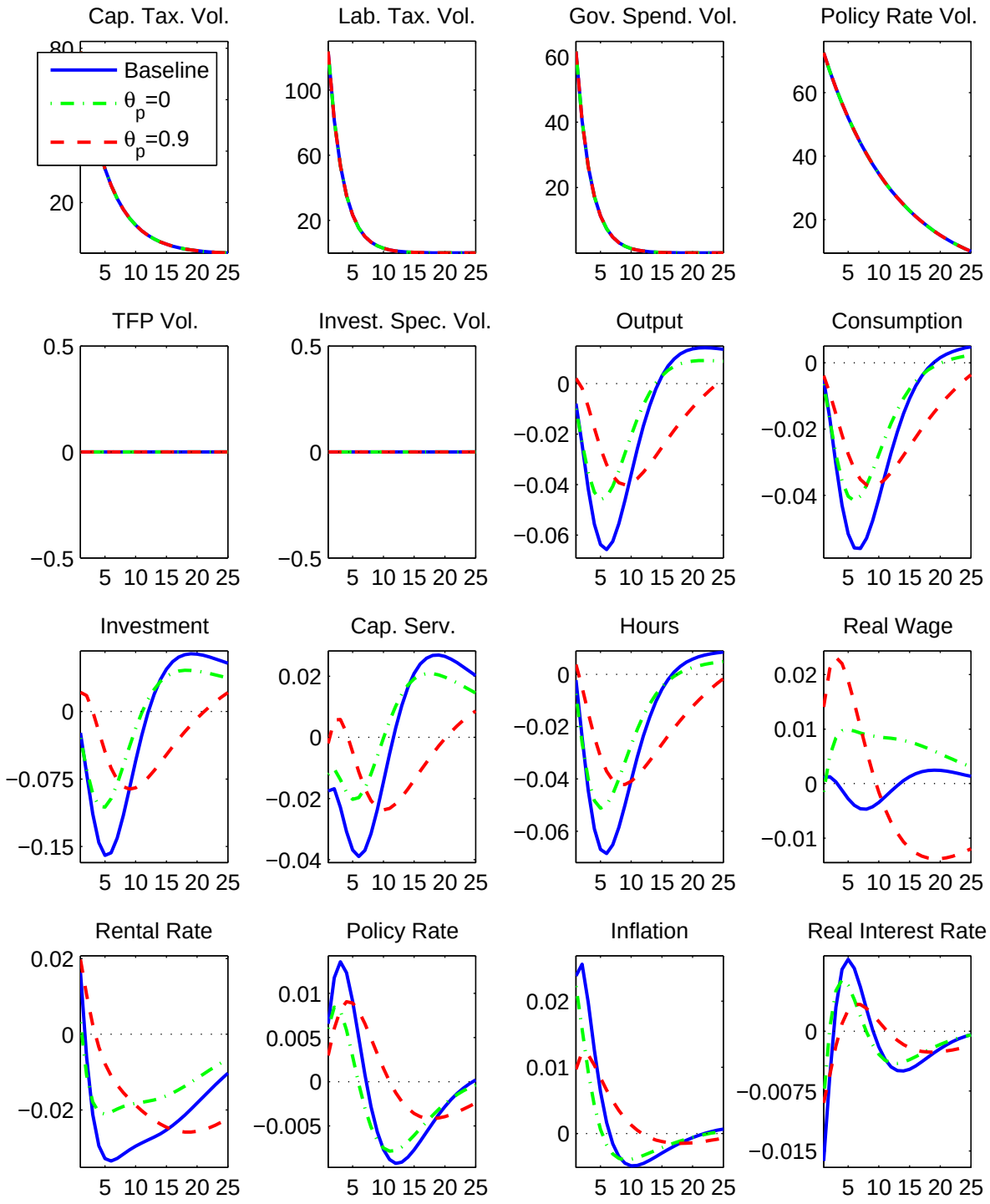


Figure D.9: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed and the green dashed dotted line indicate a lower and higher frequency of price adjustments θ_p.

