

Simulation study of NKPC estimators

Sophocles Mavroeidis

Mikkel Plagborg-Moller

James H. Stock

January 15, 2013

1 Model and estimators

1.1 Model and data generation

The NKPC is

$$\pi_t = c + \gamma E_t \pi_{t+1} + (1 - \gamma) \pi_{t-1} + \lambda x_t + u_t, \quad (1)$$

or equivalently,

$$(1 - \gamma) \Delta \pi_t = c + \gamma E_t \Delta \pi_{t+1} + \lambda x_t + u_t, \quad (2)$$

so the forcing variable is denoted x_t . Let $z_t = (y_t, x_t)'$, where y_t is either π_t or $\Delta \pi_t$. We impose the reduced-form dynamics

$$a(L)z_t = \tilde{c} + \varepsilon_t + \begin{pmatrix} b'_y \\ 0_{1 \times 2} \end{pmatrix} \varepsilon_{t-1}, \quad a(L) = \sum_{i=1}^q a_i L^i, \quad b_y, \tilde{c} \in \mathbb{R}^2.$$

Here $\varepsilon_t = (\varepsilon_{yt}, \varepsilon_{xt})'$ is a bivariate, mean-zero i.i.d. process. In the simulations it will be normally distributed with covariance matrix Ω . In case of determinacy, the MA(1) parameter b_y is set to zero.

In the following, let $Y_t = (z'_t, z'_{t-1}, \dots, z'_{t-q}, 1)'$. We denote $p = \dim Y_t = 2q + 1$. The reduced forms for y_t and x_t are then written

$$y_t = \zeta' Y_{t-1} + v_{yt}, \quad x_t = \xi' Y_{t-1} + v_{xt}, \quad (3)$$

or for short,

$$Y_t = AY_{t-1} + V_t, \quad (4)$$

where

$$A = \begin{pmatrix} \zeta' \\ \xi' \\ I_{p-3} & 0_{(p-3) \times 2} & 0_{(p-3) \times 1} \\ 0_{1 \times (p-3)} & 0_{1 \times 2} & 1 \end{pmatrix}, \quad V_t = \begin{pmatrix} v_t \\ 0_{(p-2) \times 1} \end{pmatrix} = \epsilon_t + B\epsilon_{t-1},$$

$$B = \begin{pmatrix} b'_y & 0_{1 \times (p-2)} \\ 0_{(p-1) \times 2} & 0_{(p-1) \times (p-2)} \end{pmatrix}, \quad \epsilon_t = \begin{pmatrix} \varepsilon_t \\ 0_{(p-2) \times 1} \end{pmatrix}, \quad v_t = (v_{yt}, v_{xt})'.$$

We denote by e_y , e_x and e_c the p -dimensional canonical basis vectors that are, respectively, the 1st, 2nd and p -th columns of the $p \times p$ identity matrix I_p .

When specifying a DGP, the following choices must be made. First the true values of the structural parameters (γ, λ) are selected, as well as the lag length q and whether the reduced form VARMA is expressed in terms of inflation ($y_t = \pi_t$) or the change in inflation ($y_t = \Delta\pi_t$). All our true DGPs set $c = 0$ and $\tilde{c} = 0$. The reduced form coefficients for x_t are then chosen along with the error covariance matrix Ω . Finally, the AIM algorithm of Anderson and Moore (1985) is used to calculate the implied reduced form coefficients for y_t . In case of indeterminacy (or non-existence of a stable solution), it is necessary to choose which of the Blanchard and Kahn (1980) roots to solve out, and the MA(1) parameters b_y may be set to non-zero values.¹

The MATLAB code generates a sample size of T periods of VARMA data for (y_t, x_t) , using a burn-in time of 100 periods (not counted in T). See Section 3.1 for details on the structure of the code.

1.2 Estimators

We consider four estimators: GIV, VAR-GMM, VAR-MD and VAR-ML. The three first methods are implemented as two-step estimators using optimal weight matrices. The last method uses the reverse-engineering approach of Kurmann (2007). VAR-GMM and VAR-ML require numerical optimization, although for VAR-GMM the optimization is univariate and unproblematic.

We now give a brief description of the estimators. Detailed calculations are relegated to Section 3, where it is also shown that the VAR-GMM, VAR-MD and VAR-ML estimators of (γ, λ, c) coincide for just-identified specifications ($p = 3$). Note that, except for GIV, none of the estimators account for the possibility that $b_y \neq 0$, i.e., the reduced form is assumed by the econometrician to follow a VAR(q) process.²

GIV GIV is based on the linear moment conditions

$$\begin{aligned} 0 &= E[(\pi_t - \pi_{t-1}) - \gamma(\pi_{t+1} - \pi_{t-1}) - \lambda x_t - c]Y_{t-1}] \\ &= E[(\Delta\pi_t - \gamma(\Delta\pi_{t+1} + \Delta\pi_t) - \lambda x_t - c)Y_{t-1}], \end{aligned} \quad (5)$$

where the first expression is used when $y_t = \pi_t$ and the second when $y_t = \Delta\pi_t$. We use two-step efficient GMM with the identity matrix as the first step weight matrix. The long-run variance is estimated using the Newey and West (1987) (centered) HAC estimator with $T = [4(T/100)^{1/4}]$ lags. All calculations are in closed form. For each DGP, we calculate the minimum eigenvalue of the population concentration matrix for GIV, which serves as a measure of the strength of identification. The population concentration matrix is essentially the population projection matrix in the regression of the two endogenous right-hand side variables on Y_{t-1} , scaled by the variance of Y_t and the variance of the regression error.

VAR-GMM VAR-GMM is based on the moment conditions

$$E \left[\begin{pmatrix} \pi_t - (1 - \gamma)\pi_{t-1} - \gamma\zeta'Y_t - \lambda x_t - c \\ \pi_t - \zeta'Y_{t-1} \end{pmatrix} \otimes Y_{t-1} \right] = 0 \quad (6)$$

¹Note that the condition $E_{t-1}u_t = 0$ places restrictions on b_y .

²The MATLAB code also allows the econometrician's perceived VAR to have lag length $\tilde{q}$ that may differ from the true DGP's lag length q .

if $y_t = \pi_t$, and

$$E \left[\begin{pmatrix} (1 - \gamma)\Delta\pi_t - \gamma\zeta'Y_t - \lambda x_t - c \\ \Delta\pi_t - \zeta'Y_{t-1} \end{pmatrix} \otimes Y_{t-1} \right] = 0 \quad (7)$$

if $y_t = \Delta\pi_t$. We estimate $(\gamma, \lambda, c, \zeta)$ using two-step efficient non-linear GMM. Operationally, this is done by minimizing the concentrated GMM objective function in the single parameter γ , where we use that, conditional on this parameter, the moment conditions are linear and so closed-form profile minimizers (λ, c, ζ) are available. Because the numerical optimization is univariate and smooth, it seems to be entirely unproblematic. The first step weight matrix is the identity weight matrix, while the efficient weight matrix is estimated using the Eicker-Huber-White estimator (which is consistent if $b_y = 0$). Standard errors are calculated analytically based on the second-step estimators of $(\gamma, \lambda, c, \zeta)$ and the efficient weight matrix.

VAR-MD The minimum distance estimator is based on the p -dimensional parameter restriction function $h(\gamma, \lambda, c, A)$, where

$$h(\gamma, \lambda, c, A)' = e_y' A - \gamma e_y' A^2 - (1 - \gamma)e_y' - \lambda e_x' A - c e_c' \quad (8)$$

if $y_t = \pi_t$, and

$$h(\gamma, \lambda, c, A)' = (1 - \gamma)e_y' A - \gamma e_y' A^2 - \lambda e_x' A - c e_c' \quad (9)$$

if $y_t = \Delta\pi_t$. The VAR-MD estimator of (γ, λ, c) minimizes $h(\gamma, \lambda, c, \hat{A})' W h(\gamma, \lambda, c, \hat{A})$, where $\hat{A}$ is the OLS estimator $\hat{A} = (\sum_t Y_t Y_{t-1}') (\sum_t Y_{t-1} Y_{t-1}')^{-1}$, and W is a weight matrix. We use two-step efficient MD estimation with an identity weight matrix in the first step. All calculations are in closed form.

VAR-ML As in Kurmann (2007), it may be derived that the joint validity of the structural NKPC and reduced form (4) imply

$$\xi = \frac{1}{\gamma(e_x' \zeta) + \lambda} [(I_p - \gamma \tilde{A})' \zeta - (1 - \gamma)e_y - c e_c] \quad (10)$$

if $y_t = \pi_t$, and

$$\xi = \frac{1}{\gamma(e_x' \zeta) + \lambda} [((1 - \gamma)I_p - \gamma \tilde{A})' \zeta - c e_c] \quad (11)$$

if $y_t = \Delta\pi_t$. Here $\tilde{A}$ equals the VAR(1) companion matrix A in (4) with the second row replaced with zeros, so $A = \tilde{A} + e_x \xi'$. Hence, given $\theta = (\gamma, \lambda, c, \zeta)' \in \mathbb{R}^{3+p}$, we can compute the implied ξ and thus the implied reduced-form VAR coefficient matrices. This in turn yields the implied reduced-form residuals $v_t(\theta)$. Concentrating out the residual covariance matrix Ω , the log likelihood (imposing homoskedastic, i.i.d. normally distributed errors) may be written

$$l(\theta) = -\frac{T}{2} \log |\Omega(\theta)|, \quad \Omega(\theta) = \frac{1}{T} \sum_t v_t(\theta) v_t(\theta)'$$

We calculate the MLE of θ by numerically maximizing this concentrated log likelihood. The optimizer is provided with analytical first derivatives. As is clear from formulas (10) and (11) above, the concentrated log likelihood is smooth on $\mathbb{R}^{3+p} \setminus \tilde{\Theta}$, where $\tilde{\Theta} = \{\theta: \gamma(e_x' \zeta) + \lambda = 0\}$. To better identify the global maximum, we run several separate numerical optimizations. Half of the

optimizations impose the constraint $\gamma(e'_x\zeta) + \lambda \geq 0$, while the other half imposes $\gamma(e'_x\zeta) + \lambda \leq 0$, and the initial values are determined by a combination of grid search and stochastic perturbations. The resulting restricted optima for each half space and initial values are then compared to determine the (hopefully) global MLE. Additionally, we impose lower and upper bounds on the numerically admissible values of γ and λ . ML standard errors are computed *analytically* using the outer product of the scores (not the observed information). Because the automated ML optimization routine that we implement is not perfect, our particular results should be interpreted with caution, although they do strongly underscore the need to carefully inspect the likelihood surface in empirical work.

