SUPPLEMENT TO “IDENTIFICATION USING STABILITY RESTRICTIONS”: CRITICAL VALUE TABLES, PROOFS, AND ADDITIONAL RESULTS
(Econometrica, Vol. 82, No. 5, September 2014, 1799–1851)

BY LEANDRO M. MAGNUSSON AND SOPHOCLES MAVROEIDIS

S.1. INTRODUCTION

This supplement 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.

S.2. TABLES OF CRITICAL VALUES FOR GENERALIZED S TESTS

Asymptotic critical values for the generalized and stability S tests, \(q_{\text{LL-S}}\), \(q_{\text{LL-\tilde{S}}}\), \(\exp-S\), \(\exp-\tilde{S}\), \(\text{ave-S}\), \(\text{ave-}\tilde{S}\), 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 \(p_\zeta\) 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 distributions.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>qLL-S</th>
<th>exp-S</th>
<th>ave-S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>(k/p_\zeta = 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8.59</td>
<td>9.99</td>
<td>13.03</td>
</tr>
<tr>
<td>2</td>
<td>15.32</td>
<td>17.10</td>
<td>20.78</td>
</tr>
<tr>
<td>3</td>
<td>21.76</td>
<td>23.82</td>
<td>28.02</td>
</tr>
<tr>
<td>4</td>
<td>27.98</td>
<td>30.31</td>
<td>34.79</td>
</tr>
<tr>
<td>5</td>
<td>34.21</td>
<td>36.56</td>
<td>41.81</td>
</tr>
<tr>
<td>6</td>
<td>40.12</td>
<td>42.75</td>
<td>48.03</td>
</tr>
<tr>
<td>7</td>
<td>46.10</td>
<td>48.95</td>
<td>54.90</td>
</tr>
<tr>
<td>8</td>
<td>52.16</td>
<td>55.23</td>
<td>61.17</td>
</tr>
<tr>
<td>9</td>
<td>58.09</td>
<td>61.30</td>
<td>67.68</td>
</tr>
<tr>
<td>10</td>
<td>64.09</td>
<td>67.24</td>
<td>73.80</td>
</tr>
</tbody>
</table>

*Derived from Theorems 1 and 2 in the paper, with \(\bar{\xi} = \xi = c\), 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 \(p_\zeta\) is the number of estimated parameters under the null.

© 2014 The Econometric Society DOI: 10.3982/ECTA9612
### TABLE S.II

**Asymptotic Critical Values for Generalized S Tests**

<table>
<thead>
<tr>
<th>Statistic:</th>
<th>qLL-S</th>
<th>exp-S</th>
<th>ave-S</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k \setminus p_\xi = 1$</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{1}$</td>
</tr>
<tr>
<td>2</td>
<td>14.03</td>
<td>15.74</td>
<td>19.14</td>
</tr>
<tr>
<td>3</td>
<td>20.56</td>
<td>22.51</td>
<td>26.69</td>
</tr>
<tr>
<td>4</td>
<td>26.81</td>
<td>28.97</td>
<td>33.70</td>
</tr>
<tr>
<td>5</td>
<td>33.00</td>
<td>35.35</td>
<td>40.26</td>
</tr>
<tr>
<td>6</td>
<td>39.10</td>
<td>41.71</td>
<td>46.92</td>
</tr>
<tr>
<td>7</td>
<td>45.12</td>
<td>47.75</td>
<td>53.71</td>
</tr>
<tr>
<td>8</td>
<td>51.05</td>
<td>53.97</td>
<td>59.86</td>
</tr>
<tr>
<td>9</td>
<td>57.02</td>
<td>60.10</td>
<td>66.11</td>
</tr>
<tr>
<td>10</td>
<td>62.87</td>
<td>66.17</td>
<td>72.70</td>
</tr>
</tbody>
</table>

$^a$See Table S.I for details.

### TABLE S.III

**Asymptotic Critical Values for Generalized S Tests**

<table>
<thead>
<tr>
<th>Statistic:</th>
<th>qLL-S</th>
<th>exp-S</th>
<th>ave-S</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k \setminus p_\xi = 2$</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{1}$</td>
</tr>
<tr>
<td>3</td>
<td>19.36</td>
<td>21.21</td>
<td>25.19</td>
</tr>
<tr>
<td>4</td>
<td>25.67</td>
<td>27.81</td>
<td>32.14</td>
</tr>
<tr>
<td>5</td>
<td>31.87</td>
<td>34.25</td>
<td>39.03</td>
</tr>
<tr>
<td>6</td>
<td>37.91</td>
<td>41.71</td>
<td>46.92</td>
</tr>
<tr>
<td>7</td>
<td>44.12</td>
<td>47.75</td>
<td>53.71</td>
</tr>
<tr>
<td>8</td>
<td>51.05</td>
<td>53.97</td>
<td>59.86</td>
</tr>
<tr>
<td>9</td>
<td>57.02</td>
<td>60.10</td>
<td>66.11</td>
</tr>
<tr>
<td>10</td>
<td>62.87</td>
<td>66.17</td>
<td>72.70</td>
</tr>
</tbody>
</table>

$^a$See Table S.I for details.

### TABLE S.IV

**Asymptotic Critical Values for Generalized S Tests**

<table>
<thead>
<tr>
<th>Statistic:</th>
<th>qLL-S</th>
<th>exp-S</th>
<th>ave-S</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k \setminus p_\xi = 3$</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{1}$</td>
</tr>
<tr>
<td>4</td>
<td>24.47</td>
<td>26.54</td>
<td>30.81</td>
</tr>
<tr>
<td>5</td>
<td>30.79</td>
<td>32.99</td>
<td>37.47</td>
</tr>
<tr>
<td>6</td>
<td>36.84</td>
<td>39.23</td>
<td>44.34</td>
</tr>
<tr>
<td>7</td>
<td>42.93</td>
<td>45.57</td>
<td>50.94</td>
</tr>
<tr>
<td>8</td>
<td>48.92</td>
<td>51.71</td>
<td>57.52</td>
</tr>
<tr>
<td>9</td>
<td>54.89</td>
<td>57.93</td>
<td>64.01</td>
</tr>
<tr>
<td>10</td>
<td>60.87</td>
<td>63.87</td>
<td>70.08</td>
</tr>
</tbody>
</table>

$^a$See Table S.I for details.
### Table S.V

**Asymptotic Critical Values for Generalized S Tests**

<table>
<thead>
<tr>
<th>Statistic:</th>
<th>$qLL-S$</th>
<th>$exp-S$</th>
<th>$ave-S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k / p_k = 4$</td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
</tr>
<tr>
<td>5</td>
<td>29.63 31.81 36.51</td>
<td>5.92 6.85 8.88</td>
<td>9.35 10.75 13.88</td>
</tr>
<tr>
<td>6</td>
<td>35.74 38.08 43.18</td>
<td>7.41 8.45 10.60</td>
<td>12.08 13.73 17.20</td>
</tr>
<tr>
<td>7</td>
<td>41.78 44.34 49.72</td>
<td>8.85 9.95 12.35</td>
<td>14.63 16.44 20.41</td>
</tr>
<tr>
<td>8</td>
<td>47.89 50.59 56.27</td>
<td>10.26 11.46 13.95</td>
<td>17.19 19.17 23.35</td>
</tr>
<tr>
<td>9</td>
<td>53.90 56.78 62.65</td>
<td>11.59 12.86 15.45</td>
<td>19.61 21.76 26.33</td>
</tr>
<tr>
<td>10</td>
<td>59.79 62.87 69.11</td>
<td>12.96 14.30 16.99</td>
<td>22.05 24.33 29.13</td>
</tr>
</tbody>
</table>

*See Table S.I for details.*

### Table S.VI

**Asymptotic Critical Values for Generalized S Tests**

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

*See Table S.I for details.*

### Table S.VII

**Asymptotic Critical Values for Stability Component of Generalized S Tests**

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

*Derived from Theorems 1 and 2 in the paper, with $\tilde{c} = 10, \infty$, and 0 for $qLL/exp/ave-\tilde{S}$, resp. Computed using 50,000 draws of $k$-dimensional Brownian motion, where $k$ denotes the number of moment conditions.*
\( \chi^2 \) 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\%.

S.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 \( F_T(\theta_0) \) by Lemma 6 of Elliott and Müller (2006). Independence of \( \bar{\psi}_k \) and \( \psi_c \) follows from the asymptotic independence between \( \sqrt{T} \hat{V}_{ff}(\theta_0) \) and \( [I_k \otimes (M_e - G_c)] \hat{u} \), where \( \hat{u} = (v'_1, \ldots, v'_q)' \), which is a direct consequence of Assumption 2 under \( H_0 \). Equation (26) then follows from the continuous mapping theorem. Finally, asymptotic efficiency follows from Müller (2011, Theorem 1).

\( Q.E.D. \)

PROOF OF THEOREM 2: By Assumption 2, under \( H_0 \), \( V^{-1/2}X_T(1) \Rightarrow W(1) \) and \( S_T(\theta_0) = X_T(1)'V^{-1}X_T(1) + \rho_o(1) \Rightarrow \bar{\psi}_k \). Also,

\[
\tilde{S}_T(\theta_0, \tau) = X_T(\tau)'V^{-1}_\tau X_T(\tau) \\
\quad + \left[ X_T(1) - X_T(\tau)' \right] V^{-1}_\tau \left[ X_T(1) - X_T(\tau) \right] \\
\quad - X_T(1)'V^{-1}_\tau X_T(1) + \rho_o(1) \\
= \frac{[X_T(\tau) - \tau X_T(1)]'V^{-1}_\tau [X_T(\tau) - \tau X_T(1)]}{\tau(1-\tau)} + \rho_o(1)
\]

and \( V^{-1/2}_X[X_T(\tau) - \tau X_T(1)] \Rightarrow \tilde{W}(\tau) \), which is independent of \( W(1) \) and \( \tilde{S}_T(\theta_0, \tau) \Rightarrow \tilde{\psi}_k(\tau) \). By the Neyman–Pearson lemma, the test function

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

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 Müller (2011, Theorem 1).

\( Q.E.D. \)

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 non-stochastic 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

\[
\text{cov}(Z(s_1)'Yb_0, Z(s_2)'Y\Omega^{-1}A_0) = \sum_{t=1}^{T} Z_t(s_1)'Z_t(s_2) \text{cov}(Y_t b_0, Y_t \Omega^{-1} A_0) = \sum_{t=1}^{T} Z_t(s_1)'Z_t(s_2) b_0' \Omega \Omega^{-1} A_0 = 0.
\]

\( Q.E.D. \)
PROOF OF THEOREM 3: Since the random functions $\mathcal{F}(\cdot)$ and $\mathcal{D}(\cdot)$ are independent of each other, by Lemma 1, and $\hat{\tau}$ only depends on $\mathcal{D}(\cdot)$ by (32), it follows that $\hat{\tau}$ is also independent of $\mathcal{F}(\cdot)$. Therefore, under $H_0$, Lemma 1 part 1 implies that the (conditional) distribution of $\mathcal{F}(\hat{\tau})$ is Gaussian with zero mean and variance matrix $I_k$. Part 1 follows immediately. For part 2, note that conditional on $\mathcal{D}(\hat{\tau})$ and $\hat{\tau}$, $\mathcal{D}(\hat{\tau})F(\hat{\tau})$ is Gaussian with mean zero and variance matrix $\mathcal{Q}$, and the matrix $\mathcal{D}(\hat{\tau})\mathcal{D}(\hat{\tau})$ is invertible with probability 1. Parts 3 and 4 follow from the fact that $\mathcal{F}(\hat{\tau})\mathcal{D}(\hat{\tau})$ implies that $\mathcal{F}(\hat{\tau})\mathcal{D}(\hat{\tau})$ is asymptotically independent of $\mathcal{Q}$. The asymptotic independence of $\mathcal{F}(\hat{\tau})\mathcal{D}(\hat{\tau})$ will be asymptotically independent of $\mathcal{Q}$. Splitting $\mathcal{F}(\hat{\tau})\mathcal{D}(\hat{\tau})$ into its constituent parts, Lemma 1 part 1 implies that the (conditional) distribution of $\mathcal{F}(\hat{\tau})\mathcal{D}(\hat{\tau})$ is Gaussian with zero mean and variance $\mathcal{Q}$. This establishes that $\mathcal{F}(\hat{\tau})\mathcal{D}(\hat{\tau}) \overset{d}{\rightarrow} \chi^2(2k)$. The remaining results follow by direct analogy with the proof of Theorem 3.