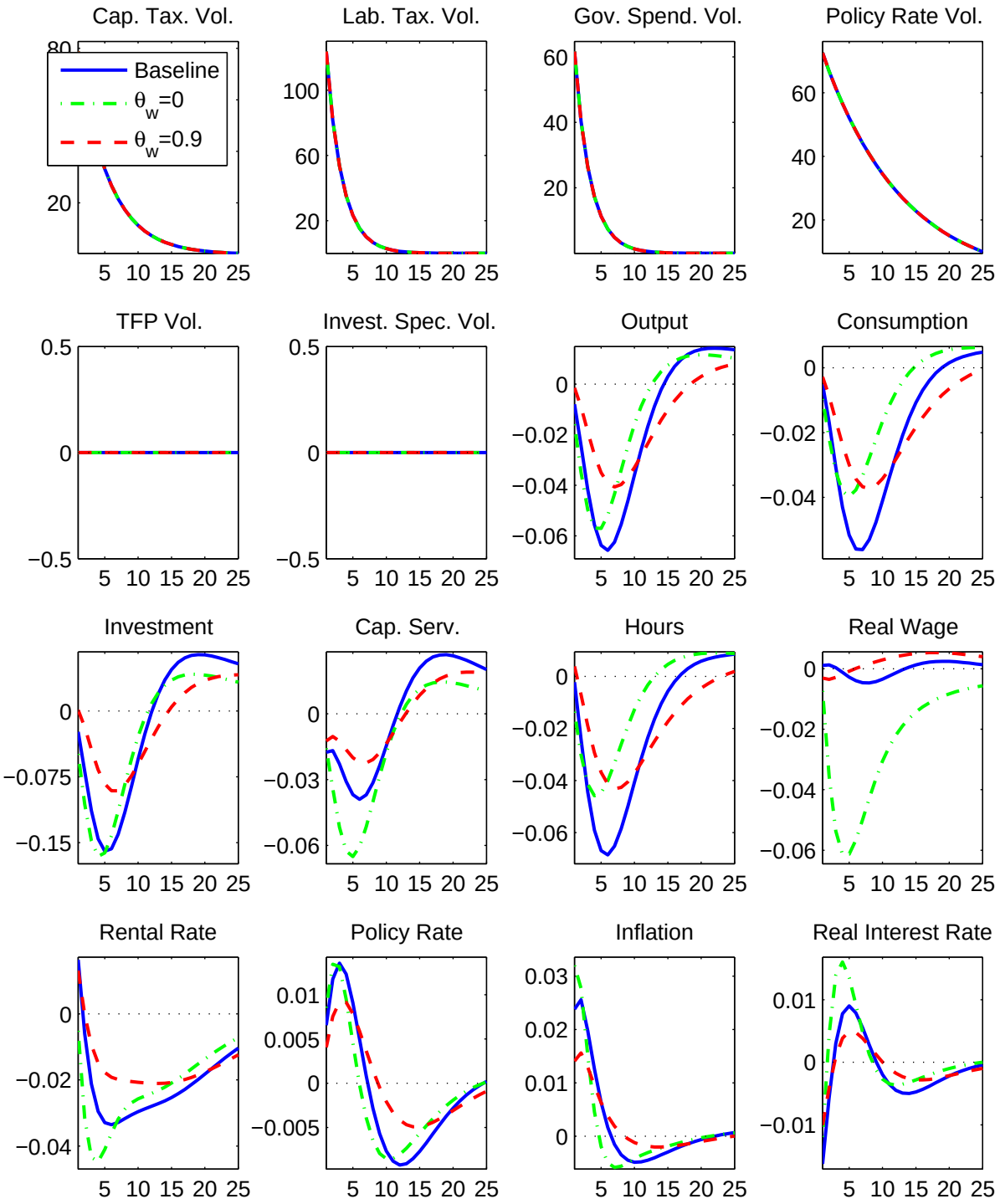


Figure D.10: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed and the green dashed dotted line indicate a lower and higher frequency of wage adjustments θ_w.