1.3 DGPs

Here we describe the specific DGPs that we provide results for in Section 2. We set $T = 200$ throughout; the Newey-West HAC estimator for GIV then uses 4 lags. The reduced form is written in terms of raw inflation ($y_t = \pi_t$), not inflation changes, and the econometrician knows the true lag length. The DGPs that we simulate are listed in Table 1. All these DGPs set $\Omega = ((0.07, 0.03)', (0.03, 0.70)')$, which is a typical estimate of the disturbance covariance matrix in reduced-form OLS regressions on the 1960–2011 sample, using most major inflation series with the labor share as forcing variable. In addition, we set $c = 0$ and $\tilde{c} = (0, 0)'$. The VAR-ML optimization routine is restricted to search over the parameter subspace $\{(\gamma, \lambda, c, \zeta) : \gamma \in [-0.5, 2], \lambda \in [-0.3, 0.3]\}$, and we compare optimization results corresponding to 8 different initial parameter values per Monte Carlo repetition.

DGP 1a has $\gamma = 0.7$ and $\lambda = 0.03$. The reduced form is a VAR(1). The coefficients in the forcing variable equation ξ are set to typical empirical OLS estimates for the NFB log labor share on the 1960–2011 sample. These parameter choices lead to indeterminacy (all Blanchard and Kahn (1980) roots are below unity in modulus), so we make the AIM routine solve out the third-largest root.³ The econometrician is assumed to correctly use one lag each of inflation and the forcing variable as instruments (along with a constant), i.e., $Y_t = (\pi_t, x_t, 1)'$ and $p = 3$. The minimum eigenvalue of the population concentration matrix is 2.42, well below 10. Hence, identification is fairly weak but certainly not more so than what has been reported in the literature (Mavroeidis, 2004; Kleibergen and Mavroeidis, 2009).

DGP 1b is identical to *1a*, except that an extra (irrelevant) lag of inflation and the forcing variable is used in the instrument set, so $Y_t = (\pi_t, x_t, \pi_{t-1}, x_{t-1}, 1)'$ and $p = 5$. *DGP 1c* has a VAR(2) reduced form, with the reduced-form coefficient vector ξ for the forcing variable equation approximately equal to typical OLS estimates for the labor share on the 1960–2011 sample. We still make AIM solve out the third-largest Blanchard-Kahn root. *DGP 1d* is identical to *1c*, except that we set $\lambda = 0.01$, leading to weaker identification. *DGP 1e* is a modification of *1c*, in which the reduced form coefficients for the forcing variable have been set to empirically unrealistic values that give much stronger identification (the minimum eigenvalue of the population concentration matrix is 108.40), while keeping the largest eigenvalue of the VAR companion matrix safely away from the unit circle.

DGP 1f is a variant of *1c* in which the MA parameter b_y is non-zero. We have set

$$\pi_t = \frac{1}{\gamma}\pi_{t-1} - \frac{1-\gamma}{\gamma}\pi_{t-2} - \frac{\lambda}{\gamma}x_{t-1} + v_{yt},$$

³The other two Blanchard-Kahn roots are complex conjugates of each other and so cannot be solved out separately.

where $v_{yt} = \varepsilon_{yt} + b_y' \varepsilon_{t-1}$. It is easy to verify that this inflation process satisfies the structural NKPC (1) with $u_t = -\gamma b_y' \varepsilon_t$. We choose $b_y = (-1.4, -0.3)'$ to obtain stark results. Note that *DGP 1f* is much more strongly identified than *1c*.

DGPs 2a–e are analogs of *1a–e* in which we set $\gamma = 0.3$, a lower value. We could not find any DGPs with reasonable reduced-form parameter ξ such that both $\lambda > 0$ and a stable RE solution exists. Consequently, we set $\lambda = -0.03$ for *DGPs 2a–c* and *2e*, while $\lambda = -0.01$ for *DGP 2d*. All the RE solutions are determinate (exactly one Blanchard-Kahn root is above unity in modulus). For *DGP 2e*, we pick reduced-form parameters ξ such that the minimum eigenvalue of the population concentration matrix is moderately large.⁴

Of the above DGPs of family *1* and *2*, we consider the *c* and *d* variants to be the most empirically relevant in terms of the reduced-form dynamics they imply. Moving from *d* to *c* to *e* gives a progression from virtual under-identification to weak identification to strong identification. The largest eigenvalue (in terms of absolute value) of the VAR companion matrix A is in the range 0.95–0.97 for all our DGPs, so the dynamics are quite persistent.

2 Simulation results

All the following results are based on 10,000 Monte Carlo repetitions per DGP. Summary statistics for the simulated sampling distributions of the various estimators of γ and λ are listed in Tables 2–3. We also list the rejection frequency (RF) of overidentification tests corresponding to each of the four estimators. For GIV and VAR-GMM, this is the Hansen (1982) J test, for VAR-MD it is the usual MD overidentification test, and for VAR-ML it is the likelihood ratio test against an unrestricted VAR(q). Additionally, the tables show the RF of the weak identification robust S test (Stock and Wright, 2000) corresponding to the GIV specification. Figures 1–11 show kernel-smoothed density plots comparing the sampling distributions of the γ estimators. As remarked above, the VAR-GMM/MD/ML estimates perfectly coincide for *DGPs 1a* and *2a*, so the three estimators are jointly referred to as “VAR methods” in these parametrizations.

3 Detailed documentation

The following calculations are meant to serve as documentation for the formulas used in the MATLAB code. Let $\mathbb{I}$ denote the indicator variable that is 1 when $y_t = \Delta\pi_t$ and 0 if $y_t = \pi_t$.

3.1 MATLAB code

The simulations are executed in MATLAB. The following files are used (each file has been extensively annotated).

- `run_mc_var.m`: The main executable where settings and DGPs are specified.
- `disp_output_var.m`: Displays summary statistics and plots after running the simulations.
- `drawdens.m`: Calculates kernel-smoothed densities of the sampling distributions.

⁴We were unable to find a value for ξ that leads to even stronger identification without also moving the largest eigenvalue of the VAR companion matrix very close to the unit circle.

- `gmm.m`: General routine for calculating linear system GMM estimates. Used for the GIV and VAR-GMM estimators (see below).
- `md.m`: Calculates VAR-MD estimates of the NKPC (see below).
- `ml_con.m`: Auxiliary function that returns the non-linear constraint used in the VAR-ML optimization (see below).
- `ml_kurmann.m`: Log likelihood function for VAR-ML à la Kurmann (2007) (see below).
- `nws.m`: General routine for Newey and West (1987) HAC estimation.
- `observ1.m`: Invokes the AIM routine of Anderson and Moore (1985).

In addition to the above, the code uses a very slight modification of Jeff Fuhrer's MATLAB implementation of the AIM procedure. The sole modification is that we allow for solving out arbitrary Blanchard and Kahn (1980) roots, whereas the original code always solves out the roots with the largest moduli.

3.2 GIV

Concentration matrix We wish to express the two endogenous regressors $q_t = (\pi_{t+1} - \pi_{t-1}, x_t)'$ as a function of the instruments Y_{t-1} and an error term. Suppose first that $y_t = \pi_t$. Then

$$\begin{aligned}
q_t &= (e_y, 0)'Y_{t+1} + (0, e_x)'Y_t - (e_y, 0)'Y_{t-1} \\
&= (e_y, 0)'(AY_t + V_{t+1}) + (0, e_x)'Y_t - (e_y, 0)'Y_{t-1} \\
&= \{(e_y, 0)'A + (0, e_x)'\} Y_t - (e_y, 0)'Y_{t-1} + (e_y, 0)'V_{t+1} \\
&= \{(e_y, 0)'A + (0, e_x)'\} (AY_{t-1} + V_t) - (e_y, 0)'Y_{t-1} + (e_y, 0)'V_{t+1} \\
&= DY_{t-1} + w_t,
\end{aligned}$$

where $D = CA - (e_y, 0)'$, $C = (e_y, 0)'A + (0, e_x)'$ and $w_t = (e_y, 0)'V_{t+1} + CV_t$. Note that when we write $(e_y, 0)$, the zero vector has dimension p .

In the case where $y_t = \Delta\pi_t$, similar calculations lead to

$$\begin{aligned}
q_t &= \begin{pmatrix} \Delta\pi_{t+1} + \Delta\pi_t \\ x_t \end{pmatrix} \\
&= (e_y, 0)'Y_{t+1} + (e_y, e_x)'Y_t \\
&= DY_{t-1} + w_t,
\end{aligned}$$

with $D = CA$, $C = (e_y, 0)'A + (e_y, e_x)'$ and $w_t = (e_y, 0)'V_{t+1} + CV_t$. Observe that the only differences between the present differenced case and the non-differenced case above is in the formulas for the matrices C and D .

