Факторный анализ

© Наследов А. Д, 2012

Пример факторного анализа

Пример 1.1. Влияние внутреннего валового продукта (ВВП) на социально-демографические показатели (европейские страны, 2008 г.)

Переменные:

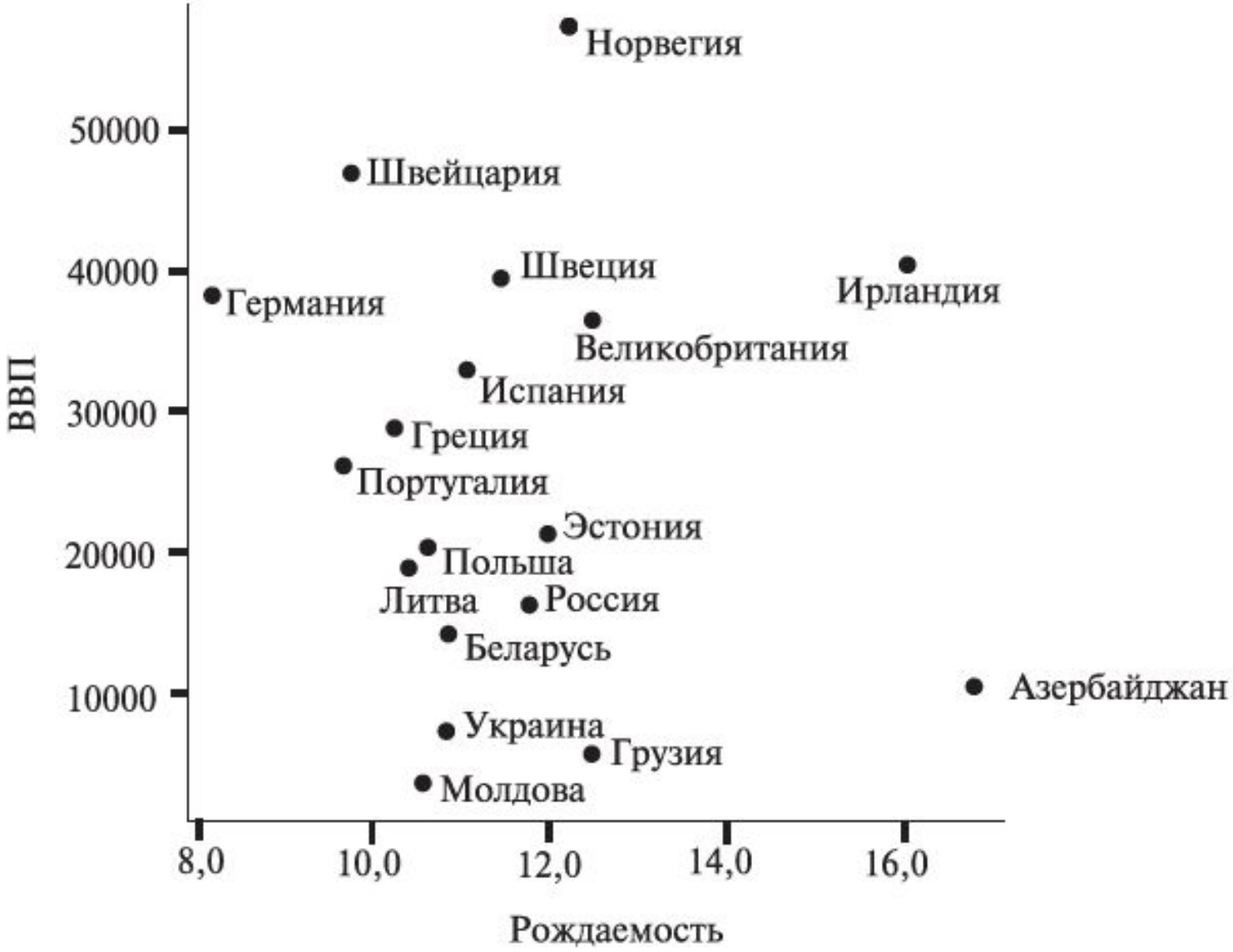
1. Медианный возраст населения страны.
2. Рождаемость (число родившихся на 1000 жителей).
3. Смертность (число умерших на 1000 жителей).
4. Естественный прирост (разность рождаемости и смертности на 1000 жителей).
5. Детская смертность (до 1 года на 1000 живорожденных).
6. Ожидаемая продолжительность жизни мужчин при рождении.
7. Ожидаемая продолжительность жизни женщин при рождении.
8. Валовой внутренний продукт на душу населения.

