Identification using stability restrictions, Supplementary material: critical value tables, proofs and additional results

Leandro M. Magnusson* — The University of Western Australia

Sophocles Mavroeidis† — University of Oxford

March 6, 2014

Contents

*Department of Economics, Business School, The University of Western Australia, 35 Stirling Highway - M251, Crawley, WA 6009, Australia. Email: leandro.magnusson@uwa.edu.au

†Department of Economics and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Manor Road, Oxford, OX1 3UQ, United Kingdom. Email: sophocles.mavroeidis@economics.ox.ac.uk.

1 Introduction

This appendix contains tables of critical values, proofs, algebraic derivations, detailed description of econometric methods and additional empirical results. If the reader is primarily interested in the derivations and empirical results, the description of the computation algorithms can be skipped. Equations in this document are numbered with the suffix 'S–'. Equations without suffix refer to the main paper.

2 Tables of critical values for generalized S tests

Asymptotic critical values for the generalized and stability S tests, qLL-S, qLL-$\tilde{\text{S}}$, exp-S, exp-$\tilde{\text{S}}$, ave-S, ave-$\tilde{\text{S}}$, defined in Section 2 of the paper, are obtained by simulation from the distributions given in Theorems 1, 2 and 6 of the paper. k denotes the number of moment conditions and p_ζ denotes the number of strongly identified parameters that have been concentrated out. The critical values are computed using 50,000 draws of Brownian motion processes and 50,000 independent draws from the appropriate χ^2 distribution. We use 4,000 points to approximate the Brownian motion process. The trimming parameter for computing the ave– and exp–S tests is 15%.

Statistic:	qLL-S			exp-S			ave-S		
$k \setminus p_\zeta = 0$	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	8.59	9.99	13.03	2.43	3.10	4.68	4.16	5.31	8.14
2	15.32	17.10	20.78	4.15	4.98	6.77	7.14	8.60	12.00
3	21.76	23.82	28.02	5.73	6.66	8.80	9.95	11.63	15.37
4	27.98	30.31	34.79	7.24	8.26	10.50	12.52	14.35	18.28
5	34.21	36.56	41.81	8.65	9.80	12.27	15.09	17.08	21.61
6	40.12	42.75	48.03	10.05	11.23	13.78	17.57	19.60	24.16
7	46.10	48.95	54.90	11.42	12.70	15.33	20.00	22.29	26.91
8	52.16	55.23	61.17	12.74	14.08	16.82	22.39	24.67	29.50
9	58.09	61.30	67.68	14.06	15.47	18.35	24.73	27.21	32.23
10	64.09	67.24	73.80	15.42	16.86	19.85	27.13	29.80	35.11

Table 1: Asymptotic critical values for generalized S tests.
Derived from Theorems 1 and 2 in the paper, with $\tilde{c} = \bar{c} = c$, and $c = 10, \infty$ and 0 for qLL/exp/ave-S, resp. Computed using 50,000 draws of k-dimensional Brownian motion and 50,000 draws of $\chi^2(k - p_\zeta)$, where k is the number of moment conditions and p_ζ is the number of estimated parameters under the null.

Statistic:	qLL-S			exp-S			ave-S		
$k \setminus p_\zeta = 1$	10%	5%	1%	10%	5%	1%	10%	5%	1%
2	14.03	15.74	19.14	3.38	4.12	5.84	5.51	6.74	9.63
3	20.56	22.51	26.69	5.01	5.91	7.83	8.42	9.93	13.26
4	26.81	28.97	33.70	6.57	7.54	9.70	11.10	12.81	16.62
5	33.00	35.35	40.26	8.03	9.10	11.52	13.79	15.63	19.86
6	39.10	41.71	46.92	9.38	10.53	12.99	16.20	18.19	22.43
7	45.12	47.75	53.71	10.77	11.99	14.48	18.63	20.78	25.39
8	51.05	53.97	59.86	12.17	13.45	16.10	21.15	23.42	28.17
9	57.02	60.10	66.11	13.48	14.85	17.78	23.55	25.92	31.01
10	62.87	66.17	72.70	14.82	16.26	19.13	25.95	28.43	33.67

Table 2: Asymptotic critical values for generalized S tests. See Table 1 for details.

Statistic:	qLL-S			exp-S			ave-S		
$k \setminus p_\zeta$=2	10%	5%	1%	10%	5%	1%	10%	5%	1%
3	19.36	21.21	25.19	4.29	5.10	6.95	6.84	8.16	11.06
4	25.67	27.81	32.14	5.86	6.82	8.79	9.66	11.23	14.62
5	31.87	34.25	39.03	7.31	8.33	10.55	12.31	14.03	17.83
6	37.91	40.43	45.50	8.76	9.89	12.29	14.86	16.80	20.86
7	44.09	46.75	52.15	10.13	11.34	13.85	17.36	19.40	23.70
8	49.86	52.86	58.94	11.51	12.78	15.36	19.84	22.03	26.56
9	56.06	59.07	65.24	12.86	14.18	16.95	22.26	24.55	29.47
10	61.95	65.13	71.64	14.22	15.56	18.39	24.65	27.10	32.00

Table 3: Asymptotic critical values for generalized S tests. See Table 1 for details.

Statistic:	qLL-S			exp-S			ave-S		
$k \backslash p_\zeta$=3	10%	5%	1%	10%	5%	1%	10%	5%	1%
4	24.47	26.54	30.81	5.15	6.02	7.94	8.14	9.49	12.59
5	30.79	32.99	37.47	6.62	7.58	9.74	10.84	12.43	15.92
6	36.84	39.23	44.35	8.13	9.22	11.52	13.55	15.38	19.46
7	42.93	45.57	50.94	9.47	10.63	13.06	15.99	17.92	22.16
8	48.92	51.71	57.52	10.83	12.07	14.62	18.40	20.52	25.03
9	54.89	57.93	64.01	12.28	13.54	16.24	21.02	23.24	27.90
10	60.87	63.87	70.08	13.60	14.95	17.75	23.38	25.69	30.56

Table 4: Asymptotic critical values for generalized S tests. See Table 1 for details.

Statistic:	qLL-S			exp-S			ave-S		
$k \backslash p_\zeta$=4	10%	5%	1%	10%	5%	1%	10%	5%	1%
5	29.63	31.81	36.51	5.92	6.85	8.88	9.35	10.75	13.88
6	35.74	38.08	43.18	7.41	8.45	10.60	12.08	13.73	17.20
7	41.78	44.34	49.72	8.85	9.95	12.35	14.63	16.44	20.41
8	47.89	50.59	56.27	10.26	11.46	13.95	17.19	19.17	23.35
9	53.80	56.78	62.65	11.59	12.86	15.45	19.61	21.76	26.33
10	59.79	62.87	69.11	12.96	14.30	16.99	22.05	24.33	29.13

Table 5: Asymptotic critical values for generalized S tests. See Table 1 for details.

Statistic:	qLL-S			exp-S			ave-S		
$k \backslash p_\zeta$=5	10%	5%	1%	10%	5%	1%	10%	5%	1%
6	34.59	36.90	41.64	6.74	7.67	9.74	10.62	12.05	15.18
7	40.70	43.19	48.45	8.15	9.25	11.60	13.21	14.92	18.66
8	46.72	49.39	54.95	9.61	10.76	13.11	15.84	17.72	21.72
9	52.70	55.51	61.45	10.97	12.21	14.76	18.32	20.32	24.79
10	58.71	61.67	67.81	12.35	13.65	16.34	20.83	23.02	27.67

Table 6: Asymptotic critical values for generalized S tests. See Table 1 for details.

Statistic:	qLL-S̃			exp-S̃			ave-S̃		
k	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	7.17	8.36	11.10	1.50	2.03	3.35	2.15	2.85	4.59
2	12.79	14.30	17.58	2.54	3.20	4.72	3.69	4.58	6.52
3	18.14	19.95	23.51	3.50	4.23	5.96	5.15	6.16	8.33
4	23.33	25.28	29.19	4.43	5.22	7.03	6.55	7.65	10.01
5	28.47	30.63	34.93	5.21	6.08	8.02	7.80	8.98	11.59
6	33.48	35.78	40.46	6.04	6.96	8.92	9.11	10.39	13.01
7	38.49	40.85	45.77	6.82	7.78	9.84	10.29	11.61	14.48
8	43.47	46.02	51.30	7.62	8.60	10.78	11.56	12.96	15.88
9	48.39	51.07	56.56	8.39	9.41	11.61	12.81	14.23	17.27
10	53.39	56.05	61.80	9.13	10.23	12.60	14.00	15.51	18.74

Table 7: Asymptotic critical values for stability component of generalized S tests. Derived from Theorems 1 and 2 in the paper, with $\tilde{c} = 10, \infty$ and 0 for qLL/exp/ave-S̃, resp. Computed using 50,000 draws of k dimensional Brownian motion, where k denotes the number of moment conditions.

3 Proofs of theorems

Proof of Theorem 1 $\sum_{i=1}^{k} \hat{v}_i' (M_e - G_c) \hat{v}_i \Rightarrow \psi_c$ follows from the consistency of $\hat{V}_{ff}(\theta_0)$ and the FCLT on $F_{\cdot T}(\theta_0)$ by Lemma 6 of Elliott and Mueller (2006). Independence of $\overline{\psi}_k$ and ψ_c follows from the asymptotic independence between $\sqrt{T}\hat{V}_{ff}(\theta_0)^{-1/2}$ $F_T(\theta_0)$ and $[I_k \otimes (M_e - G_c)]\, \hat{v}$, where $\hat{v} = (v_1', \ldots, v_k')'$, which is a direct consequence of Assumption 2 under H_0. (26) then follows from the continuous mapping theorem. Finally, asymptotic efficiency follows from Mueller (2011, Theorem 1).

Proof of Theorem 2 By Assumption 2, under H_0, $V_X^{-1/2} X_T(1) \Rightarrow W(1)$ and $S_T(\theta_0) = X_T(1)' V_X^{-1} X_T(1) + o_p(1) \Rightarrow \overline{\psi}_k$. Also,

$$\begin{aligned}
\widetilde{S}_T(\theta_0, \tau) &= X_T(\tau)' \frac{V_X^{-1}}{\tau} X_T(\tau) + [X_T(1) - X_T(\tau)]' \frac{V_X^{-1}}{1-\tau} [X_T(1) - X_T(\tau)] \\
&\quad - X_T(1)' V_X^{-1} X_T(1) + o_p(1) \\
&= \frac{[X_T(\tau) - \tau X_T(1)]' V_X^{-1} [X_T(\tau) - \tau X_T(1)]}{\tau(1-\tau)} + o_p(1)
\end{aligned}$$

and $V_X^{-1/2}\left[X_T(\tau)-\tau X_T(1)\right]\Rightarrow\widetilde{W}(\tau)$, which is independent of $W(1)$ and $\widetilde{S}_T(\theta_0,\tau)\Rightarrow\widetilde{\psi}_k(\tau)$. By the Neyman-Pearson lemma, the test function

$$\mathbf{1}\left\{\frac{\bar{c}}{1+\bar{c}}W(1)'W(1)+2\log\int_{\varsigma}\exp\left[\frac{1}{2}\frac{\tilde{c}}{1+\tilde{c}}\frac{\widetilde{W}(\tau)'\widetilde{W}(\tau)}{\tau(1-\tau)}\right]d\nu_\tau>cv\right\}$$

maximizes WAP in the limiting problem $H_0 : dX(s) = V_X^{1/2}dW(s)$ against $H_1 : dX(s)=m(\theta,s)+V_X^{1/2}dW(s)$ and it is continuous at almost all realizations of W. Asymptotic efficiency then follows from Mueller (2011, Theorem 1).

Proof of Lemma 1 The proof is analogous to Andrews, Moreira, and Stock (2006, Lemma 2). Parts 1 and 2 follow from the fact that $Z(s)$ is nonstochastic and V is Gaussian. For part 3 note that, for every s_1, s_2, $Z(s_1)'Yb_0$ and $Z(s_2)'Y\Omega^{-1}A_0$ are jointly normal, and their covariance is $cov\left(Z(s_1)'Yb_0, Z(s_2)'Y\Omega^{-1}A_0\right)=\sum_{t=1}^T Z_t(s_1)'Z_t(s_2)$ $cov\left(Y_tb_0, Y_t\Omega^{-1}A_0\right)=\sum_{t=1}^T Z_t(s_1)'Z_t(s_2)\,b_0'\Omega\Omega^{-1}A_0=0$.

Proof of Theorem 3 Since the random functions $\overline{F}(\cdot)$ and $\overline{D}(\cdot)$ are independent of each other, by Lemma 1, and $\hat{\tau}$ only depends on $\overline{D}(\cdot)$ by (32), it follows that $\hat{\tau}$ is also independent of $\overline{F}(\cdot)$. Therefore, under H_0, Lemma 1 part 1 implies that the (conditional) distribution of $\overline{F}(\hat{\tau})$ is Gaussian with zero mean and variance matrix I_k. Part 1 follows immediately. For part 2, note that conditional on $\overline{D}(\hat{\tau})$ and $\hat{\tau}$, $\overline{D}(\hat{\tau})'F(\hat{\tau})$ is Gaussian with mean zero and variance $\overline{D}(\hat{\tau})'\overline{D}(\hat{\tau})$, and the matrix $\overline{D}(\hat{\tau})'\overline{D}(\hat{\tau})$ is invertible with probability 1. Parts 3 and 4 follow from the fact that $JKLM(\hat{\tau})=\overline{F}(\hat{\tau})'\overline{D}_\perp(\hat{\tau})\left[\overline{D}_\perp(\hat{\tau})'\overline{D}_\perp(\hat{\tau})\right]^{-1}\overline{D}_\perp(\hat{\tau})'\overline{F}(\hat{\tau})$, where $\overline{D}_\perp(\hat{\tau})$ is a $2k\times(2k-p)$ matrix which is the orthogonal complement of $\overline{D}(\hat{\tau})$, i.e., $\overline{D}_\perp(\hat{\tau})'\overline{D}(\hat{\tau})=0$, so $\overline{D}_\perp(\hat{\tau})'\overline{F}(\hat{\tau})$ and $\overline{D}(\hat{\tau})'\overline{F}(\hat{\tau})$ are independent conditionally on $\overline{D}(\hat{\tau})$ and $\hat{\tau}$. Part 5 now follows by combining the above results using the continuous mapping theorem.

Proof of Theorem 4 Assumption 3 yields the asymptotic counterpart of Lemma 1 for the linear model with fixed instruments and known variance. The asymptotic independence of $D_T(\theta_0,\cdot)$ and $F_{\cdot T}(\theta_0)$ implies that $\hat{\tau}_0$ will be asymptotically independent of $F_{\cdot T}(\theta_0)$ as well, and so, conditional on $\hat{\tau}_0$, $(\hat{\tau}_0T)^{-1/2}F_T^1(\theta_0,\hat{\tau}_0)$ and $((1-\hat{\tau}_0)T)^{-1/2}$ $F_T^2(\theta_0,\hat{\tau}_0)$ are jointly asymptotically Gaussian and independent with zero mean and variance V_{ff}. This establishes that $split\text{-}S_T(\theta_0,\hat{\tau}_0)\xrightarrow{d}\chi^2(2k)$. The remaining results follow by direct analogy with the proof of Theorem 3.

Proof of Theorem 5 First, by Assumptions 7 (b) and 8 (a) we have $T^{-1/2}F_{sT}(\theta_0) = O_p(1)$ and $\hat{V}^i_{qf}(\theta_0, s) = O_p(1)$, $\hat{V}^i_{ff}(\theta_0, s) = O_p(1)$ uniformly in $s \in [0,1]$, respectively. Hence,

$$vec\left[D^i_T(\theta_0, s)\right] = vec\left[Q^i_T(\theta_0, s)\right] + o_p(T) \tag{S–1}$$

uniformly in $s \in [0,1]$. Next, consider the two cases in Assumption 8 in turn.

The case of no (large) breaks: Assumption 8 (c). Assumption 7 implies $X^* = J'V_{ff}^{-1}X(1)$ and $V_{X^*} = J'V_{ff}^{-1}J$. Equation (S–1) and Assumption 7 (c) imply $T_i^{-1}D^i_T(\theta_0, \hat{\tau}_0) \xrightarrow{p} J$ for both $i = 1, 2$. Moreover, by Assumption 8 (b) and Slutsky's theorem we have $\frac{1}{T_i}D^i_T(\theta_0, \hat{\tau}_0)' \hat{V}^i_{ff}(\theta_0, \hat{\tau}_0)^{-1} \xrightarrow{p} J'V_{ff}^{-1}$, $i = 1, 2$ and $\hat{V}_{X^*} \xrightarrow{p} V_{X^*}$, since $|\hat{\tau}_0| \leq 1$. Finally, since $T^{-1/2}F^i_T(\theta_0, s)$ is uniformly bounded, $X^*_T = \sum_{i=1}^2 \frac{D^i_T(\theta_0,\hat{\tau}_0)'}{T^i}$ $\hat{V}^i_{ff}(\theta_0, \hat{\tau}_0)^{-1} \frac{F^i_T(\theta_0,\hat{\tau}_0)}{\sqrt{T}} = J'V_{ff}^{-1}\frac{1}{\sqrt{T}}F_T(\theta_0) + o_p(1) \xrightarrow{d} X^*$ by Assumption 7 (b) and the continuous mapping theorem.

The case of a large break: Assumption 8 (d). Assumption 7 implies $X^* = J_1'V_{ff}^{-1}X(\tau) + J_2'V_{ff}^{-1}[X(1) - X(\tau)]$ and $V_{X^*} = \tau J_1'V_{ff}^{-1}J_1 + (1-\tau) J_2'V_{ff}^{-1}J_2$.

First, we show that $T^{-1/2}\left[F^i_T(\theta_0, \hat{\tau}_0) - F^i_T(\theta_0, \tau)\right] = o_p(1)$, $i = 1, 2$. Observe that $\|F^i_T(\theta_0, \hat{\tau}_0) - F^i_T(\theta_0, \tau)\| = \left\|\sum_{j=1}^n f_{t_0+j}(\theta_0)\right\| \leq \sum_{j=1}^n \|f_{t_0+j}(\theta_0)\|$, $t_0 = [\min(\tau, \hat{\tau}_0)T]$ and $n = [|\hat{\tau}_0 - \tau|T]$. So, we need to show that for all $\eta, \delta > 0$, there exist T^* such that for all $T \geq T^*$, $\Pr\left(T^{-1/2}\sum_{j=1}^n \|f_{t_0+j}(\theta_0)\| > \delta\right) < \eta$. By Assumption 8 (d.iii), $T^{-1/2}\sum_{j=1}^N \|f_{t_0+j}(\theta_0)\| = o_p(1)$ for any fixed $N < \infty$, i.e., for every $\delta > 0$ there exist T_1^* large enough such that $\Pr\left(T^{-1/2}\sum_{j=1}^N \|f_{t_0+j}(\theta_0)\| > \delta\right) < \eta/2$ for all $T \geq T_1^*$. By Assumption 8 (d.ii), there exist $N < \infty$ and T_2^* large enough such that $\Pr(n > N) < \eta/2$ for all $T \geq T_2^*$. Hence,

$$\begin{aligned}
\Pr\left(T^{-1/2}\sum_{j=1}^n \|f_{t_0+j}(\theta_0)\| > \delta\right) &= \Pr\left(T^{-1/2}\sum_{j=1}^n \|f_{t_0+j}(\theta_0)\| > \delta | n \leq N\right)\Pr(n \leq N) \\
&\quad + \Pr\left(T^{-1/2}\sum_{j=1}^n \|f_{t_0+j}(\theta_0)\| > \delta | n > N\right)\Pr(n > N) \\
&\leq \Pr\left(T^{-1/2}\sum_{j=1}^N \|f_{t_0+j}(\theta_0)\| > \delta\right) + \Pr(n > N) < \eta
\end{aligned}$$

for all $T \geq T^{*} = \max\left(T_1^{*}, T_2^{*}\right)$.

Similar arguments can be used to establish $T^{-1}\left[D_T^i\left(\theta_0, \hat{\tau}_0\right) - D_T^i\left(\theta_0, \tau\right)\right] = o_p(1)$, $i = 1,2$. By (S–1), $T^{-1}\left\|D_T^i\left(\theta_0, \hat{\tau}_0\right) - D_T^i\left(\theta_0, \tau\right)\right\| = T^{-1}\left\|Q_T^i\left(\theta_0, \hat{\tau}_0\right) - Q_T^i\left(\theta_0, \tau\right)\right\| + o_p(1)$, so the result follows by Assumption 8 (d) (ii) and (iv) substituting $q_{t_0+j}\left(\theta_0\right)$ for $f_{t_0+j}\left(\theta_0\right)$ in the previous argument. Finally, Assumption 7 (c) yields

$$T_i^{-1} D_T^i\left(\theta_0, \hat{\tau}_0\right) = J_i + o_p(1), \quad i = 1,2. \tag{S–2}$$

Combining (S–2) with Assumption 8 (b) yields $\hat{V}_{X^*} \xrightarrow{p} V_{X^*}$ using Slutsky's theorem. $X_T^* \xrightarrow{d} X^*$ follows from (S–2), Assumption 7 (b) and the continuous mapping theorem.

Proof of Theorem 6 Let $\hat{X}_T^*(s) = T^{-1/2}\mathcal{W}_T^{1/2}\sum_{t=1}^{[sT]} f_t\left(\theta_0, \hat{\zeta}_0\right)$ and $X_T^*(s) = T^{-1/2} V_{ff}^{-1/2}\sum_{t=1}^{[sT]} f_t\left(\theta_0, \zeta_0\right)$. Assumption 9 (i) implies $X_T^*(s) \Rightarrow W(s)$, while Assumptions 9 (ii) and (iii) imply $\hat{X}_T^*(1) \Rightarrow MW(1)$ and $\hat{X}_T^*(s) \Rightarrow W(s) - sPW(1)$, where $M = I_k - P$, and $P = V_{ff}^{-1/2}\Gamma\left(\Gamma' V_{ff}^{-1}\Gamma\right)^{-1}\Gamma' V_{ff}^{-1/2}$. Hence, $\hat{X}_T^*(s) - s\hat{X}_T^*(1) \Rightarrow W(s) - sW(1) = \tilde{W}(s)$. This is the same as the distribution of the statistic $X_T^*(s) - sX_T^*(1)$ that does not involve estimation of any nuisance parameters ζ. Hence, the asymptotic distribution of the stability component of the gen-S statistics, which only involves $\hat{X}_T^*(s) - s\hat{X}_T^*(1)$, is the same as in Theorems 1 and 2. On the other hand, $S_T\left(\theta_0, \hat{\zeta}_0\right) = \hat{X}_T^*(1)' \hat{X}_T^*(1) \Rightarrow W(1)' MW(1) \sim \chi^2(k - p_\zeta)$. Moreover, $\hat{X}_T^*(s) - s\hat{X}_T^*(1)$ converges to a Brownian Bridge $\tilde{W}(s)$, which is independent of $W(1)$, showing that $S_T\left(\theta_0, \hat{\zeta}_0\right)$ and *gen*-$\widetilde{S}_T^{\tilde{c}}\left(\theta_0, \hat{\zeta}_0\right)$ are asymptotically independent.

