

Е.А. Савченко, Е.В. Тутова

## Методика анализа данных, моделирования и прогнозирования уровня развития человеческого потенциала стран мира и регионов Украины

Описана методика применения корреляционно-регрессионного анализа и комбинаторного алгоритма МГУА для анализа данных, моделирования и прогнозирования уровня человеческого развития стран мира и регионов Украины. Приведены примеры решения задач восстановления данных, отбора информативных аргументов, построения моделей для анализа и прогноза.

Описано методику застосування кореляційно-регресійного аналізу та комбинаторного алгоритму МГУА для аналізу даних, моделювання і прогнозування рівня людського розвитку країн світу та регіонів України. Подано приклади розв'язання задач відновлення даних, відбору інформативних аргументів, побудови моделей для аналізу і прогнозу.

**Введение.** Макроэкономические данные зачастую содержат неопределенности, нечеткость и зашумленность. Одним из подходов, позволяющих получать приемлемые результаты по таким данным, является индуктивный подход.

Используя этот подход, можно решать задачи анализа данных, моделирования и прогнозирования. Индуктивное моделирование построено на принципах индукции, т.е. обобщения на основании некоторых фактов. Один из основных методов индуктивного моделирования – метод группового учета аргументов (МГУА). Самый популярный из них – комбинаторный алгоритм МГУА, не требующий доказательства сходимости, поскольку осуществляет полный перебор всех возможных вариантов моделей. Этот алгоритм решает задачу структурно-параметрической идентификации, т.е. позволяет найти не только параметры модели, но и ее оптимальную структуру.

Цель данной статьи: разработать на основе комбинаторного алгоритма МГУА и корреляционно-регрессионного анализа методику решения задач анализа данных, моделирования и прогнозирования уровня человеческого развития стран мира и регионов Украины. Данная методика способствует усовершенствованию процесса принятия решений в области социально-экономического развития Украины как на региональном уровне, так и для сравнения с другими странами мира, чтобы оценить ситуацию в сфере развития человеческого потенциала в Украине и в мире в целом.

### Индекс человеческого развития

Стремительное развитие научно-технического прогресса требует качественно новой ра-

бочей силы. Структурные изменения в экономике стали основой для появления современной концепции человеческого капитала.

Общая модель анализа инвестиций в человеческий капитал опирается на оценку целесообразности вложений в образование, анализ разницы в доходах и востребованности образовательных услуг. Эта теория развивается в рамках неоклассического направления экономической теории и используется также для исследования таких сфер, как образование, здравоохранение, семейные отношения и другие сферы нерыночной деятельности. Человеческий капитал состоит из приобретенных знаний, умений, мотиваций и энергии, которыми владеет человек, и которые можно использовать на протяжении какого-то промежутка времени с целью производства товаров и услуг.

Человеческий капитал – это форма капитала, так как служит источником будущих доходов. В то же время он есть совокупность знаний, умений, навыков, черт характера человека [1]. Человеческий капитал включают в категорию капитала по трем причинам:

- расходы на воспитание и образование являются реальными расходами;
- продукт этой работы увеличивает национальное богатство;
- расходы человека, которые увеличивают этот продукт, будут увеличивать национальное богатство.

Сторонники теории человеческого капитала разработали количественные методы анализа эффективности инвестиций в человеческий капитал. К факторам воспроизводства человеческого капитала относятся образование, ме-

медицинское обслуживание, подготовка на производстве, миграция, рождение детей, уход за ними и их денежная отдача для общества и семьи. Основное внимание в этих исследованиях уделяется способностям человека и дифференциации доходов, вызванным разными уровнями инвестиций в их воспроизводство. Количественный результат – это индекс человеческого развития (ИЧР).

ИЧР состоит из трех основных компонентов: продолжительности жизни, накопленных знаний и базового дохода для обеспечения достойного образа жизни. Продолжительность жизни и знания обеспечивают процесс формирования возможностей человека, его способностей, а доход – это способ измерения, который представляет круг возможностей человека, применяющего свои способности [2].

Значения всех трех компонентов ИЧР объединяются в составной индекс посредством среднего геометрического. ИЧР способствует наглядному сравнению опыта функционирования разных стран.

### **Методика анализа данных, моделирования и прогнозирования уровня человеческого развития**

Разработка данной методики начата с решения прикладных задач моделирования и поиска взаимосвязей в данных [3–6].