By the law of motion $Y_t = AY_{t-1} + \epsilon_t + B\epsilon_{t-1}$, we have

$$\begin{aligned}
V(Y_t) &= AV(Y_{t-1})A' + AE(Y_{t-1}\epsilon'_{t-1})B' + BE(\epsilon_{t-1}Y'_{t-1})A' + V(\epsilon_t) + BV(\epsilon_{t-1})B' \\
&= AV(Y_{t-1})A' + A\tilde{\Omega}B' + B\tilde{\Omega}A' + \tilde{\Omega} + B\tilde{\Omega}B',
\end{aligned}$$

where $\tilde{\Omega} \equiv V(\epsilon_t)$ (a $p \times p$ matrix of zeros, except for the upper left 2×2 block which is Ω). Hence, under stationarity of Y_t ,

$$Q \equiv V(Y_t) = (I_{p^2} - A \otimes A)^{-1} \text{vec}(\tilde{\Omega} + A\tilde{\Omega}B' + B\tilde{\Omega}A' + B\tilde{\Omega}B').$$

Observe also that

$$\begin{aligned} V(w_t) &= V[(e_y, 0)'(\epsilon_{t+1} + B\epsilon_t) + C(\epsilon_t + B\epsilon_{t-1})] \\ &= (e_y, 0)' \tilde{\Omega} (e_y, 0) + [(e_y, 0)'B + C] \tilde{\Omega} [(e_y, 0)'B + C]' + CB\tilde{\Omega}B'C', \end{aligned}$$

and

$$E(w_t Y_{t-1}') = CE(V_t Y_{t-1}') = CB\tilde{\Omega}.$$

The population projection of q_t on Y_{t-1} has coefficient matrix

$$M \equiv E(q_t Y_{t-1}') Q^{-1} = D + E(w_t Y_{t-1}') Q^{-1} = D + CB\tilde{\Omega}Q^{-1}.$$

The population projection residual is then

$$e_t \equiv q_t - MY_{t-1} = -CB\tilde{\Omega}Q^{-1}Y_{t-1} + w_t,$$

with variance

$$\begin{aligned} \Sigma &\equiv V(e_t) \\ &= CB\tilde{\Omega}Q^{-1}QQ^{-1}\tilde{\Omega}B'C' - CB\tilde{\Omega}Q^{-1}E(Y_{t-1}w_t') - E(w_t Y_{t-1}')Q^{-1}\tilde{\Omega}B'C' + V(w_t) \\ &= V(w_t) - CB\tilde{\Omega}Q^{-1}\tilde{\Omega}B'C'. \end{aligned}$$

We now have all the ingredients to compute the population concentration matrix, which is defined as

$$K \equiv T\Sigma^{-1/2}MQM'\Sigma^{-1/2'}, \quad \Sigma^{-1/2}\Sigma^{-1/2'} = \Sigma^{-1}.$$

Probability limit of OLS Under weak identification, GIV is biased toward OLS. We here calculate the probability limit of the OLS estimator in the regression of $\Delta\pi_t$ on $(\pi_{t+1} - \pi_{t-1})$ and x_t (i.e., the regression corresponding to the GIV specification). Let q_t and Q be defined as above. The probability limit of the OLS estimator is given by $[V(q_t)]^{-1}E(q_t\Delta\pi_t)$. Define $Q_1 = E(Y_t Y_{t-1}') = AQ + B\tilde{\Omega}$ and $Q_2 = E(Y_t Y_{t-2}') = AQ_1$.

If $y_t = \pi_t$, we showed above that $q_t = \Xi(Y_{t+1}', Y_t', Y_{t-1}')'$, where $\Xi = ((e_y, 0)', (0, e_x)', -(e_y, 0)')$. We have

$$V \begin{pmatrix} Y_{t+1} \\ Y_t \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} Q & Q_1 & Q_2 \\ Q_1' & Q & Q_1 \\ Q_2' & Q_1' & Q \end{pmatrix}.$$

We can then calculate $V(q_t) = \Xi V[(Y_{t+1}', Y_t', Y_{t-1}')'] \Xi'$. Furthermore,

$$E(q_t \Delta\pi_t) = E[q_t(Y_t - Y_{t-1})' e_y] = \Xi \begin{pmatrix} Q_1 - Q_2 \\ Q - Q_1 \\ Q_1' - Q \end{pmatrix} e_y.$$

If instead $y_t = \Delta\pi_t$, we have $q_t = \Xi(Y_{t+1}', Y_t')'$, where $\Xi = ((e_y, 0)', (e_y, e_x)')$. Hence, $V(q_t) = \Xi V[(Y_{t+1}', Y_t')'] \Xi'$ and $E(q_t \Delta\pi_t) = \Xi(Q_1', Q)' e_y$.

3.3 VAR-GMM

Standard errors Let the moment function be

$$h_t(\theta) = \begin{pmatrix} (1 - \mathbb{I}\gamma)y_t - (1 - \mathbb{I})(1 - \gamma)y_{t-1} - \gamma\zeta'Y_t - \lambda x_t - c \\ y_t - \zeta'Y_{t-1} \end{pmatrix} \otimes Y_{t-1},$$

where $\theta = (\gamma, \lambda, c, \zeta')'$. To estimate the asymptotic variance of the VAR-GMM estimator, we need an estimate of the derivative of the moment function. We have

$$\begin{aligned} \frac{\partial h_t(\theta)}{\partial \theta'} &= \left(\frac{\partial h_t(\theta)}{\partial \gamma}, \frac{\partial h_t(\theta)}{\partial \lambda}, \frac{\partial h_t(\theta)}{\partial c}, \frac{\partial h_t(\theta)}{\partial \zeta'} \right) \\ &= \left(\frac{(1 - \mathbb{I})y_{t-1} - \mathbb{I}y_t - \zeta'Y_t}{0} \mid \frac{-x_t}{0} \mid \frac{-1}{0} \mid \frac{-\gamma Y_t'}{-Y_{t-1}'} \right) \otimes Y_{t-1}. \end{aligned}$$

Numerical optimization The estimation of θ is carried out by noting that, conditional on γ , the moment conditions are linear in the other parameters (λ, c, ζ) . Hence, the concentrated GMM objective function is available in closed form using standard formulas, and we need only optimize over the scalar parameter γ . In the simulations, the resulting concentrated objective seems to be smooth and nicely convex, so the numerical optimization is fast and unproblematic.

3.4 VAR-MD

Optimal weight matrix and standard errors In contrast to Sbordone (2005), our parameter restrictions (6)–(7) are linear in $\theta = (\gamma, \lambda, c)'$ conditional on A , so the estimator is available in closed form.⁵ We can write

$$h(\gamma, \lambda, c, A)' = (1 - \mathbb{I}\gamma)e_y'A - (1 - \mathbb{I})(1 - \gamma)e_y' - \gamma e_y'A^2 - \lambda e_x'A - c e_c', \quad (12)$$

so

$$h(\gamma, \lambda, c, A) = D(A)\theta + C(A),$$

where

$$D(A) = [(1 - \mathbb{I})I_p - \mathbb{I}A - A^2]'e_y, -A'e_x, -e_c, \quad C(A) = [A - (1 - \mathbb{I})I_p]'e_y.$$

Given the OLS estimate $\hat{A}$ of A , the VAR-MD estimator $\hat{\theta}$ minimizes $h(\theta, \hat{A})'Wh(\theta, \hat{A})$ for a choice of weight matrix W . Standard calculations give

$$\hat{\theta} = -[D(\hat{A})'WD(\hat{A})]^{-1}D(\hat{A})'WC(\hat{A}).$$

To derive the form of the optimal weight matrix, and to calculate analytical standard errors, we need the derivatives $\partial h/\partial \theta'$ and $\partial h/\partial \text{vec}(A)'$. The former derivative is given by $D(A)$. To calculate the latter, we first note that

$$\frac{\partial \text{vec}(X^2)}{\partial \text{vec}(X)'} = (I_p \otimes X) + (X' \otimes I_p)$$

⁵We achieve linearity by keeping the NKPC in difference equation format instead of iterating it forward. See Barnes et al. (2011) for a critique of this approach.

for any unrestricted $p \times p$ dimensional matrix X (Petersen and Pedersen, 2012). We write

$$h = [((1 - \mathbb{I}\gamma)e_y - \lambda e_x)' \otimes I_p] \text{vec}(A') - \gamma(e_y' \otimes I_p) \text{vec}((A')^2) - (1 - \mathbb{I})(1 - \gamma)e_y - ce_c.$$

Then

$$\frac{\partial h}{\partial \text{vec}(A')'} = [((1 - \mathbb{I}\gamma)e_y - \lambda e_x)' \otimes I_p] - \gamma(e_y' \otimes I_p)[(I_p \otimes A') + (A \otimes I_p)].$$

The final piece in the puzzle for calculating the efficient MD weight matrix is to obtain the asymptotic variance of $\text{vec}(\hat{A}')$. But the asymptotic variance of $\text{vec}(\hat{A})$ is $[E(Y_t Y_t')]^{-1} \otimes \Omega$, so the asymptotic variance of $\text{vec}(\hat{A}')$ may be obtained by permuting the elements of this matrix.

3.5 VAR-ML

Reverse engineering à la Kurmann (2007) First suppose that $y_t = \pi_t$. The NKPC (1) and reduced form (3) for y_t imply

$$\pi_t = c + \gamma(\zeta' Y_t) + (1 - \gamma)\pi_{t-1} + \lambda x_t + u_t.$$

If we again use (3) on the left-hand side, and use the companion form representation (4) on the right-hand side, we obtain

$$\zeta' Y_{t-1} + v_{yt} = ce_c' Y_{t-1} + \gamma \zeta' (A Y_{t-1} + V_t) + (1 - \gamma)e_y' Y_{t-1} + \lambda(\zeta' Y_{t-1} + v_{xt}) + u_t,$$

or

$$[\zeta'(I_p - \gamma A) - \lambda \xi' - (1 - \gamma)e_y' - ce_c'] Y_{t-1} + [v_{yt} - \lambda v_{xt} - \gamma \zeta' V_t - u_t] = 0.$$

If we (perhaps mistakenly) assume that $E[V_t | Y_{t-1}] = E[u_t | Y_{t-1}] = 0$, then the expressions in the two square brackets above must vanish, so in particular

$$\zeta'(I_p - \gamma A) - \lambda \xi' - (1 - \gamma)e_y' - ce_c' = 0_{1 \times p}.$$

Let $\tilde{A}$ be defined as in Section 1.2. Then $A = \tilde{A} + e_x \xi'$, with $\tilde{A}$ independent of ξ , and we can solve for ξ in the equation above to obtain the formula (10) for ξ in terms of $(\gamma, \lambda, c, \zeta)$.

The calculations for the case $y_t = \Delta \pi_t$ are very similar and thus omitted.

Standard errors Let $v_t(\theta)$ be the first two elements of $V_t(\theta) \equiv Y_t - A(\theta)Y_{t-1}$, where $A(\theta) = \tilde{A}(\zeta) + e_x \xi(\theta)'$. Here $\xi(\theta)$ denotes the expression for ξ in terms of θ derived above. Also,

$$\tilde{A}(\zeta) = \tilde{\tilde{A}} + e_y \zeta', \quad \tilde{\tilde{A}} = \begin{pmatrix} & 0_{2 \times p} & \\ I_{p-3} & 0_{(p-3) \times 2} & 0_{(p-3) \times 1} \\ 0_{1 \times (p-3)} & 0_{1 \times 2} & 1 \end{pmatrix}.$$

The per-observation log likelihood is

$$l_t(\theta, \Omega) = \text{constant} - \frac{1}{2} \log |\Omega| - \frac{1}{2} v_t(\theta)' \Omega^{-1} v_t(\theta).$$

The asymptotic variance of the MLE (under correct specification) is $(E[s_t(\theta, \Omega) s_t(\theta, \Omega)'])^{-1}$, where the per-observation score vector is $s_t(\theta, \Omega) = (\partial l_t(\theta, \Omega) / \partial \theta', \partial l_t(\theta, \Omega) / \partial \text{vech}(\Omega)')'$. We estimate this as $[T^{-1} \sum_t s_t(\hat{\theta}, \Omega(\hat{\theta})) s_t(\hat{\theta}, \Omega(\hat{\theta}))']^{-1}$.

