

Приложение V

Б. Ю. Лемешко, А. А. Горбунова

Построение вероятностной модели, описывающей ГТП УЧП

Настоящее приложение содержит анализ данных, по которым в тексте построены графики 7.3, 7.27 (США. 1980–2003. Распределение ГТП отраслевой УЧП по вероятности).

Исходная выборка ГТП УЧП по всем отраслям представляет собой группированные (интервальные) данные, являющиеся результатом сжатия первичной статистической информации. Повидимому, макроэкономическая статистика всегда опирается на данные такого рода. Особенностью таких данных является и то, что границы интервалов оказываются неизвестными.

Будем рассуждать так же, как и при анализе вероятностной модели ГТП ОПФ (см. Приложение IV в [1]). Каждому значению случайной величины ГТП УЧП соответствует определенное значение УЧП, обладающее таким «средним» значением ГТП. То есть каждой единице измерения (каждому \$) из УЧП данной строки поставлена в соответствие величина ГТП УЧП из этой строки (такое среднее значение ГТП). Можно считать, что это значение ГТП определяет собой середину интервала значений (среднее значение ГТП) для соответствующей величины УЧП.

В настоящем случае статистический анализ вероятностной модели ГТП УЧП 56 отраслей проводился по исходным данным с первоначальным количеством интервалов 1344 (1346, если учесть левые и правые интервалы с бесконечными границами, в которые попало нулевое количество наблюдений). При этом в качестве граничных точек интервалов брались значения, равные средним значениям между соседними значениями ГТП УЧП.

Проблемы и особенности построения вероятностных моделей по группированным и интервальным выборкам упоминаются в [2–7].

Отметим два момента, которые считаем важными при статистическом анализе такого рода данных и о чем ранее уже упоминалось в [1]. Во-первых, попытки повторного группирования таких данных, например, с разбиением области изменения ГТП УЧП на 1% интервалы равной длины с соответствующим преобразованием выборки неминуемо приведут к потере информации о предполагаемом законе и к смещению построенной модели относительно истинного закона распределения. Во-вторых, попытки корректного описания закона распределения, соответствующего настоящей выборке ГТП УЧП или выборке ГТП ОПФ, рассмотренной в [1], некоторой параметрической моделью, наилучшей из некоторого (пусть самого широкого) множества параметрических моделей, заведомо обречены на провал. В данном случае имеется в виду идентификация закона, включающая оценивание параметров множества моделей, проверку адекватности этих моделей с использованием критериев согласия и выбор наиболее подходящей.

Анализ модели по полной выборке

На рисунках 1 и 2 приведены эмпирическая функция распределения ГТП УЧП и функции распределения двух нормальных законов, представленные на странице ГТП УЧП (США) файла в формате Excel. Смещение функции распределения нормального закона на рисунке 2 объясняется как раз «повторным» группированием, сделанным в процессе построения модели.

На рисунке 3 нормальный закон с рисунка 2 сдвинут для устранения «смещения».

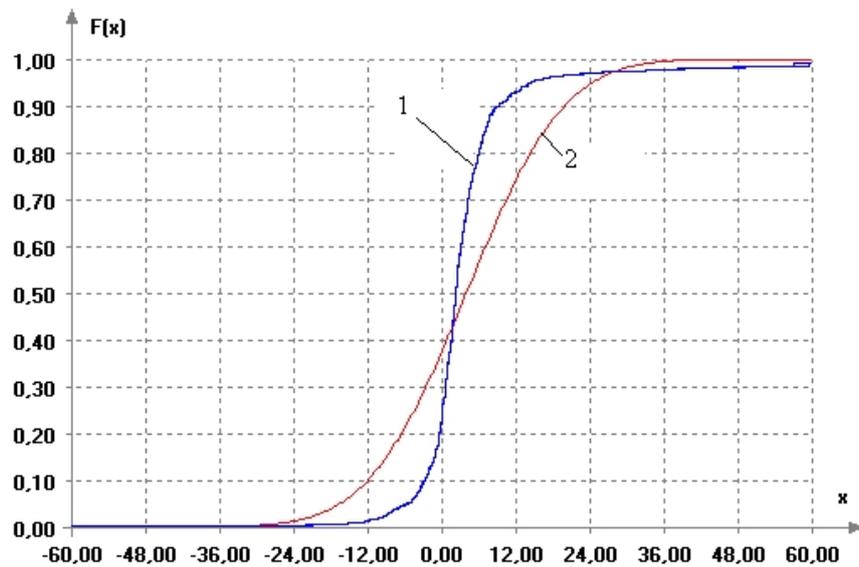


Рис.1. Эмпирическая функция распределения ГТП УЧП США по полной выборке (1) и функция распределения нормального закона с параметрами $\mu = 4,071$, $\sigma = 12,445$ (2)

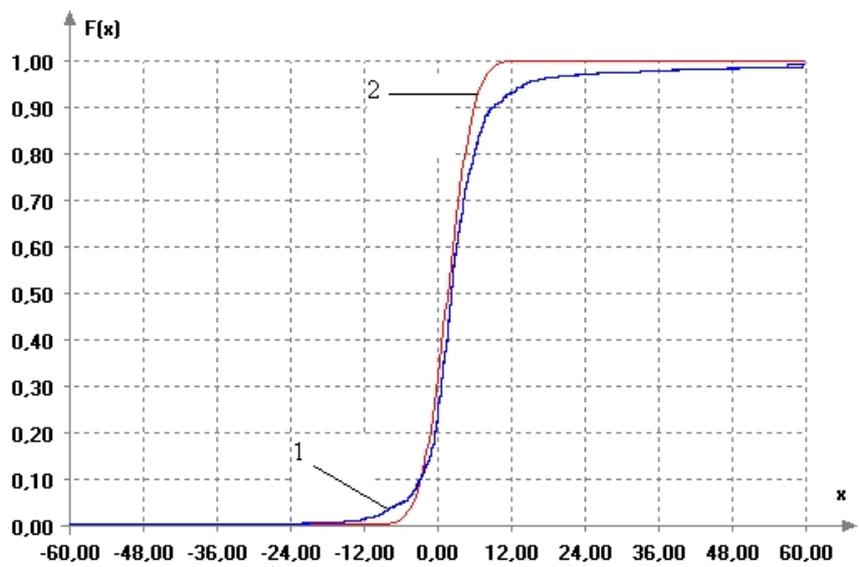


Рис.2. Эмпирическая функция распределения ГТП УЧП США по полной выборке (1) и функция распределения нормального закона с параметрами $\mu = 1,749$, $\sigma = 3,399$ (2)