Proof of Theorem 7 In the derivation of the split-sample statistics, $D_T\left(\theta_0, \cdot\right)$ and $F_{\cdot T}\left(\theta_0\right)$ are replaced with their counterparts that use $f_t\left(\theta_0, \hat{\zeta}_0\right)$ instead of $f_t\left(\theta_0\right)$ in their definition. Denote these by $D_T\left(\theta_0, \hat{\zeta}_0, \cdot\right)$ and $F_{\cdot T}\left(\theta_0, \hat{\zeta}_0\right)$. Under assumption 10, $D_T\left(\theta_0, \hat{\zeta}_0, \cdot\right)$ and $F_{\cdot T}\left(\theta_0, \hat{\zeta}_0\right)$ are asymptotically independent, and since $\hat{\tau}_0$ only depends on $D_T\left(\theta_0, \hat{\zeta}_0, \cdot\right)$, it will be asymptotically independent of $F_{\cdot T}\left(\theta_0, \hat{\zeta}_0\right)$, so we can condition on $\hat{\tau}_0$ to obtain the distribution of *split*-$S_T\left(\theta_0, \hat{\zeta}_0, \hat{\tau}_0\right) = \sum_{i=1}^{2} T_i^{-1} F_T^i\left(\theta_0, \hat{\zeta}_0, \hat{\tau}_0\right)'$ $\hat{V}_{ff}^i\left(\theta_0, \hat{\zeta}_0, \hat{\tau}_0\right)^{-1} F_T^i\left(\theta_0, \hat{\zeta}_0, \hat{\tau}_0\right)$. Now, $F_T^1\left(\theta_0, \hat{\zeta}_0, \hat{\tau}_0\right) = F_{\hat{\tau} T}\left(\theta_0, \hat{\zeta}_0\right)$, and $\hat{V}_{ff}^1\left(\theta_0, \hat{\zeta}_0, \hat{\tau}_0\right) \xrightarrow{p} V_{ff}$, and $\hat{\tau}_0 \Rightarrow \hat{\tau}_\infty$, where $\hat{\tau}_\infty$ is independent of $W_f(\cdot)$. Thus, $\hat{V}_{ff}^1\left(\theta_0, \hat{\zeta}_0, \hat{\tau}_0\right)^{-1/2}$ $F_T^1\left(\theta_0, \hat{\zeta}_0, \hat{\tau}_0\right) \Rightarrow W_f\left(\hat{\tau}_\infty\right) - \hat{\tau}_\infty PW_f(1) = \widetilde{W}_f\left(\hat{\tau}_\infty\right) + \hat{\tau}_\infty MW_f(1)$, where $\widetilde{W}_f(\cdot)$ is a

Brownian Bridge. Similarly, since $F_T^2\left(\theta_0,\hat{\zeta}_0,\hat{\tau}_0\right)=F_T\left(\theta_0,\hat{\zeta}_0\right)-F_{\hat{\tau}T}\left(\theta_0,\hat{\zeta}_0\right), \hat{V}_{ff}^2\left(\theta_0,\hat{\zeta}_0,\hat{\tau}_0\right)^{-1/2}$
$F_T^2\left(\theta_0,\hat{\zeta}_0,\hat{\tau}_0\right)\overset{d}{\to}MW_f\left(1\right)-\left[W_f\left(\hat{\tau}_\infty\right)-\hat{\tau}_\infty PW_f\left(1\right)\right]=\left(1-\hat{\tau}_\infty\right)MW_f\left(1\right)-\widetilde{W}_f\left(\hat{\tau}_\infty\right).$
So,

$$\begin{aligned} split\text{-}S_T\left(\theta_0,\hat{\zeta}_0,\hat{\tau}_0\right) &\overset{d}{\to} \tfrac{\left[\widetilde{W}_f(\hat{\tau}_\infty)+\hat{\tau}_\infty MW_f(1)\right]'\left[\widetilde{W}_f(\hat{\tau}_\infty)+\hat{\tau}_\infty MW_f(1)\right]}{\hat{\tau}_\infty} \\ &+\tfrac{\left[(1-\hat{\tau}_\infty)MW_f(1)-\widetilde{W}_f(\hat{\tau}_\infty)\right]'\left[(1-\hat{\tau}_\infty)MW_f(1)-\widetilde{W}_f(\hat{\tau}_\infty)\right]}{1-\hat{\tau}_\infty} \\ &= \tfrac{\widetilde{W}_f(\hat{\tau}_\infty)'\widetilde{W}_f(\hat{\tau}_\infty)}{\hat{\tau}_\infty(1-\hat{\tau}_\infty)}+W_f\left(1\right)'MW_f\left(1\right). \end{aligned}$$

The first result follows by noting that $W_f\left(1\right)$ and $\left[\hat{\tau}_\infty\left(1-\hat{\tau}_\infty\right)\right]^{-1/2}\widetilde{W}_f\left(\hat{\tau}_\infty\right)$ are independent standard normal vectors of dimension k, and M is idempotent with rank $k-p_\zeta$. The remaining results can be established analogously.

4 Testing general hypotheses

Let $g:\Theta\to\Re^r$, where $r\leq p=\dim\theta$, and consider the problem of testing

$$H_0:g\left(\theta\right)=0\ \text{ against }\ \ H_1:g\left(\theta\right)\neq 0. \tag{S–3}$$

If the function g is injective (one-to-one), then this can be done using the methods in the paper by setting $\theta_0=g^{-1}\left(0\right)$.

If g is not injective, then define $\Theta_0=\left\{\theta\in\Theta:g\left(\theta\right)=0\right\}$, such that (S–3) is equivalent to $H_0:\theta\in\Theta_0$ against $H_1:\theta\notin\Theta_0$. This can be tested using any of the tests in the paper by the projection method. Specifically, an η-level test can be obtained by performing η-level tests for $\theta=\theta_0$ for all points $\theta_0\in\Theta_0$, and rejecting the hypothesis if all those tests reject. In practice, since all the tests in the paper have a rejection region of the form $\mathrm{T}(\theta_0)>$crit, where $\mathrm{T}(\cdot)$ is the test statistic, the projection test can be performed simply by $\min_{\theta\in\Theta_0}\mathrm{T}(\theta)>$crit. Since $\mathrm{T}(\cdot)$ is smooth, this minimization problem can be solved using faster numerical algorithms than simple grid search over Θ_0.

An alternative to the projection method is the following. Suppose $r<p$, and there exists a partition of $\theta=\binom{\theta_1}{\theta_2}$, with $\dim\theta_1=r$, such that the function $h\left(\theta\right)=\binom{g(\theta)}{\theta_2}$ is injective, and let $h^{-1}\left(\cdot,\cdot\right)$ denote the inverse of h. Then, let $\varphi=g\left(\theta\right)$ and $\zeta=\theta_2$, so that $\theta=h^{-1}\left(\varphi,\zeta\right)$. If one is prepared to assume that ζ is strongly identified, and Assumptions 9 or 10 hold, then (S–3) can be tested by reparameterizing the problem into φ,ζ, and testing $\varphi=0$ concentrating out the strongly-identified parameter ζ, as

described in Subsection 3.5 of the paper. In practice, this can be done without explicit reparameterization as follows. Obtain the restricted estimator $\hat{\theta}_0$ by minimizing an efficient (full-sample) GMM criterion function subject to the restriction $g\left(\hat{\theta}_0\right) = 0$, evaluate all the test statistics at $\hat{\theta}_0$, and use the critical values given in Theorems 6 and 7, with $p_\zeta = p - r$.

5 Comparison to Rossi (2005)

In this section, we discuss in some detail the connection of our ave-S test to the Mean-Wald$_T^*$ test proposed by Rossi (2005) – the connection between the exp-S test and her Exp-Wald test is analogous. First, we demonstrate, using the linear IV example, that the statistics are not equivalent in general, although they are in the case of just-identified models. The intuition is that the Anderson-Rubin statistic is equivalent to the LM and LR statistics in just-identified models. Second, we show that the ave-S test in the original testing problem corresponds to Rossi's (2005) Mean-Wald$_T^*$ test applied to some *auxiliary* regression.

In the linear IV example, our methods are based on a specification of the form:

$$y_{1,t} = Y_{2,t}\theta + u_t, \tag{S–4}$$

$$Y_{2,t} = Z_t\Pi_t + V_{2,t}, \quad t = 1, ..., T \tag{S–5}$$

where $(y_{1,t}, Y_{2,t})$ is a $1 \times (1+p)$ random vector, u_t is a (structural) error, $\theta \in \Re^p$ is the unknown structural parameter vector, $Z_t \in \mathbb{R}^{1\times k}$ is the observed vector of instrumental variables, $V_{2,t} \in \mathbb{R}^{1\times p}$ is a (reduced-form) error vector, and $\Pi_t \in \mathbb{R}^{k\times p}$, $t = 1, ..., T$ is a sequence of unknown parameters. We are interested in testing

$$H_0 : \theta = \theta_0, \quad \text{against} \quad H_1 : \theta \neq \theta_0.$$

Our ave-S test in this model is based on the statistic

$$\begin{aligned} ave\text{-}S_T(\theta_0) &= S_T(\theta_0) + \int_\varsigma \widetilde{S}_T(\theta_0, s)\, d\nu_s \\ &= \frac{1}{T} F_T(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} F_T(\theta_0) + \int_\varsigma \frac{\widetilde{F}_{sT}(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} \widetilde{F}_{sT}(\theta_0)}{Ts(1-s)} d\nu_s. \end{aligned} \tag{S–6}$$

where ν_s is Uniform over $s \in \varsigma \subset (0,1)$, and its asymptotic distribution under H_0 is given by

$$ave\text{-}S_T(\theta_0) \Rightarrow \psi_k + \int_\varsigma \frac{\widetilde{W}_k(s)' \widetilde{W}_k(s)}{s(1-s)} d\nu_s,$$

where $\psi_k \sim \chi^2(k)$ and $\widetilde{W}_k(\cdot)$ is a k-dimensional standard Brownian Bridge process that is independent of ψ_k (see Theorem 2 in the paper).

Rossi's (2005) approach can be described using the specification

$$y_{1,t} = Y_{2,t}\theta_{t,T} + u_t, \tag{S–7}$$

$$Y_{2,t} = Z_t\Pi + V_{2,t}, \quad t = 1, ..., T. \tag{S–8}$$

Π is assumed to be constant over t, and it is also assumed to be of full rank, see Rossi (2005, Assumption 4). We specialize Rossi (2005, Assumption 7) to $H_0 : \theta_{t,T} = \theta_0$ for all t, T, and Rossi (2005, Assumption 2) to a single break in all parameters at some point τ:

$$H_{AT} : \theta_{t,T} = \theta_0 + \frac{1}{\sqrt{T}}\left[\theta_1 + 1_{\{s \geq \tau\}}\theta_2\right],$$

where $\theta_1, \theta_2 \in \Re^p$ and $\tau \in (0, 1)$.

The Mean-Wald test in Rossi (2005, eq. 26), with the uniform weights over the break dates used in (S–6) is given by

$$\text{Mean-Wald}_T^* = LM_1 + \int_\varsigma LM_2(s)\, d\nu_s \tag{S–9}$$

where

$$LM_1 = \frac{F_T(\theta_0)' \hat{\Sigma}^{-1/2} P_{\hat{\widetilde{M}}} \hat{\Sigma}^{-1/2} F_T(\theta_0)}{T}, \tag{S–10}$$

$$LM_2(s) = \frac{\widetilde{F}_{sT}(\theta_0)' \hat{\Sigma}^{-1/2} P_{\hat{\widetilde{M}}} \hat{\Sigma}^{-1/2} \widetilde{F}_{sT}(\theta_0)}{Ts(1-s)}, \tag{S–11}$$

$\hat{\widetilde{M}} = \hat{\Sigma}^{-1/2} Q_T(\theta_0)$, and $\hat{\Sigma} = \hat{V}_{ff}(\theta_0)$. It follows from Rossi (2005, Proposition 2(b)) that

$$\text{Mean-Wald}_T^* \Rightarrow \psi_p + \int_\varsigma \frac{\widetilde{W}(s)' \widetilde{W}(s)}{s(1-s)} d\nu_s,$$

where $\psi_p \sim \chi^2(p)$ and $\widetilde{W}(\cdot)$ is a $p \times 1$ standard Brownian Bridge process that is independent of ψ_p. It is evident that the Mean-Wald$_T^*$ statistic in (S–9) is generally different from $ave\text{-}S_T(\theta_0)$ statistic given in (S–6) above. The Mean-Wald$^*_T$ statistic involves a projection of the moment vectors $F_T(\theta_0)$ and $\widetilde{F}_{sT}(\theta_0)$ onto the space spanned by the Jacobian of the moment conditions, while the $ave\text{-}S_T(\theta_0)$ statistic uses the full vectors $F_T(\theta_0)$ and $\widetilde{F}_{sT}(\theta_0)$. An exception occurs when the model is just-identified, i.e., $k = p$, in which case, since $\hat{\overline{M}}$ is a square matrix and $P_{\hat{\overline{M}}} = I_k$, $ave\text{-}S_T(\theta_0)$ =Mean-Wald$^*_T$. This is intuitive because in this case the Jacobian of the moment conditions plays no role in the construction of the statistics, so the different assumptions about the first-stage regression have no impact on the statistics. This is exactly analogous to the fact that the Anderson-Rubin test is equivalent to the LM and LR tests in a just-identified model.[1] Finally, even when the break date is assumed to be known, i.e., the support ς of ν_s contains a single point, and identification is strong, the Mean-Wald$^*_T$ test is different from the split-KLM/CLR tests. To see this, observe that the strong-instrument asymptotic distribution of the Mean-Wald$^*_T$ statistic under the null is $\chi^2(2p)$, while that of the split-KLM and CLR statistics is $\chi^2(p)$.

Now, consider the Mean-Wald$^*_T$ test for the null hypothesis $H_0^* : \delta_t = 0$ in the following auxiliary regression model

$$y_{0,t} = Z_t\delta_t + u_{0,t}, \tag{S–12}$$

where $y_{0,t} \equiv y_{1,t} - Y_{2,t}\theta_0$, against time-varying alternatives, e.g., the local alternatives $H^*_{AT} : \delta_t = \frac{1}{\sqrt{T}}\left(\delta_1 + \delta_2 1_{\{t \geq [\tau T]\}}\right), \tau \in \varsigma$ in Rossi (2005, Assumption 2). Denote the moment function for this problem by $g_t(\delta_t) = Z_t'(y_{0,t} - Z_t\delta_t)$, to distinguish it from $f_t(\theta)$ in the original model. Note that this auxiliary model is just-identified (the number of parameters in δ is equal to the number of instruments), and the variance of the moment conditions is identical to V_{ff} under the null (since $g_t(0) = f_t(\theta_0)$). It follows that $P_{\hat{\overline{M}}} = I_k$ and $\hat{\Sigma}$ can be chosen as $\hat{V}_{ff}(\theta_0)$ in (S–10) and (S–11), so that the Mean-Wald$^*_T$ statistic in (S–9) coincides with $ave\text{-}S_T(\theta_0)$ in (S–6). In other words, the Mean-Wald$^*_T$ (S–9) for testing H_0^* in the *auxiliary regression* (S–12) is identical to the ave-S_T statistic (S–6) in the original model. The same connection holds for our exp-S statistic and the Exp-Wald$^*_T$ statistic of Rossi's (2005). This is entirely analogous to the fact that the Anderson and Rubin (1949) statistic for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ in the canonical IV regression model, which is obtained from (S–4) and

[1] In the just-identified case, the Anderson-Rubin test is also equivalent to the modified Wald test of Wang and Zivot (1998), where the variance of the structural shock is computed under the null, see Dufour (2003, p. 795).

(S–5) when we assume $\Pi_t = \Pi$ for all t and the errors are Gaussian and homoskedastic, is the same as the F statistic for testing the exclusion restrictions $H_0^* : \delta = 0$ against $H_1^* : \delta \neq 0$ in the auxiliary regression $y_{0,t} = Z_t\delta + u_{0,t}$, see Dufour (2003, p. 789).

The above connection between the ave-S test and Rossi's (2005) Mean-Wald test in some auxiliary regression holds more generally for any model specified in terms of the moment conditions (1) in the paper, for which Assumptions 1 and 2 hold. The auxiliary regression is the local level model

$$y_t = \mu_t + e_t$$

where $y_t = f_t(\theta_0)$, $\mu_t = E[f_t(\theta_0)]$ and $e_t = f_t(\theta_0) - \mu_t$. The null hypothesis in the auxiliary regression is $H_0^* : \mu_t = 0$ against a single-break alternative.

6 Derivation of the solution in Example RS

The model given by (7) and (8) in the paper is:

$$y_t = \beta E[y_{t+1}|\mathcal{I}_t] + \gamma x_t + u_t \tag{S–13}$$

$$x_t = \rho x_{t-1} + (1-\rho)\varphi y_t + \varepsilon_t. \tag{S–14}$$

A solution to the model is given by

$$y_t = \alpha_1 x_{t-1} + v_t^y \tag{S–15}$$

$$x_t = \rho_1 x_{t-1} + v_t^x, \tag{S–16}$$

where α_1, ρ_1 will be obtained later using the method of undetermined coefficients. The conditions for existence and uniqueness of a stable solution can be checked using the method of Blanchard and Kahn (1980). Equations (S–13) and (S–14) can be written in the Blanchard and Kahn (1980) canonical form as

$$E[Y_{t+1}|\mathcal{I}_t] = AY_t + Z_t,$$

where $Y_t = (y_t, x_{t-1})'$, $A = B^{-1}C$,

$$B = \begin{pmatrix} \beta & \gamma \\ 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 \\ (1-\rho)\,\varphi & \rho \end{pmatrix}, \quad Z_t = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u_t \\ \varepsilon_t \end{pmatrix},$$

(B is invertible because we assume $\beta \neq 0$ – if $\beta = 0$, the solution is trivial). Since there is one predetermined and one nonpredetermined variable in Y_t, existence of a stable solution requires that at least one of the eigenvalues of A should lie inside the unit circle. Determinacy (uniqueness) requires that the other root should lie outside the unit circle. The eigenvalues of A are the same as the roots λ_1, λ_2 of the determinantal equation

$$\det(C - \lambda B) = \rho - (1 + \beta\rho - \gamma\varphi\,(1-\rho))\,\lambda + \beta\lambda^2 = 0. \tag{S–17}$$

Since $\lambda_1\lambda_2 = \frac{\rho}{\beta}$and $\lambda_1 + \lambda_2 = \frac{1+\beta\rho-\gamma\varphi(1-\rho)}{\beta}$, it follows that for $\rho, \beta \in [0,1]$, $\gamma \geq 0$, and $\varphi \leq 0$, the roots are real and non-negative, and the smallest root is

$$\lambda_1 = \frac{1 + \beta\rho - \gamma\varphi\,(1-\rho) - \sqrt{(1 + \beta\rho - \gamma\varphi\,(1-\rho))^2 - 4\beta\rho}}{2\beta}. \tag{S–18}$$

Since λ_1 is increasing in φ, and $\lambda_1 = \rho < 1$ at $\varphi = 0$, the root λ_1 is stable, and a stable solution exists. Moreover,

$$\lambda_2 = \frac{1 + \beta\rho - \gamma\varphi\,(1-\rho) + \sqrt{(1 + \beta\rho - \gamma\varphi\,(1-\rho))^2 - 4\beta\rho}}{2\beta}$$

is decreasing in φ, and $\lambda_2 = \beta^{-1} > 1$ at $\varphi = 0$, so the solution is determinate for all $\varphi \leq 0$.

We can then determine α_1, ρ_1, v_t^y and v_t^x by the method of undetermined coefficients. Using (S–15) to substitute for $E[y_{t+1}|\mathcal{I}_t]$ in (S–13) and re-arranging yields

$$y_t = (\beta\alpha_1 + \gamma)\,x_t + u_t, \tag{S–19}$$

and substituting for x_t from (S–16) yields the equations

$$y_t = \underbrace{(\beta\alpha_1 + \gamma)\,\rho_1}_{\alpha_1} x_{t-1} + \underbrace{u_t + (\beta\alpha_1 + \gamma)\,v_t^x}_{v_t^y}. \tag{S–20}$$

So, since $\beta, \rho_1 < 1$,

$$\alpha_1 = \frac{\gamma\rho_1}{1-\beta\rho_1}. \tag{S–21}$$

Substituting for y_t into (S–14) using (S–15) and re-arranging yields

$$\begin{aligned} x_t &= \rho x_{t-1} + (1-\rho)\,\varphi\,(\alpha_1 x_{t-1} + v_t^y) + \varepsilon_t \\ &= \underbrace{(\rho + (1-\rho)\,\varphi\alpha_1)}_{\rho_1} x_{t-1} + \underbrace{(1-\rho)\,\varphi v_t^y + \varepsilon_t}_{v_t^x}. \end{aligned} \tag{S–22}$$

Upon substituting for α_1 from (S–21), the first term yields a quadratic equation for ρ_1, which is the same as (S–17) above. Hence, the smallest root corresponds to

$$\rho_1 = \lambda_1,$$

where λ_1 is given by (S–18). Substituting for ρ_1 in (S–21) yields the solution for α_1. Finally, substituting for v_t^y into (S–22) using (S–20) and re-arranging yields

$$v_t^x = \frac{\varepsilon_t + (1-\rho)\,\varphi u_t}{1-(1-\rho)\,\varphi\,(\beta\alpha_1+\gamma)},.$$

and using (S–20) yields

$$v_t^y = u_t + (\beta\alpha_1+\gamma)\,v_t^x = \frac{u_t + (\beta\alpha_1+\gamma)\,\varepsilon_t}{1-(1-\rho)\,\varphi\,(\beta\alpha_1+\gamma)}.$$

7 Supplementary material for empirical section

7.1 The baseline new Keynesian Phillips curve

The baseline model is given by:

$$\pi_t - \varrho\pi_{t-1} = \beta E\left(\pi_{t+1} - \varrho\pi_t|\mathcal{I}_t\right) + \lambda\widehat{mc}_t + \varepsilon_t, \tag{S–23}$$

where:

$$\lambda = \tfrac{(1-\alpha)(1-\beta\alpha)}{\alpha}\upsilon, \qquad \upsilon = \tfrac{a(\mu-1)}{(\mu-a)}.$$

π_t is inflation, a is the labor elasticity of a Cobb-Douglas production function (the average labor share), μ is the desired mark-up under flexible prices, β is a discount factor, ϱ is the fraction of prices that are indexed to past inflation when they cannot be optimally reset, α is the probability that a price will be fixed in a given period. $\widehat{mc}_t$ is the log deviation of real marginal costs from their steady state and ε_t is a cost-push (e.g., mark-up) shock. We shall proxy real marginal costs using the labor share, see below.

