Санкт-Петербургский Государственный Университет
Факультет прикладной математики – процессов управления
Кафедра моделирования экономических систем

Бабшукова Екатерина Викторовна

Выпускная квалификационная работа бакалавра

Решение задачи о доходах населения: анализ, моделирование и прогноз

Направление 010400

Прикладная математика и информатика

Научный руководитель:
к. ф.-м. н., доцент Евстафьева В. В.

Санкт-Петербург
2017

Оглавление

Введение

В настоящее время Российская Федерация находится в напряжённой ситуации, связанной со сложной политической атмосферой в мире и её неоднозначным положением в отношении других стран [19]. Трудности вынуждают государство укреплять свои политические взгляды, формировать направление дальнейшего развития, что не может не влиять на социально-экономическую сферу жизни страны [11]. Последствия, которые мы наблюдаем, выражаются в ускорении темпа инфляции и в обострении проблемы бедности [29].

Исследование динамики изменения уровня жизни населения может способствовать определению причин данных проблем и методов их решения. Анализ среднедушевого денежного дохода населения позволяет наглядно рассмотреть проблему социального неравенства, выделить основные тенденции изменения благосостояния жителей страны в целом или отдельного города. Он является особенно актуальным в условиях реализации стратегии социально-экономического развития Санкт-Петербурга до 2030 года [22].

Для анализа среднедушевого дохода нужно построить математическую модель, которая бы описывала его динамику и давала возможность прогнозировать будущие значения при условии отсутствия происшествий, способных вызвать резкие изменения не только в социально-экономической сфере, но и в жизни страны в целом, таких как война, революция, смена политического режима, экономический кризис.

В данной работе представлен анализ текущего уровня жизни населения Санкт-Петербурга и сравнение с уровнем жизни населения страны. Построены математические модели на основе временного ряда, который состоит из значений среднедушевых доходов жителей Санкт-Петербурга за 2009–2016 гг. С помощью модели с наилучшими прогностическими свойствами выполнен прогноз на 3 первых месяца 2017 года. Более того, с учётом основных показателей уровня жизни населения предложена спецификация регрессионной модели для формирования среднедушевого дохода.

1. Постановка задачи

Исследуемые в работе данные представляют собой значения среднедушевого дохода населения Санкт-Петербурга за период с января 2009 года по декабрь 2016 года (рис. 1).

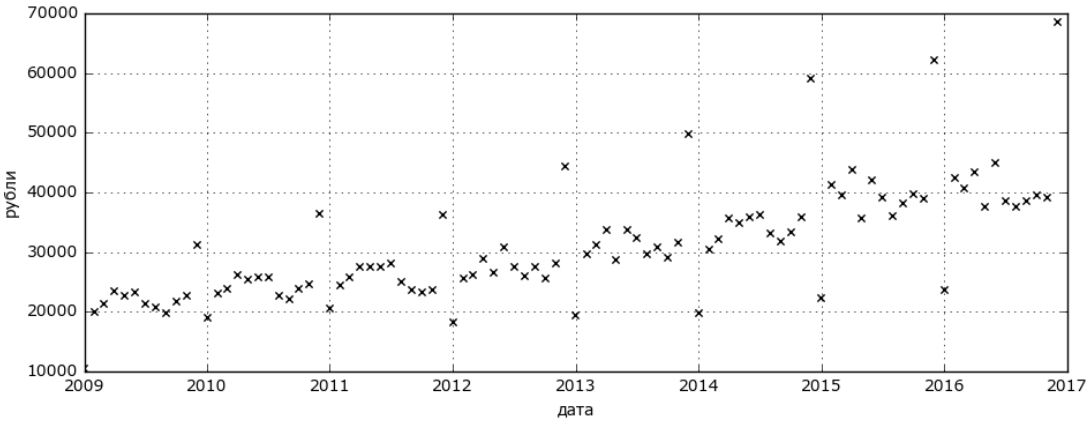


Рис. 1: Значения среднедушевого дохода [26]

Целью работы являются анализ среднедушевого дохода населения г.Санкт-Петербурга и построение математической модели, адекватно описывающей его динамику. Для достижения этой цели были поставлены следующие задачи:

- изучить методы прогнозирования временных рядов;
- построить временной ряд по статистическим данным, доступным из открытых источников;
- предложить спецификацию регрессионной модели для построенного временного ряда на основе корреляционного анализа;
- построить математические модели для временного ряда;
- сравнить прогностические свойства моделей по критериям точности прогнозирования на контрольной выборке;
- построить краткосрочный точечный прогноз и доверительный интервал для уровня значимости 0,05.

2. Обзор предметной области исследования

В разделе представлены основные термины регрессионного анализа и теории временных рядов, эконометрические методы и статистические критерии, которые использовались в настоящей работе, по материалам [16] и [10].

2.1. Основы регрессионного анализа

Линейная эконометрическая модель с p объясняющими переменными имеет вид:

$$y_i = \theta_1 x_{i1} + ... + \theta_p x_{ip} + \varepsilon_i,\ i = \overline{1,n},\ n \geq p, \tag{1}$$

где $y_i, x_{i1}, .., x_{ip},\ i = \overline{1,n}$ — *наблюдения*. Оценивание неизвестных коэффициентов производится с помощью *метода наименьших квадратов*:

$$\theta_i = \widehat{\theta}_i,\ i = \overline{1,n}: \quad Q(\widehat{\theta}_1, .., \widehat{\theta}_p) = \min_{\theta_1,..,\theta_p} \sum_{i=1}^{n} (y_i - \theta_1 x_{i1} - ... - \theta_p x_{ip})^2.$$

Остаточная сумма квадратов RSS равна этому минимальному значению Q.

Коэффициент детерминации R^2 определяется соотношением:

$$R^2 = 1 - \frac{RSS}{TSS}, \quad TSS = \sum_{i=1}^{n} (y_i - \overline{y})^2.$$

Стандартные предположения об ошибках модели:

- модель наблюдений имеет вид (1) или, в матричной форме, $y = X\theta + \varepsilon$, где $X = \{x_{ij}\}_{np}$, $\theta = (\theta_1, .., \theta_p)$, $\varepsilon = (\varepsilon_1, .., \varepsilon_n)$;
- $\varepsilon_1, .., \varepsilon_n \sim i.i.d.\ N(0, \sigma^2)$, где $i.i.d.$ – независимые, одинаково распределённые случайные величины;
- $detX^T X \neq 0$.

Рассматриваемая модель относится к классу *регрессионных моделей*:

$$Y_i = f(X_{i1}, .., X_{ip}) + \varepsilon_i,\ i = \overline{1, n}.$$

Теорема Гаусса – Маркова гласит, что если модель наблюдений имеет вид $y = X\theta + \varepsilon$ с попарно не коррелирующими случайными ошибками, имеющими нулевое математическое ожидание и одинаковые дисперсии, где столбцы X линейно независимы, а $detX^TX \neq 0$, то оценка $\widehat{\theta} = (X^TX)^{-1}X^Ty$, найденная по методу наименьших квадратов, является наилучшей линейной несмещённой оценкой. Она имеет наименьшую возможную дисперсию в классе линейных смещённых оценок коэффициента, т.е. является *эффективной*.

Для того чтобы определить значимость оценки коэффициента линейной регрессии на соответствующем уровне значимости α, используют критерий Стьюдента. Для проверки нулевой гипотезы о равенстве коэффициента θ нулю ($H_0 : \theta = 0$) против альтернативной гипотезы $H_1 : \theta \neq 0$ находят значение статистики:

$$T = \frac{\widehat{\theta} - \theta_0}{s_{\widehat{\theta}}} \sim t(n-2), \tag{2}$$

где $\widehat{\theta}$ — оценка коэффициента регрессии, $\theta_0 = 0$, $s_{\widehat{\theta}}$ — оценки дисперсии оценки $\widehat{\theta}$, n — количество наблюдений. При справедливости нулевой гипотезы статистика критерия подчиняется распределению Стьюдента с $(n-2)$ степенями свободы. Если полученное значение по модулю меньше критического значения, найденного по таблице распределения Стьюдента с учётом числа степеней свободы и уровня значимости, то нет оснований отвергнуть гипотезу H_0, т.е. его незначимости, в противном случае принимается H_1.

Гипотеза значимости регрессии в целом имеет вид:

$$H_0 : \theta_2 = ... = \theta_p = 0.$$

Соответствующий статистический критерий основан на F-статистике:

$$F = \frac{(RSS_{H_0} - RSS)/(p-1)}{RSS/(n-p)},$$

где RSS_{H_0} – остаточная сумма квадратов, которая получается при оценивании модели с наложенными нулевой гипотезой ограничениями. H_0 отвергается, если $F > F_{1-\alpha}(p-1, n-p)$, где $F_{1-\alpha}(p-1, n-p)$ — квантиль распределения $F(p-1, n-p)$ уровня $(1-\alpha)$.