Для исследования использованы данные отчета ООН о человеческом развитии [7]. Авторами [3] проведено исследование влияния на ИЧР его компонентов, не связанных с доходом. Это исследование состояло в анализе влияния макроэкономических факторов на уровень человеческого развития государств мира. Используются данные ООН о развитии стран за 2011 г. Страны поделены на группы в соответствии с исследованиями ООН. Четыре группы стран содержат неодинаковое количество участников: группы, состоящие из стран с очень высоким, высоким и средним ИЧР, содержат по 47 стран, а в группе стран с низким уровнем ИЧР всего 46 государств. Выделены и проанализированы показатели, влияющие на уровень развития каждой из групп. Для построения адекватных моделей, описывающих зависимо-

сти индекса развития человеческого потенциала от макроэкономических показателей, значения всех этих показателей нормированы. Для построения этих моделей использован комбинаторный алгоритм МГУА [8, 9]. Отметим, что, поскольку данные из отчета ООН содержат много пропусков, для построения моделей отобраны страны, данные показателей которых не содержат пропусков.

В задаче моделирования влияния макроэкономических показателей на рост *валового национального дохода* (ВНД) как одного из основных компонентов ИЧР [4], авторами проанализированы 10 таких показателей. С этой целью из всех 187 стран мира отобраны 10, достигших наибольшего прогресса в рейтинге ИЧР за 2005–2012 гг. Также во всех этих странах в этот период увеличивался ВНД. Все они принадлежат к разным группам по уровню развития человеческого потенциала. Цель исследования – определить, какие из этих показателей оказывают наибольшее влияние на рост ВНД, и как их изменение может способствовать его дальнейшему росту. В процессе формирования выборки также решалась задача восстановления пропущенных данных с использованием комбинаторного алгоритма МГУА.

Для оценки информативности показателей рассчитан коэффициент парной корреляции и отобрано четыре наиболее информативных. Построены два типа моделей: регрессионная и МГУА. Анализ влияния возможных изменений некоторых факторов на рост ВНД показал, изменение каких из них могло бы привести к росту ВНД данного региона.

В задаче исследования влияния демографических и социальных показателей на *валовой региональный продукт* (ВРП) в областях Украины с наивысшим региональным ИЧР построены модели на основе МГУА [5]. Данные для исследования взяты с сайта Государственной службы статистики Украины [10].

С этой целью выделены и проанализированы демографические и социальные показатели, а также их влияние на ВРП. Для этого путем корреляционного анализа рассчитан коэффициент парной корреляции ВРП с этими показателями.

телями. Посредством комбинаторного алгоритма МГУА восстановлены пропущенные данные для регионов. Для построения зависимости ВРП от демографических и социальных показателей были разработаны два варианта моделей: путем регрессионного анализа и МГУА. Поскольку данные выборки измеряются в разных числовых диапазонах, для них применяется нормирование. Точность моделей оценивается по данным за 2013 г., которые не используются при построении моделей (экзаменационная выборка С).

Методики решения описанных задач обобщены и разработана методика анализа данных, моделирования и прогнозирования уровня человеческого развития стран и регионов (рис. 1).