Let $E_{yx} = (e_y, e_x)'$, so $v_t(\theta) = E_{yx}(Y_t - A(\theta)Y_{t-1})$. Using formulas from Petersen and Pedersen (2012), we find

$$\frac{\partial l_t(\theta, \Omega)}{\partial \theta'} = \text{vec}(\Omega^{-1}v_t(\theta)Y_{t-1}')' \frac{\partial \text{vec}(E_{yx}A(\theta))}{\partial \theta'}.$$

Let T_p be the $p^2 \times p^2$ matrix of zeros and ones such that $\text{vec}(X') = T_p \text{vec}(X)$ for any $p \times p$ matrix X . Then

$$\begin{aligned} \frac{\partial l_t(\theta, \Omega)}{\partial \theta'} &= \text{vec}(Y_{t-1}v_t(\theta)'\Omega^{-1})' T_p' T_p \frac{\partial \text{vec}\{(E_{yx}A(\theta))'\}}{\partial \theta'} \\ &= \text{vec}(Y_{t-1}v_t(\theta)'\Omega^{-1})' \frac{\partial \text{vec}\{(E_{yx}A(\theta))'\}}{\partial \theta'}, \end{aligned} \quad (13)$$

since $T_p' T_p = I_{p^2}$. Observe that

$$E_{yx}A(\theta) = E_{yx}e_y\zeta' + E_{yx}e_x\xi(\theta)' = \tilde{e}_y\zeta' + \tilde{e}_x\xi(\theta)',$$

where $\tilde{e}_y = (1, 0)'$ and $\tilde{e}_x = (0, 1)'$. Thus,

$$\frac{\partial \text{vec}\{(E_{yx}A(\theta))'\}}{\partial \theta'} = (0_{(2p) \times 3}, \tilde{e}_y \otimes I_p) + (\tilde{e}_x \otimes I_p) \frac{\partial \xi(\theta)}{\partial \theta'}.$$

We write

$$\xi(\theta) = \frac{[(1 - \gamma\mathbb{I})I_p - \gamma\tilde{A}(\zeta)]'\zeta - (1 - \mathbb{I})(1 - \gamma)e_y - ce_c}{\gamma(e'_x\zeta) + \lambda} \equiv \frac{\text{numer}(\theta)}{\text{denom}(\theta)},$$

so

$$\frac{\partial \xi(\theta)}{\partial \theta'} = \frac{1}{\text{denom}(\theta)} \frac{\partial \text{numer}(\theta)}{\partial \theta'} - \frac{\text{numer}(\theta)}{(\text{denom}(\theta))^2} \frac{\partial \text{denom}(\theta)}{\partial \theta'}.$$

We have

$$\frac{\partial \text{denom}(\theta)}{\partial \theta'} = (e'_x\zeta, 1, 0, \gamma e'_x),$$

and

$$\frac{\partial \text{numer}(\theta)}{\partial \theta'} = \left(-(\mathbb{I}I_p + \tilde{A}(\zeta))'\zeta + (1 - \mathbb{I})e_y, 0_{p \times 1}, -e_c, \frac{\partial[(1 - \gamma\mathbb{I})I_p - \gamma\tilde{A}(\zeta)]'\zeta}{\partial \zeta'} \right),$$

where

$$\frac{\partial[(1 - \gamma\mathbb{I})I_p - \gamma\tilde{A}(\zeta)]'\zeta}{\partial \zeta'} = (1 - \gamma\mathbb{I})I_p - \gamma\tilde{A}' - \gamma \frac{\partial(\zeta e'_y\zeta)}{\partial \zeta'}, \quad \frac{\partial(\zeta e'_y\zeta)}{\partial \zeta'} = I_p(e'_y\zeta) + \zeta e'_y.$$

To derive $\partial l_t(\theta, \Omega)/\partial \text{vech}(\Omega)'$, we first find the corresponding “scalar-over-matrix” derivative for which the elements of Ω are treated as variation free (i.e., without imposing that Ω is symmetric). For lack of better notation, we call this derivative $\partial l_t(\theta, \Omega)/\partial \Omega \in \mathbb{R}^{2 \times 2}$. We have

$$\begin{aligned} \frac{\partial l_t(\theta, \Omega)}{\partial \Omega} &= -\frac{1}{2}\Omega^{-1} + \frac{1}{2}\Omega^{-1}v_t(\theta)v_t(\theta)'\Omega^{-1} \\ &= \frac{1}{2}\Omega^{-1}[v_t(\theta)v_t(\theta)' - \Omega]\Omega^{-1}. \end{aligned}$$

According to Petersen and Pedersen (2012), we then obtain

$$\begin{aligned} \frac{\partial l_t(\theta, \Omega)}{\partial \text{vech}(\Omega)'} &= \text{vech} \left\{ \frac{\partial l_t(\theta, \Omega)}{\partial \Omega} + \frac{\partial l_t(\theta, \Omega)}{\partial \Omega'} - \text{diag} \left(\frac{\partial l_t(\theta, \Omega)}{\partial \Omega} \right) \right\}' \\ &= \text{vech} \left\{ \Omega^{-1}[v_t(\theta)v_t(\theta)' - \Omega]\Omega^{-1} - \frac{1}{2} \text{diag} (\Omega^{-1}[v_t(\theta)v_t(\theta)' - \Omega]\Omega^{-1}) \right\}'. \end{aligned}$$

Concentrated log likelihood and its derivatives As used by Kurmann (2007), the concentrated log likelihood function in θ (i.e., concentrating out $\text{vech}(\Omega)$), is

$$l(\theta) = \text{constant} - \frac{T}{2} \log |\Omega(\theta)|, \quad \Omega(\theta) = \frac{1}{T} \sum_{t=1}^T v_t(\theta) v_t(\theta)'.$$

By the envelope theorem, the gradient of $l(\theta) = \sum_t l_t(\theta, \Omega(\theta))$ is given by

$$\sum_t \frac{\partial l_t(\theta, \Omega(\theta))}{\partial \theta'} = \text{vec} \left(\left\{ \sum_t Y_{t-1} v_t(\theta)' \right\} \Omega(\theta)^{-1} \right)' \frac{\partial \text{vec} \{ (E_{yx} A(\theta))' \}}{\partial \theta'}, \quad (14)$$

using (13), where $\partial \text{vec} \{ (E_{yx} A(\theta))' \} / \partial \theta'$ was computed above.

Numerical optimization The concentrated log likelihood function is infinitely smooth everywhere except on the hyperplane $\tilde{\Theta} = \{\theta: \gamma(e'_x \zeta) + \lambda = 0\}$. For fixed (λ, ζ) , it appears that the concentrated log likelihood tends to negative infinity as $\gamma \rightarrow -\lambda/(e'_x \zeta)$ from the left or the right. To improve the numerical accuracy, we run two separate types of optimizations: one on the half-space $\gamma(e'_x \zeta) + \lambda \geq 0$ and another on the half-space $\gamma(e'_x \zeta) + \lambda \leq 0$. Initial parameter values on each half space are determined by a grid search over (γ, λ) values, conditional on $c = 0, \zeta = \hat{\zeta}^{\text{OLS}}$. In addition to using the grid-determined restricted MLEs as initial values, we generate other initial values on each half space by stochastically perturbing the grid-MLEs with mean-zero Gaussian noise. We run a constrained optimization for each thus generated initial value of the parameter vector, after which the various optimized log likelihood values are compared. Optimization is carried out by the MATLAB routine `fmincon`, which is supplied with analytical first derivatives of the concentrated log likelihood (as derived above) and the constraint function. We also restrict the allowable values of (γ, λ) to a rectangle in $\mathbb{R}^2$.

The performance of the numerical optimizer is decent. It seems to do reasonably well for strongly identified specifications, since the sampling distribution of the VAR-ML estimator is typically Gaussian and close to the sampling distribution for VAR-GMM and VAR-MD, as we would expect. If the true VAR lag length is $q = 1$, we know that VAR-ML should give the exact same (γ, λ, c) estimates as VAR-GMM and VAR-MD, cf. Section 3.6. For sufficiently weak identification, our VAR-ML optimization routine does not yield exactly the same results as VAR-GMM and VAR-MD (the latter two do coincide) when $q = 1$, meaning that our procedure sometimes fails to find the global maximum of the likelihood. For example, for *DGPs 1a* and *2b* in Table 1, the numerically computed MLE differs from the VAR-GMM/MD estimate in about 15% of the simulations, although the kernel density estimates of the sampling distributions are graphically very similar.⁶ We have spent considerable effort trying to improve the VAR-ML optimization, but it seems difficult to come up with a better *automated* routine.⁷ The main problem seems to be sensitivity to initial values, rather than a failure to converge locally. This underscores the need for researchers to carefully inspect the likelihood surface when using VAR-ML to estimate the NKPC.

⁶The estimates are said to differ if the γ estimates are more than 0.01 apart or the λ estimates are more than 0.001 apart.

⁷The simulated annealing algorithm advocated by Kurmann (2007) requires extensive case-specific fine-tuning and so is not well suited for our general Monte Carlo framework.