1. *Испания*
2. ***Италия***
3. ***Латвия***
4. ***Литва***
5. ***Македония***
6. ***Молдова***
7. ***Нидерданды***
8. ***Норвегия***
9. ***Польша***
10. ***Португалия***
11. ***Россия***
12. ***Румыния***
13. ***Сербия***
14. ***Словакия***
15. ***Словения***
16. ***Украина***
17. ***Финляндия***
18. ***Франция***
19. ***Хорватия***
20. ***Чехия***
21. ***Швейцария***
22. ***Швеция***
23. ***Эстония***

Таблица 1.2

Матрица корреляций между социально-демографическими показателями

Переменная	1	2	3	4	5	6	7	8
1. Медианный возраст	1	**-0,74**	0,26	**-0,55**	**-0,78**	0,39	0,53	0,39
2. Рождаемость		1	-0,46	**0,81**	0,50	-0,01	-0,08	0,08
3. Смертность			1	**-0,90**	-0,04	**-0,75**	**-0,63**	**-0,55**
4. Естественный прирост				1	0,50	0,38	0,27	0,40
5. Детская смертность					1	-0,49	**-0,63**	**-0,58**
6. Ожидаемая продолжительность жизни мужчин						1	**0,92**	**0,77**
7. Ожидаемая продолжительность жизни женщин							1	**0,83**
8. ВВП								1


Норвегия
50000
Швейцария
40000
Швеция
Ирландия
Германия
Великобритания
Испания
ВВП
30000
Греция
Португалия
Эстония
20000
Польша
Литва
Россия
Беларусь
10000
Азербайджан
Украина
Грузия
Молдова
8,0
10,0
12,0
14,0
16,0
Рождаемость

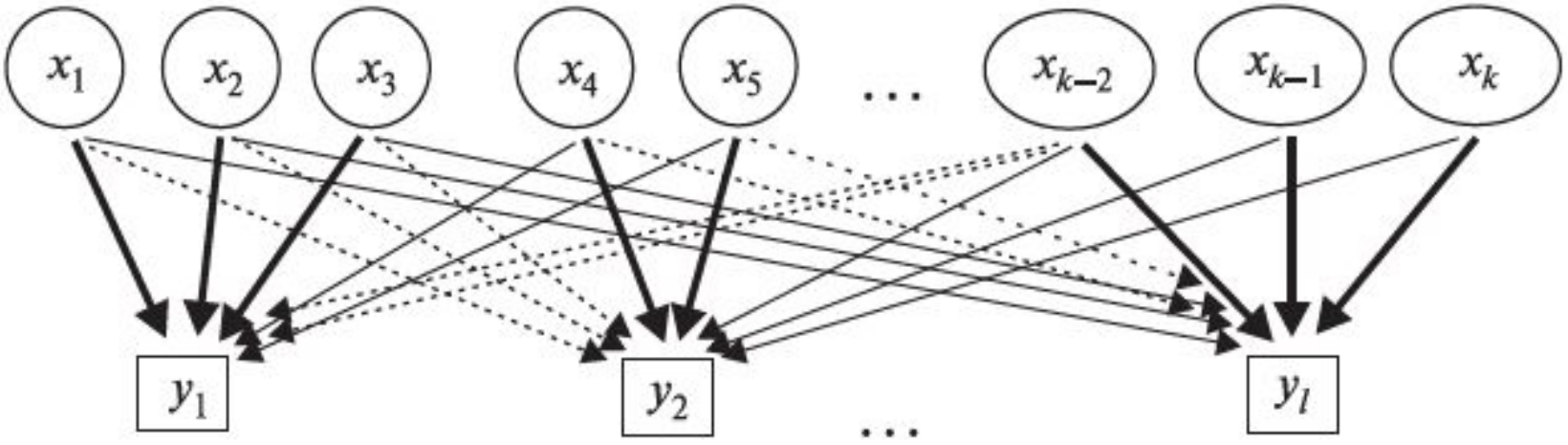


Рис. 2.6. Модель метода главных компонент:

