Table 1: Stationary BootstrapTest for GL and CL Forecast Superiority (DGPs4- 6)

	DGP2	DGP4	DGP6	DGP2	DGP4	DGP6
S_n	GL forecast superiority			CL forecast superiority		
	n=100			n=100		
0.63	0.136	0.582	0.277	0.170	0.710	0.372
0.54	0.130	0.605	0.259	0.169	0.726	0.361
0.44	0.115	0.558	0.247	0.170	0.712	0.345
0.35	0.129	0.559	0.239	0.131	0.684	0.348
0.25	0.134	0.555	0.290	0.151	0.696	0.380
0.16	0.124	0.541	0.270	0.146	0.693	0.363
DM	0.165	0.845	0.506			
	n=500			n=500		
0.54	0.140	0.978	0.516	0.161	0.996	0.718
0.45	0.127	0.974	0.517	0.118	0.997	0.715
0.36	0.126	0.981	0.491	0.142	0.998	0.678
0.27	0.106	0.975	0.482	0.130	0.999	0.679
0.17	0.105	0.971	0.462	0.125	0.995	0.672
0.08	0.109	0.976	0.478	0.109	0.995	0.661
DM	0.146	1.000	0.906			
	n=1000			n=1000		
0.50	0.129	1.000	0.717	0.147	1.000	0.898
0.41	0.130	1.000	0.698	0.121	1.000	0.891
0.33	0.130	1.000	0.697	0.128	1.000	0.877
0.24	0.090	1.000	0.687	0.108	1.000	0.876
0.15	0.098	1.000	0.699	0.088	1.000	0.879
0.06	0.095	1.000	0.696	0.119	1.000	0.859
DM	0.165	1.000	0.986			

The case where $\rho = \lambda = 0.5$

Similar results to those discussed in the paper apply to other moderate values of ρ and λ. Please see the tables here for the simulation results with different values of ρ and λ ($\rho = \lambda = 0.5$, $\rho = 0.3$, $\lambda = 0.5$, and $\rho = 0.5$, $\lambda = 0.3$). The tests perform poorly when the forecast errors are highly dependent and strongly autocorrelated (e.g., $\rho = \lambda = 0.8$). In most cases, our tests have worse power performance than the DM test, but our tests have better size properties, especially when both ρ and λ are large.

Table 2: Stationary BootstrapTest for GL and CL Forecast Superiority (DGPs4- 6)

	DGP2	DGP4	DGP6	DGP2	DGP4	DGP6
S_n	GL forecast superiority			CL forecast superiority		
	n=100			n=100		
0.63	0.116	0.580	0.222	0.120	0.751	0.352
0.54	0.125	0.579	0.219	0.146	0.750	0.341
0.44	0.109	0.579	0.208	0.130	0.762	0.355
0.35	0.112	0.589	0.232	0.123	0.735	0.328
0.25	0.120	0.580	0.259	0.118	0.720	0.348
0.16	0.127	0.615	0.241	0.136	0.770	0.361
DM	0.127	0.887	0.494			
	n=500			n=500		
0.54	0.112	0.990	0.506	0.129	1.000	0.709
0.45	0.112	0.988	0.484	0.088	1.000	0.709
0.36	0.105	0.989	0.479	0.121	1.000	0.709
0.27	0.098	0.983	0.495	0.107	1.000	0.698
0.17	0.101	0.984	0.447	0.121	1.000	0.727
0.08	0.124	0.987	0.472	0.108	1.000	0.713
DM	0.130	1.000	0.927			
	n=1000			n=1000		
0.50	0.115	1.000	0.767	0.107	1.000	0.918
0.41	0.119	1.000	0.718	0.106	1.000	0.909
0.33	0.120	1.000	0.715	0.107	1.000	0.914
0.24	0.089	1.000	0.709	0.094	1.000	0.912
0.15	0.104	1.000	0.749	0.090	1.000	0.912
0.06	0.092	1.000	0.726	0.116	1.000	0.899
DM	0.122	1.000	0.995			

The case where $\rho = 0.5$, $\lambda = 0.3$

Table 3: Stationary BootstrapTest for GL and CL Forecast Superiority (DGPs4- 6)

	DGP2	DGP4	DGP6	DGP2	DGP4	DGP6
S_n	GL forecast superiority			CL forecast superiority		
	n=100			n=100		
0.63	0.146	0.580	0.222	0.120	0.751	0.352
0.54	0.125	0.579	0.219	0.146	0.750	0.341
0.44	0.109	0.579	0.208	0.130	0.762	0.355
0.35	0.112	0.589	0.232	0.123	0.735	0.328
0.25	0.120	0.580	0.259	0.118	0.720	0.348
0.16	0.127	0.615	0.241	0.136	0.770	0.361
DM	0.127	0.887	0.494			
	n=500			n=500		
0.54	0.112	0.990	0.506	0.129	1.000	0.709
0.45	0.112	0.988	0.484	0.088	1.000	0.709
0.36	0.105	0.989	0.479	0.121	1.000	0.709
0.27	0.098	0.983	0.495	0.107	1.000	0.698
0.17	0.101	0.984	0.447	0.121	1.000	0.727
0.08	0.124	0.987	0.472	0.108	1.000	0.713
DM	0.130	1.000	0.927			
	n=1000			n=1000		
0.50	0.115	1.000	0.767	0.107	1.000	0.918
0.41	0.119	1.000	0.718	0.106	1.000	0.909
0.33	0.120	1.000	0.715	0.107	1.000	0.914
0.24	0.089	1.000	0.709	0.094	1.000	0.912
0.15	0.104	1.000	0.749	0.090	1.000	0.912
0.06	0.092	1.000	0.726	0.116	1.000	0.899
DM	0.122	1.000	0.995			