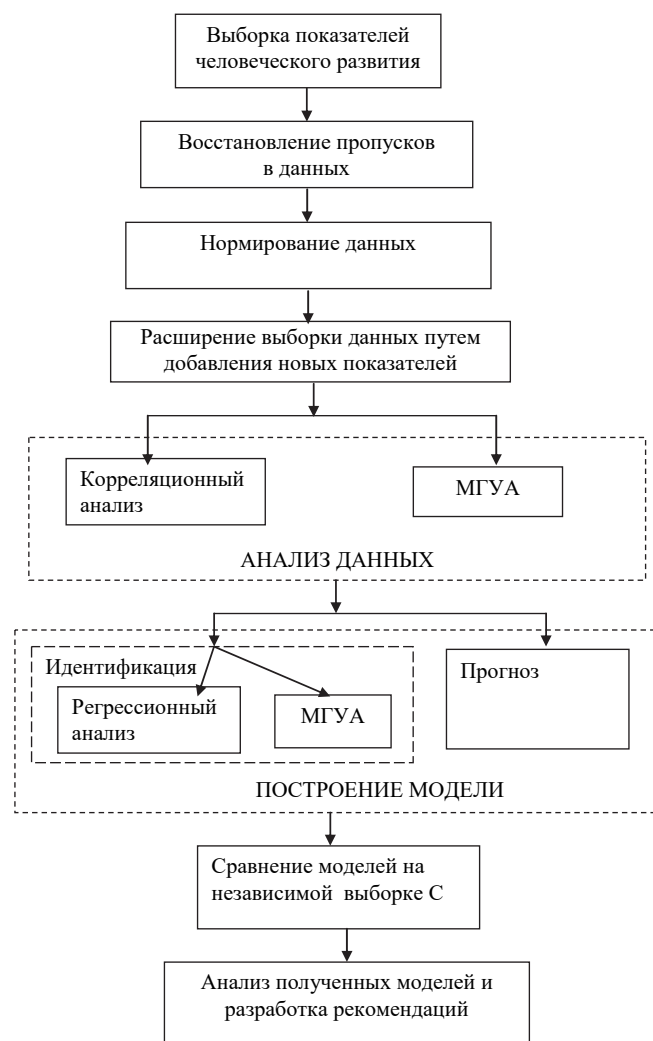


Рис. 1. Блок-схема решения задачи анализа данных, моделирования и прогнозирования

**Выборка данных** содержит информацию о показателях человеческого развития стран или одного из выбранных регионов.

**Восстановление пропусков в данных.** Существует много способов восстановления пропусков, зависящих от его формы и наличия достаточной информации для восстановления. В предложенной методике для этого использован комбинаторный алгоритм МГУА [11].

Для восстановления каждого пропуска составляется новая выборка данных и строится модель, т.е. сколько будет пропусков в данных, столько моделей потребуется построить.

**Нормирование данных.** Учитывая, что данные, как правило, измеряются в различных единицах и диапазонах, для сведения их к одному формату применяется нормирование данных, для чего значения каждого показателя делятся на максимальное значение этого показателя.

**Расширение выборки данных.** С целью возможного повышения точности модели используется способ расширения выборки введением в нее новых дополнительных показателей, которыми могут быть произведения, суммы и другие функции от исходных показателей [11].

**Анализ данных.** Этот блок предполагает два способа анализа: посредством корреляционного анализа и МГУА. Корреляционный анализ используется для оценки степени влияния каждого из показателей на выходную величину и отбора информативных показателей, который осуществляется расчетом коэффициента парной корреляции одного из показателей с выходной величиной. Используя комбинаторный алгоритм МГУА для анализа данных, можно построить модель, и, изучив, какие показатели вошли в нее и с какими коэффициентами, сделать вывод о зависимостях в данных.

**Построение модели.** Осуществляется двумя способами:

- строятся регрессионные модели, т.е. методом наименьших квадратов (МНК) определяются коэффициенты; при этом структура модели задана.

- Модели находим, используя комбинаторный алгоритм МГУА с заданными внешними критериями. При этом определяем не только параметры модели, но и ее структуру.

В зависимости от того, какой из внешних критериев задан в качестве основного критерия выбора модели, могут быть получены модели с разными свойствами. Основным внешним критерием, ответственным за точность модели, служит внешний критерий регулярности *AR*. Критерий несмещенности является критерием, отражающим требование, чтобы лучшие модели, полученные на разных частях выборки, отличались минимально. В теории МГУА известны критерий несмещенности решений *CB* и ошибок *BS* [12].

**Сравнение моделей на независимой выборке.** Полученные модели сравниваются на некоторой независимой (экзаменационной) выборке данных, не использованной для их получения. Результаты исследования подтверждают известный в индуктивном моделировании факт, что регрессионные модели дают хороший результат только на данных, на которых они получены, в отличие от моделей МГУА, дающих достаточно неплохие результаты и на новых данных.

**Анализ полученных моделей и разработка рекомендаций.** Модели анализируются как с учетом полученных параметров – как каждый из входящих в модель параметров влияет на заданную выходную величину (величина коэффициента при нем и его знак), – так и с учетом точности модели. Может быть проведен также анализ того, как изменение одного или нескольких параметров (уменьшение или увеличение) повлияет на значение выходной величины (на сколько процентов ее изменит).