We impose the restriction $\beta = 1$, so the NKPC can be written as:

$$\varrho \Delta \pi_t = \kappa + E\left(\Delta \pi_{t+1} | \mathcal{I}_t\right) + \frac{(1-\alpha)^2}{\alpha} \widetilde{x}_t + \varepsilon_t, \tag{S–24}$$

and $\widetilde{x}_t = \upsilon x_t$, where $x_t = \ln \frac{S_t}{a}$, S_t is the labor share, and υ is calibrated to 0.25, using $a = \frac{2}{3}$ and $\mu = 1.2$. The constant κ is equal to $\lambda \ln \mu$, and captures the steady-state value of real marginal costs $-\ln \mu$, see Woodford (2003) or Galí (2008).

The moment condition used for the derivation of the tests has the form $E\left[Z_t' \epsilon_t\right] = 0$ where ϵ_t is a scalar residual function,

$$\epsilon_t = \underbrace{\varrho \Delta \pi_t - \Delta \pi_{t+1} - \frac{(1-\alpha)^2}{\alpha} \widetilde{x}_t}_{Y_t b} - \underbrace{\kappa}_{X_t c}, \tag{S–25}$$

and Z_t is a $1 \times k_z$ vector of instruments. The row vector Y_t is $\left(\Delta \pi_t, \Delta \pi_{t+1}, \widetilde{x}_t\right)$, $X_t = 1$ and $b = \left(\varrho, -1, -\frac{(1-\alpha)^2}{\alpha}\right)'$. In all of our empirical results, the set of instruments Z_t includes a constant, two lags of the change in inflation and three lags of the forcing variable.

7.1.1 Data

We measure inflation as $\pi_t = \ln\left(\frac{\mathrm{P}_t}{\mathrm{P}_{t-1}}\right)$ where P_t is the GDP implicit price deflator (Index numbers, 2005=100. Seasonally adjusted). This series was obtained from the Bureau of Economic Analysis website, Table 1.1.9.

Our measure of real marginal costs is $\frac{S_t}{a}$ where S_t is the labor share. This is based on a Cobb-Douglas production function with constant returns to scale where MPL = $a \times$APL. This series, measured in levels, was obtained directly from the Bureau of Labor Statistic, Labor and Productivity Division. It is not the PRS85006173 labor share index series, which is publicly available, although they are almost perfectly correlated. The parameter a, the average labor share coefficient in the Cobb-Douglas function, is set to $\frac{2}{3}$ in all cases.

7.1.2 Computational details

In this subsection, we describe in detail the computation of the various test statistics used in the paper. In the following exposition, Z and X are $T\times k_z$ and $T\times k_x$ matrices, respectively. We use I_k to denote the $k\times k$ identity matrix, and $\mathrm{M}_{\mathsf{a}} = \mathrm{I} - \mathsf{a}\left(\mathsf{a}'\mathsf{a}\right)^{-1}\mathsf{a}'$ is a projection matrix. The vector of tested parameters is θ which has p elements.

Computation of S and CLR statistics In matrix notation, the empirical moments derived from equation (S–25) are:

$$Z'\epsilon = Z'\left(Yb - Xc\right)$$

where Y, Z, X and ϵ are matrices consisting of the T stacked rows of Y_t, Z_t, X_t and ϵ_t, respectively. Under the null assumption $H_0 : \theta = (\alpha, \varrho) = (\alpha_0, \varrho_0) = \theta_0$, $b = b\left(\theta_0\right)$ is fixed. We define $\widehat{V}_{ff}$, the estimator asymptotic variance of $\frac{1}{\sqrt{T}}Z'\left(Yb - Xc\right)$, as:

$$\widehat{V}_{ff} = \left(b' \otimes I_{k_z}\right)\widehat{\Sigma}\left(b \otimes I_{k_z}\right) \tag{S–26}$$

where $\widehat{\Sigma}$, the full sample HAC estimator, is:

$$\widehat{\Sigma} = A\left[\widehat{\Gamma}_0 + \sum_{j=1}^{T}\omega_j\left(\widehat{\Gamma}_j + \widehat{\Gamma}'_j\right)\right]A', \tag{S–27}$$

$$\widehat{\Gamma}_j = \left[\frac{1}{T}\sum_{t=j+1}^{T}\widehat{w}_t\widehat{w}'_t\right],$$

where $\widehat{w}_t$ is the 'pre-whitened' $\mathrm{vec}(Z'_t\widehat{Y}_t)$, i.e., the residuals from a VAR(1) with coefficient matrix A, $\widehat{Y}_t$ is the t^{th}-row of the matrix M_XY, ω_j is the Bartlett kernel with $\left\lfloor 4\left(T/100\right)^{2/9}\right\rfloor$ as the lag truncation parameter,[2] and A is the 'recoloring' matrix. This procedure for estimating $\mathrm{Var}\left(\frac{1}{\sqrt{T}}Z'\left(Yb - Xc\right)\right)$ is equivalent to using $\{\mathrm{vec}(Z'_t\widehat{\epsilon}_t)\}_{t=1}^{T}$ as the empirical moments where:

$$\widehat{\epsilon} = Yb - X\widehat{c}_1 = \mathrm{M}_XYb,$$

[2] $\left\lfloor 4\left(T/100\right)^{2/9}\right\rfloor$ means the largest integer smaller than $4\left(T/100\right)^{2/9}$, which is 4 lags when using the sample is 1966q1 to 2010q4.

and $\widehat{c}_1$ is the first-step estimator $\widehat{c}_1 = (X'X)^{-1} X'Yb$.

We estimate c by minimizing the following objective function:

$$\widehat{c}_2 = \arg\min_c \frac{1}{T} (Yb - Xc)' Z\widehat{V}_{ff}^{-1} Z' (Yb - Xc) \qquad \text{(S–28)}$$

i.e.,

$$\widehat{c}_2 = \left(X'Z\widehat{V}_{ff}^{-1} Z'X \right)^{-1} X'Z\widehat{V}_{ff}^{-1} Z'Yb$$

Under our maintained assumptions, the two-step estimator $\widehat{c}_2$ is $\sqrt{T}$-consistent with asymptotic distribution:

$$\sqrt{T} (\widehat{c}_2 - c) \overset{a}{=} \left(\Gamma'_{ZX} V_{ff}^{-1} \Gamma_{ZX} \right)^{-1} \Gamma'_{ZX} V_{ff}^{-\frac{1}{2}} \xi$$

where $\text{plim}_{T \to +\infty} \frac{1}{T} (Z'X) = \Gamma_{ZX}$, $\text{plim}_{T \to +\infty} \widehat{V}_{ff} = V_{ff}$, a positive definite matrix, $V_{ff}^{-\frac{1}{2}}$ is the symmetric square root matrix of V_{ff}^{-1}, and ξ is a $k_z \times 1$ standard normal random vector.

Substituting $\widehat{c}_2$ back into the objective function in (S–28) we derive:

$$b'Y'Z\widehat{V}_{ff}^{-\frac{1}{2}} \mathrm{M}_{\widehat{V}_{ff}^{-\frac{1}{2}} \widehat{\Gamma}_{ZX}} \widehat{V}_{ff}^{-\frac{1}{2}} Z'Yb$$

where $\widehat{\Gamma}_{ZX} = \frac{Z'X}{T}$. Let L is a $k_z \times (k_z - k_x)$ matrix such that:

$$LL' = \mathrm{M}_{\widehat{V}_{ff}^{-\frac{1}{2}} \widehat{\Gamma}_{ZX}} \text{ and } L'L = \mathrm{I}_{(k_z - k_x)},$$

The S statistic is:

$$S_T(\theta_0) = \frac{1}{T} \underbrace{b'Y'Z\widehat{V}_{ff}^{-\frac{1}{2}} L}_{\widehat{\xi}_T} \underbrace{L'\widehat{V}_{ff}^{-\frac{1}{2}} Z'Yb}_{\widehat{\xi}_T}.$$

Let $\nabla_{\theta_0} b = \frac{\partial b}{\partial \theta}\big|_{\theta = \theta_0}$, and define the $(k_z - k_x) \times p$ Jacobian matrix $\hat{q}_T = \hat{q}_T (\theta_0)$ as:

$$\hat{q}_T = L'\widehat{V}_{ff}^{-\frac{1}{2}} Z'Y \nabla_{\theta_0} b$$

The estimator of the variance-covariance matrix of $T^{-\frac{1}{2}}\left(\widehat{\xi}_T,\ \text{vec}\left(\hat{q}_T\right)\right)$ are:

$$\begin{aligned}
\widehat{V}_{\xi\xi} &= \mathrm{I}_{(k_z-k_x)} \\
\widehat{V}_{q\xi} &= \left(\nabla_{\theta_0} b' \otimes L'\widehat{V}_{ff}^{-\frac{1}{2}}\right)\widehat{\Sigma}\left(b \otimes \widehat{V}_{ff}^{-\frac{1}{2}} L\right) = \widehat{V}_{\xi q}' \\
\widehat{V}_{qq} &= \left(\nabla_{\theta_0} b' \otimes L'\widehat{V}_{ff}^{-\frac{1}{2}}\right)\widehat{\Sigma}\left(\nabla_{\theta_0} b \otimes \widehat{V}_{ff}^{-\frac{1}{2}} L\right)
\end{aligned}$$

We compute the $\widehat{D}_T$ statistic, which is a $(k_z - k_x) \times p$ matrix, as:

$$\widehat{D}_T = \frac{1}{T}\,\text{mat}\left(\text{vec}\left(\hat{q}_T\right) - \widehat{V}_{q\xi}\widehat{\xi}_T\right)$$

where mat is the inverse of the vec operator, such that $\text{vec}(\text{mat}\,(x)) = x$. The KLM, JKLM and CLR statistics are computed as:

$$\begin{aligned}
KLM_T(\theta_0) &= \tfrac{1}{T}\,\widehat{\xi}_T'\widehat{D}_T\left(\widehat{D}_T'\widehat{D}_T\right)^{-1}\widehat{D}_T'\widehat{\xi}_T \\
JKLM_T(\theta_0) &= S_T(\theta_0) - KLM_T(\theta_0) \\
CLR_T(\theta_0) &= \tfrac{1}{2}\left\{S_T(\theta_0) - rk(\theta_0) \times \sqrt{\left[S_T(\theta_0) + rk(\theta_0)\right]^2 - 4JKLM_T(\theta_0)rk(\theta_0)}\right\}
\end{aligned} \tag{S–29}$$

where $rk(\theta_0)$ is a rank statistic, which is function of $\widehat{D}_T$ and its variance matrix $\widehat{V}_{qq.\xi} = \widehat{V}_{qq} - \widehat{V}_{q\xi}\widehat{V}_{\xi\xi}^{-1}\widehat{V}_{q\xi}'$. The results are based on Kleibergen and Paap (2006) rank test.

Computation of the qLL-S statistic In the algorithm for computing the qLL-S statistic, we use the following $T \times T$ matrices and $T \times 1$ vector:

$$\Delta_{\mathrm{D}} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & 0 & \cdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & -1 & 1 & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ r & 1 & 0 & \cdots & \vdots \\ r^2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & r & 1 & 0 \\ r^{T-1} & \cdots & r^2 & r & 1 \end{bmatrix}, \text{ and } \quad r_T = \begin{bmatrix} r \\ r^2 \\ \vdots \\ \vdots \\ r^T \end{bmatrix}$$

where $r = 1 - \frac{10}{T}$. The Δ_{D} matrix is a first difference operator, while R is the cumulative product operator matrix. Let $\widehat{E}$ be the following $T \times k_z$ matrix:

$$\widehat{E} = \left[\widehat{\epsilon}, \ldots, \widehat{\epsilon}\right] \tag{S–30}$$

where $\widehat{\epsilon} = Yb - X\widehat{c}_2$. The qLL-S̃ statistic, which is the stability part of the qLL-S test, is obtained using the following steps:

1. Compute first the $T \times k_z$ matrices

$$\widehat{F} = \left(\widehat{E} \odot Z\right) \widehat{V}_{ff}^{-\frac{1}{2}}, \text{ and } \widehat{H} = R\Delta_{\mathrm{D}}\widehat{F} \tag{S–31}$$

   where $\odot$ denotes the direct product (element-by-element multiplication);

2. Estimate the $T \times k_z$ matrix $\widehat{G} = M_{r_T}\widehat{H}$, the OLS residuals of the following regression:

$$\widehat{H} = r_T d_G + G,$$

   where d_G is a $1 \times k_z$ row vector of parameters, and compute

$$TSSR_{\widehat{G}} = \sum_{i=1}^{k_z}\sum_{t=1}^{T} \left(\hat{g}_{i,t}\right)^2,$$

   where $\hat{g}_{i,t}$ is the (t, i) element of the matrix $\widehat{G}$.

3. Compute the $T \times k_z$ matrix $\widehat{N} = M_{\iota_T}\widehat{F}$, the OLS residuals of the following regression:

$$\widehat{F} = \iota_T d_N + N$$

   where ι_T is a $T \times 1$ vector of ones and d_N is a $1 \times k_z$ row vector of parameters. Calculate

$$TSSR_{\widehat{N}} = \sum_{i=1}^{k}\sum_{t=1}^{T} \left(\hat{n}_{i,t}\right)^2,$$

   where $\hat{n}_{i,t}$ is the (t, i) element of the matrix $\widehat{N}$.

The qLL-S̃ statistic under H_0 is:

$$\text{qLL-}\widetilde{S}_T(\theta_0) = TSSR_{\widehat{N}} - r \times TSSR_{\widehat{G}}.$$

The qLL-S test is defined as:

$$qLL\text{-}S_T(\theta_0) = qLL\text{-}\widetilde{S}_T(\theta_0) + \frac{10}{11} S_T\left(\theta_0\right).$$

Computation of ave-S and exp-S statistics Partition $Z = \left[\begin{array}{cc} Z_1' & Z_2' \end{array} \right]'$ where Z_1 and Z_2 are, respectively, $[sT] \times k_z$ and $T - [sT] \times k_z$ matrices, $s \in (0,1)$, and $[sT]$ is the largest integer lower than sT. We define the split-sample moment condition as:

$$\overline{Z}' \underbrace{(Yb - Xc)}_{\epsilon}$$

where $\overline{Z}$ is the $T \times 2k_z$ 'split-sample' instruments matrix:

$$\overline{Z} = \overline{Z}(s) = \left[\begin{array}{cc} Z_1 & 0 \\ 0 & Z_2 \end{array} \right] \tag{S–32}$$

Similar to equation (S–26), we define $\widehat{V}_s$, the estimator of the asymptotic variance of $\mathrm{Var}\left(\frac{1}{\sqrt{T}} \overline{Z}' (Yb - Xc) \right)$, as:

$$\widehat{V}_{ff,s} = (b' \otimes I_{2k_z}) \widehat{\Sigma}_s (b' \otimes I_{2k_z}) \tag{S–33}$$

where $\widehat{\Sigma}_s$ is the HAC estimator of the asymptotic variance of $\frac{1}{\sqrt{T}} \mathrm{vec}\left(\overline{Z}' Y \right)$, which is computed as:

$$\widehat{\Sigma}_s = \mathrm{P}_R \begin{pmatrix} s\widehat{\Sigma}_1 & 0 \\ 0 & (1-s)\widehat{\Sigma}_2 \end{pmatrix} \mathrm{P}_C$$

where, for $i = 1,\, 2$, $\widehat{\Sigma}_i$ is an estimator of the asymptotic variance of $T_i^{-\frac{1}{2}} \mathrm{vec}(Z_i' Y_i)$, $T_1 = [sT]$ and $T_2 = T - [sT]$. P_R and P_C are permutation matrices such that:

$$\mathrm{Var}\left[\frac{1}{\sqrt{T}} \mathrm{vec}\left(\overline{Z}' Y \right) \right] = \mathrm{P}_R \, \mathrm{Var}\left[\frac{1}{\sqrt{T}} \mathrm{vec}\left(\begin{array}{cc} Z_1' Y_1 & Z_2' Y_2 \end{array} \right) \right] \mathrm{P}_C \tag{S–34}$$

Under the assumptions of the model, $\underset{T_i \to +\infty}{\mathrm{plim}} \widehat{\Sigma}_i = \underset{T \to +\infty}{\mathrm{plim}} \widehat{\Sigma} = \Sigma$. So we estimate $\widehat{\Sigma}_s$ as:

$$\widehat{\Sigma}_s = \mathrm{P}_R \begin{pmatrix} s\widehat{\Sigma} & 0 \\ 0 & (1-s)\widehat{\Sigma} \end{pmatrix} \mathrm{P}_C \tag{S–35}$$

where $\widehat{\Sigma}$ is defined in equation (S–27).

The estimator of the strongly identified parameter is the minimizer of the split-sample objective function:

$$\widehat{c}_s = \arg\min_c \frac{1}{T} (Yb - Xc)' \overline{Z} \widehat{V}_{ff,s}^{-1} \overline{Z}' (Yb - Xc) \tag{S–36}$$

i.e.,

$$\widehat{c}_s = \left(X' \overline{Z} \widehat{V}_{ff,s}^{-1} \overline{Z}' X \right)^{-1} X' \overline{Z} \widehat{V}_{ff,s}^{-1} \overline{Z}' Yb$$

($\widehat{c}_s$ is asymptotically equivalent to $\widehat{c}_2$ defined earlier). Substituting $\widehat{c}_s$ into the split-sample objective function we derive:

$$b'Y'\overline{Z}\widehat{V}_{ff,s}^{-\frac{1}{2}} \mathrm{M}_{\widehat{V}_{ff,s}^{-\frac{1}{2}} \widehat{\Gamma}_{\overline{Z}X}} \widehat{V}_{ff,s}^{-\frac{1}{2}} \overline{Z}' Yb$$

where $\widehat{\Gamma}_{\overline{Z}X} = \frac{\overline{Z}'X}{T}$. Similar to the derivation of the weak instruments of Subsection 7.1.2, we define $\bar{L}$ matrix such that:

$$\bar{L}\bar{L}' = \mathrm{M}_{\widehat{V}_{ff,s}^{-\frac{1}{2}} \widehat{\Gamma}_{\overline{Z}X}}, \text{ and } \bar{L}'\bar{L} = \mathrm{I}_{(2k_z - k_x)}.$$

The split-sample S statistic under $H_0 : \theta = \theta_0$ at a fixed date s is

$$S(\theta_0; s) = \frac{1}{T} \underbrace{(Yb)' \overline{Z} \widehat{V}_{ff,s}^{-\frac{1}{2}} \overline{L}}_{\widehat{\bar{\xi}}_T} \underbrace{\overline{L}' \widehat{V}_{ff,s}^{-\frac{1}{2}} \overline{Z}' Yb}_{\widehat{\bar{\xi}}_T}.$$

The average and exponential S statistics are defined as:

$$\begin{aligned}
\text{ave-}S_T(\theta_0) &= \frac{1}{t_u - t_l + 1} \sum_{t_b = t_l}^{t_u} S(\theta_0, \frac{t_b}{T}) \\
\text{exp-}S_T(\theta_0) &= \log \left(\frac{1}{t_u - t_l + 1} \sum_{\tau = t_l}^{t_u} \exp \left[-.5 \times S(\theta_0, \frac{t_b}{T}) \right] \right)
\end{aligned}$$

where $t_l = [.15T]$, and $t_u = [.85T]$. The stability parts of the ave-S and exp-S statistics are computed, respectively, as

$$\begin{aligned}
\text{ave-}\widetilde{S}_T(\theta_0) &= \text{ave-}S_T(\theta_0) - S_T(\theta_0) \\
\text{exp-}\widetilde{S}_T(\theta_0) &= \text{exp-}S_T(\theta_0) - S_T(\theta_0).
\end{aligned}$$

Computation of split-sample statistics Define the $(2k_z - k_x) \times p$ Jacobian matrix $\widehat{\bar{q}}_T$ as:

$$\widehat{\bar{q}}_T = \bar{L}' \widehat{V}_{ff,s}^{-\frac{1}{2}} \overline{Z}' Y \nabla_{\theta_0} b$$

The estimators of the variance-covariance matrix of $T^{-\frac{1}{2}} \left(\widehat{\bar{\xi}}_T, \text{vec}\left(\widehat{\bar{q}}_T\right) \right)$ are:

$$\begin{aligned}
\widehat{V}_{\bar{\xi}\bar{\xi}} &= I_{(2k_z - k_x)} \\
\widehat{V}_{\bar{q}\bar{\xi}} &= \left(\nabla_{\theta_0} b' \otimes \bar{L}' \widehat{V}_{ff,s}^{-\frac{1}{2}} \right) \widehat{\Sigma}_s \left(b \otimes \widehat{V}_{ff,s}^{-\frac{1}{2}} \bar{L} \right) = \widehat{V}_{\bar{\xi}\bar{q}}' \\
\widehat{V}_{\bar{q}\bar{q}} &= \left(\nabla_{\theta_0} b' \otimes \bar{L}' \widehat{V}_{ff,s}^{-\frac{1}{2}} \right) \widehat{\Sigma}_s \left(\nabla_{\theta_0} b \otimes \widehat{V}_{ff,s}^{-\frac{1}{2}} \bar{L} \right)
\end{aligned}$$

We compute the $\widehat{\bar{D}}_T$ statistic and its variance matrix as, respectively:

$$\widehat{\bar{D}}_T = \frac{1}{T} \text{mat} \left(\text{vec}\left(\widehat{\bar{q}}_T\right) - \widehat{V}_{\bar{q}\bar{\xi}} \widehat{\bar{\xi}}_T \right), \quad \text{and} \quad \widehat{V}_{\bar{q}\bar{q}.\bar{\xi}} = \widehat{V}_{\bar{q}\bar{q}} - \widehat{V}_{\bar{q}\bar{\xi}} \widehat{V}_{\bar{q}\bar{\xi}}'.$$

The estimate of the break date $\widehat{t}_b$ is the solution of the following maximization problem:

$$\widehat{t}_b = \underset{t_b \in [t_l, t_u]}{\arg\max} \quad \text{vec}\left(\widehat{\bar{D}}_T\right)' \widehat{V}_{\bar{q}\bar{q}.\bar{\xi}}^{-1} \text{vec}\left(\widehat{\bar{D}}_T\right).$$

The split-sample KLM, JKLM and CLR statistics are computed using the formulas given in (S–29), with $\widehat{\bar{\xi}}_T$ and $\widehat{\bar{D}}_T$ in place of $\widehat{\xi}_T$ and $\widehat{D}_T$, and the $rk(\theta_0)$ statistic evaluated using the split sample instruments $\overline{Z} = \overline{Z}\left(\hat{s}\right)$, where $\hat{s} = \frac{\widehat{t}_b}{T}$.

