West, Wong and Anatolyev Last revised October 2006

"Instrumental Variables Estimation of Heteroskedastic Linear Models Using All Lags of Instruments"

Additional Appendix, Part 1

This appendix contains results omitted from the submitted paper to save space. It includes:

1. Tables with additional simulation results (in a separate document called Additional Appendix, part 2)

2. Tables with additional asymptotic comparisons for GMM

3. Construction of positive semidefinite $\hat{S}$

4. Approximation error in computing asymptotic variances of optimal estimator

5. Details on procedures used in simulations

6. Formal statement of asymptotic results

7. Some worked out examples

8. Some details on maximum likelihood calculations

2. Tables with additional asymptotic comparisons

The first of the three sets of entries we present includes those in Table 1:

ϕ	θ	γ	γ1	GMM1	GMM4	GMM12
0.500	0.000	0.500	0.100	1.0007e+00	1.0000e+00	1.0000e+00
0.500	0.000	0.500	0.300	1.0085e+00	1.0000e+00	1.0000e+00
0.500	0.000	0.600	0.100	1.0007e+00	1.0000e+00	1.0000e+00
0.500	0.000	0.600	0.300	1.0090e+00	1.0000e+00	1.0000e+00
0.500	0.000	0.700	0.100	1.0006e+00	1.0000e+00	1.0000e+00
0.500	0.000	0.700	0.300	1.0084e+00	1.0000e+00	1.0000e+00
0.500	0.000	0.800	0.100	1.0004e+00	1.0000e+00	1.0000e+00
0.500	0.000	0.800	0.300	1.0068e+00	1.0001e+00	1.0000e+00
0.500	0.000	0.900	0.100	1.0002e+00	1.0000e+00	1.0000e+00
0.500	0.000	0.900	0.300	1.0045e+00	1.0001e+00	1.0000e+00
0.900	0.000	0.500	0.100	1.0015e+00	1.0000e+00	1.0000e+00
0.900	0.000	0.500	0.300	1.0207e+00	1.0003e+00	1.0000e+00
0.900	0.000	0.600	0.100	1.0021e+00	1.0001e+00	1.0000e+00
0.900	0.000	0.600	0.300	1.0328e+00	1.0013e+00	1.0000e+00
0.900	0.000	0.700	0.100	1.0029e+00	1.0002e+00	1.0000e+00
0.900	0.000	0.700	0.300	1.0519e+00	1.0057e+00	1.0000e+00
0.900	0.000	0.800	0.100	1.0036e+00	1.0006e+00	1.0000e+00
0.900	0.000	0.800	0.300	1.0877e+00	1.0228e+00	1.0003e+00
0.900	0.000	0.900	0.100	1.0038e+00	1.0012e+00	1.0001e+00
0.900	0.000	0.900	0.300	1.2308e+00	1.1396e+00	1.0353e+00
0.500	0.500	0.500	0.100	1.3864e+00	1.0044e+00	1.0000e+00
0.500	0.500	0.500	0.300	1.5430e+00	1.0060e+00	1.0000e+00
0.500	0.500	0.600	0.100	1.3869e+00	1.0046e+00	1.0000e+00
0.500	0.500	0.600	0.300	1.5513e+00	1.0071e+00	1.0000e+00
0.500	0.500	0.700	0.100	1.3837e+00	1.0048e+00	1.0000e+00
0.500	0.500	0.700	0.300	1.5467e+00	1.0087e+00	1.0000e+00
0.500	0.500	0.800	0.100	1.3761e+00	1.0049e+00	1.0000e+00
0.500	0.500	0.800	0.300	1.5268e+00	1.0102e+00	1.0000e+00
0.500	0.500	0.900	0.100	1.3633e+00	1.0048e+00	1.0000e+00
0.500	0.500	0.900	0.300	1.4899e+00	1.0109e+00	1.0000e+00
0.900	0.500	0.500	0.100	1.2009e+00	1.0024e+00	1.0000e+00
0.900	0.500	0.500	0.300	1.3502e+00	1.0045e+00	1.0000e+00
0.900	0.500	0.600	0.100	1.2058e+00	1.0029e+00	1.0000e+00
0.900	0.500	0.600	0.300	1.3900e+00	1.0081e+00	1.0000e+00
0.900	0.500	0.700	0.100	1.2099e+00	1.0037e+00	1.0000e+00
0.900	0.500	0.700	0.300	1.4431e+00	1.0177e+00	1.0000e+00
0.900	0.500	0.800	0.100	1.2115e+00	1.0050e+00	1.0000e+00
0.900	0.500	0.800	0.300	1.5263e+00	1.0462e+00	1.0004e+00
0.900	0.500	0.900	0.100	1.2068e+00	1.0066e+00	1.0001e+00
0.900	0.500	0.900	0.300	1.8070e+00	1.2130e+00	1.0464e+00
0.500	0.700	0.500	0.100	1.9682e+00	1.0587e+00	1.0002e+00
0.500	0.700	0.500	0.300	2.2499e+00	1.0688e+00	1.0002e+00
0.500	0.700	0.600	0.100	1.9727e+00	1.0600e+00	1.0002e+00
0.500	0.700	0.600	0.300	2.2838e+00	1.0758e+00	1.0002e+00
0.500	0.700	0.700	0.100	1.9710e+00	1.0615e+00	1.0002e+00
0.500	0.700	0.700	0.300	2.3021e+00	1.0856e+00	1.0003e+00
0.500	0.700	0.800	0.100	1.9604e+00	1.0626e+00	1.0002e+00
0.500	0.700	0.800	0.300	2.2949e+00	1.0970e+00	1.0003e+00
0.500	0.700	0.900	0.100	1.9373e+00	1.0623e+00	1.0002e+00
0.500	0.700	0.900	0.300	2.2486e+00	1.1055e+00	1.0007e+00
0.900	0.700	0.500	0.100	1.8223e+00	1.0487e+00	1.0001e+00
0.900	0.700	0.500	0.300	2.2860e+00	1.0659e+00	1.0002e+00
0.900	0.700	0.600	0.100	1.8322e+00	1.0508e+00	1.0001e+00
0.900	0.700	0.600	0.300	2.3813e+00	1.0794e+00	1.0002e+00
0.900	0.700	0.700	0.100	1.8418e+00	1.0545e+00	1.0002e+00
0.900	0.700	0.700	0.300	2.5221e+00	1.1087e+00	1.0003e+00
0.900	0.700	0.800	0.100	1.8471e+00	1.0600e+00	1.0002e+00
0.900	0.700	0.800	0.300	2.7621e+00	1.1790e+00	1.0015e+00
0.900	0.700	0.900	0.100	1.8364e+00	1.0660e+00	1.0006e+00
0.900	0.700	0.900	0.300	3.5979e+00	1.5081e+00	1.0762e+00
0.500	0.900	0.500	0.100	3.1150e+00	1.3531e+00	1.0398e+00
0.500	0.900	0.500	0.300	3.4094e+00	1.3501e+00	1.0376e+00
0.500	0.900	0.600	0.100	3.1281e+00	1.3572e+00	1.0400e+00
0.500	0.900	0.600	0.300	3.4704e+00	1.3654e+00	1.0381e+00
0.500	0.900	0.700	0.100	3.1390e+00	1.3635e+00	1.0406e+00
0.500	0.900	0.700	0.300	3.5454e+00	1.3942e+00	1.0403e+00
0.500	0.900	0.800	0.100	3.1438e+00	1.3717e+00	1.0419e+00
0.500	0.900	0.800	0.300	3.6454e+00	1.4451e+00	1.0479e+00
0.500	0.900	0.900	0.100	3.1326e+00	1.3789e+00	1.0442e+00
0.500	0.900	0.900	0.300	3.8806e+00	1.5681e+00	1.0961e+00

