5th International Workshop on Combinatorial Testing (IWCT 2016)
10 April 2016, Chicago, USA

Should We Care about "Don't Care" Testing Inputs? Empirical Investigation of Pair-Wise Testing

Sergiy Vilkomir and Galen Pennell

East Carolina University

Department of Computer Science

Tolerance (stability) of testing criteria

- This research is part of a more general problem about the tolerance (stability) of the testing coverage criteria.
- During testing according to some testing criterion (branch coverage, MC/DC, pairwise, etc.), only one such test case is usually required
- At the same time, many such test sets are possible for the same software
- Can we select any test set?
- How close are the characteristics of different test sets which satisfy to the same criterion?

Tolerance (stability) of testing criteria

- The main characteristic is effectiveness - the ability to detect failures in a software program
- Tolerance (stability) - the ability of every test set satisfying the criterion to provide a similar level of effectiveness
- For criteria with high tolerance, effectiveness of separate test sets does not vary much and is sufficiently close to the average effectiveness.
- For criteria with low tolerance effectiveness of separate test sets can vary significantly.

Tolerance (stability) of testing criteria

Figure 2: Distributions of effectiveness for criteria CC, DC, D/CC

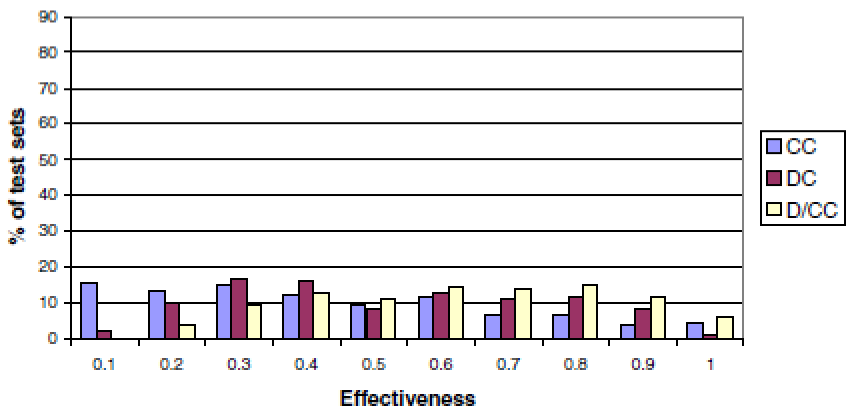
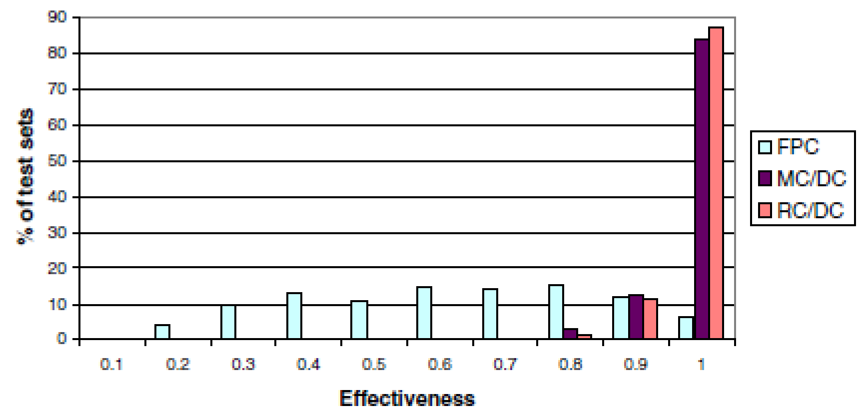


Figure 3: Distributions of effectiveness for criteria FPC, MC/DC, RC/DC



From: S. A. Vilkomir, K. Kapoor and J. P. Bowen. Tolerance of Control-Flow Testing Criteria, Proceedings of 27th IEEE Annual International Computer Software and Applications Conference (COMPSAC 2003), Dallas, Texas, USA, 3-6 November 2003, pp. 182-187

This research

How close are the characteristics of different pairwise test sets with the same inputs?

