ПРИМЕНЕНИЕ КОНВЕРГЕНТНОЙ ТЕХНОЛОГИИ ДЛЯ РЕШЕНИЯ ЗАДАЧ ПО ОЦЕНКЕ ЧЕЛОВЕЧЕСКОГО КАПИТАЛА НА ПРИМЕРЕ МНОГОФАКТОРНОЙ МОДЕЛИ МИНЦЕРА[13]

Людмила Малахова,
Научный сотрудник,
Институт математики и информатики им. "Владимира Андрунакевича",
Республика Молдова
E-mail: lmalahov@gmail.com

Татьяна Колесникова,
Конференциар исследователь, к.э.н.,
Национальный Институт Экономических Исследований, Республика Молдова
E-mail: ctania@gmail.com

Александр Колесников,
Конференциар исследователь, к.ф-м.н
Институт математики и информатики им. "Владимира Андрунакевича",
Республика Молдова
E-mail: acolesnicov@gmx.com

Summary: *The paper describes the application of convergent technology for resolving the problems of assessing the human capital using the extended model of D. Mincer as an example. The software package was developed on the XWiki platform (Java, Velocity, Groovy, generation of dynamic pages, etc.) using libraries of statistics and processing tabular data from the Python language.*

---

[13] Данная статья разработана по результатам научно-исследовательского проекта *„Innovative approaches to applied computations and software development for gender equality regulation on labour market”* (2018-2020 гг.) в рамках совместной программы STCU-ASM и зарегистрированного в Государственном реестре проектов в области науки и инноваций Республики Молдова под номером 18.80013.0807.06.STCU.

Key words: *convergent technology, assessment of human capital, extended Mincerian equation, XWiki platform, Python language.*

Введение

Под конвергентной технологией мы понимаем процесс интеграции различных программных технологий, обеспечивающих решение экономических задач по типу «все-в-одном», что предполагает только ввод пользователем данных, а их проверка и обработка производится программой автоматически, с выдачей результата в удобном для пользователя виде.

Одной из ключевых проблем концепции человеческого капитала, является оценка влияния, оказываемого на величину текущих доходов (заработной платы) различных форм человеческого капитала: продолжительность обучения, общий профессиональный опыт работы и др. Основоположником данного подхода является Дж. Минцер [1]. Данный подход базируется на количественной оценке параметров, так называемой, «производственной функции заработков».

Многофакторная модель описывает зависимость заработков человека от уровня его образования, трудового стажа, продолжительности отработанного времени, социально-демографических и других факторов.

Уравнение заработной платы Дж. Минцера, признанное традиционной оценкой выгодности инвестиций в человеческий капитал, в процессе развития производственных отношений может видоизменяться, т.е. удаляются привычные параметры, появляются новые элементы. В каждом конкретном случае это будет новая модель.

Расчет регрессии Минцера возможен в ряде программных продуктов: Excel, EViews, SPSS, Stata[14] и др. Однако все они требуют определенных знаний в эконометрике и знаний самого продукта и его функций [2; 3; 4; 5 и др.]. В отличие от них, разработанный нами программный модуль упрощает подготовку данных, автоматически проверяет их корректность, а также автоматически выполняет расчеты, используя современные информационные технологии.

Данные могут предоставляться в нескольких форматах, включая CSV и таблицы Excel.

Программный комплекс разработан на платформе XWiki (Java, Velocity, Groovy, генерация динамических страниц и др.) с использованием библиотек статистики и обработки табличных данных языка Python.

Об используемом программном обеспечении

Программный комплекс состоит из ряда модулей: модуль импорта данных, модуль проверки данных, модуль обработки данных, модуль представления результатов. Для программирования большинства модулей используются стандартные возможности платформы XWiki. Мотивация данного выбора была нами опубликована [6; 7; 8; 9; 10; 11].

Кроме того, для программирования сложных экономических расчётов, в рамках модуля обработки данных, используются более удобные инструменты а именно Python 3[15] с библиотекой Pandas (Python Data Analysis) и другими необходимыми модулями. Возможность использования дополнительных инструментов разработки является одной из мощных функций подхода Web 2.0[16], обеспечиваемого платформой XWiki.