ϕ	θ	γ	γ1	GMM1	GMM4	GMM12
0.500	0.900	0.900	0.300	3.8806e+00	1.5681e+00	1.0961e+00
0.900	0.900	0.500	0.100	5.9907e+00	1.8148e+00	1.0905e+00
0.900	0.900	0.500	0.300	8.2621e+00	1.9855e+00	1.1005e+00
0.900	0.900	0.600	0.100	6.0084e+00	1.8256e+00	1.0909e+00
0.900	0.900	0.600	0.300	8.5692e+00	2.0579e+00	1.1036e+00
0.900	0.900	0.700	0.100	6.0377e+00	1.8431e+00	1.0919e+00
0.900	0.900	0.700	0.300	9.1515e+00	2.1982e+00	1.1109e+00
0.900	0.900	0.800	0.100	6.0832e+00	1.8723e+00	1.0948e+00
0.900	0.900	0.800	0.300	1.0550e+01	2.5427e+00	1.1420e+00
0.900	0.900	0.900	0.100	6.1319e+00	1.9168e+00	1.1053e+00
0.900	0.900	0.900	0.300	2.0983e+01	5.1149e+00	1.7430e+00
0.500	0.950	0.500	0.100	3.5313e+00	1.5062e+00	1.1039e+00
0.500	0.950	0.500	0.300	3.7512e+00	1.4743e+00	1.0921e+00
0.500	0.950	0.600	0.100	3.5417e+00	1.5093e+00	1.1037e+00
0.500	0.950	0.600	0.300	3.7898e+00	1.4827e+00	1.0903e+00
0.500	0.950	0.700	0.100	3.5530e+00	1.5158e+00	1.1042e+00
0.500	0.950	0.700	0.300	3.8465e+00	1.5052e+00	1.0903e+00
0.500	0.950	0.800	0.100	3.5639e+00	1.5264e+00	1.1063e+00
0.500	0.950	0.800	0.300	3.9411e+00	1.5534e+00	1.0974e+00
0.500	0.950	0.900	0.100	3.5701e+00	1.5408e+00	1.1119e+00
0.500	0.950	0.900	0.300	4.2748e+00	1.7130e+00	1.1686e+00
0.900	0.950	0.500	0.100	1.0453e+01	2.8519e+00	1.3738e+00
0.900	0.950	0.500	0.300	1.4453e+01	3.1823e+00	1.4006e+00
0.900	0.950	0.600	0.100	1.0450e+01	2.8683e+00	1.3749e+00
0.900	0.950	0.600	0.300	1.4805e+01	3.2994e+00	1.4084e+00
0.900	0.950	0.700	0.100	1.0462e+01	2.8935e+00	1.3768e+00
0.900	0.950	0.700	0.300	1.5527e+01	3.5107e+00	1.4248e+00
0.900	0.950	0.800	0.100	1.0510e+01	2.9372e+00	1.3829e+00
0.900	0.950	0.800	0.300	1.7463e+01	4.0197e+00	1.4868e+00
0.900	0.950	0.900	0.100	1.0647e+01	3.0224e+00	1.4076e+00
0.900	0.950	0.900	0.300	3.6363e+01	8.5720e+00	2.6159e+00
0.500	−0.500	0.500	0.100	1.1028e+00	1.0014e+00	1.0000e+00
0.500	−0.500	0.500	0.300	1.0709e+00	1.0016e+00	1.0000e+00
0.500	−0.500	0.600	0.100	1.1046e+00	1.0015e+00	1.0000e+00
0.500	−0.500	0.600	0.300	1.0759e+00	1.0018e+00	1.0000e+00
0.500	−0.500	0.700	0.100	1.1070e+00	1.0015e+00	1.0000e+00
0.500	−0.500	0.700	0.300	1.0815e+00	1.0020e+00	1.0000e+00
0.500	−0.500	0.800	0.100	1.1099e+00	1.0015e+00	1.0000e+00
0.500	−0.500	0.800	0.300	1.0879e+00	1.0021e+00	1.0000e+00
0.500	−0.500	0.900	0.100	1.1134e+00	1.0015e+00	1.0000e+00
0.500	−0.500	0.900	0.300	1.0951e+00	1.0021e+00	1.0000e+00
0.900	−0.500	0.500	0.100	1.0136e+00	1.0003e+00	1.0000e+00
0.900	−0.500	0.500	0.300	1.0134e+00	1.0007e+00	1.0000e+00
0.900	−0.500	0.600	0.100	1.0138e+00	1.0004e+00	1.0000e+00
0.900	−0.500	0.600	0.300	1.0227e+00	1.0020e+00	1.0000e+00
0.900	−0.500	0.700	0.100	1.0145e+00	1.0007e+00	1.0000e+00
0.900	−0.500	0.700	0.300	1.0387e+00	1.0065e+00	1.0000e+00
0.900	−0.500	0.800	0.100	1.0158e+00	1.0011e+00	1.0000e+00
0.900	−0.500	0.800	0.300	1.0703e+00	1.0227e+00	1.0002e+00
0.900	−0.500	0.900	0.100	1.0176e+00	1.0017e+00	1.0000e+00
0.900	−0.500	0.900	0.300	1.1991e+00	1.1304e+00	1.0325e+00
0.500	−0.900	0.500	0.100	1.2390e+00	1.0438e+00	1.0049e+00
0.500	−0.900	0.500	0.300	1.1550e+00	1.0312e+00	1.0032e+00
0.500	−0.900	0.600	0.100	1.2478e+00	1.0457e+00	1.0051e+00
0.500	−0.900	0.600	0.300	1.1750e+00	1.0359e+00	1.0037e+00
0.500	−0.900	0.700	0.100	1.2571e+00	1.0478e+00	1.0054e+00
0.500	−0.900	0.700	0.300	1.1967e+00	1.0415e+00	1.0044e+00
0.500	−0.900	0.800	0.100	1.2669e+00	1.0498e+00	1.0057e+00
0.500	−0.900	0.800	0.300	1.2204e+00	1.0483e+00	1.0057e+00
0.500	−0.900	0.900	0.100	1.2767e+00	1.0515e+00	1.0061e+00
0.500	−0.900	0.900	0.300	1.2495e+00	1.0589e+00	1.0106e+00
0.900	−0.900	0.500	0.100	1.0299e+00	1.0065e+00	1.0007e+00
0.900	−0.900	0.500	0.300	1.0218e+00	1.0042e+00	1.0003e+00
0.900	−0.900	0.600	0.100	1.0312e+00	1.0070e+00	1.0007e+00
0.900	−0.900	0.600	0.300	1.0327e+00	1.0065e+00	1.0004e+00
0.900	−0.900	0.700	0.100	1.0330e+00	1.0078e+00	1.0008e+00
0.900	−0.900	0.700	0.300	1.0501e+00	1.0123e+00	1.0005e+00
0.900	−0.900	0.800	0.100	1.0354e+00	1.0087e+00	1.0009e+00
0.900	−0.900	0.800	0.300	1.0826e+00	1.0300e+00	1.0010e+00
0.900	−0.900	0.900	0.100	1.0387e+00	1.0098e+00	1.0010e+00
0.900	−0.900	0.900	0.300	1.2103e+00	1.1376e+00	1.0336e+00

The second set of entries allows non-normal eta:

φ	θ	γ	γ1	unconditional excess kurtosis	κη	GMM1	GMM4	GMM12
0.500	−0.900	0.500	0.100	7.000	5.941	1.1550e+00	1.0312e+00	1.0032e+00
0.500	−0.900	0.900	0.100	7.000	3.897	1.2624e+00	1.0527e+00	1.0068e+00
0.500	−0.500	0.500	0.100	7.000	5.941	1.0709e+00	1.0016e+00	1.0000e+00
0.500	−0.500	0.900	0.100	7.000	3.897	1.1053e+00	1.0017e+00	1.0000e+00
0.500	0.000	0.500	0.100	7.000	5.941	1.0085e+00	1.0000e+00	1.0000e+00
0.500	0.000	0.900	0.100	7.000	3.897	1.0012e+00	1.0000e+00	1.0000e+00
0.500	0.500	0.500	0.100	7.000	5.941	1.5430e+00	1.0060e+00	1.0000e+00
0.500	0.500	0.900	0.100	7.000	3.897	1.4089e+00	1.0065e+00	1.0000e+00
0.500	0.900	0.500	0.100	7.000	5.941	3.4094e+00	1.3501e+00	1.0376e+00
0.500	0.900	0.900	0.100	7.000	3.897	3.3407e+00	1.4221e+00	1.0517e+00
0.500	0.950	0.500	0.100	7.000	5.941	3.7512e+00	1.4743e+00	1.0921e+00
0.500	0.950	0.900	0.100	7.000	3.897	3.7596e+00	1.5760e+00	1.1184e+00
0.900	−0.900	0.500	0.100	7.000	5.941	1.0218e+00	1.0042e+00	1.0003e+00
0.900	−0.900	0.900	0.100	7.000	3.897	1.0466e+00	1.0195e+00	1.0017e+00
0.900	−0.500	0.500	0.100	7.000	5.941	1.0134e+00	1.0007e+00	1.0000e+00
0.900	−0.500	0.900	0.100	7.000	3.897	1.0294e+00	1.0112e+00	1.0006e+00
0.900	0.000	0.500	0.100	7.000	5.941	1.0207e+00	1.0003e+00	1.0000e+00
0.900	0.000	0.900	0.100	7.000	3.897	1.0285e+00	1.0112e+00	1.0007e+00
0.900	0.500	0.500	0.100	7.000	5.941	1.3502e+00	1.0045e+00	1.0000e+00
0.900	0.500	0.900	0.100	7.000	3.897	1.3149e+00	1.0260e+00	1.0010e+00
0.900	0.900	0.500	0.100	7.000	5.941	8.2621e+00	1.9855e+00	1.1005e+00
0.900	0.900	0.900	0.100	7.000	3.897	8.0517e+00	2.2904e+00	1.1543e+00
0.900	0.950	0.500	0.100	7.000	5.941	1.4453e+01	3.1823e+00	1.4006e+00
0.900	0.950	0.900	0.100	7.000	3.897	1.3902e+01	3.6701e+00	1.5155e+00
0.500	−0.900	0.500	0.100	1.167	1.000	1.2195e+00	1.0409e+00	1.0045e+00
0.500	−0.900	0.900	0.100	1.750	1.000	1.2721e+00	1.0518e+00	1.0063e+00
0.500	−0.500	0.500	0.100	1.167	1.000	1.0957e+00	1.0015e+00	1.0000e+00
0.500	−0.500	0.900	0.100	1.750	1.000	1.1109e+00	1.0016e+00	1.0000e+00
0.500	0.000	0.500	0.100	1.167	1.000	1.0015e+00	1.0000e+00	1.0000e+00
0.500	0.000	0.900	0.100	1.750	1.000	1.0004e+00	1.0000e+00	1.0000e+00
0.500	0.500	0.500	0.100	1.167	1.000	1.4128e+00	1.0046e+00	1.0000e+00
0.500	0.500	0.900	0.100	1.750	1.000	1.3764e+00	1.0053e+00	1.0000e+00
0.500	0.900	0.500	0.100	1.167	1.000	3.1669e+00	1.3521e+00	1.0394e+00
0.500	0.900	0.900	0.100	1.750	1.000	3.1896e+00	1.3901e+00	1.0459e+00
0.500	0.950	0.500	0.100	1.167	1.000	3.5692e+00	1.4993e+00	1.1015e+00
0.500	0.950	0.900	0.100	1.750	1.000	3.6222e+00	1.5498e+00	1.1133e+00
0.900	−0.900	0.500	0.100	1.167	1.000	1.0248e+00	1.0058e+00	1.0006e+00
0.900	−0.900	0.900	0.100	1.750	1.000	1.0381e+00	1.0114e+00	1.0011e+00
0.900	−0.500	0.500	0.100	1.167	1.000	1.0111e+00	1.0004e+00	1.0000e+00
0.900	−0.500	0.900	0.100	1.750	1.000	1.0183e+00	1.0033e+00	1.0001e+00
0.900	0.000	0.500	0.100	1.167	1.000	1.0032e+00	1.0000e+00	1.0000e+00
0.900	0.000	0.900	0.100	1.750	1.000	1.0082e+00	1.0028e+00	1.0001e+00
0.900	0.500	0.500	0.100	1.167	1.000	1.2240e+00	1.0027e+00	1.0000e+00
0.900	0.500	0.900	0.100	1.750	1.000	1.2331e+00	1.0103e+00	1.0002e+00
0.900	0.900	0.500	0.100	1.167	1.000	6.3606e+00	1.8438e+00	1.0923e+00
0.900	0.900	0.900	0.100	1.750	1.000	6.5910e+00	2.0041e+00	1.1160e+00
0.900	0.950	0.500	0.100	1.167	1.000	1.1109e+01	2.9082e+00	1.3787e+00
0.900	0.950	0.900	0.100	1.750	1.000	1.1429e+01	3.1754e+00	1.4321e+00