— значительный вклад; — вклад средней величины;
--- незначительный вклад

Таблица 2.8

Собственные значения матрицы корреляций

Компонента	Объясненная дисперсия		
	λ_i	% дисперсии	накопленный %
y_1	3,82	54,60	54,60
y_2	2,32	33,12	87,72
y_3	0,45	6,37	94,09
y_4	0,19	2,70	96,79
y_5	0,15	2,15	98,94
y_6	0,06	0,83	99,77
y_7	0,02	0,23	100,00
Сумма	7,00	100,00	

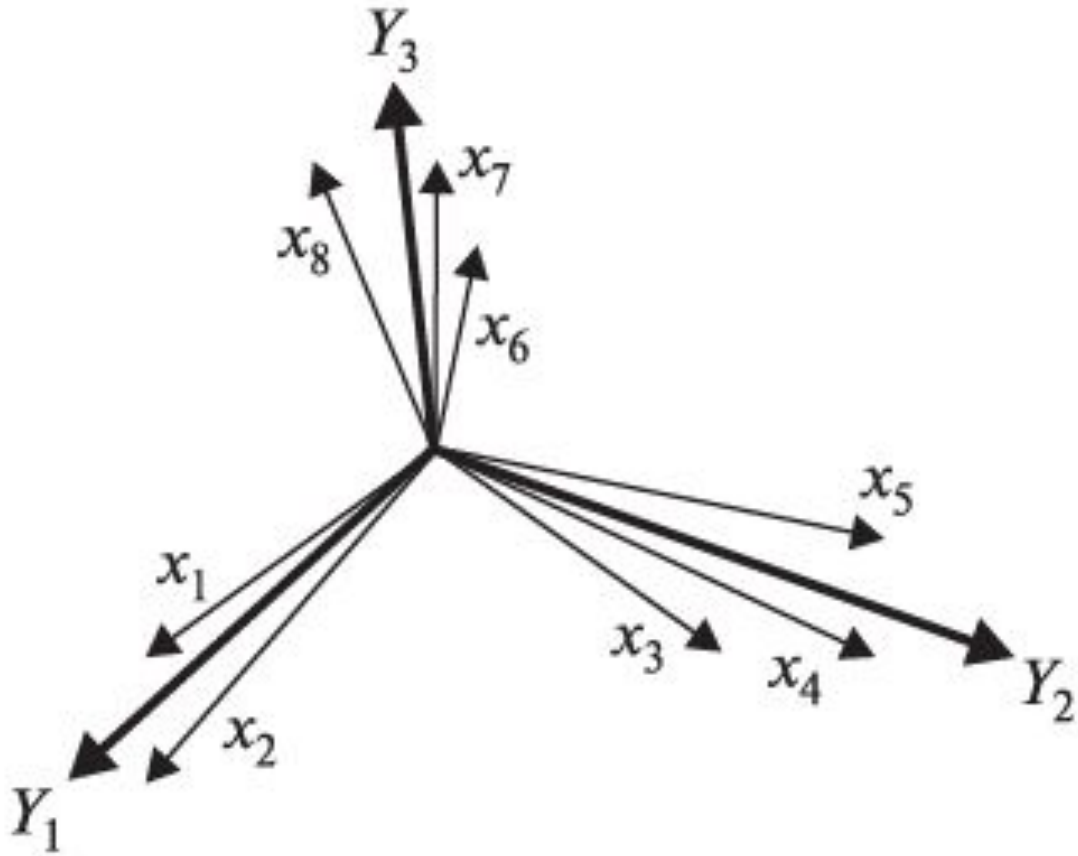


Рис. 2.3. Снижение размерности пространства переменных

Последовательность ФА

- Выбор исходных данных
- Вычисление корреляционной матрицы
- АГК и оценка числа факторов по графику собственных значений
- Выбор и применение метода ФА
- Вращение факторов
- Интерпретация факторов, принятие решения о качестве факторной структуры
- (Вычисление факторных оценок)

Анализ главных компонент (МГК) и ФА

МГК в матричной форме:

$$R = A \cdot A',$$

кол-во компонент = числу переменных

Факторный анализ:

