

Начать стоит с теоретического обоснования перехода к дополненной модели с человеческим капиталом. Что имеется: необходимо оценить кросс-секцию показателей российских регионов на основании производственной модели Кобба-Дугласа. Это в явном виде модель Солоу про экзогенный экономический рост.

Одним из ее недостатков (*помимо экзогенного характера нормы сбережений*) является несовпадение модели эмпирическим данным. Объяснение на основе модели Солоу, использовавшейся, чтобы объяснить более быстрый темп роста Японии по сравнению с США в послевоенное время, плохо согласовывалось с оценкой межстрановых различий в реальных ставках процента, которая в Японии по приблизительным расчетам согласно Солоу должна была быть равна 400%. Понятно, что это нереальная оценка.

Существует два направления решения поставленной задачи. Первое - эндогенизировать научно-технический прогресс, т. е. определять темпы его роста g в процессе решения задачи. Можно пойти другим путем и изменить понятие капитала, представив его шире, чем просто физический капитал. Тогда параметр, представляющий собой долю дохода на капитал в ВВП, окажется существенно больше априорного значения $1/3$, и выводы о динамике показателей в процессе перехода к устойчивому состоянию не будут приходить в противоречие с оценкой реальной ставки процента.

То есть за счет расширения понятия капитала, который теперь состоит из физического и человеческого соответственно, мы рассчитываем увеличить значимость этого фактора в производственной функции.

Так, мы приходим к модели Менкью-Ромера-Вейла. О важности добавления человеческого капитала в производственную функцию говорили и другие исследователи, но именно эти ученые формализовали это понятие, построили регрессионные модели и оценили данные.

Поскольку будут оцениваться обе модели, план работы заключается в следующем:

- Вывод уравнений регрессии из экономических моделей;
- Разбор используемых данных;
- Построение моделей по нашей выборке и соответствующие результаты.

Задачами же исследования было не только доказать значимость тех или иных факторов в производственной функции российских регионов, но и понять, подтверждает ли эмпирика теоретические модели Солоу и Мэнкью.

Модель Солоу.

Модель Солоу имеет следующий вид:

$$Y(t) = K(t)^\alpha (A(t)L(t))^{1-\alpha}, \quad 0 < \alpha < 1$$

Предполагается, что L и A растут во времени экзогенно со скоростью n и g , то есть

$$\begin{aligned} L(t) &= L(0)e^{nt}; \\ A(t) &= A(0)e^{gt} \end{aligned}$$

Если разделить имеющуюся функцию производства на $L(t)$ и логарифмировать, мы получим следующее:

$$\ln\left[\frac{Y(t)}{L(t)}\right] = \ln A(0) + gt + \frac{\alpha}{1-\alpha} \ln(s) - \frac{\alpha}{1-\alpha} \ln(n + g + \delta).$$

Это еще уравнение еще не представляет собой регрессионную модель, но поможет нам к ней прийти.

Поговорим о спецификации модели.

В модель закладываются следующие предпосылки:

- Прежде всего мы полагаем, что g и δ не изменяются между регионами. Действительно, учитывая, что g - отражает в первую очередь развитие знаний, которое не зависит от конкретного региона (также, как и темп амортизации).
- Слагаемое $\ln A(0)$ здесь отражает не только начальный технологический уровень, но и обеспеченность ресурсами, географическое расположение, развитие институтов и т.д., что уже может изменяться у регионов. Поэтому, в предположении, что $\ln A(0) = a + \epsilon$, а $t = 0$, получим

$$\ln\left[\frac{Y}{L}\right] = a + \frac{\alpha}{1-\alpha} \ln(s) - \frac{\alpha}{1-\alpha} \ln(n + g + \delta) + \epsilon$$

, где $a = \text{const}$, а ϵ - шок, характерный для конкретного региона,

И вот это уже представляет собой уравнение регрессии.

- Мы предполагаем, что темпы сбережений и прироста населения не зависят от специфических для региона факторов, изменяющих производственную функцию. Другими словами, s и n не зависят от ошибки. Доводы к этому допущению мы опустим ввиду нехватки времени. Скажем только, что в противном случае, если бы s и n были эндогенны и зависели от дохода, то оценки, полученные МНК, могли бы быть несостоятельными.
- Поскольку по общепринятому мнению $\alpha \sim \frac{1}{3}$, то мы ожидаем получить значения коэффициентов регрессии около 0.5 и -0.5.