На основе анализа моделей можно разработать рекомендации относительно взаимосвязей между параметрами для специалистов в данной области. Анализируя коэффициенты модели и их знаки, можно сказать, какой вклад вносит каждый показатель в выходную величину.

Рассмотрим решение некоторых задач по приведенной методике.

**Восстановление пропущенных данных.** С 1990 г. сотрудники ООН готовят доклад о проблемах человеческого развития во всем мире [3]. Он охватывает все сферы человеческой деятельности. Однако, несмотря на огромный массив

данных, которые собирают во всем мире, некоторые из параметров для ряда стран оказываются пропущенными, что затрудняет анализ представленной информации.

Так, при исследовании влияния макроэкономических показателей на рост национального дохода возникла необходимость восстановить отсутствующие данные [4]. Для проведения этого исследования выбраны 10 стран с разным уровнем человеческого развития, достигших наибольшего прогресса в этом рейтинге в 2005–2012 гг.

Одна из этих 10 стран – Сингапур. Для Сингапура отсутствуют значения ИЧР с 2006 по 2008 гг. Для восстановления этих значений использован комбинаторный алгоритм МГУА. С этой целью получены модели по выборке данных, содержащих значения для 10 стран: Гонконга, Сингапура, Чили, Беларуси, Саудовской Аравии, Венесуэлы, Ирана, Эквадора, Ганы и Танзании – за восемь лет. По этим данным получена модель, по которой рассчитаны значения ИЧР для Сингапура в 2006 г.

Построена модель вида

$$HDI_{2006} = f(HDI_{2005}, HDI_{2007}, HDI_{2008}, \\ HDI_{2009}, HDI_{2010}, HDI_{2011}, HDI_{2012}),$$

где *HDI* – значение ИЧР.

Для восстановления каждого пропуска строится модель, например по данным, приведенным в [4], получена модель для расчета ИЧР в 2007 г.:

$$HDI_{2007} = -0,0538 + 0,4818HDI_{2005} + 0,575HDI_{2012}.$$

Таким же способом были восстановлены пропущенные значения ИЧР Сингапура за другие годы.

Проблема восстановления пропущенных данных возникает также и в задачах исследования развития человеческого потенциала на региональном уровне. Начиная с 2002 г. Государственный комитет статистики Украины ежегодно готовит и распространяет статистический бюллетень «Региональний людський розвиток». К сожалению, эта статистическая отчетность неполная и содержит пропущенные данные. Для исследования влияния демографических и социальных показателей на ВРП требуется по-

строить соответствующие выборки за период с 2005 по 2012 г. для Киева, Харьковской, Черновицкой, Львовской, Киевской и Закарпатской областей, т.е. для регионов с наивысшим ИЧР в Украине согласно данным Государственного комитета статистики Украины за 2012 г. Например, для Харьковской области отсутствуют данные о количестве умерших от отравлений и действия алкоголя за 2005–2007 гг. Эти пропущенные данные восстановлены посредством комбинаторного алгоритма МГУА [11].

### Корреляционный анализ

Для исследования влияния демографических и социальных показателей на ВРП построены соответствующие выборки для Киева, Харьковской, Черновицкой, Львовской, Киевской и Закарпатской областей, т.е. для регионов с наивысшим ИЧР в Украине согласно данным Государственного комитета статистики Украины за 2012 г. [6].

Рассчитан коэффициент парной корреляции ВРП с демографическими и социальными показателями для этих областей, что дает возможность определить показатели, наиболее влияющие на рост ВРП для каждой из этих областей.

В задаче моделирования влияния макроэкономических показателей на рост национального дохода в 10 странах, достигших наибольшего прогресса в рейтинге ИЧР за период с 2005 по 2012 г. из ряда макроэкономических показателей выбраны наиболее весомые [6].

Для этого были рассчитаны значения коэффициентов парной корреляции между ИЧР и всеми остальными макроэкономическими показателями. Для определения наиболее информативных из них отобраны те, для которых значение парной корреляции превышало 0,9. Поскольку валовой национальный доход показывает уровень благосостояния в стране, он выбран в качестве выходной переменной.

### Построение моделей

В задаче моделирования влияния макроэкономических показателей на рост национального дохода в 10 странах построены два варианта моделей – регрессионные с помощью МНК и оптимальные с помощью комбинаторного алгоритма МГУА (*Combi*).