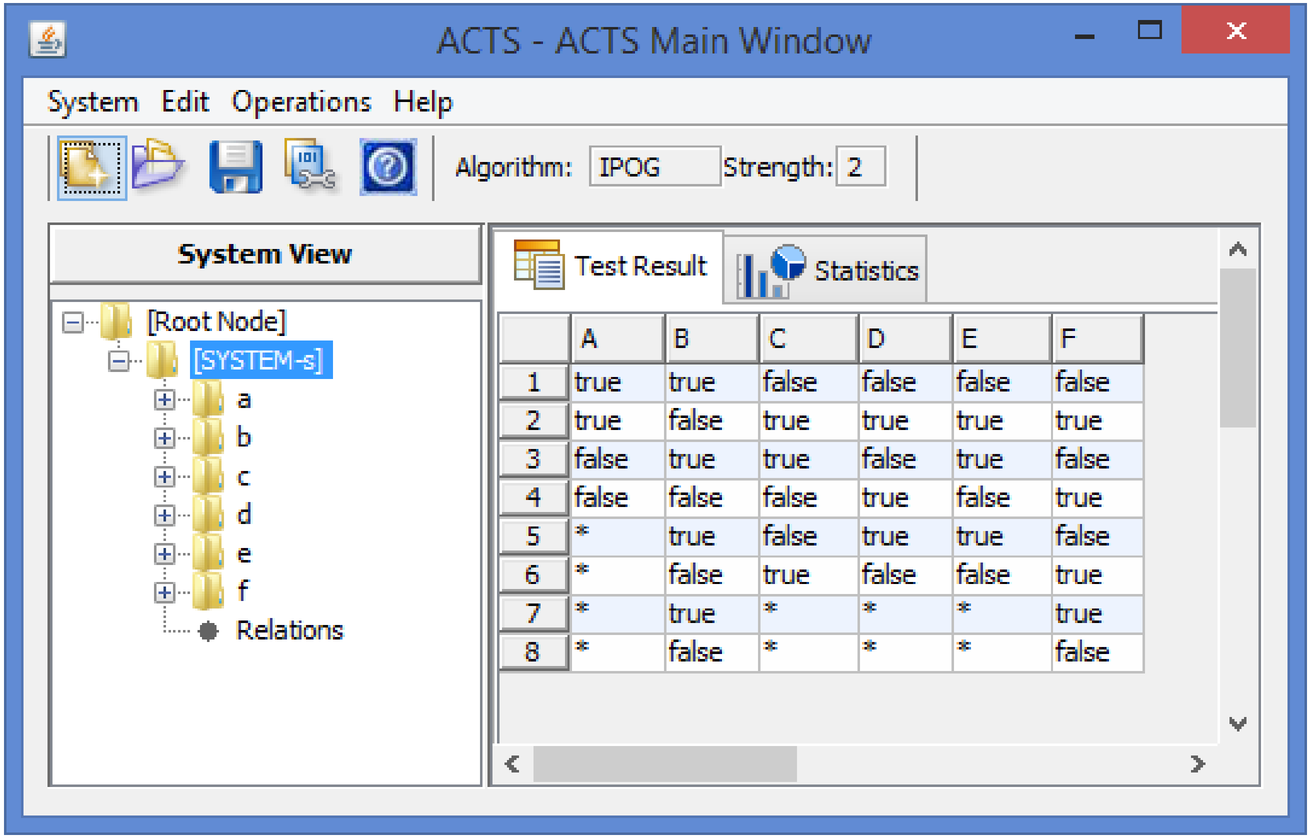
We consider:

- Only pairwise testing
- Only logical Inputs: logical variables with “true” and “false” values

It is not necessary that the results would be the same for general (not logical) inputs

- Two characteristics:
  - effectiveness
  - MC/DC level
- Pairwise sets with different “don’t care” values
- We started step 2: Different pairwise tests from different tools (ACTS, Pict, Jenny, Pairwiser, VPTag)

“don’t care” values



	A	B	C	D	E	F
1	true	true	false	false	false	false
2	true	false	true	true	true	true
3	false	true	true	false	true	false
4	false	false	false	true	false	true
5	*	true	false	true	true	false
6	*	false	true	false	false	true
7	*	true	*	*	*	true
8	*	false	*	*	*	false

20% values - “don’t care”

1024 different test sets

“don’t care” values

Expr. Size	IPOG			IPOG-F			IPOG-F2		
	Number of Test Inputs	Don't Care Inputs		Number of Test Inputs	Don't Care Inputs		Number of Test Inputs	Don't Care Inputs	
		Number	%		Number	%		Number	%
4	24	4	16.7	24	4	16.7	24	2	8.3
5	30	2	6.7	30	2	6.7	60	2	3.3
6	48	10	20.8	42	5	11.9	42	4	9.5
7	56	8	14.3	49	3	6.1	49	3	6.1
8	64	6	9.4	64	8	12.5	64	4	6.3

Empirical investigation

Empirical investigation included the following steps:

1. Selection of logical expressions for experimental testing. Together, 10 sets with 102 logical expressions were used.

2. Generation of different pair-wise test sets with 4 to 8 logical inputs. Test case generation was repeated using 3 different algorithms in ACTS: IPOG, IPOG-F, and IPOG-F2. Together, 122 pair-wise test sets were generated.