После того, как мы проговорили спецификацию модели, попробуем ее видоизменить.

$$\ln\left[\frac{Y}{L}\right] = a + \frac{\alpha}{1-\alpha} (\ln(s) - \ln(n + g + \delta)) + \epsilon$$

В общем-то записано то же самое, что у нас и так было, но к этому уравнению мы применим другой подход. Предполагая, что $\alpha \sim 1/3$, и, соответственно, $\frac{\alpha}{1-\alpha} = 0.5$, мы тем самым можем «наложить ограничение» на эту модель, проверяя нулевую гипотезу о том, что наше априорное

значение доли фактора капитала в выпуске в действительности согласно реальным данным равно 1/3.

Таким образом, для каждой модели будут оцениваться простое и constrained уравнения регрессии.

Теперь перейдем к **данным**.

В нашей модели нам нужно определить данные по $Y/L, s, n, g$ и δ .

- Под n мы понимаем прирост населения трудоспособного возраста за год;
- В качестве s был взят параметр реальных региональных инвестиций за год, включая государственные;
- Y/L – реальный ВВП.
- Что касается g и δ : мы будем считать далее, что их сумма равна 0.05 для всех регионов.

Важный момент: все параметры выше нормируются на трудовое население. Не на душу населения – этого фактора в модели попросту нет – а именно на экономически-активное население. Неправильная нормировка факторов может привести к смещенности получаемых результатов.

Помимо извлечения данных было выполнено разделение выборки на 2 части: на те регионы, где ВВП больше среднего и меньше среднего значения соответственно. Данное распределение мотивировано тем, что в богатых регионах ожидается большее влияние человеческого капитала, нежели в бедных. Это предположение также будет проверено.

Оценка регрессии

Были оценены две регрессии: простая и constrained. Результаты показаны *ниже*.

Простая:

Residuals:

Min	1Q	Median	3Q	Max
-0.45244	-0.14298	0.00833	0.13859	0.79237

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.21258	0.63692	5.044	3.08e-06 ***
log(inv)	0.69132	0.04357	15.867	< 2e-16 ***
log(dem_rate_)	-0.51266	0.21327	-2.404	0.0187 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2202 on 75 degrees of freedom

Multiple R-squared: 0.8023, Adjusted R-squared: 0.797

F-statistic: 152.2 on 2 and 75 DF, p-value: < 2.2e-16

Что можно увидеть:

1. Знаки при коэффициентах получились такими, какими и ожидалось. Чем выше уровень сбережений, тем богаче регион. Чем выше темпы прироста населения, тем беднее регион.
2. Оба регрессора значимы на 5% уровне.
На самом деле, учитывая около нулевой прирост населения в регионах, у нас не было высоких ожиданий относительно влияния этого фактора на выпуск, но оказалось, что подобная связь есть.
3. Adjusted R-squared (исправленный р-квадрат) равен почти 0.8.
Это неожиданно высокий результат, поскольку у Мэнкью с коллегами вышло на уровне 0.6, и это побудило их включать человеческий капитал в модель.

Оценивая Constrained уравнение, мы получили:

$$\frac{\alpha}{1 - \alpha} = 0.72 \Rightarrow \alpha = 0.42$$

Residuals:

Min	1Q	Median	3Q	Max
-0.43679	-0.13838	0.02517	0.15026	0.76921

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.39111	0.41915	10.48	<2e-16 ***
log(inv)	0.72132	0.04303	16.76	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.227 on 76 degrees of freedom

Multiple R-squared: 0.7871, Adjusted R-squared: 0.7843

F-statistic: 280.9 on 1 and 76 DF, p-value: < 2.2e-16

Это, конечно, не 1/3, но достаточно близко; и это ощутимо лучше, нежели 0.6 у Мэнкью. Однако действительно ли лучше? Используя t-test, можно ответить на этот вопрос:

$$\text{Теоретическое } \alpha: \frac{\alpha}{1 - \alpha} = \frac{1}{3} \Rightarrow \alpha = 1/2$$

$$\text{Мэнкью: } t_{\text{набл.}} = \frac{(1.48 - 1/2)}{0.12} = 8.16 > t_{\text{кр.}}$$

$$\text{Наша оценка (с Тюменью): } t_{\text{набл.}} = \frac{0.42 - 1/2}{0.043} = |-1.86| < t_{\text{кр.}}$$

Таким образом, мы получили более близкую оценку, вероятность правильности нулевой гипотезы у нас выше, и она не отклоняется.

Тем не менее попробуем улучшить результаты.

Ошибка: использовать не робастные SE.

Если говорить про другие предпосылки ТГМ, то в модели нет мультиколлинеарности, автокорреляция будет присутствовать (должна, по крайней мере), эндогенности нет.