The case where $\rho = 0.3$, $\lambda = 0.5$

Table 4: Stationary BootstrapTest for GL and CL Forecast Superiority (DGPs4- 6)

	DGP2	DGP4	DGP6	DGP2	DGP4	DGP6
S_n	GL forecast superiority			CL forecast superiority		
	n=100			n=100		
0.63	0.213	0.399	0.216	0.280	0.522	0.327
0.54	0.185	0.389	0.198	0.286	0.450	0.338
0.44	0.155	0.378	0.183	0.220	0.480	0.299
0.35	0.181	0.368	0.180	0.223	0.435	0.298
0.25	0.170	0.320	0.209	0.205	0.432	0.292
0.16	0.162	0.315	0.195	0.206	0.389	0.281
DM	0.283	0.566	0.404			
	n=500			n=500		
0.54	0.197	0.660	0.276	0.258	0.763	0.439
0.45	0.184	0.632	0.274	0.210	0.753	0.391
0.36	0.166	0.589	0.259	0.181	0.733	0.369
0.27	0.164	0.563	0.248	0.170	0.685	0.358
0.17	0.131	0.514	0.217	0.141	0.663	0.317
0.08	0.128	0.507	0.192	0.144	0.653	0.300
DM	0.276	0.860	0.546			
	n=1000			n=1000		
0.50	0.181	0.817	0.347	0.221	0.908	0.509
0.41	0.178	0.776	0.309	0.185	0.909	0.489
0.33	0.160	0.763	0.294	0.186	0.889	0.466
0.24	0.144	0.753	0.289	0.158	0.854	0.428
0.15	0.151	0.706	0.269	0.128	0.844	0.396
0.06	0.118	0.663	0.286	0.137	0.799	0.369
DM	0.280	0.953	0.650			

The case where $\rho = 0.8$, $\lambda = 0.8$

We also consider pairwise forecast comparison with parameter estimation errors. The DGPs (DGP PEE1 - DGP PEE 16) are taken from Corradi and Swanson (2007). Please see the table below.

In the setup, the benchmark model (denoted by *DGP PEE1 below*) is an AR(1). (The benchmark model is also called the "small" model.) For the DM test, the null hypothesis is that the smaller (size) model is MSFE-better than the "big" alternative model. Ten of the DGPs include (non)linear functions of x_{t-1} (see below for further discussion). We set $P = 0.5T$, and $T = 600$ as in Corradi and Swanson (2007).

Data Generating Processes with Parameter Estimation Errors

$u_{1,t} \sim iidN(0,1)$, $u_{2,t} \sim iidN(0,1)$

$x_t = 1 + 0.3x_{t-1} + u_{1,t}$

DGP PEE1: $y_t = 1 + 0.3y_{t-1} + u_{2,t}$

DGP PEE2: $y_t = 1 + 0.3y_{t-1} + 0.3u_{3,t-1} + u_{3,t}$

DGP PEE3: $y_t = 1 + 0.6y_{t-1} + u_{2,t}$

DGP PEE4: $y_t = 1 + 0.6y_{t-1} + 0.3u_{3,t-1} + u_{3,t}$

DGP PEE5 : $y_t = 1 + 0.3y_{t-1} + \exp(\tan^{-1}(x_{t-1}/2)) + u_{3,t}$

DGP PEE6 : $y_t = 1 + 0.3y_{t-1} + \exp(\tan^{-1}(x_{t-1}/2)) + 0.3u_{3,t-1} + u_{3,t}$

DGP PEE7 : $y_t = 1 + 0.3y_{t-1} + x_{t-1} + u_{3,t}$

DGP PEE8 : $y_t = 1 + 0.3y_{t-1} + x_{t-1} + 0.3u_{3,t-1} + u_{3,t}$

DGP PEE9 : $y_t = 1 + 0.3y_{t-1} + x_{t-1}1\{x_{t-1} > 1/(1-0.3)\} + u_{3,t}$

DGP PEE10 : $y_t = 1 + 0.3y_{t-1} + x_{t-1}1\{x_{t-1} > 1/(1-0.3)\} + 0.3u_{3,t-1} + u_{3,t}$

DGP PEE11 : $y_t = 1 + 0.6y_{t-1} + \exp(\tan^{-1}(x_{t-1}/2)) + u_{3,t}$

DGP PEE12 : $y_t = 1 + 0.6y_{t-1} + \exp(\tan^{-1}(x_{t-1}/2)) + 0.3u_{3,t-1} + u_{3,t}$