3. Evaluation of the effectiveness of pair-wise test sets. The Fault Evaluator research tool was used.

Empirical investigation

Empirical investigation included the following steps:

4. Evaluation of MC/DC levels of pair-wise test sets. The CodeCover tool was used.

5. Analysis of variations of the effectiveness and MC/DC levels. We compared these variations for different test sets separately for the same sets of expressions. The standard deviation and the coefficient of variation (the ratio of standard deviation to mean) were used.

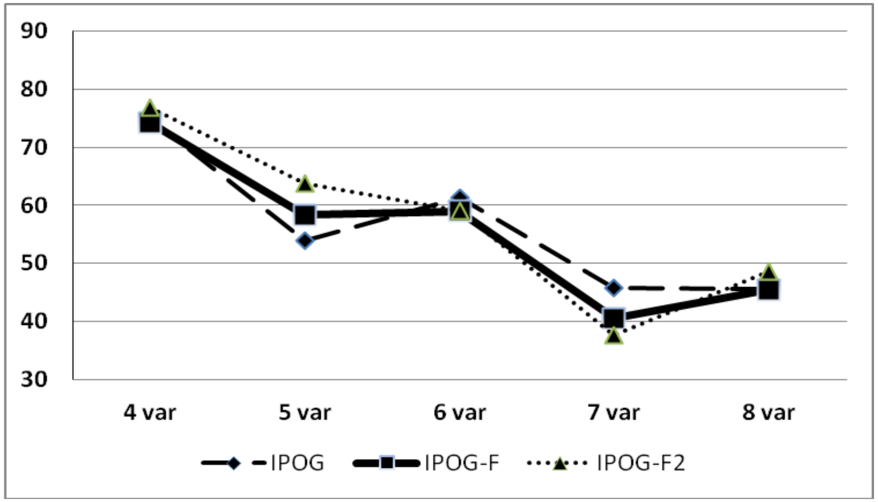
Empirical investigation

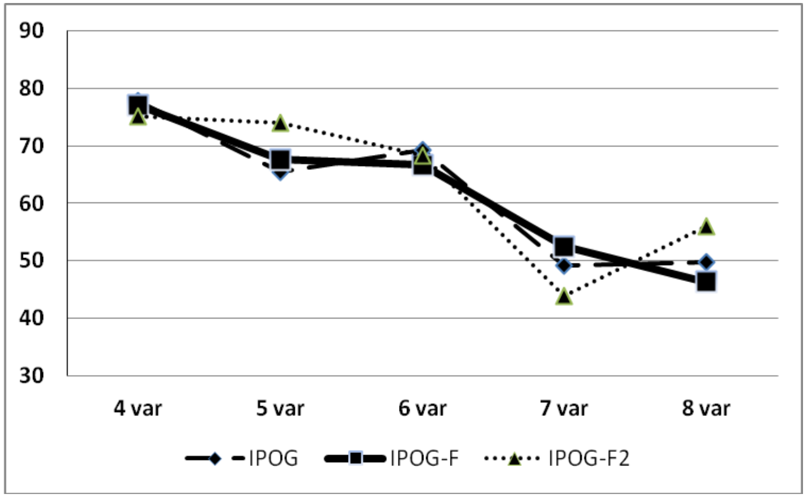
SCOPE OF THE EMPIRICAL TESTING

Expr. Size	Number of Expr.	IPOG			IPOG-F			IPOG-F2			Total Runs
		Test Sets	Test Cases	Runs	Test Sets	Test Cases	Runs	Test Sets	Test Cases	Runs	
4	22	10	60	1320	10	60	1320	4	24	528	3168
5	20	4	24	480	4	24	480	4	24	480	1440
6	20	10	80	1600	10	70	1400	10	70	1400	4400
7	20	10	80	1600	8	56	1120	8	56	1120	3840
8	20	10	80	1600	10	80	1600	10	80	1600	4800
Total	102	44	324	6600	42	290	5920	36	254	5128	**17648**