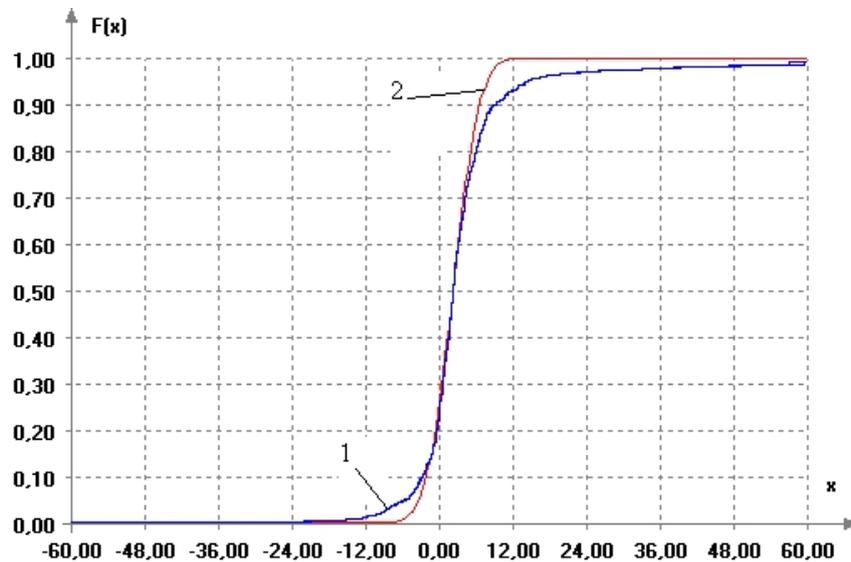


Рис.3. Эмпирическая функция распределения ГТП УЧП США по полной выборке (1) и функция распределения нормального закона с параметрами $\mu = 2,3243$, $\sigma = 3,399$ (2)

На рисунке 4 представлена эмпирическая функция распределения ГТП УЧП по группированным данным (1346 интервалов) и функция распределения нормального закона, с оценками параметров, найденными методом максимального правдоподобия по группированным данным. Очевидно, что использование для описания ГТП УЧП нормального распределения с такими параметрами (как и законов, приведенных на рисунках выше) не даст возможности для более или менее точных предсказаний вероятностей. Эмпирическое распределение обладает «тяжелыми хвостами» и к тому же является асимметричным – с более выраженным правым хвостом.

Если использовать для описания различные параметрические модели, то наиболее подходящим законом оказывается распределение Коши с плотностью

$$f(x) = \frac{\theta_0}{\pi \left[\theta_0^2 + (x - \theta_1)^2 \right]}$$

и с оценками максимального правдоподобия (ОМП) параметров $\theta_0 = 2,2793$, $\theta_1 = 2,2216$ (см. рисунок 5). Однако и в этом случае

слишком очевидно отличие «левого хвоста» теоретического закона от эмпирического.

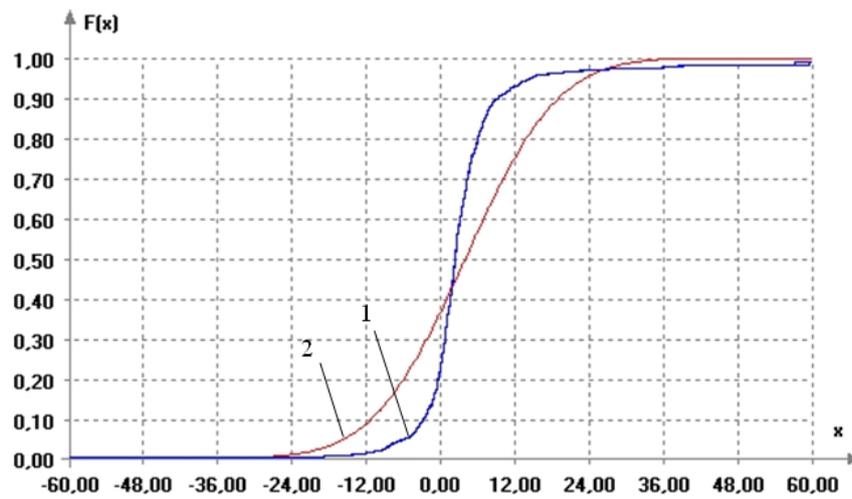


Рис. 4. Эмпирическая функция распределения ГТП УЧП США по полной выборке (1) и функция распределения нормального закона с параметрами $\mu = 4,0735$, $\sigma = 11,7658$ (2)

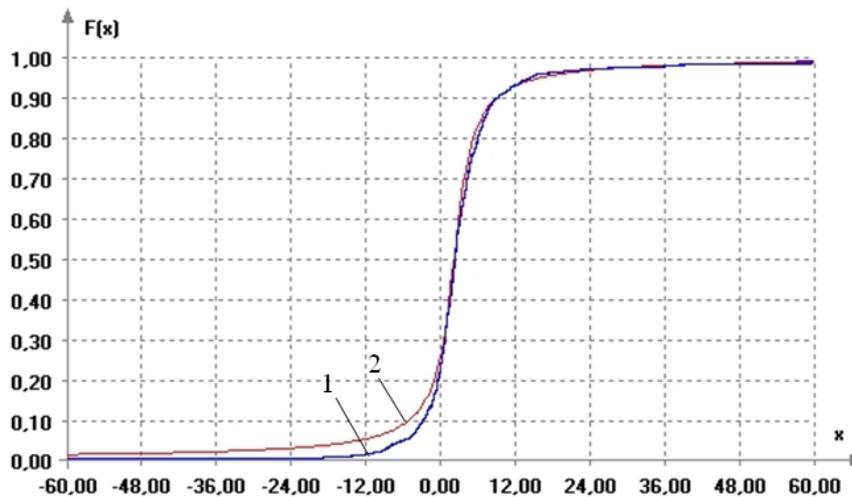


Рис. 5. Эмпирическая функция распределения ГТП УЧП США по полной выборке (1) и функция распределения Коши с оценками параметров $\theta_0 = 2,2793$, $\theta_1 = 2,2216$ (2)

В то же время вид эмпирической функции распределения ГТП УЧП (США) по всем отраслям позволяет утверждать, что эту функцию распределения невозможно с достаточной точностью описать в рамках некоторой простой параметрической модели (нормальным или каким-либо другим законом).