Q.E.D.

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 $\mathcal{D}_{ij}(\theta_0, \cdot)$ and $\mathcal{F}_{ij}(\theta_0)$ implies that $\hat{\tau}_0$ will be asymptotically independent of $\mathcal{F}_{ij}(\theta_0)$ as well, and so, conditional on $\hat{\tau}_0$, $(\hat{\tau}_0T)^{-1/2}\mathcal{F}_{ij}(\theta_0, \hat{\tau}_0)$ and $((1 - \hat{\tau}_0)T)^{-1/2}\mathcal{F}_{ij}(\theta_0, \hat{\tau}_0)$ are jointly asymptotically Gaussian and independent with zero mean and variance $\mathcal{V}$, that is, $\mathcal{D}_{ij}(\hat{\tau}_0)\mathcal{F}(\hat{\tau}_0) = 0$, so $\mathcal{D}_{ij}(\hat{\tau}_0)\mathcal{F}(\hat{\tau}_0)$ and $\mathcal{D}(\hat{\tau})\mathcal{F}(\hat{\tau})$ are independent conditionally on $\mathcal{D}(\hat{\tau})$ and $\hat{\tau}$. Part 5 now follows by combining the above results using the continuous mapping theorem.

Q.E.D.

PROOF OF THEOREM 5: First, by Assumptions 7(b) and 8(a), we have $T^{-1/2}\mathcal{F}_{ij}(\theta_0) = \mathcal{O}_{p}(1)$ and $\mathcal{V}_{ij}(\theta_0, s) = \mathcal{O}_{p}(1)$, $\mathcal{V}_{ij}(\theta_0, s) = \mathcal{O}_{p}(1)$ uniformly in $s \in [0, 1]$, respectively. Hence,

$$(S-1) \quad \text{vec}[\mathcal{D}_{ij}(\theta_0, s)] = \text{vec}[\mathcal{Q}_{ij}(\theta_0, s)] + o_p(T)$$

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 $X^* = J'\mathcal{V}_{ij}^{-1}X(1)$ and $V_{X^*} = J'\mathcal{V}_{ij}^{-1}J$. Equation (S-1) and Assumption 7(c) imply $T_i^{-1}\mathcal{D}_{ij}(\theta_0, \hat{\tau}_0) \overset{p}{\rightarrow} J$ for both $i = 1, 2$. Moreover, by Assumption 8(b) and Slutsky’s theorem, we have $\frac{1}{T_i}\mathcal{D}_{ij}(\theta_0, \hat{\tau}_0)\mathcal{V}_{ij}(\theta_0, \hat{\tau}_0)^{-1} \overset{p}{\rightarrow} J\mathcal{V}_{ij}^{-1}$, $i = 1, 2$ and $\mathcal{V}_{X^*} \overset{p}{\rightarrow} \mathcal{V}_{X^*}$, since $|\hat{\tau}_0| \leq 1$. Finally, since $T^{-1/2}\mathcal{F}_{ij}(\theta_0, s)$ is uniformly bounded, $X^* = \sum_{i=1}^{2} \frac{1}{T_i}\mathcal{D}_{ij}(\theta_0, \hat{\tau}_0)\mathcal{V}_{ij}(\theta_0, \hat{\tau}_0)^{-1} \mathcal{F}_{ij}(\theta_0, \hat{\tau}_0) \overset{d}{\rightarrow} \mathcal{V}_{X^*} = \mathcal{V}_{X^*} = \mathcal{V}_{X^*} = J'\mathcal{V}_{ij}^{-1}J + (1 - \tau)J_1\mathcal{V}_{ij}^{-1}J_1$. The continuous mapping theorem.

THE CASE OF A LARGE BREAK—Assumption 8(d): Assumption 7 implies $X^* = J_1'\mathcal{V}_{ij}^{-1}X(\tau) + J_2'\mathcal{V}_{ij}^{-1}[X(1) - X(\tau)]$ and $V_{X^*} = \tau J_1'\mathcal{V}_{ij}^{-1}J_1 + (1 - \tau)J_2'\mathcal{V}_{ij}^{-1}J_2$. The continuous mapping theorem.
First, we show that \( T^{-1/2}[F^i_T(\theta_0, \hat{\tau}_0) - F^i_T(\theta_0, \tau)] = o_p(1), \ i = 1, 2. \) Observe that \( \|F^i_T(\theta_0, \hat{\tau}_0) - F^i_T(\theta_0, \tau)\| = \| \sum_{j=1}^n f_{0+j}(\theta_0) \| \leq \sum_{j=1}^n \| f_{0+j}(\theta_0) \|, \ t_0 = [\min(\tau, \hat{\tau}_0)T] \text{ and } n = |[\hat{\tau}_0 - \tau]T|. \) So, we need to show that for all \( \eta, \delta > 0, \) there exist \( T^* \) such that, for all \( T \geq T^*, \) \( \Pr(T^{-1/2} \sum_{j=1}^n \| f_{0+j}(\theta_0) \| > \delta < \eta. \)

By Assumption 8(d.iii), \( T^{-1/2} \sum_{j=1}^N \| f_{0+j}(\theta_0) \| = o_p(1) \) for any fixed \( N < \infty, \) that is, for every \( \delta > 0, \) there exist \( T^*_1 \) large enough such that \( \Pr(T^{-1/2} \sum_{j=1}^n \| f_{0+j}(\theta_0) \| > \delta) < \frac{\eta}{2} \) for all \( T \geq T^*_1. \) By Assumption 8(d.ii), there exist \( N < \infty \) and \( T^*_2 \) large enough such that \( \Pr(n > N) < \frac{\eta}{2} \) for all \( T \geq T^*_2. \) Hence,

\[
\Pr\left( T^{-1/2} \sum_{j=1}^n \| f_{0+j}(\theta_0) \| > \delta \right) = \Pr\left( T^{-1/2} \sum_{j=1}^n \| f_{0+j}(\theta_0) \| > \delta \mid n \leq N \right) \Pr(n \leq N) + \Pr\left( T^{-1/2} \sum_{j=1}^n \| f_{0+j}(\theta_0) \| > \delta \mid n > N \right) \Pr(n > N) \leq \Pr\left( T^{-1/2} \sum_{j=1}^N \| f_{0+j}(\theta_0) \| > \delta \right) + \Pr(n > N) < \eta
\]

for all \( T \geq T^* = \max(T^*_1, T^*_2). \)

Similar arguments can be used to establish \( T^{-1}[D^T(\theta_0, \hat{\tau}_0) - D^T(\theta_0, \tau)] = o_p(1), \ i = 1, 2. \) By (S-1), \( T^{-1}[D^T(\theta_0, \hat{\tau}_0) - D^T(\theta_0, \tau)] = T^{-1}[\Omega^T(\theta_0, \hat{\tau}_0) - \Omega^T(\theta_0, \tau)] + o_p(1), \) so the result follows by Assumptions 8(d.ii) and (d.iv), substituting \( q_{0+j}(\theta_0) \) for \( f_{0+j}(\theta_0) \) in the previous argument. Finally, Assumption 7(c) yields

(S-2) \( T^{-1} D^T(\theta_0, \hat{\tau}_0) = J_i + o_p(1), \ i = 1, 2. \)

Combining (S-2) with Assumption 8(b) yields \( \hat{\nu}_{X^*} \xrightarrow{P} V_{X^*} \) using Slutsky’s theorem. \( X^*_T \xrightarrow{d} X^* \) follows from (S-2), Assumption 7(b), and the continuous mapping theorem.

**Q.E.D.**

**Proof of Theorem 6:** Let \( \hat{X}^*_T(s) = T^{-1/2} W^{1/2}_T \sum_{i=1}^{[sT]} f_i(\theta_0, \tilde{\xi}_0) \) and \( X^*_T(s) = T^{-1/2} V^{1/2}_{ff} \sum_{i=1}^{[sT]} f_i(\theta_0, \xi_0). \) Assumption 9(i) implies \( X^*_T(s) \Rightarrow W(s), \) while Assumptions 9(ii) and (iii) imply \( \hat{\nu}_{X^*} \Rightarrow MW(1) \) and \( \hat{\nu}_{X^*} \Rightarrow W(s) - sPW(1), \) where \( M = I_k - P, \) and \( P = V^{1/2}_{ff} \Gamma(V^{1/2}_{ff})^{-1} \Gamma^{1/2} V^{1/2}_{ff}. \) Hence, \( \hat{\nu}_{X^*} - \tilde{s} \hat{\nu}_{X^*} \Rightarrow W(s) - sW(1) = \tilde{W}(s). \) 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 $\zeta$. Hence, the asymptotic distribution of the stability component of the gen-S statistics, which only involves $\tilde{X}_T^*(s) - s\tilde{X}_T^*(1)$, is the same as in Theorems 1 and 2. On the other hand, $S_T(\theta_0, \hat{\zeta}_0) = \tilde{X}_T^*(1)\tilde{X}_T^*(1) \Rightarrow W(1)MWf(1) \sim \chi^2(k - p_\zeta)$. Moreover, $\tilde{X}_T^*(s) - s\tilde{X}_T^*(1)$ converges to a Brownian Bridge $\tilde{W}(s)$, which is independent of $W(1)$, showing that $S_T(\theta_0, \hat{\zeta}_0) \text{ and } \text{gen-S}_T(\theta_0, \hat{\zeta}_0)$ are asymptotically independent.}