Empirical investigation - Effectiveness

Effectiveness of pair-wise testing





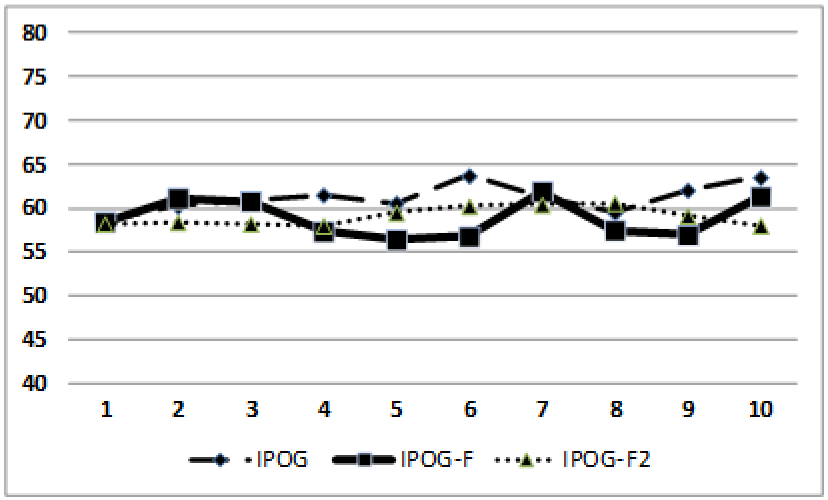
Simple expressions

Complex expressions

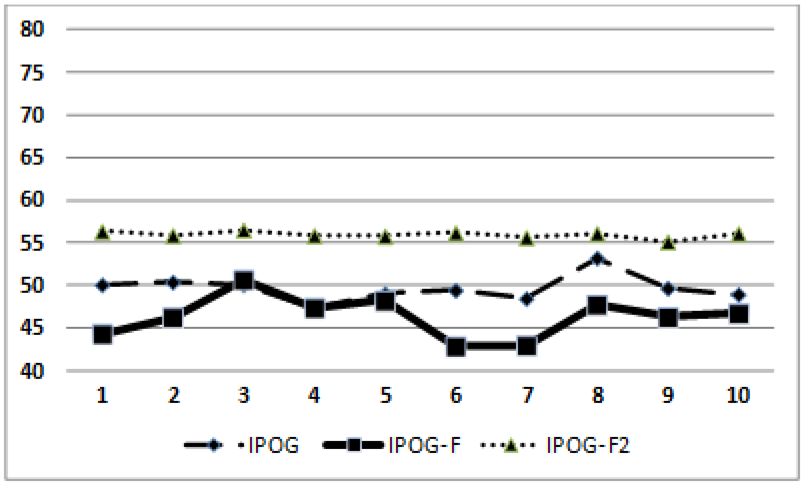
but this is not about tolerance

Empirical investigation - Effectiveness

Effectiveness of pair-wise testing



Simple expressions size 6



Complex expressions size 8

this is tolerance

Empirical investigation - Effectiveness

Effectiveness of pair-wise testing (IPOG Algorithm)

Expression		Pair-wise test sets - IPOG algorithm										Effectiveness		Range			Coeff. of
Size		1	2	3	4	5	6	7	8	9	10	Avg.	St. Dev.	Max	Min	Size	var., %
4	Simple	72.07	73.27	77.48	75.38	74.17	75.68	79.88	76.88	71.17	72.67	74.87	2.71	79.88	71.17	8.71	3.6
	Complex	79.2	80.09	80.97	82.08	74.12	75.22	75.88	76.55	75.88	76.77	77.68	2.70	82.08	74.12	7.96	3.5
5	Simple	52.13	56.39	51.63	55.64	N/A	N/A	N/A	N/A	N/A	N/A	53.95	2.42	56.39	51.63	4.76	4.5
	Complex	63.78	68.98	62.05	66.72	N/A	N/A	N/A	N/A	N/A	N/A	65.38	3.08	68.98	62.05	6.93	4.7
6	Simple	59.49	60.22	60.95	61.5	60.58	63.69	61.31	59.49	62.04	63.5	61.28	1.47	63.69	59.49	4.2	2.4
	Complex	66.02	69.05	68.23	69.88	67.95	70.7	69.46	70.29	70.01	70.56	69.22	1.46	70.7	66.02	4.68	2.1
7	Simple	46.67	44.4	47.38	44.54	44.26	44.26	46.81	48.37	47.09	44.68	45.85	1.57	48.37	44.26	4.11	3.4
	Complex	51.55	50.74	49.26	48.05	48.32	48.32	48.18	48.18	53.84	46.16	49.26	2.20	53.84	46.16	7.68	4.5
8	Simple	44.59	46.4	44.59	46.28	46.85	44.71	45.95	48.31	42.68	45.27	45.56	1.55	48.31	42.68	5.63	3.4
	Complex	50.08	50.4	50.08	47.27	49.04	49.44	48.47	53.22	49.68	48.95	49.66	1.55	53.22	47.27	5.95	3.1