Одним из критериев выбора "наилучшей" модели может выступать коэффициент R^2 либо, если модели имеют разное число объясняющих переменных, *скорректированный коэффициент детерминации*:

$$R^2_{adj} = 1 - \frac{RSS/(n-p)}{TSS/(n-1)}.$$

Также часто применяют *информационные критерии*. При использовании, например, *критерия Акаике* линейной модели с p объясняющими переменными сопоставляется значение:

$$AIC = ln\Big(\frac{RSS}{n}\Big) + \frac{2p}{n} + 1 + ln2\pi. \tag{3}$$

Если $detX^TX$ близок к 0, так что между двумя или более объясняющими переменными есть высокая степень линейной корреляции, то говорят о наличии *мультиколлинеарности* в модели. В таком случае бывает невозможно оценить влияние отдельных переменных.

Статистические критерии проверки выполнения стандартных предположений о модели нацелены на проверку гипотезы:

$$H_0 : \varepsilon_1, .., \varepsilon_n \sim i.i.d.\ N(0, \sigma^2),$$

при этом каждый из них есть способ определения специфических нарушений стандартных предположений о модели.

Критерий Голдфелда – Квандта позволяет выявить гетероскедастичность остатков, т.е. возможную неоднородность их дисперсий. Статистика критерия имеет вид:

$$F = \frac{RSS_2}{RSS_1}.$$

Если остатки гомоскедастичны, т. е. однородны, то статистика имеет распределение Фишера $F\left(\frac{n-r}{2}-p, \frac{n-r}{2}-p\right)$. Гипотеза $H_0 : D(\varepsilon_i) = \sigma^2,\ i = \overline{1,n}$, отвергается, если найденное значение больше критического, соответствующего уровню значимости α.

Критерий Дарбина – Уотсона используется, когда график указывает на автокоррелированность последовательности ошибок ε_i. Предполагается, что её структура определяется следующим образом:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \delta_i,\ i = \overline{1,n},$$

где $|\rho| < 1,\ \delta_i \sim i.i.d.\ N(0, \sigma_\delta^2)$.

Статистика критерия:

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2},$$

где $e_i, i = \overline{1,n}$ — остатки модели.

Нулевая гипотеза $H_0 : \rho = 0$ соответствует независимости $\varepsilon_i,\ i = \overline{1,n}$. Она отвергается в пользу $H_1 : \rho > 0$, если $DW < d_{L\alpha}$, и принимается, если $DW > d_{U\alpha}$, где $d_{L\alpha}, d_{U\alpha}$ — критические значения статистики критерия, определяемые числом наблюдений и уровнем значимости. В ином случае вывод о справедливости нулевой гипотезы не делается.

Критерий Бройша – Годфри используется для проверки гипотезы о некоррелированности ошибок, однако при этом он допускает зависимость случайных составляющих ε_i в виде процесса авторегрессии K-ого порядка:

$$\varepsilon_i = a_1\varepsilon_{i-1} + ... + a_K\varepsilon_{i-K} + \delta_i,\ i = \overline{1,n},$$

где $\delta_1, .., \delta_n \sim i.i.d.\ N(0, \sigma^2)$. Статистика критерия равна nR^2, где R^2 — коэффициент детерминации вспомогательной модели:

$$e_i = \gamma_1 x_{i1} + ... + \gamma_p x_{ip} + a_1 e_{i-1} + ... + a_K e_{i-K} + \nu_i,\ i = \overline{1,n},$$

где $\nu_1, .., \nu_n \sim i.i.d.\ N(0, \sigma^2)$, $\gamma_i = \theta_1 x_{i1} + ... + \theta_p x_{ip} + \varepsilon_i$.

Нулевая гипотеза $H_0 : a_1 = ... = a_K = 0$ отклоняется, если $nR^2 > (nR^2)_{crit} = \chi^2_{1-\alpha}(K)$, где $\chi^2_{1-\alpha}(K)$ — квантиль распределения $\chi^2(K)$ уров-

ня $(1-\alpha)$.

Критерий Харке – Бера применяется для проверки выполнения условия нормальности ошибок. Если оно выполняется, то при большом числе наблюдений статистика:

$$JB = n\Big(\frac{S^2}{6} + \frac{(K-3)^2}{24}\Big),$$

где $S = \frac{\sum_{i=1}^{n} e_i^3}{n\widehat{\sigma}^3}$ — выборочный коэффициент асимметрии, $K = \frac{\sum_{i=1}^{n} e_i^4}{n\widehat{\sigma}^4}$ — выборочный коэффициент эксцесса, e_i — остатки модели, n – количество наблюдений, подчиняется распределению $\chi^2(2)$. Гипотеза о нормальности ошибок отвергается, если $JB > \chi^2_{1-\alpha}(2)$, где $\chi^2_{1-\alpha}(2)$ — квантиль распределения $\chi^2(2)$, соответствующая уровню $(1-\alpha)$.

2.2. Процессы авторегрессии и скользящего среднего

Временной ряд — ряд значений некоторой переменной, которые были измерены через последовательные равные промежутки времени. Если принять длину промежутка за единицу времени, то можно считать, что наблюдения $x_i, i = \overline{1,n}$, некоторой переменной x производились в моменты $t_i, t = \overline{1,n}$.

Пусть есть последовательность n наблюдений $x_i, i = \overline{1,n}$, некоторого признака X, выбранных случайным образом из некоторой совокупности, называемой *генеральной совокупностью*, так что $x_i, i = \overline{1,n}$, есть реализации независимых, одинаково распределённых случайных величин $X_i, i = \overline{1,n}$.

Закон распределения этих величин характеризуется функцией распределения $F(x) = P(X < x), -\infty < x < \infty$. Если $F(x)$ задаёт *непрерывное* распределение, то для него определена фунция плотности $p(x)$: $F(x) = \int_{-\infty}^{\nu} p(x)dx$. Тогда для $\forall a, b : -\infty < a \leq b < \infty, P(a \leq X < b) = F(b) - F(a)$.

В статистическом анализе временных рядов наблюдения $x_i, i = \overline{1,n}$, рассматриваются как реализации статистически зависимых случайных

величин $X_i, i = \overline{1,n}$, подчиняющихся некоторому совместному распределению с функцией распределения:

$$F(\nu_1, .., \nu_n) = P\{X_1 < \nu_1, .., X_n < \nu_n\}.$$

В основном рассматриваются временные ряды, у которых совместное распределение $X_i, i = \overline{1,n}$ имеет совместную плотность $p(x_1, .., x_n)$:

$$F(\nu_1, .., \nu_n) = \int_{-\infty}^{\nu_1} ... \int_{-\infty}^{\nu_n} p(x_1, .., x_n) dx_1 ... dx_n.$$

Последовательность $X_i, i = \overline{1,n}$, образует *случайный процесс с дискретным временем.*

Если

$$F(\nu_1, .., \nu_n) = P\{X_1 < \nu_1\} \cdot .. \cdot P\{X_n < \nu_n\} = \prod_{i=1}^{n} F(\nu_i),$$

$$p(x_1, .., x_n) = \prod_{i=1}^{n} p(x_i),$$

то $X_i, i = \overline{1,n}$, представляет собой *случайную выборку* из распределения с функцией F.

Случайный процесс X_t, который порождает временной ряд $x_t, t = \overline{1,n}$, — *строго стационарный*, если его свойства не изменяются при изменении начала отсчёта времени. Для $m = 1$ это означает, что закон распределения X_t не зависит от t, а значит, не зависят и $E(X_t) = \mu, D(X_t) = \sigma^2$.

Степень тесноты статистической связи между случайными величинами X_t и $X_{t+\tau}$ может быть определена с помощью *коэффициента парной корреляции*:

$$Corr(X_t, X_{t+\tau}) = \frac{Cov(X_t, X_{t+\tau})}{\sqrt{D(X_t)}\sqrt{D(X_{t+\tau})}}, \tag{4}$$

$$Cov(X_t, X_{t+\tau}) = E[(X_t - E(X_t))(X_{t+\tau} - E(X_{t+\tau}))].$$

Если X_t — строго стационарный случайный процесс, то $Cov(X_t, X_{t+\tau})$

является функцией от τ: $Cov(X_t, X_{t+\tau}) = \gamma(\tau)$. Тогда коэффициент корреляции тоже зависит только от τ:

$$Corr(X_t, X_{t+\tau}) = \frac{\gamma(\tau)}{\gamma(0)} = \rho(\tau).$$

Случайный процесс — *слабо стационарный*, если для него выполнены условия:

$$\begin{aligned} &E(X_t) = \mu, \\ &D(X_t) = \sigma^2, \\ &Cov(X_t, X_{t+\tau}) = \gamma(\tau). \end{aligned}$$

На практике, как правило, под стационарным процессом подразумевают слабо стационарный.