Анализ ГТП УЧП отдельных отраслей

Распределения ГТП УЧП по отдельным отраслям имеют слишком большие различия, в том числе отличаются области определения случайных величин, соответствующих различным отраслям (см., например, рис. 6). Поэтому их объединение в одну выборку не приводит (не выполняются условия центральной предельной теоремы) к нормальному закону для всей совокупности отраслей.

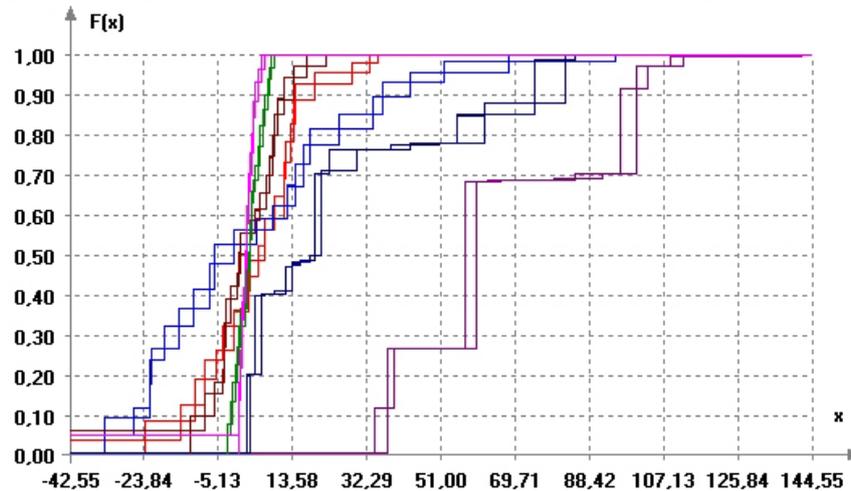


Рис.6. Эмпирические функции распределения ГТП УЧП США по выборкам из нескольких отраслей

Анализ ГТП УЧП по каждой из 56 отраслей показал, что в одних случаях нормальный закон оказывается приемлемым для описания закона распределения ГТП УЧП соответствующей отрасли (например, рис. 7–8), в других – истинный закон распределения весьма далек от нормального (см., например, рис.9–11).

В целом же, *нормальный закон оказывается достаточно плохим приближением для эмпирических распределений ГТП УЧП большинства отраслей.*

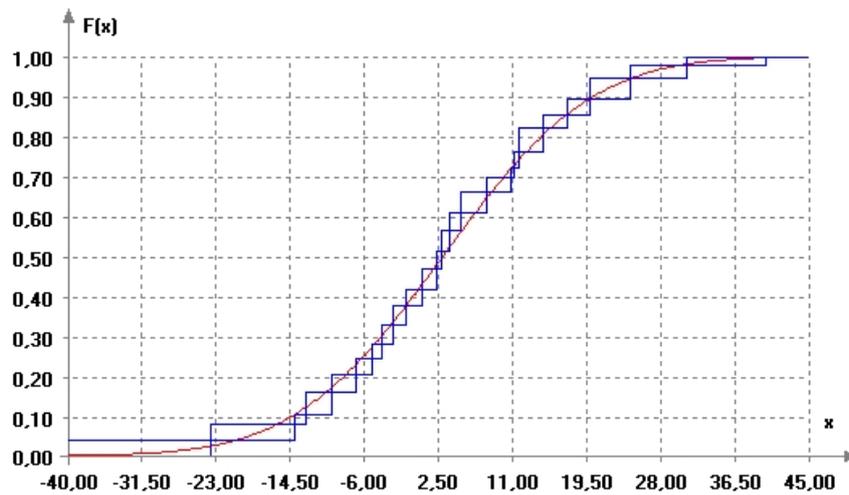


Рис.7. Эмпирическая функция распределения ГТП УЧП США по выборке из отрасли 27 (Motor vehicles) и функция распределения нормального закона с параметрами $\mu = 3,0654$, $\sigma = 13,4449$

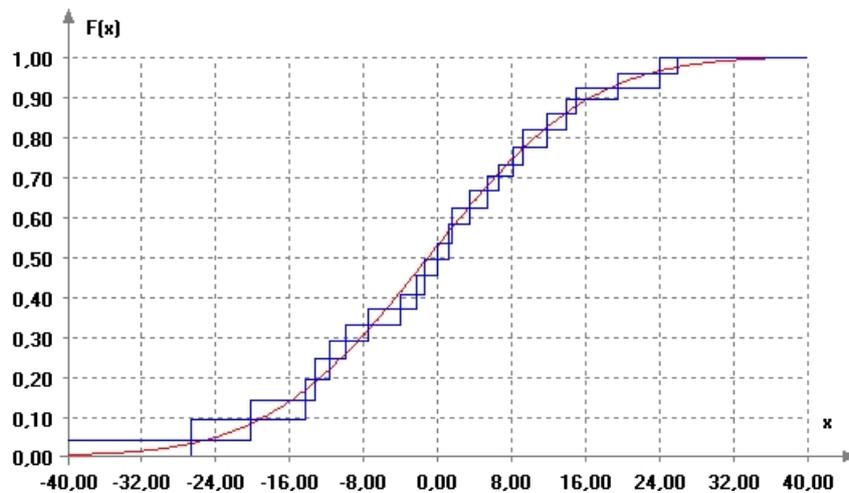


Рис.8. Эмпирическая функция распределения ГТП УЧП США по выборке из отрасли 30 (Railroad equipment and transport equipment) и функция распределения нормального закона с параметрами $\mu = -0,8848$, $\sigma = 13,7479$