**Q.E.D.**

**Proof of Theorem 7:** In the derivation of the split-sample statistics, $D_T(\theta_0, \cdot)$ and $F_T(\theta_0)$ are replaced with their counterparts that use $f_i(\theta_0, \hat{\zeta}_0)$ instead of $f_i(\theta_0)$ in their definition. Denote these by $D_T(\theta_0, \hat{\zeta}_0, \cdot)$ and $F_T(\theta_0, \hat{\zeta}_0)$. Under Assumption 10, $D_T(\theta_0, \hat{\zeta}_0, \cdot)$ and $F_T(\theta_0, \hat{\zeta}_0)$ are asymptotically independent, and since $\hat{\tau}_0$ only depends on $D_T(\theta_0, \hat{\zeta}_0, \cdot)$, it will be asymptotically independent of $F_T(\theta_0, \hat{\zeta}_0)$, so we can condition on $\hat{\tau}_0$ to obtain the distribution of split-$S_T(\theta_0, \hat{\zeta}_0, \hat{\tau}_0) = \sum_{i=1}^2 T_i^{-1}F_T(\theta_0, \hat{\zeta}_0, \hat{\tau}_0)\tilde{V}_{ij}^2(\theta_0, \hat{\zeta}_0, \hat{\tau}_0) - 1^{-1/2}F_T(\theta_0, \hat{\zeta}_0, \hat{\tau}_0) \Rightarrow W_f(\hat{\tau}_0) = \tilde{W}_f(\hat{\tau}_0) + \tau_\infty MW_f(1)$, where $\tilde{W}_f(\cdot)$ is a Brownian Bridge. Similarly, since $F_T(\theta_0, \hat{\zeta}_0, \hat{\tau}_0) = F_T(\theta_0, \hat{\zeta}_0, \hat{\tau}_0) - 1^{-1/2}F_T(\theta_0, \hat{\zeta}_0, \hat{\tau}_0) \Rightarrow W_f(\hat{\tau}_0) = \tilde{W}_f(\hat{\tau}_0) + \tau_\infty MW_f(1)$. So,

$$\frac{d}{\hat{\tau}_\infty} \left[ \tilde{W}_f(\hat{\tau}_\infty) + \tau_\infty MW_f(1) \right] \left[ \tilde{W}_f(\hat{\tau}_\infty) + \tau_\infty MW_f(1) \right]$$

$$= \tilde{W}_f(\hat{\tau}_\infty) + \tau_\infty MW_f(1)$$

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

**S.4. Testing General Hypotheses**

Let $g : \Theta \rightarrow \mathbb{R}^r$, where $r \leq p = \dim \theta$, and consider the problem of testing

(S-3) \[ H_0 : g(\theta) = 0 \] \text{ against } \[ H_1 : g(\theta) \neq 0. \]
If the function $g$ is injective (one-to-one), then this can be done using the methods in the paper by setting $\theta_0 = g^{-1}(0)$.

If $g$ is not injective, then define $\Theta_0 = \{\theta \in \Theta : g(\theta) = 0\}$, 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 $\eta$-level test can be obtained by performing $\eta$-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 $T(\theta_0) > \text{crit}$, where $T(\cdot)$ is the test statistic, the projection test can be performed simply by $\min_{\theta_0 \in \Theta_0} T(\theta) > \text{crit}$. Since $T(\cdot)$ is smooth, this minimization problem can be solved using faster numerical algorithms than simple grid search over $\Theta_0$.

An alternative to the projection method is the following. Suppose $r < p$, and there exists a partition of $\theta = (\theta_1, \theta_2)$, with $\dim \theta_1 = r$, such that the function $h(\theta) = (g(\theta), \theta_2)$ is injective, and let $h^{-1}(\cdot, \cdot)$ denote the inverse of $h$. Then, let $\varphi = g(\theta)$ and $\zeta = \theta_2$, so that $\theta = h^{-1}(\varphi, \zeta)$. If one is prepared to assume that $\zeta$ is strongly identified, and Assumptions 9 or 10 hold, then (S-3) can be tested by reparameterizing the problem into $\varphi, \zeta$, and testing $\varphi = 0$ concentrating out the strongly identified parameter $\zeta$, as described in Section 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 $g(\hat{\theta}_0) = 0$, evaluate all the test statistics at $\hat{\theta}_0$, and use the critical values given in Theorems 6 and 7, with $p_\zeta = p - r$.

S.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

\[ y_{1,t} = Z_t \theta + u_t, \]
\[ Y_{2,t} = Z_t \Pi_t + V_{2,t}, \quad t = 1, \ldots, T, \]

where $(y_{1,t}, Y_{2,t})$ is a $1 \times (1 + p)$ random vector, $u_t$ is a (structural) error, $\theta \in \mathbb{R}^p$ 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 $\Pi_t \in \mathbb{R}^{k \times p}$, $t = 1, \ldots, T$ 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

$$(S-6) \quad \text{ave-}S_T(\theta_0) = S_T(\theta_0) + \int_\varsigma \tilde{S}_T(\theta_0, s) \, d\nu_s$$

$$= \frac{1}{T} F_T(\theta_0) \tilde{V}_{ff}(\theta_0)^{-1} F_T(\theta_0)$$

$$+ \int_\varsigma \frac{\tilde{F}_T(\theta_0) \tilde{V}_{ff}(\theta_0)^{-1} \tilde{F}_T(\theta_0)}{Ts(1-s)} \, d\nu_s,$$

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

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

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

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

$$(S-7) \quad y_{1,t} = Y_{2,t} \theta_{t,T} + u_t,$$

$$(S-8) \quad Y_{2,t} = Z_t \Pi + V_{2,t}, \quad t = 1, \ldots, T.$$  

$\Pi$ 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 $\tau$:

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

where $\theta_1, \theta_2 \in \mathbb{R}^p$ and $\tau \in (0, 1)$.

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

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

\[
LM_1 = \frac{F_T(\theta_0)' \hat{\Sigma}^{-1/2} P_{\hat{M}} \hat{\Sigma}^{-1/2} F_T(\theta_0)}{T},
\]

\[
LM_2(s) = \frac{\tilde{F}_{ST}(\theta_0)' \hat{\Sigma}^{-1/2} P_{\hat{M}} \hat{\Sigma}^{-1/2} \tilde{F}_{ST}(\theta_0)}{Ts(1-s)},
\]

\[\hat{M} = \hat{\Sigma}^{-1/2} Q_T(\theta_0), \text{ and } \hat{\Sigma} = \hat{V}_{ff}(\theta_0).\]

It follows from Rossi (2005, Proposition 2(b)) that

\[
\text{Mean-Wald}_T^* \Rightarrow \psi_p + \int_{\varsigma} \hat{W}(s) \hat{W}(s)' s(1-s) \, dv_s,
\]

where \(\psi_p \sim \chi^2(p)\) and \(\hat{W}(\cdot)\) is a \(p \times 1\) standard Brownian Bridge process that is independent of \(\psi_p\). It is evident that the \(\text{Mean-Wald}_T^*\) statistic in (S-9) is generally different from the \(\text{ave-S}_T(\theta_0)\) statistic given in (S-6) above. The \(\text{Mean-Wald}_T^*\) statistic involves a projection of the moment vectors \(F_T(\theta_0)\) and \(\tilde{F}_{ST}(\theta_0)\) onto the space spanned by the Jacobian of the moment conditions, while the \(\text{ave-S}_T(\theta_0)\) statistic uses the full vectors \(F_T(\theta_0)\) and \(\tilde{F}_{ST}(\theta_0)\). An exception occurs when the model is just-identified, that is, \(k = p\), in which case, since \(\hat{M}\) is a square matrix and \(P_{\hat{M}} = I_k\), \(\text{ave-S}_T(\theta_0) = \text{Mean-Wald}_T^*\). 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. Finally, even when the break date is assumed to be known, that is, the support \(\varsigma\) of \(\nu_s\) contains a single point, and identification is strong, the \(\text{Mean-Wald}_T^*\) test is different from the split-KLM/CLR tests. To see this, observe that the strong-instrument asymptotic distribution of the \(\text{Mean-Wald}_T^*\) statistic under the null is \(\chi^2(2p)\), while that of the split-KLM and CLR statistics is \(\chi^2(p)\).

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

\[
y_{0,t} = Z_t \delta_t + u_{0,t},
\]

where \(y_{0,t} \equiv y_{1,t} - Y_{2,t} \theta_0\), against time-varying alternatives, for example, the local alternatives \(H_{AT}^* : \delta_t = \frac{1}{\sqrt{T}} (\delta_1 + \delta_2 1_{(t \geq \tau T)})\), \(\tau \in \varsigma\) in Rossi (2005, Assumption 2). Denote the moment function for this problem by \(g_t(\delta_t) = Z_t'(y_{0,t} -

\footnote{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).}
Identifying $Z, \delta,$ to distinguish it from $f_t(\theta)$ in the original model. Note that this auxiliary model is just-identified (the number of parameters in $\delta$ 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 $P_{\hat{M}} = I_k$ and $\hat{\Sigma}$ can be chosen as $\hat{V}_{ff}(\theta_0)$ in (S-10) and (S-11), so that the $\text{Mean-Wald}^*_T$ statistic in (S-9) coincides with $\text{ave-S}^*_T(\theta_0)$ in (S-6). In other words, the $\text{Mean-Wald}^*_T$ (S-9) for testing $H^*_0$ in the auxiliary regression (S-12) is identical to the $\text{ave-S}^*_T$ statistic (S-6) in the original model. The same connection holds for our exp-S statistic and the $\text{Exp-Wald}^*_T$ 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 (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

\[ y_t = \mu_t + u_t, \]

where $y_t = f_t(\theta_0), \mu_t = E[f_t(\theta_0)],$ and $u_t = f_t(\theta_0) - \mu_t$. The null hypothesis in the auxiliary regression is $H^*_0 : \mu_t = 0$ against a single-break alternative.

**S.6. Derivation of the Solution in Example RS**

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

\begin{align*}
\text{(S-13)} & \quad y_t = \beta E[y_{t+1} | I_t] + \gamma x_t + \epsilon_t, \\
\text{(S-14)} & \quad x_t = \rho x_{t-1} + (1 - \rho) \phi y_t + \eta_t.
\end{align*}

A solution to the model is given by

\begin{align*}
\text{(S-15)} & \quad y_t = \alpha_1 x_{t-1} + v^y_t, \\
\text{(S-16)} & \quad x_t = \rho_1 x_{t-1} + v^x_t,
\end{align*}

where $\alpha_1, \rho_1$ 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

\[ E[Y_{t+1} | I_t] = AY_t + Z_t, \]
where \( Y_t = (y_t, x_{t-1})' \), \( A = B^{-1}C \),

\[
B = \begin{pmatrix} \beta & \gamma \\ 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 \\ (1 - \rho)\varphi & \rho \end{pmatrix}, \quad Z_t = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}.
\]

(\( B \) 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 \( \lambda_1, \lambda_2 \) of the determinantal equation

\[
\det(C - \lambda B) = \rho - (1 + \beta \rho - \gamma \varphi(1 - \rho))\lambda + \beta \lambda^2 = 0.
\]

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 nonnegative, and the smallest root is

\[
\lambda_1 = \frac{1 + \beta \rho - \gamma \varphi(1 - \rho)}{2\beta} - \sqrt{\left(1 + \beta \rho - \gamma \varphi(1 - \rho)\right)^2 - 4\beta \rho}.
\]

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

\[
\lambda_2 = \frac{1 + \beta \rho - \gamma \varphi(1 - \rho)}{2\beta} + \sqrt{\left(1 + \beta \rho - \gamma \varphi(1 - \rho)\right)^2 - 4\beta \rho}
\]

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

We can then determine \( \alpha_1, \rho_1, v_t^y, \) and \( v_t^x \) by the method of undetermined coefficients. Using (S-15) to substitute for \( E[y_{t+1}|z_t] \) in (S-13) and rearranging yields

\[
y_t = (\beta \alpha_1 + \gamma) x_t + \varepsilon_t,
\]

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

\[
y_t = (\beta \alpha_1 + \gamma) \rho_1 x_{t-1} + \varepsilon_t + (\beta \alpha_1 + \gamma) v_t^x.
\]