Среда XWiki предлагает ряд расширений, некоторые из них мы используем. Однако, этой функциональности не всегда достаточно для реализации задач. Например, возможно включение программы Python непосредственно в код динамической веб-страницы. Мы опробовали это расширение и выяснили, что его функциональность не полная.

Указанное расширение реализовано с использованием библиотек Jython, которые реализуют Python в среде Java. При таком подходе перевод с языка Jython выполняется не в коды интерпретатора Python, а в коды виртуальной машины Java для среды выполнения Java. Отсюда вытекают два ограничения: на версию языка и технологию библиотек и плагинов. Jython реализует более старую версию Python 2.7. На ней разработка Python 2 была остановлена). Таким образом, все нововведения из Python 3 являются недоступными.

Кроме этого, только те библиотеки Python, которые используют двоичный код интерпретатора Python, легко переводимый в код виртуальной машины Java, могут быть использованы.

Однако, большинство современных библиотек Python широко используют машинный код в целях оптимизации. Такие библиотеки в принципе не могут быть связаны с Jython. К сожалению, это касается самых необходимых и нужных в нашем случае библиотек, реализующих сложные вычисления, как например, метод наименьших квадратов. К ним относятся библиотеки NumPy, SciPy, Pandas и несколько других.

Наконец, с июля 2019 года поддержка библиотеки Pandas для Python 2 была прекращена.

---

[14] https://www.stata.com/
[15] https://www.python.org/
[16] http://wikireality.ru/wiki/%D0%92%D0%B5%D0%B1_2.0

Гибкость платформы XWiki помогла нам и в этом случае. Для сложных вычислений (регрессионного анализа и т. д.) использовалась самая современная стандартную реализация Python 3, а именно CPython, основанная на C.

Эта реализация имеет доступ к полному набору библиотек, загружаемых при необходимости из репозитория PyPI, и к инструментам обновления. Она специально устанавливается на сервере XWiki под Linux (в нашем случае), или под Windows.

Это дало нам полезную возможность отлаживать программы на Python независимо от среды XWiki.

На сервере XWiki была установлена библиотека JPserve. Это реализация инструментов вызова Python из Java-программы. Эта библиотека состоит из сервера, написанного на Python, и клиента, реализованного на Java. Будучи модулем Python, серверная часть устанавливается обычным способом: с помощью команды pip из репозитория PyPI.

Клиентская библиотека добавляется (копируется) в корневой каталог сервера веб-приложений (в нашем случае Apache Tomcat) вместе с другими архивами Java, которые совместно реализуют платформу XWiki и, таким образом, становятся частью XWiki.

Для запуска программы экономичных вычислений на Python и включения результатов вычислений в динамическую страницу XWiki в код этой страницы используется макрос Groovy. Groovy есть расширение Java. Фактически, таким образом, Java-код включён в страницу.

В этом случае все, что выводится командами print() или println(), считается кодом XWiki (расширенный код HTML) и отображается браузером. Очень важно, что, если на странице несколько таких фрагментов на разных языках программирования, то все они используют общий пул переменных.

В коде Groovy вызывается клиентская программа, одним из параметров которой является имя программы Python, которая должна быть выполнена. Клиент Java (то есть XWiki) отправляет запрос на выполнение программы на порт, выделенный для этой цели, серверу Python, который принимает запрос и запускает указанную python-программу.

Результат должен быть присвоен переменной _result_ и возвращен клиенту (в программе Java) в стандартной кодировке JSON. Превращение строки JSON в значения переменных XWiki или в код HTML выполняется отдельным макросом, написанным на стандартном для XWiki расширении Java - Velocity.

Пример многофакторной модели Дж. Минцера

Функция, показывающая зависимость индивидуальных заработков от трудового стажа или возраста, а также других переменных, получила название «уравнение Минцера» [1].