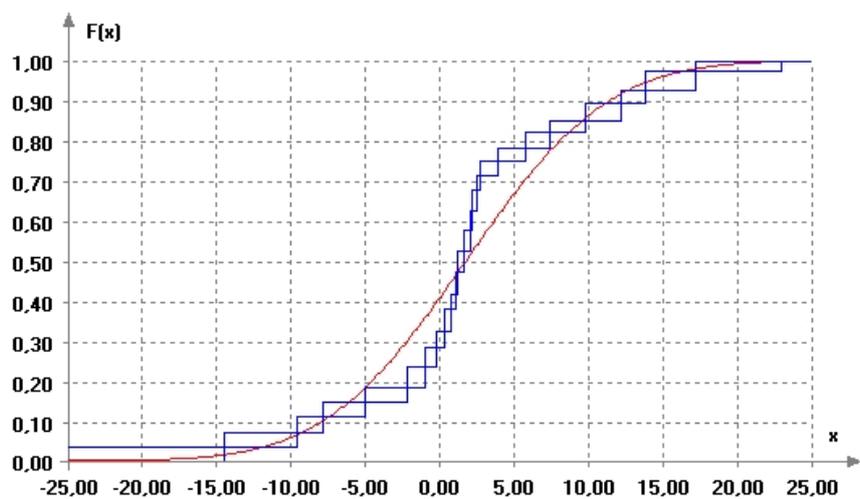


Рис.9. Эмпирическая функция распределения ГТП УЧП США по выборке из отрасли 15 (Non-metallic mineral products) и функция распределения нормального закона с параметрами $\mu = 1,7803$, $\sigma = 7,5092$

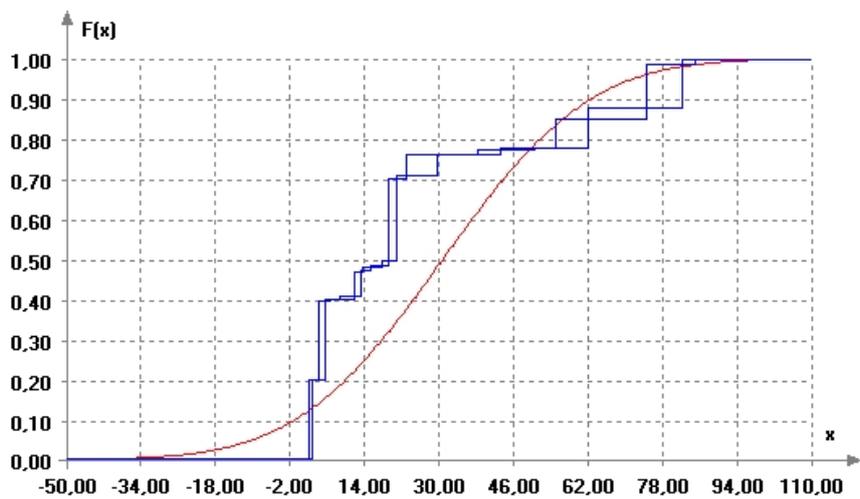


Рис.10. Эмпирическая функция распределения ГТП УЧП США по выборке из отрасли 22 (Electronic valves and tubes) и функция распределения нормального закона с параметрами $\mu = 30,9127$, $\sigma = 24,7828$

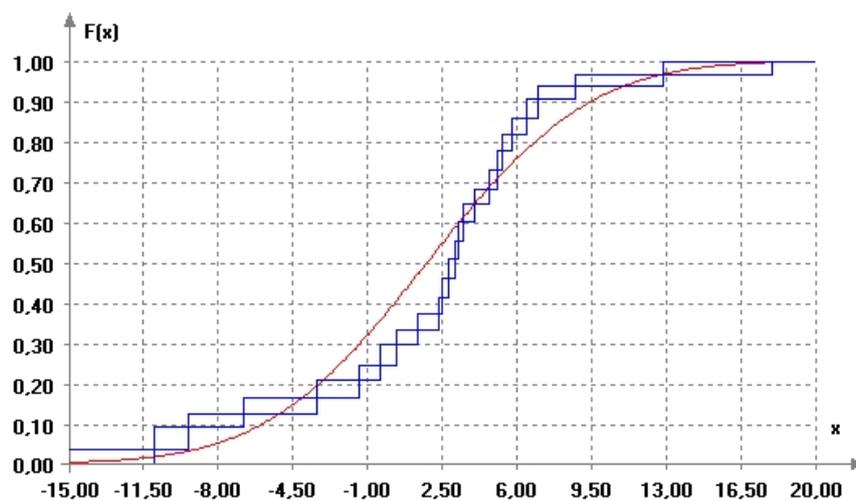


Рис.11. Эмпирическая функция распределения ГТП УЧП США по выборке из отрасли 31 (Furniture, miscellaneous manufacturing; recycling) и функция распределения нормального закона с параметрами $\mu = 1,8219$, $\sigma = 5,9879$

Построение модели для полной выборки ГТП УЧП всех отраслей

Если поставить целью построение достаточно хорошей модели для описания ГТП УЧП всех отраслей, можно пойти по следующему пути: подобрать для описания ГТП УЧП каждой из 56 отраслей наиболее подходящую параметрическую модель, а для описания всех отраслей использовать модель в виде смеси. Однако такая вероятностная модель будет очень громоздкой и неудобной для использования.

Можно разбить множество отраслей на группы, в которых ГТП УЧП имеют достаточно близкие законы, подобрав для описания групп свои параметрические модели, а общую модель построить уже в виде более простой смеси. Такой путь кажется более предпочтительным.

В таблице 1 рассматриваемые отрасли упорядочены по величине параметра «сдвига» аппроксимирующих нормальных зависимостей.

Анализ характеристик нормальных законов (таблица 1) позволяет предположить, что множество отраслей можно разбить (это не

строго) на 6 групп, в которых характеристики сдвига и масштаба «относительно близки». Выделенные группы отраслей приведены в таблице 2. Эмпирические функции распределения ГТП УЧП для выборок, соответствующих полученным (выделенным) группам показаны на рисунке 12.