So, since \( \beta, \rho_1 < 1 \),

\[
\alpha_1 = \frac{\gamma \rho_1}{1 - \beta \rho_1}.
\]
Substituting for $y_t$ into (S-14) using (S-15) and rearranging yields

\[
(S-22) \quad x_t = \rho x_{t-1} + (1 - \rho) \varphi (\alpha_1 x_{t-1} + v_t^y) + \eta_t
= \left( \rho + (1 - \rho) \varphi \alpha_1 \right) x_{t-1} + (1 - \rho) \varphi v_t^y + \eta_t.
\]

Upon substituting for $\alpha_1$ from (S-21), the first term yields a quadratic equation for $\rho_1$, which is the same as (S-17) above. Hence, the smallest root corresponds to

\[
\rho_1 = \lambda_1,
\]

where $\lambda_1$ is given by (S-18). Substituting for $\rho_1$ in (S-21) yields the solution for $\alpha_1$. Finally, substituting for $v_t^y$ into (S-22) using (S-20) and rearranging yields

\[
v_t^y = \frac{\eta_t + (1 - \rho) \varphi \varepsilon_i}{1 - (1 - \rho) \varphi (\beta \alpha_1 + \gamma)}.
\]

and using (S-20) yields

\[
v_t^y = \varepsilon_i + (\beta \alpha_1 + \gamma) v_t^y = \frac{\varepsilon_i + (\beta \alpha_1 + \gamma) \eta_t}{1 - (1 - \rho) \varphi (\beta \alpha_1 + \gamma)}.
\]

S.7. SUPPLEMENTAL MATERIAL FOR EMPIRICAL SECTION

S.7.1. The Baseline New Keynesian Phillips Curve

The baseline model is given by

\[
(S-23) \quad \pi_t = \varphi \pi_{t-1} = \beta E(\pi_{t+1} - \varphi \pi_t | I_t) + \lambda \widehat{m} c_t + \varepsilon_t,
\]

where

\[
\lambda = \frac{(1 - \alpha)(1 - \beta \alpha)}{\alpha} v, \quad v = \frac{a(\mu - 1)}{(\mu - a)},
\]

$\pi_t$ is inflation, $a$ is the labor elasticity of a Cobb–Douglas production function (the average labor share), $\mu$ is the desired mark-up under flexible prices, $\beta$ is a discount factor, $\varphi$ is the fraction of prices that are indexed to past inflation when they cannot be optimally reset, $\alpha$ is the probability that a price will be fixed in a given period, $\widehat{m} c_t$ is the log deviation of real marginal costs from their steady state, and $\varepsilon_t$ 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

$$\rho \Delta \pi_t = \kappa + E(\Delta \pi_{t+1} | \bar{Z}_t) + \frac{(1-\alpha)^2}{\alpha}\bar{x}_t + \varepsilon_t,$$

and $\bar{x}_t = \nu x_t$, where $x_t = \ln \frac{S_t}{a}$, $S_t$ is the labor share, and $\nu$ is calibrated to 0.25, using $a = \frac{2}{3}$ and $\mu = 1.2$. The constant $\kappa$ 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 $E[Z'_t u_t] = 0$, where $u_t$ is a scalar residual function,

$$u_t = \rho \Delta \pi_t - \Delta \pi_{t+1} - \frac{(1-\alpha)^2}{\alpha}\bar{x}_t - \frac{\kappa}{x_{tc}},$$

and $Z_t$ is a $1 \times k_z$ vector of instruments. The row vector $Y_t$ is $(\Delta \pi_t, \Delta \pi_{t+1}, \bar{x}_t)$, $X_t = 1$, and $b = (\rho, -1, -\frac{(1-\alpha)^2}{\alpha})'$. 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.

S.7.1.1. Data

We measure inflation as $\pi_t = \ln \left( \frac{P_t}{P_{t-1}} \right)$, 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 $S_t$, 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.

S.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 $M_a = I - a(a'a)^{-1}a'$ is a projection matrix. The vector of tested parameters is $\theta$, which has $p$ elements.

S.7.1.2.1. Computation of $S$ and CLR Statistics. In matrix notation, the empirical moments derived from equation (S-25) are

$$Z'u = Z'(Yb - Xc),$$
where \( Y, Z, X, \) and \( u \) are matrices consisting of the \( T \) stacked rows of \( Y_t, Z_t, X_t, \) and \( u_t, \) respectively. Under the null assumption \( H_0: \theta = (\alpha, \varphi) = (\alpha_0, \varphi_0) = \theta_0, \) \( b = b(\theta_0) \) is fixed. We define \( \hat{V}_{ff} \), the estimator asymptotic variance of \( \frac{1}{\sqrt{T}} Z'(Yb - Xc), \) as

(S-26) \[
\hat{V}_{ff} = (b' \otimes I_{k_z}) \hat{\Sigma}(b \otimes I_{k_z}),
\]

where \( \hat{\Sigma} \), the full sample HAC estimator, is

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

where \( \hat{\Gamma}_j \) is the “pre-whitened” \( \text{vec}(Z'_t \hat{\gamma}_i) \), that is, the residuals from a VAR(1) with coefficient matrix \( A, \) \( \hat{\gamma}_i \) is the \( r \)th row of the matrix \( M_X Y, \) \( \omega_j \) is the Bartlett kernel with \( [4(T/100)^{2/9}] \) as the lag truncation parameter, and \( A \) is the “recoloring” matrix. This procedure for estimating \( \text{var}(\frac{1}{\sqrt{T}} Z'(Yb - Xc)) \) is equivalent to using \( \text{vec}(Z'_t \hat{u}_t) \) as the empirical moments, where

\[
\hat{u} = Yb - X\hat{c}_1 = M_X Yb,
\]

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

We estimate \( c \) by minimizing the following objective function:

(S-28) \[
\hat{c}_2 = \arg \min_c \frac{1}{T} (Yb - Xc)' Z\hat{V}_{ff}^{-1} Z'(Yb - Xc),
\]

that is,

\[
\hat{c}_2 = (X'Z\hat{V}_{ff}^{-1}Z'X)^{-1}X'Z\hat{V}_{ff}^{-1}Z'Yb.
\]

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

\[
\sqrt{T} (\hat{c}_2 - c) \overset{d}{\rightarrow} (\Gamma'_{XZ} V_{ff}^{-1} \Gamma_{XZ})^{-1} \Gamma'_{XZ} V_{ff}^{-1/2} \xi,
\]

where \( \text{plim}_{T \rightarrow +\infty} \frac{1}{T} (Z X) = \Gamma_{XZ}, \) \( \text{plim}_{T \rightarrow +\infty} \hat{V}_{ff} = V_{ff}, \) a positive definite matrix, \( V_{ff}^{-1/2} \) is the symmetric square root matrix of \( V_{ff}^{-1}, \) and \( \xi \) is a \( k_z \times 1 \) standard normal random vector.

\(^2[4(T/100)^{2/9}] \) means the largest integer smaller than \( 4(T/100)^{2/9} \), which is four lags when using the sample 1966q1 to 2010q4.
Substituting \( \hat{c}_2 \) back into the objective function in (S-28), we derive

\[
b'Y'Z\hat{V}_{ff}^{-1/2}M\hat{r}_{ff}^{-1/2}\hat{r}_{ZX}\hat{V}_{ff}^{-1/2}Z'Yb,
\]
where \( \hat{r}_{ZX} = \frac{ZX}{T} \). Let \( L \) be a \( k_z \times (k_z - k_x) \) matrix such that

\[
LL' = M\hat{r}_{ff}^{-1/2}\hat{r}_{ZX} \quad \text{and} \quad L'L = I_{(k_z-k_x)}.
\]

The S statistic is

\[
S_T(\theta_0) = \frac{1}{T} b'Y'Z\hat{V}_{ff}^{-1/2}LL\hat{V}_{ff}^{-1/2}Z'Yb.
\]

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

\[
\hat{q}_T = L'\hat{V}_{ff}^{-1/2}Z'Y\nabla_{\theta_0} b.
\]

The estimators of the variance–covariance matrix of \( T^{-1/2}(\hat{\xi}_T, \vec{\hat{q}_T}) \) are

\[
\hat{V}_{\hat{\xi}\hat{\xi}} = I_{(k_z-k_x)},
\]

\[
\hat{V}_{\hat{q}\hat{q}} = (\nabla_{\theta_0} b' \otimes L'\hat{V}_{ff}^{-1/2})\hat{\Sigma}(b \otimes \hat{V}_{ff}^{-1/2}L) = \hat{V}_{\hat{q}\hat{q}},
\]

\[
\hat{V}_{\hat{q}\hat{\xi}} = (\nabla_{\theta_0} b' \otimes L'\hat{V}_{ff}^{-1/2})\hat{\Sigma}(\nabla_{\theta_0} b \otimes \hat{V}_{ff}^{-1/2}L).
\]

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

\[
\hat{D}_T = \frac{1}{T} \text{mat}(\vec{\hat{q}_T} - \hat{V}_{\hat{q}\hat{q}}\hat{\xi}_T),
\]

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

\[
\text{KLM}_T(\theta_0) = \frac{1}{T} \hat{\xi}_T\hat{D}_T(\hat{D}_T'\hat{D}_T)^{-1}\hat{D}_T'\hat{\xi}_T,
\]

\[
\text{JKLM}_T(\theta_0) = S_T(\theta_0) - \text{KLM}_T(\theta_0),
\]

\[
\text{CLR}_T(\theta_0) = \frac{1}{2} \left\{ S_T(\theta_0) - \frac{\mathbf{rk}(\theta_0)}{\mathbf{rk}(\theta_0)\sqrt{[S_T(\theta_0) + \mathbf{rk}(\theta_0)]^2 - 4\text{JKLM}_T(\theta_0)\mathbf{rk}(\theta_0)}} \right\},
\]

where \( \mathbf{rk}(\theta_0) \) is a rank statistic, which is a function of \( \hat{D}_T \) and its variance matrix \( \hat{V}_{\hat{q}\hat{q}} = \hat{V}_{\hat{q}\hat{q}} - \hat{V}_{\hat{q}\hat{q}}\hat{\xi}_T^{-1}\hat{\xi}_T\hat{V}_{\hat{q}\hat{q}} \). The results are based on the Kleibergen and Paap (2006) rank test.
S.7.1.2.2. 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:

$$
\Delta_D = \begin{bmatrix}
1 & 0 & \cdots & \cdots & 0 \\
-1 & 1 & 0 & \cdots & \vdots \\
0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & -1 & 1 & 0 \\
0 & \cdots & 0 & -1 & 1
\end{bmatrix}, \quad R = \begin{bmatrix}
1 & 0 & \cdots & \cdots & 0 \\
\cdot & 1 & 0 & \cdots & \vdots \\
\cdot & \cdot & \ddots & \ddots & \vdots \\
\cdot & \cdot & \cdot & r & 1 \\
\cdot & \cdot & \cdot & \cdot & r^{T-1}
\end{bmatrix},
$$

and

$$
r_T = \begin{bmatrix}
r \\
r^2 \\
\vdots \\
r^T
\end{bmatrix},
$$

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

(S-30) \hspace{1cm} \hat{U} = [\hat{u}, \ldots, \hat{u}],

where $\hat{u} = Yb - X\hat{c}_2$. The qLL-$\tilde{S}$ 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

