Задача 1

(а)

<table>
<tr><th rowspan="2">Утверждение</th><th>Модель 1
2000–2017</th><th>Модель 2
2000–2008</th></tr>
<tr></tr>
<tr><td>1</td><td colspan="2">Оценка параметра λ положительная и значимая. Утверждение верно для стран, не таргетирующих инфляцию.
Оценка суммы $\lambda + \beta$ положительная, однако для проверки значимости нам не хватает данных (чтобы тестировать незначимость суммы коэффициентов нам недостаточно знать стандартные ошибки каждого из них). Так что мы не можем быть уверены в корректности утверждения для стран, таргетирующих инфляцию.</td></tr>
<tr><td>2</td><td colspan="2">Оценка параметра β положительная и значимая. Утверждение верно.</td></tr>
<tr><td>3</td><td colspan="2">Оценки параметров θ и τ незначимые. Так что данные не подтверждают эту гипотезу.</td></tr>
<tr><td>4</td><td colspan="2">Оценки параметров τ незначимые. Так что данные не подтверждают эту гипотезу.</td></tr>
<tr><td>5</td><td colspan="2">Доверительный интервал для параметра λ не содержит единицу, так что данные не отвергают эту гипотезу.</td></tr>
</table>

(б) Единственное изменение касается четвертого утверждения для стран, не таргетирующих инфляцию, для периода 2000-2008 годов. Для них при 10-процентном уровне значимости с утверждением можно согласиться.

Задача 2

(а) Отметим, что по условию задачи единственное отклонение от предпосылок классической модели регрессии состоит в наличии гетероскедастичности, поэтому МНК-оценки остаются несмещенными.

Будем подбирать m так, чтобы дисперсия оценки $\hat{\beta}_2$ получилась минимальной. Для этого вычислим её.

Обозначим $a = \frac{m}{n}$, где n — это общий размер выборки (в нашем случае он равен 100). Тогда $\bar{x} = a$ и $\sum (x_i - \bar{x})^2 = na(1-a)$. Договоримся, что икс равен единице для первых m наблюдений.

$$var\left(\hat{\beta}_2\right) = var\left(\frac{\sum (x_i - \bar{x})\varepsilon_i}{\sum (x_i - \bar{x})^2}\right) = \frac{1}{n^2a^2(1-a)^2} var\left(\sum (x_i - \bar{x})\varepsilon_i\right)$$

$$= \frac{1}{n^2a^2(1-a)^2} var\left((1-a)\sum_{i=1}^{m} \varepsilon_i - a\sum_{i=m+1}^{n} \varepsilon_i\right) =$$

$$= \frac{\sigma^2}{n^2a^2(1-a)^2}\left(9m(1-a)^2 + (n-m)a^2\right) =$$

$$= \frac{1}{n} * \frac{\sigma^2}{a^2(1-a)^2}\left(9a(1-a)^2 + (1-a)a^2\right) =$$

$$= \frac{\sigma^2}{n} * \frac{9(1-a)+a}{a(1-a)} = \frac{\sigma^2}{n} * \left(\frac{9}{a} + \frac{1}{1-a}\right)$$

$$\frac{\partial var\left(\hat{\beta}_2\right)}{\partial a} = \frac{\sigma^2}{n} * \left(-\frac{9}{a^2} + \frac{1}{(1-a)^2}\right) = \frac{\sigma^2}{n} * \frac{a^2 - 9(1-a)^2}{a^2(1-a)^2} =$$

$$= -\frac{\sigma^2}{n} * \frac{8a^2 - 18a + 9}{a^2(1-a)^2} = -\frac{\sigma^2}{n} * \frac{8(a-0{,}75)(a-1{,}5)}{a^2(1-a)^2}$$

Мы знаем, что a — это доля, так что лежит между нулем и единицей. Эта производная отрицательна при $a < 0.75$, равна нулю при $a = 0.75$ и больше нуля при $a > 0.75$. Таким образом, дисперсия оценки минимальна при $a = 0.75$. Следовательно, $m = 0.75 \cdot 100 = 75$.

(б) Нужно применить взвешенный метод наименьших квадратов.