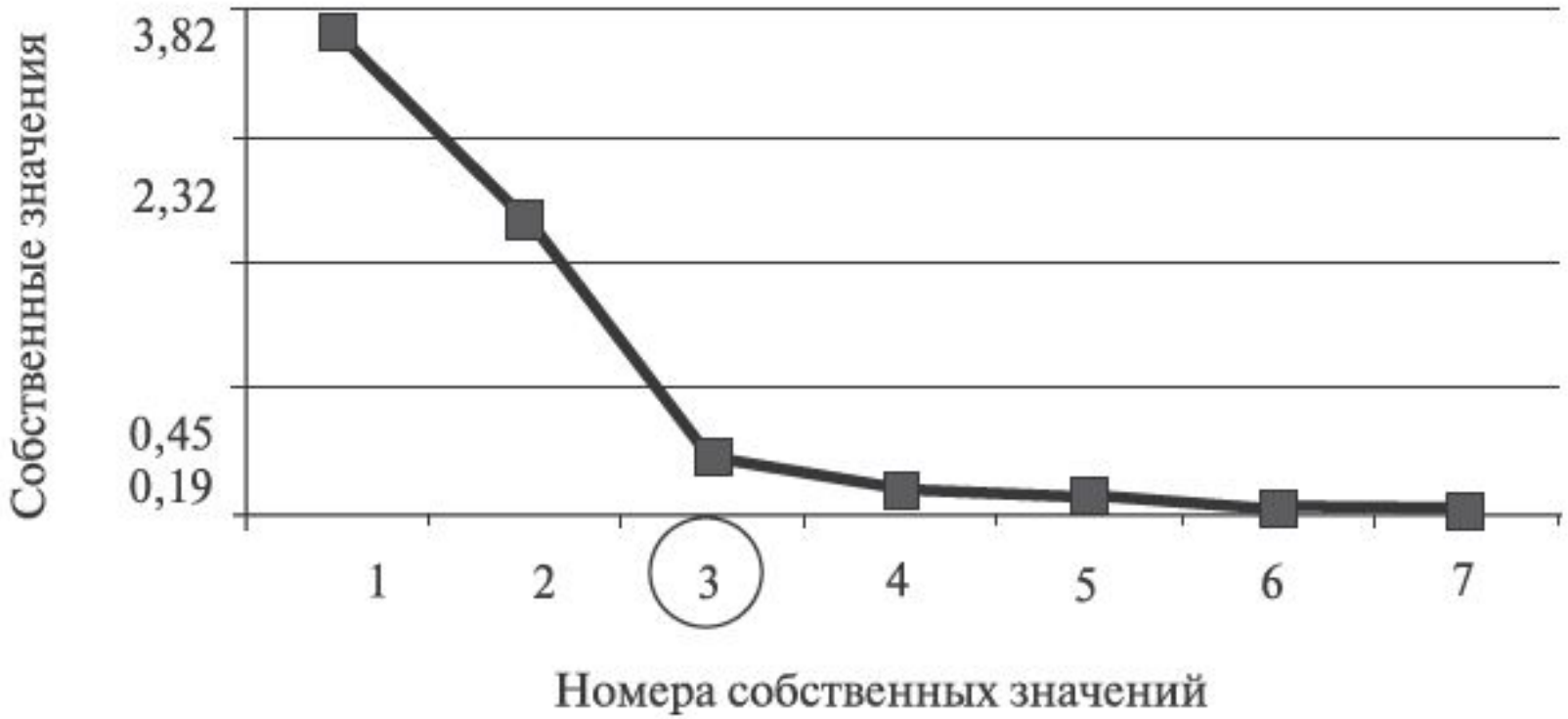
$$\hat{R} = A \cdot A' \text{ при условии } \hat{R} \to R$$

кол-во факторов (*M*) существенно меньше кол-ва переменных (*P*)

© *Наследов А. Д, 2012*

График собственных значений

Критерий числа факторов:
Р.Кеттелла (scree-test), «каменистой осыпи»