Приведем построение моделей для Беларуси (табл. 1) [4].

Таблица 1. Исходные данные для построения модели Беларуси и результаты моделирования

Год	$x_1$	$x_2$	$x_3$	$x_4$	$y$	$y_{\text{mod}}(\text{МНК})$	$y_{\text{mod}}(\text{Comby})$
2005	55,6	1	43	3,8	8,54	8,489	8,535
2006	55,2	1	42,3	2,9	9,407	9,320	9,355
2007	54,8	4	41,5	8,7	10,187	10,408	10,385
2008	54,4	3,6	40,8	4,4	11,311	11,395	11,205
2009	53,8	3,8	40,3	9,8	11,329	11,251	11,274
2010	54,2	2,5	40	6,2	12,245	12,268	12,564
2011	54,4	7,2	40,2	10,9	12,77	12,658	12,471
Средняя ошибка						<b>0,0122</b>	<b>0,0354</b>

Регрессионная модель, построенная МНК, имеет вид

$$y = 19,292 + 1,181x_1 + 0,174x_2 - 1,777x_3 - 0,066x_4, \quad (1)$$

где  $y$  – валовой национальный доход на душу населения;  $x_1$  – отношение части занятого населения к общему числу населения;  $x_2$  – прямые иностранные инвестиции, чистый приток (проценты от ВВП);  $x_3$  – отношение числа нетрудоспособного населения к трудоспособному;  $x_4$  – общие запасы за вычетом золота.

Полученная посредством комбинаторного алгоритма МГУА модель имеет вид

$$y = 7,926 + 1,643x_1 - 2,109x_3. \quad (2)$$

Анализируя результаты, можно сделать вывод, что модель, полученная посредством МНК для данных табл. 1, оказалась точнее, однако такая модель может давать противоположный результат на независимых данных, не использованных для построения модели, как будет видно из примера.

Обобщая результат построения моделей (1) и (2), можно сказать, что для Беларуси наибольшее влияние на рост валового национального дохода оказывают такие показатели, как занятость и отношение числа нетрудоспособного населения к трудоспособному.

Аналогично, для поиска зависимости ВРП от демографических и социальных показателей регионов Украины с наивысшим ИЧР по результатам 2012 г. построены два варианта моделей: регрессионная и МГУА [5].

Поскольку данные выборки находятся в разных числовых диапазонах, применяется норми-

рование по максимальному значению каждого из параметров, т.е. входные данные каждой переменной делятся на максимальное значение этой переменной за период с 2005 по 2012 г.

Точность моделей оценивается по данным за 2013 г., которые не используются при построении моделей (экзаменационная выборка  $C$ ).

Регрессионная модель для Киева выглядит так:

$$y = -7,144 + 11,852x_1 - 2,267x_2 + 0,532x_4 + 0,442x_5 + 0,881x_6 + 1,615x_7 - 3,794x_8, \\ AR(C) = 0,223, \quad (3)$$

где  $y$  – валовой региональный продукт на душу населения;  $x_1$  – отношение числа нетрудоспособного населения к трудоспособному;  $x_2$  – количество погибших на транспорте;  $x_4$  – количество самоубийств;  $x_5$  – миграционный прирост;  $x_6$  – количество разводов;  $x_7$  – смертность детей в возрасте до одного года;  $x_8$  – суммарный коэффициент рождаемости (на одну женщину).  $AR(C)$  – точность на экзаменационной выборке.

Модель, полученная по МГУА:

$$y = -2,026 + 0,169x_4 + 0,046x_5 + 2,722x_8, \\ AR(C) = 0,026. \quad (4)$$

Таким образом, для каждой из областей с наивысшим ИЧР по результатам 2012 г. построены модели на основе регрессионного анализа (3) и МГУА (4). Проверка результатов моделирования по данным за 2013 г., показала, что во всех случаях МГУА-модели оказались точнее по значению критерия  $AR(C)$ . Точность модели на экзаменационной выборке, полученной по комбинаторному алгоритму МГУА – 0,026, в то время, как точность, полученная по модели с использованием регрессионного анализа – 0,223.

### Построение моделей для прогноза показателей

Данная методика может быть использована также для построения прогнозирующих моделей. В задаче [6] построены модели ВРП для шести областей Украины на 2014–2016 гг. с применением комбинаторного алгоритма МГУА. Данные

о ВРП и социально-демографических показателях Киева представлены в табл. 2.

