Supplementary Material

Using the Gini coefficient to characterize the shape of computational chemistry error distributions

Pascal PERNOT (pascal.pernot@universite-paris-saclay.fr)
Institut de Chimie Physique, CNRS, Univ. Paris-Sud,
Université Paris-Saclay, 91405, Orsay, France
Email: pascal.pernot@universite-paris-saclay.fr

Andreas SAVIN (andreas.savin@lct.jussieu.fr)
Laboratoire de Chimie Théorique,
CNRS and UPMC Université Paris 06,
Sorbonne Universités, F-75252 Paris, France
Email: andreas.savin@lct.jussieu.fr

Tuesday 24th November, 2020

Contents

Case	Property	Figure
BiasNorm	Normal distributions with the mean μ varying from 0.0 to 5.0 by 0.5	1
Student	Student's-t distributions with degrees of freedom in $\{2\text{-}10,20,50,100\}$	2
BiasStudent1	Same as above, with a shift of +1	3
BiasStudent2	Same as above, with a shift of +2	-
GandH	g-and-h distributions with asymmetry parameter g varying from 0.1 to 1.0 by 0.1	4
BiasGandH1	Same as above, with a shift of +1	5
BiasGandH2	Same as above, with a shift of +2	-

Table 1: Reference datasets. All sets were generated with $N = 1000$ samples.

1 Probability density functions and Lorenz curves

bla

1.1 Reference datasets

The ensemble of datasets designed to illustrate the dependence of the Gini coefficient on various parameters of distribution shape is summarized in Table 1. Figures have been generated for representative cases: BiasNorm (Fig. 1), Student (Fig. 2), StudentBias1 (Fig. 3), GAndH (Fig. 4) and GAndHBias1 (Fig. 5). The probability density functions are generated analytically, while the Lorenz curves and statistics in Table 2 are generated from random samples of $N = 10^3$ points.

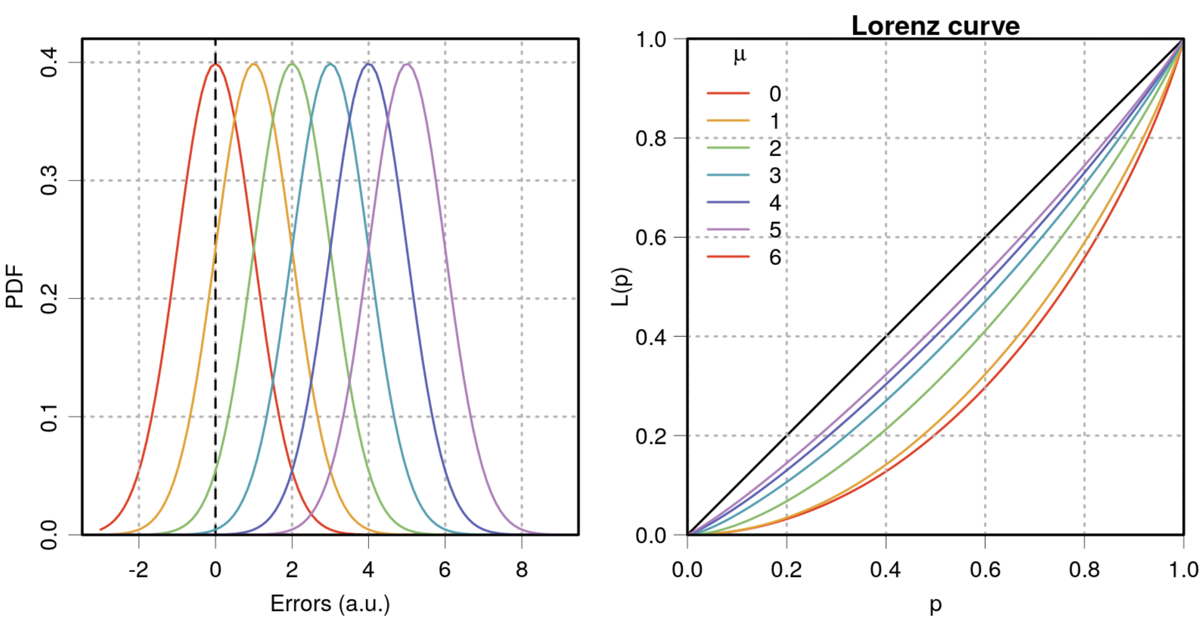

Figure 1: Probability density functions (left) Lorenz curves (right) for standard normal error sets with shifted mean value μ.

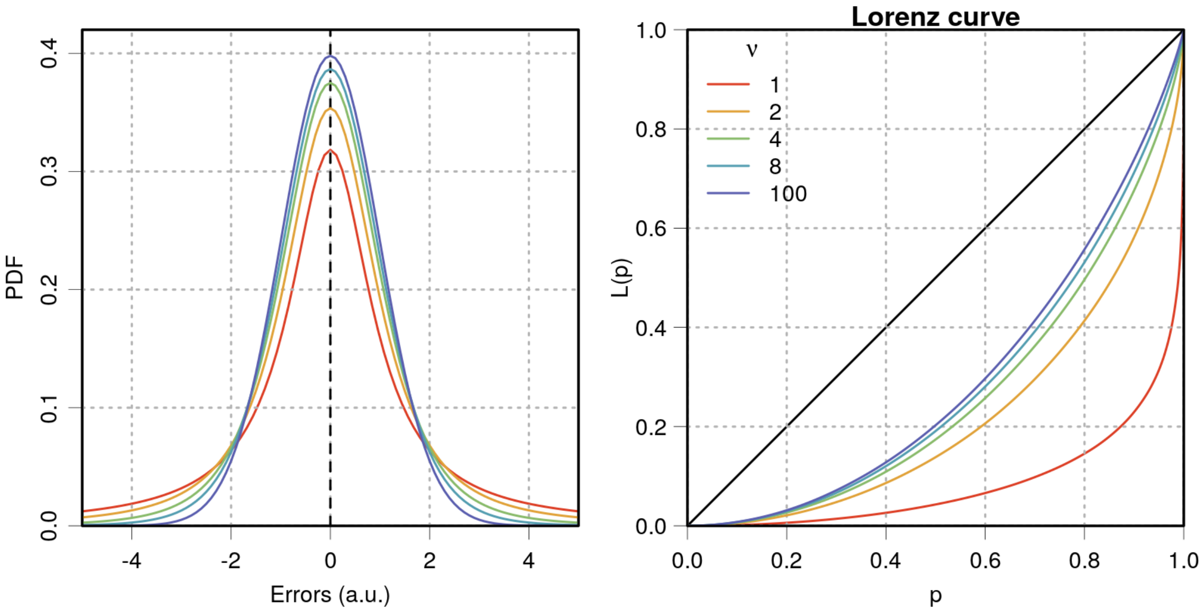

Figure 2: Probability density functions (left) Lorenz curves (right) for Student's-$t(\nu)$ distributions with different degrees of freedom ν. The case $\nu = 1$ (Cauchy distribution) has been included as an extreme case.