Пусть x_t — стационарный ряд. В этом случае $\rho(\tau)$ называют *коэффициентом автокорреляции*, так как он измеряет корреляцию между значениями одного ряда. При анализе его изменения говорят об *автокорреляционной функции* $\rho(\tau)$, а график этой функции называют *коррелограммой*.

Белый шум — стационарный случайный процесс $X_t, t = 0, \pm1, \pm2, ..,$ такой что:

$$E(X_t) = 0, D(X_t) = \sigma^2 > 0,\ \rho(\tau) = 0, \tau \neq 0.$$

Процесс авторегрессии p-ого порядка (AR(p)) описывает порождение ряда следующим образом:

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + ... + a_p X_{t-p} + \varepsilon_t, \tag{5}$$

где ε_t — белый шум, X_0 — случайная величина, $a_p \neq 0$ — постоянные коэффициенты.

При рассмотрении таких процессов иногда используют *оператор сдвига* L: $LX_t = X_{t-1}$. С помощью $L^k, k = \overline{1, p}$, можно переписать (5) в виде:

$$X_t = a_1 L X_t + a_2 L^2 X_t + ... + a_p L^p X_t + \varepsilon_t$$

или

$$a(L)X_t = \varepsilon_t.$$

Для того чтобы процесс AR(p) был стационарым, необходимо и достаточно, чтобы все корни $a(z) = 0$ лежали вне единичного круга на комплексной плоскости.

Другой моделью порождения ряда является *процесс скользящего среднего порядка q* (MA(q)). Согласно нему,

$$X_t = \varepsilon_t + b_1\varepsilon_{t-1} + ... + b_q\varepsilon_{t-q}, \tag{6}$$

где ε_t — белый шум, X_0 — случайная величина, $b_p \neq 0$ — постоянные коэффициенты.

Для того чтобы он был обратим, необходимо и достаточно, чтобы все корни $b(z) = 0$ лежали вне единичного круга на комплексной плоскости. В этом случае если X_t имеет представление (6), то для него также существует AR-представление:

$$d(L)(X_t - \nu) = \varepsilon_t,$$

где

$$d(L) = 1 - \sum_{j=1}^{\infty} d_j L^j = \frac{1}{b(L)}, \quad d_0 = 1, \ \sum_{j=1}^{\infty} |d_j| < \infty.$$

Процесс X_t принадлежит классу *процессов авторегрессии — скользящего среднего* ARMA(p,q), если:

$$X_t = \sum_{i=1}^{p} a_j X_{t-i} + \sum_{j=0}^{q} b_j \varepsilon_{t-j}, \quad a_p \neq 0, \ b_q \neq 0 \ b_0 = 1. \tag{7}$$

В операторной форме:

$$a(L)X_t = b(L)\varepsilon_t.$$

Интегрированная модель авторегрессии — скользящего среднего (ARIMA) является расширением моделей ARMA для нестационарных временных рядов, которые можно привести к стационарному виду пу-

тём взятия разности некоторого порядка. Она имеет вид:

$$\Delta^d X_t = \sum_{i=1}^{p} a_j \Delta^d X_{t-i} + \sum_{j=0}^{q} b_j \varepsilon_{t-j}, \quad a_p \neq 0,\ b_q \neq 0\ b_0 = 1.$$

где Δ^d — оператор разности временного ряда порядка d. Представление с оператором сдвига L:

$$a(L)(1-L)^d X_t = b(L)\varepsilon_t.$$

Если нестационарный временной ряд имеет сезонную составляющую, то на его значениях строят *сезонную интегрированную модель авторегрессии — скользящего среднего* (SARIMA). В операторной форме SARIMA(p,d,q)(P,D,Q)$_s$ выглядит следующим образом:

$$a(L)\widetilde{a}(L^s)(1-L)^d(1-L^s)^D X_t = b(L)\widetilde{a}(L^s)\varepsilon_t,$$

где s — период сезонности, p — порядок авторегрессии, d — порядок разности, q — порядок скользящего среднего, P — порядок сезонной авторегрессии, D — порядок сезонной разности, Q — порядок сезонного скользящего среднего.

2.3. Подбор модели ARMA

Выбор модели ARMA(p,q) включает 3 этапа:

- идентификацию модели;
- оценивание модели;
- диагностику модели.

На первом этапе определяются значения p и q, а также предварительные оценки коэффициентов модели. Уточнение этих оценок производится на втором этапе. На третьем осуществляется проверка адекватности выбранной модели относительно имеющихся данных.

Прежде всего нужно учесть то, что построить такую модель можно только на основе стационарного временного ряда. Если имеющийся ряд

не удовлетворяет данному условию, то его нужно привести к стационарному одним из существующих способов, например, с помощью взятия разностей или устранением сезонной/трендовой компоненты. Проверку выполнения условия можно провести с помощью критерия Дики – Фуллера [16].

Провести идентификацию модели помогают отличия в поведении автокорреляционных и частных автокорреляционных функций рядов, подчиняющихся различным процессам.

Вопрос о порядке авторегрессии решается с помощью *частной автокорреляционной функции*. Её значение на лаге k, $\rho_{part}(k)$, находится как значение коэффициента корреляции между X_t и X_{t+k} с устранённым влиянием промежуточных случайных величин, соответствующее решению относительно a_k системы первых k уравнений Юла – Уокера:

$$\rho(s-1)a_1 + \rho(s-2)a_2 + ... + \rho(s-k)a_k = \rho(s),\ s = \overline{1,k}.$$

Тогда если X_t — процесс типа AR(p), то

$$\rho_{part}(p) \neq 0, \quad \rho_{part}(k) = 0,\ k > p.$$

Этот факт позволяет по графику частной автокорреляционной функции определить значение p.

С помощью похожего свойства автокорреляционной функции находится порядок процесса скользящего среднего:

$$\rho(q) \neq 0, \quad \rho(q) = 0,\ k > q.$$

На практике вместо $\rho(k)$ и $\rho_{part}(k)$ известны их оценки — выборочные автокорреляция $r(k)$:

$$r(k) = \frac{\frac{1}{T-k}\sum_{t=1}^{T-k}(x_t - \widehat{\mu})(x_{t+k} - \widehat{\mu})}{\frac{1}{T}\sum_{t=1}^{T}(x_t - \widehat{\mu})^2},\ k = \overline{1, T-1}$$

и частная автокорреляция $r_{part}(k)$, которую можно получить, заменив автокорреляции в $\rho_{part}(k)$ их оценками. Во многих случаях по выборочным оценкам можно судить о поведении теоретических функций.

Наряду со значениями выборочных автокорреляционных функций, вычисляют значения *Q-статистик Льюнга – Бокса*:

$$Q = T(T+2)\sum_{k=1}^{M}\frac{r^2(k)}{T-k}.$$

Эта величина имеет асимптотическое распределение $\chi^2(M)$ при $T \to \infty$. Q-статистики относятся к критерию для проверки гипотезы о том, что имеющиеся данные — реализация белого шума.

Если в результате идентификации моделей подходящими оказываются несколько вариантов, то выбор между ними можно произвести с помощью информационных критериев. Одним из таких является *критерий Шварца*, статистика которого имеет вид:

$$BIC = ln\widehat{\sigma}^2_{p,q} + (p+q)\frac{lnT}{T},$$

где T — число наблюдений, $\widehat{\sigma}^2_{p,q}$ — оценка дисперсии.

Модель выбирают ту, для которой BIC имеет наименьшее значение.

На этапе оценивания модели обычно используют метод максимального правдоподобия, сводящийся к методу наименьших квадратов. Эта задача решается итерационными методами, где в качестве начальных значений берут оценки, полученные на предыдущем этапе.

При оценивании моделей с MA(q) компонентой важным является условие обратимости процесса. При его выполнении можно положить $\varepsilon_0 = \varepsilon_{-1} = ... = \varepsilon_{-q+1} = 0$. Для получения более точной аппроксимации существует процедура *backcasting*, в которой значения $\varepsilon_0, \varepsilon_{-1}, .., \varepsilon_{-q+1}$ находятся посредством построения обратного прогноза.

При диагностике модели с помощью различных критериев проверяются статистические гипотезы, которые сводятся к предположению, что последовательность ε_t образует белый шум.

Пусть на первом этапе была выбрана модель ARMA(p,q):

$$a(L)X_t = b(L)\varepsilon_t,$$

а на втором получили её оценку:

$$\widehat{a}(L)X_t = \widehat{b}(L)\varepsilon_t.$$

Если MA-компонента обратима, то

$$\varepsilon_t = \frac{a(L)}{b(L)}X_t,$$

$$\widehat{\varepsilon_t} = \frac{\widehat{a}(L)}{\widehat{b}(L)}X_t.$$

Если ε_t — белый шум, то $\widehat{\varepsilon_t}$ должны быть на него похожи. Взяв за основу это соображение, было предложено определять значимость выборочных автокорреляций для ошибок:

$$r_\varepsilon(k) = \frac{\sum_{t=1}^{T-k}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t+k}}{\sum_{t=1}^{T}\widehat{\varepsilon}_t^2}$$

с помощью Q-статистики Льюнга – Бокса:

$$Q_{LB} = T(T+2)\sum_{k=1}^{M}\frac{r_\varepsilon^2}{T-k},$$

которая имеет асимптотическое распределение $\chi^2(M-p-q)$. Гипотезу адекватности модели нет оснований принять, если $Q_{LB} > \chi^2_{0,95}(M-p-q)$.