Социально-демографические показатели описаны в (3). Пропущенные значения восстановлены с применением комбинаторного алгоритма МГУА и нормированы. По полученным данным построены модели для прогноза каждого из показателей табл. 2 на 2014–2016 гг. (табл. 3).

Прогноз, полученный для Киева, приведен на рис. 2.

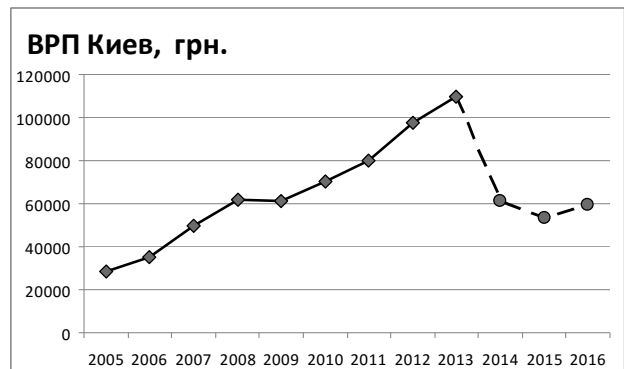


Рис. 2. Действительные (сплошная линия) и прогнозные значения (пунктир) ВРП Киева

Следует отметить, что прогнозирующие модели для расчета значений показателей на 2014–2016 гг. строились с использованием внешнего критерия несмещенности решений. Поэтому полученный результат может быть менее точным, чем с использованием критерия регулярности, но указывает тенденцию изменения выходной величины, которая неявно уже присутствовала в социально-демографических данных.

Таблица 2. Социально-демографические показатели для Киева

№	Год	$y$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$
1	2005	28780	323,00	464,00		45,00	11,40	3,90	7,50	1,1090
2	2006	35210	324,00	463,00		60,00	9,80	3,90	8,10	1,1340
3	2007	49795	327,00	560,00		132,00	9,20	3,90	9,70	1,1610
4	2008	61592	327,00	419,00	15,00	109,00	8,60	3,70	8,90	1,2250
5	2009	61088	329,00	270,00	3,00	101,00	5,60	3,30	7,50	1,2970
6	2010	70424	334,00	230,00	3,00	91,00	3,80	2,90	7,30	1,2950
7	2011	79729	339,00	215,00	9,00	99,00	3,60	1,70	8,00	1,2920
8	2012	97429	351,00	245,00	5,00	113,00	8,70	1,50	7,80	1,3810
9	2013	109402	364,00	208,00	4,00	125,00	6,40	1,40	8,20	1,3640
10	2014		379,00	264,00	1,00	143,00			7,00	
11	2015		399,00							

Таблица 3. Результаты расчетов

№	Год	$y$	$x_1$	$x_2$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$
1	2014	0,5614	0,90693	0,89688	0,59965	1,4777	0,32694	0,94425	0,98269
2	2015	0,4871	0,9335	0,91705	1,099	1,078	0,38885	1,1166	1,0218
3	2016	0,543	0,84211	0,56469	1,1275	1,1866	0,40663	1,1792	0,8489

Как видно из рисунка, для Киева максимальное значение ВРП было зафиксировано в 2013 г., после чего наблюдается резкое падение в 2014 г., приблизительно до уровня 2008–2009 гг. Изменение социально-демографических показателей в течение последних лет наряду с другими факторами также повлияло на спад ВРП в 2014–2015 гг. Эта негативная тенденция была заложена в неблагоприятных процессах, происходивших в социальной и демографической сферах. А эти изменения являются отражением экономической и социально-политической ситуации в стране в целом.

В 2015 г. ВРП незначительно упал, а в 2016 г. уже наблюдается тенденция небольшого роста ВРП. Этот результат можно считать обнадеживающим, поскольку ВРП Киева составляет более 20 процентов в ВВП Украины. Из этого можно сделать вывод о наметившейся тенденции роста ВРП Киева.

**Заключение.** На основании индуктивного подхода и корреляционно-регрессионного анализа с учетом особенности задачи в сфере уровня человеческого развития стран мира и регионов Украины разработана методика анализа данных, моделирования и прогнозирования. Используя данную методику, можно решать задачи восстановления данных, отбора информативных аргументов, построения модели для идентификации и прогноза, а также расширения пространства моделирования добавлением новых аргументов, например, произведений или сумм исходных аргументов.