Таблица 1. Ранжирование отраслей по величине параметра «сдвига» нормальных законов

Номер отрасли	Название отрасли	Значение параметра сдвига	Значение параметра масштаба
8	Leather and footwear	-2,888	7,697
7	Clothing	-1,458	4,685
16	Basic metals	-1,387	9,023
18	Mechanical engineering	-0,923	7,835
30	Railroad equipment and transport equipment nec	-0,885	13,748
11	Printing & publishing	-0,367	3,607
20	Insulated wire	-0,316	8,830
28	Building and repairing of ships and boats	-0,112	12,664
10	Pulp, paper & paper products	0,011	7,085
21	Other electrical machinery and apparatus nec	0,491	5,936
6	Textiles	0,536	5,981
4	Mining and quarrying	0,537	5,566
25	Scientific instruments	0,659	5,263
33	Construction	0,773	5,627
9	Wood & products of wood and cork	0,843	8,023
49	Research and development	0,847	11,661
52	Public administration and defence; compulsory social security	0,947	1,184

5	Food, drink & tobacco	1,192	4,713
17	Fabricated metal products	1,198	6,116
44	Insurance and pension funding, except compulsory social security	1,393	4,583
56	Private households with employed persons	1,424	6,781
2	Forestry	1,566	11,657
53	Education	1,574	1,026
26	Other instruments	1,652	5,064
15	Non-metallic mineral products	1,780	7,509
31	Furniture, miscellaneous manufacturing; recycling	1,822	5,988
29	Aircraft and spacecraft	1,931	12,496
38	Inland transport	2,011	3,731
54	Health and social work	2,047	1,723
12	Mineral oil refining, coke & nuclear fuel	2,184	27,170
37	Hotels & catering	2,341	2,671
13	Chemicals	2,390	5,098
46	Real estate activities	2,504	1,536
50	Legal, technical and advertising	2,573	3,356
3	Fishing	2,648	19,607
32	Electricity, gas and water supply	2,689	4,736
47	Renting of machinery and equipment	2,752	7,016
41	Supporting and auxiliary transport activities; activities of travel agencies	2,971	4,166
27	Motor vehicles	3,065	13,445
39	Water transport	3,096	7,999

55	Other community, social and personal services	3,430	2,863
42	Communications	3,432	4,009
43	Financial intermediation, except insurance and pension funding	3,451	5,597
34	Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel	3,718	6,652
36	Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods	3,825	3,778
24	Radio and television receivers	3,861	26,084
14	Rubber & plastics	3,885	5,509
1	Agriculture	3,990	13,084
35	Wholesale trade and commission trade, except of motor vehicles and motorcycles	5,163	4,890
51	Other business activities, nec	5,947	8,059
23	Telecommunication equipment	6,581	13,961
40	Air transport	7,784	8,295
48	Computer and related activities	9,452	5,979
45	Activities auxiliary to financial intermediation	11,869	18,104
22	Electronic valves and tubes	30,913	24,783
19	Office machinery	67,733	23,662

Таблица 2. Разбиение множества отраслей на группы с «относительно близкими» законами распределения

Номер группы	Номер отрасли	Название отрасли	Значение параметра сдвига	Значение параметра масштаба
1	30	Railroad equipment and transport equipment nec	-0,885	13,748
	28	Building and repairing of ships and boats	-0,112	12,664
	49	Research and development	0,847	11,661
2	8	Leather and footwear	-2,888	7,697
	7	Clothing	-1,458	4,685
	16	Basic metals	-1,387	9,023
	18	Mechanical engineering	-0,923	7,835
	11	Printing & publishing	-0,367	3,607
	20	Insulated wire	-0,316	8,830
	10	Pulp, paper & paper products	0,011	7,085
	21	Other electrical machinery and apparatus nec	0,491	5,936
	6	Textiles	0,536	5,981
	4	Mining and quarrying	0,537	5,566
	25	Scientific instruments	0,659	5,263
	33	Construction	0,773	5,627
	9	Wood & products of wood and cork	0,843	8,023
	52	Public administration and defence; compulsory social security	0,947	1,184

3	5	Food, drink & tobacco	1,192	4,713
	17	Fabricated metal products	1,198	6,116
	44	Insurance and pension funding, except compulsory social security	1,393	4,583
	56	Private households with employed persons	1,424	6,781
	53	Education	1,574	1,026
	26	Other instruments	1,652	5,064
	15	Non-metallic mineral products	1,780	7,509
	31	Furniture, miscellaneous manufacturing; recycling	1,822	5,988
	38	Inland transport	2,011	3,731
	54	Health and social work	2,047	1,723
	37	Hotels & catering	2,341	2,671
	13	Chemicals	2,390	5,098
	46	Real estate activities	2,504	1,536
	50	Legal, technical and advertising	2,573	3,356
	32	Electricity, gas and water supply	2,689	4,736
	47	Renting of machinery and equipment	2,752	7,016
	41	Supporting and auxiliary transport activities; activities of travel agencies	2,971	4,166

	39	Water transport	3,096	7,999
	55	Other community, social and personal services	3,430	2,863
	42	Communications	3,432	4,009
	43	Financial intermediation, except insurance and pension funding	3,451	5,597
	34	Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel	3,718	6,652
	36	Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods	3,825	3,778
	14	Rubber & plastics	3,885	5,509
4	2	Forestry	1,566	11,657
	29	Aircraft and spacecraft	1,931	12,496
	12	Mineral oil refining, coke & nuclear fuel	2,184	27,170
	3	Fishing	2,648	19,607
	27	Motor vehicles	3,065	13,445
	24	Radio and television receivers	3,861	26,084
5	35	Wholesale trade and commission trade, except of motor vehicles and motorcycles	5,163	4,890

	51	Other business activities, nec	5,947	8,059
	40	Air transport	7,784	8,295
	48	Computer and related activities	9,452	5,979
6	1	Agriculture	3,990	13,084
	23	Telecommunications equipment	6,581	13,961
	45	Activities auxiliary to financial intermediation	11,869	18,104
	22	Electronic valves and tubes	30,913	24,783
	19	Office machinery	67,733	23,662

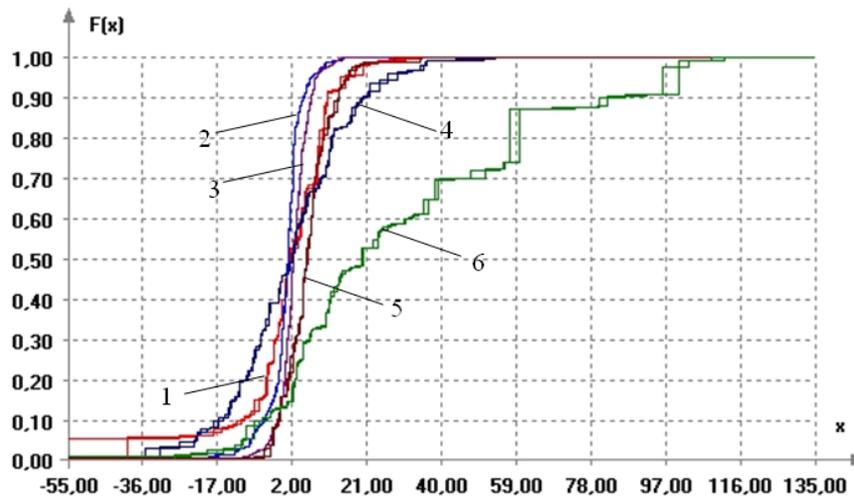


Рис.12. Эмпирические функции распределения ГТП УЧП США, соответствующие 6 группам

Проанализируем полученные выборки для групп ГТП УЧП, подбирая для них теоретические распределения вероятностей, а затем построим для полной выборки ГТП УЧП модель в виде смеси теоретических законов, построенных для отдельных групп.