В нашем случае модель формирования заработков работников имеет следующий вид:

$$\ln w = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \beta_5 x_2 + \beta_6 x_2^2 + \beta_7 x_2 + \beta_8 x_5 + \beta_9 x_6$$

где $\ln w$ – натуральный логарифм реальных заработков;

β_1–β_9 – коэффициенты (веса переменных и их квадратов),

β_0 – постоянное смещение (intercept);

x_1 – общий трудовой стаж;

x_2 – трудовой стаж в отрасли;

x_3 – возраст;

x_4 – семейное положение (замужем/не замужем);

x_5 – количество детей;

x_6 – режим работы (полный/неполный рабочий день).

Входные данные

Используемые данные берутся из общей таблицы показателей. Никакая дополнительная обработка данных, а именно - логарифмирование, возведение в квадрат, сортировка, выборка и т.д. не требуются.

Данные могут предоставляться в нескольких форматах, включая CSV и таблицы Excel.

Мы использовали данные за 2018 год одного из крупных предприятий страны (более 6800 сотрудников, 30 различных показателей), предоставленные нам на условиях полной анонимности. Фрагмент данных показан на рис. 1.

1	sc_murc	clckadr_pol_r_5	virst	clckadr_merital_status_r_5	stajasigurat	stajramurai	staj_intre	childrens	TIPclccnas
12	12	F	54	1.Casatorit	37.320	37.320	37.320	0	женский
13	13	M	60	2.Necasatorit	41.508	10.008	10.008	2	мужской
14	14	F	53	2.Necasatorit	39.000	19.167	19.167	0	смешанный
15	15	F	43	2.Necasatorit	24.400	24.400	24.400	0	смешанный
16	16	F	53	2.Necasatorit	34.635	31.798	31.798	2	женский
17	17	F	51	2.Necasatorit	30.849	30.849	18.000	2	смешанный
18	18	F	37	2.Necasatorit	17.349	17.349	7.500	2	женский
19	19	F	38	2.Necasatorit	19.126	19.126	7.500	2	женский
20	20	M	40	1.Casatorit	15.715	15.715	15.715	2	мужской
21	21	M	34	1.Casatorit	16.161	11.083	3.661	2	мужской
22	22	M	47	1.Casatorit	25.680	8.513	8.513	2	мужской
23	23	M	23	2.Necasatorit	2.476	2.476	2.476	1	мужской
24	24	M	25	2.Necasatorit	6.212	6.212	6.212	0	мужской
25	25	M	51	2.Necasatorit	30.855	2.438	2.438	2	мужской
26	26	M	55	2.Necasatorit	11.941	11.941	2.417	0	мужской
27	27	M	30	1.Casatorit	6.712	6.710	6.710	1	мужской
28	28	M	41	1.Casatorit	9.005	9.005	9.005	2	мужской
29	29	M	26	2.Necasatorit	7.422	4.419	4.419	0	мужской
30	30	F	36	2.Necasatorit	18.417	18.417	7.500	2	женский
31	31	F	51	1.Casatorit	34.319	34.008	34.008	0	женский
32	32	M	40	1.Casatorit	17.056	17.056	17.056	3	мужской
33	33	M	49	2.Necasatorit	28.234	28.234	27.011	2	мужской
34	34	M	59	1.Casatorit	13.274	13.274	7.151	0	мужской
35	35	M	38	1.Casatorit	24.000	16.688	16.610	2	смешанный
36	36	M	50	1.Casatorit	2.685	2.685	2.685	2	смешанный
37	37	M	53	1.Casatorit	14.390	6.863	4.390	2	мужской
38	38	M	60	1.Casatorit	41.457	33.132	33.132	1	мужской
39	39	F	53	2.Necasatorit	33.836	27.116	8.419	0	смешанный
40	41	M	33	1.Casatorit	11.911	11.911	11.911	1	мужской
41	43	M	21	2.Necasatorit	0.454	0.452	0.452	0	мужской
42	44	M	26	1.Casatorit	9.253	9.253	2.056	1	мужской
43	46	M	50	2.Necasatorit	20.099	20.099	19.457	1	мужской
44	47	M	30	1.Casatorit	9.874	8.874	8.874	2	мужской
45	48	F	39	1.Casatorit	3.739	1.156	1.156	2	женский
46	49	M	30	2.Necasatorit	10.417	10.417	10.417	0	мужской
47	50	F	33	1.Casatorit	11.078	11.078	8.129	1	женский
48	51	F	55	1.Casatorit	37.530	19.454	19.454	0	женский
49	52	M	40	1.Casatorit	22.134	19.833	19.051	2	мужской
50	53	M	52	2.Necasatorit	11.651	11.651	11.651	2	мужской
51	54	M	38	1.Casatorit	17.495	17.495	17.495	2	мужской
52	55	F	49	1.Casatorit	28.056	28.056	23.694	0	мужской
53	56	F	43	1.Casatorit	29.000	19.126	16.935	1	смешанный
54	57	F	45	2.Necasatorit	26.274	1.022	1.022	3	женский
55	58	F	24	2.Necasatorit	5.333	5.333	5.333	0	женский
56	59	M	52	1.Casatorit	38.000	17.796	17.796	0	мужской
57	60	M	24	2.Necasatorit	3.500	3.500	3.500	0	мужской
58	61	M	36	1.Casatorit	14.747	6.997	6.997	2	мужской
59	63	M	31	1.Casatorit	13.247	13.247	13.247	2	мужской
60	64	F	31	2.Necasatorit	12.301	12.301	12.301	0	женский
61	65	F	52	1.Casatorit	30.371	30.371	30.371	1	мужской
62	66	F	47	1.Casatorit	33.000	23.642	22.460	0	смешанный
63	67	M	47	2.Necasatorit	23.755	23.755	7.478	1	смешанный
64	68	M	53	1.Casatorit	19.887	19.142	12.970	2	мужской
65	69	F	27	1.Casatorit	6.255	3.761	3.761	1	женский
66	70	M	24	2.Necasatorit	2.220	2.054	2.054	0	смешанный
67	71	M	48	1.Casatorit	26.290	26.290	25.417	0	смешанный