3.6 Just-identified specifications

If the specification is just-identified (i.e., $p = 3$), then the VAR-GMM, VAR-MD and VAR-ML estimators of (γ, λ, c) coincide. To see this, note first from the moment conditions (6)–(7) that when $p = 3$, VAR-GMM has $\hat{\zeta}^{\text{VAR}} = \hat{\zeta}^{\text{OLS}} = e'_y S_1 S_0^{-1}$, where $S_0 = \sum_t Y_{t-1} Y'_{t-1}$ and $S_1 = \sum_t Y_t Y'_{t-1}$. The VAR-GMM estimators of (γ, λ, c) satisfy the first p sample moment conditions exactly, i.e.,

$$\sum_t \left[(1 - \hat{\gamma}^{\text{VAR}} \mathbb{I}) y_t - (1 - \mathbb{I})(1 - \hat{\gamma}^{\text{VAR}}) y_{t-1} - \hat{\gamma}^{\text{VAR}} \hat{\zeta}^{\text{VAR}'} Y_t - \hat{\lambda}^{\text{VAR}} x_t - \hat{c}^{\text{VAR}} \right] Y_{t-1} = 0,$$

or

$$(1 - \hat{\gamma}^{\text{VAR}} \mathbb{I}) e'_y S_1 - (1 - \mathbb{I})(1 - \hat{\gamma}^{\text{VAR}}) e'_y S_0 - \hat{\gamma}^{\text{VAR}} \hat{\zeta}^{\text{VAR}'} S_1 - \hat{\lambda}^{\text{VAR}} e'_x S_1 - \hat{c}^{\text{VAR}} e'_c S_0 = 0.$$

Because $S_1 = \hat{A}^{\text{OLS}} S_0$ and $\hat{\zeta}^{\text{VAR}} = \hat{A}^{\text{OLS}'} e_y$, the left-hand side above may equivalently be written

$$\left\{ (1 - \hat{\gamma}^{\text{VAR}} \mathbb{I}) e'_y \hat{A}^{\text{OLS}} - (1 - \mathbb{I})(1 - \hat{\gamma}^{\text{VAR}}) e'_y - \hat{\gamma}^{\text{VAR}} e'_y (\hat{A}^{\text{OLS}})^2 - \hat{\lambda}^{\text{VAR}} e'_x \hat{A}^{\text{OLS}} - \hat{c}^{\text{VAR}} e'_c \right\} S_0,$$

so

$$(1 - \hat{\gamma}^{\text{VAR}} \mathbb{I}) e'_y \hat{A}^{\text{OLS}} - (1 - \mathbb{I})(1 - \hat{\gamma}^{\text{VAR}}) e'_y - \hat{\gamma}^{\text{VAR}} e'_y (\hat{A}^{\text{OLS}})^2 - \hat{\lambda}^{\text{VAR}} e'_x \hat{A}^{\text{OLS}} - \hat{c}^{\text{VAR}} e'_c = 0. \quad (15)$$

But this is exactly the equations determining the VAR-MD estimators of (γ, λ, c) , cf. (12).

To show that the VAR-MLE also equals these values, we show that if $\theta = \hat{\theta}^{\text{VAR}}$, the first-order conditions for maximization of the concentrated VAR-ML log likelihood function are satisfied. These FOCs are

$$\text{vec} \left(\left\{ \sum_t Y_{t-1} (Y_t - A(\theta) Y_{t-1})' \right\} E'_{yx} \Omega(\theta)^{-1} \right)' \frac{\partial \text{vec} \{ (E_{yx} A(\theta))' \}}{\partial \theta'} = 0,$$

cf. (14). Note that if $A(\hat{\theta}^{\text{VAR}}) = \hat{A}^{\text{OLS}}$, then the sum in the first curly braces above vanishes, so the FOCs will be satisfied for $\theta = \hat{\theta}^{\text{VAR}}$. We now show that $A(\hat{\theta}^{\text{VAR}}) = \hat{A}^{\text{OLS}}$ by demonstrating that $\xi(\hat{\theta}^{\text{VAR}}) = \hat{\xi}^{\text{OLS}} = \hat{A}^{\text{OLS}'} e_x$.

Using $\tilde{A}(\hat{\zeta}^{\text{OLS}}) = \hat{A}^{\text{OLS}} - e_x \hat{\xi}^{\text{OLS}'}$, where $\hat{\zeta}^{\text{OLS}} = \hat{A}^{\text{OLS}'} e_y$, we find

$$\begin{aligned} \xi(\hat{\theta}^{\text{VAR}}) &= \frac{[(1 - \hat{\gamma}^{\text{VAR}} \mathbb{I}) I_p - \hat{\gamma}^{\text{VAR}} \tilde{A}(\hat{\zeta}^{\text{OLS}})]' \hat{\zeta}^{\text{OLS}} - (1 - \mathbb{I})(1 - \hat{\gamma}^{\text{VAR}}) e_y - \hat{c}^{\text{VAR}} e_c}{\hat{\gamma}^{\text{VAR}} (e'_x \hat{\zeta}^{\text{OLS}}) + \hat{\lambda}^{\text{VAR}}} \\ &= \frac{(1 - \hat{\gamma}^{\text{VAR}} \mathbb{I}) \hat{A}^{\text{OLS}'} e_y - (1 - \mathbb{I})(1 - \hat{\gamma}^{\text{VAR}}) e_y - \hat{\gamma}^{\text{VAR}} (\hat{A}^{\text{OLS}'})^2 e_y - \hat{\lambda}^{\text{VAR}} \hat{A}^{\text{OLS}'} e_x - \hat{c}^{\text{VAR}} e_c}{\hat{\gamma}^{\text{VAR}} (e'_x \hat{\zeta}^{\text{OLS}}) + \hat{\lambda}^{\text{VAR}}} \\ &\quad + \frac{\hat{\gamma}^{\text{VAR}} \hat{\xi}^{\text{OLS}} e'_x \hat{\zeta}^{\text{OLS}} + \hat{\lambda}^{\text{VAR}} \hat{\xi}^{\text{OLS}}}{\hat{\gamma}^{\text{VAR}} (e'_x \hat{\zeta}^{\text{OLS}}) + \hat{\lambda}^{\text{VAR}}}. \end{aligned}$$

When $p = 3$, the numerator in the first fraction vanishes according to equation (15). Hence,

$$\xi(\hat{\theta}^{\text{VAR}}) = \frac{\hat{\gamma}^{\text{VAR}} (e'_x \hat{\zeta}^{\text{OLS}}) + \hat{\lambda}^{\text{VAR}}}{\hat{\gamma}^{\text{VAR}} (e'_x \hat{\zeta}^{\text{OLS}}) + \hat{\lambda}^{\text{VAR}}} \hat{\xi}^{\text{OLS}} = \hat{\xi}^{\text{OLS}},$$

as needed.

References

- Anderson, G. and Moore, G. (1985), ‘A linear algebraic procedure for solving linear perfect foresight models’, *Economics Letters* **17**(3), 247–252.
- Barnes, M. L., Gumbau-Brisa, F., Lie, D. and Olivei, G. P. (2011), ‘Estimation of Forward-Looking Relationships in Closed Form: An Application to the New Keynesian Phillips Curve’, *Federal Reserve Bank of Boston Working Papers* (11-3). June version.
- Blanchard, O. J. and Kahn, C. M. (1980), ‘The Solution of Linear Difference Models under Rational Expectations’, *Econometrica* **48**(5), 1305–1311.
- Hansen, L. P. (1982), ‘Large Sample Properties of Generalized Method of Moments Estimators’, *Econometrica* **50**(4), 1029–1054.
- Kleibergen, F. and Mavroeidis, S. (2009), ‘Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve’, *Journal of Business & Economic Statistics* **27**(3), 293–311.
- Kurmann, A. (2007), ‘VAR-based estimation of Euler equations with an application to New Keynesian pricing’, *Journal of Economic Dynamics and Control* **31**(3), 767–796.
- Mavroeidis, S. (2004), ‘Weak Identification of Forward-looking Models in Monetary Economics’, *Oxford Bulletin of Economics and Statistics* **66**(supplement), 609–635.
- Newey, W. K. and West, K. D. (1987), ‘A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix’, *Econometrica* **55**(3), 703–708.
- Petersen, K. B. and Pedersen, M. S. (2012), ‘The Matrix Cookbook’.
- Sbordone, A. M. (2005), ‘Do expected future marginal costs drive inflation dynamics?’, *Journal of Monetary Economics* **52**(6), 1183–1197.
- Stock, J. H. and Wright, J. (2000), ‘GMM with Weak Identification’, *Econometrica* **68**(5), 1055–1096.

DGP	NKPC		Red. form: forcing var.				Red. form: inflation				MA				plim OLS	
	γ	λ	ξ_{y1}	ξ_{x1}	ξ_{y2}	ξ_{x2}	ζ_{y1}	ζ_{x1}	ζ_{y2}	ζ_{x2}	b'_y	p	LE	Conc.	γ	λ
<i>1a</i>	0.7	0.03	0.10	0.90	—	—	1.01	-0.08	—	—	—	3	0.96	2.42	0.53	0.01
<i>1b</i>	0.7	0.03	0.10	0.90	0.00	0.00	1.01	-0.08	0.00	0.00	—	5	0.96	2.42	0.53	0.01
<i>1c</i>	0.7	0.03	0.20	0.80	-0.10	0.10	1.00	-0.07	0.01	-0.01	—	5	0.96	1.83	0.53	0.01
<i>1d</i>	0.7	0.01	0.20	0.80	-0.10	0.10	1.00	-0.02	0.00	0.00	—	5	0.96	0.12	0.51	0.00
<i>1e</i>	0.7	0.03	0.10	0.70	0.10	-0.10	0.79	0.25	0.05	-0.05	—	5	0.96	108.40	0.52	-0.01
<i>1f</i>	0.7	0.03	0.20	0.80	-0.10	0.10	1.43	-0.05	-0.43	0.00	$(-1.4, -0.3)$	5	0.96	88.69	0.54	0.02
<i>2a</i>	0.3	-0.03	0.10	0.90	—	—	0.99	-0.06	—	—	—	3	0.95	0.89	0.52	0.01
<i>2b</i>	0.3	-0.03	0.10	0.90	0.00	0.00	0.99	-0.06	0.00	0.00	—	5	0.95	0.89	0.52	0.01
<i>2c</i>	0.3	-0.03	0.20	0.80	-0.10	0.10	0.98	-0.06	0.01	-0.01	—	5	0.95	0.67	0.52	0.01
<i>2d</i>	0.3	-0.01	0.20	0.80	-0.10	0.10	0.99	-0.02	0.00	0.00	—	5	0.96	0.06	0.50	0.00
<i>2e</i>	0.3	-0.03	0.70	1.05	0.08	-0.08	0.91	-0.07	-0.01	0.01	—	5	0.97	30.08	0.49	0.00

Table 1: List of our DGPs. Columns 2 and 3 list the true values of (γ, λ) . The subscripts on the reduced-form coefficient vectors ζ and ξ signify the time series (y_t or x_t) and lag (first or second) they correspond to. For example, ζ_{x2} is the coefficient on x_{t-2} in the reduced-form equation for y_t . The last six columns list the MA(1) parameter b_y (if non-zero), the number of instruments p , the largest eigenvalue modulus of the reduced-form VAR companion matrix A (excluding the trivial eigenvalue 1 associated with the constant $\tilde{c}$), the minimum eigenvalue of the population concentration matrix for GIV (cf. Section 3.2), and the probability limits of the OLS estimators of γ and λ (cf. Section 3.2). Numbers in columns 8–11 and 14–17 have been rounded off to 2 decimal points.