The third set of entries allows u(t) to impound two shocks: $u_t = e_{t+2} - \theta e_{t+1} + v_{t+2} - dv_{t+1}$, $v_t \sim$ i.i.d. $(0,\sigma_v^2)$, v_t independent of e_t. σ_v^2 chosen so that half the variance of u_t was due to each signal.

φ	θ	γ	γ1	−d	GMM1	GMM4	GMM12
0.900	0.950	0.900	0.100	2.000	1.0031e+00	1.0000e+00	1.0000e+00
0.900	0.950	0.900	0.100	1.111	1.0011e+00	1.0000e+00	1.0000e+00
0.900	0.950	0.900	0.100	1.000	1.0011e+00	1.0000e+00	1.0000e+00
0.900	0.950	0.900	0.100	0.900	1.0011e+00	1.0000e+00	1.0000e+00
0.900	0.950	0.900	0.100	0.500	1.0031e+00	1.0000e+00	1.0000e+00
0.900	0.950	0.900	0.100	−0.500	1.4666e+00	1.0197e+00	1.0000e+00
0.900	0.950	0.900	0.100	−0.900	7.0517e+00	2.1574e+00	1.1707e+00
0.900	0.950	0.900	0.100	−0.950	9.9322e+00	2.8845e+00	1.3860e+00
0.900	0.950	0.900	0.100	−1.000	1.2041e+01	3.4372e+00	1.5763e+00
0.900	0.950	0.900	0.100	−1.053	9.9322e+00	2.8845e+00	1.3860e+00
0.900	0.950	0.900	0.100	−1.111	7.0517e+00	2.1574e+00	1.1707e+00
0.900	0.950	0.900	0.100	−2.000	1.4666e+00	1.0197e+00	1.0000e+00

3. Construction of positive semidefinite $\hat{S}$ may be insured by the following algorithm, applicable in the symmetric case in which $Eu_t u_{t+n} e_{t-i} e_{t-j}' = Eu_t e_{t-i} e_{t-j}' = 0$ for $i \neq j$. Assume the regression disturbance may be written

$$u_t = c_{q+1}'e_{t+q} + c_q'e_{t+q-1} + ... + c_1'e_{t+1} + u_{2t} \equiv u_{1t} + u_{2t}$$

where the c_i's are $(r\times1)$ and the u_{it}'s are scalars. After a parametric model is fit to z_t, yielding $\hat{e}_t$, a regression of $\hat{u}_t$ on $\hat{e}_{t+q}$,..., and $\hat{e}_{t+1}$ delivers $\hat{c}_j$'s, a fitted value $\hat{u}_{1t}$ and a residual $\hat{u}_{2t}$. We can use these sample quantities to construct p.s.d. $\hat{S}$. (To keep notation uncluttered, we use $\hat{e}$ rather than $\hat{e}^{\dagger}$.)

As in the DGP used in the third set of asymptotic calculations described above, let u_{2t} be conditionally homoskedastic and mean independent of e_{t+q-j} for $j\geq0$. Then $S=S_1+S_2$, $S_i = \Sigma_{j=-q}^{q} Ee(t-j)u_{it-j}u_{it}e(t)'$. A sufficient condition for $\hat{S}$ to be p.s.d. is that each $\hat{S}_i$ is p.s.d.. It may help to note that if u_t is conditionally homoskedastic, one can set $\hat{u}_{1t}=0$, $\hat{S}_1=0$ and $\hat{u}_{2t}=\hat{u}_t$.

In each S_i, the first row and column is zero except for $S_i(1,1) = \Sigma_{j=-q}^{q} Eu_{it}u_{it-j}$. The $(Tr\times Tr)$ submatrix that remains is band diagonal, with blocks on the diagonal $= Ee_{t-j}e_{t-j}'u_{it}^2$, blocks on the immediate off-diagonal $= Ee_{t-j}e_{t-j}'u_{it}u_{it+1}$, ..., and, finally, blocks $= Ee_{t-j}e_{t-j}'u_{it}u_{it+q}$. For i=1,2, estimate Eu_{it}^2, $Eu_{it}u_{it+1}$, ..., $Eu_{it}u_{it+q}$ so that the estimates are consistent with u_{it} following an MA(q) process. Possible techniques for doing so include fitting an MA(q), or using sample autocovariances after checking that they obey the necessary inequalities (see, e.g., Box and Jenkins (1976, p71) for the inequalities when q=2). Then the implied estimate of $\hat{S}_i(1,1)$ will be positive for i=1,2. In S_2, $Ee_te_t'u_{2t}u_{2t+j} = Ee_te_t'Eu_{2t}u_{2t+j}$, and $\hat{S}_2$ will be p.s.d. if one constructs the $(Tr\times Tr)$ submatrix using $T^{-1}\Sigma\hat{e}_t\hat{e}_t'$ for Ee_te_t' and the estimates of Eu_{2t}^2, $Eu_{2t}u_{2t+1}$, ..., $Eu_{2t}u_{2t+q}$ that were just mentioned.

For the $(Tr\times Tr)$ submatrix of S_1: One can use the $\hat{c}_j$'s and estimates of a GARCH model for e_t. Alternatively, one can estimate $\hat{S}_1$ as $\hat{S}_1=T^{-1}\Sigma_{t=p}^{T}\hat{R}_t\otimes(\hat{e}_t\hat{e}_t')$, where $\hat{R}_t$ is a $T\times T$ symmetric band diagonal matrix with an estimate of $E_tu_{1t+j}^2$ on the diagonal (j=0,...,T-1), of $E_tu_{1t+j}u_{1t+j+1}$ on the first off-diagonal (j=0,...,T-2), ... , of $E_tu_{1t+j}u_{1t+j+q}$ on the q'th off-diagonal (j=0,...,T-q-1). In the spirit of the procedure in Schwert (1989), one can construct $\hat{R}_t$ as follows. (1)Let abs($\hat{e}_t$) denote the $r\times1$ vector obtained by taking the absolute value of $\hat{e}_t$ element by element. Estimate a VAR(p) in abs($\hat{e}_t$), abs($\hat{e}_t$) $= \hat{a}_0 + \hat{a}_1'$abs($\hat{e}_{t-1}$) + ... + $\hat{a}_p'$abs($\hat{e}_{t-p}$) + residual $\equiv \hat{A}_1'W_{t-1}$ + residual, where a_0 is $(r\times1)$, a_1, ..., a_p are $(r\times r)$, $\hat{A}_1'$ is $r\times(pr+1)$ and W_{t-1} is $(pr+1)\times1$. (2)Compute $\hat{A}_j$ for j=1,...,T+1, where $\hat{A}_j'W_t$ is the j-period ahead forecast of abs($\hat{e}_{t+j}$) implied by the VAR. (3)Observe that an estimate of $E_t e_{t+j} e_{t+j}'$ consistent with Schwert (1989) is $\frac{\pi}{2}\hat{A}_j' W_t W_t' \hat{A}_j$. Corresponding estimates of conditional moments of u_{1t} are thus: $E_t u_{1t+j}^2$: $\frac{\pi}{2}(W_t'\hat{A}_{j+1}\hat{c}_1)^2 + ... + \frac{\pi}{2}(W_t'\hat{A}_{j+q+1}\hat{c}_{q+1})^2$; $E_t u_{1t+j} u_{1t+j+1}$: $\frac{\pi}{2}(W_t'\hat{A}_{j+2}\hat{c}_1)(W_t'\hat{A}_{j+2}\hat{c}_2) + ... + \frac{\pi}{2}(W_t'\hat{A}_{j+q+1}\hat{c}_q)(W_t'\hat{A}_{j+q+1}\hat{c}_{q+1})$; ... ; $E_t u_{1t+j} u_{1t+j+q}$: $\frac{\pi}{2}(W_t'\hat{A}_{j+q+1}\hat{c}_1)(W_t'\hat{A}_{j+q+1}\hat{c}_{q+1})$.