Empirical investigation - Effectiveness

Effectiveness of pair-wise testing (IPOG-F Algorithm)

Expression Size		Pair-wise test sets - IPOG-F algorithm 1	2	3	4	5	6	7	8	9	10	Effectiveness Avg.	St. Dev.	Range Max	Min	Size	Coeff. of var., %
4	Simple	71.47	74.17	73.27	76.28	76.88	75.38	79.28	74.47	68.47	72.37	74.2	3.04	79.28	68.47	10.8	4.1
	Complex	78.98	79.2	73.67	74.34	80.75	80.31	76.55	76.77	73.89	76.33	77.08	2.63	80.75	73.67	7.08	3.4
5	Simple	57.39	59.15	57.64	59.15	N/A	N/A	N/A	N/A	N/A	N/A	58.33	0.95	59.15	57.39	1.76	1.6
	Complex	68.11	70.88	64.82	66.72	N/A	N/A	N/A	N/A	N/A	N/A	67.63	2.55	70.88	64.82	6.06	3.8
6	Simple	59.49	61.13	60.77	57.3	56.39	56.75	61.86	57.48	56.93	61.31	58.94	2.18	61.86	56.39	5.47	3.7
	Complex	63.96	66.16	64.1	70.01	65.47	69.88	65.2	65.2	66.71	69.74	66.64	2.38	70.01	63.96	6.05	3.6
7	Simple	42.27	40.14	42.27	39.57	41.42	39.01	40.99	38.01	N/A	N/A	40.46	1.55	42.27	38.01	4.26	3.8
	Complex	51.14	54.64	51.41	54.64	50.2	53.7	50.2	53.57	N/A	N/A	52.44	1.90	54.64	50.2	4.44	3.6
8	Simple	43.24	46.62	46.06	46.51	45.27	43.47	43.02	45.72	47.75	46.73	45.44	1.65	47.75	43.02	4.73	3.6
	Complex	44.29	46.22	50.64	47.35	48.23	42.85	42.93	47.75	46.3	46.7	46.33	2.43	50.64	42.85	7.79	5.2

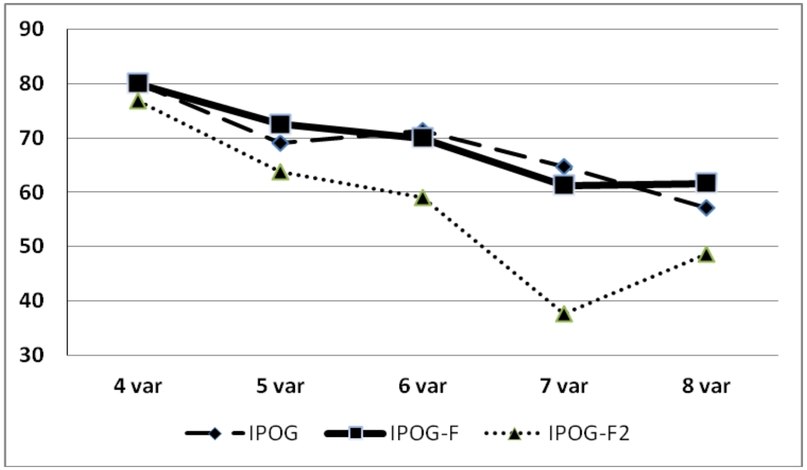
Empirical investigation - Effectiveness

Effectiveness of pair-wise testing (IPOG-F2 Algorithm)