Estimator	γ estimates			λ estimates			RF	
	med.	5–95	RF t	med.	5–95	RF t	Overid.	S test
<i>DGP 1a</i>								
GIV	0.658	2.110	0.043	0.024	0.366	0.043	—	0.017
VAR methods	0.666	2.143	0.113	0.025	0.365	0.111	—	
<i>DGP 1b</i>								
GIV	0.618	0.796	0.072	0.017	0.137	0.071	0.053	0.041
VAR-GMM	0.813	0.755	0.533	0.048	0.135	0.506	0.292	
VAR-MD	0.960	0.522	0.790	0.075	0.093	0.750	0.049	
VAR-ML	0.904	0.960	0.591	0.063	0.171	0.548	0.006	
<i>DGP 1c</i>								
GIV	0.606	0.822	0.084	0.015	0.143	0.079	0.047	0.042
VAR-GMM	0.799	0.795	0.544	0.047	0.143	0.523	0.353	
VAR-MD	0.979	0.605	0.753	0.077	0.109	0.722	0.053	
VAR-ML	0.939	1.277	0.524	0.069	0.230	0.497	0.010	
<i>DGP 1d</i>								
GIV	0.500	0.790	0.201	−0.001	0.053	0.113	0.037	0.047
VAR-GMM	0.586	0.775	0.817	0.004	0.059	0.601	0.361	
VAR-MD	0.954	0.715	0.835	0.026	0.057	0.481	0.070	
VAR-ML	0.975	0.928	0.815	0.023	0.074	0.497	0.013	
<i>DGP 1e</i>								
GIV	0.670	0.195	0.128	0.040	0.067	0.060	0.063	0.077
VAR-GMM	0.675	0.193	0.113	0.039	0.064	0.055	0.068	
VAR-MD	0.675	0.191	0.107	0.039	0.064	0.049	0.066	
VAR-ML	0.676	0.198	0.090	0.042	0.068	0.044	0.057	
<i>DGP 1f</i>								
GIV	0.696	0.255	0.057	0.030	0.066	0.042	0.067	0.047
VAR-GMM	1.056	0.452	0.944	0.126	0.148	0.924	0.378	
VAR-MD	1.052	0.385	0.989	0.126	0.133	0.965	0.336	
VAR-ML	1.096	0.378	0.993	0.134	0.128	0.952	0.299	

Table 2: Simulation results for *DGP*s 1a–f based on 10,000 Monte Carlo repetitions per *DGP*. Columns 2 and 3 list the median and 5–95% quantile range of the $\hat{\gamma}$ estimates, while column 4 gives the rejection frequency (RF) of the t -test of the null hypothesis $H_0: \gamma = \gamma_0$ (γ_0 being the true value) at a 5% nominal significance level. Columns 5–7 show the corresponding figures for $\hat{\lambda}$. Columns 8 and 9 list, respectively, the RF of the test of overidentifying restrictions and (for GIV) the S test. For just-identified specifications the VAR-GMM, VAR-MD and VAR-ML estimates coincide and are therefore referred to jointly as “VAR methods” (the t tests use VAR-MD standard errors).

Estimator	γ estimates			λ estimates			RF	
	med.	5–95	RF t	med.	5–95	RF t	Overid.	S test
<i>DGP 2a</i>								
GIV	0.415	1.958	0.093	−0.015	0.278	0.079	—	0.028
VAR methods	0.407	2.100	0.234	−0.017	0.295	0.214	—	
<i>DGP 2b</i>								
GIV	0.449	0.781	0.157	−0.011	0.112	0.136	0.045	0.048
VAR-GMM	0.456	0.821	0.580	−0.010	0.121	0.550	0.214	
VAR-MD	0.924	0.749	0.896	0.054	0.110	0.884	0.140	
VAR-ML	0.498	1.043	0.631	−0.006	0.152	0.608	0.009	
<i>DGP 2c</i>								
GIV	0.463	0.830	0.163	−0.009	0.120	0.143	0.044	0.052
VAR-GMM	0.469	0.875	0.625	−0.008	0.130	0.593	0.251	
VAR-MD	0.952	0.789	0.905	0.059	0.119	0.890	0.126	
VAR-ML	0.522	1.265	0.625	−0.003	0.191	0.601	0.011	
<i>DGP 2d</i>								
GIV	0.493	0.705	0.240	−0.001	0.043	0.101	0.040	0.061
VAR-GMM	0.519	0.771	0.885	−0.000	0.053	0.788	0.378	
VAR-MD	0.931	0.710	0.963	0.020	0.051	0.897	0.104	
VAR-ML	0.940	0.913	0.870	0.016	0.066	0.762	0.017	
<i>DGP 2e</i>								
GIV	0.310	0.218	0.052	−0.029	0.031	0.042	0.067	0.046
VAR-GMM	0.314	0.207	0.104	−0.029	0.029	0.088	0.096	
VAR-MD	0.323	0.489	0.212	−0.028	0.068	0.197	0.171	
VAR-ML	0.304	0.225	0.075	−0.030	0.031	0.063	0.057	

Table 3: Simulations results for *DGP*s 2a–e. See the caption under Table 2 for details.

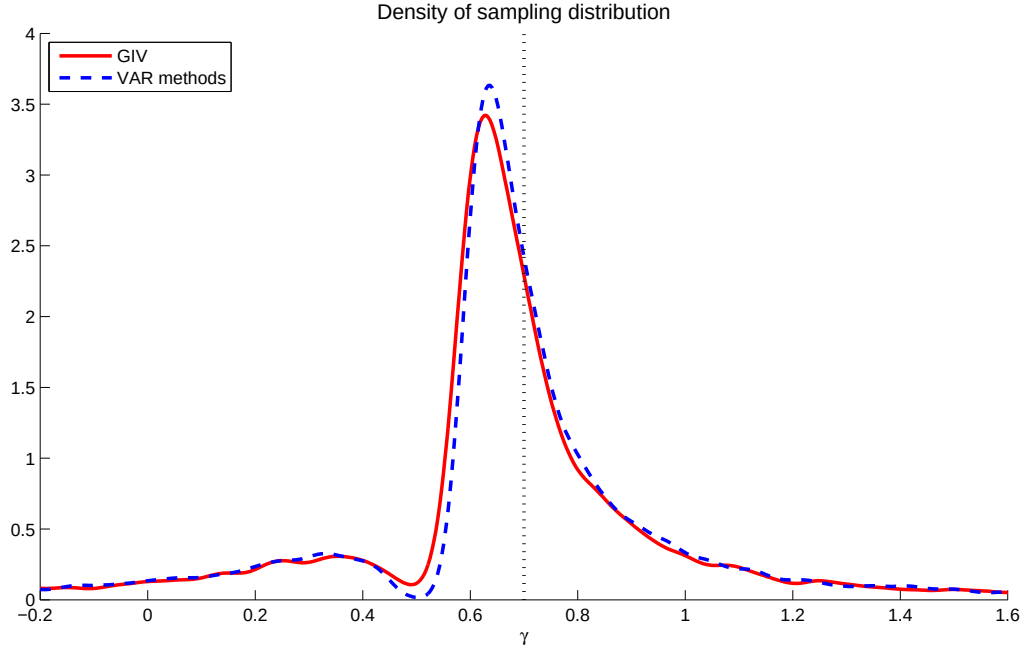


Figure 1: Sampling distribution of γ estimators, *DGP 1a*. The density is estimated using MATLAB's kernel smoother `ksdensity` over 500 grid points with default settings.

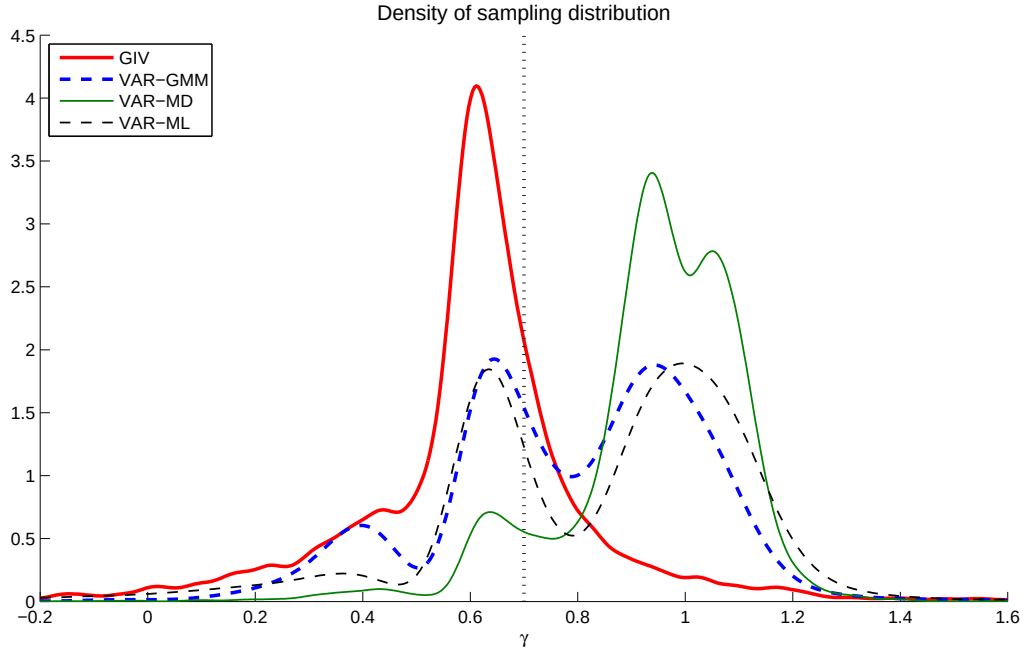


Figure 2: Sampling distribution of γ estimators, *DGP 1b*.