4. Approximation error in computing asymptotic variances of optimal estimator

To compute the asymptotic variance of the optimal estimator, we truncated the infinite sum $\Sigma_{j=0}^{\infty} g_j^* e_{t-j}$ at lag j=100, and approximated the optimal variance by that resulting when 101 g^*'s and lagged e_t's were used. Analytical arguments and numerical computations about to be detailed were used to bound the difference between the asymptotic variance of the estimates using $\Sigma_{j=0}^{\infty} g_j^* e_{t-j}$ and those using 101 g^*'s, for several DGPs. The two variances were invariably within a fraction of a percent of one another. Thus, in our DGPs, the decline in the elements of Ψ^* (in our DGPs, the second column of Ψ^* is $\sigma_e^2 \phi^j$) and reversion of the elements of S^* to their conditionally homoskedastic counterparts (in our DGPs, $Ee_{t-j}^2 u_t^2 = Ee_t^2 Eu_t^2 + c\gamma^j$ for a suitable constant c) are sufficiently rapid that terms beyond the hundredth have negligible effect on asymptotic variances. (A clarifying note for those who have read the asymptotic theory presented in section 6 below: as explained below equation (AA6.5), when we compute the g^*'s from S^* and Ψ^* of dimension 101, the result should properly be denoted $\{g_{j,100}^*\}$ to distinguish it from $\{g_j^*\}$. Our calculations do take into account the difference between $\{g_{j,100}^*\}$ and $\{g_j^*\}$, as well as the effects of g_j^* for $j \geq 100$.)

In bounding the approximation error, we do not attempt to establish as tight a bound as possible, but merely to check whether the error is acceptably small. In presenting our formulas, we use certain notation not used elsewhere in the paper or Additional Appendix. We define "g_j^*" and "g_{jJ}^*" to be the scalar second elements of the (1×2) vectors elsewhere called "g_j^*" and "$g_{j,T}^*$". As stated in the text, we set J=101 in our calculations. Also, $\sigma^2 \equiv Ee_t^2$, $\lambda_i \equiv Ee_t^2 e_{t-i}^2$. We consider the case when $\phi > 0$, $\theta > 0$ and $\phi \neq \theta$.

When $u_t = e_{t+2} - \theta e_{t+1}$, $|\theta| < 1$, and $z_t = \phi z_{t-1} + e_t$, it may be shown that the g_j^*'s satisfy:

(AA4.1)
$$g_0^* = \frac{\sigma^2}{\phi-\theta}\frac{N_0^*}{D_0^*},\ N_0^* = \sum_{i=1}^{\infty}\frac{\theta^{i-1}(\phi^i-\theta^i)}{\lambda_{i+1}},\ D_0^* = 1+\lambda_1\sum_{i=1}^{\infty}\frac{\theta^{2i}}{\lambda_{i+1}},$$
$$g_j^* = -g_0^*\lambda_1\sum_{i=1}^{\infty}\frac{\theta^{j+2i}}{\lambda_{j+i+1}} + \frac{\sigma^2}{\phi-\theta}\sum_{i=1}^{\infty}\frac{\theta^{i-1}(\phi^{j+i}-\theta^{j+i})}{\lambda_{j+i+1}}$$

Let $\tilde{g}_j^*$ be the g_j^*-series calculated using (AA4.1), but with infinite sums truncated at K-1 for some K>J:

(AA4.2) $\tilde{g}_0^* = \dfrac{\sigma^2}{\phi-\theta}\dfrac{\tilde{N}_0^*}{\tilde{D}_0^*}$, $\tilde{N}_0^* = \sum_{i=1}^{K-1}\dfrac{\theta^{i-1}(\phi^i-\theta^i)}{\lambda_{i+1}}$, $\tilde{D}_0^* = 1+\lambda_1\sum_{i=1}^{K-1}\dfrac{\theta^{2i}}{\lambda_{i+1}}$,

$$\tilde{g}_j^* = -\tilde{g}_0^*\lambda_1\sum_{i=1}^{K-1}\frac{\theta^{j+2i}}{\lambda_{j+i+1}} + \frac{\sigma^2}{\phi-\theta}\sum_{i=1}^{K-1}\frac{\theta^{i-1}(\phi^{j+i}-\theta^{j+i})}{\lambda_{j+i+1}}.$$

Let $V_J=[\sigma^2\Sigma_{j=0}^{J-1}(g_{jJ}^*\phi^j)]^{-1}$ be the value we compute for the asymptotic variance, $V=[\sigma^2\Sigma_{j=0}^{\infty}(g_j^*\phi^j)]^{-1}$ the exact value. Our aim is to compute a numerical upper bound to the absolute value of the percentage error in our calculation of the asymptotic variance V,

(AA4.3) $|V_J\text{-}V|/V = |[\sigma^2\Sigma_{j=0}^{J-1}(g_{jJ}^*\phi^j)]^{-1}-[\sigma^2\Sigma_{j=0}^{\infty}(g_j^*\phi^j)]^{-1}|/[\sigma^2\Sigma_{j=0}^{\infty}(g_j^*\phi^j)]^{-1} \equiv \Xi[\Sigma_{j=0}^{J-1}(g_{jJ}^*\phi^j)]^{-1}$,

$$\Xi \equiv |\Sigma_{j=0}^{J-1}(g_{jJ}^*\phi^j)-\Sigma_{j=0}^{\infty}(g_j^*\phi^j)| \le$$

$$|\Sigma_{j=0}^{J-1}(g_{jJ}^*\phi^j)-\Sigma_{j=0}^{J-1}(\tilde{g}_j^*\phi^j)|+|\Sigma_{j=0}^{J-1}[(\tilde{g}_j^*\phi^j)-(g_j^*\phi^j)]|+|\Sigma_{j=J}^{\infty}(g_j^*\phi^j)| \equiv \Xi_1+\Xi_2+\Xi_3.$$

Since $[\Sigma_{j=0}^{J-1}(g_{jJ}^*\phi^j)]^{-1}$ and Ξ_1 may be computed directly, our task is to bound Ξ_2 and Ξ_3 by functions that may be calculated from the parameters of the DGP. Tedious but straightforward algebra may be used to do so. To present the final result, we define the following notation:

(AA4.4) $\lambda^* \equiv \sup_i Ee_t^2e_{t-i}^2$, $\lambda_* \equiv \inf_i Ee_t^2e_{t-i}^2$,

$C \equiv [\lambda_*(1\text{-}\phi\theta)(1\text{-}\theta^2)]^{-1}\{1+[(\lambda_*/\lambda^*)\theta^2/(1\text{-}\theta^2)]\}^{-1}$,

$A \equiv C\lambda^*\theta^2/[\lambda_*(1\text{-}\theta^2)] - \{\theta/[\lambda_*(1\text{-}\theta^2)(\phi\text{-}\theta)]\}$,

$B \equiv \phi/[\lambda_*(1\text{-}\phi\theta)(\phi\text{-}\theta)]$,

$\Lambda \equiv \sigma^4\{[\theta^{2J}A^2/(1\text{-}\theta^2)] + [2(\theta\phi)^J AB/(1\text{-}\phi\theta)] + [\phi^{2J}B^2/(1\text{-}\phi^2)]\}$.

$M_2 \equiv \theta^{2K}/[\lambda_*(1\text{-}\theta^2)]$,

$M_3 \equiv \{[\theta^{K-1}\phi^K/(1\text{-}\phi\theta)] + [\theta^{2K-1}/(1\text{-}\theta^2)]\} / \lambda_*$

$M_1 \equiv \sigma^2|\phi\text{-}\theta|^{-1}\{M_3(1+[\lambda^*\theta^2/\lambda_*(1\text{-}\theta^2)]) + M_2(\lambda^*/\lambda_*)[\phi/(1\text{-}\phi\theta)+\theta/(1\text{-}\theta^2)]\}$,

$M_4 \equiv \sigma^2M_2C\lambda^* + M_1M_2\lambda^* + \sigma^2M_3|\phi\text{-}\theta|^{-1}$,

$M_5 \equiv \theta^2M_1\lambda^*/[\lambda_*(1\text{-}\theta^2)]$.

Then

(AA4.5) $\Xi_2 \le [M_4(1\text{-}\phi^J)/(1\text{-}\phi)] + \{M_5[1\text{-}(\phi\theta)^J]/(1\text{-}\phi\theta)\}$,

$\Xi_3 \le [\Lambda\phi^{2J}/(1\text{-}\phi^2)]^{1/2}$.

As stated above, we set J=101 in the results presented in the paper. Some numerical results for two values of J and several parameter sets, with $\sigma^2=1$ and K=500 throughout:

1. $\phi=.9$, $\theta=.95$, $\gamma=.9$, $\gamma_1=.1$: J=100, $|V_J-V|/V < 3x10^{-4}$; J=300, $|V_J-V|/V < 3x10^{-13}$.
2. $\phi=.9$, $\theta=.95$, $\gamma=.9$, $\gamma_1=.3$: J=100, $|V_J-V|/V < 1x10^{-3}$; J=300, $|V_J-V|/V < 8x10^{-13}$.
3. $\phi=.5$, $\theta=.51$, $\gamma=.9$, $\gamma_1=.1$: J=100 and J=300, $|V_J-V|/V < 4x10^{-16}$.

5. Details on procedures used in the simulations

Each sample was of size $M+T_{max}+2$, where the first M observations were presample values generated so that the actual data were not directly influenced by initial conditions, and observations M+1 through M+T+2 observations were used in the actual estimation. T_{max}=10,000, and T=250, 500, 1000 or 10,000. Thus the results for T=250, for example, rely on the first 250 observations of the samples used for T=500, 1000 and 10,000. We proceeded as follows:

1. Generate a $(M+T_{max}+2)$x1 vector of i.i.d. N(0,1) variables. Call these $\{\eta_t\}$, t=1,...,M+T+2.

2. Set $\sigma_1^2=Ee_t^2$, $e_1=\sigma_1\eta_1$. (Here and subsequently, $\sigma_t>0$ is the positive square root of σ_t^2.) For t=2,...,M+T+2, set $\sigma_t^2 = \omega + \gamma_1 e_{t-1}^2 + \gamma_2\sigma_{t-1}^2$, $e_t = \sigma_t\eta_t$, $z_t = \phi z_{t-1}+e_t-\varsigma e_{t-1}$.