Задача 3

(а) Для решения надо убедиться, что ожидаемая инфляция является стационарным процессом (так как все корни характеристического уравнения лежат вне единичного круга) и заметить, что безработица является стационарным процессом (как и любой MA-процесс конечного порядка). Получается, что фактическая инфляция равна сумме двух независимых друг от друга стационарных процессов и, следовательно, сама является стационарным процессом.

(б) Прогноз уровня безработицы на два периода вперед равен $\hat{u}_{2022} = 5$, а дисперсия его ошибки составляет $var\left(5 + \varepsilon_{2022} - 0.25\varepsilon_{2021}\right) = 1.0625$.

Прогноз уровня ожидаемой инфляции на два периода вперед равен

$$\widehat{\pi^e_{2022}} = 0.6\left(0.6\pi^e_{2020} + 0.3\pi^e_{2019}\right) + 0.3\pi^e_{2020} =$$

$$= 0.6(0.6 \cdot 20 + 0.3 \cdot 20) + 0.3 \cdot 20 = 5.4 + 3 = 16.8$$

Дисперсия его ошибки составляет $var\left(0.6\left(0.6\pi^e_{2020} + 0.3\pi^e_{2019} + v_{2021}\right) + 0.3\pi^e_{2020} + v_{2022}\right) = var\left(0.6v_{2021} + v_{2022}\right) = 1.36$

Следовательно, прогноз фактической инфляции составляет $16.8 + 2 \cdot (5 - 5) = 16.8$, а дисперсия его ошибки равна $1.36 + 2^2 \cdot 1.0625 = 5.61$.

Таким образом, доверительный интервал имеет вид:

$$\left(16.8 - 1.96 \cdot \sqrt{5.61}\ \ ;\ \ 16.8 + 1.96 \cdot \sqrt{5.61}\right)$$

$$(16.8 - 1.96 \cdot 2.37\ \ ;\ \ 16.8 + 1.96 \cdot 2.37)$$

$$(12.15\ \ ;\ \ 21.45)$$

Задача 4

1. Можно ожидать, что коэффициент β будет положительным, если студенты участвуют в некотором соревновании, где победителю достаётся приз (например, ВСОШ по экономике). Отрицательного значения β следует ожидать, если слишком хороший результат не считается приемлемым в учебной группе.

2. Рассмотрим фиксированную школу s, вычислим матожидания по потенциальным исходам для всех студентов и усредним их. Получим такое выражение:

$$\frac{1}{n_s}\sum_{j\in s} E\left[Y_j\left(\{\omega_l\}_{l\in s(i)}\right)\right] = \alpha_s + \beta\sum_{j\in s}\frac{E\left[Y_j\left(\{\omega_l\}_{l\in s(i)}\right)\right]}{n_s} + \frac{\tau}{n_s}\sum_{j\in s}\omega_j$$

   Решим это уравнение относительно $\frac{1}{n_s}\sum_{j\in s} E\left[Y_j\left(\{\omega_l\}_{l\in s(i)}\right)\right]$ и подставим в исходное выражение для модели. Тогда для студента i получим такое представление:

$$Y_i\left(\{\omega_j\}_{j\in s(i)}\right) = \frac{\alpha_{s(i)}}{1-\beta} + \frac{\beta\tau}{1-\beta}\sum_{j\in s(i)}\frac{\omega_j}{n_s} + \tau\omega_i + u_i$$

3. Для каждой школы возьмём среднее значение исхода в группе воздействия, вычтем из него среднее значение в контрольной группе, затем усредним по всем школам. Легко видеть, что это несмещённая оценка. В таком случае невозможно оценить β (а следовательно и γ), поскольку отсутствует вариация в доле студентов из группы воздействия по школам.

4. Мы можем оценить τ тем же способом, что и в прошлом пункте. Затем мы можем оценить γ, сравнив исходы в контрольной группы в разных школах: мы берём средние значения исходов в контрольных группах и используем их как зависимую переменную в регрессии на долю учеников в группе воздействия и константу. Так мы получим несмещённую оценку для γ. Зная оценки для τ и γ, мы можем оценить β, используя результат из пункта 2.