Единственная проблема может быть с гетероскедастичностью, но, чтобы не запариваться с тестами, лучше использовать сразу исправленные SE.

Модель Мэнкью.

Модель дополняется человеческим капиталом:

$$Y(t) = K(t)^\alpha H(t)^\beta (A(t)L(t))^{1-\alpha-\beta}$$

Здесь доля капитала в выпуске представлена как сумма человеческого и физического капитала, то есть $\alpha + \beta$.

Вывод уравнения регрессии:

$$\ln \left[\frac{Y}{L} \right] = \alpha + \frac{\alpha}{1 - \alpha - \beta} \ln(s_k) + \frac{\beta}{1 - \alpha - \beta} \ln(s_h) - \frac{\alpha + \beta}{1 - \alpha - \beta} \ln(n + g + \delta) + \epsilon$$

Помимо предпосылок, заложенных Солоу, Мэнкью, Ромер и Вейл вкладывают в модель следующие предпосылки:

- $\alpha + \beta < 1$.
Это условие необходимо для существования решения. В противном случае у нас не будет 'steady state'.
- Как в модели Солоу, так и в текущей модели ожидается $\alpha \sim 1/3$.
Оценить разумное значение β , доли человеческого капитала, сложнее. В США минимальная заработная плата - грубо говоря, отдача от труда без человеческого капитала - составляет в среднем от 30 до 50 процентов средней заработной платы в обрабатывающей промышленности. Этот факт говорит о том, что от 50 до 70 процентов общего трудового дохода представляет собой отдачу от человеческого капитала. Отсюда мы примерно получаем значение β в следующем диапазоне $1/3 < \beta < 1/2$.
Можно было бы так и оставить, но нет. Поскольку мы все-таки заинтересованы в согласовании теории с практикой можно попробовать отрегулировать значение β .
<https://rosstat.gov.ru/storage/mediabank/situaz.pdf> (36 слайд)
В таком случае мы ожидаем β в следующем диапазоне: **$\beta \sim 0.11-0.12$** . Эти значения, кстати, почти идентичны сведениям Мирового банка, который оценил в конце 90-х вклад человеческого капитала на уровне 14%.
Понятно, что это все, конечно, все очень приблизительно и примерно, но всяко лучше, чем априорная информация от дедов с прошлого тысячелетия.

По поводу данных

Все факторы будем считать также, как и считали в модели Солоу. Что касается человеческого капитала, то он подсчитывается в форме инвестиций в образование, то есть без здравоохранения, социальную культуры и т.д. Такое прокси, конечно, спорное и одностороннее. Но при этом данные собрать нетрудно, а понятие определено - форма упущенных трудовых доходов со стороны студентов.

Для фактора человеческого капитала был взят выпуск молодых людей из высших учебных заведений в этом году на 1.000 человек трудоспособного населения. Этот подсчет тоже спорный, однако если наша прокси-переменная пропорциональна s_h , то это уже здорово: мы можем оценивать регрессию, а значение фактора просто изменится на константу.

Также можно было бы сказать, что, расширяя понятие капитала, мы убираем проблему эндогенности тем, что убираем из ошибки человеческий капитал, который по идее должен коррелировать с инвестициями и приростом фактора труда, однако значения корреляции были - 0.10 и 0.29 соответственно, поэтому проблемы в этом плане нет. В дополнение, исходя из вышесказанного, не должно быть в модели серьезной мультиколлинеарности, что также здорово.

Простая:

Residuals:

Min	1Q	Median	3Q	Max
-0.44857	-0.15698	0.03921	0.14851	0.60703

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.0763	0.6146	5.005	3.66e-06	***
log(inv)	0.6895	0.0419	16.455	< 2e-16	***
log(dem_rate_)	-0.4423	0.2068	-2.139	0.03574	*
log(ed)	0.1649	0.0619	2.665	0.00946	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2117 on 74 degrees of freedom

Multiple R-squared: 0.8196, Adjusted R-squared: 0.8123

F-statistic: 112.1 on 3 and 74 DF, p-value: < 2.2e-16

Из заметного:

1. Уменьшилась ошибка регрессии: с 0.227 до 0.2117.
2. Несильно, но улучшился Adjusted R-squared (исправленный р-квадрат): с 0.797 до 0.8123.
3. Значимость на 5% всех факторов производства. Переменная человеческого капитала значима на 1% уровне.
4. Знаки при коэффициентах такие, какие мы и ожидали получить.