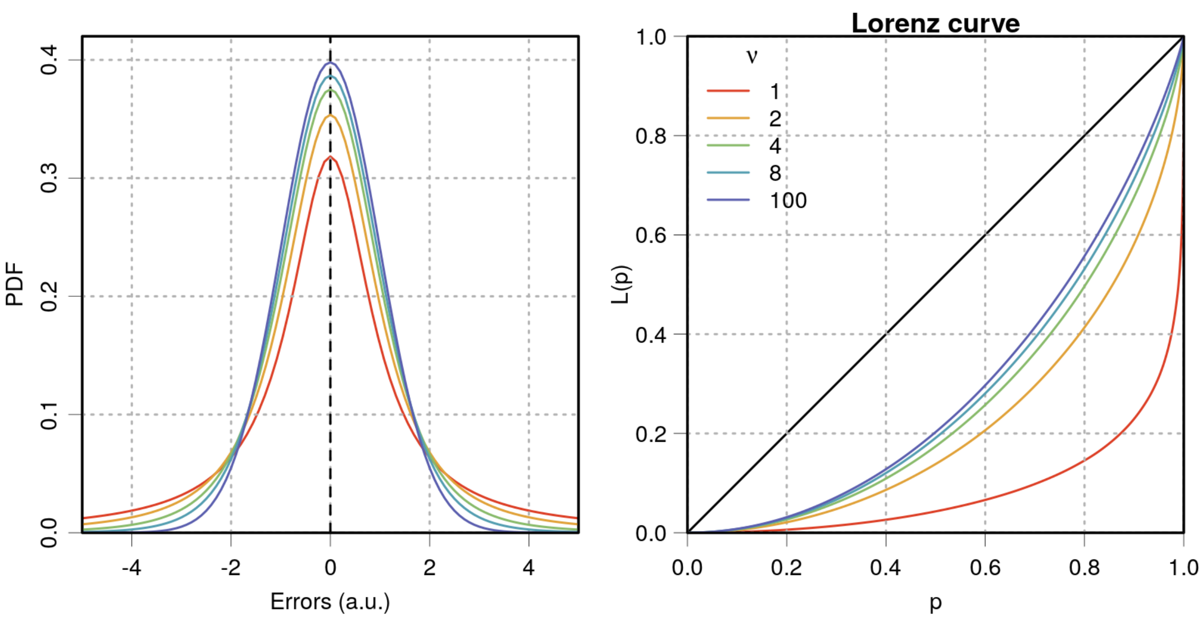

Figure 3: Probability density functions (left) Lorenz curves (right) for shifted Student's-$t(\nu)$ distributions with different degrees of freedom ν. The smaller ν, the heavier the tails.

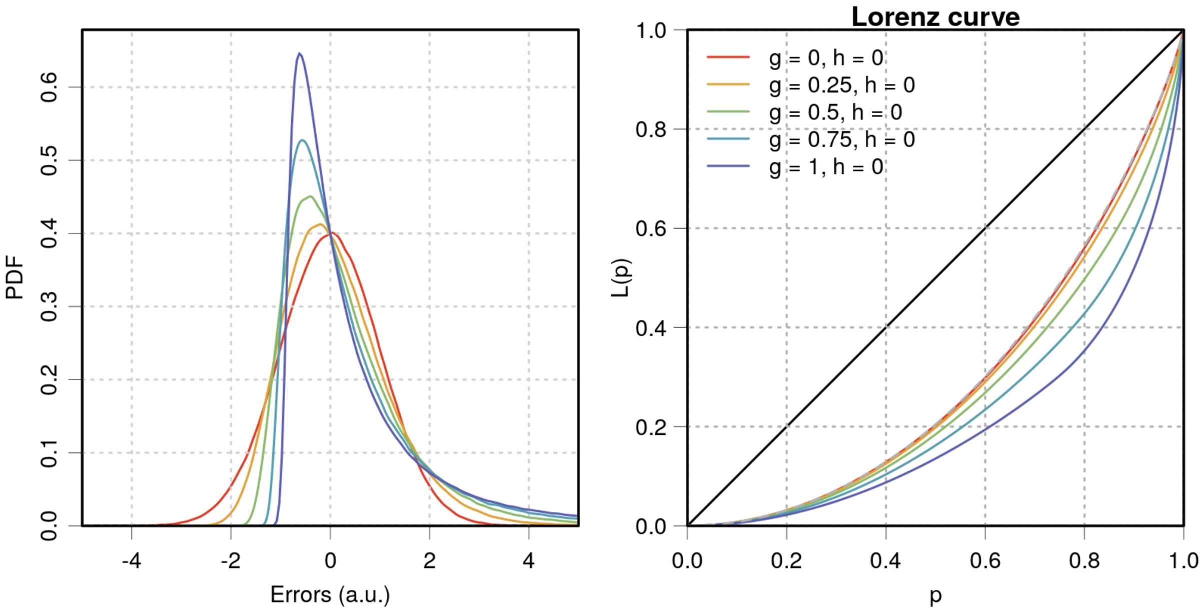

Figure 4: Probability density functions (left) Lorenz curves (right) for g-and-h distributions with different asymmetry parameters g [1, 2, 3].