3. For t=M+1,...,M+T, set $y_t = \beta_0+z_t\beta_1 + c_2e_{t+2}+c_1e_{t+1} = c_2e_{t+2}+c_1e_{t+1}$.

The vector of η_t's was held fixed across DGPs, though different DGPs of course involved different z_t's and y_t's.

For convenience in presentation, in the text and in the remainder of this discussion, we renumber the observations used in estimation as 1, ..., T+2. (rather than M+1,..,M+T+2).

Some details on how we implemented the proposed estimator, $\hat{\beta}^* \equiv (\Sigma_{t=1}^T\hat{Z}_t^*X_t')^{-1}(\Sigma_{t=1}^T\hat{Z}_t^*y_t)$. We obtained $\hat{Z}_t^*$ as follows.

(a)Estimate β in (4.1a) (t=1 to T) by least squares. Call the residuals $\hat{u}_t$, with sample variance $\hat{\sigma}_u^2$ and first autocorrelation $\hat{\rho}_u$. The estimate of the first autocovariance of $\hat{u}_t$ was: $\hat{\sigma}_{u,1}=\hat{\rho}_u\hat{\sigma}_u^2$ if $|\hat{\rho}_u|<.4999$, $\hat{\sigma}_{u,1}=\text{sgn}(\hat{\rho}_u)(.4999\hat{\sigma}_u^2)$ if $|\hat{\rho}_u|\geq.4999$.

(b)Estimate $\phi_0,...,\phi_4$ in (4.2a) (t=5 to T) by least squares. Let $\hat{e}_t^\dagger$ denote the residual, with sample variance $\hat{\sigma}_{e\dagger}^2$.

(c)Estimate (4.2b) with a least squares regression of $|\hat{e}_t^\dagger|$ on a constant and $|\hat{e}_{t-1}^\dagger|,...,|\hat{e}_{t-4}^\dagger|$, t=9 to T. The j-period ahead autoregressive forecast of $|\hat{e}_{t+j}^\dagger|$ was used as described in part 3 of the Additional Appendix to obtain an estimate of $Ee_t^{\dagger 2}e_{t+j}^{\dagger 2}$. An estimate of the covariance between $e_t^{\dagger 2}e_{t+j}^{\dagger 2}$, call it $\hat{\lambda}_j$, was defined as $\hat{\lambda}_j = \max(0,[\text{estimate of } Ee_t^{\dagger 2}e_{t+j}^{\dagger 2} - \hat{\sigma}_{e\dagger}^4])$.

(d)Estimate (4.1b) with a regression of $\hat{u}_t$ on $\hat{e}_{t+2}$ and $\hat{e}_{t+1}$, t=5,...,T-2, obtaining $\hat{c}_1$ and $\hat{c}_2$.

(e)As illustrated in equation (2.5), $\hat{S}\equiv S(\hat{b})$ (101×101) involves estimates of $Ee_{t-j}^2u_t^2$ and $Ee_{t-j}^2u_tu_{t+1}$. Note that $Ee_{t-j}^2u_t^2 = \sigma_e^2\sigma_u^2 + c_2^2\text{cov}(e_{t+2}^2,e_{t-j}^2) + c_1^2\text{cov}(e_{t+1}^2,e_{t-j}^2)$, $Ee_{t-j}^2u_tu_{t+1} = \sigma_e^2\sigma_{u,1} + c_1c_2\text{cov}(e_{t+2}^2,e_{t-j}^2)$, and that these relationships hold even if, as in Table 2B, u_t = ($c_2e_{t+2}+c_1e_{t+1}+$ conditionally homoskedastic MA(1) term). We constructed $\hat{S}$ in a fashion robust to the possible presence of such a conditionally homoskedastic term, by setting $\hat{S}(j,j) = \hat{\sigma}^2_{e\dagger}\hat{\sigma}^2_u + \hat{c}^2_2\hat{\lambda}_j + \hat{c}^2_1\hat{\lambda}_{j-1}$, j=2,...,101, $\hat{S}(j,j+1) = \hat{S}(j+1,j) = \hat{\sigma}^2_{e\dagger}\hat{\sigma}_{u,1} + \hat{c}_1\hat{c}_2\hat{\lambda}_j$, j=2,...,100. We finished constructing $\hat{S}$ by setting $\hat{S}(1,1)=\hat{\sigma}^2_u+2\hat{\sigma}_{u,1}$, $\hat{S}(1,j)=0$ j>1.

(f)Construct $\hat{\Psi}\equiv\Psi(\hat{b})$ (101×2) by setting $\hat{\Psi}(1,1)=1$, $\hat{\Psi}(1,2)$=sample mean of z_t, $\hat{\Psi}(j,1)=0$ j>1, $\hat{\Psi}(j,2)=\hat{\sigma}^2_{e\dagger}$ times the (j-2) weight in moving average representation implied by the step (b) estimates of $\hat{\phi}_0,\hat{\phi}_1,...,\hat{\phi}_4$.

(g)Compute $\hat{G} \equiv (\hat{\mu},\hat{g}_0,...,\hat{g}_{100})' = \hat{S}^{-1}\hat{\Psi}$. Write $\hat{g}_j=(0,\hat{g}_{2j})'$ (the first element of each (2×1) $\hat{g}_j$ will be zero by construction).

(h)Set $\hat{Z}^*_t = \hat{\mu}+(0,\Sigma^{t-1}_{j=0}\hat{g}_{2j}\hat{e}^\dagger_{t-j})'$, t=1,...,T, with $\hat{e}^\dagger_{t-j} \equiv 0$ for t-j≤0.

6. Formal statement of asymptotic results

This appendix presents a formal statement of our asymptotic results. We derive our results under what we feel are reasonable conditions, without attempting to be as general as possible. We note some limitations and possible extensions in our discussion. Proofs are in an earlier version of this paper.

Notation: For a matrix A, $|A|$ denotes the largest element of A in absolute value; for $x \in \mathbf{R}^n$ and a differentiable function $f:\mathbf{R}^n \rightarrow \mathbf{R}^p$, $\frac{\partial f}{\partial x'}$ denotes the (p×n) matrix of first derivatives of f; $\|.\|_1$ denotes 1-norm: for an (n×1) vector x, $\|x\|_1 \equiv \Sigma_{j=1}^{n}|x_j|$, for a (p×n) matrix A, $\|A\|_1$ is the max column sum, $\|A\|_1 = \max_{1\le j\le n} \Sigma_{i=1}^{p}|a_{ij}|$; b_i is the i'th element of a (m×1) vector b. For precise statement of technical conditions and results, on occasion we shall need to add superscripts, subscripts and arguments to matrices that in the paper are left uncluttered. In particular, we add a "*" superscript and, sometimes, a "T" subscript to the following quantities defined in section 2 of the paper: $\Psi_T^* \equiv Ee(t)X_t'$; $S_T^* = \Sigma_{i=-q}^{q}[Ee(t\text{-}i)u_{t-i}u_te(t)']$.

Assumption 1: (a) $\{e_t', u_t, X_t'\}'$ is strictly stationary and ergodic, with finite fourth moments and absolutely summable autocovariances. (b) $E(e_t|e_{t-1}, e_{t-2}, \ldots) = 0$. (c) For some integer $q \ge 0$, $E(u_t|e_t, e_{t-1}, e_{t-2}, \ldots, u_{t-q-1}, u_{t-q-2}, \ldots) = 0$. (d) Ee_te_t' is positive definite.

Assumption 2: (a) The class of allowable instruments are ones of the form:

(AA6.1)
$$\underset{(k\times1)}{Z_t} = \underset{(k\times1)}{\mu} + \Sigma_{j=0}^{\infty}\underset{(k\times r)}{g_j}\underset{(r\times1)}{e_{t-j}},$$

$$\Sigma_{j=0}^{\infty}|g_j| < \infty,\ EZ_tX_t' \text{ of rank k},\ \lim_{T\rightarrow\infty} E(T^{-1/2}\Sigma_{t=1}^{T}Z_tu_t)(T^{-1/2}\Sigma_{t=1}^{T}Z_tu_t)' \text{ positive definite.}$$

(b) There is an allowable instrument $Z_t^* \equiv \mu^* + \Sigma_{j=0}^{\infty}g_j^*e_{t-j}$ that satisfies

(AA6.2)
$$EZ_tX_t' = \Sigma_{i=-q}^{q}E(Z_{t-i}u_{t-i}u_tZ_t^{*\prime})$$

for all allowable Z_t.

Remark 1: In 1(a), the stationarity assumption is for convenience and concreteness; analogous results could be obtained under mixing conditions. From Assumption 1(b), the (r×1) vector e_t is a martingale difference sequence; it is the innovation in an observable (r×1) series of basic instruments z_t. This and subsequent assumptions make no formal reference to z_t because no explicit reference is necessary.

Assumption 1(c) says that the regression disturbance follows an MA(q) process. Note that 1(c) does not say that the univariate Wold innovation in u_t is uncorrelated with e_t, an assumption that will be violated in many applications (Hayashi and Sims (1983)). Assumption 1(d) rules out degeneracies such as including a linear combination of two instruments as a third instrument.

Remark 2: In conjunction with Assumption 1, Assumption 2(a) guarantees that EZ_tX_t' is finite and does not depend on t and similarly for the limiting variance of $T^{-1/2}\Sigma_{t=1}^{T}Z_tu_t$. The additional assumption that EZ_tX_t' and the limiting variance are of full rank are essentially identification conditions.