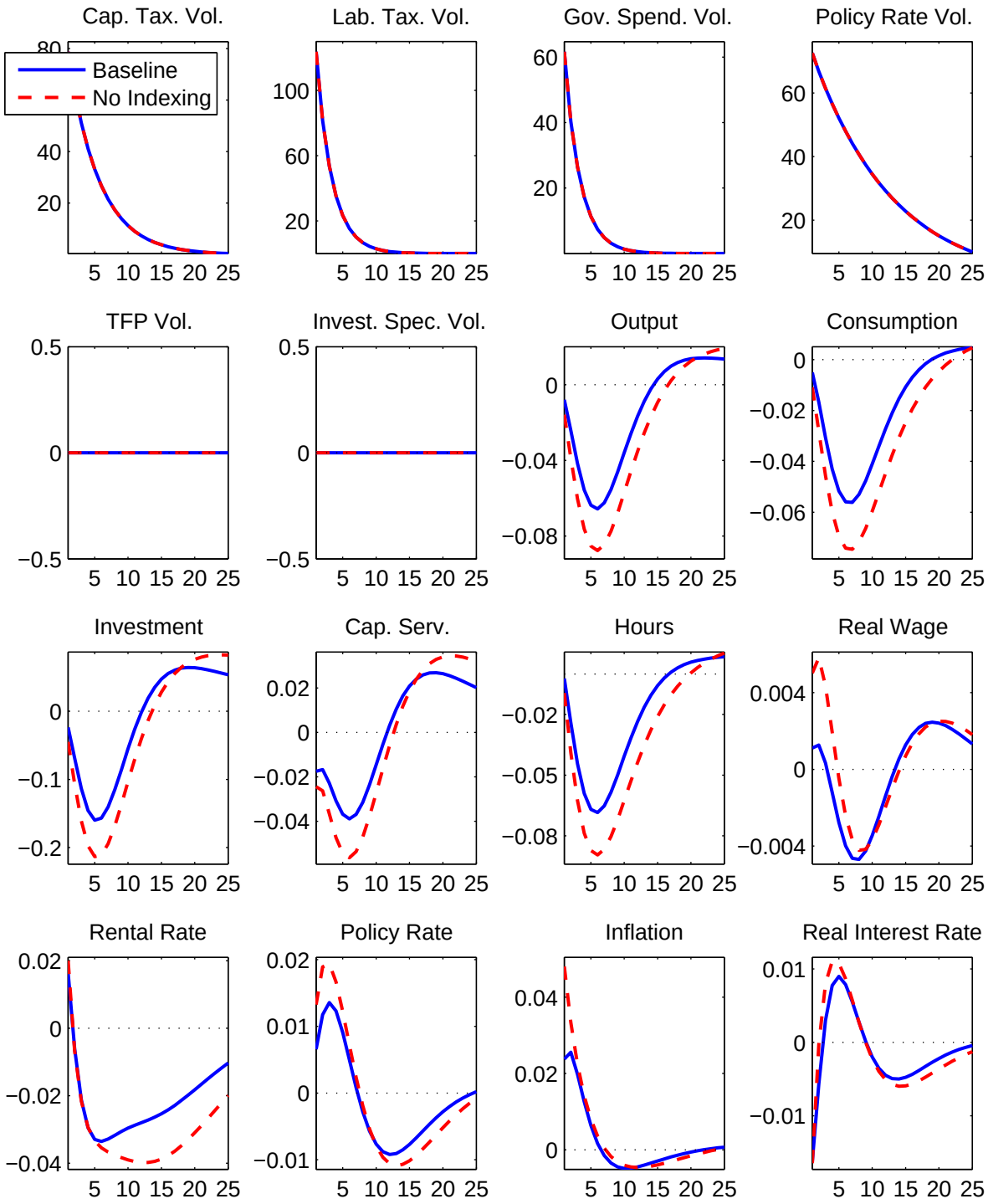


Figure D.11: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration with indexing, while the red dashed line indicates the absence of wage and price indexing.

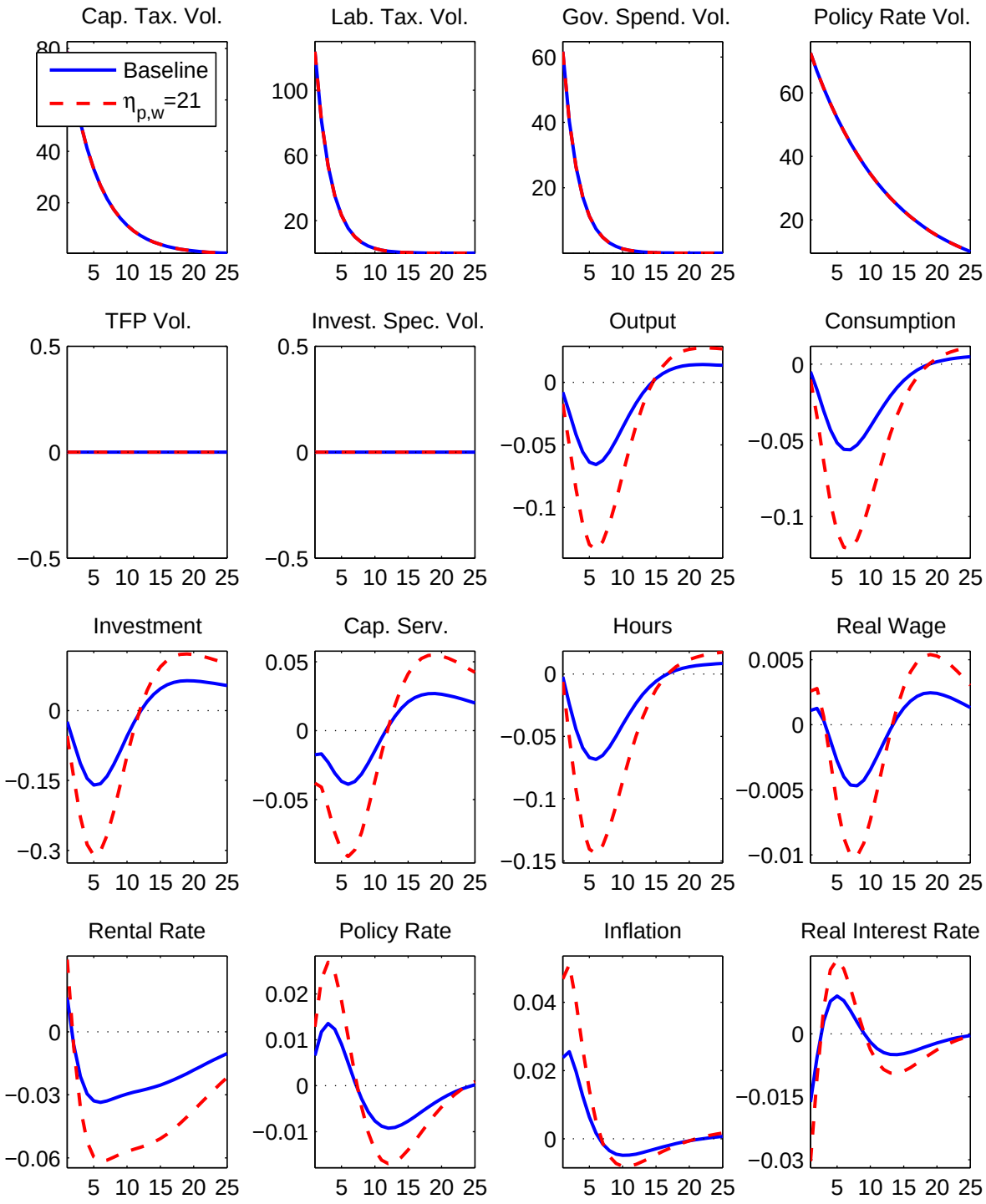


Figure D.12: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration with 10% steady state markups, while the red dashed line indicates a markup of 5%.