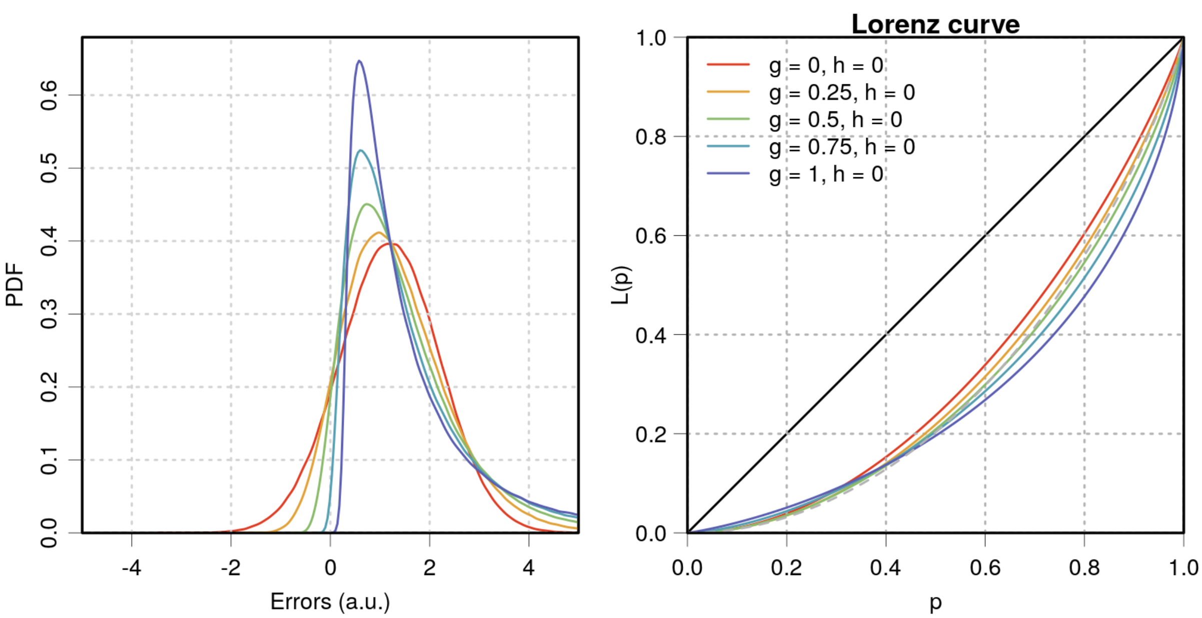

Figure 5: Probability density functions (left) Lorenz curves (right) for shifted g-and-h distributions with different asymmetry parameters g [1, 2, 3].

Table 2: Statistics of the reference datasets

Dataset	Parameters	MUE	MSE	$RMSD$	Q_{95}	κ_{CS}	G	$1/c_v$
NormBias	mu=0	0.786(19)	-0.004(31)	0.985(22)	1.936(50)	-0.07(12)	0.4178(79)	0.004(31)
NormBias	mu=0.5	0.909(22)	0.519(32)	1.009(21)	2.191(39)	-0.05(11)	0.4143(84)	0.514(33)
NormBias	mu=1	1.188(26)	1.039(32)	1.004(24)	2.718(57)	0.22(14)	0.3856(81)	1.034(40)
NormBias	mu=1.5	1.535(29)	1.477(32)	0.999(22)	3.106(80)	-0.02(12)	0.3347(73)	1.479(46)
NormBias	mu=2	2.051(30)	2.029(31)	1.030(22)	3.748(59)	-0.04(11)	0.2737(63)	1.969(52)
NormBias	mu=2.5	2.465(31)	2.460(31)	0.972(23)	4.122(72)	0.19(12)	0.2195(54)	2.530(68)
NormBias	mu=3	3.018(29)	3.017(29)	0.979(22)	4.610(58)	-0.02(13)	0.1827(46)	3.080(75)
NormBias	mu=3.5	3.493(32)	3.493(32)	1.046(22)	5.182(62)	-0.06(11)	0.1696(41)	3.341(77)
NormBias	mu=4	3.964(31)	3.964(31)	0.972(22)	5.520(65)	0.17(13)	0.1376(34)	4.080(98)
NormBias	mu=4.5	4.489(33)	4.489(33)	1.040(22)	6.257(64)	-0.07(12)	0.1310(30)	4.318(97)
NormBias	mu=5	4.995(31)	4.995(31)	0.980(24)	6.545(55)	-0.05(11)	0.1104(28)	5.10(13)
Student	df=2	1.415(86)	0.138(98)	3.12(53)	4.51(29)	2.29(39)	0.581(21)	0.044(32)
Student	df=3	1.107(55)	0.055(64)	2.04(45)	3.17(14)	1.27(22)	0.505(19)	0.027(32)
Student	df=4	1.030(34)	0.017(47)	1.493(95)	2.80(12)	0.79(20)	0.468(12)	0.011(31)
Student	df=5	0.907(28)	0.014(43)	1.280(86)	2.39(12)	0.52(21)	0.457(12)	0.011(34)
Student	df=6	0.913(24)	0.031(35)	1.174(32)	2.371(83)	0.29(14)	0.4258(80)	0.026(30)
Student	df=7	0.875(22)	-0.040(35)	1.131(30)	2.311(88)	0.32(15)	0.4319(83)	0.035(31)
Student	df=8	0.871(25)	0.016(36)	1.165(39)	2.305(94)	0.53(17)	0.4523(92)	0.014(31)
Student	df=9	0.869(23)	0.043(36)	1.123(30)	2.286(89)	0.27(14)	0.4355(89)	0.038(32)
Student	df=10	0.858(22)	-0.073(36)	1.093(30)	2.132(91)	0.23(17)	0.4178(89)	0.067(33)
Student	df=20	0.865(20)	0.075(35)	1.084(25)	2.184(70)	-0.02(12)	0.4132(85)	0.070(32)
Student	df=50	0.821(20)	0.012(32)	1.028(23)	2.019(49)	-0.04(11)	0.4098(77)	0.012(31)
Student	df=100	0.795(20)	-0.049(31)	1.011(24)	1.981(59)	0.15(14)	0.4297(78)	0.049(31)
StudentBias1	df=2	1.666(61)	0.992(71)	2.34(16)	4.50(31)	2.65(49)	0.483(13)	0.424(42)
StudentBias1	df=3	1.474(41)	1.001(51)	1.667(73)	3.89(15)	1.38(26)	0.4304(91)	0.600(40)
StudentBias1	df=4	1.367(38)	1.029(46)	1.459(72)	3.32(12)	0.85(21)	0.4209(97)	0.705(47)
StudentBias1	df=5	1.263(33)	0.987(41)	1.310(66)	3.10(12)	0.92(17)	0.4143(98)	0.753(49)
StudentBias1	df=6	1.263(30)	0.981(38)	1.230(36)	3.016(94)	0.74(20)	0.3940(84)	0.798(39)
StudentBias1	df=7	1.268(30)	1.028(39)	1.176(31)	2.949(92)	0.35(14)	0.3921(80)	0.875(40)
StudentBias1	df=8	1.233(27)	0.950(36)	1.173(35)	2.893(83)	0.40(15)	0.3887(79)	0.810(39)
StudentBias1	df=9	1.231(27)	0.998(35)	1.123(27)	2.874(73)	0.24(14)	0.3884(77)	0.889(38)
StudentBias1	df=10	1.214(27)	0.972(35)	1.115(30)	2.825(68)	0.44(15)	0.3884(75)	0.872(39)
StudentBias1	df=20	1.202(26)	1.027(32)	1.033(25)	2.721(45)	-0.04(12)	0.3820(80)	0.994(39)
StudentBias1	df=50	1.180(26)	0.994(32)	1.022(24)	2.665(88)	-0.05(12)	0.3789(79)	0.973(39)
StudentBias1	df=100	1.168(24)	0.963(32)	1.020(26)	2.596(74)	0.49(15)	0.3683(77)	0.944(40)