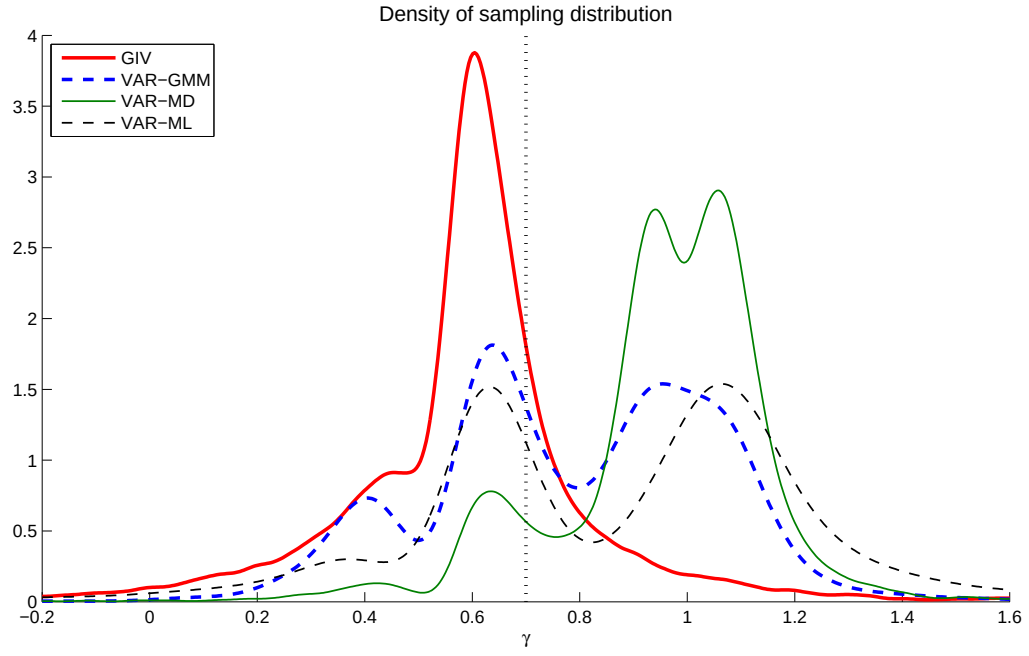


Figure 3: Sampling distribution of γ estimators, *DGP 1c*.

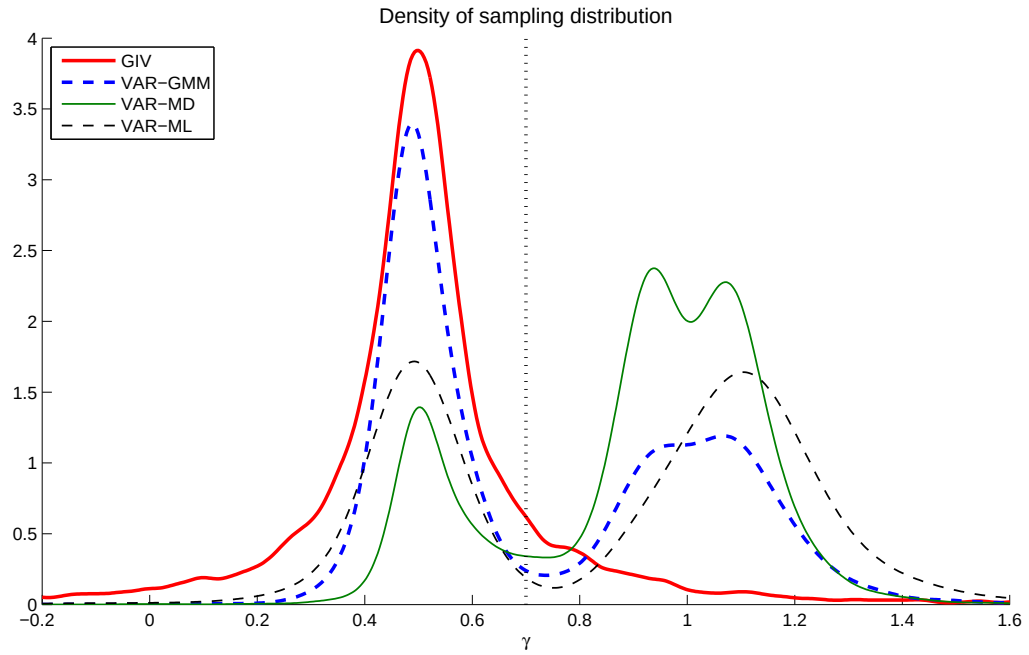


Figure 4: Sampling distribution of γ estimators, *DGP 1d*.

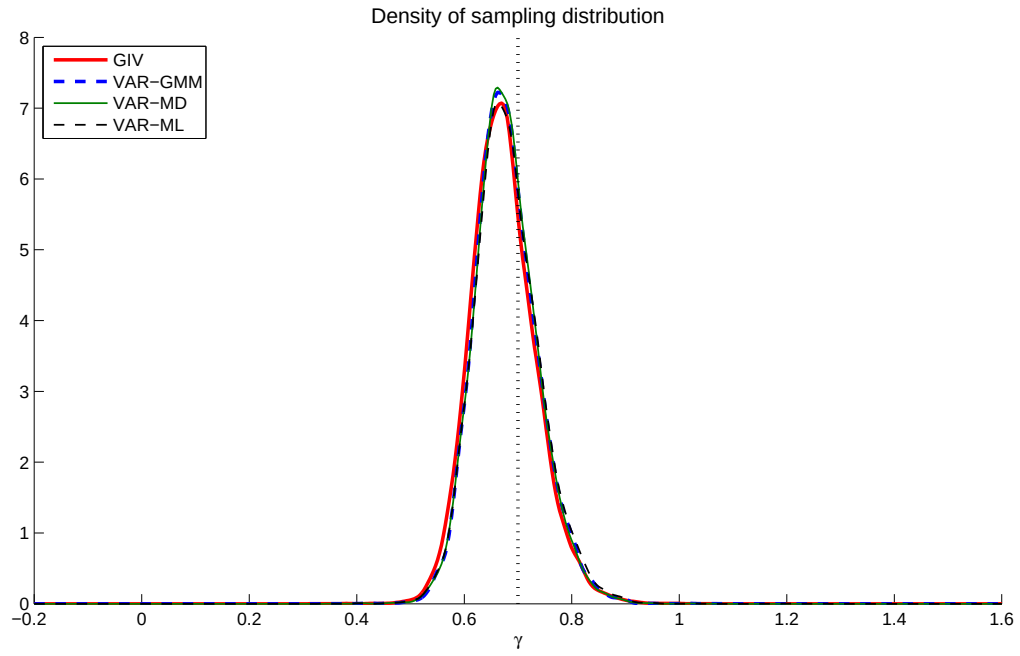


Figure 5: Sampling distribution of γ estimators, *DGP 1e*.

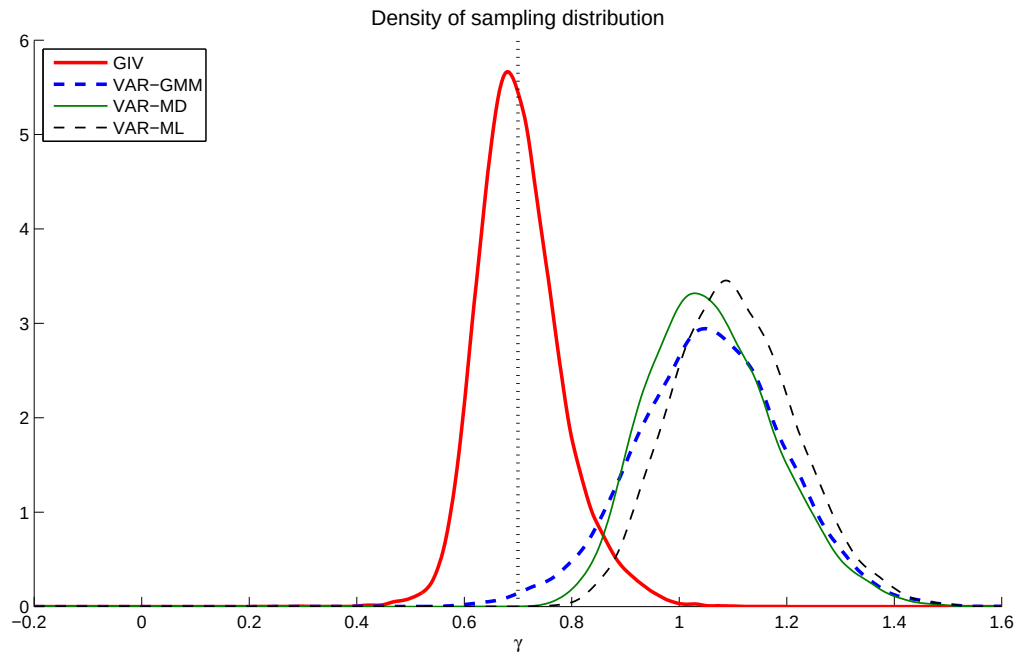


Figure 6: Sampling distribution of γ estimators, *DGP 1f*.