7.1.3 Empirical results

Confidence sets at the 90% and 95% level for the deep structural parameters (α, ϱ) in (S–24) are constructed by inverting the various weak-identification robust tests, and they are plotted in Figures 1-3 for the sample 1966q1-2010q4 and Figures 4-6 for the sample 1984q1-2010q4.

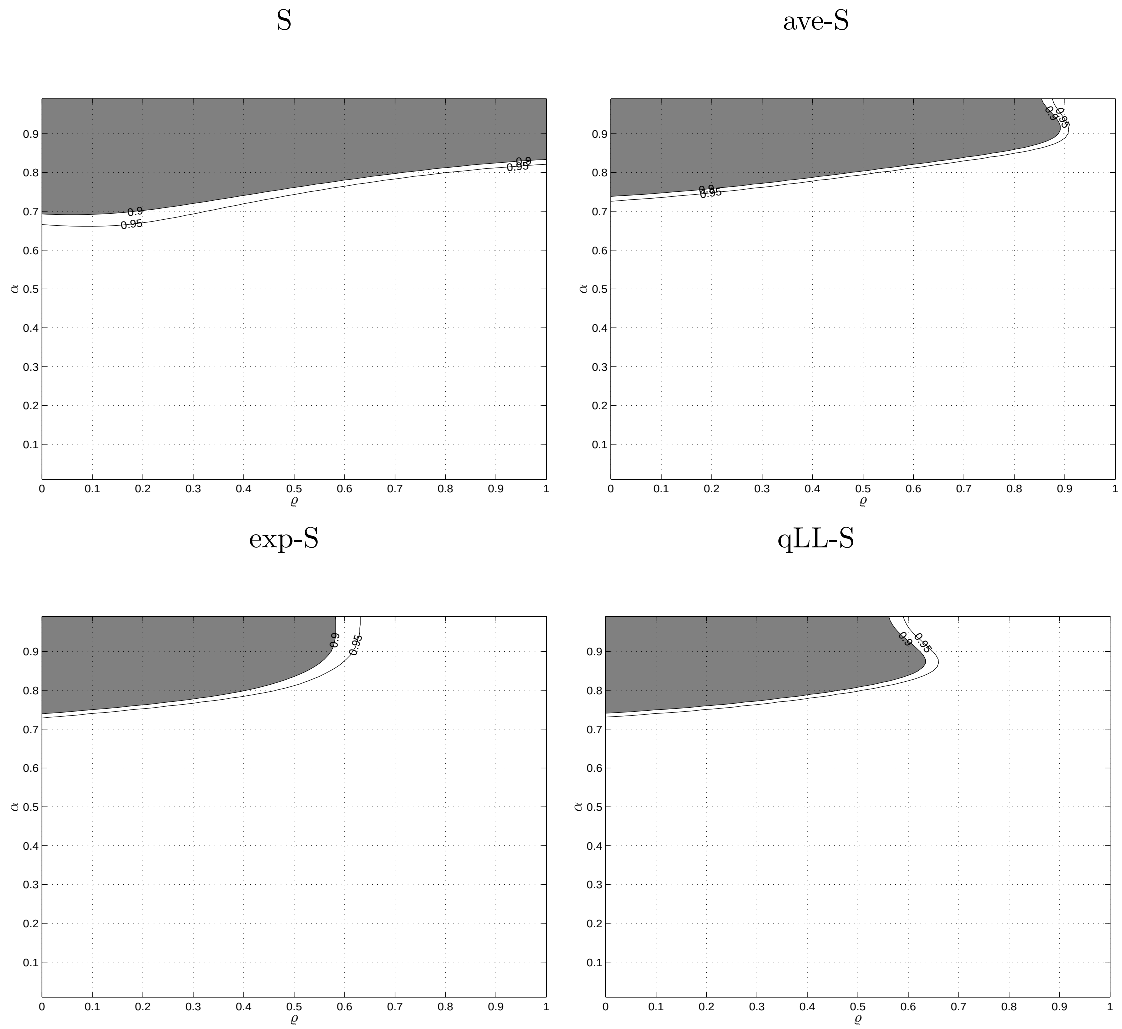

Figure 1: 90% and 95% S and qLL/exp/ave-S confidence sets for α and ϱ in the NKPC: $\varrho\Delta\pi_t = E\left(\Delta\pi_{t+1}|\mathcal{I}_t\right) + \frac{(1-\alpha)^2}{\alpha}\widetilde{x}_t + \kappa + \epsilon_t$, $\widetilde{x}_t$ is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1966q1-2010q4.

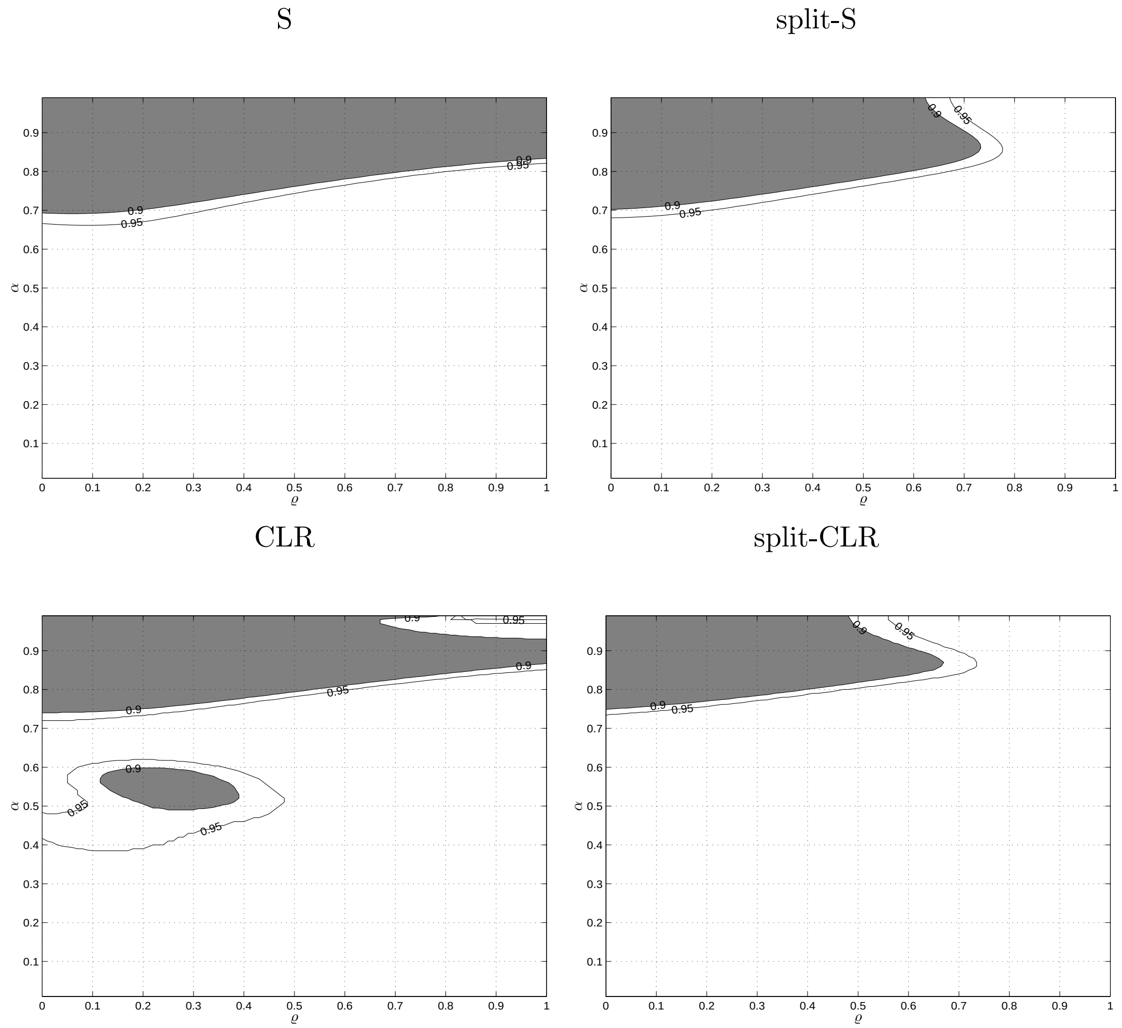

Figure 2: 90% and 95% S, split-S, CLR and split-CLR confidence sets for α and ϱ in the NKPC: $\varrho\Delta\pi_t = E\left(\Delta\pi_{t+1}|\mathcal{I}_t\right) + \frac{(1-\alpha)^2}{\alpha}\widetilde{x}_t + \kappa + \epsilon_t$, $\widetilde{x}_t$ is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1966q1-2010q4.

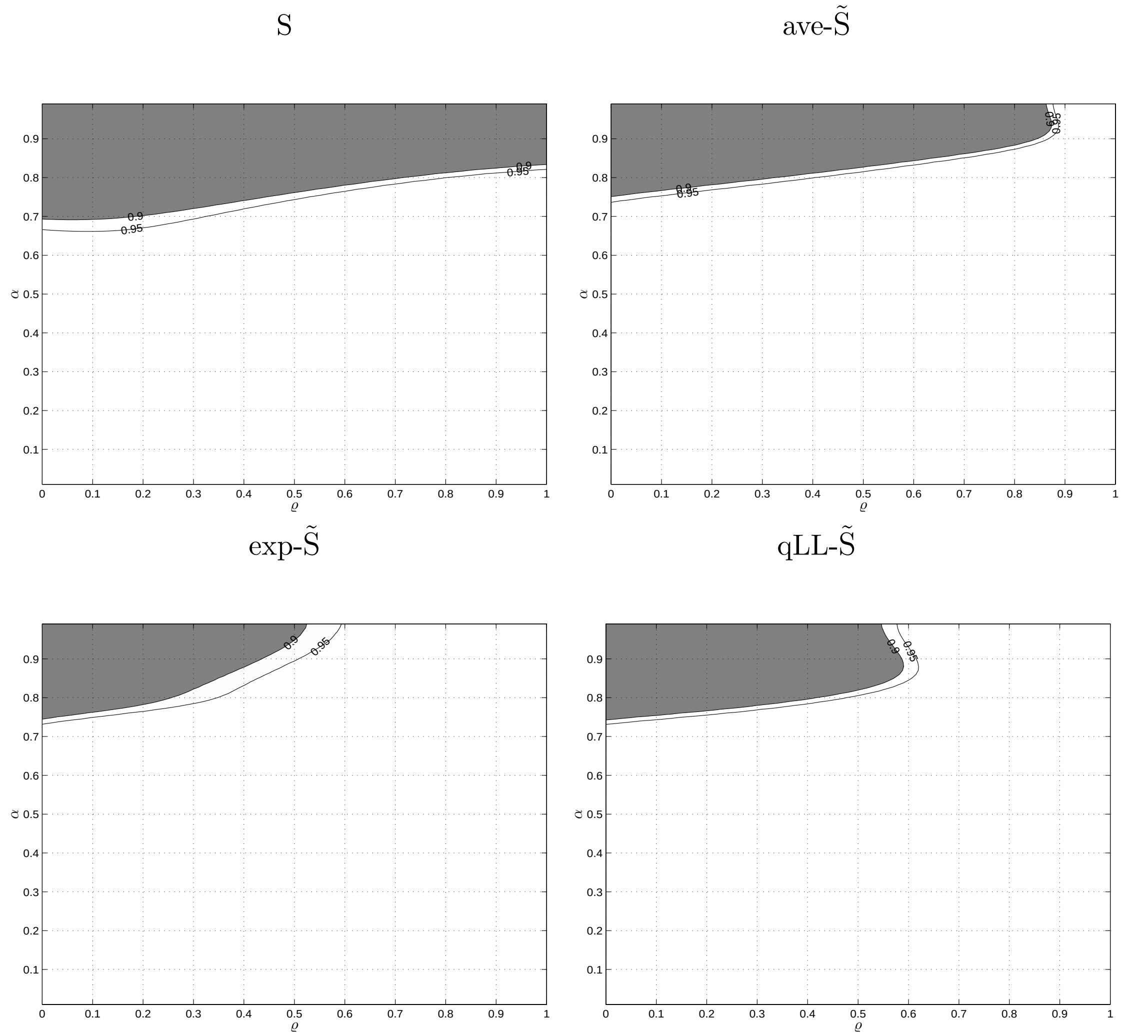

Figure 3: 90% and 95% S and qLL/exp/ave-S̃ confidence sets for α and ϱ in the NKPC: $\varrho\Delta\pi_t = E\left(\Delta\pi_{t+1}|\mathcal{I}_t\right) + \frac{(1-\alpha)^2}{\alpha}\widetilde{x}_t + \kappa + \epsilon_t$, $\widetilde{x}_t$ is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1966q1-2010q4.

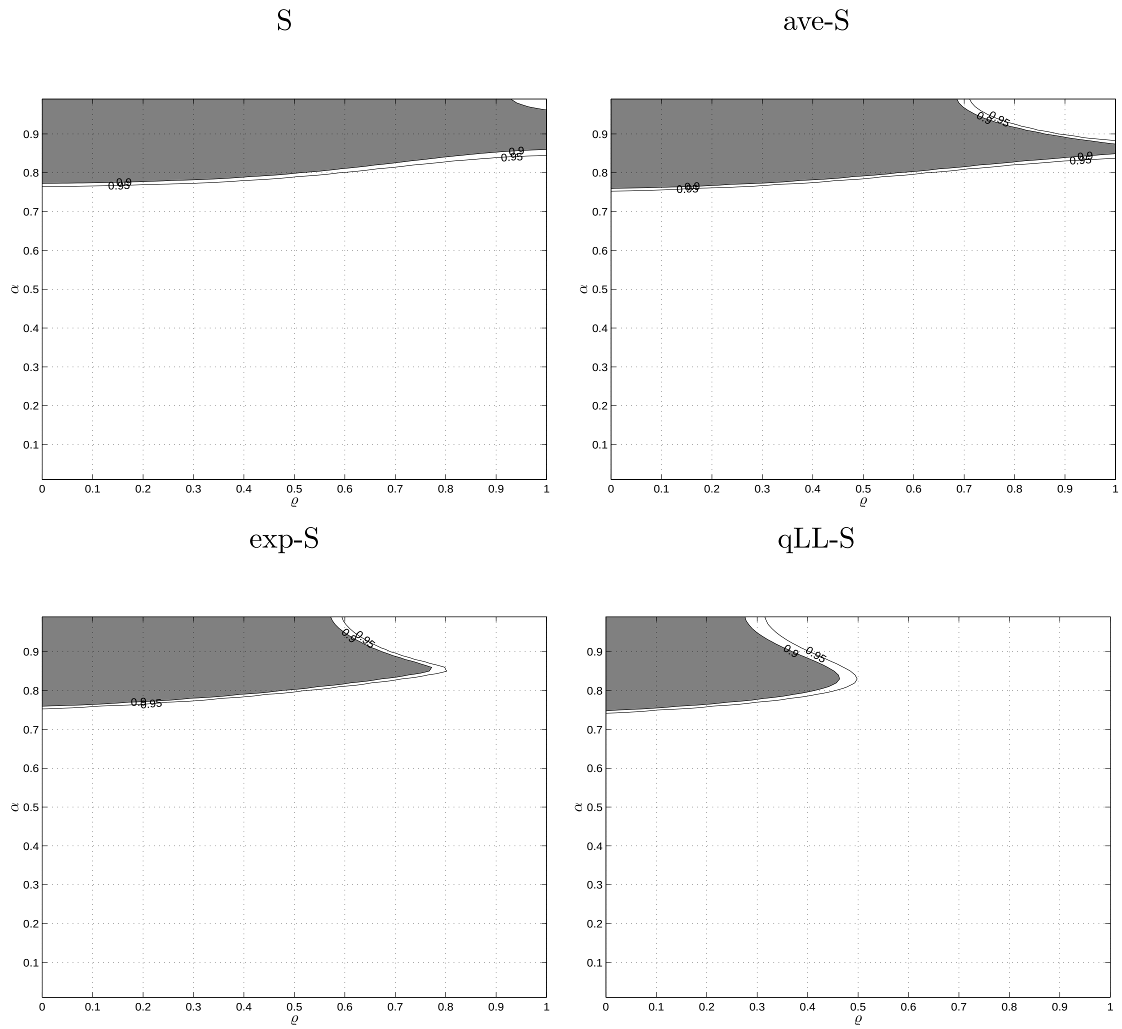

Figure 4: 90% and 95% S and qLL/exp/ave-S confidence sets for α and ϱ in the NKPC: $\varrho\Delta\pi_t = E\left(\Delta\pi_{t+1}|\mathcal{I}_t\right) + \frac{(1-\alpha)^2}{\alpha}\widetilde{x}_t + \kappa + \epsilon_t$, $\widetilde{x}_t$ is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1984q1-2010q4.

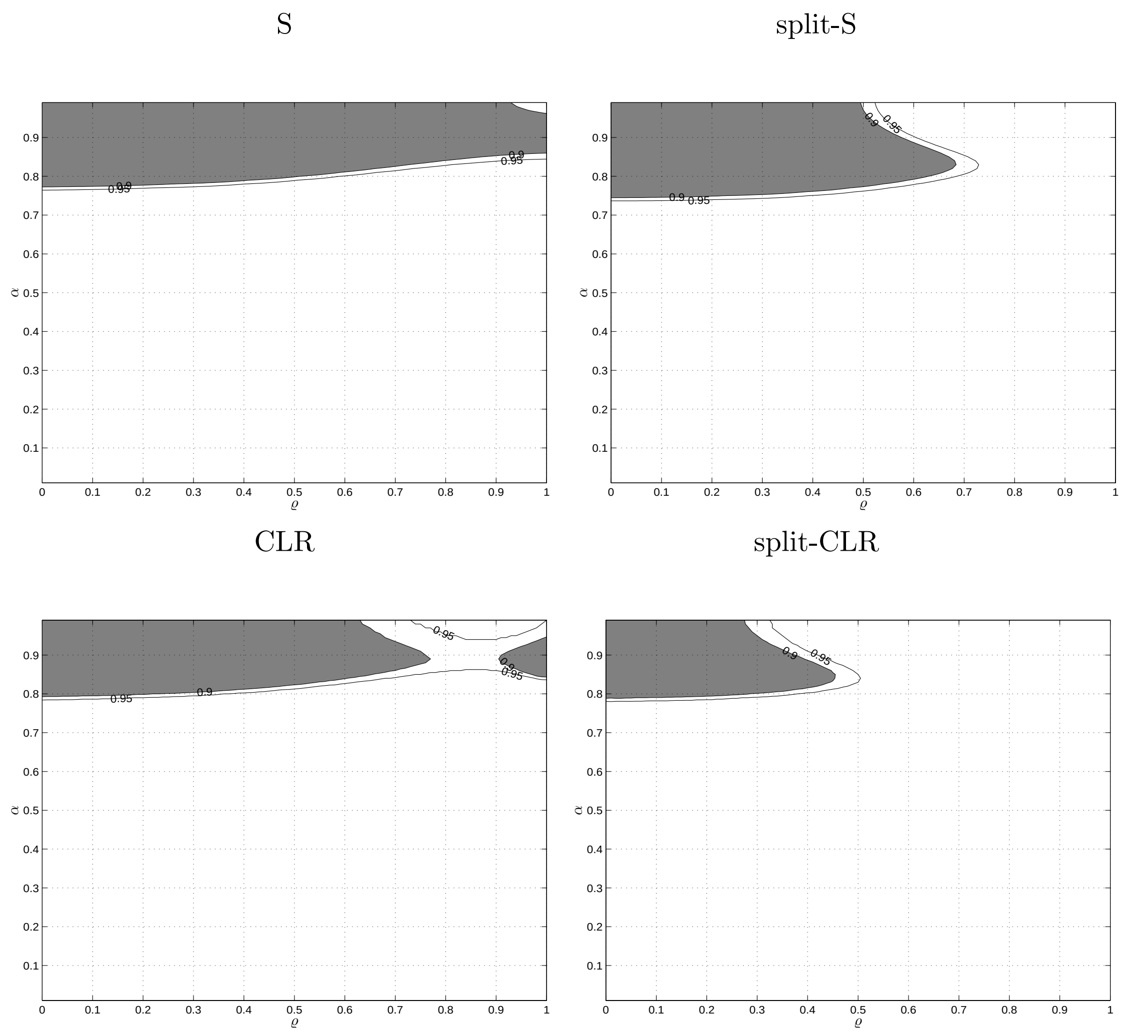

Figure 5: 90% and 95% S, split-S, CLR and split-CLR confidence sets for α and ϱ in the NKPC: $\varrho\Delta\pi_t = E\left(\Delta\pi_{t+1}|\mathcal{I}_t\right) + \frac{(1-\alpha)^2}{\alpha}\widetilde{x}_t + \kappa + \epsilon_t$, $\widetilde{x}_t$ is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1984q1-2010q4.

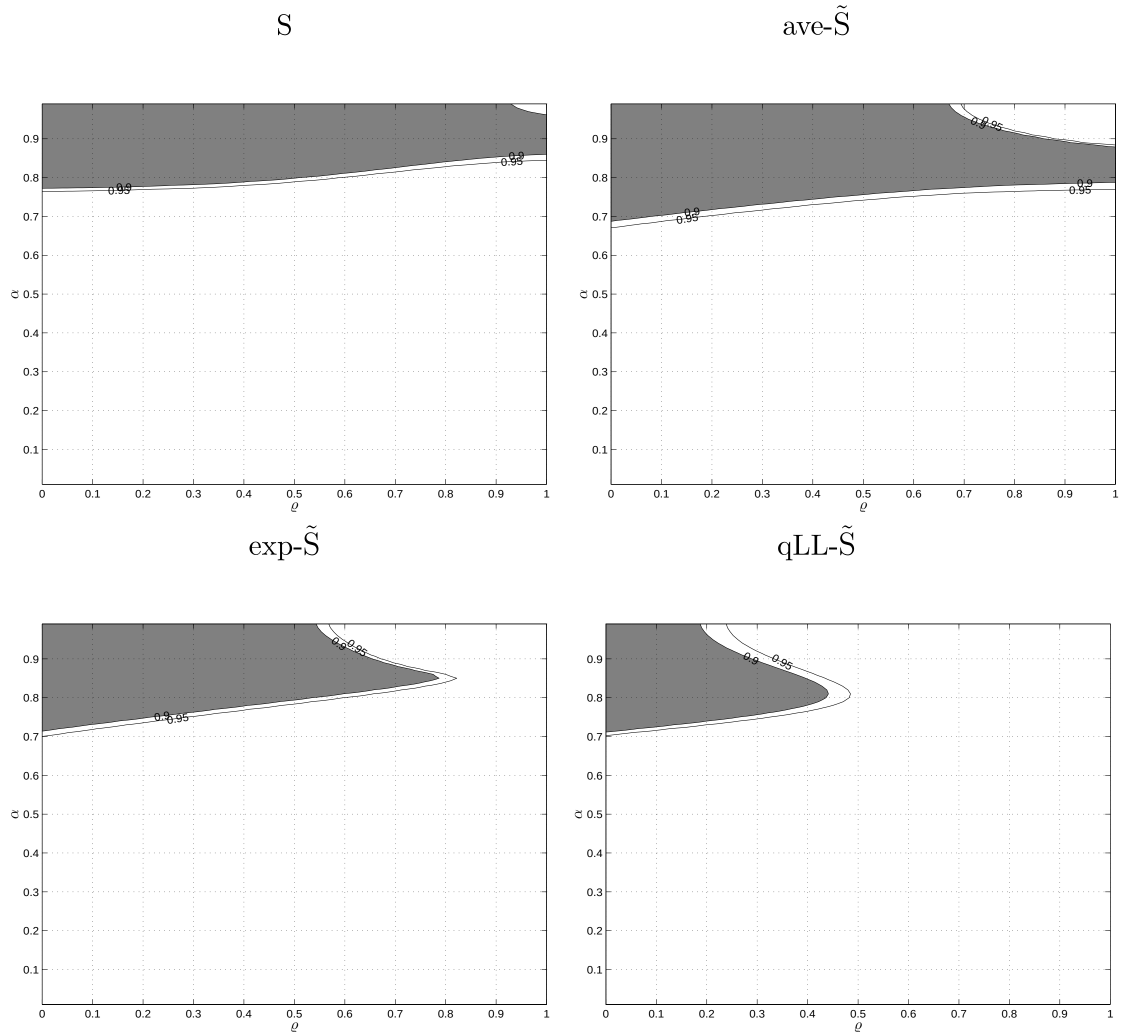

Figure 6: 90% and 95% S and qLL/exp/ave-S̃ confidence sets for α and ϱ in the NKPC: $\varrho \Delta\pi_t = E\left(\Delta\pi_{t+1}|\mathcal{I}_t\right) + \frac{(1-\alpha)^2}{\alpha}\widetilde{x}_t + \kappa + \epsilon_t$, $\widetilde{x}_t$ is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1984q1-2010q4.