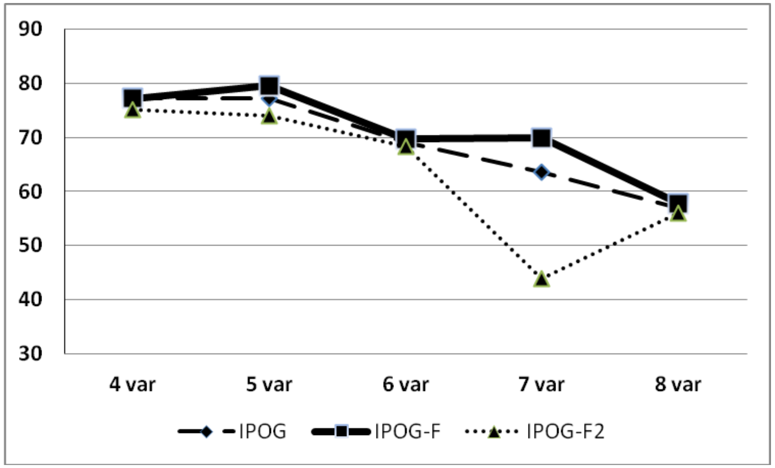
Size	Expression	Pair-wise test sets - IPOG-F2 algorithm										Effectiveness		Range			Coeff. of var., %
		1	2	3	4	5	6	7	8	9	10	Avg.	St. Dev.	Max	Min	Size	
4	Simple	76.58	75.68	76.58	78.08	N/A	N/A	N/A	N/A	N/A	N/A	76.73	0.99	78.08	75.68	2.4	1.3
	Complex	76.77	72.12	73.01	78.32	N/A	N/A	N/A	N/A	N/A	N/A	75.06	2.97	78.32	72.12	6.2	4
5	Simple	65.66	63.91	61.4	64.16	N/A	N/A	N/A	N/A	N/A	N/A	63.78	1.77	65.66	61.4	4.26	2.8
	Complex	76.26	74.18	71.4	73.83	N/A	N/A	N/A	N/A	N/A	N/A	73.92	1.99	76.26	71.4	4.86	2.7
6	Simple	58.58	58.39	58.21	58.03	59.49	60.22	60.4	60.58	59.12	58.03	59.11	1.01	60.58	58.03	2.55	1.7
	Complex	67.95	68.36	66.57	66.85	69.74	70.84	68.64	69.74	69.74	65.34	68.38	1.72	70.84	65.34	5.5	2.5
7	Simple	36.45	37.87	38.16	38.01	38.3	40	35.89	36.88	N/A	N/A	37.7	1.28	40	35.89	4.11	3.4
	Complex	43.61	44.01	47.11	44.15	42.8	44.55	43.88	40.51	N/A	N/A	43.83	1.83	47.11	40.51	6.6	4.2
8	Simple	47.97	48.31	46.96	47.07	49.66	48.65	50	49.66	48.54	48.42	48.52	1.04	50	46.96	3.04	2.1
	Complex	56.35	55.87	56.51	55.87	55.79	56.19	55.63	56.11	55.14	56.11	55.96	0.39	56.51	55.14	1.37	0.7