(S-31) \hspace{1cm} \hat{F} = (\hat{U} \odot Z)\hat{\Phi}^{1/2} \quad \text{and} \quad \hat{H} = R\Delta_D \hat{F},

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

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

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

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

$$
TSSR_G = \sum_{i=1}^{k_z} \sum_{t=1}^{T} (\hat{g}_{i,t})^2,
$$

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

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

$$
\hat{F} = \nu_T d_N + N,
$$

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

\[
TSSR_N = \sum_{i=1}^{k} \sum_{t=1}^{T} (\hat{n}_{i,t})^2,
\]

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

The qLL-\( \tilde{S} \) statistic under \( H_0 \) is

\[
qLL-\tilde{S}_T(\theta_0) = TSSR_N - r \times TSSR_G.
\]

The qLL-S test is defined as

\[
qLL-S_T(\theta_0) = qLL-\tilde{S}_T(\theta_0) + \frac{10}{11} S_T(\theta_0).
\]

S.7.1.2.3. Computation of ave-S and exp-S Statistics. Partition \( Z = [Z_1', Z_2']' \), 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

\[
\bar{Z}'(Yb - Xc),
\]

where \( \bar{Z} \) is the \( T \times 2k_z \) “split-sample” instruments matrix:

\[
(S-32) \quad \bar{Z} = \bar{Z}(s) = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}.
\]

Similarly to equation (S-26), we define \( \hat{V}_s \), the estimator of the asymptotic variance of \( \text{Var}(\sqrt{T} \vec{Z}(Yb - Xc)) \), as

\[
(S-33) \quad \hat{V}_{ff,s} = (b' \otimes I_{2k_z}) \hat{\Sigma}_s (b' \otimes I_{2k_z}),
\]

where \( \hat{\Sigma}_s \) is the HAC estimator of the asymptotic variance of \( \frac{1}{\sqrt{T}} \vec{Z}Y \), which is computed as

\[
\hat{\Sigma}_s = P_R \left( \begin{bmatrix} s\hat{\Sigma}_1 & 0 \\ 0 & (1-s)\hat{\Sigma}_2 \end{bmatrix} \right) P_C,
\]

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

\[
(S-34) \quad \text{Var} \left[ \frac{1}{\sqrt{T}} \vec{Z}Y \right] = P_R \text{Var} \left[ \frac{1}{\sqrt{T}} \vec{Z}_1 Y_1 \quad \vec{Z}_2 Y_2 \right] P_C.
\]
Under the assumptions of the model, \( \text{plim}_{T \to +\infty} \hat{\Sigma}_s = \text{plim}_{T \to +\infty} \hat{\Sigma} = \Sigma \). So we estimate \( \hat{\Sigma}_s \) as

\[
(S-35) \quad \hat{\Sigma}_s = P_R \left( \begin{array}{cc} s\hat{\Sigma} & 0 \\ 0 & (1-s)\hat{\Sigma} \end{array} \right) P_C.
\]

where \( \hat{\Sigma} \) is defined in equation (S-27).

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

\[
(S-36) \quad \hat{c}_s = \arg \min_c \frac{1}{T} (Yb - Xc)' Z \hat{V}^{-1} Z (Yb - Xc),
\]

that is,

\[
\hat{c}_s = (X' \hat{Z} \hat{V}^{-1} Z' X)^{-1} X' \hat{Z} \hat{V}^{-1} Z' Yb
\]

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

\[
b' Y' \hat{Z} \hat{V}^{-1/2} M_{ff,s}^{-1/2} \hat{L}_X \hat{V}^{-1/2} Z' Yb,
\]

where \( \hat{L}_X \) is equal to \( \hat{X}' \). Similarly to the derivation of the weak instruments of Section S.7.1.2.1, we define the \( L \) matrix such that

\[
\hat{L} \hat{L}' = M_{ff,s}^{-1/2} \hat{L}_X \hat{V}^{-1/2} Z' Yb.
\]

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

\[
S(\theta_0; s) = \frac{1}{T} \left[ (Yb)' Z \hat{V}^{-1/2} \hat{L} \hat{L}' \hat{V}^{-1/2} Z' Yb \right].
\]

The average and exponential S statistics are defined as

\[
\text{ave}-S_T(\theta_0) = \frac{1}{t_u - t_l + 1} \sum_{t_0=t_l}^{t_u} S\left( \theta_0, \frac{t_b}{T} \right),
\]

\[
\text{exp}-S_T(\theta_0) = \log \left( \frac{1}{t_u - t_l + 1} \sum_{t_0=t_l}^{t_u} \exp \left[ -0.5 \times S\left( \theta_0, \frac{t_b}{T} \right) \right] \right),
\]
where $t_l = [0.15T]$, and $t_u = [0.85T]$. The stability parts of the ave-S and exp-S statistics are computed, respectively, as

\[
\text{ave-} S_T(\theta_0) = \text{ave-S}_T(\theta_0) - S_T(\theta_0),
\]

\[
\text{exp-} S_T(\theta_0) = \text{exp-S}_T(\theta_0) - S_T(\theta_0).
\]

S.7.1.2.4. Computation of Split-Sample Statistics. Define the $(2k_z - k_x) \times p$ Jacobian matrix $\hat{\bar{q}}_T$ as

\[
\hat{\bar{q}}_T = L' \hat{V}^{-1/2} Y' \nabla_{\theta_0} b.
\]

The estimators of the variance–covariance matrix of $T^{-1/2}(\hat{\xi}_T, \text{vec}(\hat{q}_T))$ are

\[
\hat{V}_{\hat{\xi}\hat{\xi}} = I_{(2k_z - k_x)},
\]

\[
\hat{V}_{\hat{q}\hat{q}} = (\nabla_{\theta_0} b' \otimes L' \hat{V}^{-1/2}_{ff,s}) \hat{\Sigma}(b \otimes \hat{V}^{-1/2}_{ff,s} L) = \hat{V}_{\hat{q}\hat{q}},
\]

\[
\hat{V}_{\hat{q} \hat{\xi}} = (\nabla_{\theta_0} b' \otimes L' \hat{V}^{-1/2}_{ff,s}) \hat{\Sigma}(\nabla_{\theta_0} b \otimes \hat{V}^{-1/2}_{ff,s} L).
\]

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

\[
\hat{D}_T = \frac{1}{T} \text{mat}(\text{vec}(\hat{q}_T) - \hat{V}_{\hat{q}\hat{q}} \hat{\xi}_T) \quad \text{and} \quad \hat{V}_{\hat{q}\hat{q},\hat{\xi}} = \hat{V}_{\hat{q}\hat{q}} - \hat{V}_{\hat{q} \hat{\xi}} \hat{V}_{\hat{\xi}\hat{\xi}}'.
\]

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

\[
\hat{t}_b = \arg \max_{t_b \in [t_l, t_u]} \text{vec}(\hat{D}_T)' \hat{V}^{-1}_{\hat{q}\hat{q},\hat{\xi}} \text{vec}(\hat{D}_T).
\]

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

S.7.1.3. Empirical Results

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

S.7.2. The NKPC With Autocorrelated Errors

Suppose the error term $\epsilon_t$ in (S-24) follows

\[
\epsilon_t = \phi \epsilon_{t-1} + \epsilon_t,
\]
FIGURE S.1.—90% and 95% $S$ and $qLL/exp/ave-S$ confidence sets for $\alpha$ and $\varrho$ in the NKPC: $\varrho\Delta\pi_t = E(\Delta\pi_{t+1}|I_t) + \frac{(1-\alpha)^2}{\alpha} \tilde{x}_t + \kappa + \epsilon_t$, $\tilde{x}_t$ is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta\pi_t$, and three lags of $\tilde{x}_t$. Newey and West (1987) HAC with prewhitening. Period: 1966q1–2010q4.
FIGURE S.2.—90% and 95% S, split-S, CLR, and split-CLR confidence sets for $\alpha$ and $\varrho$ in the NKPC:

$$\varrho \Delta \pi_t = E(\Delta \pi_{t+1} | I_t) + \frac{1-\omega}{\alpha} \tilde{x}_t + \kappa + \epsilon_t$$

$\tilde{x}_t$ is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $\tilde{x}_t$. Newey and West (1987) HAC with prewhitening. Period: 1966q1–2010q4.
Figure S.3.—90% and 95% $S$ and qLL/exp/ave-$\tilde{S}$ confidence sets for $\alpha$ and $\varrho$ in the NKPC:

$$\varrho \Delta \pi_t = E(\Delta \pi_{t+1}|\mathcal{I}_t) + (1-\alpha)\tilde{x}_t + \kappa + \epsilon_t,$$

$\tilde{x}_t$ is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $\tilde{x}_t$. Newey and West (1987) HAC with prewhitening. Period: 1966q1–2010q4.
$\rho \Delta \pi_t = E(\Delta \pi_{t+1} | I_t) + \frac{1 - \alpha^2}{\alpha} \bar{x}_t + \kappa + \bar{\varepsilon}_t$, $\bar{x}_t$, is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $\bar{x}_t$. Newey and West (1987) HAC with prewhitening. Period: 1984q1–2010q4.

Figure S.4.—90% and 95% S and qLL/exp/ave-S confidence sets for $\alpha$ and $\varphi$ in the NKPC.
Figure S.5.—90% and 95% S, split-S, CLR, and split-CLR confidence sets for $\alpha$ and $\varrho$ in the NKPC: $
abla \pi_t = E(\Delta \pi_{t+1} | Z_t) + \frac{\sqrt{\pi^2}}{\alpha} \tilde{x}_t + \kappa + \epsilon_t$, $\tilde{x}_t$ is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $\tilde{x}_t$. Newey and West (1987) HAC with prewhitening. Period: 1984q1–2010q4.
FIGURE S.6.—90% and 95% S and qLL/exp/ave-Ś confidence sets for $\alpha$ and $\varrho$ in the NKPC:

$\varrho \Delta \pi_t = E(\Delta \pi_{t+1}|I_t) + \frac{1-\alpha^2}{\alpha} \tilde{x}_t + \kappa + \varepsilon_t$, $\tilde{x}_t$ is 0.25 times the log of the labor share. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $\tilde{x}_t$. Newey and West (1987) HAC with prewhitening. Period: 1984q1–2010q4.
where \( e_t \) satisfies \( E(e_t | I_{t-1}) = 0 \). Lagging equation (S-24) by one period, multiplying by \( \phi \), and subtracting from (S-24) yields

\[
(S-37) \quad \rho(\Delta \pi_t - \phi \Delta \pi_{t-1}) = (1 - \phi) \kappa + E(\Delta \pi_{t+1} | I_t) - \phi E(\Delta \pi_t | I_{t-1})
\]

\[
+ \frac{(1 - \alpha)^2}{\alpha} (x_t - \phi x_{t-1}) + e_t.
\]

Define \( \nu_t = u_t - \phi u_{t-1} \), where \( u_t \) is given in (S-25). Equation (S-37) and \( E(e_t | I_{t-1}) = 0 \) imply \( E(\nu_t | I_{t-1}) = 0 \). The unconditional moment conditions are

\[
E[Z_t \nu_t] = 0.
\]

The structural parameter vector includes \( \phi \) and the confidence sets are three-dimensional (the constant is unrestricted and concentrated out as before).