На рис. 13–18 приведены функции распределения, «наилучшим» образом аппроксимирующие эмпирические распределения, соответствующие 6 выделенным группам отраслей. Отметим, что в каждом случае использование модели нормального закона оказывалось менее подходящим.

Для описания первой группы наиболее подходящей моделью оказалось (см. рис. 13) распределение Су–Джонсона с плотностью

$$f(x) = \frac{\theta_1}{\sqrt{2\pi} \sqrt{x - \theta_3}^2 + \theta_2^2} \exp \left\{ -\frac{1}{2} \left[\theta_0 + \theta_1 \ln \left\{ \frac{x - \theta_3}{\theta_2} + \sqrt{\left(\frac{x - \theta_3}{\theta_2} \right)^2 + 1} \right\} \right]^2 \right\},$$

для описания второй группы (см. рис. 14) – двустороннее экспоненциальное распределение с плотностью

$$f(x) = \frac{\theta_0}{2\theta_1 \Gamma(1/\theta_0)} \exp \left\{ -\left(\frac{|x - \theta_2|}{\theta_1} \right)^{\theta_0} \right\},$$

для описания третьей (см. рис. 15) – распределение Лапласа с плотностью

$$f(x) = \frac{1}{2\theta_0} e^{-|x - \theta_1|/\theta_0},$$

для описания четвертой (см. рис. 16) – двустороннее экспоненциальное распределение, для описания пятой (см. рис. 17) – распределение Су–Джонсона, для описания шестой группы (см. рис. 18) – Г–распределение с плотностью

$$f(x) = \frac{\theta_0}{\Gamma(\theta_1)} \left(\frac{x - \theta_3}{\theta_2} \right)^{\theta_0 \theta_1 - 1} \exp \left\{ -\left(\frac{x - \theta_3}{\theta_2} \right)^{\theta_0} \right\}.$$

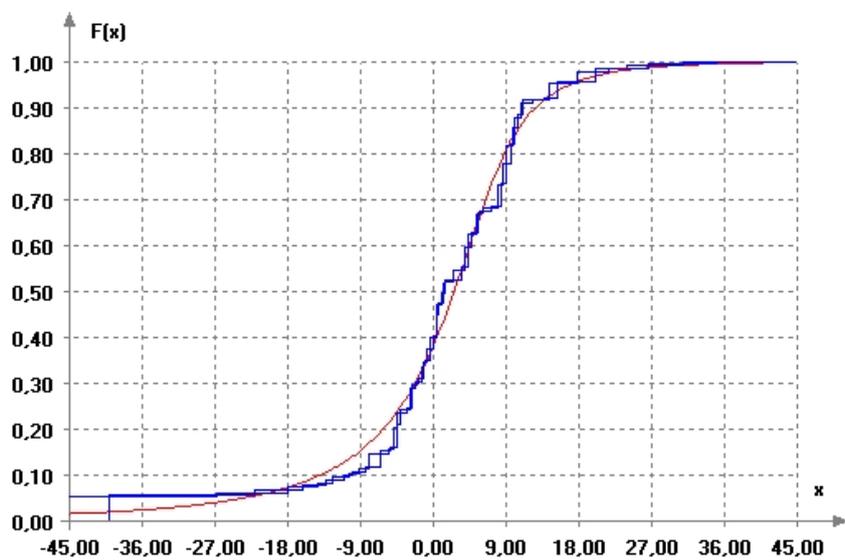


Рис. 13. Эмпирическая функция распределения ГТП УЧП США, соответствующая 1-й группе, и функция распределения Су–Джонсона с оценками параметров $\theta_0 = 0,4329$, $\theta_1 = 0,9829$, $\theta_2 = 7,0735$, $\theta_3 = 5,9275$

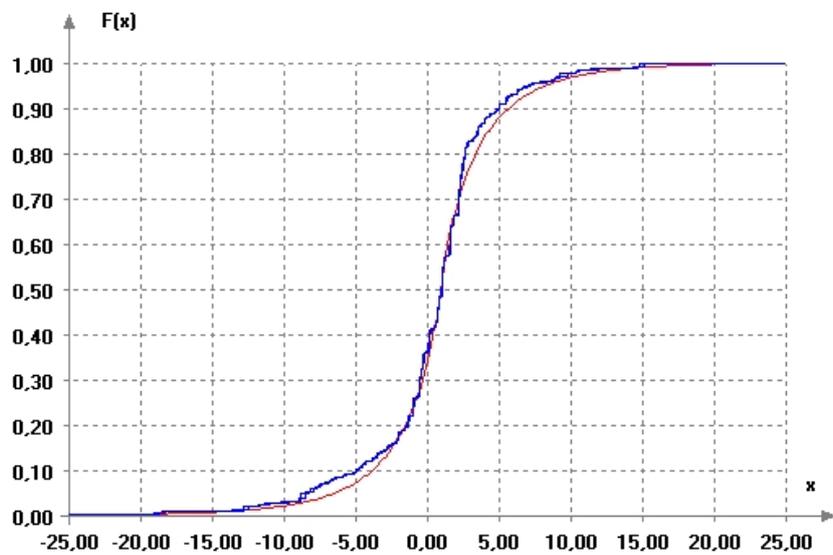


Рис. 14. Эмпирическая функция распределения ГТП УЧП США, соответствующая 2-й группе, и функция двустороннего экспоненциального распределения с оценками параметров $\theta_0 = 0,7054$, $\theta_1 = 1,5473$, $\theta_2 = 0,9700$.