Используя данную методику, специалисты, формирующие политику в социально-экономической сфере, могут разработать рекомендации по возможной корректировке отдельных аспектов человеческого развития, что позволило бы усовершенствовать процесс принятия решений.

1. *Тутова О.В.* Концептуальні основи формування людського капіталу // Економіко-математичне моделювання соціально-економічних систем. – 2012. – 17. – С. 119–129.

2. *Tutova O.V.* Conceptual Fundamentals of Assessment of a Human Development Level // УСиМ. – 2014. – № 6. – С. 65–71.
3. *Savchenko E., Tutova O.* Use of GMDH for investigation of impact of non-income components on HDI // Індуктивне моделювання складних систем. – 2012. – 4. – С. 28–37.
4. *Tutova O., Savchenko Ie.* Modeling of Impact of Macroeconomic Indicators on the Growth of National Income // Proc. of 4th Int. Conf. in Inductive Modelling ICIM 2013. – Kyiv: IRTC, 2013. – P. 315–319.
5. *Тутова О.В., Савченко Є.А.* Вплив демографічних і соціальних показників на валовий регіональний продукт в областях України // Індуктивне моделювання складних систем. – 2015. – 7. – С. 271–281.
6. *Савченко Е.А., Тутова Е.В.* Прогнозирование валового регионального продукта областей Украины на основе индуктивного подхода // УСиМ. – 2016. – № 3. – С. 84–92.
7. *Human Development Report 2011.* – <http://hdr.undp.org/en/reports>
8. *Ивахненко А.Г., Степанко В.С.* Помехоустойчивость моделирования. – Киев : Наук. думка, 1985. – 215 с.
9. *Степанко В.С.* Комбинаторный алгоритм МГУА с оптимальной схемой перебора моделей // Автоматика. – 1981. – № 3. – С. 31–36.
10. *Стат. зб. «Україна у цифрах».* – <http://www.ukrstat.gov.ua>
11. *Савченко Е.А.* Предварительная обработка данных в задаче индуктивного моделирования // УСиМ. – 2015. – № 2. – С. 82–87.
12. *Савченко Е.А., Степанко В.С.* Анализ селективных свойств критериев МГУА при их последовательном применении // Моделювання та керування станом еколого-економічних систем регіону. Зб. праць. – К.: МННЦ ІТ та С, 2008. – С. 199–210.

*E-mail: savchenko\_e@meta.ua*  
© Е.А. Савченко, Е.В. Тутова, 2016

UDC 364.2:331; 681.513

Ye.A. Savchenko, Ye.V. Tutova

## **A Procedure for the Human Potential Development Level Data Analysis, Modeling and Forecasting for Countries of the World and Ukrainian Regions**

**Introduction.** In spite of existing techniques, a problem of data analysis, modeling, and forecasting of social and economic processes is of current importance since the macroeconomic data often contain uncertainty. They also can be fuzzy and noisy. Therefore, it is often difficult to find a solution for such problems with the acceptable accuracy.

**Purpose.** The purpose of this article is to develop a procedure to solve the problems of the human development level data analysis, modeling, and forecasting for the countries of the world and the regions of Ukraine based on the combinatorial GMDH algorithm and the correlation-regression analysis. This procedure promotes improvement to the decision-making process in the field of the social and economic development policy in Ukraine both on the regional level and to compare the results achieved by our state on the way of progress in the field of the human potential development with other countries.

**Methods.** By the development of this procedure the combinatorial GMDH algorithm and regression analysis are used to build these models. Correlation analysis was applied to analyze the level of influence of indicators on the output quantity and to choose the most informative factors.

**Results.** A procedure for data analysis, modeling and forecasting to assess the human development level in the countries of the world and the Ukrainian regions was elaborated. This procedure can be used of recovering the data, selecting the informative arguments, building a model for the identification and forecasting, and extending the modeling space owing to the new arguments' addition.

**Conclusion.** The application results of this procedure are recommendations for the possible correction of some aspects of the human development. It gives a possibility to improve the decision making process in the social and economic sphere. The worked out procedure can be used by the specialists in the field of the social and economical development, international relations to analyze and forecast the main trends in the regional development and international policy.