7.2 The NKPC with Autocorrelated Errors

Suppose the error term ε_t in (S–24) follows

$$\varepsilon_t = \phi\varepsilon_{t-1} + v_t,$$

where v_t satisfies $E\left(v_t|\mathcal{I}_{t-1}\right) = 0$. Lagging equation (S–24) by one period, multiplying by ϕ and subtracting from (S–24) yields:

$$\begin{aligned}\varrho\left(\Delta\pi_t - \phi\Delta\pi_{t-1}\right) &= (1-\phi)\,\kappa + E\left(\Delta\pi_{t+1}|\mathcal{I}_t\right) - \phi E\left(\Delta\pi_t|\mathcal{I}_{t-1}\right) \\ &\quad + \frac{(1-\alpha)^2}{\alpha}\left(\widetilde{x}_t - \phi\widetilde{x}_{t-1}\right) + v_t. \end{aligned} \tag{S–37}$$

Define $\nu_t = \epsilon_t - \phi\epsilon_{t-1}$, where ϵ_t is given in (S–25). Equation (S–37) and $E\left(v_t|\mathcal{I}_{t-1}\right) = 0$ imply $E\left(\nu_t|\mathcal{I}_{t-1}\right) = 0$. The unconditional moment conditions are $E\left[Z_t'\nu_t\right] = 0$. The structural parameter vector includes ϕ and the confidence sets are three-dimensional (the constant is unrestricted and concentrated out as before).

7.2.1 Computational details

The empirical moment condition is $\frac{1}{T}\sum_{t=1}^{T} Z_t\nu_t$, where $\nu_t = \epsilon_t - \phi\epsilon_{t-1}$ is:

$$\begin{aligned}\nu_t &= \underbrace{\left(\varrho\Delta\pi_t - \Delta\pi_{t+1} - \frac{(1-\alpha)^2}{\alpha}\widetilde{x}_t - \kappa\right)}_{\epsilon_t} - \phi\underbrace{\left(\varrho\Delta\pi_{t-1} - \Delta\pi_t - \frac{(1-\alpha)^2}{\alpha}\widetilde{x}_{t-1} - \kappa\right)}_{\epsilon_{t-1}} \\ &= \underbrace{\varrho\left(\Delta\pi_t - \phi\Delta\pi_{t-1}\right) - \left(\Delta\pi_{t+1} - \phi\Delta\pi_t\right) - \frac{(1-\alpha)^2}{\alpha}\left(\widetilde{x}_t - \phi\widetilde{x}_{t-1}\right)}_{Y_t(\phi)b} - \underbrace{(1-\phi)\,\kappa}_{X_t c}\end{aligned}$$

By defining $Y_t(\phi)$ as $\left(\Delta\pi_t - \phi\Delta\pi_{t-1}, \Delta\pi_{t+1} - \phi\Delta\pi_t, \widetilde{x}_t - \phi\widetilde{x}_{t-1}\right)$, $X_t = 1$ and $b = \left(\varrho, -1, -\frac{(1-\alpha)^2}{\alpha}\right)'$, we rewrite the empirical moments as $Z'\left(Y(\phi)\,b - Xc\right)$. Let $\bar{\theta} = (\varrho, \alpha, \phi)' = (\theta', \phi)'$ The Jacobian of the moment condition after concentrating out c is $Z'\begin{pmatrix} Y(\phi) & Y_1 \end{pmatrix}\nabla_{\bar{\theta}_0}b$ where $\nabla_{\bar{\theta}_0}b = \begin{pmatrix} \nabla_{\theta_0}b & 0 \\ 0 & -b \end{pmatrix}$, $\nabla_{\theta_0}b$ is the derivative of b with respect to "deep" parameters ϱ and α , see Subsection 7.1.2, and Y_1 is the lagged values of matrix Y. By using $\bar{b} = (b, 0)$ in place of b and $\nabla_{\bar{\theta}_0}b$ in place of $\nabla_{\theta_0}b$ and $\begin{pmatrix} Y(\phi) & Y_1 \end{pmatrix}$ in place of Y, we compute the weak instruments, the generalized and split-sample tests following the same steps as described in Subsection 7.1.2.

7.2.2 Empirical results

Three dimensional 95%-level confidence sets for (α, ϱ, ϕ) in (S–37) are constructed by inverting the various weak-identification robust tests, and they are plotted in Figures 7, 8 and 9 for the sample 1966q1-2010q4, and Figures 10 and 11 for the sample 1984q1-2010q4. Table 8 provides the proportion of volume of the 95% confidence region relative to volume of the parallelepiped, and Table 9 provides the point estimates of the structural parameters. The point estimates correspond to the values that minimize the S statistic.

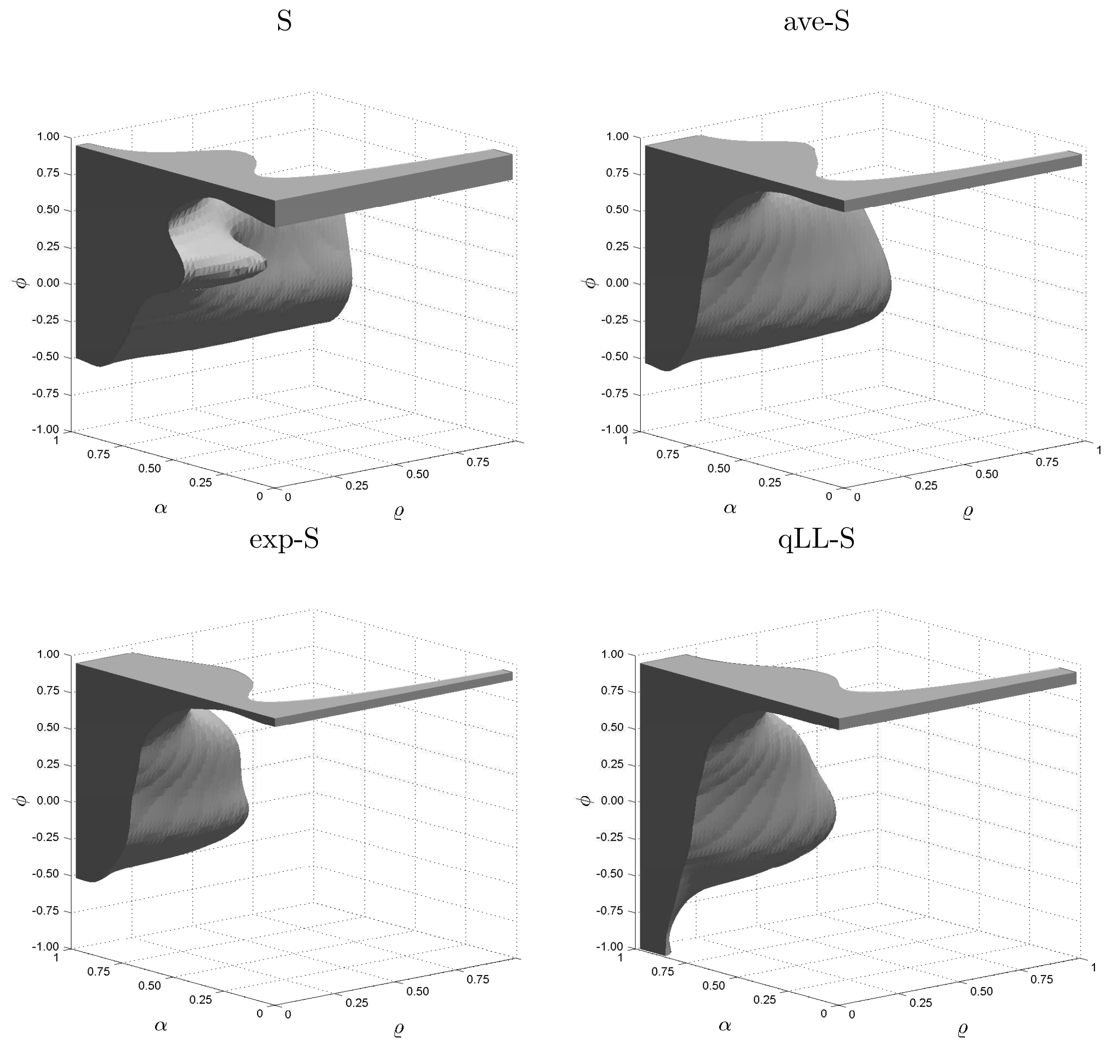

Figure 7: 95% S and qLL/exp/ave-S confidence sets for (α, ϱ, ϕ) in the NKPC with autocorrelated errors. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1966q1-2010q4.

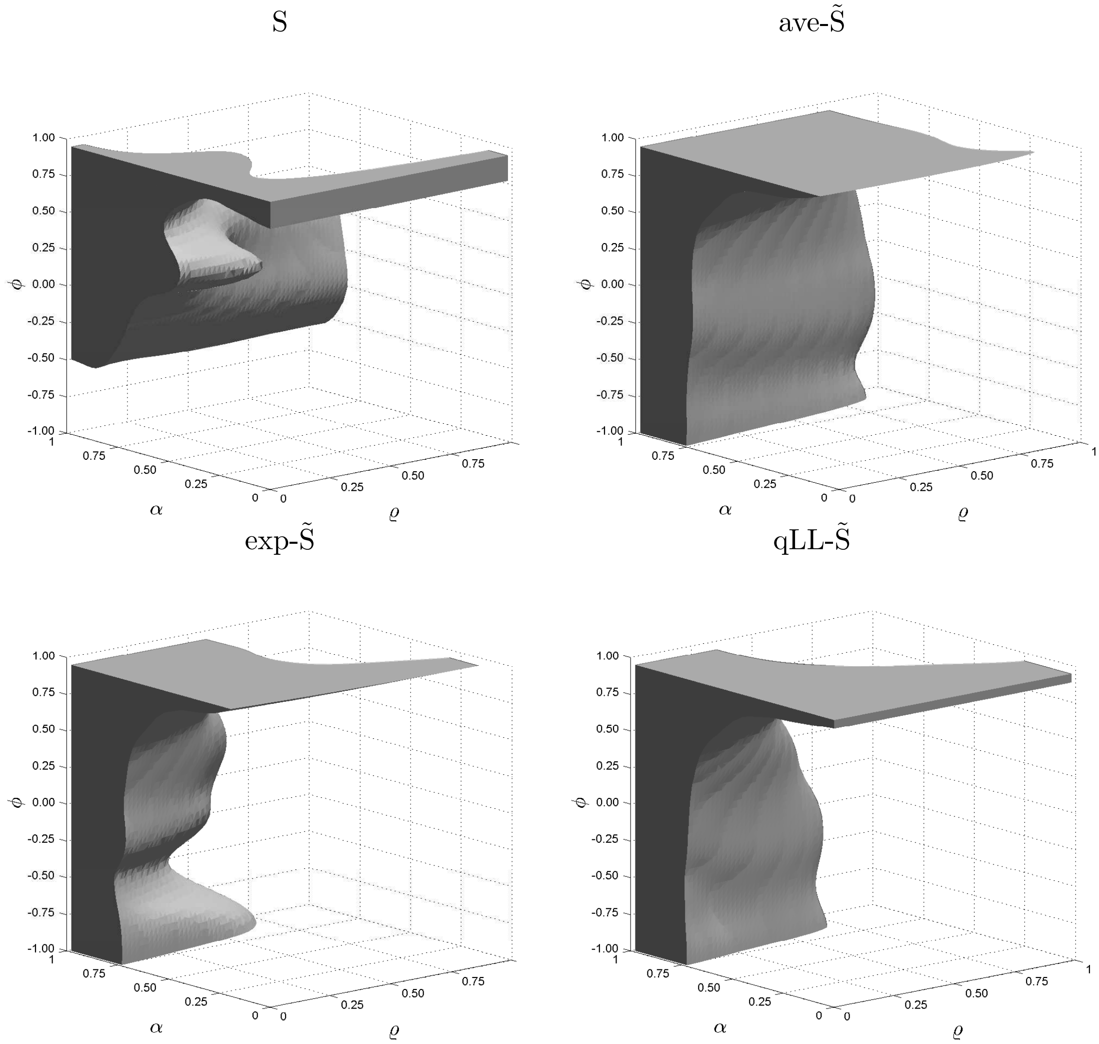

Figure 8: 95%-level S and qLL/exp/ave-S̃ confidence sets for (α, ϱ, ϕ) in the NKPC with autocorrelated errors. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1966q1-2010q4.

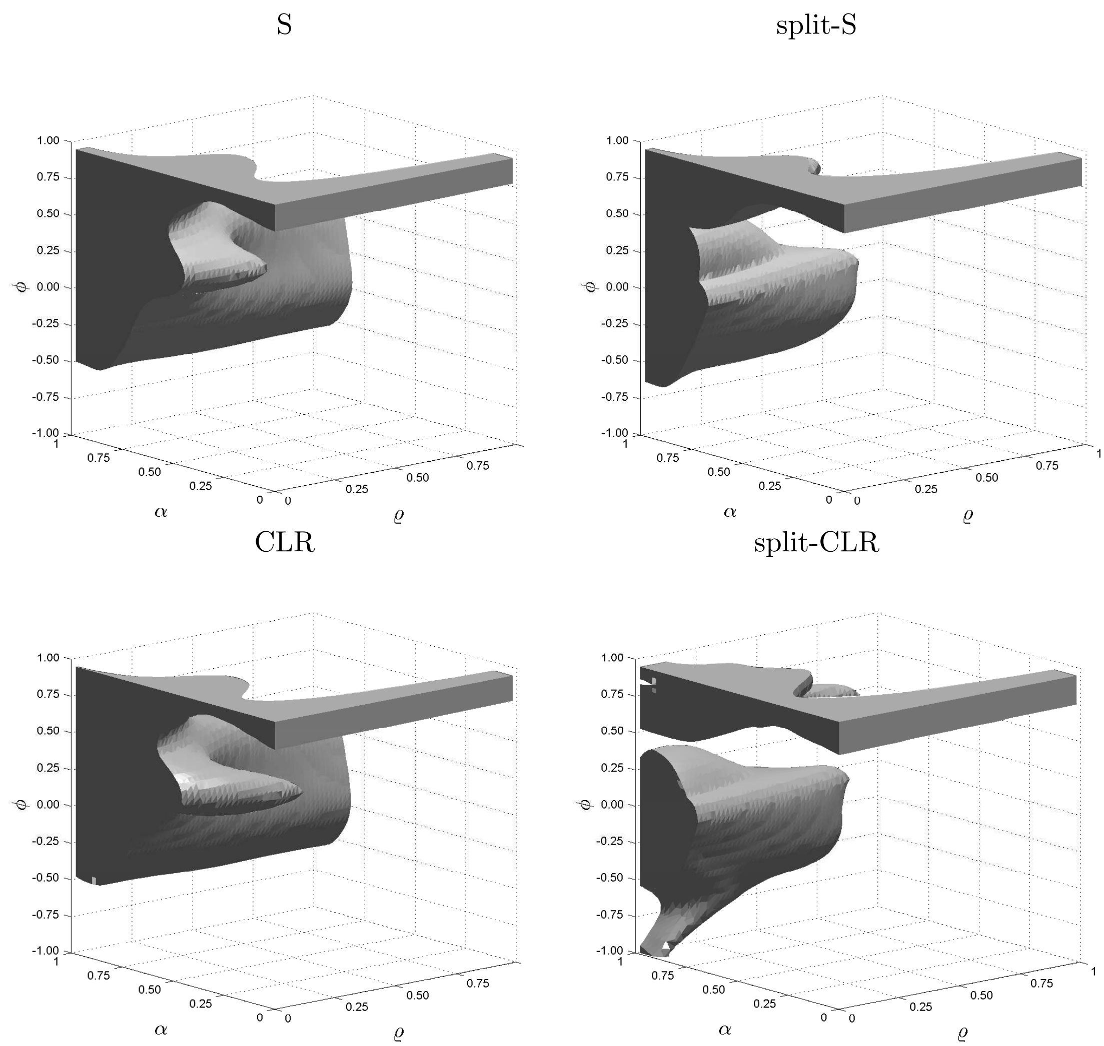

Figure 9: 95%-level S, split-S, CLR and split-CLR confidence sets for (α, ϱ, ϕ) in the NKPC with autocorrelated errors. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1966q1-2010q4.

7.3 The NKPC with Trend Inflation

This section describes how to deal with potentially time-varying trend inflation in the NKPC, using the model of Cogley and Sbordone (2008). The idea is to treat variation in trend inflation as relatively moderate in the sense of Stock and Watson (1998), i.e.,

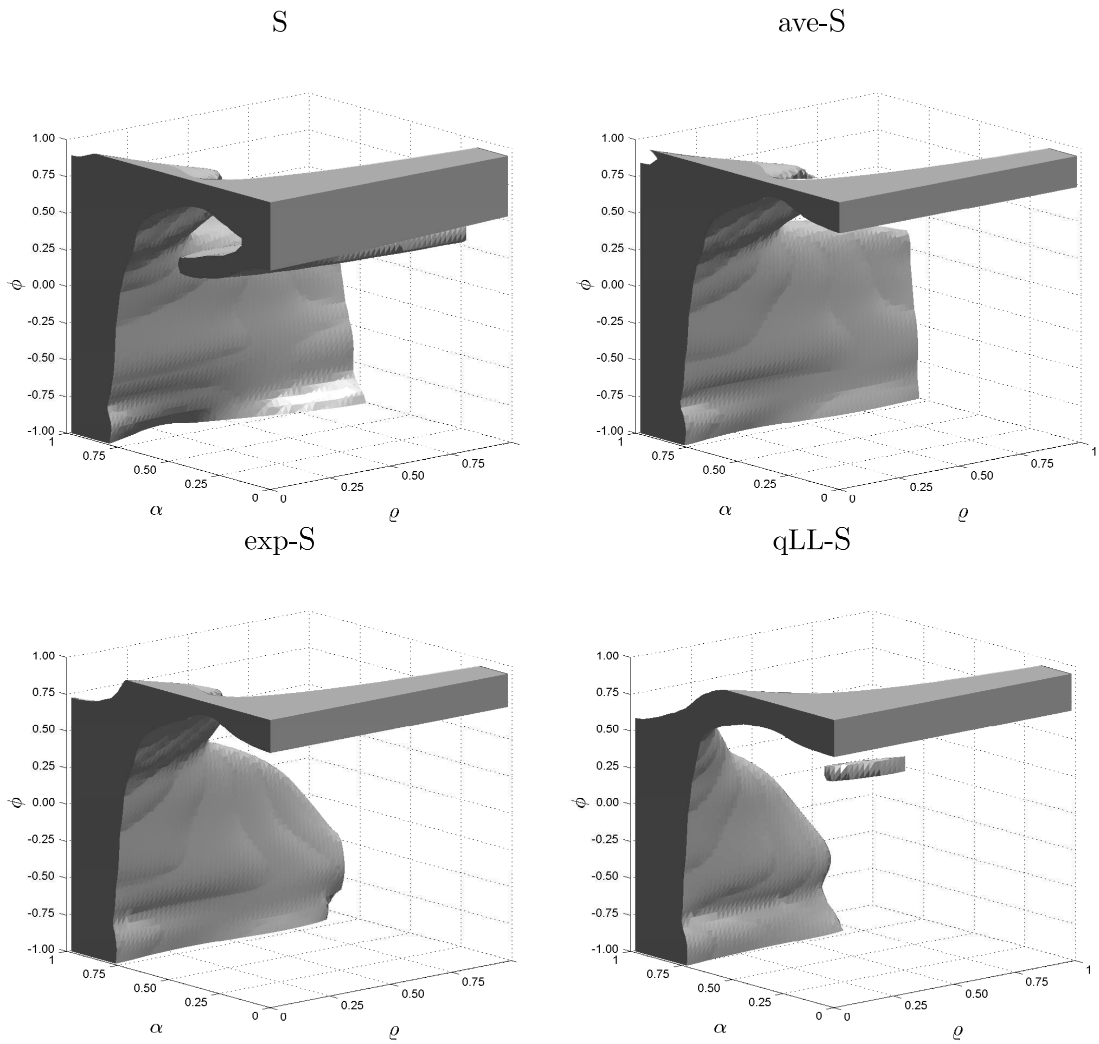

Figure 10: 95%-level S and qLL/exp/ave-S confidence sets for (α, ϱ, ϕ) in the NKPC with autocorrelated errors. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1984q1-2010q4.