Dataset	Parameters	MUE	MSE	$RMSD$	Q_{95}	κ_{CS}	G	$1/c_v$
StudentBias2	df=2	2.48(17)	1.78(18)	5.8(2.5)	4.94(25)	1.87(42)	0.407(38)	0.31(14)
StudentBias2	df=3	2.148(47)	2.012(52)	1.67(13)	4.38(18)	1.08(25)	0.330(10)	1.208(99)
StudentBias2	df=4	2.141(37)	1.995(44)	1.442(56)	4.12(16)	1.18(27)	0.3049(81)	1.384(62)
StudentBias2	df=5	2.073(36)	1.974(41)	1.298(44)	3.96(11)	0.70(21)	0.2993(76)	1.521(60)
StudentBias2	df=6	2.072(36)	2.009(40)	1.214(40)	4.008(87)	0.07(14)	0.2997(69)	1.655(64)
StudentBias2	df=7	1.996(32)	1.926(36)	1.175(36)	3.95(11)	0.48(19)	0.2934(71)	1.640(59)
StudentBias2	df=8	2.029(32)	1.949(36)	1.170(39)	3.770(74)	0.48(16)	0.2826(70)	1.666(63)
StudentBias2	df=9	2.066(34)	2.023(37)	1.134(29)	3.905(68)	0.38(15)	0.2854(72)	1.783(56)
StudentBias2	df=10	2.016(34)	1.984(36)	1.108(28)	3.79(10)	0.18(12)	0.2927(72)	1.790(56)
StudentBias2	df=20	2.037(31)	2.015(33)	1.026(26)	3.775(73)	0.12(13)	0.2717(69)	1.965(59)
StudentBias2	df=50	2.068(30)	2.046(32)	1.015(22)	3.823(64)	0.08(12)	0.2665(61)	2.016(54)
StudentBias2	df=100	2.070(29)	2.056(30)	0.984(22)	3.703(44)	0.00(12)	0.2614(65)	2.089(56)
GandH	g=0.1	0.779(19)	0.092(30)	0.979(22)	1.946(55)	-0.01(12)	0.4173(82)	0.094(31)
GandH	g=0.2	0.798(19)	0.098(31)	1.013(26)	1.967(56)	-0.03(11)	0.4186(86)	0.096(31)
GandH	g=0.3	0.787(22)	0.134(32)	1.036(32)	2.03(11)	0.16(14)	0.4428(97)	0.130(31)
GandH	g=0.4	0.873(26)	0.222(36)	1.152(37)	2.44(12)	0.24(15)	0.4478(93)	0.192(32)
GandH	g=0.5	0.832(25)	0.239(36)	1.135(44)	2.28(11)	0.30(20)	0.455(11)	0.210(33)
GandH	g=0.6	0.888(32)	0.305(40)	1.308(69)	2.73(19)	0.85(24)	0.488(12)	0.233(33)
GandH	g=0.7	0.921(36)	0.356(43)	1.415(80)	3.02(18)	1.00(29)	0.514(12)	0.251(34)
GandH	g=0.8	0.954(37)	0.427(46)	1.425(72)	3.36(17)	0.52(22)	0.515(12)	0.300(36)
GandH	g=0.9	1.041(49)	0.506(56)	1.75(12)	3.95(28)	1.40(29)	0.553(13)	0.289(38)
GandH	g=1	1.132(72)	0.634(79)	2.45(39)	3.85(34)	1.89(45)	0.598(20)	0.259(52)
GandHBias1	g=0.1	1.215(27)	1.075(32)	1.036(23)	2.846(74)	0.02(13)	0.3963(81)	1.038(39)
GandHBias1	g=0.2	1.173(28)	1.074(31)	1.024(28)	2.97(12)	0.41(16)	0.4155(83)	1.049(42)
GandHBias1	g=0.3	1.236(33)	1.172(35)	1.134(53)	3.23(11)	0.23(15)	0.4399(90)	1.033(57)
GandHBias1	g=0.4	1.226(35)	1.174(37)	1.137(42)	3.38(18)	0.36(17)	0.4563(91)	1.032(50)
GandHBias1	g=0.5	1.281(38)	1.250(39)	1.230(55)	3.38(15)	0.41(25)	0.4700(98)	1.016(55)
GandHBias1	g=0.6	1.349(40)	1.331(40)	1.248(50)	3.70(13)	0.37(19)	0.4658(90)	1.067(53)
GandHBias1	g=0.7	1.364(43)	1.356(43)	1.352(99)	3.75(15)	0.67(24)	0.472(11)	1.002(80)
GandHBias1	g=0.8	1.456(52)	1.454(52)	1.64(11)	4.48(28)	1.12(32)	0.503(11)	0.889(68)
GandHBias1	g=0.9	1.443(49)	1.443(49)	1.52(11)	4.04(19)	1.58(41)	0.470(11)	0.951(76)
GandHBias1	g=1	1.664(67)	1.664(67)	2.12(18)	5.25(34)	2.14(48)	0.522(12)	0.786(74)
GandHBias2	g=0.1	2.122(32)	2.114(33)	1.046(26)	3.959(78)	-0.07(12)	0.2717(61)	2.021(59)
GandHBias2	g=0.2	2.094(32)	2.092(32)	1.004(28)	3.932(73)	0.13(12)	0.2658(63)	2.083(66)
GandHBias2	g=0.3	2.135(34)	2.135(34)	1.062(36)	4.08(13)	0.33(16)	0.2678(60)	2.010(75)
GandHBias2	g=0.4	2.210(37)	2.210(37)	1.164(42)	4.25(13)	0.33(18)	0.2798(63)	1.899(76)