S.7.2.1. Computational Details

The empirical moment condition is \( \frac{1}{T} \sum_{t=1}^{T} Z_t \nu_t \), where \( \nu_t = u_t - \phi u_{t-1} \) is

\[
\nu_t = \left( \frac{\rho(\Delta \pi_t - \Delta \pi_{t+1} - (1 - \alpha)^2}{\alpha} \tilde{x}_t - \kappa) 
\right)
\]

\[
- \phi \left( \frac{\rho(\Delta \pi_{t-1} - \Delta \pi_t - (1 - \alpha)^2}{\alpha} \tilde{x}_{t-1} - \kappa) 
\right)
\]

\[
= \rho(\Delta \pi_t - \phi \Delta \pi_{t-1}) - (\Delta \pi_{t+1} - \phi \Delta \pi_t) - \frac{(1 - \alpha)^2}{\alpha} (\tilde{x}_t - \phi \tilde{x}_{t-1})
\]

\[
- (1 - \phi) \kappa.
\]

By defining \( Y_t(\phi) \) as \((\Delta \pi_t - \phi \Delta \pi_{t-1}, \Delta \pi_{t+1} - \phi \Delta \pi_t, \tilde{x}_t - \phi \tilde{x}_{t-1}) \), \( X_t = 1 \), and \( b = (\rho, -1, -\frac{(1-\alpha)^2}{\alpha}) \), we rewrite the empirical moments as \( Z'(Y(\phi)b - Xc) \).

Let \( \theta = (\rho, \alpha, \phi)' \) and \( \phi \)' is the derivative of \( \phi \) with respect to “deep” parameters \( \rho \) and \( \alpha \) (see Section S.7.1.2.1), and \( Y_1 \) is the lagged values of matrix \( Y \). By using \( \tilde{b} = (b, 0, 0) \) in place of \( b \) and \( \nabla_{\theta_b} b \) in place of \( \nabla_{\theta_b} b \) and \( (Y(\phi) \ Y_1) \) in place of \( Y \), we compute the weak instruments, the generalized and split-sample tests following the same steps as described in Section S.7.1.2.1.
S.7.2.2. *Empirical Results*

Three-dimensional 95%-level confidence sets for \((\alpha, \rho, \phi)\) in (S-37) are constructed by inverting the various weak-identification robust tests, and they are plotted in Figures S.7, S.8, and S.9 for the sample 1966q1–2010q4, and Figures S.10, S.11, and S.12 for the sample 1984q1–2010q4. Table S.VIII provides the proportion of volume of the 95% confidence region relative to volume of the parallelepiped, and Table S.IX provides the point estimates of the structural parameters. The point estimates correspond to the values that minimize the S statistic.

![Confidence sets](image)

*Figure S.7.—95% S and qLL/exp/ave-S confidence sets for \((\alpha, \rho, \phi)\) in the NKPC with autocorrelated errors. Instruments: constant, two lags of \(\Delta\pi_t\), and three lags of \(\hat{x}_t\). Newey and West (1987) HAC with prewhitening. Period: 1966q1–2010q4.*
FIGURE S.8.—95%-level $S$ and qLL/exp/ave-$\tilde{S}$ confidence sets for $(\alpha, \varrho, \phi)$ in the NKPC with autocorrelated errors. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $\tilde{x}_t$. Newey and West (1987) HAC with prewhitening. Period: 1966q1–2010q4.
FIGURE S.9.—95%-level S, split-S, CLR, and split-CLR confidence sets for $(\alpha, \varrho, \phi)$ in the NKPC with autocorrelated errors. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $\hat{x}_t$. Newey and West (1987) HAC with prewhitening. Period: 1966q1–2010q4.
Figure S.10.—95%-level $S$ and $qLL/exp/ave-S$ confidence sets for $(\alpha, \Omega, \varphi)$ in the NKPC with autocorrelated errors. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $\tilde{x}_t$. Newey and West (1987) HAC with prewhitening. Period: 1984q1–2010q4.
Figure S.11.—95%-level $S$ and qLL/exp/ave-$\tilde{S}$ confidence sets for $(\alpha, \varrho, \phi)$ in the NKPC with autocorrelated errors. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $\tilde{x}_t$. Newey and West (1987) HAC with prewhitening. Period: 1984q1–2010q4.
FIGURE S.12.—95%-level S, split-S, CLR, and split-CLR confidence sets for \((\alpha, \varrho, \varphi)\) in the NKPC with autocorrelated errors. Instruments: constant, two lags of \(\Delta \pi_t\), and three lags of \(x_t\). Newey and West (1987) HAC with prewhitening. Period: 1984q1–2010q4.
S.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), that is, \( O_p(T^{-1/2}) \). This is the type of “partial instability” discussed in Li and Müller (2009).

S.7.3.1. The Model of Cogley and Sbordone (2008)

Cogley and Sbordone (2008) obtained 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, equations (7), (46), and (47)) translated to our notation, which are sufficient to write down a NKPC with time-varying trend inflation:

\[
\hat{\pi}_t - \varphi \hat{\pi}_{t-1} = \lambda_t \tilde{m}_t + \beta_t \left[ E(\hat{\pi}_{t+1} | I_t) - \varphi \hat{\pi}_t \right] - \varphi \Delta \tilde{r}_t + \gamma_t \tilde{D}_t + \epsilon_t, \tag{CS 46}
\]

\[
\tilde{D}_t = \varphi_t E[\hat{r}_{t,t+1} + \Delta \hat{y}_{t+1} + (\mu - 1)^{-1}(\hat{\pi}_{t+1} - \varphi \hat{\pi}_t) + \tilde{D}_{t+1} | I_t], \tag{CS 47}
\]

### TABLE S.IX

<table>
<thead>
<tr>
<th></th>
<th>1966q1–2010q4</th>
<th>1984q1–2010q4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \alpha )</td>
<td>( \varphi )</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.85</td>
<td>0.27</td>
</tr>
<tr>
<td>( \varphi )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \phi )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE S.VIII

**Volume of Confidence Regions as a Proportion of the Volume of the Parallelepiped**

\((\alpha, \varphi, \phi) \in [0.01, 0.99] \times [0, 1] \times [-0.99, 0.99]\)

<table>
<thead>
<tr>
<th>Confidence Regions</th>
<th>1966q1–2010q4</th>
<th>1984q1–2010q4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95%</td>
<td>90%</td>
</tr>
<tr>
<td>S</td>
<td>17.55</td>
<td>14.10</td>
</tr>
<tr>
<td>CLR</td>
<td>15.92</td>
<td>12.15</td>
</tr>
<tr>
<td>ave-S</td>
<td>13.19</td>
<td>11.30</td>
</tr>
<tr>
<td>exp-S</td>
<td>9.66</td>
<td>7.71</td>
</tr>
<tr>
<td>qLL-S</td>
<td>11.14</td>
<td>9.49</td>
</tr>
<tr>
<td>ave-( \tilde{S} )</td>
<td>20.33</td>
<td>17.50</td>
</tr>
<tr>
<td>exp-( \tilde{S} )</td>
<td>13.71</td>
<td>10.57</td>
</tr>
<tr>
<td>qLL-( \tilde{S} )</td>
<td>15.94</td>
<td>13.71</td>
</tr>
<tr>
<td>split-S</td>
<td>11.46</td>
<td>8.39</td>
</tr>
<tr>
<td>split-CLR</td>
<td>10.75</td>
<td>8.09</td>
</tr>
</tbody>
</table>
where $\hat{\pi}_t = \ln \Pi_t$, $\Pi_t$ is gross trend inflation, $r_{t,t+1} = \pi_{t+1} - i_t \approx \ln R_{t,t+1}$, $R_{t,t+1} = \frac{P_{t+1}}{P_t} \frac{1}{1+i_t}$, and $Y_t$ is real output. Hatted variables indicate stationary log-deviations of variables from their steady state or trend, that is, $\hat{\pi}_t = \pi_t - \bar{\pi}_t$, $\hat{mc}_t = \ln(mc_t) - \ln(\bar{mc}_t)$, and $\hat{r}_{t,t+1} + \Delta \hat{y}_{t+1} = r_{t,t+1} + \Delta y_{t+1} - \ln \frac{R_{t,t+1} Y_{t+1}}{Y_t}$. The term $R_{t,t+1} Y_{t+1}$ is the effective discount factor from period $t$ to $t+1$, whose steady-state value $R_{t,t+1} Y_{t+1}$ was denoted by $\beta$ 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 $\varphi_{1t}$ are functions of $\bar{\pi}_t$, and they are given by

$$\lambda_t = \chi_t (1 - \varphi_{2t}), \quad \beta_t = \Pi_t^{(1-\varrho)/\upsilon}, \quad \gamma_t = \chi_t (\Pi_t^{(1-\varrho)/\upsilon} - 1),$$

$$\chi_t = \upsilon \left( 1 - \frac{\alpha \Pi_t^{(1-\varrho)/(\mu-1)}}{\alpha \Pi_t^{(1-\varrho)/(\mu-1)}} \right), \quad \varphi_{1t} = \alpha \Pi_t^{(1-\varrho)/(\mu-1)}, \quad \varphi_{2t} = \varphi_{1t} \Pi_t^{(1-\varrho)/\upsilon}.$$

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

$$\xi_t = \lambda_t / \Delta_t, \quad \rho_t = \varrho / \Delta_t, \quad b_{1t} = \frac{\beta_t + \gamma_t (1 - \varrho \varphi_{1t}) \varphi_{1t} / (\mu - 1)}{\Delta_t},$$

$$b_{2t} = \frac{\gamma_t (1 - \varrho \varphi_{1t}) / (\mu - 1)}{\Delta_t}, \quad b_{3t} = \gamma_t \varphi_{1t} / \Delta_t,$$

$$\Delta_t = 1 + \varrho \left( \beta_t + \frac{\gamma_t \varphi_{1t}}{\mu - 1} \right).$$

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 $\hat{D}_t$, 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 $E(Z_t u_t) = 0$, 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.
S.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, $\bar{\Pi} = 1$, is

$$s_t = s + s_{\bar{\pi}}\bar{\pi}_t + o(\bar{\pi}_t).$$

The intercept and slope of this approximation for each of the coefficients in (S-38) are

$$\lambda = \frac{(1 - \alpha)^2}{\alpha} v, \quad \beta = 1, \quad \gamma = 0,$$

$$\lambda_{\bar{\pi}} = -\frac{(1 - \alpha)}{\alpha} (1 - \varrho) \frac{(a + \alpha \mu)}{(\mu - a)}, \quad \beta_{\bar{\pi}} = \frac{(1 - \varrho)}{\nu},$$

$$\gamma_{\bar{\pi}} = \frac{(1 - \alpha)(1 - \varrho)}{\alpha}, \quad \chi = \frac{(1 - \alpha)}{\alpha} v, \quad \varphi_1 = \alpha, \quad \varphi_2 = \alpha,$$

$$\chi_{\bar{\pi}} = -\frac{(1 - \varrho)}{\alpha(\mu - 1)} v, \quad \varphi_{1\bar{\pi}} = \frac{(1 - \varrho)}{(\mu - 1)}, \quad \varphi_{2\bar{\pi}} = \frac{\alpha(1 - \varrho)}{(\mu - 1)a}.$$