Empirical investigation - MC/DC level

MC/DC level of pair-wise testing



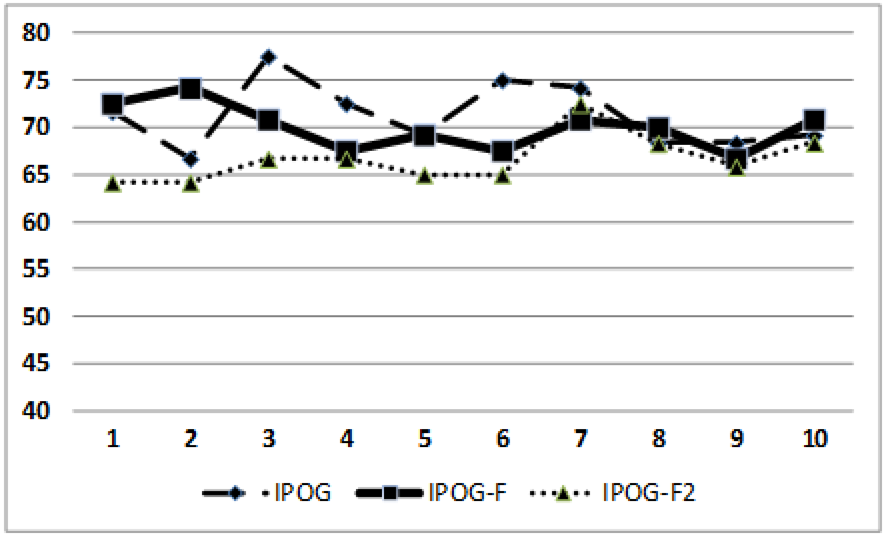
Simple expressions



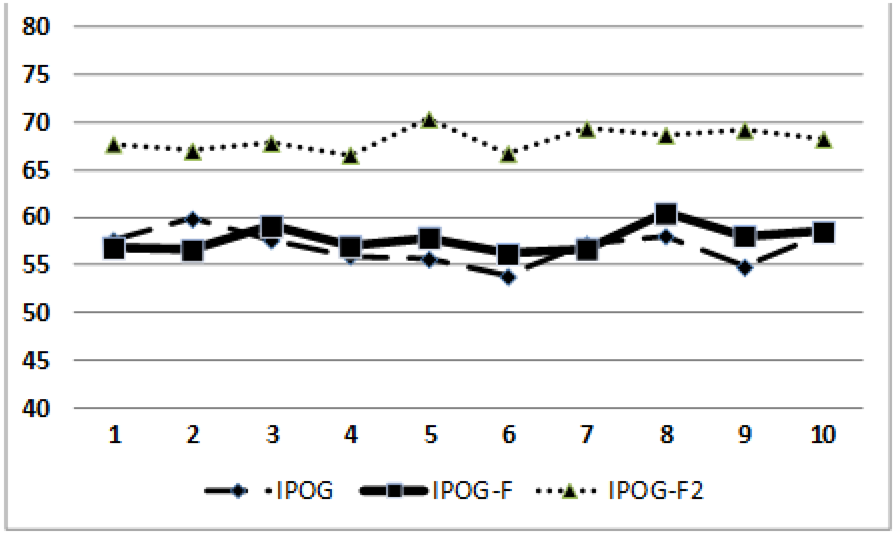
Complex expressions

Empirical investigation - MC/DC level

MC/DC level of pair-wise testing



Simple expressions size 6



Complex expressions size 8

this is tolerance

Empirical investigation - MC/DC level

MC/DC level of pair-wise testing (IPOG Algorithm)

Expression		Pair-wise test sets - IPOG algorithm										MC/DC Cov.		Range			Coeff. of
Size		1	2	3	4	5	6	7	8	9	10	Avg.	St. Dev.	Max	Min	Size	var., %
4	Simple	81.25	77.08	83.33	77.08	83.33	79.17	85.42	79.17	79.17	75	80	3.29	85.42	75	10.4	4.1
	Complex	79.11	76.94	76.26	78.92	78.94	76.77	76.09	78.76	76.27	74.1	77.22	1.66	79.11	74.1	5.01	2.2
5	Simple	66	68	70	72	N/A	N/A	N/A	N/A	N/A	N/A	69	2.58	72	66	6	3.7
	Complex	75.36	78.56	75.72	78.9	N/A	N/A	N/A	N/A	N/A	N/A	77.14	1.85	78.9	75.36	3.54	2.4
6	Simple	71.67	66.67	77.5	72.5	69.16	75	74.16	68.33	68.33	69.16	71.25	3.50	77.5	66.67	10.8	4.9
	Complex	69.37	69.37	72.5	71.39	67.01	71.09	73	67.51	62.02	67.71	69.1	3.25	73	62.02	11	4.7
7	Simple	60.28	59.56	63.86	62.84	68.01	68.01	71.57	66.85	60.98	65.86	64.78	3.93	71.57	59.56	12	6.1
	Complex	64.59	62.21	64.09	61.72	63.53	63.53	63.32	66.17	64.88	61.72	63.58	1.44	66.17	61.72	4.45	2.3
8	Simple	54.39	56.26	54.39	60	57.49	55	61.88	58.13	53.77	60	57.13	2.82	61.88	53.77	8.11	4.9
	Complex	57.62	59.84	57.62	55.94	55.62	53.83	57.2	58.04	54.8	58.23	56.87	1.80	59.84	53.83	6.01	3.2

Empirical investigation - MC/DC level

MC/DC level of pair-wise testing (IPOG-F Algorithm)