Можно воспользоваться точным критерием Дарбина – Уотсона или Бройша – Годфри.

Многие статистические методы, которые применяются для анализа временных рядов, предполагают нормальность имеющихся данных. Для её проверки обычно используется критерий Харке – Бера.

Достоинствами модели являются подробная методика и простота построения модели и проверки её на адекватность. К недостаткам относят её неадаптивность и требование к длине временного ряда: например, для построения адекватной модели SARIMA нужно не менее 6 сезонов [27], которые на практике не всегда можно получить. Тем не менее, на сегодняшний день этот класс моделей является одним из

наиболее популярных.

2.4. Метод Хольта – Уинтерса

Метод Хольта – Уинтерса эффективен в прогнозировании временных рядов, содержащих трендовую и сезонную компоненты [5]. Суть метода состоит в тройном экспоненциальном сглаживании ряда.

Прогноз на h шагов в мультипликативной версии метода находится по формуле:

$$\widetilde{y}_{t+h} = (\ell_t + b_t h) s_{t+h-m}, \quad t = \overline{1, n}. \tag{8}$$

Коэффициенты уточняются следующим образом:

$$\begin{aligned} \ell_t &= \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}), \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1}, \\ s_t &= \gamma \frac{y_t}{\ell_{t-1}} + (1-\gamma) s_{t-m}, \end{aligned} \tag{9}$$

где ℓ_t — общий уровень ряда, b_t — трендовая компонента, s_t — сезонная составляющая, m — период сезонности, n — размер ряда, α, β, γ – параметры.

Начальный тренд b_0 определяется по формуле:

$$b_0 = \frac{1}{m}\left(\frac{y_{m+1} - y_1}{m} + \frac{y_{m+2} - y_2}{m} + ... + \frac{y_{m+m} - y_m}{m}\right). \tag{10}$$

Для нахождения начальных значений сезонных коэффициентов $s_i, i = \overline{1, m}$, вычисляется средний уровень каждого имеющегося m, затем наблюдения делятся на соответствующие им сезонные средние, а нужные значения получаются как средние по каждому номеру i.

К достоинству метода Хольта – Уинтерса можно отнести простоту анализа и построения модели. Метод часто используется для выполнения долгосрочного прогноза [8]. Недостатком считается отсутствие гибкости [1].

3. Экономический анализ денежных доходов населения

В разделе представлен анализ уровня жизни жителей Санкт-Петербурга и сравнение с уровнем жизни населения страны. Кроме того, рассматриваются исследуемые данные с корректировкой на инфляционный процесс.

3.1. Неравенство доходов и сравнение с показателями по Российской Федерации

Показатель денежных доходов является ключевым индикатором, определяющим уровень жизни населения региона и всей страны. Уровень жизни отражает обеспеченность людей материальными и духовными благами, степень удовлетворения их различных потребностей. С ростом доходов увеличивается потребление продуктов рыночной экономики, оказывая влияние на развитие её сегментов. К тому же, доходы населения являются одним из источников налоговых поступлений в бюджеты различных уровней, которые затем распределяются на нужды государственных масштабов. Таким образом, доходы играют важную роль не только в жизни отдельных людей, но и в реализации целей и задач различных сфер жизни страны.

Неравномерность распределения доходов и, как следствие, благ оценивается показателем дифференциации доходов, связанной различиями в положении людей. В зависимости от этого показателя находится стабильность общества и уровень социального напряжения в нём, структура и объём потребления. Поэтому установление масштабов различий членов общества в доходах, улучшение качества жизни населения есть актуальная правительственная задача [15].

Одним из способов оценки дифференциации денежных доходов является расчёт коэффициента Джини[1] (индекса концентрации доходов)

[1]Коэффициент Джини — статистический показатель степени расслоения населения страны или региона по отношению к некоторому признаку. Он может быть рассчитан по формуле Джини: $G =$

[3]. Он может принимать значения из (0;1), и чем ближе значение к единице, тем в большей степени денежные доходы находятся в расположении наиболее обеспеченного слоя населения. Самое большое значение этого коэффициента среди регионов и городов федерального значения Российской Федерации отмечено в Москве и в 2015 году составило 0,432. На втором месте по величине дифференциации доходов находится Санкт-Петербург, где коэффициент Джини равен 0,417 по данным того же года. Высокий уровень расслоения общества в доходах наблюдается по России в целом (таблица 1).

год / город/страна	2011	2012	2013	2014	2015
г.Москва	0,503	0,486	0,481	0,452	0,432
г.Санкт-Петербург	0,442	0,443	0,443	0,437	0,417
Российская Федерация	0,417	0,42	0,419	0,416	0,412

Таблица 1: Значения коэффициентов Джини [25]

Другим показателем дифференциации денежных доходов населения служит коэффициент фондов (децильный коэффициент) — отношение среднедушевого дохода 10% наиболее состоятельной части населения к доходу 10% наименее состоятельной. В 2015 году на долю 10% наиболее обеспеченного населения Санкт-Петербурга пришлось 30,6% доходов, на долю 10% наименее обеспеченного — 1,9%. Децильный коэффициент был равен 16,2. В 2014 году на долю 10% богатейшей и 10% беднейшей части горожан пришлось 32,3% и 1,7% от общей суммы среднедушевых доходов соответственно, а коэффициент фондов составил 18,7 [21].

Существующее в России расслоение общества по величине доходов отражено в значениях коэффициента фондов за период с 2005 по 2015 гг. (таблица 2). Несмотря на высокий уровень неравенства по доходам, стоит отметить его постепенное уменьшение с 2007 года. Однако, остаётся значительным отставание России в степени дифференциации дохо-

$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i-y_j|}{2n^2\overline{y}}$, где n — число домохозяйств, y_k — доля дохода домохозяйства в общем доходе, $\overline{y}$ — среднее долей доходов домохозяйств.

дов от уровня большинства развитых стран, в частности стран Европы [2].

год	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
к.ф.	15,9	16,7	16,6	16,6	16,6	16,2	16,4	16,3	16,0	15,6

Таблица 2: Значения коэффициента фондов по России [26]

Наиболее показательными для анализа уровня жизни населения являются среднедушевые и реальные денежные доходы. Среднедушевой денежный доход находится путём деления общей суммы денежного дохода S за отчётный период на численность населения P, включая детей и пенсионеров:

$$I_a = \frac{S}{P}. \tag{11}$$

Реальный денежный доход — относительный показатель, который определяется путём деления индекса реального дохода J_r на индекс потребительских цен J_p:

$$I_r = \frac{J_r}{J_p}. \tag{12}$$

Индекс потребительских цен, в свою очередь, является одним из важнейших показателей, характеризующих уровень инфляции. Статистика по среднедушевым доходам предоставляется в абсолютных величинах, по реальным — в процентном отношении к тому же периоду в предыдущем году. Для удобства в таблице 3 все показатели по Санкт-Петербургу занесены в процентах на основании официальных данных Росстата и Петростата [20, 21].

Как видно из таблицы, в 2014 году имел место значительный рост индекса потребительских цен. Среднедушевые доходы увеличиваются по сравнению с предыдущими периодами, чего нельзя сказать о реальных доходах. Однако факт снижения покупательской способности рубля не обязательно сопровождается уменьшением реальных доходов, если только номинальные или среднедушевые доходы не отстают от уровня инфляции [28].

По данным за 2015 год (таблица 4), большая часть населения в

год / показатель	2012	2013	2014	2015	2016
Среднедушевой доход, %	106,8	112,8	110,6	115,0	110,1
Реальный денежный доход, %	103,0	107,6	103,7	98,9	103,1
Индекс потребительских цен, %	106,1	106,7	113,3	113,2	105,2

Таблица 3: Значения показателей в % к декабрю предыдущего года

Санкт-Петербурге (70,6%) имеет значение среднедушевого дохода меньше 45 тыс. рублей в месяц. 8% жителей города получают в месяц меньше 10 тыс. рублей. По сравнению с данными 2014 года, среднедушевые доходы населения г.Санкт-Петербург увеличились: 13,2% жителей имели доход меньше 10 тыс. руб./мес, а у 23,4% он был больше 45 тыс. рублей. Численность населения с денежными доходами ниже величины прожиточного минимума в 2015 году составила 415,4 тыс. человек или 8% от общего численности населения, что на 0,3% меньше, чем в 2014 году [21].