From equation (CS 7), we have

$$\ln m_{c_t} = -\ln \mu - \left(\frac{\mu - a}{a}\right) \ln \left(\frac{1 - \alpha\bar{\Pi}_t^{(1-\varrho)/(\mu-1)}}{1 - \alpha}\right),$$

$$+ \ln \left(\frac{1 - \alpha\bar{\Pi}_t^{(1-\varrho)/(\mu-1)a}}{1 - \alpha\bar{\Pi}_t^{(1-\varrho)/(\mu-1)}}\right).$$
Note that

\[
\frac{\partial m_c_t}{\partial \Pi_t} \bigg|_{\Pi=1} = \left( \frac{\mu - a}{a} \right) \frac{\alpha \left( \frac{1 - \varrho}{\mu - 1} \right)}{(1 - \alpha)} - \frac{\alpha \mu(1 - \varrho)}{(\mu - 1)a} + \frac{\alpha (1 - \varrho)}{(\mu - 1)}
\]

Therefore,

\[
\ln m_c_t = - \ln \mu + o_p(\bar{\pi}_t).
\]

Ignoring terms of order smaller than \(\bar{\pi}_t\), using \(\hat{\pi}_t = \pi_t - \bar{\pi}_t\), \(\Delta \hat{\pi}_{t+1} = o_p(\bar{\pi}_t)\), 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

(S-39) \[\pi_t - \varrho \pi_{t-1} = \lambda m_c_t + E(\pi_{t+1}|\mathcal{I}_t) - \varrho \pi_t \]

\[+ \pi_t \left[ \lambda m_c_{t+1} + \beta \pi_{t+2} - \varrho \pi_{t+1}|\mathcal{I}_t \right] \quad + \bar{\pi}_t \left[ \lambda m_c_{t+1} + \beta \pi_{t+2} - \varrho \pi_{t+1} + \gamma \mathcal{D}_{t+1} \right] + \varepsilon_t,\]

where

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

Next, we need to remove the infinite terms in \(D_t\) from the above equation. Lead (S-39) one period to get

\[\pi_{t+1} - \varrho \pi_t = \lambda m_c_{t+1} + E(\pi_{t+2} - \varrho \pi_{t+1}|\mathcal{I}_{t+1}) \]

\[+ \pi_{t+1} \left[ \lambda m_c_{t+1} + \beta \pi_{t+2} - \varrho \pi_{t+1} + \gamma \mathcal{D}_{t+1} \right] + \varepsilon_{t+1}.\]

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

(S-41) \[E(\pi_{t+1}|\mathcal{I}_t) - \varrho \pi_t \]

\[= \lambda E(m_c_{t+1}|\mathcal{I}_t) + E(\pi_{t+2} - \varrho \pi_{t+1}|\mathcal{I}_t) \]

\[+ \pi_{t+1} \left[ \lambda m_c_{t+1} + \beta \pi_{t+2} - \varrho \pi_{t+1} + \gamma \mathcal{D}_{t+1} \right].\]

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

\[\pi_t - \varrho \pi_{t-1} = \lambda m_c_t + E(\pi_{t+1} - \varrho \pi_t|\mathcal{I}_t) \]

\[+ \alpha E \left[ \pi_{t+1} - \varrho \pi_t - (\pi_{t+2} - \varrho \pi_{t+1}) - \lambda m_c_{t+1}|\mathcal{I}_t \right] \]
Collecting the constant and all the terms in $\bar{\pi}_t$, and substituting $x_t + \ln \mu$ for $\check{mc}_t$, where $x_t = \ln \frac{\bar{S}_t}{a}$ as before, the above can be written more compactly as

\[
\phi \Delta \pi_t = E(\Delta \pi_{t+1}|I_t) + \frac{(1 - \alpha)^2}{\alpha} \nu x_t + \kappa \nonumber \\
+ \alpha E \left( \phi \Delta \pi_{t+1} - \Delta \pi_{t+2} - \frac{(1 - \alpha)^2}{\alpha} \nu x_{t+1} - \kappa |I_t \right) + \sigma_t(\theta) \bar{\pi}_t \nonumber \\
+ \epsilon_t, \nonumber
\]

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

\[
\sigma_t(\theta) = (1 - \alpha) \lambda \ln \mu + \lambda \bar{\pi}_{t+1} - \phi \bar{\pi}_t \nonumber \\
- \alpha \left[ \lambda \bar{x}_{t+1} + \delta \bar{\pi}_{t+2} - \phi \bar{\pi}_{t+1} \right] \nonumber \\
+ \alpha \gamma \left[ (r_{t,t+1} + \Delta y_{t+1}) + (\mu - 1)^{-1} (\pi_{t+1} - \phi \bar{\pi}_t) \right]. \nonumber
\]

If we impose $\bar{\pi}_t = 0$, then $E(\phi \Delta \pi_{t+1} - \Delta \pi_{t+2} - \frac{(1 - a)^2}{\alpha} \nu x_{t+1} - \kappa |I_t) = 0$ and equation (S-42) collapses to the NKPC in (S-24).

We are interested in making inference about the structural parameter vector $\theta = (\alpha, \phi, \rho, \mu)^\prime$. 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: $\alpha \in (0, 1)$, $\rho \in [0, 1]$, and $\mu \in (1, +\infty)$.

Let $\eta^\pi_{t+1} = \pi_{t+1} - E(\pi_{t+1}|I_t)$ and $\eta^x_{t+1} = x_{t+1} - E(x_{t+1}|I_t)$. Then, substituting for the expectations terms in equation (S-39), we can define the residual function

\[
\nu_t \equiv u_t - \alpha u_{t+1} - \sigma_t(\theta) \bar{\pi}_t \nonumber \\
= \epsilon_t - \eta^\pi_{t+1} - \alpha \left( \phi \eta^\pi_{t+1} - \eta^\pi_{t+2} - \frac{(1 - \alpha)^2}{\alpha} \nu \eta^x_{t+1} \right), \nonumber
\]

where $u_t$ and $\sigma_t(\theta)$ are defined in equations (S-25) and (S-43), respectively. The assumption $E(\epsilon_t|I_{t-1}) = 0$ yields $E[\nu_t|I_t] = 0$, from which we can obtain unconditional moment restrictions of the form

\[
E[Z_i \nu_t] = E[f_i(\theta) - g_i(\theta) \bar{\pi}_t] = 0, \quad \text{where} \quad f_i(\theta) = Z_i'(u_t - \alpha u_{t+1}) \quad \text{and} \quad g_i(\theta) = Z_i \sigma_i(\theta). \nonumber
\]
S.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 \( T^{-1/2} \sum_{t=1}^{T} f_t(\theta) \), where \( f_t(\theta) \) is defined in (S-45), that is, ignoring the term \( g_t(\theta) \bar{\pi}_t \) that is unobserved. This depends on the behavior of \( T^{-1/2} \sum_{i=1}^{T} g_t(\theta) \bar{\pi}_t \), which we show is not negligible in general for moderate instability in \( \bar{\pi}_t \).

The model’s residual \( \nu_t \), defined in (S-44), satisfies \( E(\nu_t | I_{t-1}) = 0 \) 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 \( \text{var}(Z_t' \nu_t) = V_{ff} \) is finite and consistently estimable and the partial sum process \( T^{-1/2} \sum_{t=1}^{T} Z_t' \nu_t \) satisfies a FCLT.

Now, notice that \( \bar{\pi}_t(\theta) \) 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 \( \theta \), 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 \( T^{-1} \sum_{i=1}^{T} Z_t' Y_{\bar{\pi},t} \overset{p}{\rightarrow} s \Gamma_{ZY_{\bar{\pi}}} \) (which needs to hold uniformly in \( s \) for \( Y_{\bar{\pi},t} \) to be “asymptotically mse stationary”). Suppose also that \( \bar{\pi}_t = \frac{1}{\sqrt{T}} h(\frac{t}{T}) \), where \( h(\cdot) \) is a cadlag function with at most a finite number of discontinuities. Then,

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t(\theta) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t' \nu_t + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t' \bar{\pi}_t(\theta) \bar{\pi}_t
\]

\[
\overset{d}{\Rightarrow} V_{ff}^{1/2} \xi + \Gamma_{ZY_{\bar{\pi}}} l_{\bar{\pi}} \int_{0}^{1} h(r) \, dr,
\]

where \( \xi \) is a standard normal vector and the convergence of the second term follows from Li and Müller (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 \( T^{-1} \sum_{i=1}^{T} Z_t Y_{\bar{\pi},t} \overset{p}{\rightarrow} s \Gamma_{ZY_{\bar{\pi}}} \) is uniform in \( s \in [0, 1] \), then the above argument extends to the weak convergence of the partial sums:

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t(\theta) \Rightarrow V_{ff}^{1/2} W(s) + \Gamma_{ZY_{\bar{\pi}}} l_{\bar{\pi}} \int_{0}^{s} h(r) \, dr,
\]

where \( W(\cdot) \) is a multivariate standard Wiener process. So our stability and generalized tests also have power against this misspecification.

S.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 $V_{ff}^{-1/2} \Gamma_{ZY} l_z$. Given consistent estimators of $\hat{V}_{ff}$ and $\hat{\Gamma}_{ZY}$ of $V_{ff}$ and $\Gamma_{ZY}$, respectively, this can be done by premultiplying $rac{1}{\sqrt{T}} \hat{V}_{ff}^{-(1/2)_{:t}} \sum_{t=1}^{T} f_t(\theta_0)$ with the orthogonal projection matrix $M_{\hat{V}_{ff}^{-1/2} \hat{\Gamma}_{ZY} l_z}$ defined in (S-49) below. Note that this orthogonalization depends on the null value of the structural parameters $\theta_0$, so it has to be repeated for each value of $\theta_0 \in \Theta$.

S.7.3.5. Identification Fails When $\alpha \to 0$

We will show that when $\alpha$ 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 $\sum_{t=1}^{T} g_t(\theta)$ leads to sample moment conditions that are exactly zero over the full sample for all $\theta$ 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 $\alpha$. Rescaling $(u_t - \alpha u_{t+1})$ and $\sigma_t(\theta)$ by $\alpha$ yields

$$
\alpha \varrho \Delta \pi_t - \alpha [(1 + \varrho)(1 - \alpha)^2 \nu x_t]
+ \alpha^2 \Delta \pi_{t+2} + \alpha (1 - \alpha)^2 \nu x_{t+1} - (1 - \alpha)^3 \nu \ln \mu
$$

and

$$
\alpha \sigma_t(\theta) = -(1 - \alpha)(1 - \varrho) \frac{(a + \alpha \mu)}{(\mu - a)} [(1 - \alpha) \ln \mu + x_t]
+ \alpha \lambda_z (\pi_{t+1} - \varrho \pi_t)
+ \alpha (1 - \alpha)(1 - \varrho) \frac{(a + \alpha \mu)}{(\mu - a)} x_{t+1}
- \alpha^2 \frac{(1 - \varrho)}{\nu} (\pi_{t+2} - \varrho \pi_{t+1})
+ \alpha (1 - \alpha)(1 - \varrho)
\times [r_{t+1} + \Delta y_{t+1} + (\mu - 1)^{-1}(\pi_{t+1} - \varrho \pi_t)].
$$

Hence, when $\alpha \to 0$, the rescaled moment vector $\alpha [f_t(\theta) + g_t(\theta) \pi_t]$ in (S-45) becomes

$$
\frac{a(\mu - 1)}{\mu - a} \frac{Z_t'(x_t + \ln \mu)}{f_t(\theta=0)} + (1 - \varrho) \frac{(a + \mu)}{\mu - a} \frac{Z_t'(x_t + \ln \mu)}{g_t(\theta=0) \pi_t}.
$$
Since \( f_t(\theta) \) and \( g_t(\theta) = Z_t'Y_{\pi,t}l_{\pi} \) are collinear, and \( \hat{f}_{ZY_{\pi}l_{\pi}} = T^{-1} \sum_{t=1}^{T} g_t(\theta) \), \( M_f^{-1/2} \hat{f}_{ZY_{\pi}l_{\pi}} \hat{V}^{-1/2} = \sum_{t=1}^{T} f_t(\theta) = 0 \) for all \( \theta \in \Theta \cap \{ \alpha = 0 \} \). There is still information in the stability restrictions for \( \mu \), but not for \( \varrho \) because \( f_t(\theta) \) does not depend on \( \varrho \) in the limit as \( \alpha \to 0 \). So, stability restrictions lead to partial identification in this case. This discussion helps explain the empirical results.