Expression		Pair-wise test sets - IPOG-F algorithm										MC/DC Cov.		Range			Coeff. of
Size		1	2	3	4	5	6	7	8	9	10	Avg.	St. Dev.	Max	Min	Size	var., %
4	Simple	76.04	75	82.29	80.21	79.17	77.08	85.42	83.33	80.21	81.25	80	3.29	85.42	75	10.4	4.1
	Complex	78.22	77.76	73.72	73.26	79.27	79.64	75.6	78.77	76.89	78.67	77.18	2.27	79.64	73.26	6.38	2.9
5	Simple	72	71	74	73	N/A	N/A	N/A	N/A	N/A	N/A	72.5	1.29	74	71	3	1.8
	Complex	79.65	80.96	78.13	79.45	N/A	N/A	N/A	N/A	N/A	N/A	79.55	1.16	80.96	78.13	2.83	1.5
6	Simple	72.49	74.16	70.83	67.5	69.16	67.5	70.83	69.99	66.66	70.83	70	2.36	74.16	66.66	7.5	3.4
	Complex	66.67	68.33	66.46	74.1	69.5	76.31	67.07	69.5	69.5	69.11	69.66	3.20	76.31	66.46	9.85	4.6
7	Simple	63.72	58.09	64.44	58.71	63.72	58	64.44	58.72	N/A	N/A	61.23	3.07	64.44	58	6.44	5
	Complex	68.93	69.82	71.47	71.91	67.75	68.64	69.8	70.24	N/A	N/A	69.82	1.40	71.91	67.75	4.16	2
8	Simple	58.76	63.13	63.76	62.51	61.39	58.76	58.76	63.13	61.89	64.39	61.65	2.17	64.39	58.76	5.63	3.5
	Complex	56.81	56.53	59.11	56.94	57.86	56.15	56.65	60.42	58.02	58.47	57.7	1.35	60.42	56.15	4.27	2.3

Empirical investigation - MC/DC level

MC/DC level of pair-wise testing (IPOG-F2 Algorithm)

Expression		Pair-wise test sets - IPOG-F2 algorithm										MC/DC Cov.		Range			Coeff. of
Size		1	2	3	4	5	6	7	8	9	10	Avg.	St. Dev.	Max	Min	Size	var., %
4	Simple	79.17	79.17	80.21	80.21	N/A	N/A	N/A	N/A	N/A	N/A	79.69	0.60	80.21	79.17	1.04	0.8
	Complex	74.18	74.35	78.02	77.85	N/A	N/A	N/A	N/A	N/A	N/A	76.1	2.12	78.02	74.18	3.84	2.8
5	Simple	66	68	70	72	N/A	N/A	N/A	N/A	N/A	N/A	69	2.58	72	66	6	3.7
	Complex	75.46	80.66	75.72	78.9	N/A	N/A	N/A	N/A	N/A	N/A	77.69	2.53	80.66	75.46	5.2	3.3
6	Simple	64.16	64.16	66.65	66.65	64.99	64.99	72.48	68.31	65.81	68.33	66.65	2.54	72.48	64.16	8.32	3.8
	Complex	65.42	65.98	69.31	69.92	67.79	67.8	71.67	71.67	70.57	61.61	68.17	3.17	71.67	61.61	10.1	4.6
7	Simple	55.87	57.3	55.87	56.58	57.3	60.15	55.87	58.72	N/A	N/A	57.21	1.55	60.15	55.87	4.28	2.7
	Complex	63.41	63.44	66.91	64.04	66.96	64.47	62.55	61.13	N/A	N/A	64.11	2.01	66.96	61.13	5.83	3.1
8	Simple	61.25	63.13	61.89	64.38	60.64	60	59.39	62.5	63.76	61.88	61.88	1.61	64.38	59.39	4.99	2.6
	Complex	67.7	67	67.83	66.58	70.33	66.72	69.35	68.66	69.21	68.24	68.16	1.23	70.33	66.58	3.75	1.8

Conclusions

- Should we care about “don’t care” testing inputs?
- Based on our results, our answer is “**No**.”
- The pair-wise test sets with different “don’t care” values are stable and both the effectiveness and the level of MC/DC coverage are very close within such test sets.
- Our results support current practice when one test set with randomly selected “don’t care” values is used for testing.

Conclusions

Limitations

• we limit the results to logical expressions only.

• we considered only the random selection of “don’t care” values. In principle, it is possible to select some specific values, for example, maximizing the 3-way coverage of a test set.

Future work

• could include the investigation of the pair-wise stability for a general combinatorial test space.

• different pairwise tests from different tools