Рис. 1. Фрагмент таблицы исходных данных (Excel)

Источник: фрагмент базы данных предприятия

Описание алгоритма

Вычисления производятся отдельно по двум выборкам - мужчины и женщины, чтобы выявить разницу в коэффициентах уравнения Минцера. Выборка мужчин и женщин из общей таблицы производится программой с помощью сортировки.

Зависимость от переменных, включая логарифмирование и возведение в квадрат, задаётся формулой зависимости в стиле языка R.

Всё это указывается в аргументах стандартной программы регрессии, доступной в библиотеках языка Python 3. Результаты показаны на рис.2.

14. Mincer regression

Dependent variable: the natural logarithm of hourly wage

	M		F	
	Coef.	Std.err.	Coef.	Std.err.
Type of education (1-4)	0.1581	0.0037	0.2211	0.0056
Work experience	0.0093	0.0026	0.0065	0.0046
Squared work experience	-0.0001	0.0001	-0.0001	0.0001
Work experience in domain	0.0157	0.0023	0.0323	0.0035
Squared work experience in domain	-0.0002	0.0001	-0.0005	0.0001
Age	0.0036	0.0046	-0.0316	0.0089
Squared age	-0.0000	0.0001	0.0003	0.0001
Marital status (not married = 1)	0.0735	0.0097	0.0628	0.0135
Nr. of children	0.0058	0.0050	0.0173	0.0075
Work arrangement (full time = 1)	0.1541	0.0226	0.1136	0.0295
Intercept	2.6218	0.0843	3.0293	0.1621
Nr. of observations	4192		2623	
R-squared	0.3882		0.4338	
R-squared adjusted	0.3867		0.4316	
F statistics	265.2983		200.1296	
Probability > F	0.0000		0.0000	
Square root of MSE	0.3657		0.4455	

Рис. 2. Результаты расчета многофакторной модели Минцера

Источник: авторские расчеты

Выводы

Разработка программного обеспечения в рамках платформы XWiki с использованием внутренних возможностей Groovy, Velocity является достаточно удобной, в основном из-за ее высокой функциональной совместимости. При этом, использование мощных современных библиотек Python 3 позволило эффективно реализовать необходимый алгоритм и автоматзировать процесс реализации задачи по оценке человеческого капитала.