	в 2015 году		в 2014 году	
	тыс. чел.	в %	тыс. чел.	в %
Всё население, в том числе со среднедушевым доходом в месяц, руб	5191,7	100	5131,9	100
до 10000	415,3	8,0	677,4	13,2
от 10000,1 до 19000	1053,9	20,3	1226,5	23,9
от 19000,1 до 27000	882,6	17,0	867,3	16,9
от 27000,1 до 45000	1313	25,3	1159,8	22,6
от 45000,1	1526,4	29,4	1200,9	23,4

Таблица 4: Распределение населения Санкт-Петербурга по размеру среднедушевого дохода [20]

Ситуация распределения населения по размеру среднедушевого дохода в г.Санкт-Петербург обстоит немного лучше, чем в целом по России (таблица 5). На 81,4% жителей страны приходится доход меньше 45 тыс. руб./мес. Среднедушевой доход 14,2% населения составляет меньше 10 тысяч рублей в месяц. В сравнении с 2014 годом, отмечается рост доходов: число людей, которые получают меньше 19 тысяч рублей в ме-

сяц, уменьшилось на 5,1%, при этом категории с доходом выше увеличились на 0,3%, 1,8% и 3% соответственно. Однако, величину денежных доходов меньше прожиточного минимума в 2015 году получили 13,3%, в то время как в 2014 эта цифра составляла 11,2%.

	в 2015 году		в 2014 году	
	тыс. чел.	в %	тыс. чел.	в %
Всё население,	146267,3	100	143667	100
в том числе со среднедушевым				
доходом в месяц, руб				
до 10000	20769	14,2	25141,7	17,5
от 10000,1 до 19000	38907,1	26,6	40801,4	28,4
от 19000,1 до 27000	26474,4	18,1	25572,7	17,8
от 27000,1 до 45000	32910,1	22,5	29739	20,7
от 45000,1	27205,7	18,6	22412	15,6

Таблица 5: Распределение населения России по размеру среднедушевого дохода [20]

Таким образом, величина среднедушевого дохода с течением времени имеет возрастающую тенденцию, распределение населения по его размеру изменяется в сторону увеличения числа жителей с бóльшим доходом, причём это относится как к Санкт-Петербургу, так и к России.

3.2. Учёт инфляционного процесса в денежных доходах

Если J_c — индекс покупательской способности, S — сумма денег, измеренная по номиналу, то эта же сумма с учётом её обесценивания равна [14]

$$C = S \cdot J_c.$$

J_c связан с индексом потребительских цен (индексом инфляции) J_p следующим соотношением:

$$J_c = \frac{1}{J_p}.$$

Темп инфляции — относительный прирост цен за период, определяющийся по формуле

$$h = 100(J_p - 1). \tag{13}$$

Инфляция является цепным процессом. Следовательно, индекс цен за несколько периодов равен произведению

$$J_p = \prod_{i=1}^{n} \left(1 + \frac{h_i}{100}\right). \tag{14}$$

Пусть базовый месяц в расчётах — декабрь 2008 года. Сравним реальную покупательскую способность по значениям среднедушевого дохода в рублях с покупательской способностью денег в базовом месяце. Для этого нужно каждое значение дохода разделить на индекс инфляции за весь предшествующий период, найденный по формуле (14) на основе данных [24].

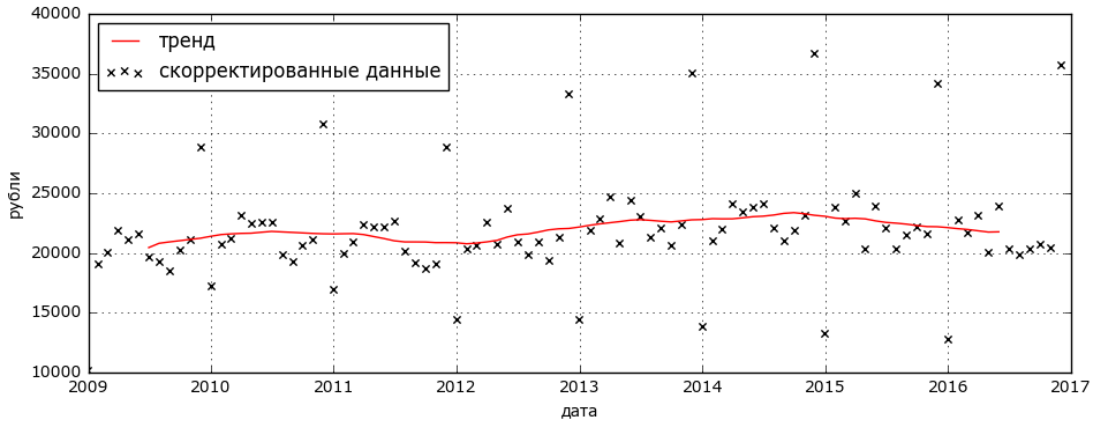


Рис. 2: Значения среднедушевого дохода населения с учётом обесценивания денежной массы

Как видно из рис. 2, на протяжении 8 лет величина среднедушевого дохода не убывала, но и не имела серьёзного возрастания. На протяжении последних двух лет покупательская способность денег снижается. За весь рассматриваемый период темп инфляции в г.Санкт-Петербург составил примерно 92%, после 2009-ого года — 77%. Среднедушевой доход за 2016 год относительно 2009 года больше на 65%-125% в зависимости от месяца. Следовательно, он не всегда опережает рост цен.

4. Корреляционно-регрессионный анализ основных экономических показателей уровня жизни населения

На формирование среднедушевого денежного дохода населения оказывает влияние большое количество различных факторов. В их число входят как базовые показатели, отражающие наиболее общие условия и сложившиеся в результате длительного развития (географическое положение, природные условия и др.), так и частные, которые характеризуют современные процессы и выступают индикаторами уровня жизни населения (среднемесячная номинальная заработная плата, структура доходов и др.). Показатели из второй группы способны динамично меняться, поэтому в качестве факторных признаков возьмём часть из них [12].

Введём обозначения (см. приложение С).

признак	описание признака, единицы измерения
Y	СДД, руб. в месяц
X_1	СНЗП, руб. в месяц
X_2	ПМ, руб. в месяц
X_3	МРОТ, руб. в месяц
X_4	ЧЗ, млн. человек
X_5	ИПЦ, % в месяц
X_6	ЧН, млрд. человек

Таблица 6: Состав признаков (переменных)

На основе данных по выбранным показателям за период с января 2009 года по декабрь 2016 года [17] построим динамические ряды, состоящие из 96-ти значений, с шагом в 1 месяц (рис. 3). Отдельно взятый ряд есть выборка, элементы которой составляют область определения конкретной объясняющей переменной (таблица 6).

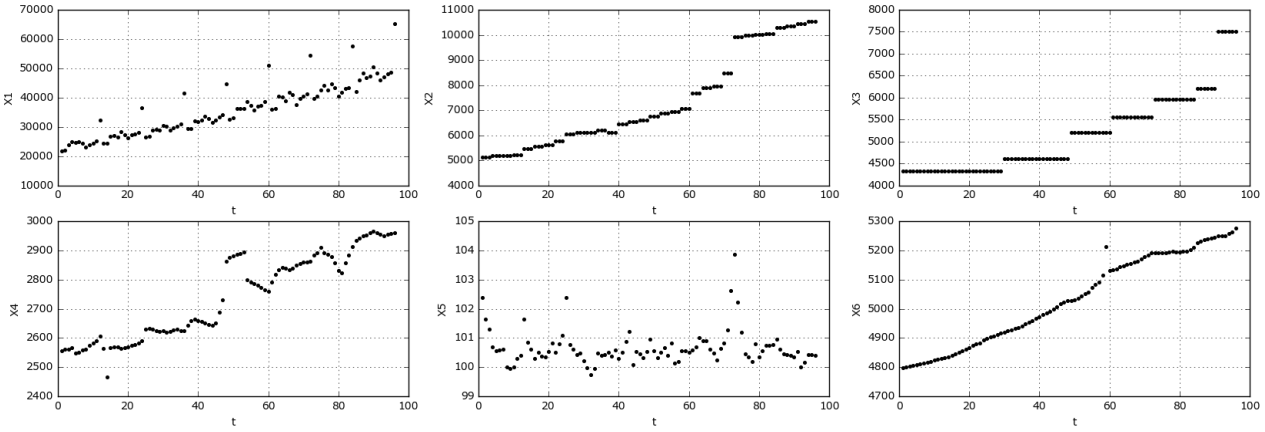


Рис. 3: Временные ряды, построенные на значениях показателей уровня жизни населения

Для того чтобы осуществить отбор факторных признаков, проведём корреляционный анализ, предварительно устранив сезонные компоненты из Y и X_1 и переобозначив соответствующие переменные. Диаграммы рассеяния для каждой пары "факторный признак — результативный признак", являющиеся графическим представлением корреляционной связи, изображены на рис. 4. По ним можно предположить наличие корреляции между всеми парами, кроме Y и X_5. Корреляция между Y и X_6 объясняется тем, что между ними присутствует функциональная зависимость, поэтому признак X_6 далее мы не будем рассматривать. Для того чтобы подтвердить предположения относительно других пар, найдём матрицу коэффициентов парной корреляции (таблица 7).