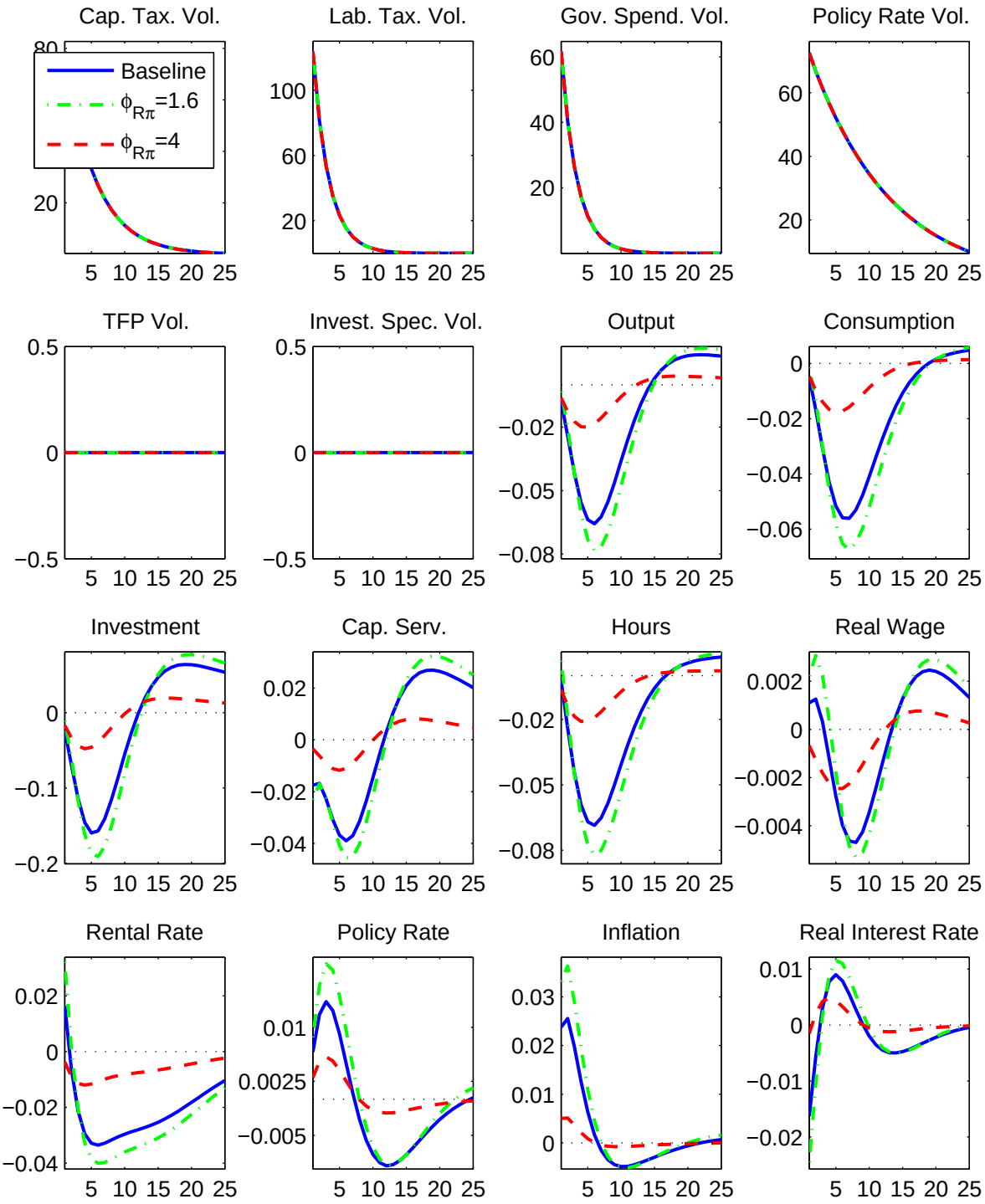


Figure D.13: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed and the green dashed dotted line indicate a lower and higher inflation feedback in the Taylor Rule $\phi_{R\pi}$.