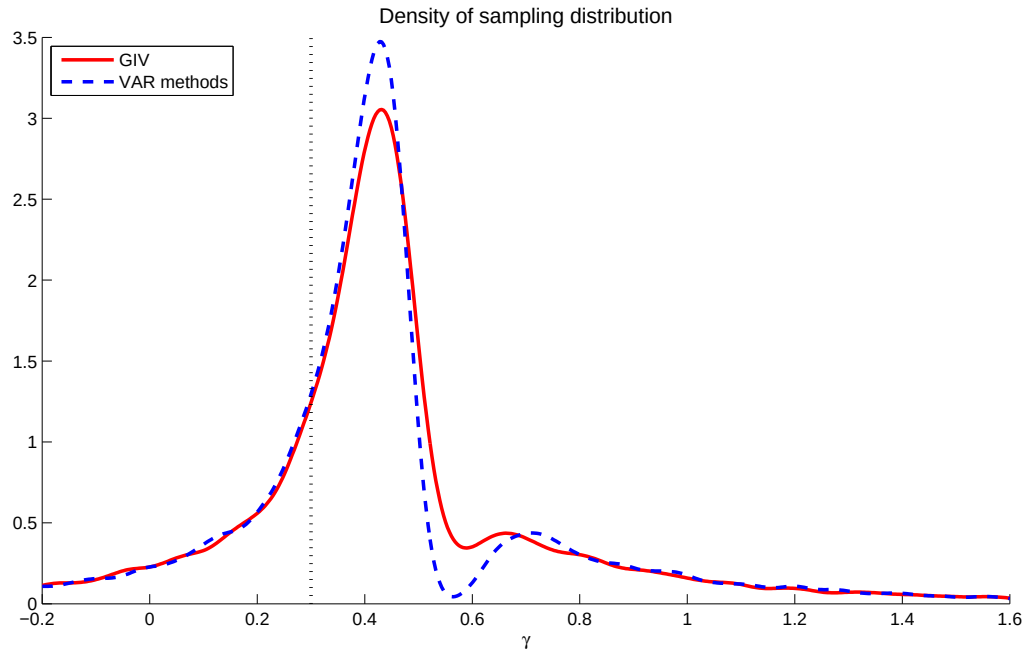


Figure 7: Sampling distribution of γ estimators, *DGP 2a*.

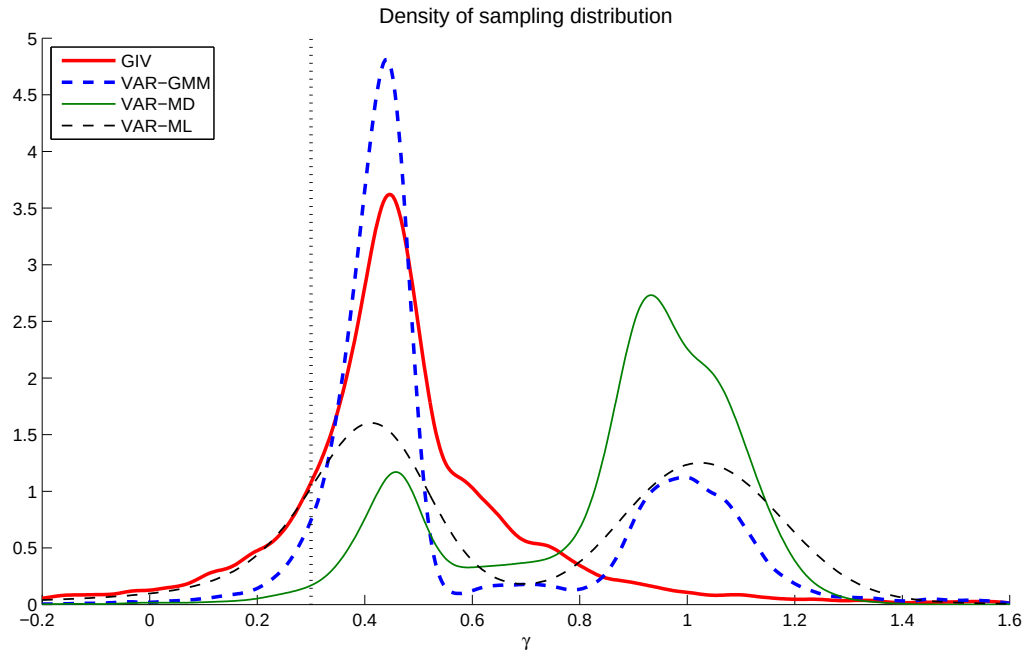


Figure 8: Sampling distribution of γ estimators, *DGP 2b*.

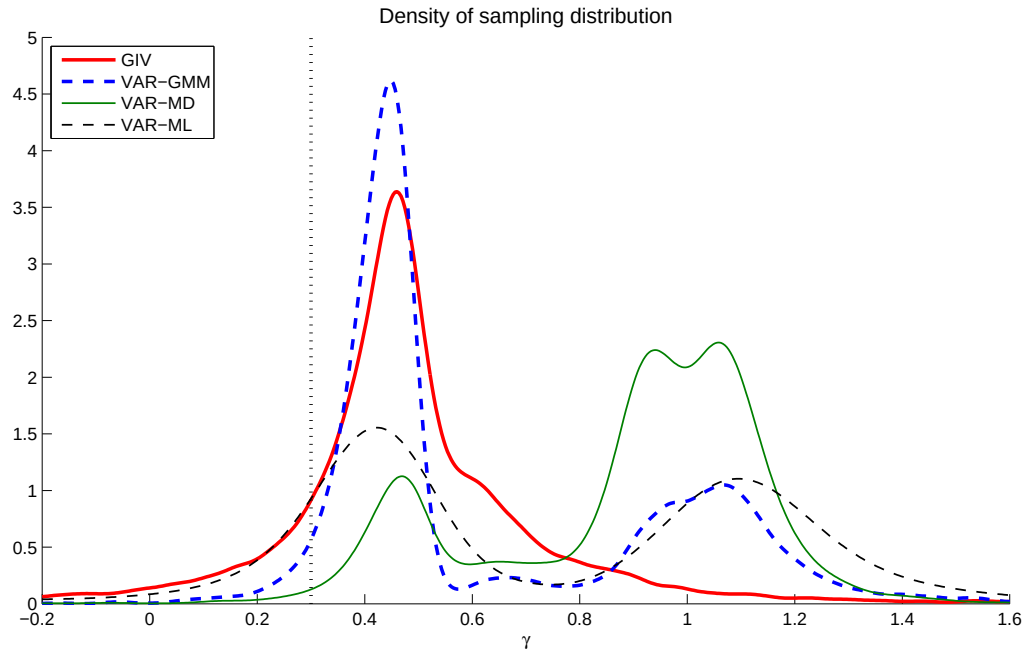


Figure 9: Sampling distribution of γ estimators, *DGP 2c*.

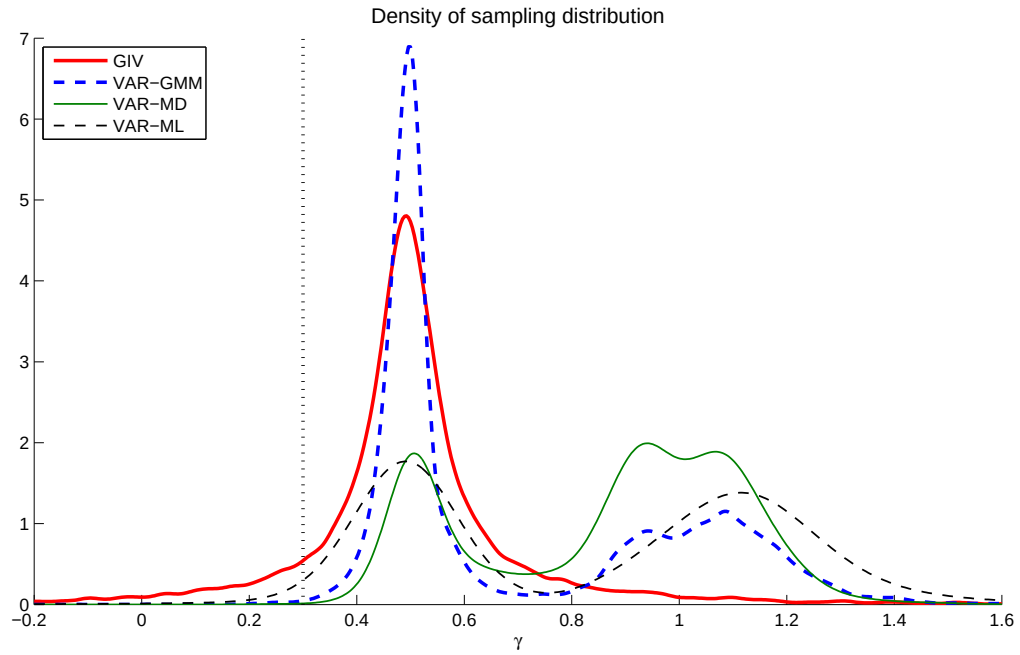


Figure 10: Sampling distribution of γ estimators, *DGP 2d*.

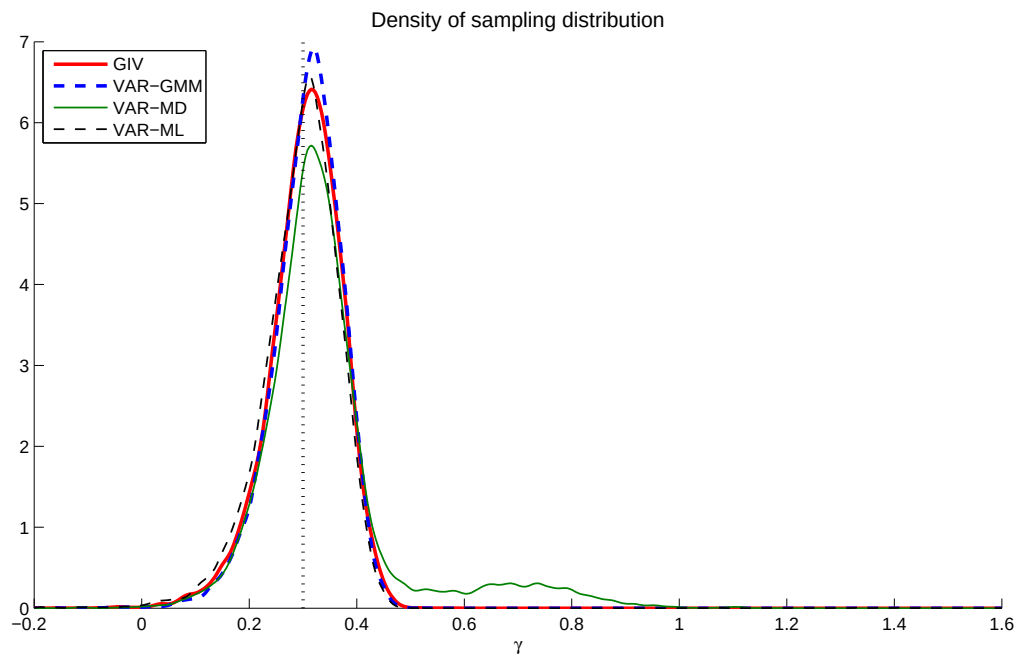


Figure 11: Sampling distribution of γ estimators, *DGP 2e*.