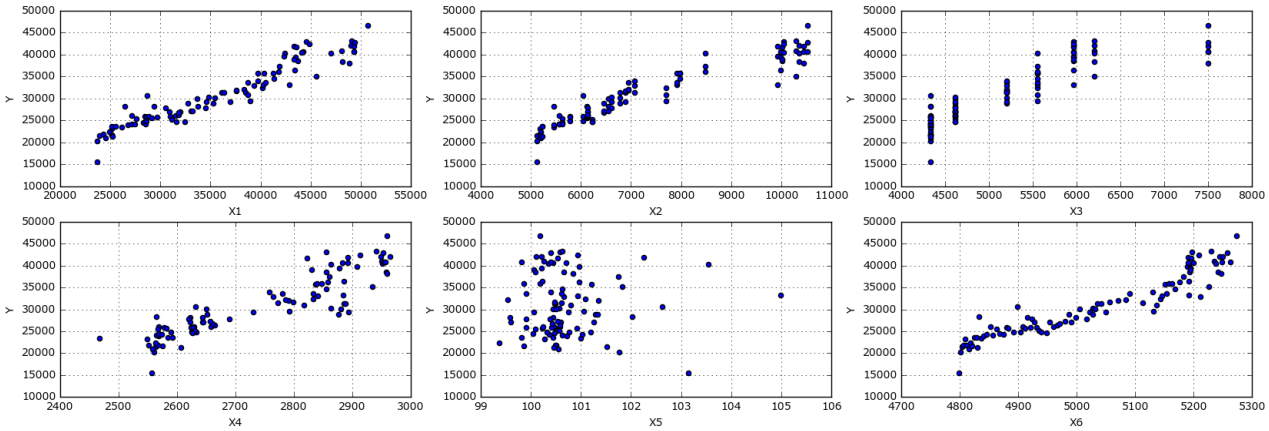


Рис. 4: Диаграммы рассеяния

Наибольшие значения коэффициентов корреляции $r_{YX_i}, i = \overline{1,5}$ со-

	Y	X_1	X_2	X_3	X_4	X_5
Y	1	0,9551	0,9462	0,8918	0,8989	0,0069
X_1	0,9551	1	0,9611	0,9391	0,9500	0,0090
X_2	0,9462	0,9611	1	0,9348	0,9021	0,0644
X_3	0,8989	0,9391	0,9348	1	0,8950	0,0069
X_4	0,8918	0,9500	0,9021	0,8950	1	0,0562
X_5	0,0069	0,0080	0,0644	0,0069	0,0562	1

Таблица 7: Результаты корреляционного анализа

ответствуют X_1, X_2, X_3. Их мы включим во многофакторную регрессионную модель формирования среднедушевого дохода населения

$$\widehat{y_t} = ax_{t-1}^{(1)} + bx_{t-1}^{(2)} + cx_{t-1}^{(3)},\ t = \overline{2,n}, \tag{15}$$

где $(\widehat{y_2}, .., \widehat{y_t}, .., \widehat{y_n})^T = \widehat{Y}$, $(x_1^{(1)}, .., x_{n-1}^{(1)})^T = X_1$, $(x_1^{(2)}, .., x_{n-1}^{(2)})^T = X_2$, $(x_1^{(3)}, .., x_{n-1}^{(3)})^T = X_3$, $a, b, c \in \mathbb{R}$. В модель включены лаги объясняющих переменных, чтобы избежать проблемы эндогенности [6].

Оценка уравнения множественной регрессии (15) методом наименьших квадратов имеет вид:

$$\widehat{y_t} = 0{,}3687x_{t-1}^{(1)} + 1{,}6008x_{t-1}^{(2)} + 1{,}2188x_{t-1}^{(3)},\ t = \overline{2,n}. \tag{16}$$

Поскольку регрессоры коррелируют друг с другом, в модели присутствует мультиколлинеарность. Она имеет умеренную степень: число обусловленности матрицы (X_1, X_2, X_3) равно 89. Результаты оценивания модели с помощью пакетной реализации МНК в языке Python представлены в таблице 12 приложения А.

Было принято во внимание то, что данные имеют общую тенденцию, отчего величины коэффициентов корреляции могут быть завышенными. Однако между рассмотренными показателями присутствует смысловая, очевидная связь, поэтому анализ проводился на основе временных рядов, не подвергавшихся предварительно устранению трендовых

компонент.

Несмотря на невысокие величины коэффициентов корреляции между индексом потребительских цен и другими показателями, нельзя сделать вывод, что между инфляцией и среднедушевым доходом нет связи. На рис. 3 видно, что почти все значения индекса больше 100. Это значит, что цены в некотором периоде выше, чем в предыдущем. Если бы имеющиеся данные определялись как соотношения цен каждого периода с одним, конкретно взятым (базовым), то график и результаты могли быть другими.

5. Построение математических моделей

В разделе анализируется временной ряд, построенный на значениях среднедушевого дохода жителей Санкт-Петербурга, и на его основе выполняется построение трёх математических моделей: трендовой модели с учётом сезонности, модели Хольта – Уинтерса и сезонной интегрированной модели авторегрессии — скользящего среднего.

5.1. Трендовая модель

Рассмотрим временной ряд y_t, состоящий из 96-ти значений среднедушевого дохода населения Санкт-Петербурга за период с января 2009 по декабрь 2016 года, с шагом в 1 месяц (рис. 5).

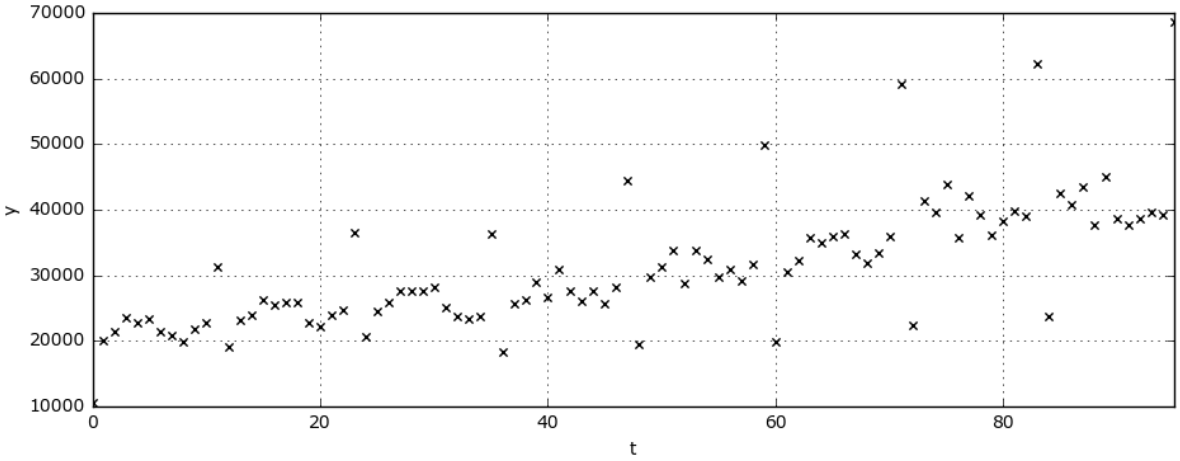


Рис. 5: Исходный временной ряд

При визуальном анализе можно отметить несколько особенностей временного ряда: наличие тренда во времени и сезонной составляющей — ежегодно повторяющейся закономерности, — как следствие, он не является станционарным. Для того чтобы подтвердить эти предположения, рассмотрим график автокорреляционной функции (рис. 6).

Коррелограмма содержит всплески, по которым можно предположить период сезонных колебаний. В данном случае наибольшее значение автокорреляционной функции наблюдается при 12-ом лаге, следующее по величине значение соответствует 24-ому лагу. Поэтому период сезонности возьмём равный 12 месяцам.

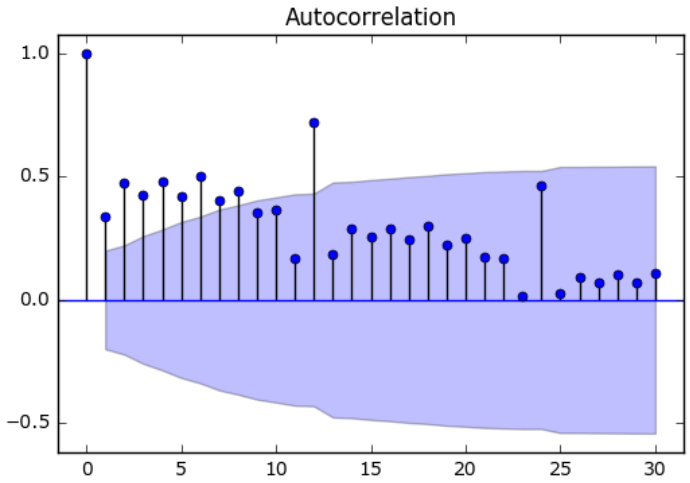


Рис. 6: Значения автокорреляционной функции для исходного временного ряда

Будем производить дальнейшие вычисления и построения на временном ряде, состоящем из 72-х значений среднедушевого дохода за период с января 2009 по декабрь 2014 года, который будем далее понимать как исходный, чтобы иметь возможность сравнить прогнозные значения, рассчитанные с помощью построенных далее моделей, с реальными.