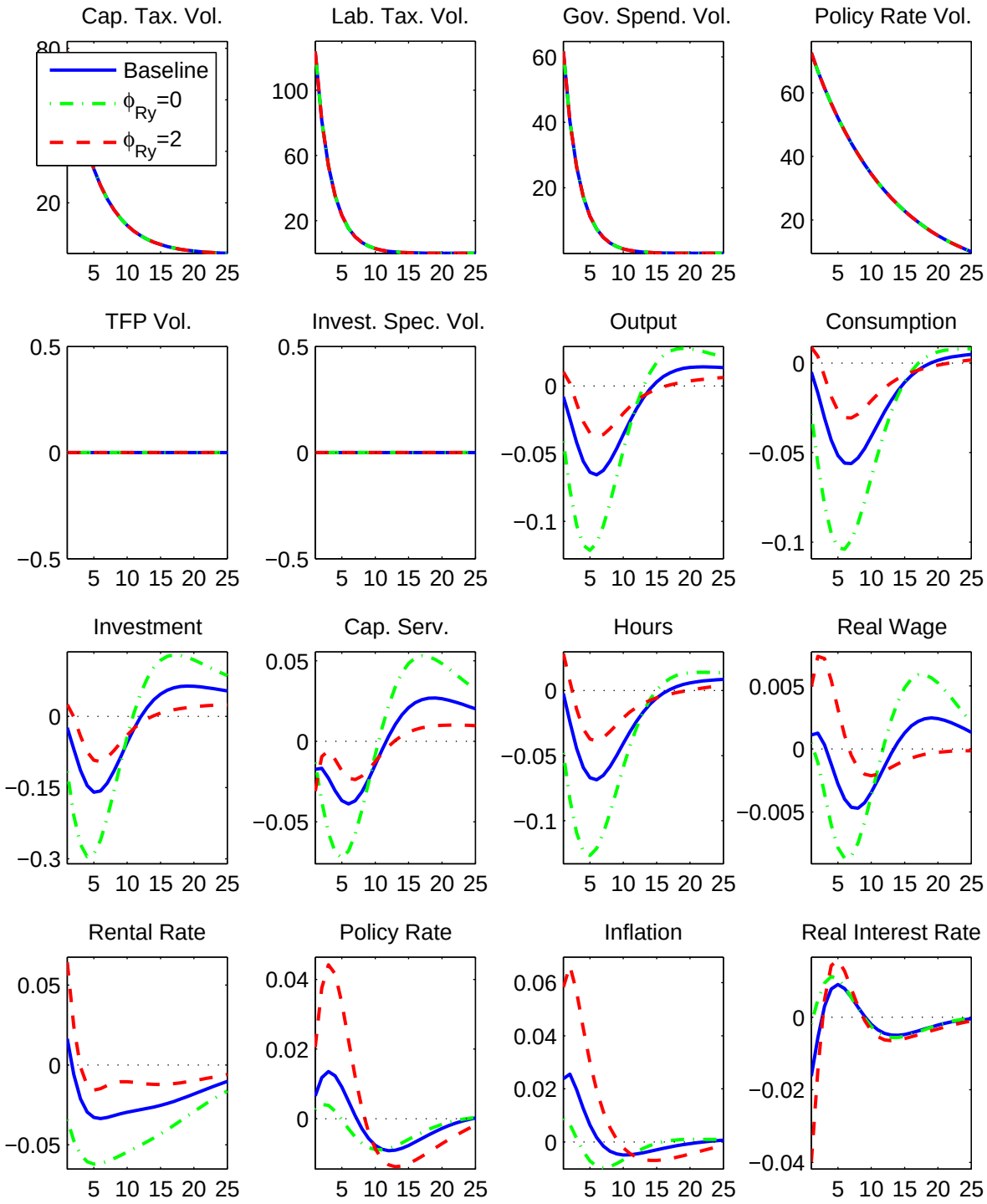


Figure D.14: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed and the green dashed dotted line indicate a lower and higher output feedback in the Taylor Rule ϕ_{Ry}.

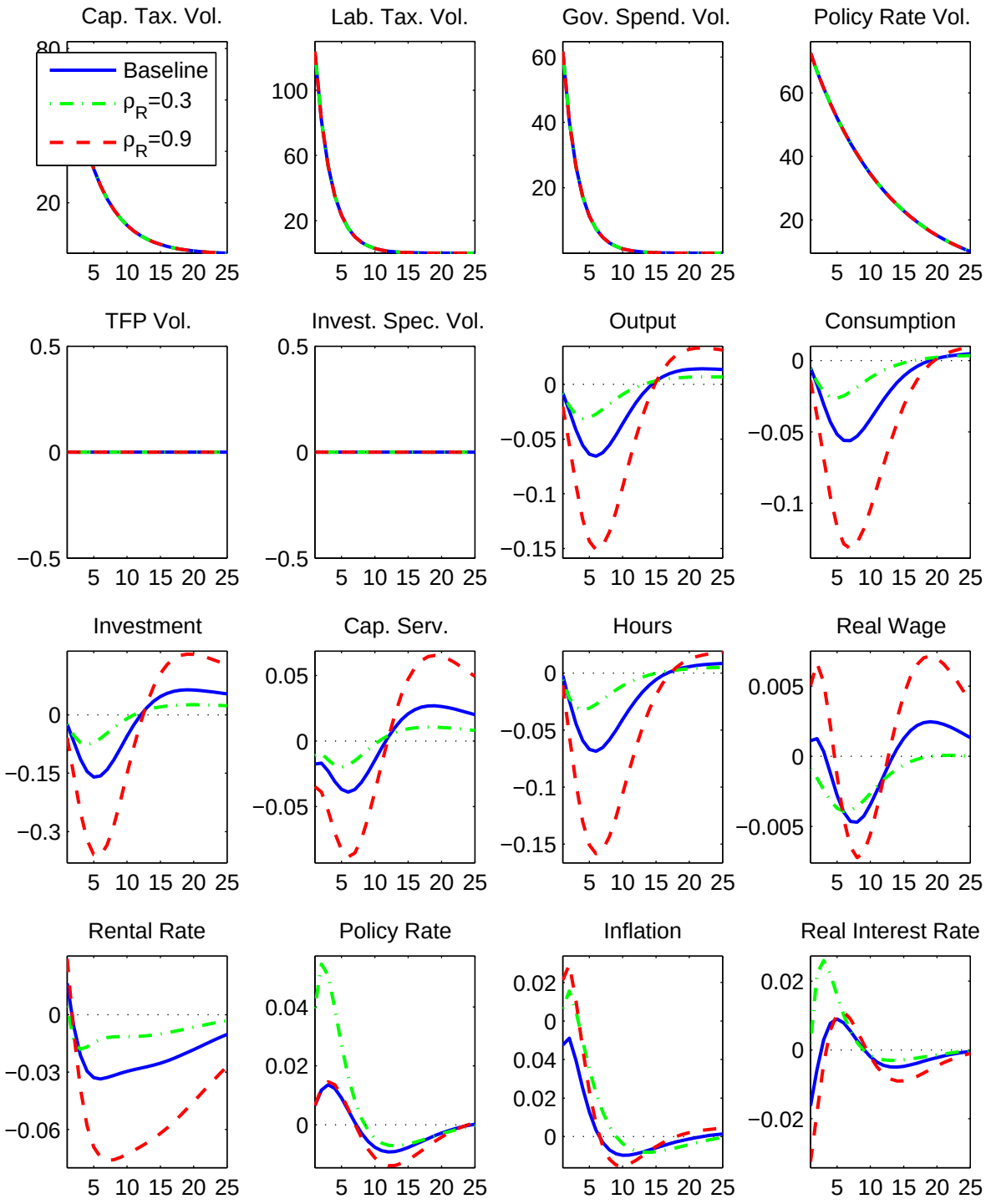


Figure D.15: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed and the green dashed dotted line indicate higher and lower interest rate smoothing.