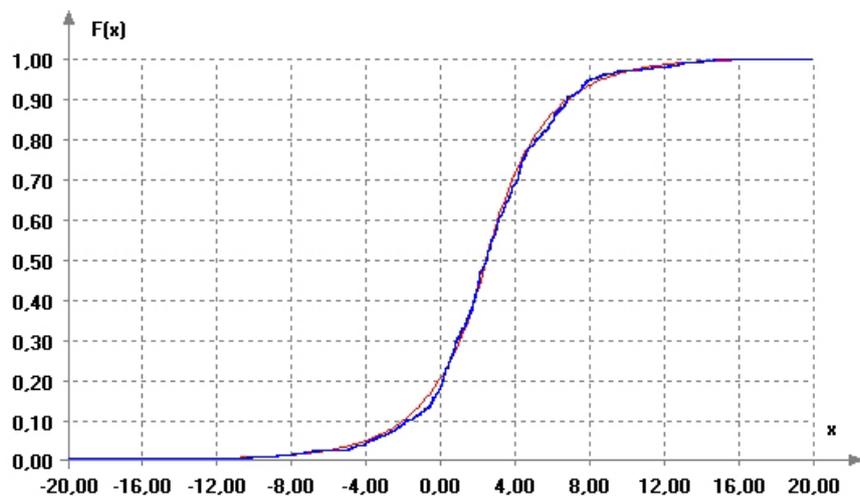


Рис. 15. Эмпирическая функция распределения ГТП УЧП США, соответствующая 3-й группе, и функция распределения Лапласа с оценками параметров $\theta_0 = 2,7328$, $\theta_1 = 2,4862$

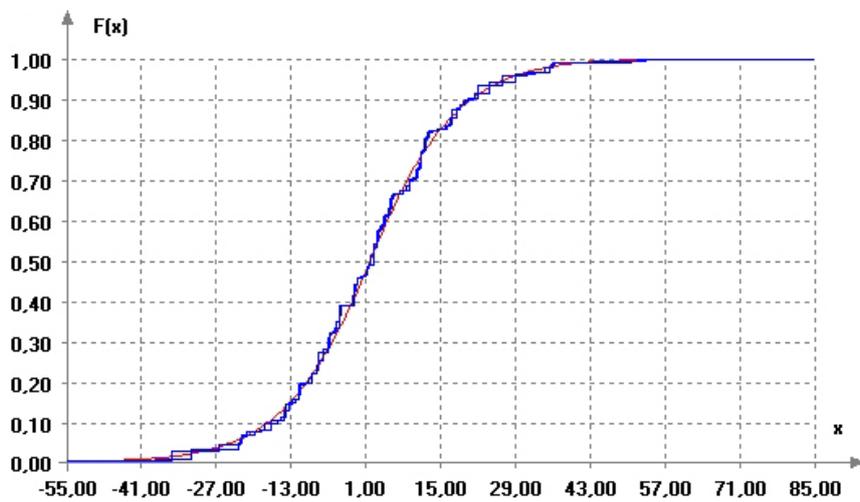


Рис. 16. Эмпирическая функция распределения ГТП УЧП США, соответствующая 4-й группе, и функция двустороннего экспоненциального распределения с оценками параметров $\theta_0 = 1,4175$, $\theta_1 = 17,1206$, $\theta_2 = 1,8659$

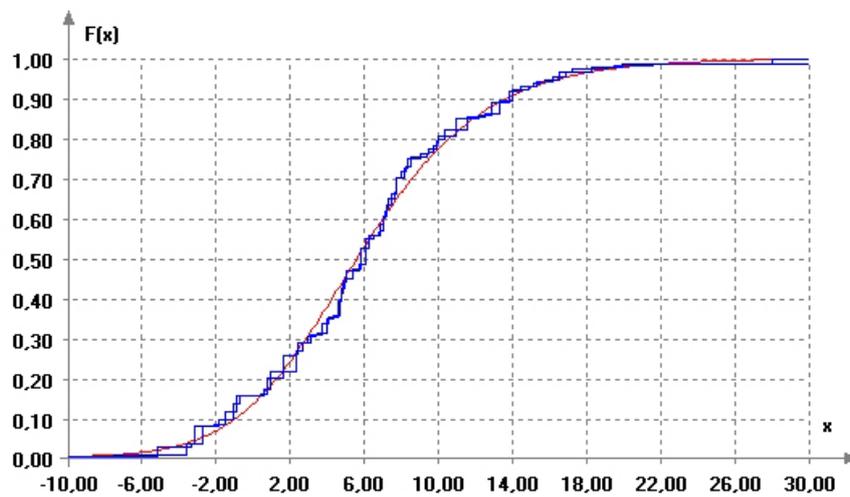


Рис. 17. Эмпирическая функция распределения ГТП УЧП США, соответствующая 5-й группе, и функция распределения Су-Джонсона с оценками параметров $\theta_0 = -1,1240$, $\theta_1 = 2,1779$, $\theta_2 = 10,3168$, $\theta_3 = -0,0579$

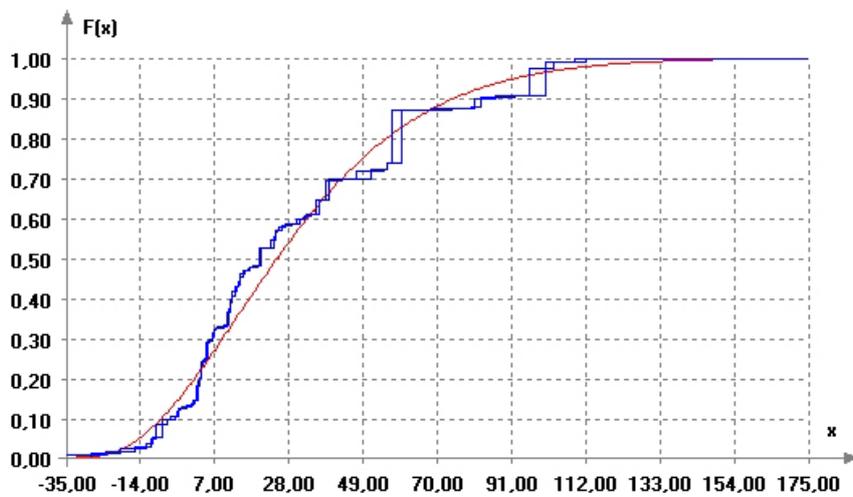


Рис. 18. Эмпирическая функция распределения ГТП УЧП США, соответствующая 6-й группе, и функция Γ -распределения с оценками параметров $\theta_0 = 2,2365$, $\theta_1 = 1,2574$, $\theta_2 = 33,9201$, $\theta_3 = -31,9409$