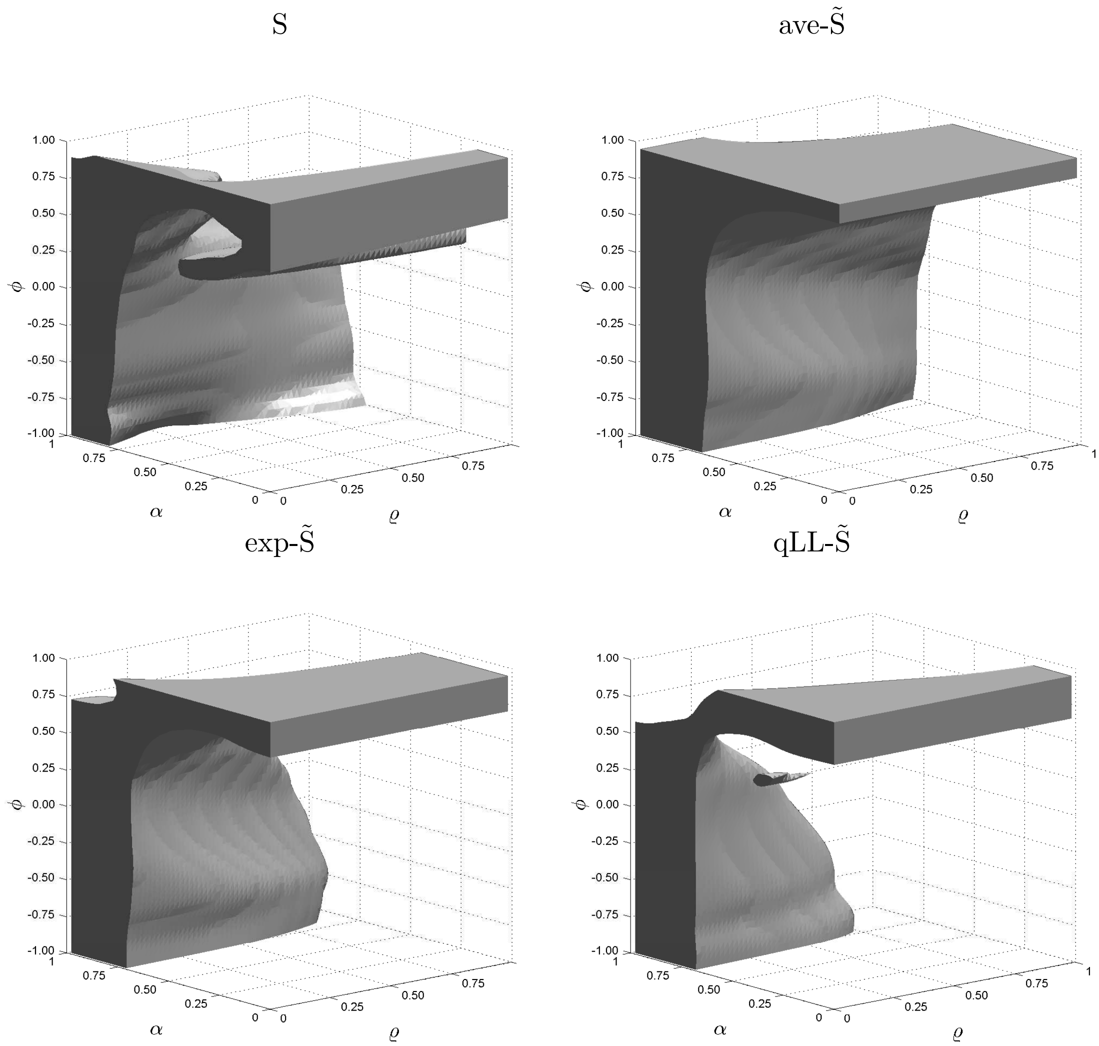

Figure 11: 95%-level S and qLL/exp/ave-$\tilde{\mathrm{S}}$ confidence sets for (α, ϱ, ϕ) in the NKPC with autocorrelated errors. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1984q1-2010q4.

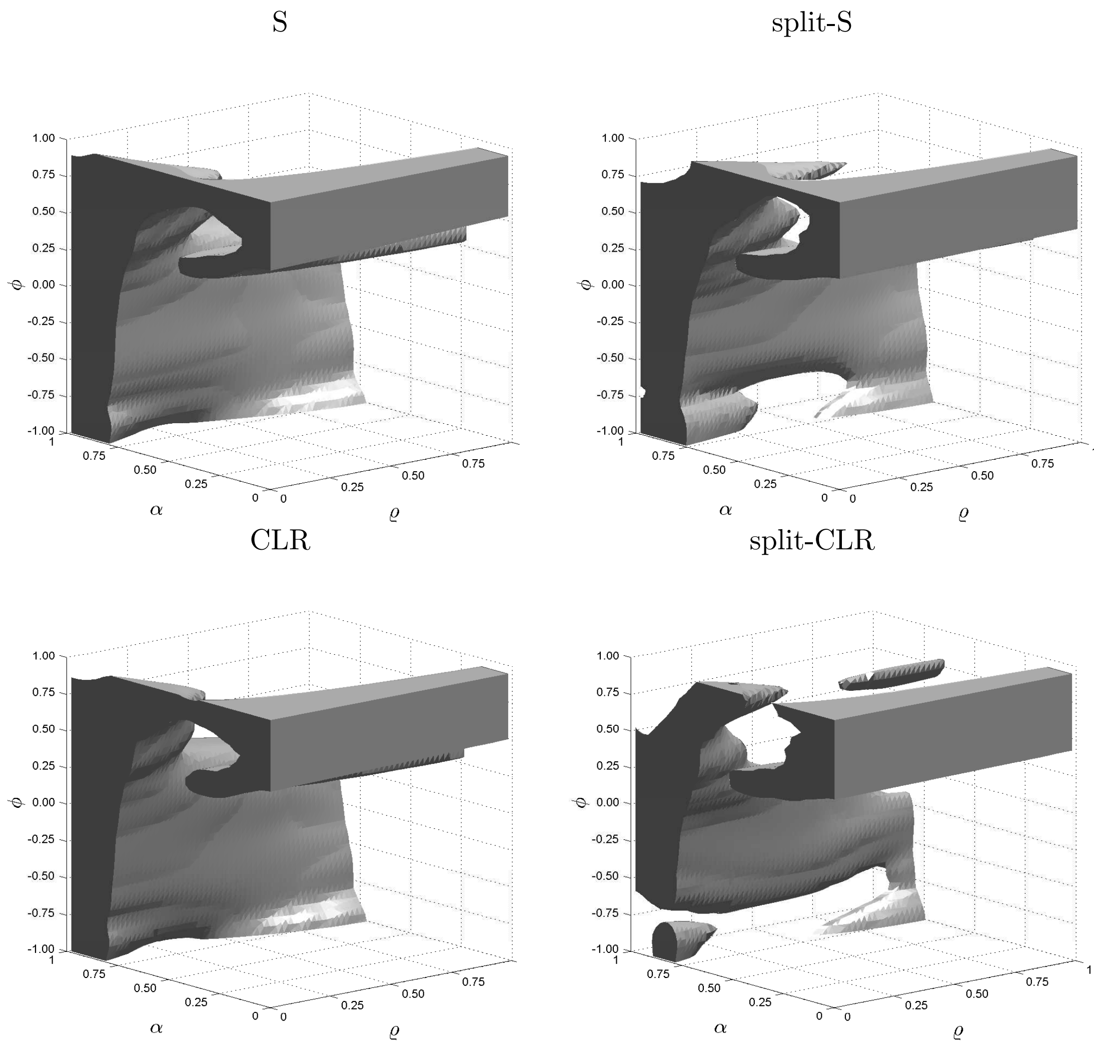

Figure 12: 95%-level S, split-S, CLR and split-CLR confidence sets for (α, ϱ, ϕ) in the NKPC with autocorrelated errors. Instruments: constant, two lags of $\Delta\pi$ and three lags of $\widetilde{x}_t$. Period: 1984q1-2010q4.

	1966q1 - 2010q4		1984q1 - 2010q4	
Confidence Regions	95%	90%	95%	90%
S	17.55	14.10	18.95	15.08
CLR	15.92	12.15	18.03	13.77
ave-S	13.19	11.30	16.88	14.55
exp-S	9.66	7.71	13.71	11.59
qLL-S	11.14	9.49	10.60	8.79
ave-$\tilde{S}$	20.33	17.50	30.99	26.73
exp-$\tilde{S}$	13.71	10.57	21.05	15.84
qLL-$\tilde{S}$	15.94	13.71	14.27	11.23
split-S	11.46	8.39	15.45	10.99
split-CLR	10.75	8.09	11.77	8.63

Table 8: Volume of Confidence Regions as a proportion of the Volume of the Parallelepiped $(\alpha, \varrho, \phi) \in [0.01, 0.99] \times [0, 1] \times [-0.99, 0.99]$

1966q1 - 2010q4			1984q1 - 2010q4		
α	ϱ	ϕ	α	ϱ	ϕ
0.85	0.27	0.00	0.89	0.29	-0.25

Table 9: Point estimates - New Keynesian Phillips Curve with Autoregressive Errors

$O_p\left(T^{-1/2}\right)$. This is the type of 'partial instability' discussed in Li and Mueller (2009).

7.3.1 The model of Cogley and Sbordone (2008)

Cogley and Sbordone (2008) obtain a log-linear approximation of the firms' optimizing conditions around a time-varying trend inflation. The necessary steps are given in the Appendix of their paper. We rewrite Cogley and Sbordone (2008, Eqs. (7), (46) and (47)) translated to our notation that are sufficient to write down a NKPC with time-varying trend inflation:

$$\hat{\pi}_t - \varrho\hat{\pi}_{t-1} = \lambda_t\widehat{mc}_t + \beta_t\left[E\left(\hat{\pi}_{t+1}|\mathcal{I}_t\right) - \varrho\hat{\pi}_t\right] - \varrho\Delta\bar{\pi}_t + \gamma_t\hat{\mathrm{D}}_t + \varepsilon_t \qquad \text{(CS 46)}$$

$$\hat{\mathrm{D}}_t = \varphi_{1t}E\left[\hat{r}_{t,t+1} + \Delta\hat{y}_{t+1} + (\mu-1)^{-1}\left(\hat{\pi}_{t+1} - \varrho\hat{\pi}_t\right) + \hat{\mathrm{D}}_{t+1}|\mathcal{I}_t\right], \qquad \text{(CS 47)}$$

$$(1-\alpha)^{(a-\mu)/a}\,\mu\overline{mc}_t = \left(1-\alpha\overline{\Pi}_t^{(1-\rho)/(\mu-1)}\right)^{(a-\mu)/a}\frac{1-\alpha\overline{\Pi}_t^{(1-\varrho)\mu/(a\mu-a)}}{1-\alpha\overline{\Pi}_t^{(1-\varrho)(\mu-1)}}, \qquad \text{(CS 7)}$$

where $\bar{\pi}_t = \ln\overline{\Pi}_t$, $\overline{\Pi}_t$ is gross trend inflation, $r_{t,t+1} = \pi_{t+1} - i_t \approx \ln R_{t,t+1}$, $R_{t,t+1} = \left(\frac{P_{t+1}}{P_t}\right)\times\frac{1}{1+i_t}$, i_t is the nominal interest rate, $\Delta y_{t+1} = \Delta\ln\left(Y_{t+1}\right)$ and Y_t is real output. Hatted variables indicate stationary log-deviations of variables from their steady state or trend, i.e., $\hat{\pi}_t = \pi_t - \bar{\pi}_t$, $\widehat{mc}_t = \ln\left(mc_t\right) - \ln\left(\overline{mc}_t\right)$ and $\hat{r}_{t,t+1} + \Delta\hat{y}_{t+1} = r_{t,t+1} + \Delta y_{t+1} - \ln\overline{R_{t,t+1}\frac{Y_{t+1}}{Y_t}}$. The term $R_{t,t+1}\frac{Y_{t+1}}{Y_t}$ is the effective discount factor from period t to $t+1$, whose steady-state value $\overline{R_{t,t+1}\frac{Y_{t+1}}{Y_t}}$ was denoted by β in the canonical constant-parameter NKPC. We maintain the assumption $\beta = 1$ that we used earlier, which implies that $\hat{r}_{t,t+1} + \Delta\hat{y}_{t+1} = r_{t,t+1} + \Delta y_{t+1}$ in (CS 47). The time-varying parameters $\lambda_t, \beta_t, \gamma_t$ and φ_{1t} are functions of $\bar{\pi}_t$, and they are given by

$$\begin{array}{lll} \lambda_t = \chi_t\left(1-\varphi_{2t}\right) & \beta_t = \overline{\Pi}_t^{\frac{(1-\varrho)}{\upsilon}} & \gamma_t = \chi_t\left(\overline{\Pi}_t^{\frac{(1-\varrho)}{\upsilon}} - 1\right) \\ \chi_t = \upsilon\left(\frac{1-\alpha\overline{\Pi}_t^{\frac{(1-\varrho)}{(\mu-1)}}}{\alpha\overline{\Pi}_t^{\frac{(1-\varrho)}{(\mu-1)}}}\right) & \varphi_{1t} = \alpha\overline{\Pi}_t^{\frac{(1-\varrho)}{(\mu-1)}} & \varphi_{2t} = \varphi_{1t}\overline{\Pi}_t^{\frac{(1-\varrho)}{\upsilon}}. \end{array} \qquad \text{(S–38)}$$

Algebraic manipulations of equations (CS 46) and (CS 47), similar to those in Cogley and Sbordone (2008, Appendix A), yield the specification (39) in the paper, with

$$\zeta_t = \lambda_t / \Delta_t \qquad \rho_t = \varrho / \Delta_t \qquad b_{1t} = \tfrac{\beta_t + \gamma_t (1-\varrho\varphi_{1t})\varphi_{1t}/(\mu-1)}{\Delta_t}$$

$$b_{2t} = \tfrac{\gamma_t (1-\varrho\varphi_{1t})/(\mu-1)}{\Delta_t} \qquad b_{3t} = \gamma_t \varphi_{1t}/\Delta_t \qquad \Delta_t = 1 + \varrho\left(\beta_t + \tfrac{\gamma_t\varphi_{1t}}{\mu-1}\right).$$

When the true model is given by equations (CS 46), (CS 47) and (S–38), the baseline specification (S–24) is misspecified in two ways. First, it omits the term $\hat{\mathrm{D}}_t$, which involves an infinite stream of future inflation, real interest rates and real output growth. Since these variables are correlated with predetermined instruments, this would lead to a violation of the identifying restrictions $E\left(Z_t'\epsilon_t\right) = 0$, so the usual full-sample S test will have power against it. Second, the coefficients of the model are time-varying. So, the stability S tests will have power against this time-variation in the parameters, while the generalized (joint full-sample and stability) S tests have power against both types of misspecification.

7.3.2 A specification with moderate instability in trend inflation

To correct for these two sources of misspecification, it suffices to consider a specification of the model where trend inflation is within a $\sqrt{T}$ neighborhood of zero, because the S tests are consistent for larger instabilities. There is both a theoretical and an empirical motivation for focusing on the neighborhood of zero as opposed to some unknown positive trend inflation. The theoretical motivation is that this approach can be justified by full indexation of non-optimally reset prices to any perfectly predictable long-run inflation target, as in Yun (1996), which is reasonable. The empirical motivation is that the resulting confidence sets are actually very large, so this assumption is not at odds with the data, and relaxing it will most likely make confidence sets even larger. In any case, it is conceptually straightforward to consider deviations from some unknown trend inflation $\bar{\pi}$, at the cost of more complicated algebra and an additional unknown parameter ($\bar{\pi}$).

A general representation of a log-linear approximation of the coefficients in (S–38) around zero trend inflation, $\overline{\Pi} = 1$, is:

$$\varsigma_t = \varsigma + \varsigma_{\bar{\pi}}\bar{\pi}_t + o\left(\bar{\pi}_t\right).$$

The intercept and slope of this approximation for each of the coefficients in (S–38) is:

$$
\begin{array}{lll}
\lambda \; = \frac{(1-\alpha)^2}{\alpha}\upsilon & \beta \; = 1 & \gamma \; = 0 \\
\lambda_{\bar{\pi}} = -\frac{(1-\alpha)}{\alpha}(1-\varrho)\frac{(a+\alpha\mu)}{(\mu-a)} & \beta_{\bar{\pi}} = \frac{(1-\varrho)}{\upsilon} & \gamma_{\bar{\pi}} = \frac{(1-\alpha)(1-\varrho)}{\alpha} \\
 & & \\
\chi \; = \frac{(1-\alpha)}{\alpha}\upsilon & \varphi_1 \; = \alpha & \varphi_2 \; = \alpha \\
\chi_{\bar{\pi}} = -\frac{(1-\varrho)}{\alpha(\mu-1)}\upsilon & \varphi_{1\bar{\pi}} = \alpha\frac{(1-\varrho)}{(\mu-1)} & \varphi_{2\bar{\pi}} = \alpha\frac{\mu(1-\varrho)}{(\mu-1)a}
\end{array}
$$

From eq. (CS 7), we have:

$$
\ln \overline{mc}_t = -\ln \mu - \left(\frac{\mu - a}{a}\right) \ln \left(\frac{1 - \alpha \overline{\Pi}_t^{\frac{(1-\varrho)}{(\mu-1)}}}{1-\alpha} \right) + \ln \left(\frac{1 - \alpha \overline{\Pi}_t^{\frac{\mu(1-\varrho)}{(\mu-1)a}}}{1 - \alpha \overline{\Pi}_t^{\frac{(1-\varrho)}{(\mu-1)}}} \right)
$$

Note that:

$$
\begin{aligned}
\left.\frac{\partial \overline{mc}_t}{\partial \overline{\Pi}_t}\right|_{\overline{\Pi}=1} &= \left(\frac{\mu-a}{a}\right) \frac{\alpha\left(\frac{(1-\varrho)}{\mu-1}\right)}{(1-\alpha)} - \frac{\alpha\frac{\mu(1-\varrho)}{(\mu-1)a}}{(1-\alpha)} + \frac{\alpha\frac{(1-\varrho)}{(\mu-1)}}{(1-\alpha)} \\
&= \frac{\alpha(1-\varrho)}{1-\alpha}\left[\frac{(\mu-a)}{a(\mu-1)} - \frac{\mu}{(\mu-1)a} + \frac{1}{(\mu-1)}\right] = 0
\end{aligned}
$$

Therefore,

$$
\ln \overline{mc}_t = -\ln \mu + o_p(\bar{\pi}_t).
$$

Ignoring terms of order smaller than $\bar{\pi}_t$, using $\hat{\pi}_t = \pi_t - \bar{\pi}_t$, $\Delta\bar{\pi}_{t+1} = o_p(\bar{\pi}_t)$ and $\hat{r}_{t,t+1} + \Delta\hat{y}_{t+1} = r_{t,t+1} + \Delta y_{t+1}$ (which follows from the assumption $\beta = 1$), equation (CS 46) can be written as:

$$
\begin{aligned}
\pi_t - \varrho\pi_{t-1} = \lambda\widehat{mc}_t + E(\pi_{t+1}|\mathcal{I}_t) - \varrho\pi_t \\
+ \bar{\pi}_t\left[\lambda_{\bar{\pi}}\widehat{mc}_t + \beta_{\bar{\pi}}E(\pi_{t+1} - \varrho\pi_t|\mathcal{I}_t) + \gamma_{\bar{\pi}}\bar{\mathrm{D}}_t\right] + \varepsilon_t,
\end{aligned} \tag{S–39}
$$

where:

$$
\bar{\mathrm{D}}_t = \alpha\left[E(r_{t,t+1} + \Delta y_{t+1}|\mathcal{I}_t) + (\mu-1)^{-1}E(\pi_{t+1} - \varrho\pi_t|\mathcal{I}_t) + E\left(\bar{\mathrm{D}}_{t+1}|\mathcal{I}_t\right)\right]. \tag{S–40}
$$

Next, we need to remove the infinite terms in $\bar{\mathrm{D}}_t$ from the above equation. Lead

(S–39) one period to get

$$\pi_{t+1} - \varrho\pi_t = \lambda\widehat{mc}_{t+1} + E\left(\pi_{t+2} - \varrho\pi_{t+1}|\mathcal{I}_{t+1}\right) + \bar{\pi}_{t+1}\left[\lambda_{\bar{\pi}}\widehat{mc}_{t+1} + \beta_{\bar{\pi}}E\left(\pi_{t+2} - \varrho\pi_{t+1}|\mathcal{I}_{t+1}\right) + \gamma_{\bar{\pi}}\bar{\mathrm{D}}_{t+1}\right] + \varepsilon_{t+1}$$

Take expectations conditional on $\mathcal{I}_t$, and use the fact that terms in $\Delta\bar{\pi}_{t+1}$ are negligible to get:

$$E\left(\pi_{t+1}|\mathcal{I}_t\right) - \varrho\pi_t = \lambda E\left(\widehat{mc}_{t+1}|\mathcal{I}_t\right) + E\left(\pi_{t+2} - \varrho\pi_{t+1}|\mathcal{I}_t\right) + \bar{\pi}_t E\left[\lambda_{\bar{\pi}}\widehat{mc}_{t+1} + \beta_{\bar{\pi}}\left(\pi_{t+2} - \varrho\pi_{t+1}\right) + \gamma_{\bar{\pi}}\bar{\mathrm{D}}_{t+1}|\mathcal{I}_t\right] \quad \text{(S–41)}$$

Multiply (S–41) by α, subtract from (S–39) and use (S–40) to substitute for $\bar{\mathrm{D}}_t - \alpha E\left(\bar{\mathrm{D}}_{t+1}|\mathcal{I}_t\right)$ to get

$$\pi_t - \varrho\pi_{t-1} = \lambda\widehat{mc}_t + E\left(\pi_{t+1} - \varrho\pi_t|\mathcal{I}_t\right) + \alpha E\left[\pi_{t+1} - \varrho\pi_t - \left(\pi_{t+2} - \varrho\pi_{t+1}\right) - \lambda\widehat{mc}_{t+1}|\mathcal{I}_t\right] + \bar{\pi}_t\left\{\lambda_{\bar{\pi}}\left[\widehat{mc}_t - \alpha E\left(\widehat{mc}_{t+1}|\mathcal{I}_t\right)\right] + \beta_{\bar{\pi}}E\left[\left(\pi_{t+1} - \varrho\pi_t\right) - \alpha\left(\pi_{t+2} - \varrho\pi_{t+1}\right)|\mathcal{I}_t\right] + \gamma_{\bar{\pi}}\alpha E\left[\left(r_{t,t+1} + \Delta y_{t+1}\right) + \left(\mu - 1\right)^{-1}\left(\pi_{t+1} - \varrho\pi_t\right)|\mathcal{I}_t\right]\right\} + \varepsilon_t.$$

Collecting the constant and all the terms in $\bar{\pi}_t$, and substituting $x_t + \ln\mu$ for $\widehat{mc}_t$, where $x_t = \ln\frac{S_t}{a}$ as before, the above can be written more compactly as:

$$\varrho\Delta\pi_t = E\left(\Delta\pi_{t+1}|\mathcal{I}_t\right) + \frac{\left(1-\alpha\right)^2}{\alpha}\upsilon\, x_t + \kappa + \alpha E\left(\varrho\Delta\pi_{t+1} - \Delta\pi_{t+2} - \frac{\left(1-\alpha\right)^2}{\alpha}\upsilon x_{t+1} - \kappa|\mathcal{I}_t\right) + \varpi_t\left(\theta\right)\bar{\pi}_t + \varepsilon_t. \quad \text{(S–42)}$$

where $\kappa = \frac{(1-\alpha)^2}{\alpha}\upsilon\ln\mu$, $\upsilon = \frac{a(\mu-1)}{\mu-a}$, and

$$\varpi_t\left(\theta\right) = \left(1-\alpha\right)\lambda_{\bar{\pi}}\ln\mu + \lambda_{\bar{\pi}}x_t + \beta_{\bar{\pi}}\left(\pi_{t+1} - \varrho\pi_t\right) - \alpha\left[\lambda_{\bar{\pi}}x_{t+1} + \delta_{\bar{\pi}}\left(\pi_{t+2} - \varrho\pi_{t+1}\right)\right] + \alpha\gamma_{\bar{\pi}}\left[\left(r_{t,t+1} + \Delta y_{t+1}\right) + \left(\mu - 1\right)^{-1}\left(\pi_{t+1} - \varrho\pi_t\right)\right]. \quad \text{(S–43)}$$

If we impose $\bar{\pi}_t = 0$, then $E\left(\varrho\Delta\pi_{t+1} - \Delta\pi_{t+2} - \frac{(1-\alpha)^2}{\alpha}\upsilon x_{t+1} - \kappa|\mathcal{I}_t\right) = 0$ and equation (S–42) collapses to the NKPC in (S–24).