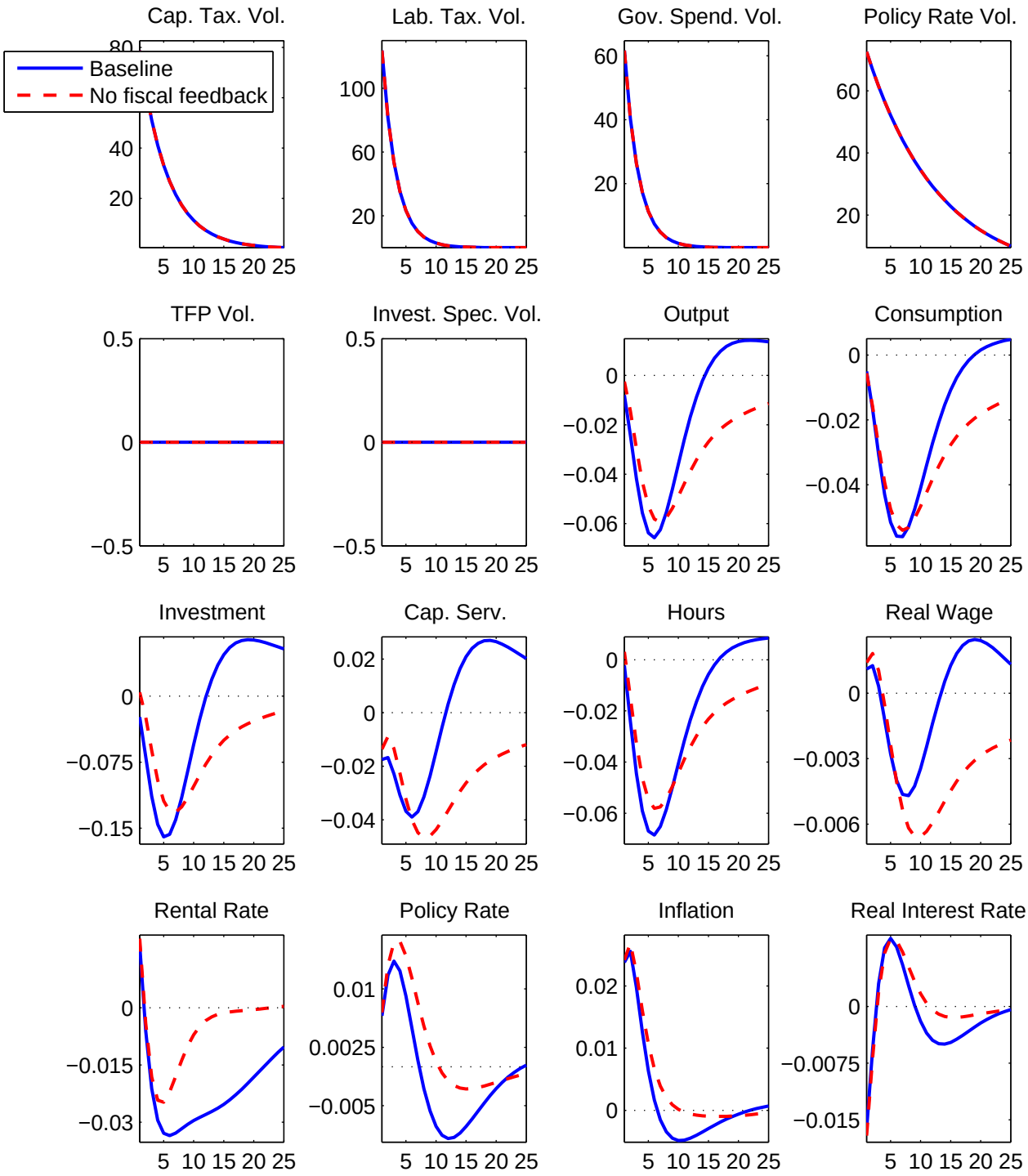


Figure D.16: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed line shows the IRFs when fiscal feedback is shut off.