S.7.3.6. Data

Real output growth (\( \Delta y_t \)) and the nominal interest rate (\( i_t \)) are measured as \( \Delta y_t = \ln(\frac{Y_{t+1}}{Y_t}) - 1 \) and \( i_t = \frac{b_t}{100000} \), 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).

S.7.3.7. Computational Details

Note that \( (u_t - \alpha u_{t+1}) \) can be rewritten as

\[
(S-46) \quad \varrho \Delta \pi_t - (1 + \alpha \varrho) \Delta \pi_{t+1} = \frac{(1 - \alpha)^2}{\alpha} \nu x_t + \alpha \Delta \pi_{t+2} + (1 - \alpha)^2 \nu x_{t+1} \frac{y_{tb}}{X_t c}
\]

where \( Y_t = (\Delta \pi_t, \Delta \pi_{t+1}, \Delta \pi_{t+2}, x_t, x_{t+1}) \), \( b = (\varrho, -(1 + \alpha \varrho), \alpha, -\frac{(1 - \alpha)^2}{\alpha} \nu, (1 - \alpha)^2 \nu) \), \( X_t = 1 \), and \( c = \kappa \). We use the following relationships: 

\[
\pi_{t+1} - \varrho \pi_t = (1 - \varrho) \pi_{t+1} + \varrho \Delta \pi_{t+1}; \quad \pi_{t+2} - \varrho \pi_{t+1} = \Delta \pi_{t+2} + (1 - \varrho) \pi_{t+1}; \quad \lambda_{\pi} + \gamma_{\pi}(\mu - 1)^{-1} = [\nu^{-1} + (\mu - 1)^{-1}(1 - \alpha)](1 - \varrho); \quad \text{and} \quad \alpha \lambda_{\pi} = \alpha(1 - \varrho) \nu^{-1}
\]

to rewrite \( \sigma_t(\theta) \) as

\[
\sigma_t(\theta) = (1 - \alpha) \lambda_{\pi} \ln \mu + Y_t b_{\pi} + (1 - \alpha)(1 - \varrho)(r_{t,t+1} + \Delta y_{t+1}) + \vartheta \pi_{t+1}, \quad Y_{\pi,t} l_{\pi}
\]

where

\[
b_{\pi} = \begin{bmatrix} 0 & v^{-1} + (\mu - 1)^{-1}(1 - \alpha)(1 - \varrho) \end{bmatrix},
\]

\[
\lambda_{\pi} = \begin{bmatrix} -v^{-1} \alpha(1 - \varrho) & -\alpha \lambda_{\pi} \end{bmatrix},
\]

\[
\vartheta = [v^{-1} + (\mu - 1)^{-1}] (1 - \alpha)^2 (1 - \varrho),
\]

\[
(S-47) \quad Y_{\pi,t} = \begin{bmatrix} X_t & Y_t & (r_{t,t+1} + \Delta y_{t+1}) & \pi_{t+1} \end{bmatrix},
\]

\[
(S-48) \quad l_{\pi} = \begin{bmatrix} (1 - \alpha) \lambda_{\pi} \ln \mu & b_{\pi} & \lambda_{\pi} & (1 - \alpha)(1 - \varrho) \end{bmatrix}.
\]
Computation of S and CLR Statistics. The estimator of the asymptotic variance of \( \frac{1}{\sqrt{T}} Z'(Yb - Xc) \), \( \hat{V}_{ff} \), is the one described in equation (S-26).

Let \( Y = (Y, X) \) and \( \hat{b} = (\hat{b}) \), such that the moment vector can be written more compactly as \( Z\hat{Y}b \). We eliminate the effect of trend inflation on the tests by premultiplying the sample moments by \( M_{V_{ff}^{-1/2}Z\hat{Y}b} \), where

\[
M_{V_{ff}^{-1/2}Z\hat{Y}b} = \text{I}_{k_z} - V_{ff}^{-1/2} \Gamma Z\hat{Y}b V_{ff}^{-1/2}.
\]

We replace \( V_{ff} \), \( \Gamma Z\hat{Y}b \) with their estimators \( \hat{V}_{ff} \), and \( \hat{\Gamma Z\hat{Y}b} \) to compute the tests.

Let \( L \) be a \( k_z \times (k_z - 1) \) matrix such that \( M_{\hat{V}_{ff}^{-1/2}Z\hat{Y}b} = LL' \) and \( L'L = \text{I}_{(k_z - 1)} \).

The S statistic is

\[
S_T(\theta_0) = \frac{1}{T} \left( \frac{1}{\hat{\xi}_T} \right) T(\hat{Y}b)Z\hat{V}_{ff}^{-1/2}LL'\hat{V}_{ff}^{-1/2}Z\hat{Y}b. \]

\( \frac{1}{\sqrt{T}} \hat{\xi}_T \) is an asymptotically standard normal vector of dimension \( k_z - 1 \). Note that, in computing the S statistic, the use of the projection matrix \( M_{\hat{V}_{ff}^{-1/2}Z\hat{Y}b} \) to annihilate the trend inflation effect in the limiting distribution is equivalent to first estimating \( c_{\hat{b}} \) by solving the following minimization problem:

\[
\hat{c}_{\hat{b}} = \text{arg min}_{c_{\hat{b}}} \frac{1}{T} (\hat{Y}b - Y_{\hat{b}}l_{\hat{b}}c_{\hat{b}})'Z\hat{V}_{ff}^{-1}Z'(\hat{Y}b - Y_{\hat{b}}l_{\hat{b}}c_{\hat{b}}),
\]

and then substituting \( \hat{c}_{\hat{b}} \) back into the objective function in (S-50). The KLM, JKLM, and CLR statistics are obtained using the formulas in Section S.7.1.2.1; see equation (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 Section S.7.1.2.2, but redefining \( \hat{u} \) in equation (S-30) and \( \hat{F} \) in equation (S-31) as

\[
\hat{u} = Yb - Xc = \hat{Y}b \quad \text{and} \quad \hat{F} = (\hat{U} \odot Z)\hat{V}_{ff}^{-1/2}M_{\hat{V}_{ff}^{-1/2}Z\hat{Y}b}^{-1/2}.
\]

respectively.

Computation of the exprave-S and Split-Sample Statistics. Let the \( 2k_z \times 1 \) split-sample moment vector be defined as \( Z\hat{Y}b \), where \( Z \) is the split-sample instrument matrix defined in (S-32). The estimator of the asymptotic variance of \( \frac{1}{\sqrt{T}} Z\hat{Y}b \) 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 \(\mathcal{L}'\), defined as \(\mathcal{L}' = I_2 \otimes M_{PP^{-1/2}F_{ZYz}l_*}\), where \(M_{PP^{-1/2}F_{ZYz}l_*}\) is given by (S-49). The remaining calculations are the same as in Section S.7.1.2.3.

S.7.3.8. Empirical Results

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

![Figure S.13](image_url)

**Figure S.13.**—95\%-level S and generalized S confidence sets for \((\alpha, \varrho, \mu)\) in the NKPC with trend inflation. Instruments: constant, two lags of \(\Delta \pi_t\), and three lags of \(x_t\). Newey and West (1987) HAC with prewhitening. Period: 1966q1–2010q4.
FIGURE S.14.—95%-level S and stability S confidence sets for $\alpha, \phi, \mu$ in the NKPC with trend inflation. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $x_t$, Newey and West (1987) HAC with prewhitening. Period: 1966q1–2010q4.
FIGURE S.15.—95%-level $S$, split-$S$, CLR, and split-CLR confidence sets for $\alpha$, $\sigma$, $\mu$ in the NKPC with trend inflation. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $x_t$. Newey and West (1987) HAC with prewhitening. Period: 1966q1–2010q4.
Figure S.16.—95%-level S and generalized S confidence sets for $\alpha$, $\varrho$, $\mu$ in the NKPC with trend inflation. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $x_t$. Newey and West (1987) HAC with prewhitening. Period: 1984q1–2010q4.
Figure S.17.—95%-level S and generalized S confidence sets for $\alpha$, $\varrho$, $\mu$ in the NKPC with trend inflation. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $x_t$. Newey and West (1987) HAC with prewhitening. Period: 1984q1–2010q4.
Figure S.18.—95%-level S, split-S, CLR, and split-CLR confidence sets for $\alpha$, $\varrho$, $\mu$ in the NKPC with trend inflation. Instruments: constant, two lags of $\Delta \pi_t$, and three lags of $x_t$. Newey and West (1987) HAC with prewhitening. Period: 1984q1–2010q4.
TABLE S.X
VOLUME OF CONFIDENCE REGIONS AS A PROPORTION OF THE VOLUME OF THE PARALLELEPIPED \((\alpha, \varphi, \mu) \in [0.01, 0.99] \times [0, 1] \times [1.01, 1.40]\)

<table>
<thead>
<tr>
<th>Confidence Regions:</th>
<th>1966q1—2010q4</th>
<th>1984q1—2010q4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95%</td>
<td>90%</td>
</tr>
<tr>
<td>S</td>
<td>90.78</td>
<td>89.71</td>
</tr>
<tr>
<td>CLR</td>
<td>89.78</td>
<td>88.42</td>
</tr>
<tr>
<td>ave-S</td>
<td>29.78</td>
<td>27.90</td>
</tr>
<tr>
<td>exp-S</td>
<td>23.73</td>
<td>21.26</td>
</tr>
<tr>
<td>qLL-S</td>
<td>23.54</td>
<td>21.49</td>
</tr>
<tr>
<td>ave-\tilde{S}</td>
<td>28.84</td>
<td>27.02</td>
</tr>
<tr>
<td>exp-\tilde{S}</td>
<td>20.15</td>
<td>13.87</td>
</tr>
<tr>
<td>qLL-\tilde{S}</td>
<td>23.27</td>
<td>19.92</td>
</tr>
<tr>
<td>split-S</td>
<td>30.73</td>
<td>29.03</td>
</tr>
<tr>
<td>split-CLR</td>
<td>26.50</td>
<td>24.13</td>
</tr>
</tbody>
</table>

REFERENCES


Dept. of Economics, Business School, The University of Western Australia, 35 Stirling Highway—M251, Crawley, WA 6009, Australia; leandro.magnusson@uwa.edu.au

and

Dept. of Economics and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Manor Road, Oxford, OX1 3UQ, U.K.; sophocles.mavroeidis@economics.ox.ac.uk.

Manuscript received October, 2010; final revision received March, 2014.