Remark 3: Assumption 2(b) states that the space of allowable instruments includes an optimal instrument: equation (AA6.2) is Hansen's (1985) optimality condition.

Remark 4: It follows that the minimum possible asymptotic variance is

$$\text{(AA6.3)} \qquad V^* = (EZ_t^*X_t')^{-1}\Omega^*(EX_tZ_t^{*\prime})^{-1} = (EZ_t^*X_t')^{-1} = \Omega^{*-1};\ \Omega^* \equiv \Sigma_{i=-q}^{q}E(Z_{t-i}^*u_{t-i}u_tZ_t^{*\prime}).$$

The equality $EZ_t^*X_t' = \Omega^*$ comes from setting $Z_t=Z_t^*$ in (AA6.2). Of course, if Z_t^* is optimal, so is CZ_t^* for any k×k nonsingular nonstochastic matrix C.

Remark 5: Since allowable Z_t's are linear in a constant and lagged e_t's, (AA6.2) will hold for all allowable Z_t's if it holds when Z_t is replaced by arbitrary e_{t-j} ($j\geq 0$) or by a constant. After using $Z_t^*=\mu^*+\Sigma_{n=0}^{\infty}g_n^*e_{t-n}$, (AA6.2) is:

$$\text{(AA6.4a)} \qquad Z_t \text{ replaced by } e_{t-j}: \qquad \underset{(r\times k)}{Ee_{t-j}X_t'} = \Sigma_{i=-q}^{q}E[\underset{(r\times 1)}{e_{t-j-i}}\underset{(1\times 1)}{u_{t-i}u_t}(\underset{(k\times 1)}{\mu^*}+\Sigma_{n=0}^{\infty}\underset{(k\times r)(r\times 1)}{g_n^*e_{t-n}})'].$$

$$\text{(AA6.4b)} \qquad Z_t \text{ replaced by } 1: \qquad \underset{(1\times k)}{EX_t'} = \Sigma_{i=-q}^{q}E[\underset{(1\times 1)}{u_{t-i}u_t}(\underset{(k\times 1)}{\mu^*}+\Sigma_{n=0}^{\infty}\underset{(k\times r)(r\times 1)}{g_n^*e_{t-n}})'].$$

Additional notation: Stack (AA6.4b) and, for j=0,...,T-1, (AA6.4a) into a set of T+1 matrix equations. Let $G^*=(\mu^*,g_0^*,...,g_{T-1}^*)'$. The result may be written as

$$\text{(AA6.5)} \qquad \underset{(1+Tr)\times k}{\Psi_T^*} = \underset{(1+Tr)\times(1+Tr)}{S_T^*}\ \underset{(1+Tr)\times k}{G^*} + \underset{(1+Tr)\times k}{R_T^*},$$

$$R_T^* = (\Sigma_{i=-q}^{q}E[u_{t-i}u_t(\Sigma_{n=T}^{\infty}g_n^*e_{t-n})],\ \Sigma_{i=-q}^{q}E[(\Sigma_{n=T}^{\infty}g_n^*e_{t-n})e_{t-i}'u_{t-i}u_t],\ ...,\ \Sigma_{i=-q}^{q}E[(\Sigma_{n=T}^{\infty}g_n^*e_{t-n})e_{t-T+1-i}'u_{t-i}u_t])'.$$

Let

$$Z_{t,T} = \mu_T+\Sigma_{j=0}^{T-1}g_{j,T}^*e_{t-j},\ (\mu_T,g_{0,T}^*,...,g_{T-1,T}^*)' \equiv G_T^* \equiv S_T^{*-1}\Psi_T^*.$$

We model S_T^* and $\mathbf{\Psi}_T^*$ as smooth functions of a m×1 parameter vector "b", and then solve for the weights $\{g_{j,T}\}$:

$$\text{(AA6.6)}\qquad \underset{(1+Tr)\times k}{G_T(b)} = \underset{(1+Tr)\times(1+Tr)}{S_T(b)^{-1}}\ \underset{(1+Tr)\times k}{\mathbf{\Psi}_T(b)}, \qquad G_T(b) \equiv (\underset{(k\times 1)}{\mu_T}, \underset{(k\times r)}{g_{0,T}}, \ldots, \underset{(k\times r)}{g_{T-1,T}})'$$

In (AA6.6) and below, we drop the "*" to allow for the possibility of misspecification. The feasible estimator uses a sequence of estimates $\{\hat{b}\}$, assumed to converge to an (m×1) parameter vector $b^\dagger$. Write

$$\text{(AA6.7)}\qquad \hat{G}_T \equiv (\hat{\mu}_T, \hat{g}_{0,T}, \ldots, \hat{g}_{T-1,T})' = S_T(\hat{b})^{-1}\mathbf{\Psi}_T(\hat{b}),$$
$$\hat{Z}_{t,T} = \hat{\mu}_T + \Sigma_{j=0}^{t-1}\hat{g}_{j,T}\hat{e}_{t-j},\ \hat{\beta} = (\Sigma_{t=1}^{T}\hat{Z}_{t,T}X_t')^{-1}\Sigma_{t=1}^{T}\hat{Z}_{t,T}y_t,$$

where, again, the double subscripting serves as a reminder that the variables in general depend not only on $\hat{b}$ but also T. In (AA6.7), $\hat{e}_{t-j} \equiv e_{t-j}(\hat{b})$ is the residual after a parametric model is fitted to z_t.

<u>Assumption 3:</u> For some $b^\dagger \in \mathbf{R}^m$ and some $\delta > .25$, $T^\delta(\hat{b} - b^\dagger) = O_p(1)$.

<u>Assumption 4</u>: There is an open neighborhood $\mathbb{N}$ around $b^\dagger$ and a scalar positive constant c such that:

(a)$S_T(b)$ and $\mathbf{\Psi}_T(b)$ are twice continuously differentiable, and for i,n=1,...,m $\sup_{b\in\mathbb{N}, T>0} \|\frac{\partial}{\partial b_i} S_T(b)\|_1 < c$, $\sup_{b\in\mathbb{N}, T>0} \|\frac{\partial^2}{\partial b_i \partial b_n} S_T(b)\|_1 < c$, $\sup_{b\in\mathbb{N}, T>0} \|\mathbf{\Psi}_T(b)\|_1 < c$, $\sup_{b\in\mathbb{N}, T>0} \|\frac{\partial}{\partial b_i} \mathbf{\Psi}_T(b)\|_1 < c$, $\sup_{b\in\mathbb{N}, T>0} \|\frac{\partial^2}{\partial b_i \partial b_n} \mathbf{\Psi}_T(b)\|_1 < c$.

(b)$\det(S_T(b)) > 0$, $\sup_{b\in\mathbb{N}, T>0} \|S_T(b)\|_1 < c$, $\sup_{b\in\mathbb{N}, T>0} \|S_T(b)^{-1}\|_1 < c$.

(c)$e_t(b)$ is twice continuously differentiable in b, with the fourth moments of e_t and its first and second derivatives uniformly bounded: for i,n=1,...,m $\sup_{b\in\mathbb{N}, t>0} E|e_t(b)|^4 < c$, $\sup_{b\in\mathbb{N}, t>0} E|\frac{\partial}{\partial b_i} e_t(b)|^4 < c$, $\sup_{b\in\mathbb{N}, t>0} E|\frac{\partial^2}{\partial b_i \partial b_n} e_t(b)|^4 < c$.

(d)$e_t(b)$ is a measurable function of $\{e_{t-j} | j \geq 0\}$, as are $\frac{\partial}{\partial b_i} e_t(b) - E\frac{\partial}{\partial b_i} e_t(b)$ and $\frac{\partial^2}{\partial b_i \partial b_n} e_t(b) - E\frac{\partial^2}{\partial b_i \partial b_n} e_t(b)$ for i,n = 1,...,m.

<u>Remark 6</u>: Assumption 3 says that the sequence $\{\hat{b}\}$ converges to $b^\dagger$. The rate of convergence will be $\delta=½$ in most applications. Note that we have <u>not</u> assumed that $S_T^* = S_T(b^\dagger)$, $\mathbf{\Psi}_T^* = \mathbf{\Psi}_T(b^\dagger)$, or $e_t(b^\dagger) = e_t$, nor

even that one can recover e_t from an observable series z_t with a finite dimensional parameter vector: we allow misspecification of the process generating the data. Since moments of u_t figure into S_T, this will generally require a $\{\hat{u}_t\}$ series constructed using a T^{δ} consistent estimator of β.

Remark 7: Assumption 4 suffices to insure that $G_T(b) \equiv S_T(b)^{-1}\Psi_T(b)$ is smooth enough to allow us to apply a mean value argument around $b=b^{\dagger}$. It is satisfied by standard ARMA and GARCH models, apart from those with unit moving average roots.

Theorem 1: Suppose that for all T, $S_T(b^{\dagger})=S_T^*$, $\Psi_T(b^{\dagger})=\Psi_T^*$, $e_t(b^{\dagger})=e_t$. Then under assumptions 1-4, $\sqrt{T}(\hat{\beta}-\beta) \sim_A N(0, V^*)$, for V^* defined in (AA6.3); as well, $V^* = \lim_{T\to\infty} (\Psi_T^{*\prime} S_T^{*-1} \Psi_T^*)^{-1}$.

Remark 8: Theorem 1 states that if the specification is correct, the estimator obtains the efficiency bound. The theorem of course follows as well if $S_T(b^{\dagger})$ and $\Psi_T(b^{\dagger})$ differ from S_T^* and Ψ_T^* by a quantity that goes to zero at a sufficiently fast rate.