Для начала устраним сезонную составляющую из рассматриваемого ряда. В качестве первого этапа проведём сглаживание ряда методом скользящего среднего [9, 7]. Величина параметра метода характеризует то, сколько памяти (прошлых значений) включает в себя среднее. При его небольшом значении сглаживание повторяет изменение данных, а при значительном отражает общую картину без влияния краткосрочных колебаний. Для устранения резких всплесков и выявления тренда нужно взять окно длиною в 12 месяцев (рис. 7). Поскольку выбранное значение параметра — чётное число, нужно совершить ещё одно сглаживание, заменив каждые две соседние точки на среднее, чтобы получить коэффициенты для конкретных месяцев.

Далее, чтобы определить сезонные коэффициенты, рассмотрим отношение исходного ряда к сглаженному. Полученные значения усредним для каждого месяца и нормируем таким образом, чтобы их сумма была равна 12. С помощью итоговых мультипликативных коэффициен-

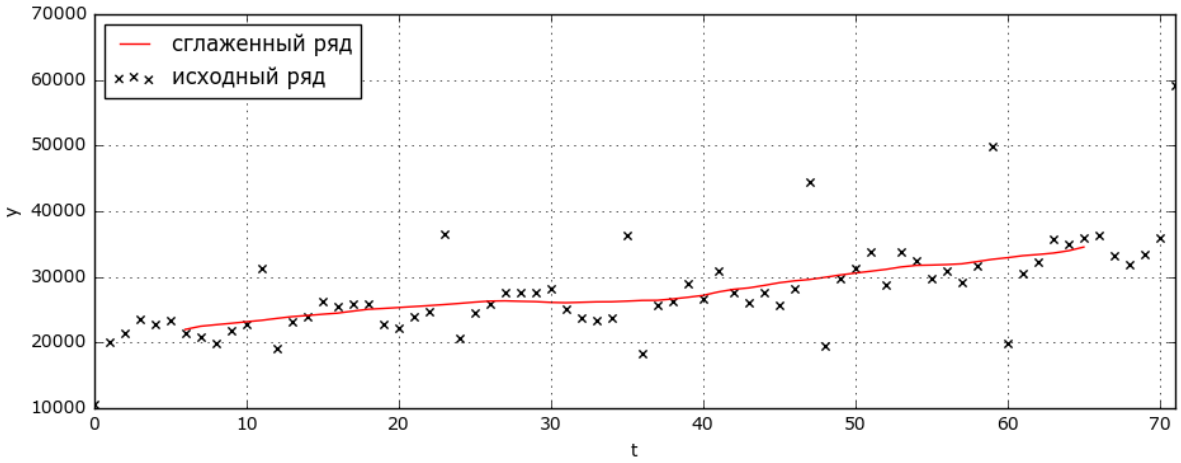


Рис. 7: График исходного временного ряда и сглаженного методом скользящего среднего

тов избавим ряд от сезонной составляющей путём деления фактических значений на них для каждого месяца.

Процедура устранения сезонной компоненты из временного ряда реализована на языке Python. Исходный код представлен в приложении В.

Сглаженная кривая позволяет чётко увидеть тренд в исходном временном ряде. Рассмотрим несколько его моделей и выберем из них наиболее адекватно отражающую динамику данных.

Уравнение линейного тренда, полученное с помощью метода наименьших квадратов, для преобразованного временного ряда выглядит следующим образом:

$$\widehat{y}_t = 208{,}9083t + 20180. \tag{17}$$

Анализ адекватности моделей проводился с помощью средств библиотеки *statsmodels* языка Python. Результаты работы пакетных реализаций МНК и отдельных статистических критериев сведены в таблицу 13 приложения А.

Для того чтобы проверить значимость коэффициентов линейной регрессии, рассмотрим критерий Стьюдента. Поскольку p-значения для каждого коэффициента равны нулю, что меньше уровня значимости $\alpha = 0{,}05$, нулевую гипотезу о незначимости коэффициентов нет основа-

ний принять. Модель в целом также является значимой: это показывает высокое значение F-статистики критерия Фишера (418) и нулевое соответствующее ему p-значение. Коэффициент детерминации для данной модели составляет 0,857. Его больше значение указывает на высокую точность приближения построенной модели к исходному ряду данных.

Уравнение квадратичного тренда, полученное с помощью метода наименьших квадратов, для исходного временного ряда имеет следующий вид:

$$\widehat{y}_t = 1{,}3268t^2 + 112{,}05t + 21370. \tag{18}$$

Все его коэффициенты являются значимыми: p-значения для каждого меньше уровня значимости, следовательно, нулевую гипотезу о незначимости коэффициентов нет оснований принять. Значение F-статистики равно 228, соответствующее p-значение — нулю, значит, модель в целом значима. Коэффициент детерминации равен 0,869, т.е. эта модель более точно описывает исходные данные, чем предыдущая, однако в данном случае значение коэффициента детерминации не является определяющим выбор модели критерием. Если модель будет слишком хорошо соответствовать исходному ряду, то может пострадать качество прогноза. Нужно ориентироваться не только на точность модели, но и на её простоту, и так как линейный тренд не намного уступает полиномиальному, остановимся на нём.

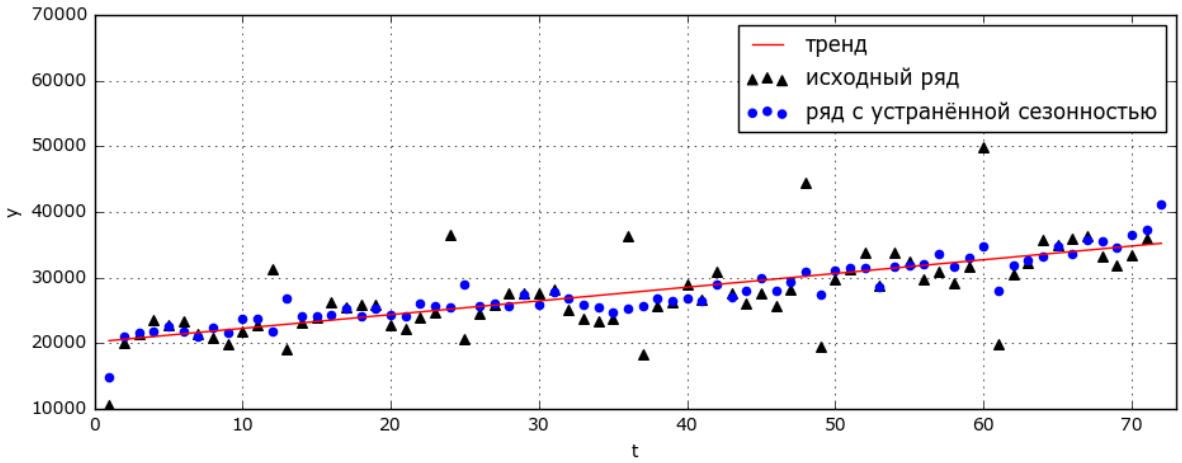


Рис. 8: Исходный ряд, скорректированный ряд и тренд

Проверим остатки трендовой модели (17) на выполнение стандарт-

ных предположений: равенство нулю математического ожидания, отсутствие автокорреляции, гомоскедастичность и нормальность распределения дополнительно. Все расчёты приведены в таблице 13 приложения А. Математическое ожидание остатков равно $4{,}72 \cdot 10^{-12}$, что очень близко к нулю, следовательно, первая предпосылка выполнена. Автокорреляции в остатках нет: значение статистики критерия Дарбина – Уотсона составляет 1,73, что больше верхнего критического значения 1,67, а значит, нулевую гипотезу об отсутствии автокоррелярии нет оснований отвергнуть. Гомоскедастичность остатков подтверждает тест Голдфелда – Квандта незначимостью вычисленной статистики: p-значение для неё равно 0,598, что позволяет принять нулевую гипотезу об отсутствии гетероскедастичности.

При выполнении этих трёх условий оценки МНК являются наилучшими в классе линейных несмещённых оценок в соответствии с теоремой Гаусса – Маркова. Однако остатки не нормально распределены: p-значение статистики Харке – Бера меньше уровня значимости.