Dataset	Parameters	MUE	MSE	$RMSD$	Q_{95}	κ_{CS}	G	$1/c_v$
GandHBias2	g=0.5	2.294(35)	2.294(35)	1.176(39)	4.60(11)	0.07(15)	0.2737(56)	1.951(71)
GandHBias2	g=0.6	2.286(38)	2.286(38)	1.213(50)	4.70(16)	0.75(21)	0.2704(64)	1.884(84)
GandHBias2	g=0.7	2.388(47)	2.388(47)	1.52(13)	5.09(17)	0.77(28)	0.2882(88)	1.57(14)
GandHBias2	g=0.8	2.460(54)	2.460(54)	1.66(11)	5.40(23)	1.01(31)	0.3056(89)	1.48(10)
GandHBias2	g=0.9	2.548(57)	2.548(57)	1.82(16)	5.77(30)	1.59(36)	0.3100(99)	1.40(13)
GandHBias2	g=1	2.577(61)	2.577(61)	1.99(17)	5.94(37)	1.73(41)	0.318(11)	1.29(11)

Case	Property (units)	N	K	Source	Figure
BOR2019	Band gaps (eV)	471	15	[4]	
HAI2018	Dipole moments (Debye)	149	5	[5]	
NAR2019	Enthalpies of formation (kcal/mol)	469	4	[6]	
PER2018	Intensive atomization energies (kcal/mol)	222	9	[7]	
SCH2018	Chemisorption energies (kcal/mol)	195	7	[8]	
THA2015	Polarizability	135	7	[9]	
WU2015	Polarizability	145	34	[10]	
ZAS2019	Effective atomization energies (kcal/mol)	6211	3	[11]	
ZHA2018	Solid formation enthalpies (kcal/mol)	196	2	[12]	

Table 3: Case studies: N is the number of systems in the dataset and K is the number of methods.

1.2 Literature datasets

bla

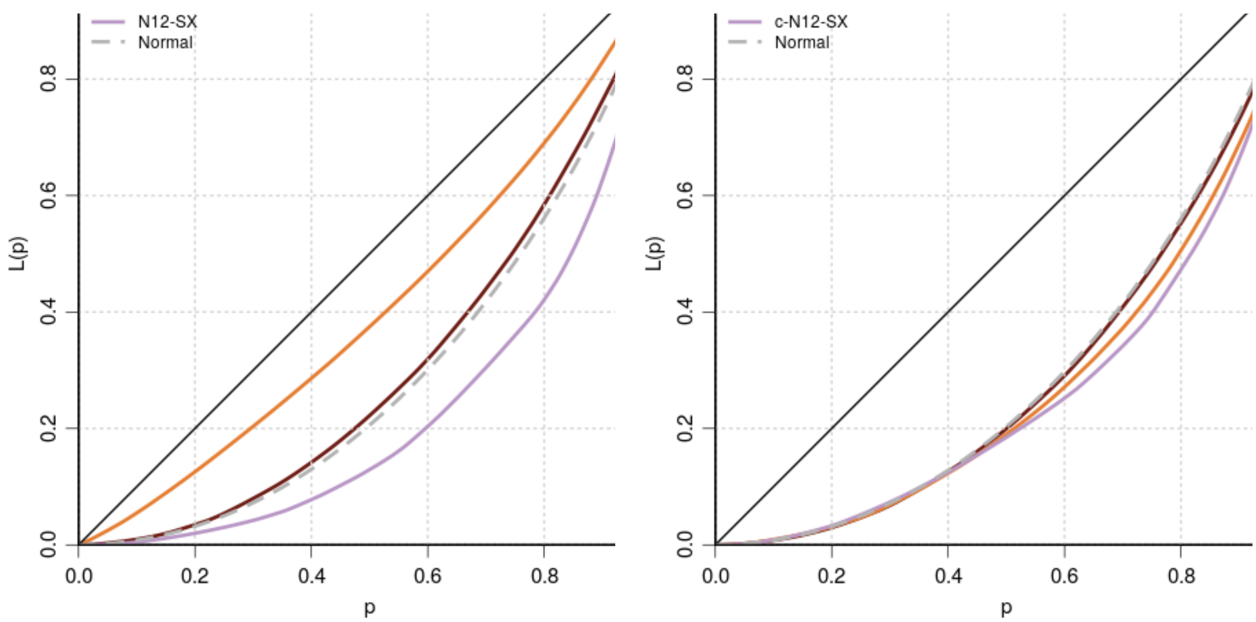

Figure 6: Lorenz curves for three typical error sets from WU2015 (left), and the impact of data centering (right). MP2 is practically normal, M11-L is strongly biased and N12-SX has heavy tails. After centering, the heavy tails of M11-L are revealed.