Remark 9: Note that nothing explicit has been said about R_T^* (defined in (AA6.5)); the assumptions made so far are sufficient to guarantee that weights derived from an estimate of $S_T^{*-1}\Psi_T^*$ are asymptotically equivalent to those from $S_T^{*-1}\Psi_T^*$-$S_T^{*-1}R_T^*$.

Remark 10: The natural estimator of V^* is $(\hat{\Psi}_T'\hat{S}_T^{-1}\hat{\Psi}_T)^{-1}$.

Assumption 5: $Z_{t,T}(b^{\dagger})$ may be written as $Z_{t,T}(b^{\dagger}) = \mathring{\mu}_T(b^{\dagger})+\Sigma_{j=0}^{\infty}\mathring{g}_{j,T}(b^{\dagger})e_{t-j}$ for $(k\times 1)$ $\mathring{\mu}_T$ and $(k\times r)$ $\{\mathring{g}_{j,T}\}$ that satisfy: (a)for all T, $\Sigma_{j=0}^{\infty}|\mathring{g}_{j,T}(b^{\dagger})|<c<\infty$ for some finite constant c; (b)there is an allowable instrument

$$Z_t^{\dagger}=\mu^{\dagger} +\Sigma_{j=0}^{\infty}g_j^{\dagger}e_{t-j} \text{ with } \lim_{T\to\infty} \mathring{\mu}_T(b^{\dagger}) = \mu^{\dagger} \text{ and } \lim_{T\to\infty} \Sigma_{j=0}^{T-1}|\mathring{g}_{j,T}(b^{\dagger})-g_j^{\dagger}| = 0.$$

Remark 11: Since we have already defined $Z_{t,T}(b^{\dagger}) = \mu_T(b^{\dagger})+\Sigma_{j=0}^{T-1}g_{j,T}(b^{\dagger})e_{t-j}(b^{\dagger})$ (note the upper bound here is T-1, in the Assumption 5 distributed lag on e_{t-j} is infinity), Assumption 5 essentially states that one can write $e_t(b^{\dagger})$ as a (possibly) infinite order distributed lag on e_t. This holds when the parametric model for z_t is an ARMA model.

Theorem 2: Under assumptions 1-5: (a)$\sqrt{T}(\hat{\beta}-\beta) \sim_A N(0, V)$, $V=(EZ_t^{\dagger}X_t')^{-1}\Omega(EX_tZ_t^{\dagger\prime})^{-1}$, $\Omega=\Sigma_{i=-q}^{q}EZ_{t-i}^{\dagger}u_{t-i}u_tZ_t^{\dagger\prime}$; (b)$V$-$V^*$ is positive semidefinite.

Remark 12: The natural estimator of $EZ_t^{\dagger}X_t'$ is $T^{-1}\Sigma_{t=1}^{T}\hat{Z}_{t,T}X_t'$. Ω may be estimated in a variety of familiar ways (e.g., Andrews (1991), Newey and West (1994), den Haan and Levin (1996)).

Remark 13: It may be shown that if the parametric specification is correct, as in Theorem 1, assumption 5 follows from previous assumptions, with $\mathring{g}_{j,T}=g^*_{j,T}$ ($g^*_{j,T}\equiv 0$ for $j\geq T$) and $g^{\dagger}_j=g^*_j$. So under correct specification, Theorems 1 and 2 make the same statement.

7. Some worked out examples

The examples in subsections (a) and (b) are intended in part to make concrete the form of the instrument, in part to illustrate that the optimal instrument typically puts nonzero weight on all lags. Subsection (a) derives the population form of the optimal instrument Z_t^* for a univariate AR(1) with a conditionally heteroskedastic disturbance. This will serve as well to illustrate that our approach has potential benefits even in the absence of serial correlation. Subsection (b) derives the population form of the optimal instrument Z_t^* in a conditionally homoskedastic special case of the DGP used in the simulations. This will serve as well to allow us to illustrate that in conditionally homoskedastic environments our estimator is asymptotically the same as that of West and Wilcox (1996). Subsection (c) shows how the DGP used in the simulations is an example of a forecasting application. This is intended to help motivate that DGP. In all three examples we suppress constant terms throughout.

(a)Univariate AR(1): Suppose we are interested in estimating β in an AR(1) model with an innovation that is a conditionally heteroskedastic martingale difference sequence:

(AA7.1) $\quad y_t = \beta y_{t-1}+u_t,\ |\beta|<1,\ Eu_t=0,\ Eu_t^2=1,$

$$E[u_t|u_{t-1}, u_{t-2}, ...]=0,\ \ E[u_t^2 u_{t-j} u_{t-i}]=0 \text{ for } i\neq j,$$

(AA7.2) $\quad E[u_t^2|u_{t-1}, u_{t-2}, ...]\neq Eu_t^2.$

Conditional heteroskedasticity is stated in abstract form in (AA7.2). We have here an uninteresting special case of (2.1)–uninteresting because weighted least squares, or maximum likelihood, would be computationally straightforward and the additional specification required relative to our estimator is in practice unlikely to be questionable. We use this example because it is particularly simple.

This example may be mapped into the notation of (2.1) by setting $X_t=z_t=y_{t-1}$, $e_t=u_{t-1}$. To facilitate reference to (AA7.1), we do not work in terms of z or e but instead stick to y and u. In this notation, then, the matrix S is diagonal with j'th diagonal element $Eu_{t-j}^2u_t^2$; the vector Ψ has j'th element $Eu_{t-j}y_{t-1}$ $(=\beta^{j-1})$:

$$S = \begin{pmatrix} Eu_{t-1}^2u_t^2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ & Eu_{t-2}^2u_t^2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & Eu_{t-3}^2u_t^2 & \cdots & 0 & 0 & 0 \\ \cdots & & & & & & \\ & 0 & 0 & \cdots & 0 & Eu_{t-T+1}^2u_t^2 & 0 \\ & 0 & 0 & \cdots & 0 & 0 & Eu_{t-T}^2u_t^2 \end{pmatrix},$$

$$\Psi = \begin{pmatrix} Eu_{t-1}y_{t-1} \\ Eu_{t-2}y_{t-1} \\ \ldots \\ Eu_{t-T}y_{t-1} \end{pmatrix}.$$

Then for given T one can solve explicitly for the instrument $\Psi' S^{-1}$. In the limit, as $T\rightarrow\infty$, we have

(AA7.3) $$Z_t^* = \sum_{j=1}^{\infty}[(Eu_{t-j}y_{t-1})u_{t-j}/Eu_{t-j}^2u_t^2] = \sum_{j=1}^{\infty}[\beta^{j-1}u_{t-j}/Eu_{t-j}^2u_t^2].$$

In u_t were conditionally homoskedastic (i.e, if we replace (AA7.2) with the condition $E[u_t^2 | u_{t-1}, u_{t-2}, \ldots]=Eu_t^2$), then $Eu_{t-j}^2u_t^2 = Eu_{t-j}^2Eu_t^2 = 1 \Rightarrow Z_t^*=\sum_{j=1}^{\infty}\beta^{j-1}u_{t-j}$. But $\sum_{j=1}^{\infty}\beta^{j-1}u_{t-j} = y_{t-1}$. So if u_t were conditionally homoskedastic, $Z_t^*=y_{t-1}$. That is, an unnecessarily tedious argument has been used to show that in the model (AA7.1), with the additional condition that u_t is conditionally homoskedastic, the best instrumental variables estimator is least squares. (To prevent confusion, we note that this argument also holds when we allow the variance of u_t to be something other than unity, say $Eu_t^2=\sigma^2$, upon noting that IV estimation with $Z_t^*=(1/\sigma^4)y_{t-1}$ is numerically identical to IV estimation with $Z_t^*=y_{t-1}$.)

But given conditional heteroskedasticity, least squares is no longer the optimal IV estimator. In constructing Z_t^*, one weights the innovations not only by β^{j-1}, but also (inversely) by $Eu_{t-j}^2u_t^2$. The specific model for conditional heteroskedasticity considered in Broze et al. (2001) is $u_t=\epsilon_t\epsilon_{t-1}$ for an i.i.d. (0,1) r.v. ϵ_t. In this case, $Eu_{t-1}^2u_t^2 = E\epsilon_t^2\epsilon_{t-1}^4\epsilon_{t-2}^2 = E\epsilon_t^2E\epsilon_{t-1}^4E\epsilon_{t-2}^2 = E\epsilon_{t-1}^4 > 1$, while $Eu_{t-j}^2u_t^2 = 1$ for $j>1$. Then

$$Z_t^*= u_{t-1}/E\epsilon_{t-1}^4 + \sum_{j=2}^{\infty}\beta^{j-1}u_{t-j}.$$

Alternatively, if u_t~ARCH(1), $E[u_t^2 | u_{t-1}, u_{t-2}, \ldots] =\omega+\alpha u_{t-1}^2$, $0<\alpha<1$, $\omega=1-\alpha$ (this conditions insures the simplification $Eu_t^2=1$), then

$$Z_t^*=\sum_{j=1}^{\infty}[\beta^{j-1}u_{t-j}/(1-\alpha^j+\alpha^jEu_t^4)].$$

Observe that all lags of u_t receive nonzero weight in Z_t^*, a result that also carries over if we rewrite Z_t^* as a distributed lag on past y_t's.

(b)Conditionally homoskedastic special case of the DGP used in the simulations.

Let consider the form of the optimal estimator when the model is:

(AA7.4) $y_t = z_t\beta + u_t$, $u_t = e_{t+2}-\theta e_{t+1}$, $|\theta<1$,

(AA7.5) $z_t=\phi z_{t-1}+e_t$, $|\phi|<1$,

(AA7.6) $e_t \sim$ i.i.d (0,1).