Это все здорово, но теперь посчитаем restricted регрессии вида

$$\ln \left[\frac{Y}{L} \right] = a + \frac{\alpha}{1-\alpha-\beta} [\ln(s_k) - \ln(n + g + \delta)] + \frac{\beta}{1-\alpha-\beta} [\ln(s_h) - \ln(n + g + \delta)] + \epsilon$$

для каждой подгруппы, с включением Тюмени и без. Нас интересует именно restricted уравнения, поскольку, подставляя полученные коэффициенты в систему уравнений, мы сможем получить доли факторов в выпуске, и, соответственно, сравнить теорию и эмпирику.

Априорные теоретические значения коэффициентов **restricted regression** (для t – теста):

ошибка, нужно использовать F-test

$$\begin{cases} \frac{1/3}{1 - 1/3 - 0.11} = \frac{\alpha}{1 - \alpha - \beta} \\ \frac{0.11}{1 - 1/3 - 0.11} = \frac{\beta}{1 - \alpha - \beta} \end{cases} \Rightarrow \frac{\alpha}{1 - \alpha - \beta} = 0.59; \frac{\beta}{1 - \alpha - \beta} = 0.19$$

Результаты вычислений представлены ниже:

- Все регионы:

Коэфф-ты: 4.06 (intercept), $\begin{pmatrix} 0.714 & 0.182 \\ (0.041)' & (0.06) \end{pmatrix}$

$$\alpha = 0.376, \beta = 0.096;$$

$$t_{\text{набл.}\alpha} = \frac{0.714 - 0.59}{0.041} = 3.02 > t_{\text{кр.}\alpha} = 1.96; t_{\text{набл.}\beta} = \frac{0.182 - 0.19}{0.06} = -0.133 < t_{\text{кр.}\beta}$$

Отвержение нулевой гипотезы о равенстве $\alpha \sim 1/3$ на 5% уровне значимости.

Неотвержение нулевой гипотезы о равенстве $\beta \sim 0.11$ на 5% уровне значимости.

+ Значимость регрессоров на 5% уровне значимости (16.455, -2.139, 2.665)

- Выборка для регионов с ВВП выше среднего:

Коэфф-ты: 5.08 (intercept), $\begin{pmatrix} 0.622 & 0.191 \\ (0.08)' & (0.069) \end{pmatrix}$

Сразу оговориться, что у нас незначимость s_h темпа прироста трудоспособного населения (n) на 5% уровне

$$\alpha = 0.343, \beta = 0.0105;$$

$$t_{\text{набл.}\alpha} = \frac{0.622 - 0.59}{0.08} = 0.4 < t_{\text{кр.}\alpha}; t_{\text{набл.}\beta} = \frac{0.182 - 0.19}{0.06} = -0.01 < t_{\text{кр.}\beta}$$

Неотвержение нулевых гипотез о равенстве параметров модели априорным значениям на 5% уровне значимости.

- Выборка для регионов с ВВП ниже среднего:

Коэфф-ты: 5.5679 (intercept), $\begin{pmatrix} 0.5487 & 0.1879 \\ (0.069)' & (0.1144) \end{pmatrix}$

Сразу оговориться, что у нас полная незначимость s_h (возможно из-за мультиколлинеарности, но вряд ли, поскольку корреляция s_h с другими регрессорами невысока)

$$\alpha = 0.315, \beta = 0.0108$$

$$t_{\text{набл.}\alpha} = \frac{0.5487 - 0.59}{0.069} = -0.59 < t_{\text{кр.}\alpha}; t_{\text{набл.}\beta} = \frac{0.1879 - 0.19}{0.114} = -0.01 < t_{\text{кр.}\beta}$$

Неотвержение нулевых гипотез о равенстве параметров модели априорным значениям на 5% уровне значимости

+ Сильная значимость n (на 1% уровне)

- Без Тюмени:

Коэфф-ты: 4.24345 (intercept), $\begin{pmatrix} 0.69556 & 0.18318 \\ (0.04548)' & (0.06286) \end{pmatrix}$

$$\alpha = 0.37, \beta = 0.097$$

$$t_{\text{набл.}\alpha} = \frac{0.696 - 0.59}{0.046} = 2.28 > t_{\text{кр.}\alpha}; \quad t_{\text{набл.}\beta} = \frac{0.183 - 0.19}{0.063} = |-0.011| < t_{\text{кр.}\alpha}$$

*Ситуация стала лучше, значимость инвестиций в выпуске снизилась, но мы все равно склоняемся к альтернативной гипотезе
+ n стал впервые значим на 1% уровне значимости*

< Отдельное замечание про выполнение предпосылок Гаусса-Маркова >