5. Мы всё ещё можем оценить τ, но не можем оценить γ и β, поскольку решения директоров могут коррелировать с α_s. Это делает оценку из регрессии, описанной в предыдущем пункте, несостоятельной. В общем случае, если нет рандомизации на уровне школ, мы не можем отделить фиксированные эффекты школ от эффектов окружения.

Задача 5

1. Справедлива следующая цепочка равенств:

$$E\left[Y_i \mid X_i, D_i=1\right]=E\left[Y_i^{\star} \mid X_i, D_i=1\right]=h\left(X_i\right)+E\left[\epsilon_i \mid X_i, D_i=1\right]= h\left(X_i\right)+\frac{E[\epsilon_i D_i|X_i]}{E[D_i|X_i]}$$

   Условно на X_i, D_i зависит от u_i и Z_i. По построению Z_i и ϵ_i независимы, однако u_i и ϵ_i могут быть зависимы. Следовательно, в общем случае $E\left[Y_i \mid X_i, D_i=1\right] \neq h\left(X_i\right)$.

2. Поскольку Z_i не зависит от u_i, X_i мы получаем, что распределение u_i условно на X_i и Z_i тоже равномерное. Остальное следует из определения матожидания и из $p(z,x) \in [0,1]$:

$$E\left[D_i \mid Z_i, X_i\right]=E\left[\left\{u_i<p\left(Z_i, X_i\right)\right\} \mid X_i, Z_i\right]=p\left(Z_i, X_i\right)$$

3. По той же логике, что и в пункте 1 мы получаем следующее:

$$E\left[Y_i \mid X_i, Z_i, D_i=1\right]=h\left(X_i\right)+\frac{E\left[\epsilon_i D_i \mid X_i, Z_i\right]}{E\left[D_i \mid X_i, Z_i\right]}$$

   Из пункта 2 получаем $E\left[D_i \mid X_i, Z_i\right] = p\left(Z_i, X_i\right)$ и по определению имеем: $E\left[\epsilon_i D_i \mid X_i, Z_i\right]=E\left[\epsilon_i\left\{u_i<p\left(X_i, Z_i\right)\right\} \mid X_i, Z_i\right]$. Поскольку $\left(u_i, \epsilon_i\right)$ не зависят от X_i, Z_i, последнее матожидание – функция только от $p\left(X_i, Z_i\right)$. Также если $p\left(X_i, Z_i\right)=1$, то $E\left[\epsilon_i D_i \mid X_i, Z_i\right]=E\left[\epsilon_i \mid X_i, Z_i\right]=0$, что означает $\lambda(0)=0$. На практике предпосылка о независимости может не выполняться, так как распределение ϵ_i может быть различным для групп с разными значениями X_i. Поскольку это распределение никак не связано с Z_i, её случайность здесь не помогает.

4. Используя результат из пункта 3 и тот факт, что $p\left(0, z_m\right)=p\left(1, z_w\right)$, мы получаем следующее:

$$E\left[Y_i \mid X_i=1, Z_w=z_w, D_i=1\right]-E\left[Y_i \mid X_i=0, Z_w=z_m, D_i=1\right]= h(1)-h(0)+\lambda\left(p\left(1, z_w\right)\right)-\lambda\left(p\left(0, z_m\right)\right)=h(1)-h(0)$$

   Отсюда следует, что мы можем определить гендерный разрыв $h(1)-h(0)$ из данных. Чтобы определить $h(0)$ нам нужно найти наблюдения с такой z, что $p(0,z)=1$, поскольку только в этом случае смещение отбора рано 0. То же самое справедливо для $h(1)$.

5. Чтобы что-то понять про $h(\cdot)$, нам возможно хотелось бы применить логику из пункта 3. Однако, по построению, распределение ϵ_i зависит от $h\left(X_i\right)$ (потому что Y_i би-

нарная), так что предположение никогда не выполняется. Поэтому, в общем случае мы не можем использовать этот подход. Даже если бы исход был непрерывной переменной, и предположения были бы выполнены, чтобы определить $E\left[h\left(X_i\right)\right]$ нам нужно было бы иметь наблюдения, которые тестируются с единичной вероятностью (как уже обсуждалось в пункте 4).