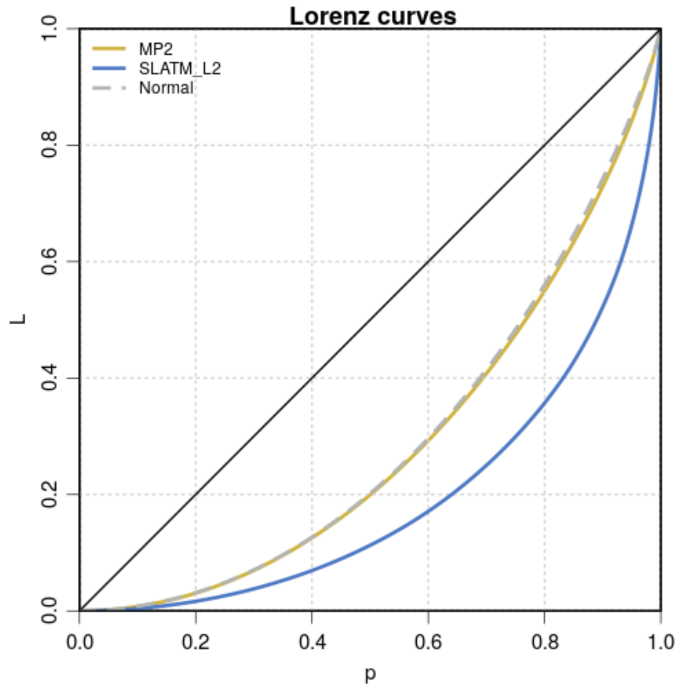

Figure 7: Lorenz curves for error sets of MP2 and SLATM-L2

Table 4: Statistics of the literature datasets

Dataset	Methods	MUE	MSE	$RMSD$	Q_{95}	κ_{CS}	G	$1/c_v$
BOR2019	LDA	1.158(47)	1.143(47)	1.051(95)	3.14(20)	1.43(45)	0.421(16)	1.09(11)
BOR2019	LDA + SOC	1.229(46)	1.222(47)	1.042(96)	3.21(20)	1.51(46)	0.397(15)	1.17(12)
BOR2019	PBE	1.043(44)	1.017(45)	1.006(95)	2.92(18)	1.46(46)	0.439(16)	1.01(11)
BOR2019	PBE + SOC	1.115(44)	1.104(44)	0.993(96)	3.00(19)	1.50(46)	0.412(16)	1.11(12)
BOR2019	PBE_SOL	1.108(45)	1.089(46)	1.020(91)	3.04(18)	1.39(45)	0.428(16)	1.07(11)
BOR2019	HLE16	0.588(34)	0.363(41)	0.865(79)	1.85(17)	1.48(38)	0.534(17)	0.420(61)
BOR2019	HLE16 + SOC	0.599(35)	0.500(38)	0.827(85)	1.91(17)	1.53(39)	0.544(17)	0.605(77)
BOR2019	BJ	0.782(36)	0.718(38)	0.845(72)	2.27(18)	1.60(46)	0.463(16)	0.849(85)
BOR2019	mBJ	0.487(22)	0.212(30)	0.647(35)	1.390(75)	0.67(28)	0.482(14)	0.328(50)
BOR2019	SCAN	0.803(38)	0.733(40)	0.896(85)	2.39(18)	1.44(44)	0.469(17)	0.818(90)
BOR2019	HSE06	0.517(29)	0.089(36)	0.820(73)	1.66(15)	2.55(48)	0.540(16)	0.108(45)
BOR2019	HSE14	0.564(28)	-0.182(36)	0.802(54)	1.79(12)	2.28(39)	0.506(14)	0.227(47)
BOR2019	HSE_mix	0.635(31)	-0.380(38)	0.830(45)	1.95(12)	1.93(38)	0.512(14)	0.458(52)
BOR2019	PBE0	0.770(27)	-0.482(37)	0.842(70)	1.80(12)	2.74(47)	0.374(14)	0.573(65)
BOR2019	PBE0_mix	0.811(36)	-0.604(43)	0.946(49)	2.42(14)	1.73(37)	0.472(13)	0.639(56)
HAI2018	LSDA	0.1301(98)	0.056(14)	0.171(12)	0.355(52)	0.73(48)	0.491(27)	0.330(85)
HAI2018	GGA	0.102(13)	0.061(14)	0.183(41)	0.340(76)	2.3(1.5)	0.571(40)	0.33(11)
HAI2018	mGGA	0.0841(51)	0.0557(74)	0.0919(56)	0.223(17)	0.83(46)	0.421(20)	0.606(89)
HAI2018	hGGA	0.0509(39)	-0.0121(56)	0.0696(54)	0.161(19)	1.07(53)	0.482(22)	0.173(82)
HAI2018	hmGGA	0.0511(37)	0.0155(56)	0.0682(49)	0.141(18)	1.07(47)	0.475(23)	0.227(84)
NAR2019	G4MP2	0.794(33)	-0.061(48)	1.057(46)	2.209(87)	0.94(26)	0.457(12)	0.057(45)
NAR2019	B3LYP	3.99(15)	3.61(17)	3.67(32)	9.25(55)	0.92(30)	0.400(16)	0.983(97)
NAR2019	M06-2X	2.705(99)	-0.09(17)	3.46(20)	6.10(46)	-0.21(18)	0.395(14)	0.025(49)
NAR2019	wB97X-D	1.854(92)	0.52(13)	2.65(25)	5.18(38)	0.53(23)	0.468(18)	0.197(52)
PER2018	B3LYP	1.182(90)	0.98(10)	1.51(12)	4.49(49)	4.93(96)	0.515(20)	0.648(84)
PER2018	B97-1	0.855(52)	0.506(69)	1.051(82)	2.67(35)	3.48(72)	0.444(21)	0.481(76)
PER2018	BH&HLYP	4.83(23)	4.79(23)	3.50(24)	11.74(61)	0.60(48)	0.368(15)	1.37(11)
PER2018	BLYP	1.63(10)	0.41(14)	2.23(17)	5.33(56)	3.47(78)	0.465(20)	0.184(64)
PER2018	CAM-B3LYP	0.904(91)	0.55(10)	1.54(12)	4.13(37)	5.1(1.3)	0.663(15)	0.357(71)
PER2018	LC-wPBE	1.094(95)	0.73(11)	1.67(14)	4.34(51)	2.82(82)	0.603(18)	0.438(75)
PER2018	PBE	2.79(16)	-2.53(17)	2.67(20)	8.06(83)	1.97(52)	0.423(18)	0.949(96)
PER2018	PBE0	0.925(74)	0.236(96)	1.44(12)	3.27(50)	2.99(86)	0.565(21)	0.164(68)
PER2018	PW86PBE	1.64(12)	-0.54(16)	2.46(20)	6.14(89)	2.68(59)	0.528(21)	0.217(67)
SCH2018	LDA	0.716(31)	0.685(35)	0.486(24)	1.515(70)	0.14(23)	0.345(17)	1.41(10)
SCH2018	PBE	0.1869(99)	0.093(16)	0.2126(94)	0.418(23)	-0.09(22)	0.414(20)	0.439(78)