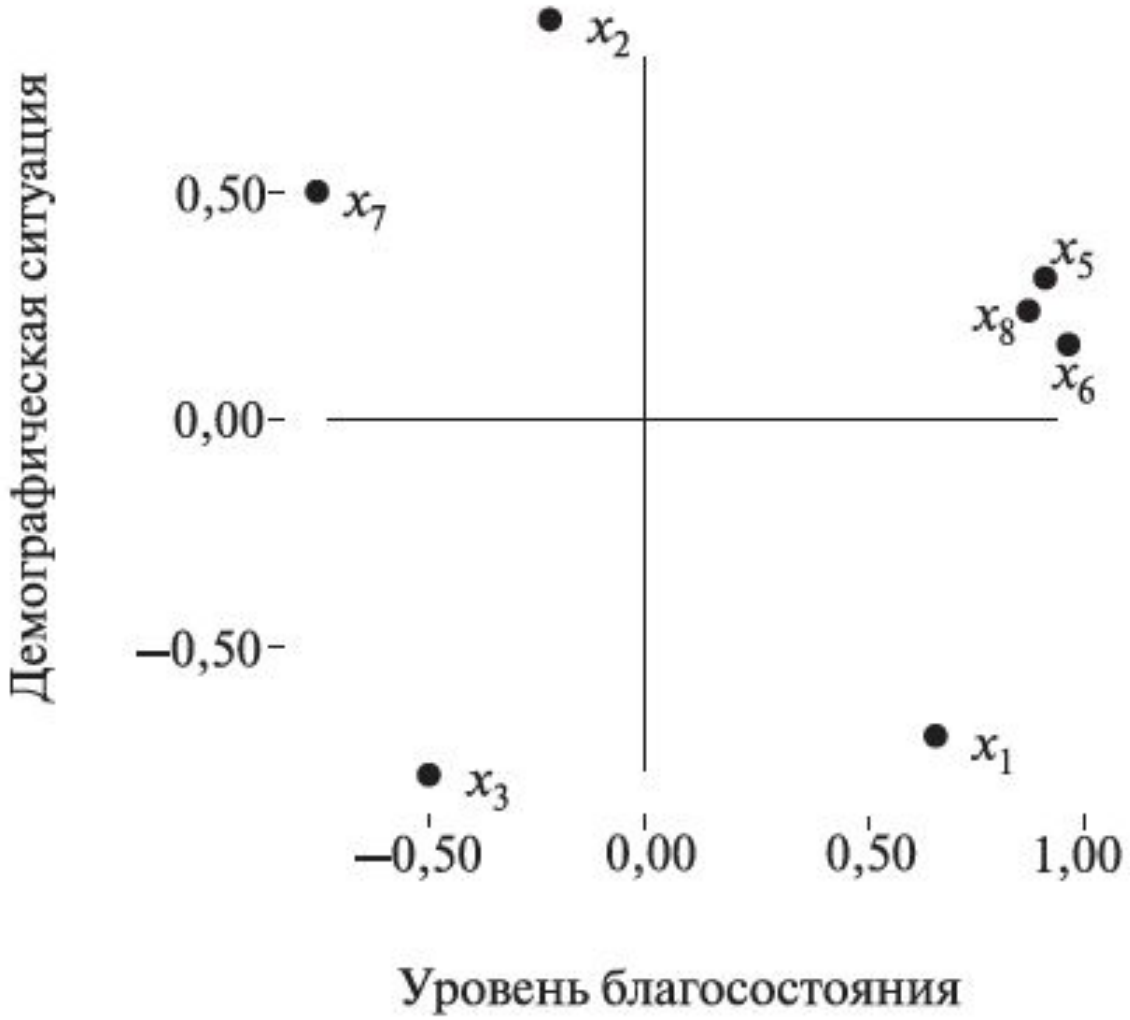


Рис. 2.10. График факторных нагрузок

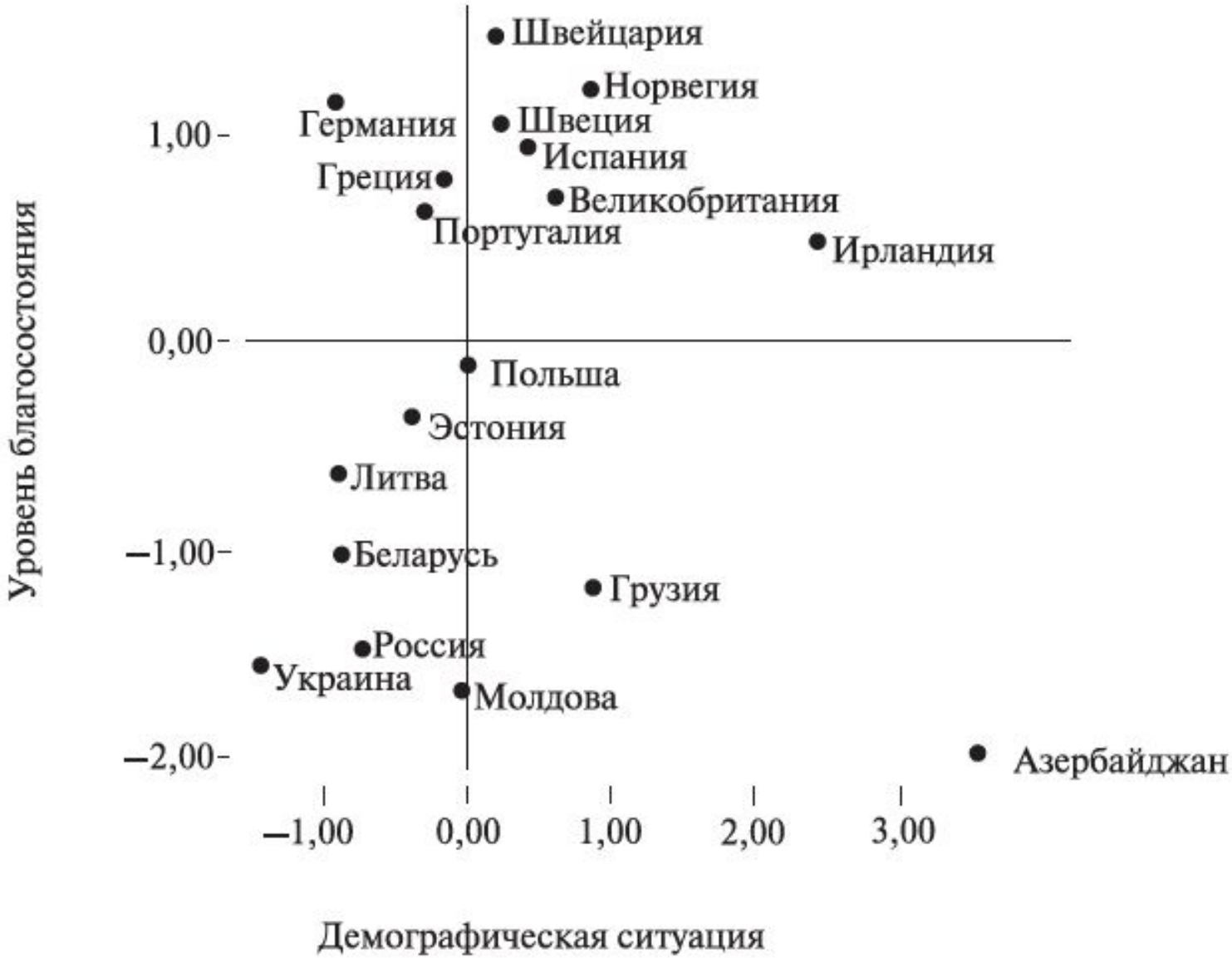


Рис. 2.11. Европейские страны в пространстве главных компонент

Проблема оценки значений факторов

Факторные оценки – новые переменные, значения факторов для объектов (случаев). Переход от *P* исходных переменных к *M* новым переменным (факторам) с минимальными потерями исходной

$$f_{ik} = \sum_{j=1}^{P} \beta_{i,j} x_{jk} = \beta_{1i} x_{1k} + \beta_{2i} x_{2k} + \ldots + \beta_{Pi} x_{Pk}$$

© Наследов А. Д, 2012

Классические проблемы ФА

- Проблема числа факторов
- Проблема общности и метода ФА
- Проблема вращения и интерпретации
- Проблема факторных оценок

© Наследов А. Д, 2012

Пример интерпретации факторов

Фактор 1 имеет наибольший вес или наибольшую информативность (27%). Его положительный полюс определяется положительными полюсами переменных 4 (независимый), 8 (спокойный), 10 (невозмутимый) и отрицательным полюсом переменной 7 (расслабленный). Отрицательный полюс этого фактора определяется про-

© Наследов А. Д, 2012