Полную выборку ГТП УЧП можно попытаться описать смесью законов распределения, описывающих составляющие выборки, с весами, пропорциональными объемам выборок, с плотностью

$$f(x) = 0,0088f_1(x) + 0,2168f_2(x) + 0,6009f_3(x) + 0,0273f_4(x) + \\ + 0,0851f_5(x) + 0,0609f_6(x),$$

где $f_i(x)$ – плотность распределения, описывающая выборку ГТП УЧП, соответствующую i -й группе отраслей.

В случае использования для описания полной выборки смеси, состоящей из логистического распределения с параметрами $\theta_0 = 6,3369$, $\theta_1 = 2,1318$ для первой выборки, двустороннего экспоненциального распределения с параметрами $\theta_0 = 3,1167$, $\theta_1 = 0,8566$, $\theta_2 = 0,9608$ для второй выборки, распределения Лапласа с параметрами $\theta_0 = 2,7328$, $\theta_1 = 2,4862$ для третьей выборки, двустороннего экспоненциального распределения с параметрами $\theta_0 = 17,1206$, $\theta_1 = 1,8659$, $\theta_2 = 1,4175$ для четвертой выборки, распределения Су-Джонсона с параметрами $\theta_0 = -1,1240$, $\theta_1 = 2,1779$, $\theta_2 = 10,3168$, $\theta_3 = -0,0579$ для пятой выборки и Г-распределения с параметрами $\theta_0 = 2,2365$, $\theta_1 = 1,2574$, $\theta_2 = 33,9201$, $\theta_3 = -31,9409$ для шестой выборки с весами, пропорциональными объемам выборок (объемам УЧП), получим картину, представленную на рис. 19.

Выводы. Рассмотренная попытка построения вероятностной модели для ГТП УЧП демонстрирует типичные особенности, связанные с задачами экономики (на макроуровне).

Во-первых. Статистические данные, связанные с процессами, регистрируемыми макроэкономической статистикой, характеризуются очень большими объемами накопленной информации, хранимой в сжатом виде группированных или интервальных данных. Причем границы интервалов, как правило, неизвестны.

Во-вторых. Если пренебречь фактом группирования данных и при дальнейшем статистическом анализе проводить повторное группирование с последующим построением вероятностной модели, это пренебрежение почти наверное приведет к смещенности построенной модели закона.

В-третьих. Очень редко, как и во многих других приложениях прикладной математической статистики, нормальный закон оказывается наилучшей моделью для наблюдаемых случайных величин.

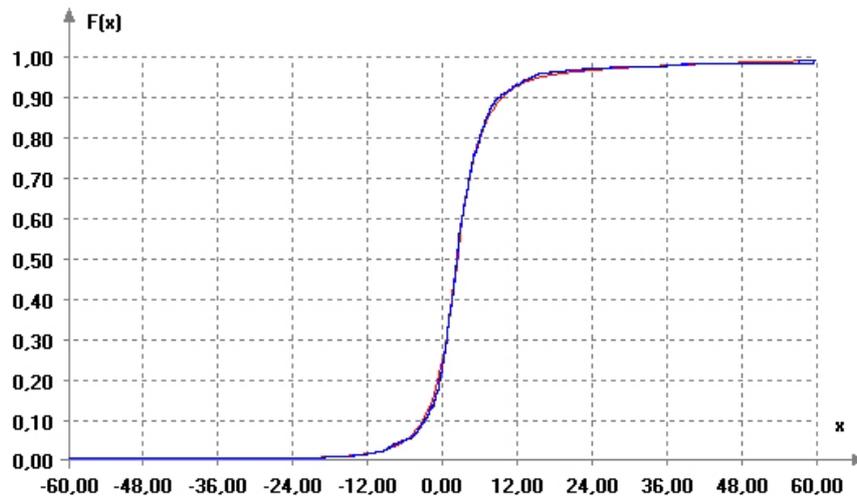


Рис.19. Эмпирическая функция распределения ДТП США по полной выборке и функция распределения смеси 6 законов

В-четвертых. По-видимому, законы распределения ДТП США зависят от отраслей и, возможно, от объемов ОПФ.

В-пятых. Модели в виде смесей параметрических законов распределения могут достаточно хорошо описывать наблюдаемые процессы.

И последнее. Окончательные выводы об адекватности построенной вероятностной модели, о том, насколько хорошо она объясняет случайные отклонения от имеющих место детерминированных законов, описывающих экономические процессы, остаются за экономической теорией.

Литература

- [1] Вальтух К.К. Воспроизводство и ценообразование. Теория. Исследование системной статистики. Том 1. Динамика основных производственных фондов. М.: «Янус-К», 2009. – 788 с.

- [2] Куллдорф Г. Введение в теорию оценивания по группированным и частично группированным выборкам. – М.: Наука, 1966. – 176 с.
- [3] Денисов В.И., Лемешко Б.Ю., Цой Е.Б. Оптимальное группирование, оценка параметров и планирование регрессионных экспериментов. В 2-х ч. / Новосиб. гос. техн. ун-т. – Новосибирск, 1993. – 347 с.
- [4] Лемешко Б.Ю., Постовалов С.Н. О решении задач статистического анализа интервальных наблюдений // Вычислительные технологии. – 1997. – Т.2. – № 1. – С. 28–36.
- [5] Лемешко Б.Ю., Постовалов С.Н. Об оценивании параметров распределений по интервальным наблюдениям // Вычислительные технологии. – 1998. – Т.3. – № 2. – С. 31–38.
- [6] Лемешко Б.Ю. Об оценивании параметров распределений по группированным наблюдениям // Вопросы кибернетики. – М., 1977. – Вып. 30. – С. 80–96.
- [7] Лемешко Б.Ю. Робастные методы оценивания и отбраковка аномальных измерений // Заводская лаборатория. – 1997. – Т.63. – № 5. – С. 43–49.