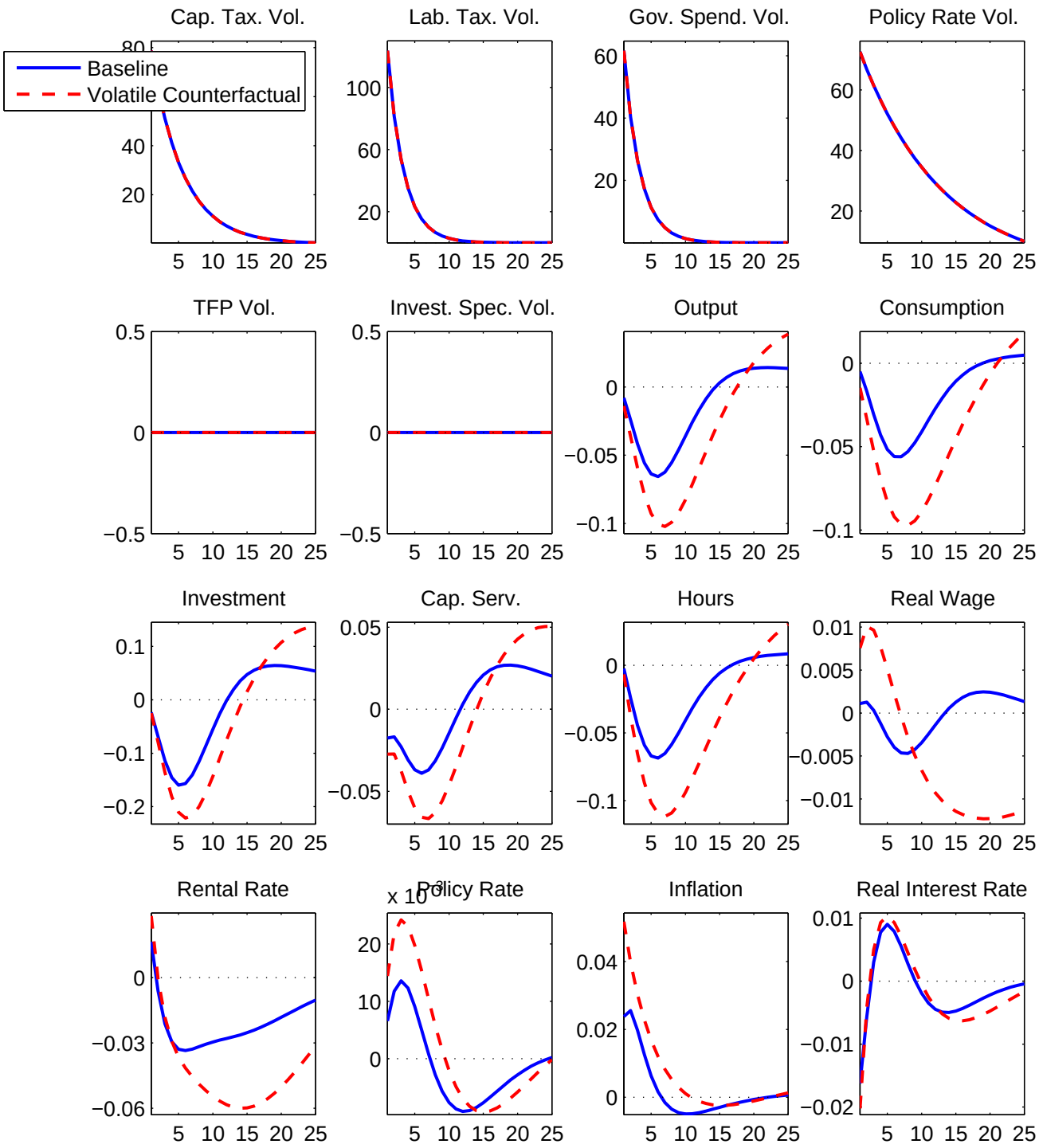


Figure D.17: The figure displays the IRFs to a two-standard deviation joint policy risk shock. The blue solid line is the benchmark calibration, while the red dashed line shows the IRFs using $\delta_2/\delta_1 = 0.001$, $\kappa = 1$, $\chi_p = \chi_w = 0$, $\phi_{Ry} = 0.1$, $\rho_R = 0.9$.

Appendix References

Abramowitz, M. and I. A. Stegun (1965). *Handbook of mathematical functions*. New York: Dover Publications.

An, Sungbae and Frank Schorfheide (2007). "Bayesian analysis of DSGE models". *Econometric Reviews 26* (2-4), 113–172.

Andreasen, Martin M. (2010). "How to maximize the likelihood function for a DSGE model". *Computational Economics 35* (2), 127–154.

Andreasen, Martin M., Jesús Fernández-Villaverde, and Juan F. Rubio-Ramírez (2013). "The pruned state-space system for non-linear DSGE models: theory and empirical applications". NBER Working Papers 18983.

Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp (2002). "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking". *IEEE Transactions on Signal Processing 50*, 174–188.

Basu, Susanto, John G. Fernald, and Miles S. Kimball (2006). "Are technology improvements contractionary?" *American Economic Review 96* (5), 1418–1448.

Binsbergen, Jules H. van, Jesús Fernández-Villaverde, Ralph S.J. Koijen, and Juan F. Rubio-Ramírez (2012). "The term structure of interest rates in a DSGE model with recursive preferences". *Journal of Monetary Economics 59* (7), 634–648.

Born, Benjamin and Johannes Pfeifer (2014). "Risk matters: a comment". CESifo Working Paper Series 4793.

Breusch, Trevor S. and Adrian R. Pagan (1979). "A simple test for heteroscedasticity and random coefficient variation". *Econometrica 47* (5), 1287–1294.

Chib, Siddharta and Srikanth Ramamurthy (2010). "Tailored randomized block MCMC methods with application to DSGE models". *Journal of Econometrics 155* (1), 19–38.

Chib, Siddhartha and Edward Greenberg (1995). "Understanding the Metropolis-Hastings algorithm". *The American Statistician 49* (4), 327–335.

Clarida, Richard, Jordi Galí, and Mark Gertler (2000). "Monetary policy rules and macroeconomic stability: evidence and some theory". *Quarterly Journal of Economics 115* (1), 147–180.