We first use the approach of the present paper to solve for the sequence $\{g_j^*\}$ such that $Z_t^*=\sum_{j=0}^{\infty} g_j^* e_{t-j}$. Then we use the approach of West and Wilcox (1996) to solve for the sequence $\{g_j^*\}$. Then we show that the two are identical.

The approach of the present paper: As in (2.5), S is tridiagonal. Under the assumptions of conditional homoskedasticity and unit variance for e_t, the diagonal elements are $Ee_{t-j}^2u_t^2 = Ee_{t-j}^2Eu_t^2 = Eu_t^2 = 1+\theta^2$, while the off-diagonal elements are $Ee_{t-j}^2u_{t+1}u_t = Ee_{t-j}^2Eu_{t+1}u_t = Eu_{t+1}u_t = -\theta$. As for Ψ, since $Ee_{t-j}z_t = \phi^j$, the jth element of Ψ is ϕ^{j-1}:

$$S = \begin{pmatrix} 1+\theta^2 & -\theta & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\theta & 1+\theta^2 & -\theta & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\theta & 1+\theta^2 & -\theta & \cdots & 0 & 0 & 0 \\ \cdots & & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & -\theta & 1+\theta^2 & -\theta \\ 0 & 0 & 0 & 0 & \cdots & 0 & -\theta & 1+\theta^2 \end{pmatrix}$$

$$\Psi = \begin{pmatrix} 1 \\ \phi \\ \cdots \\ \phi^{T-1} \end{pmatrix}$$

Since Z_t^* results in the limit as $T\rightarrow\infty$ of $\Psi' S^{-1}e(t)$, we have the following: the sequence $\{g_j^*\}$ satisfies the difference equation

(AA7.7) $-\theta g_{j+1}^* + (1+\theta^2)g_j^* -\theta g_{j-1}^* = \phi^j, j=0,1,2, \ldots$

We solve (AA7.7) subject to the initial condition $g_{-1}^*=0$ and the terminal condition $\lim_{j\rightarrow\infty} g_j^*=0$. The unique solution is

(AA7.8) $g_j^* = \theta g_{j-1}^* + \phi^j(1-\theta\phi)^{-1}$, $j=0,1,2,\ldots$; $g_{-1}^*\equiv 0$.

Thus, $g_0^*=(1-\theta\phi)^{-1}$, $g_1^*=(\theta+\phi)(1-\theta\phi)^{-1}$, etc.

The approach of West and Wilcox (1996): Here we apply equation (2.10) in West and Wilcox. To translate the notation of equation (2.10) in West and Wilcox (1996) into the notation of the present paper: $\theta_1\rightarrow\theta$, $\theta_2=0$, $P^*\equiv 1$, $F^*=\phi$, $R_t^*=z_t$. Then equation (2.10) in West and Wilcox gives

(AA7.9) $Z_t^*=\theta Z_{t-1}^* + (1-\theta\phi)^{-1}z_t$.

With some straightforward algebra it may be verified that Z_t^* defined in (AA7.9) satisfies $Z_t^*=\sum_{j=0}^{\infty}g_j^* e_{t-j}$ for $\{g_j^*\}$ defined in (AA7.8).

In a conditionally homoskedastic environment, the feasible procedure described here is asymptotically equivalent to the procedure described in West and Wilcox (1996) (provided, of course, that the approximating parametric models are the same asymptotically). In our view, in a conditionally homoskedastic environment, the West and Wilcox (1996) procedure will usually be computationally simpler.

Observe once again that the optimal instrument puts nonzero weight on all lags of e_t (from (AA7.7), $g_j^*\neq 0$ for all j) and on all lags of z_t (from (AA7.9), $Z_t^*=(1-\theta\phi)^{-1}\sum_{j=0}^{\infty}\theta^j z_{t-j}$).

<u>(c)The DGP in the simulations as an example of a forecasting application</u>

Suppose that there is a variable w_t that evolves according to

(AA7.10) $\quad w_t=\phi w_{t-1}+\epsilon_{1t} + \epsilon_{2t-1}$,

where ϵ_{1t} and ϵ_{2t} are martingale difference sequences that are independent at all leads and lags. Suppose further than in forming expectations of future w's, private agents see both ϵ_{1t} and ϵ_{2t}, while the economist sees only w_t and thus cannot distinguish between ϵ_{1t} and ϵ_{2t}. This is the simplest possible parameterization of the presumption that private agents have information available for forecasting that is not available to econometricians.

Then $E(w_{t+2}\,|\,\epsilon_{1t},\epsilon_{2t}, \epsilon_{1t-1}, \epsilon_{2t-2}, ...)$–i.e., the two period ahead private agent forecast of w_t–is readily seen to be $\phi^2 w_t+\phi\epsilon_{2t}$. In a typical forecasting application, we have a time series on the realization and on the forecast. Let

(AA7.11) $\quad y_t = w_{t+2}$, z_t=forecast=$\phi^2 w_t+\phi\epsilon_{2t}$.

Then we have

(AA7.12) $\quad y_t = z_t\beta_1 + u_t$,

$$z_t = \phi z_{t-1}+e_t,$$

where β_1=1 and

(AA7.13) $\quad e_t\equiv\phi^2\epsilon_{1t}+\phi\epsilon_{2t}$, $u_t = c_2 e_{t+2}+c_1 e_{t+1}+v_{t+2}$,

$$c_2 e_{t+2} \equiv E(\epsilon_{1t+2}\,|\,e_{t+2}),\ c_1=1/\phi,\ v_{t+2} = \epsilon_{1t+2}-E(\epsilon_{1t+2}\,|\,e_{t+2}).$$

8. Some details on maximum likelihood calculations

This section presents some algebraic details on the ML calculations.

Without loss of generality, we drop the constant term in (3.1a), relabel β_1 as β, and write (3.1a), (3.1b) as

(AA8.1) $e_t = y_{t-2} - \beta z_{t-2} + \theta e_{t-1}$,

$e_t | t\text{-}1 \sim \mathrm{N}(0,\ \sigma_t^2 \equiv \omega + \gamma_1 e_{t-1}^2 + \gamma_2\ \sigma_{t-1}^2)$.

We define the (5×1) parameter vector of interest as

(AA8.2) $\zeta \equiv (\beta, \theta, \omega, \gamma_1, \gamma_2)'$.

Apart from a constant, the log likelihood for one observation is ℓ_t =-0.5[log σ_t^2+(e_t^2/σ_t^2)]. The object of interest is the asymptotic variance of the ML estimator of β. This is the (1,1) element of the expectation of the inverse of the outer product of the score, i.e., the (1,1) element of the inverse of $E(\partial\ell_t/\partial\zeta)(\partial\ell_t/\partial\zeta)'$.

To obtain this quantity we proceed as follows. (1)For M=10,000 and T=990,000, we generate a time series of data of length $M+T$, setting the initial values of certain quantities to zero. (2)Use that time series to generate a time series of observations on the score, again setting the initial values of certain quantities to zero. (3)Discard the initial M observations. (4)Estimate $E(\partial\ell_t/\partial\zeta)(\partial\ell_t/\partial\zeta)'$ as the average of the remaining T (non-discarded) observations, setting to zero terms with expectation zero. (5)Report the (1,1) value of the inverse of the estimate of $E(\partial\ell_t/\partial\zeta)(\partial\ell_t/\partial\zeta)'$.

Specifically, in step (1) we initialize e_0=z_0=z_{-1}=σ_0=0 and then generate data recursively via

(AA8.3) $\sigma_t^2 = \omega + \gamma_1 e_{t-1}^2 + \gamma_2 \sigma_{t-1}^2$, $e_t = \sigma_t \eta_t$. $\eta_t \sim$ i.i.d.N(0,1), $z_t = \phi z_{t-1} + e_t$.

In step (2) we initialize $\partial e_0/\partial\beta = \partial e_0/\partial\theta = \partial\sigma_0^2/\partial\beta = \partial\sigma_0^2/\partial\theta = \partial\sigma_0^2/\partial\gamma_1 = \partial\sigma_0^2/\partial\gamma_2 = 0$ and then generate data recursively via

$$\begin{aligned}
\frac{\partial e_t}{\partial \beta} &= -z_{t-2} + \theta \frac{\partial e_{t-1}}{\partial \beta} \\
\frac{\partial e_t}{\partial \theta} &= e_{t-1} + \theta \frac{\partial e_{t-1}}{\partial \theta} \\
\frac{\partial \sigma_t^2}{\partial \beta} &= 2\gamma_1 e_{t-1} \frac{\partial e_{t-1}}{\partial \beta} + \gamma_2 \frac{\partial \sigma_{t-1}^2}{\partial \beta} \\
\frac{\partial \sigma_t^2}{\partial \theta} &= 2\gamma_1 e_{t-1} \frac{\partial e_{t-1}}{\partial \theta} + \gamma_2 \frac{\partial \sigma_{t-1}^2}{\partial \theta} \\
\frac{\partial \sigma_t^2}{\partial \omega} &= 1 + \gamma_2 \frac{\partial \sigma_{t-1}^2}{\partial \omega} \Rightarrow \frac{\partial \sigma_t^2}{\partial \omega} = \frac{1}{1-\gamma_2} \\
\frac{\partial \sigma_t^2}{\partial \gamma_1} &= e_{t-1}^2 + \gamma_2 \frac{\partial \sigma_{t-1}^2}{\partial \gamma_1} \\
\frac{\partial \sigma_t^2}{\partial \gamma_2} &= \sigma_{t-1}^2 + \gamma_2 \frac{\partial \sigma_{t-1}^2}{\partial \gamma_2}
\end{aligned}$$