Dataset	Methods	MUE	MSE	$RMSD$	Q_{95}	κ_{CS}	G	$1/c_v$
SCH2018	RPBE	0.212(10)	-0.172(13)	0.1886(76)	0.463(21)	-0.64(20)	0.376(18)	0.911(78)
SCH2018	vdW-DF2	0.209(10)	0.023(17)	0.254(11)	0.503(35)	-0.06(27)	0.388(18)	0.092(67)
SCH2018	BEEF-vdW	0.1457(80)	-0.009(13)	0.1824(83)	0.340(12)	-0.09(27)	0.421(17)	0.052(71)
SCH2018	mBEEF	0.2056(99)	0.058(18)	0.242(14)	0.486(30)	0.24(32)	0.376(18)	0.239(76)
SCH2018	mBEEF-vdW	0.312(13)	0.226(21)	0.285(14)	0.636(23)	0.00(29)	0.342(16)	0.794(83)
THA2015	M11	0.0257(20)	0.0080(31)	0.0339(26)	0.0777(60)	1.80(46)	0.486(21)	0.235(93)
THA2015	M06-2X	0.0272(20)	0.0127(30)	0.0332(26)	0.0775(71)	1.51(43)	0.456(25)	0.382(95)
THA2015	wB97	0.0280(20)	0.0077(32)	0.0352(28)	0.0740(76)	0.87(46)	0.431(25)	0.217(93)
THA2015	LC-tHCTH	0.0254(19)	0.0040(30)	0.0332(25)	0.0722(74)	0.72(44)	0.451(23)	0.119(91)
THA2015	HISS	0.0335(20)	0.0235(30)	0.0325(26)	0.0736(78)	1.42(46)	0.363(21)	0.72(11)
THA2015	LC-wPBE	0.0339(19)	0.0206(31)	0.0348(28)	0.0771(74)	1.17(54)	0.354(20)	0.59(10)
THA2015	MP2	0.0270(21)	0.0086(32)	0.0351(34)	0.0713(76)	1.08(64)	0.447(27)	0.245(94)
WU2015	HF	0.0424(18)	0.0378(22)	0.0295(24)	0.0803(33)	0.46(39)	0.301(18)	1.28(13)
WU2015	MP2	0.00529(33)	-0.00413(42)	0.00497(32)	0.01197(98)	0.26(37)	0.387(21)	0.83(10)
WU2015	LSDA	0.0654(18)	-0.0644(21)	0.0247(26)	0.1120(72)	1.50(77)	0.177(12)	2.61(29)
WU2015	PBE	0.0608(18)	-0.0598(21)	0.0244(26)	0.1080(70)	1.93(84)	0.186(13)	2.45(28)
WU2015	HCTH	0.0489(18)	-0.0475(21)	0.0248(26)	0.0962(69)	1.55(80)	0.235(16)	1.92(22)
WU2015	N12	0.0397(20)	-0.0378(22)	0.0268(26)	0.0920(76)	2.27(86)	0.313(19)	1.41(16)
WU2015	revTPSS	0.0455(18)	-0.0442(20)	0.0239(24)	0.0930(75)	2.10(92)	0.244(17)	1.85(20)
WU2015	tHCTH	0.0456(19)	-0.0441(21)	0.0253(26)	0.0947(74)	1.89(84)	0.257(17)	1.74(20)
WU2015	M11-L	0.0832(23)	-0.0822(25)	0.0310(29)	0.1378(58)	0.99(79)	0.181(10)	2.65(26)
WU2015	MN12-L	0.0490(22)	-0.0476(24)	0.0296(24)	0.1085(75)	1.12(66)	0.297(16)	1.61(15)
WU2015	PBE0	0.0196(14)	-0.0163(17)	0.0206(19)	0.0584(44)	2.05(70)	0.452(20)	0.79(11)
WU2015	mPW1PW	0.0175(14)	-0.0135(17)	0.0206(18)	0.0570(42)	2.06(70)	0.494(21)	0.66(10)
WU2015	B3LYP	0.0310(15)	-0.0296(17)	0.0210(20)	0.0722(49)	1.98(78)	0.317(18)	1.41(16)
WU2015	B3PW91	0.0214(15)	-0.0186(17)	0.0211(20)	0.0623(50)	2.18(74)	0.435(21)	0.88(12)
WU2015	B97-2	0.0195(15)	-0.0159(17)	0.0215(20)	0.0606(47)	2.33(77)	0.470(21)	0.74(11)
WU2015	M06	0.0289(13)	-0.0276(14)	0.0181(18)	0.0602(44)	1.16(79)	0.298(16)	1.53(17)
WU2015	M06-2X	0.01244(97)	-0.0027(13)	0.0167(13)	0.0385(37)	0.62(52)	0.468(22)	0.159(79)
WU2015	M06-HF	0.01749(99)	0.0081(15)	0.0195(14)	0.0416(45)	0.62(42)	0.363(22)	0.418(83)
WU2015	SOGGA11X	0.0150(13)	-0.0084(16)	0.0198(15)	0.0516(32)	1.25(56)	0.516(20)	0.423(87)
WU2015	tHCTHhyb	0.0297(17)	-0.0279(19)	0.0228(22)	0.0753(60)	2.45(83)	0.354(20)	1.22(14)
WU2015	BMK	0.0162(13)	-0.0078(17)	0.0212(15)	0.0523(38)	0.76(46)	0.493(21)	0.366(84)
WU2015	TPSSh	0.0301(16)	-0.0283(18)	0.0225(22)	0.0746(61)	2.40(85)	0.342(20)	1.26(15)
WU2015	LC-rTPSS	0.0250(12)	0.0240(13)	0.01638(99)	0.0525(29)	0.09(30)	0.335(22)	1.47(12)
WU2015	LC-wPBE	0.0201(10)	0.0177(13)	0.01589(89)	0.0462(26)	0.08(28)	0.358(21)	1.12(10)