Cochrane, John H. (2005). *Asset pricing*. Revised. Princeton, NJ: Princeton University Press.

Doucet, Arnaud and Adam M. Johansen (2009). "A tutorial on particle filtering and smoothing: fifteen years later". *Handbook of Nonlinear Filtering*.

Duffie, Darrell and Kenneth J. Singleton (1993). "Simulated moments estimation of Markov models of asset prices". *Econometrica 61* (4), 929–952.

Fernald, John (2012). "A quarterly, utilization-adjusted series on total factor productivity". Federal Reserve Bank of San Francisco Working Paper 2012-19.

Fernández-Villaverde, Jesús and Juan F. Rubio-Ramírez (2007). "Estimating macroeconomic models: a likelihood approach". *Review of Economic Studies 74* (4), 1059–1087.

——— (2010). "Macroeconomics and volatility: data, models, and estimation". Mimeo. University of Pennsylvania.

Fernández-Villaverde, Jesús, Pablo A. Guerrón-Quintana, and Juan F. Rubio-Ramírez (2010). "Fortune or virtue: time-variant volatilities versus parameter drifting in U.S. data". NBER Working Papers 15928.

Fernández-Villaverde, Jesús, Pablo A. Guerrón-Quintana, Juan F. Rubio-Ramírez, and Martín Uribe (2011). "Risk matters: the real effects of volatility shocks". *American Economic Review 101* (6), 2530–61.

Fisher, Jonas D.M. (2006). "The dynamic effects of neutral and investment specific technology shocks". *Journal of Political Economy 114* (3), 413–451.

Geweke, John (1992). "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments". *Bayesian Statistics*. Ed. by José M. Bernardo, James O. Berger, A. Philip Dawid, and Adrian F. M. Smith. Vol. 4. Oxford: Clarendon Press, 641–649.

Godsill, Simon J., Arnaud Doucet, and Mike West (2004). "Monte Carlo smoothing for nonlinear time series". *Journal of the American Statistical Association 99* (465), 156–168.

Hansen, Nikolaus, Sybille D. Müller, and Petros Koumoutsakos (2003). "Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)". *Evolutionary Computation 11* (1), 1–18.

Jarque, Carlos M. and Anil K. Bera (1987). "A test for normality of observations and regression residuals". *International Statistical Review / Revue Internationale de Statistique 55* (2), 163–172.

Jones, John Bailey (2002). "Has fiscal policy helped stabilize the postwar U.S. economy?" *Journal of Monetary Economics 49* (4), 709–746.

Justiniano, Alejandro and Giorgio E. Primiceri (2008). "The time-varying volatility of macroeconomic fluctuations". *American Economic Review 98* (3), 604–41.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998). "Stochastic volatility: likelihood inference and comparison with ARCH models". *Review of Economic Studies 65* (3), 361–93.

Koenker, Roger (1981). "A note on studentizing a test for heteroscedasticity". *Journal of Econometrics 17* (1), 107–112.

Kolmogorov, Andrey N. (1933). "Sulla determinazione empirica di una legge di distribuzione". *Giornale dell'Istituto Italiano degli Attuari 4* (1933), 83–91.

Leeper, Eric M., Michael Plante, and Nora Traum (2010). "Dynamics of fiscal financing in the United States". *Journal of Econometrics 156* (2), 304–321.

Mendoza, Enrique G., Assaf Razin, and Linda L. Tesar (1994). "Effective tax rates in macroeconomics: cross-country estimates of tax rates on factor incomes and consumption". *Journal of Monetary Economics 34* (3), 297–323.

Pitt, Michael K. (2002). "Smooth particle filters for likelihood evaluation and maximisation". Warwick Economic Research Papers 651.

Primiceri, Giorgio E. (2005). "Time varying structural vector autoregressions and monetary policy". *Review of Economic Studies 72* (3), 821–852.

Rosenblatt, Murray (1952). "Remarks on a multivariate transformation". *Annals of Mathematical Statistics 23* (3), 470–472.

Ruge-Murcia, Francisco J. (2010). "Estimating nonlinear DSGE models by the simulated method of moments". CIREQ Working Paper Series 19-2010.

——— (2012). "Estimating nonlinear DSGE models by the simulated method of moments: with an application to business cycles". *Journal of Economic Dynamics and Control 36* (6), 914–938.

Schmitt-Grohé, Stephanie and Martín Uribe (2004). "Solving dynamic general equilibrium models using a second-order approximation to the policy function". *Journal of Economic Dynamics and Control 28* (4), 755–775.

——— (2011). "Business cycles with a common trend in neutral and investment-specific productivity". *Review of Economic Dynamics 14* (1), 122–135.

Shapiro, Samuel S. and Martin B. Wilk (1965). "An analysis of variance test for normality (complete samples)". *Biometrika 52* (3/4), 591–611.

Smets, Frank and Rafael Wouters (2007). "Shocks and frictions in US business cycles: a Bayesian DSGE approach". *American Economic Review 97* (3), 586–606.

Smirnov, Nikolai (1948). "Table for estimating the goodness of fit of empirical distributions". *Annals of Mathematical Statistics 19* (2), 279–281.

White, Halbert (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity". *Econometrica 48* (4), 817–838.

Wooldridge, Jeffrey M. (1990). "A unified approach to robust, regression-based specification tests". *Econometric Theory 6* (1), 17–43.