Таким образом, именно применение конвергентных технологий и техник, когда различные средства разработки, программирования, разные платформы и среды интегрируются в единый программный комплекс, позволяет обеспечивать наиболее высокий уровень автоматизации и абстрагирования для конечных пользователей, по сравнению с прямыми вычислениями в Excel или другом эконометрическом приложении.

Литература

1. Mincer J. Schooling, Experience and Earnings. New York: Columbia University Press for National Bureau of Economic Research, 1974. 167 p.
2. Liyousew G. Borga, Labor Economics with STATA, December 2, 2015, CERGE, EI http://home.cerge-ei.cz/munich/labor15/Resources/Excercises/6/ES6_slides.pdf
3. Ricardo Nicolás Pérez Truglia, Applied Econometrics using Stata, Harvard University - Department of Economics, Extremely early draft: March 2009, https://www.dphu.org/uploads/attachements/books/books_3316_0.pdf
4. Jean-Louis Brillet, Structural econometric modelling: Methodology and tools with applications under EViews. https://www.eviews.com/StructModel/structmodel.html
5. Landau Sabine, Brian S. Everitt. A handbook of statistical analyses using SPSS. A CRC Press Company Boca Raton London New York Washington, D.C., ISBN 1-58488-369-3, 2004 by Chapman & Hall/CRC Press LLC, http://www.fao.org/tempref/AG/Reserved/PPLPF/ftpOUT/Gianluca/stats/Statistics%20-%20A%20Handbook%20of%20Statistical%20Analyses%20using%20SPSS%20-%20Excellent%20!!!.pdf
6. Colesnicova T., Malahov L., Colesnicov A. Some applications of IT in gender studies. În: Economic growth in conditions of internationalization=Creşterea economică în condiţiile internaţionalizării: Conferinţa X ştiinţifico-practică internaţională din 15-16 octombrie 2015: (în 2 vol.). Ch.: INCE, 2015, vol. II, - 252 p. ISBN 978-9975-4185-2-2, p.201-204
7. Colesnicov A., Malahov L., Lucasenco E. On implementation platform for gender-oriented econometrical calculations. Proceeding of the International Symposium “Experience. Knowledge. Contemporary Challenges”, 3rd Edition “Romania in the Year of the Centenary. The European and global socio-economic context”. December 13th-14th, 2018, Editura “Artifex”, Bucharest, Romania. - 696 p. pp. 232-244, ISBN 978-606-8716-43-5
8. Naval E., Colesnicova T. Application of the econometric apparatus for estimating some parameters in the field of gender equality on the Moldovan labour market. In: Proceedings of the Scientific International Conference, Iasi, 2019 “Dezvoltarea economic-sosială durabilă a zonelor transfrontaliere”, ediţia a XV-a, 25 octombrie 2019, Iași, ICES ”Gh. Zane”. Performantica. Vol XXXV, Iași, 2019, ISBN 978-606-685-687-4. pp. 315-322.
9. Colesnicov A., Malahov L. Analysis of software tools used in calculations in gender-related scientific research. In: Creşterea economică în condiţiile globalizării: competitivitate, inovativitate, sustenabilitate = Economic growth in conditions of globalization: competitiveness, innovation, sustainability. Conferinţă Internaţională Ştiinţifico-Practică, 13th edition, October 11-12 2018: [2 vol.], Ch.: INCE, 2018. Vol. 2, pp.264-267, ISBN 978-9975-3202-9-0
10. Colesnicova T. Gender equality regulation in the sphere of employment in the Republic of Moldova. Monografie. Ch.: Complex Ed. al IEFS. 2012, CZU 316.346.2(478), C62, ISBN 978-9975-4326-5-8, - 175 p.
11. Colesnicov A., Malahov L. Using second generation wiki-platform in development of online economic applications. . In: Proceedings of the Scientific International Conference, Iasi, 2019 “Dezvoltarea economico-sosială durabilă a zonelor transfrontaliere”, ediţia a XV-a, 25.X.2019