Dataset	Methods	MUE	MSE	$RMSD$	Q_{95}	κ_{CS}	G	$1/c_v$
WU2015	LC-N12	0.0380(13)	0.0376(14)	0.0179(11)	0.0676(30)	0.10(31)	0.254(17)	2.10(15)
WU2015	LC-HCTH	0.01763(98)	0.0131(13)	0.0169(12)	0.0435(29)	0.45(36)	0.374(21)	0.775(95)
WU2015	LC-tHCTH	0.0138(10)	-0.0011(14)	0.0182(13)	0.0367(25)	0.37(35)	0.451(22)	0.059(77)
WU2015	M11	0.01329(87)	-0.0024(13)	0.0166(11)	0.0352(33)	0.10(36)	0.408(21)	0.146(79)
WU2015	CAM-B3LYP	0.0151(12)	-0.0103(14)	0.0180(13)	0.0466(28)	0.83(43)	0.497(22)	0.571(88)
WU2015	wB97XD	0.0159(12)	-0.0112(15)	0.0183(13)	0.0478(27)	0.74(42)	0.484(22)	0.612(93)
WU2015	wB97	0.01446(87)	0.0014(14)	0.01769(97)	0.0339(19)	-0.07(29)	0.391(18)	0.079(79)
WU2015	wB97X	0.01409(98)	-0.0055(14)	0.0173(10)	0.0363(19)	0.27(29)	0.442(19)	0.314(83)
WU2015	N12-SX	0.0158(15)	-0.0073(18)	0.0224(19)	0.0579(50)	2.05(61)	0.542(21)	0.327(85)
WU2015	MN12-SX	0.0482(18)	-0.0472(19)	0.0241(21)	0.0959(60)	1.27(72)	0.242(14)	1.96(19)
WU2015	HSE06	0.0209(15)	-0.0180(17)	0.0208(19)	0.0605(47)	2.09(73)	0.436(20)	0.87(11)
WU2015	HISS	0.0187(11)	0.0105(16)	0.0202(16)	0.0425(40)	1.36(50)	0.373(22)	0.520(89)
ZAS2019	HF	2.370(25)	0.062(41)	3.111(40)	6.010(99)	0.324(58)	0.4360(38)	0.020(13)
ZAS2019	MP2	1.305(13)	-0.004(21)	1.660(16)	3.318(51)	0.119(64)	0.4236(35)	0.002(13)
ZAS2019	SLATM_L2	1.238(25)	0.108(30)	2.354(75)	4.61(15)	4.34(24)	0.6051(54)	0.046(13)
ZHA2018	PBE	0.2071(96)	-0.204(10)	0.1408(70)	0.442(29)	-0.34(25)	0.368(19)	1.45(10)
ZHA2018	SCAN	0.0968(64)	-0.0164(98)	0.1313(86)	0.270(15)	1.29(53)	0.487(19)	0.125(75)

References

[1] D. C. Hoaglin. *Exploring data tables, trends, and shapes*, chapter Summarizing shape numerically: The g-and-h distributions, pages 461–513. Wiley, New York, 1985.

[2] R. R. Wilcox and D. M. Erceg-Hurn. Comparing two dependent groups via quantiles. *J. App. Stat.*, 39:2655–2664, 2012. doi:10.1080/02664763.2012.724665.

[3] P. Pernot and A. Savin. Probabilistic performance estimators for computational chemistry methods: Systematic improvement probability and ranking probability matrix. I. Theory. *J. Chem. Phys.*, 152:164108, 2020. doi:10.1063/5.0006202.

[4] P. Borlido, T. Aull, A. W. Huran, F. Tran, M. A. Marques, and S. Botti. Large-scale benchmark of exchange–correlation functionals for the determination of electronic band gaps of solids. *J. Chem. Theory Comput.*, 15:5069–5079, 2019. doi:10.1021/acs.jctc.9b00322.

[5] D. Hait and M. Head-Gordon. How accurate is density functional theory at predicting dipole moments? an assessment using a new database of 200 benchmark values. *J. Chem. Theory Comput.*, 14:1969–1981, 2018. doi:10.1021/acs.jctc.7b01252.

[6] B. Narayanan, P. C. Redfern, R. S. Assary, and L. A. Curtiss. Accurate quantum chemical energies for 133000 organic molecules. *Chem. Sci.*, 10:7449–7455, 2019. doi:10.1039/c9sc02834j.

[7] P. Pernot and A. Savin. Probabilistic performance estimators for computational chemistry methods: the empirical cumulative distribution function of absolute errors. *J. Chem. Phys.*, 148:241707, 2018. doi:10.1063/1.5016248.

[8] P. S. Schmidt and K. S. Thygesen. Benchmark database of transition metal surface and adsorption energies from many-body perturbation theory. *J. Phys. Chem. C*, 122:4381–4390, 2018. doi:10.1021/acs.jpcc.7b12258.

[9] A. J. Thakkar and T. Wu. How well do static electronic dipole polarizabilities from gas-phase experiments compare with density functional and MP2 computations? *J. Chem. Phys.*, 143:144302, 2015. doi:10.1063/1.4932594.

[10] T. Wu, Y. N. Kalugina, and A. J. Thakkar. Choosing a density functional for static molecular polarizabilities. *Chem. Phys. Lett.*, 635:257–261, 2015. doi:10.1016/j.cplett.2015.07.003.

[11] P. Zaspel, B. Huang, H. Harbrecht, and O. A. von Lilienfeld. Boosting quantum machine learning models with a multilevel combination technique: Pople diagrams revisited. *J. Chem. Theory Comput.*, 15(3):1546–1559, 2019. doi:10.1021/acs.jctc.8b00832.

[12] Y. Zhang, D. A. Kitchaev, J. Yang, T. Chen, S. T. Dacek, R. A. Sarmiento-Perez, M. A. L. Marques, H. Peng, G. Ceder, J. P. Perdew, and J. Sun. Efficient first-principles prediction of solid stability: Towards chemical accuracy. *npj Comput. Mater.*, 4:9, 2018. doi:10.1038/s41524-018-0065-z.