DGP PEE13 : $y_t = 1 + 0.6y_{t-1} + x_{t-1} + u_{3,t}$

DGP PEE14: $y_t = 1 + 0.6y_{t-1} + x_{t-1} + 0.3u_{3,t-1} + u_{3,t}$

DGP PEE15: $y_t = 1 + 0.6y_{t-1} + x_{t-1}1\{x_{t-1} > 1/(1-0.3)\} + u_{3,t}$

DGP PEE16: $y_t = 1 + 0.6y_{t-1} + x_{t-1}1\{x_{t-1} > 1/(1-0.3)\} + 0.3u_{3,t-1} + u_{3,t}$.

DGPs PEE1 - PEE4 satisfy the null hypothesis whereas the other DGPs satisfy the alternative hypothesis. Note that the benchmark or "small" model in our test statistic calculations is always estimated as $y_t = \alpha + \beta y_{t-1} + \epsilon_t$; and the "big" model is the same, but with x_{t-1} added as an additional regressor. Thus, only in DGP PEE7 and DGP PEE13 is the alternative model "correct". In all other Powers, the model estimated only includes x_{t-1} and so the alternative models in these cases are misspecified. Thus the power in these cases might be low. From the table below, our tests in general have comparable power performance relative to the DM test; and DM test has more conservative sizes compared to our tests in all DGPs.

Table 5: Stationary BootstrapTest for GL and CL Forecast Superiority (DGPs1- 6)

	DGP PEE1	DGP PEE2	DGP PEE3	DGP PEE4	DGP PEE5	DGP PEE6	DGP PEE7	DGP PEE8
S_n				GL forecast superiority				
0.53	0.045	0.042	0.051	0.057	0.777	0.675	0.999	0.999
0.44	0.045	0.055	0.052	0.043	0.795	0.679	1.000	0.999
0.35	0.056	0.047	0.044	0.047	0.760	0.689	1.000	0.999
0.26	0.052	0.052	0.049	0.048	0.810	0.683	1.000	1.000
0.17	0.053	0.050	0.055	0.059	0.766	0.721	1.000	1.000
0.08	0.063	0.052	0.064	0.056	0.813	0.704	1.000	0.999
				CL forecast superiority				
0.53	0.057	0.055	0.065	0.048	0.990	0.976	1.000	1.000
0.44	0.059	0.053	0.047	0.063	0.992	0.973	1.000	1.000
0.35	0.055	0.050	0.065	0.061	0.992	0.973	1.000	1.000
0.26	0.059	0.054	0.055	0.067	0.991	0.969	1.000	1.000
0.17	0.065	0.067	0.062	0.087	0.981	0.978	1.000	1.000
0.08	0.059	0.082	0.061	0.075	0.990	0.984	1.000	1.000
DM	0.011	0.003	0.008	0.019	0.999	0.997	1.000	1.000

DGPs PEE1 - PEE4 satisfy the null hypothesis whereas the other DGPs satisfy the alternative hypothesis. n=600. Entry numbers are the rejection frequency in 1000 repetitions. The number of bootstrap resamples is 300. S_n is the bootstrap smoothing parameter. The nominal test size is 10%.

Table 6: Stationary BootstrapTest for GL and CL Forecast Superiority (DGPs1- 6)

	DGPPEE9	DGPPEE10	DGPPEE11	DGPPEE12	DGPPEE13	DGPPEE14	DGPPEE15	DGPPEE16
S_n				GL forecast superiority				
0.53	1.000	1.000	0.776	0.711	1.000	1.000	0.839	0.799
0.44	1.000	1.000	0.780	0.672	1.000	1.000	0.827	0.762
0.35	1.000	1.000	0.775	0.692	1.000	0.999	0.837	0.761
0.26	1.000	1.000	0.778	0.706	1.000	1.000	0.848	0.759
0.17	1.000	1.000	0.793	0.719	1.000	1.000	0.838	0.788
0.08	1.000	1.000	0.813	0.736	1.000	1.000	0.859	0.778
				CL forecast superiority				
0.53	1.000	1.000	0.990	0.978	1.000	1.000	0.989	0.979
0.44	1.000	1.000	0.986	0.974	1.000	1.000	0.981	0.985
0.35	1.000	1.000	0.995	0.979	1.000	1.000	0.992	0.968
0.26	1.000	1.000	0.991	0.983	1.000	1.000	0.985	0.978
0.17	1.000	1.000	0.994	0.985	1.000	1.000	0.988	0.981
0.08	1.000	1.000	0.991	0.987	1.000	1.000	0.986	0.977
DM	1.000	1.000	0.998	0.997	1.000	1.000	0.988	0.984

DGPs PEE9 - PEE16 satisfy the alternative hypothesis. n=600.

Entry numbers are the rejection frequency in 1000 repetitions. The number of bootstrap resamples is 300.

S_n is the bootstrap smoothing parameter. The nominal test size is 10%.