We are interested in making inference about the structural parameter vector $\theta = (\alpha, \varrho, \mu)'$. As before, we calibrate the output elasticity to hours of work in the production function to $a = \frac{2}{3}$. The admissible ranges for each of the structural parameter are: $\alpha \in (0,1)$, $\varrho \in [0,1]$, and $\mu \in (1, +\infty)$.

Let $\eta^{\pi}_{t+1} = \pi_{t+1} - E(\pi_{t+1}|\mathcal{I}_t)$ and $\eta^{x}_{t+1} = x_{t+1} - E(x_{t+1}|\mathcal{I}_t)$. Then, substituting for the expectations terms in Equation (S–39), we can define the residual function

$$u_t \equiv \epsilon_t - \alpha\epsilon_{t+1} - \varpi_t(\theta)\bar{\pi}_t = \varepsilon_t - \eta^{\pi}_{t+1} - \alpha\left(\varrho\eta^{\pi}_{t+1} - \eta^{\pi}_{t+2} - \frac{(1-\alpha)^2}{\alpha}\upsilon\eta^{x}_{t+1}\right), \quad \text{(S–44)}$$

where ϵ_t and $\varpi_t(\theta)$ are defined in equations (S–25) and (S–43), respectively. The assumption $E(\varepsilon_t|\mathcal{I}_{t-1}) = 0$ yields $E[u_t|\mathcal{I}_t] = 0$, from which we can obtain unconditional moment restrictions of the form:

$$\begin{aligned} &E[Z_t'u_t] = E[f_t(\theta) - g_t(\theta)\bar{\pi}_t] = 0, \text{ where} \\ &f_t(\theta) = Z_t'(\epsilon_t - \alpha\epsilon_{t+1}), \text{ and } g_t(\theta) = Z_t'\varpi_t(\theta). \end{aligned} \quad \text{(S–45)}$$

7.3.3 Implications of time-varying trend inflation when it is ignored

We now ask what happens when we do inference using the sample moments $T^{-1/2} \sum_{t=1}^{[sT]} f_t(\theta)$, where $f_t(\theta)$ is defined in (S–45), i.e., ignoring the term $g_t(\theta)\bar{\pi}_t$ that is unobserved. This depends on the behavior of $T^{-1/2}\sum_{t=1}^{[sT]} g_t(\theta)\bar{\pi}_t$, which we show is not negligible in general for moderate instability in $\bar{\pi}_t$.

The model's residual u_t, defined in (S–44) satisfies $E(u_t|\mathcal{I}_{t-1}) = 0$ and is at most MA(2). We assume throughout that the instruments Z_t are asymptotically mse stationary (Hansen (2000)), which does not contradict the existence of moderate time-varying trend inflation, and var$(Z_t'u_t) = V_{ff}$ is finite and consistently estimable and the partial sum process $T^{-1/2}\sum_{t=1}^{[sT]} Z_t'u_t$ satisfies a FCLT.

Now, notice that $\varpi_t(\theta)$ in (S–43) can be written as $Y_{\bar{\pi},t}l_{\bar{\pi}}$, where $Y_{\bar{\pi},t}$ is a row vector of data, defined in (S–47), and $l_{\bar{\pi}}$ is a corresponding column vector of functions of the structural parameters θ, given in (S–48). Let $\Gamma_{ZY_{\bar{\pi}}} = \text{plim}\, T^{-1}\sum_{t=1}^{T} Z_t'Y_{\bar{\pi},t}$, and suppose $T^{-1}\sum_{t=1}^{[sT]} Z_t'Y_{\bar{\pi},t} \xrightarrow{p} s\Gamma_{ZY_{\bar{\pi}}}$ (which needs to hold uniformly in s for $Y_{\bar{\pi},t}$ to be 'asymptotically mse stationary'). Suppose also that $\bar{\pi}_t = \frac{1}{\sqrt{T}}h\left(\frac{t}{T}\right)$, where $h(\cdot)$ is a cadlag function with at most a finite number of discontinuities. Then,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t(\theta)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t'u_t+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t'\varpi_t(\theta)\,\bar{\pi}_t$$
$$\xrightarrow{d} V_{ff}^{1/2}\xi+\Gamma_{ZY_{\bar{\pi}}}l_{\bar{\pi}}\int_0^1 h(r)\,dr,$$

where ξ is a standard normal vector and the convergence of the second term follows from Li and Mueller (2009, Lemma 4). Hence, ignoring the terms involving $\bar{\pi}_t$ violates the condition that the full-sample moment vector has mean zero under the null, which is necessary for size control of the full-sample tests. Moreover, the full-sample S test has power against this type of misspecification. If the convergence $T^{-1}\sum_{t=1}^{[sT]}Z_t'Y_{\bar{\pi},t}\xrightarrow{p} s\Gamma_{ZY_{\bar{\pi}}}$ is uniform in $s\in[0,1]$, then the above argument extends to the weak convergence of the partial sums: $\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}f_t(\theta)\Rightarrow V_{ff}^{1/2}W(s)+\Gamma_{ZY_{\bar{\pi}}}l_{\bar{\pi}}\int_0^s h(r)\,dr$, where $W(\cdot)$ is a multivariate standard Wiener process. So our stability and generalized tests also have power against this misspecification.

7.3.4 Correcting for time-varying trend inflation

We can eliminate the effect of the trend inflation on the tests by recentering the moment conditions to make them orthogonal to the space spanned by the vector $V_{ff}^{-\frac{1}{2}}\Gamma_{ZY_{\bar{\pi}}}l_{\bar{\pi}}$. Given consistent estimators of $\widehat{V}_{ff}$ and $\widehat{\Gamma}_{ZY_{\bar{\pi}}}$ of V_{ff} and $\Gamma_{ZY_{\bar{\pi}}}$, resp., this can be done by premultiplying $\frac{1}{\sqrt{T}}\widehat{V}_{ff}^{-\frac{1}{2}}\sum_{t=1}^{[sT]}f_t(\theta_0)$ with the orthogonal projection matrix $\mathrm{M}_{\widehat{V}_{ff}^{-1/2}\widehat{\Gamma}_{ZY_{\bar{\pi}}}l_{\bar{\pi}}}$ defined in (S–49) below. Note that this orthogonalization depends on the null value of the structural parameters θ_0, so it has to be repeated for each value of $\theta_0\in\Theta$.

7.3.5 Identification fails when $\alpha\to 0$

We will show that when α gets small, the terms $f_t(\theta)$ and $g_t(\theta)$ in the moment conditions (S–45) become collinear. Hence, orthogonalizing the sample moments $\sum_{t=1}^{T}f_t(\theta)$ to $T^{-1}\sum_{t=1}^{T}g_t(\theta)$ leads to sample moment conditions that are exactly zero over the full sample for all θ such that $\alpha=0$ and any choice of instruments Z_t. So, full-sample identification breaks down completely as $\alpha\to 0$. Stability restrictions yield partial identification.

Since the various test statistics are invariant to rescaling the moment vectors, consider the moment conditions multiplied by α. Rescaling $(\epsilon_t-\alpha\epsilon_{t+1})$ and $\varpi_t(\theta)$ by α

yields

$$\alpha\varrho\Delta\pi_t - \alpha\left(1+\alpha\varrho\right)\Delta\pi_{t+1} - \left(1-\alpha\right)^2 \upsilon x_t + \alpha^2\Delta\pi_{t+2} + \alpha\left(1-\alpha\right)^2 \upsilon x_{t+1} - \left(1-\alpha\right)^3 \upsilon \ln\mu$$

and

$$\begin{aligned}\alpha\varpi_t(\theta) &= -\left(1-\alpha\right)\left(1-\varrho\right)\frac{\left(a+\alpha\mu\right)}{\left(\mu-a\right)}\left[\left(1-\alpha\right)\ln\mu + x_t\right] + \alpha\lambda_{\bar{\pi}}\left(\pi_{t+1}-\varrho\pi_t\right) \\ &\quad + \alpha\left(1-\alpha\right)\left(1-\varrho\right)\frac{\left(a+\alpha\mu\right)}{\left(\mu-a\right)}x_{t+1} - \alpha^2\frac{\left(1-\varrho\right)}{\upsilon}\left(\pi_{t+2}-\varrho\pi_{t+1}\right) \\ &\quad + \alpha\left(1-\alpha\right)\left(1-\varrho\right)\left[\left(r_{t,t+1}+\Delta y_{t+1}\right) + \left(\mu-1\right)^{-1}\left(\pi_{t+1}-\varrho\pi_t\right)\right].\end{aligned}$$

Hence, when $\alpha \to 0$, the rescaled moment vector $\alpha\left[f_t(\theta) + g_t(\theta)\bar{\pi}_t\right]$ in (S–45) becomes

$$\underbrace{-\frac{a\left(\mu-1\right)}{\mu-a}Z_t'\left(x_t+\ln\mu\right)}_{f_t(\theta|\alpha=0)} + \underbrace{\left(1-\varrho\right)\frac{\left(a+\mu\right)}{\mu-a}Z_t'\left(x_t+\ln\mu\right)}_{g_t(\theta|\alpha=0)}\bar{\pi}_t.$$

Since $f_t(\theta)$ and $g_t(\theta) = Z_t'Y_{\bar{\pi},t}l_{\bar{\pi}}$ are collinear, and $\widehat{\Gamma}_{ZY_{\bar{\pi}}l_{\bar{\pi}}} = T^{-1}\sum_{t=1}^{T} g_t(\theta)$, $\mathrm{M}_{\widehat{V}_{ff}^{-\frac{1}{2}}\widehat{\Gamma}_{ZY_{\bar{\pi}}l_{\bar{\pi}}}}\widehat{V}_{ff}^{-\frac{1}{2}}\sum_{t=1}^{T} f_t(\theta) = 0$ for all $\theta \in \Theta \cap \{\alpha = 0\}$. There is still information in the stability restrictions for μ, but not for ϱ because $f_t(\theta)$ does not depend on ϱ in the limit as $\alpha \to 0$. So, stability restrictions lead to partial identification in this case. This discussion helps explain the empirical results.

7.3.6 Data

Real output growth (Δy_t), and the nominal interest rate (i_t) are measured as $\Delta y_t = \ln\left(\frac{\mathrm{Y}_t}{\mathrm{Y}_{t-1}}\right)$, and $\mathrm{i}_t = \frac{\mathrm{I}_t}{400}$, where Y_t is real GDP non-farm business sector (Billions of chained (2005) dollars. Seasonally adjusted at annual rates); and I_t is quarterly average of Effective Federal Funds Rate (annual percentage).

7.3.7 Computational details

Note that $(\epsilon_t - \alpha\epsilon_{t+1})$ can be rewritten as:

$$\underbrace{\varrho\Delta\pi_t - \left(1+\alpha\varrho\right)\Delta\pi_{t+1} - \frac{\left(1-\alpha\right)^2}{\alpha}\upsilon x_t + \alpha\Delta\pi_{t+2} + \left(1-\alpha\right)^2\upsilon x_{t+1}}_{Y_t b} - \underbrace{\left(1-\alpha\right)\kappa}_{X_t c} \quad \text{(S–46)}$$

where $Y_t = (\Delta\pi_t, \Delta\pi_{t+1}, \Delta\pi_{t+2}, x_t, x_{t+1})$, $b = \left(\varrho, -(1+\alpha\varrho), \alpha, -\frac{(1-\alpha)^2}{\alpha}\upsilon, (1-\alpha)^2\upsilon\right)'$, $X_t = 1$ and $c = \kappa$. We use the following relationships: $\pi_{t+1} - \varrho\pi_t = (1-\varrho)\pi_{t+1} + \varrho$ $\Delta\pi_{t+1}$; $\pi_{t+2} - \varrho\pi_{t+1} = \Delta\pi_{t+2} + (1-\varrho)\pi_{t+1}$; $\lambda_{\bar\pi} + \gamma_{\bar\pi}\alpha(\mu-1)^{-1} = \left[\upsilon^{-1} + (\mu-1)^{-1}(1-\alpha)\right]$ $(1-\varrho)$; and $\alpha\lambda_{\bar\pi} = \alpha(1-\varrho)\upsilon^{-1}$ to rewrite $\varpi_t(\theta)$ as:

$$\varpi_t(\theta) = \underbrace{(1-\alpha)\lambda_{\bar\pi}\ln\mu + Y_t b_{\bar\pi} + (1-\alpha)(1-\varrho)(r_{t,t+1} + \Delta y_{t+1}) + \vartheta\pi_{t+1}}_{Y_{\bar\pi,t}l_{\bar\pi}}$$

where:

$$\begin{aligned}
b_{\bar\pi} &= \begin{pmatrix} 0 & \left[\upsilon^{-1} + (\mu-1)^{-1}(1-\alpha)\right](1-\varrho)\varrho & \lambda_{\bar\pi} & -\upsilon^{-1}\alpha(1-\varrho) & -\alpha\lambda_{\bar\pi} \end{pmatrix}', \\
\vartheta &= \left[\upsilon^{-1} + (\mu-1)^{-1}\right](1-\alpha)(1-\varrho)^2, \\
Y_{\bar\pi,t} &= \begin{pmatrix} X_t & Y_t & (r_{t,t+1} + \Delta y_{t+1}) & \pi_{t+1} \end{pmatrix}, \qquad \text{(S–47)} \\
l_{\bar\pi} &= \begin{pmatrix} (1-\alpha)\lambda_{\bar\pi}\ln\mu & b_{\bar\pi} & \lambda_{\bar\pi} & (1-\alpha)(1-\varrho) & \vartheta \end{pmatrix}'. \qquad \text{(S–48)}
\end{aligned}$$

Computation of S and CLR statistics The estimator of the asymptotic variance of $\frac{1}{\sqrt{T}}Z'(Yb - Xc)$, $\widehat{V}_{ff}$, is the one described in equation (S–26).

Let $\overline{Y} = (Y, X)$ and $\bar{b} = \binom{b}{c}$, such that the moment vector can be written more compactly as $Z'\bar{Y}\bar{b}$. We eliminate the effect of trend inflation on the tests by premultiplying the sample moments by $\mathrm{M}_{V_{ff}^{-1/2}\Gamma_{ZY_{\bar\pi}}l_{\bar\pi}}$ where:

$$\mathrm{M}_{V_{ff}^{-1/2}\Gamma_{ZY_{\bar\pi}}l_{\bar\pi}} = \mathrm{I}_{k_z} - V_{ff}^{-\frac{1}{2}}\Gamma_{ZY_{\bar\pi}}l_{\bar\pi}\left[l'_{\bar\pi}\Gamma_{ZY_{\bar\pi}}V_{ff}^{-1}\Gamma_{ZY_{\bar\pi}}l_{\bar\pi}\right]^{-\frac{1}{2}} l'_{\bar\pi}\Gamma_{ZY_{\bar\pi}}V_{ff}^{-\frac{1}{2}} \qquad \text{(S–49)}$$

We replace V_{ff}, $\Gamma_{ZY_{\bar\pi}}$ with their estimators $\widehat{V}_{ff}$, and $\widehat{\Gamma}_{ZY_{\bar\pi}} = T^{-1}Z'Y_{\bar\pi}$ to compute the tests.

Let L be a $k_z \times (k_z - 1)$ matrix such that $\mathrm{M}_{\widehat{V}^{-\frac{1}{2}}\widehat{\Gamma}_{ZY_{\bar\pi}l_{\bar\pi}}} = LL'$ and $L'L = \mathrm{I}_{(k_z-1)}$. The S statistic is:

$$S_T(\theta_0) = \frac{1}{T}\underbrace{(\overline{Y}\,\bar{b})'\,Z\widehat{V}_{ff}^{-\frac{1}{2}}L}_{\hat\xi_T}\underbrace{L'\widehat{V}_{ff}^{-\frac{1}{2}}Z'\overline{Y}\,\bar{b}}_{\hat\xi_T}$$

$\frac{1}{\sqrt{T}}\hat\xi_T$ is an asymptotically standard normal vector of dimension $k_z - 1$. Note that, in the computing the S statistic, the use of the projection matrix $\mathrm{M}_{\widehat{V}_{ff}^{-1/2}\widehat{\Gamma}_{ZY_{\bar\pi}}l_{\bar\pi}}$ to annihilate the trend inflation effect in the limiting distribution is equivalent to first estimating $c_{\bar\pi}$

by solving the following minimization problem:

$$\widehat{c}_{\bar{\pi}} = \arg\min_{c_{\bar{\pi}}} \frac{1}{T} \left(\overline{Y}\,\bar{b} - Y_{\bar{\pi}} l_{\bar{\pi}} c_{\bar{\pi}}\right)' Z \widehat{V}_{ff}^{-1} Z' \left(\overline{Y}\,\bar{b} - Y_{\bar{\pi}} l_{\bar{\pi}} c_{\bar{\pi}}\right) \tag{S–50}$$

and then substituting $\widehat{c}_{\bar{\pi}}$ back into the objective function in (S–50). The KLM, JKLM and CLR statistics are obtained using the formulae in Subsection 7.1.2, see eq. (S–29).

Computation of the qLL-S statistic The computation of the qLL-S under the presence of trend inflation follows the steps described in Subsection 7.1.2, but redefining $\widehat{\epsilon}$ in equation (S–30) and $\widehat{F}$ in equation (S–31) as

$$\widehat{\epsilon} = Yb - Xc = \overline{Y}\,\bar{b}, \quad \text{and} \quad \widehat{F} = \left(\widehat{E} \odot Z\right) \widehat{V}_{ff}^{-\frac{1}{2}} \mathrm{M}_{\widehat{V}_{ff}^{-\frac{1}{2}} \widehat{\Gamma}_{ZY_{\bar{\pi}}} l_{\bar{\pi}}},$$

respectively.

Computation of the exp/ave-S and split-sample statistics Let the $2k_z \times 1$ split-sample moment vector be defined as $\overline{Z}'\overline{Y}\,\bar{b}$ where $\overline{Z}$ is the split-sample instrument matrix defined in (S–32). The estimator of the asymptotic variance of $\frac{1}{\sqrt{T}}\overline{Z}'\overline{Y}\,\bar{b}$ is defined in (S–33). We eliminate the effect of trend inflation on the tests by premultiplying the sample moments by the $2(k_z - 1) \times 2k_z$ matrix $\bar{L}'$, defined as $\bar{L}\bar{L}' = \mathrm{I}_2 \otimes \mathrm{M}_{\widehat{V}_{ff}^{-1/2} \widehat{\Gamma}_{ZY_{\bar{\pi}}} l_{\bar{\pi}}}$ where $\mathrm{M}_{\widehat{V}_{ff}^{-1/2} \widehat{\Gamma}_{ZY_{\bar{\pi}}} l_{\bar{\pi}}}$ is given by (S–49). The remaining calculations are the same as in Subsection 7.1.2.

7.3.8 Empirical results

Three-dimensional 95%-level confidence sets for (α, ϱ, μ) in (S–42) are constructed by inverting the various weak-identification robust tests, and they are plotted in Figures 13, 14 and 15 for the sample 1966q1-2010q4 and Figures 16, 17 and 18 for the sample 1984q1-2010q4. Table 10 shows the proportion of the confidence regions with respect to the parallelepiped.

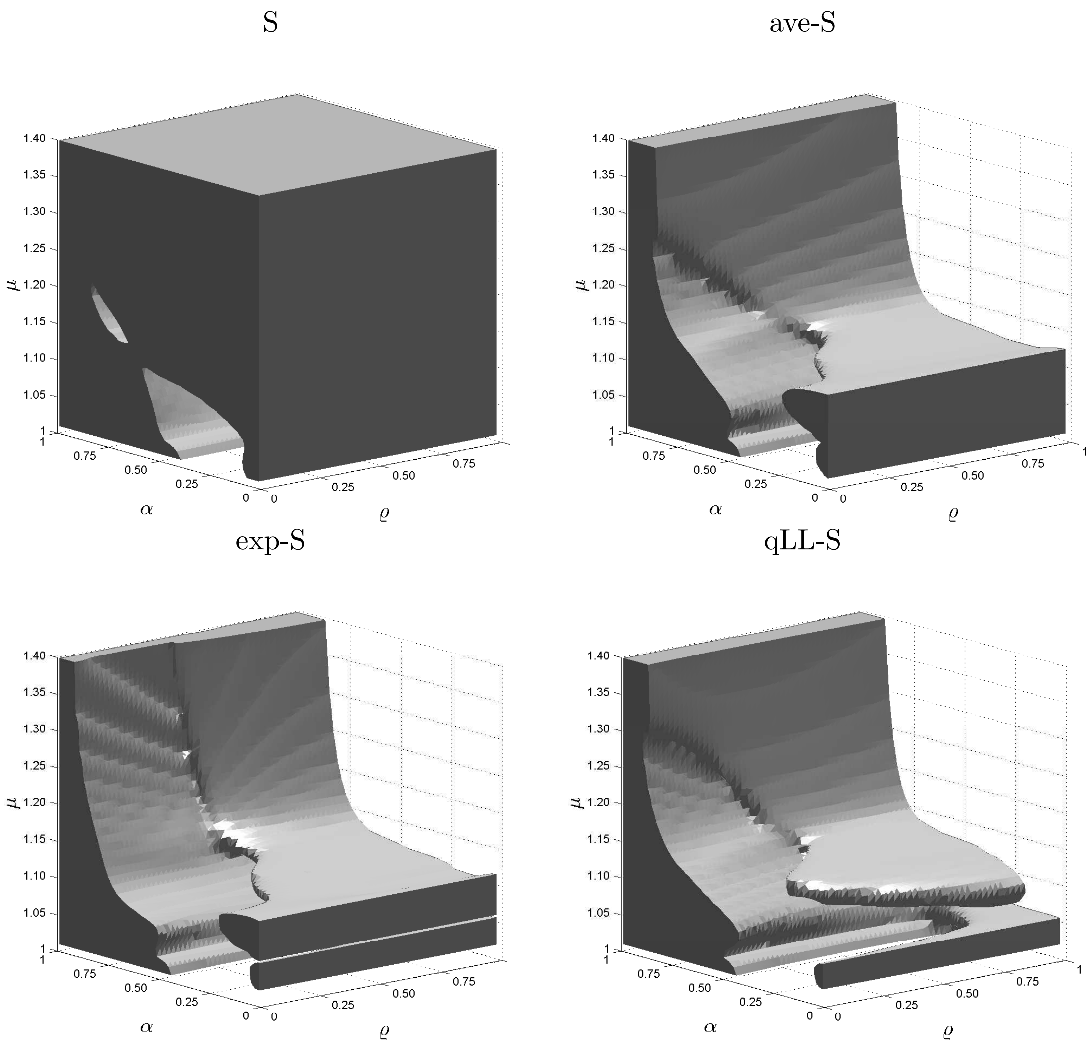

Figure 13: 95%-level S and generalized S confidence sets for α, ϱ, μ in the NKPC with trend inflation. Instruments: constant, two lags of $\Delta\pi$ and three lags of x_t. Period: 1966q1-2010q4.

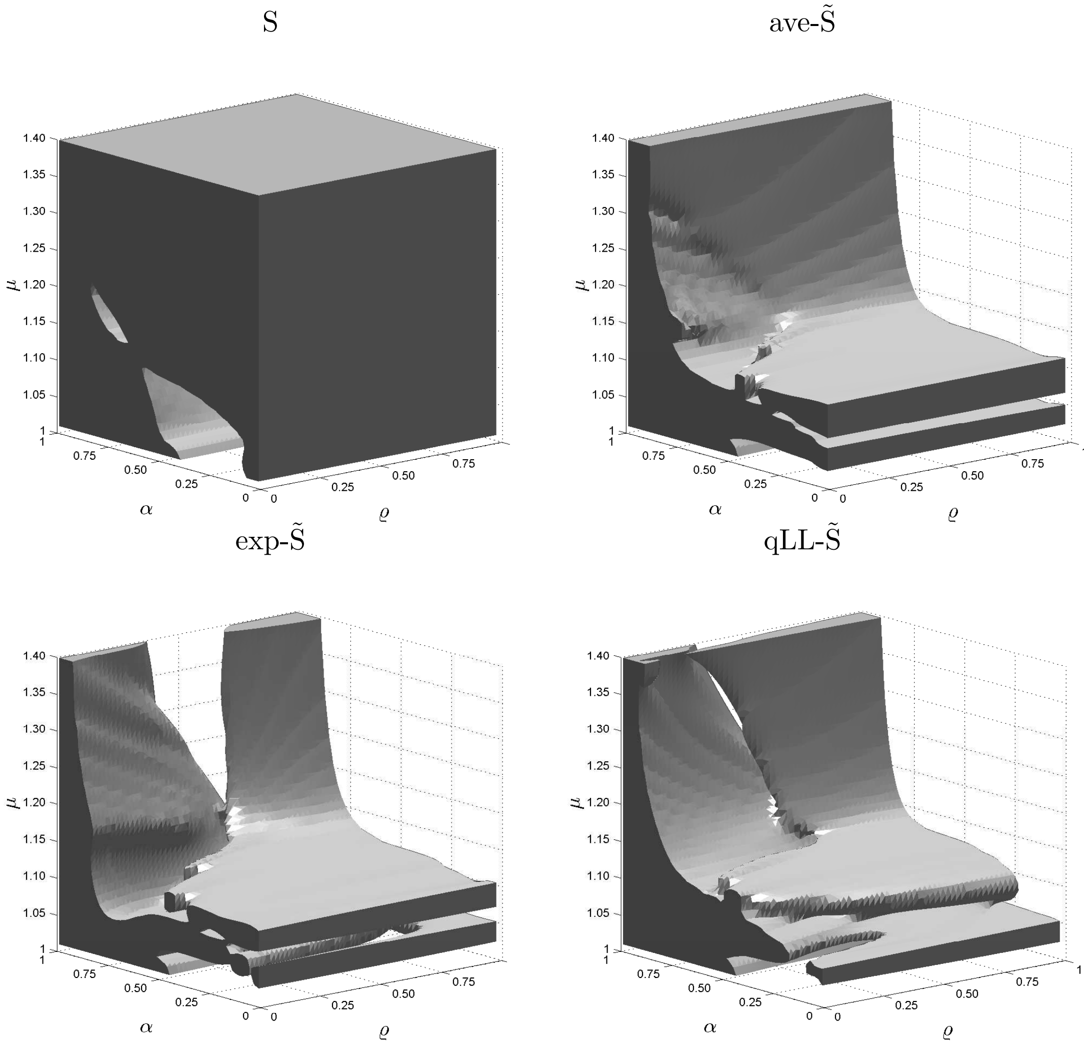

Figure 14: 95%-level S and stability S confidence sets for α, ϱ, μ in the NKPC with trend inflation. Instruments: constant, two lags of $\Delta\pi$ and three lags of x_t. Period: 1966q1-2010q4.

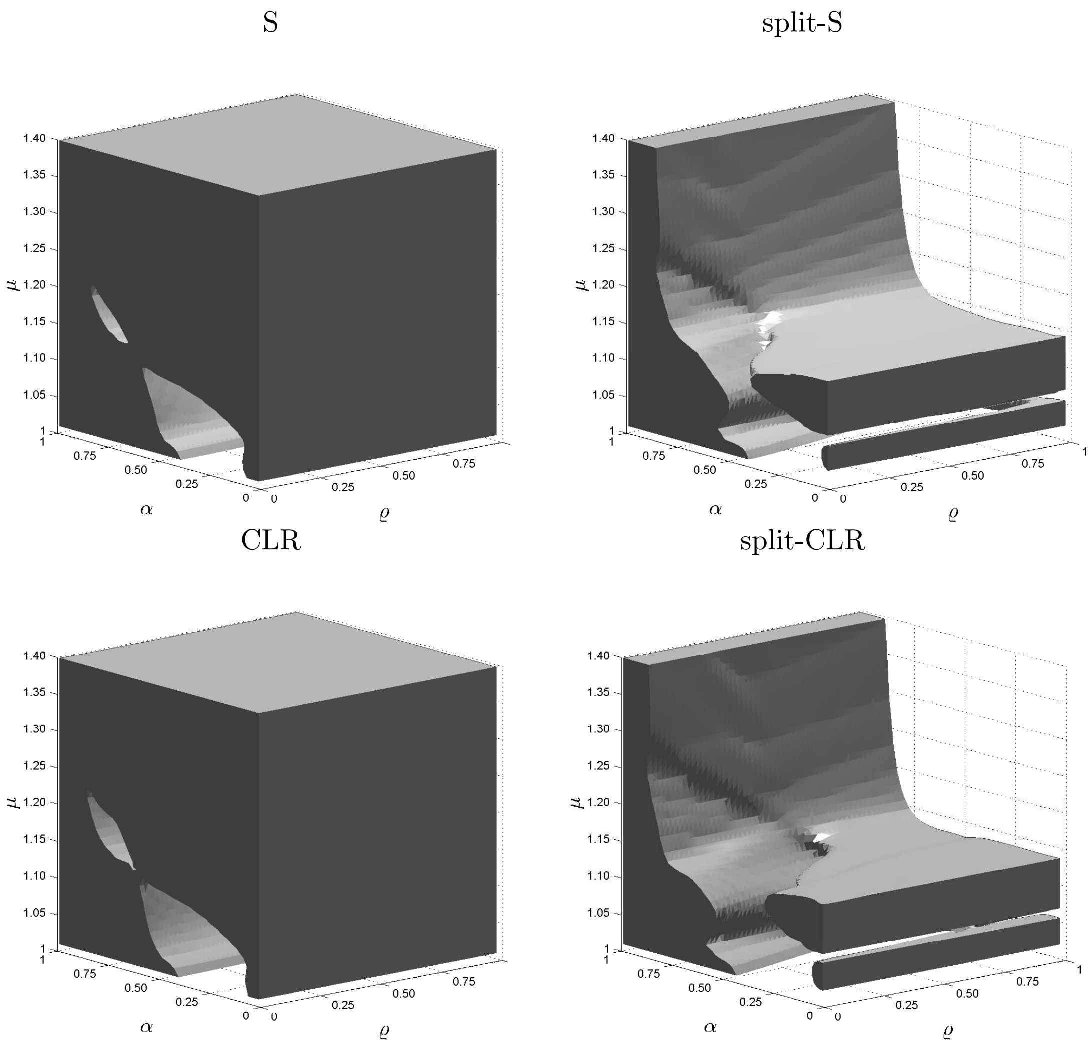

Figure 15: 95%-level S, split-S, CLR and split-CLR confidence sets for α, ϱ, μ in the NKPC with trend inflation. Instruments: constant, two lags of $\Delta\pi$ and three lags of x_t. Period: 1966q1-2010q4.

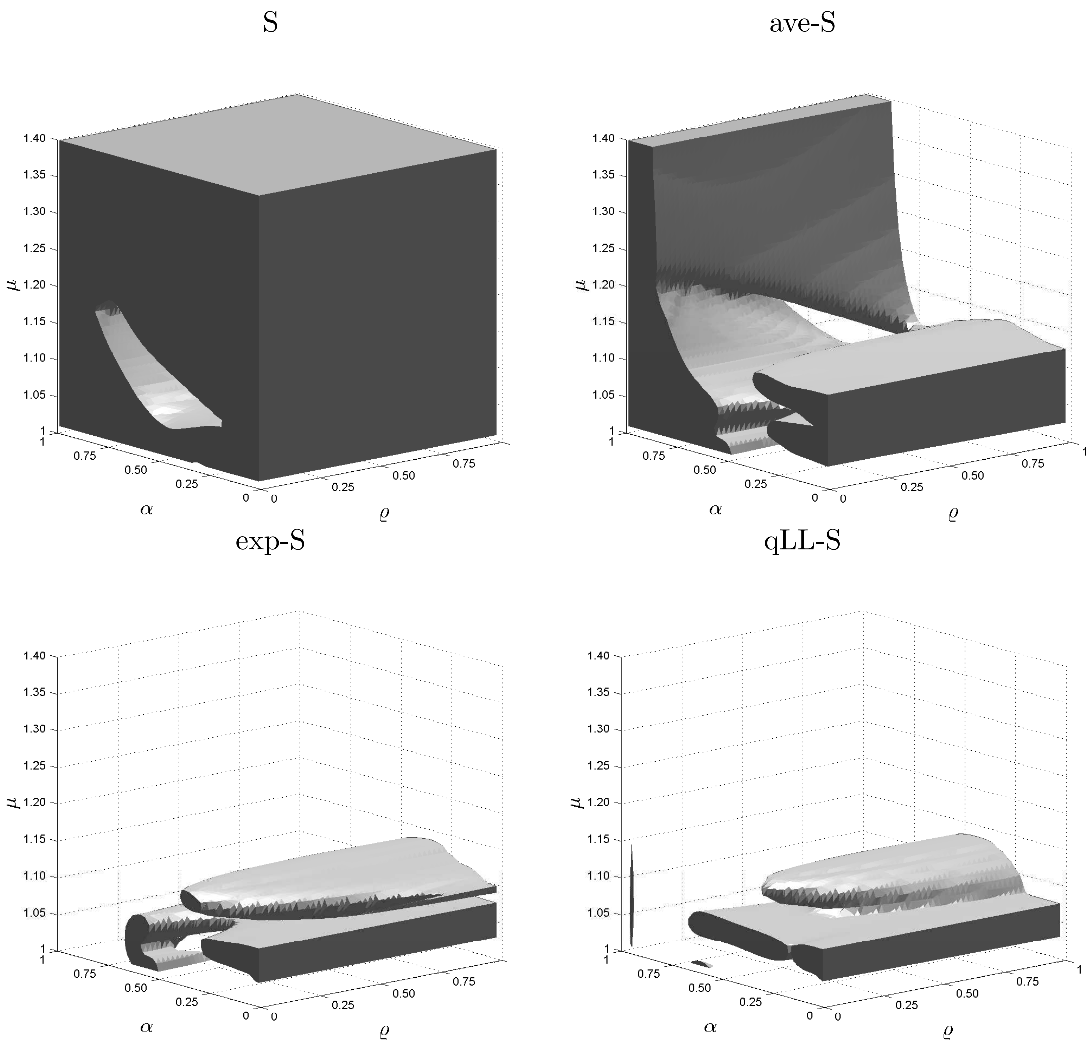

Figure 16: 95%-level S and generalized S confidence sets for α, ϱ, μ in the NKPC with trend inflation. Instruments: constant, two lags of $\Delta\pi$ and three lags of x_t. Period: 1984q1-2010q4.

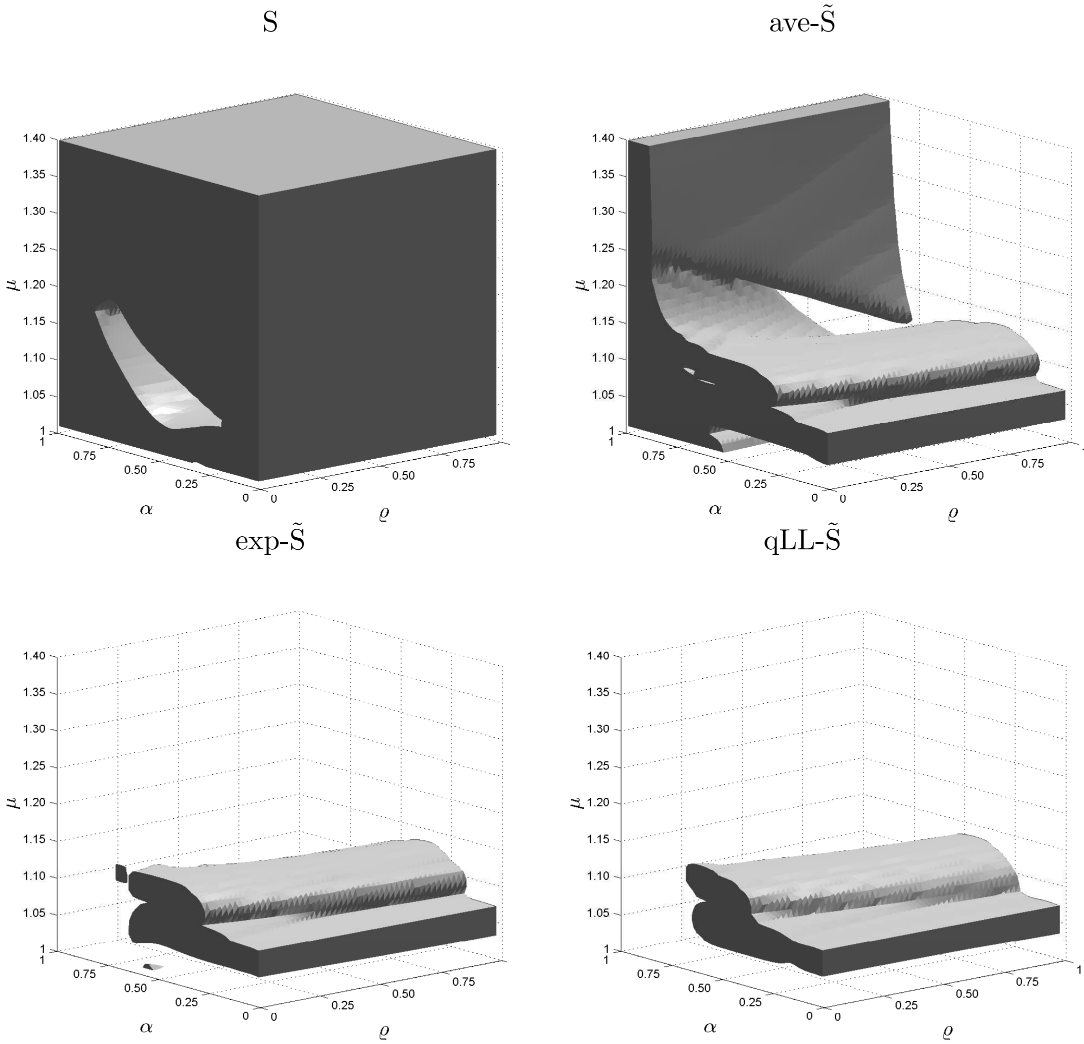

Figure 17: 95%-level S and generalized S confidence sets for α, ϱ, μ in the NKPC with trend inflation. Instruments: constant, two lags of $\Delta\pi$ and three lags of x_t. Period: 1984q1-2010q4.

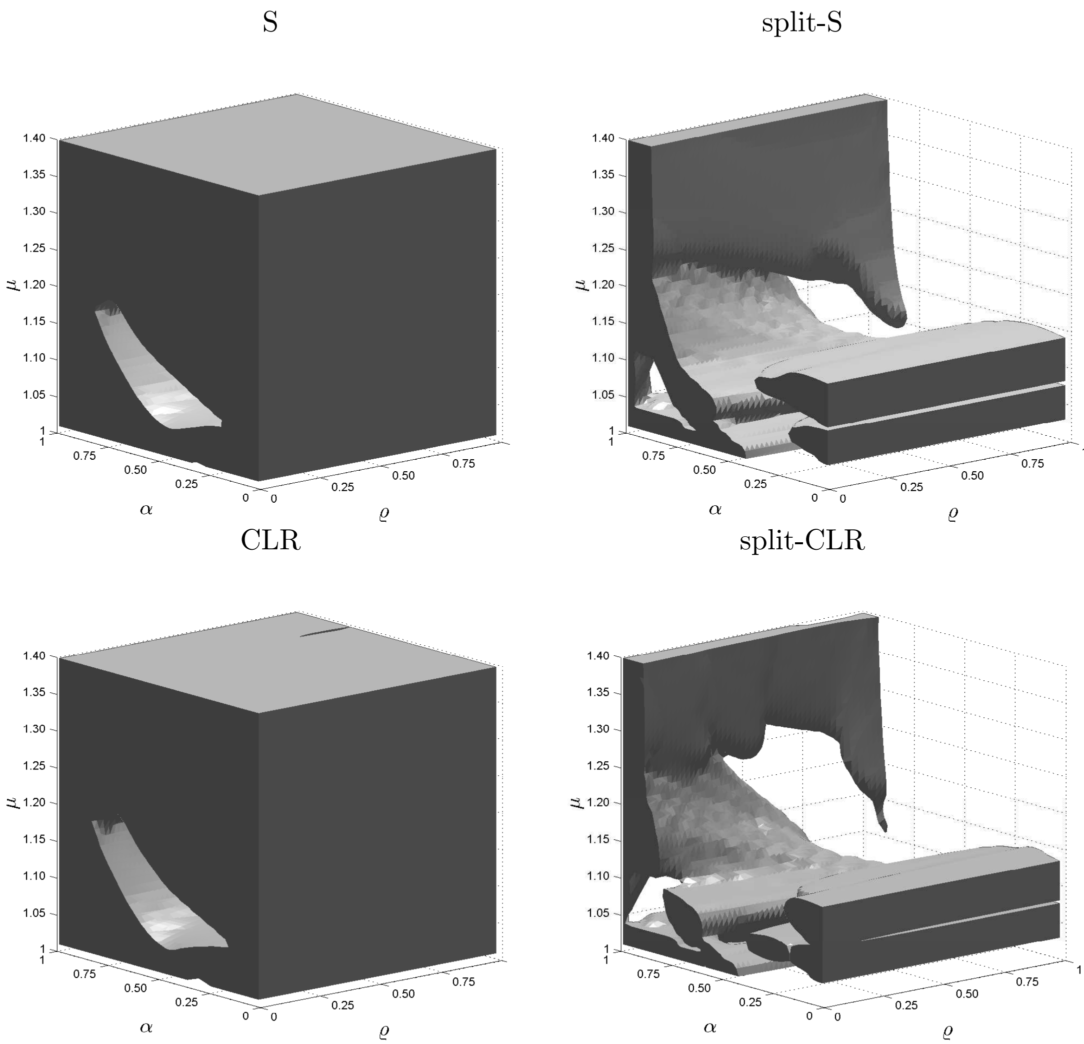

Figure 18: 95%-level S, split-S, CLR and split-CLR confidence sets for α, ϱ, μ in the NKPC with trend inflation. Instruments: constant, two lags of $\Delta\pi$ and three lags of x_t. Period: 1984q1-2010q4.

Confidence Regions	1966q1 - 2010q4 95%	90%	1984q1 - 2010q4 95%	90%
S	90.78	89.71	92.00	86.36
CLR	89.78	88.42	90.17	83.54
ave-S	29.78	27.90	27.55	24.33
exp-S	23.73	21.26	7.07	5.37
qLL-S	23.54	21.49	5.77	5.13
ave-$\tilde{S}$	28.84	27.02	24.52	22.31
exp-$\tilde{S}$	20.15	13.87	6.86	5.99
qLL-$\tilde{S}$	23.27	19.92	6.82	6.21
split-S	30.73	29.03	20.32	17.82
split-CLR	26.50	24.13	15.62	12.11

Table 10: Volume of Confidence Regions as a proportion of the Volume of the Parallelepiped $(\alpha, \varrho, \mu) \in [0.01, 0.99] \times [0, 1] \times [1.01, 1.40]$

References

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statistics 20*, 46–63.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica 74*(3), 715–752.

Cogley, T. and A. M. Sbordone (2008). Trend inflation and inflation persistence in the New Keynesian Phillips Curve. *American Economic Review 98*(5), 2101–26.

Dufour, J.-M. (2003). Identification, Weak Instruments and Statistical Inference in Econometrics. *Canadian Journal of Economics 36*(4), 767–808. Presidential Address to the Canadian Economics Association.

Galí, J. (2008). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics 127*(1), 97–126.

Li, H. and U. K. Mueller (2009). Valid Inference in Partially Unstable GMM Models. *Review of Economic Studies 76*, 343–365.

Mueller, U. K. (2011). Efficient tests under a weak convergence assumption. *Econometrica 79*(2), 395–435.

Newey, W. K. and K. D. West (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Rossi, B. (2005). Optimal tests for nested model selection with underlying parameter instability. *Econometric Theory 21*(05), 962–990.

Stock, J. H. and M. W. Watson (1998). Median unbiased estimation of coefficient variance in a time-varying parameter model. *J. Amer. Statist. Assoc. 93*(441), 349–358.

Wang, J. and E. Zivot (1998). Inference on structural parameters in instrumental variables regression with weak instruments. *Econometrica 66*(6), 1389–1404.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.