Итак, устранено влияние сезонных колебаний на исходный временной ряд и построена адекватная модель тренда, описывающую тенденцию среднедушевых доходов. На графике исходных значений есть точки, которые сильно отличаются от прочих и соответствуют декабрю и январю каждого года. Их происхождение можно объяснить выдачей денежных премий работникам в конце года, из-за чего доход в виде заработной платы в декабре больше, чем в другие месяцы, и наличием выходной недели в начале года, за которую нет выплат. Данные значения не являются выбросами, но есть неотъемлемая характеристика сезонной компоненты. По этой причине сглаживание было проведено не с помощью медианы, преимущество которой состоит в устойчивости к выбросам, а с помощью средней, иначе было бы искажено адекватное описание ряда.

5.2. Модель Хольта – Уинтерса

В данном случае период сезонности m равен 12. Оценки параметров сглаживания α^*, β^*, γ^* определим путём минимизации суммы квадратов ошибок по следующей формуле:

$$MSE(\alpha^*, \beta^*, \gamma^*) = \min_{\alpha,\beta,\gamma} MSE = \min_{\alpha,\beta,\gamma} \sum_{t=1}^{n} (y_t - \widetilde{y}_t)^2. \tag{19}$$

По расчётам получены следующие оценки параметров: $\alpha^* = 0,9$, $\beta^* = 0,001$, $\gamma^* = 0,001$.

С учётом найденных значений параметров, коэффициенты мультипликативной модели Хольта – Уинтерса для рассматриваемого временного ряда приняли следующий вид:

$$\begin{aligned} \ell_t &= 0{,}9\frac{y_t}{s_{t-12}} + 0{,}1(\ell_{t-1} + b_{t-1}), \\ b_t &= 0{,}001(\ell_t - \ell_{t-1}) + 0{,}999 b_{t-1}, \\ s_t &= 0{,}001\frac{y_t}{\ell_{t-1}} + 0{,}999 s_{t-12}, \end{aligned} \tag{20}$$

где ℓ_t — общий уровень ряда, b_t — трендовая компонента, s_t — сезонная составляющая.

В процессе написания работы метод Хольта – Уинтерса был реализован на языке Python. Исходный код представлен в приложении В.

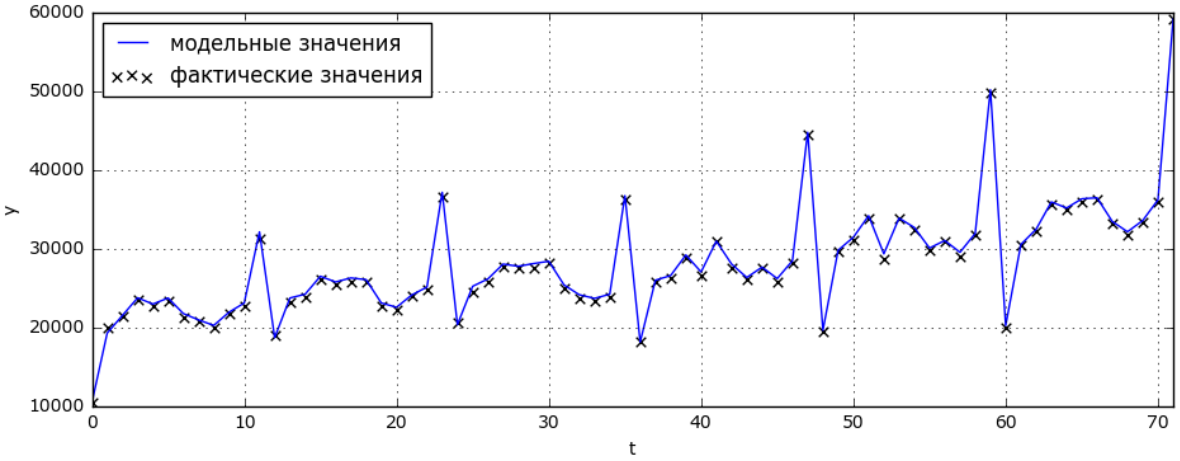


Рис. 9: Исходный ряд и сглаженный ряд методом Хольта – Уинтерса

Значения параметров сглаживания подобраны таким образом, что-

бы модель хорошо сглаживала исходный временной ряд, однако в этом случае может пострадать качество прогноза. Поэтому в зависимости от результатов на контрольной выборке их можно будет скорректировать.

Итак, успешно применено тройное экспоненциальное сглаживание на рассматриваемом временном ряде. Модельные и фактические значения представлены на рис. 9.

5.3. Сезонная интегрированная модель авторегрессии — скользящего среднего

Рассмотрим исходный временной ряд (рис. 5). Для того чтобы добиться его стационарности, нужно взять одну конечную разность и одну сезонную: после этого p-значение для вычисленной статистики Дики–Фуллера оказывается равным нулю, что позволяет отвергнуть нулевую гипотезу о наличии единичного корня, а значит, — о нестационарности ряда. Порядки авторегрессии и скользящего среднего определим по графикам автокорреляционной (АКФ) и частной автокорреляционной (ЧАКФ) функций разностного ряда (рис. 10).

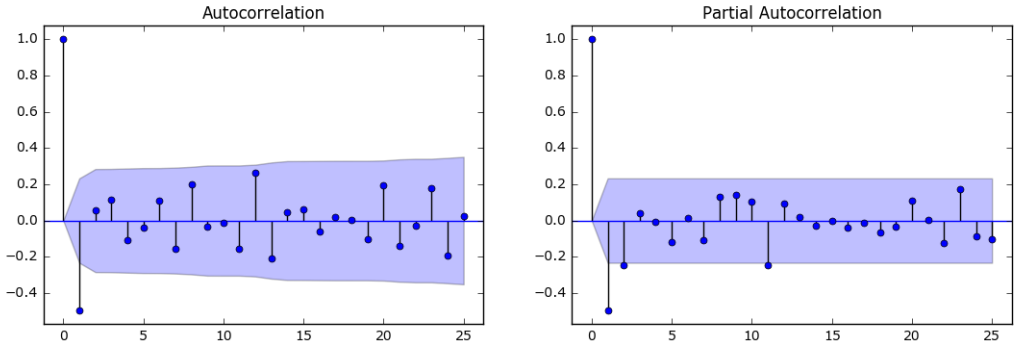


Рис. 10: АКФ и ЧАКФ для разностного временного ряда

Модель SARIMA$(1,1,2)(0,1,0)_{12}$ для рассматриваемого временного ряда в операторной форме имеет вид:

$$(1 - 0{,}7747L)(1 - L)(1 - L^{12})y_t = (1 + 1{,}4635L - 0{,}4943L^2)\epsilon_t. \qquad (21)$$

Коэффициенты модели определены с помощью метода наименьших квадратов. Оценивание модели проводилось для исходного ряда после

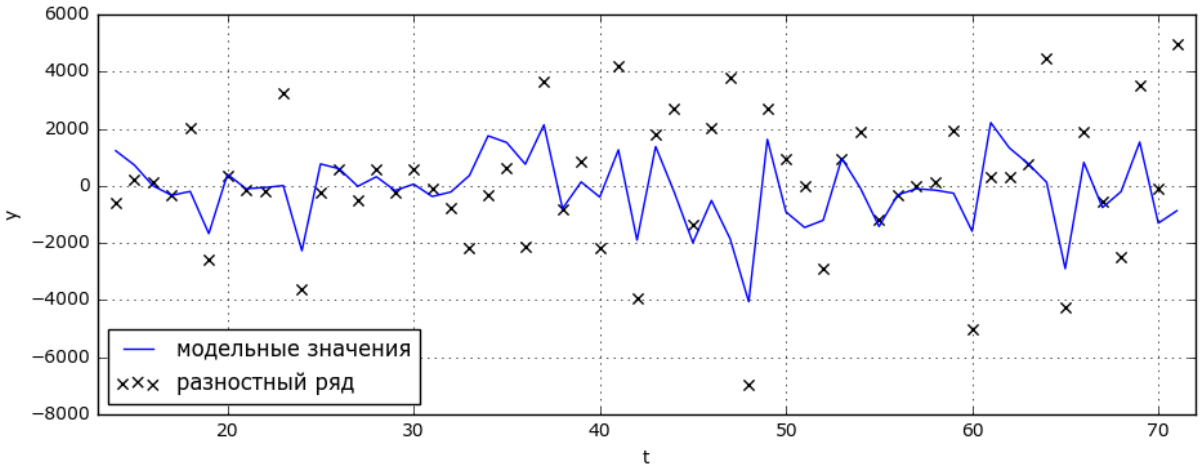


Рис. 11: Разностный ряд и модельные значения

взятия разностей. Коэффициент R^2 составил 0,30. Все коэффициенты регрессии являются значимыми: p-значения t-статистик меньше уровня значимости 0,05. Остатки модели удовлетворяют всем стандартным предположениям: результаты, полученные с помощью функций библиотеки *statsmodels*, представлены в таблице 14 приложения А. Также SARIMA$(1,1,2)(0,1,0)_{12}$ отличилась наименьшими значениями статистик информационных критериев Акаике и Шварца по сравнению с моделями других порядков (таблица 8).

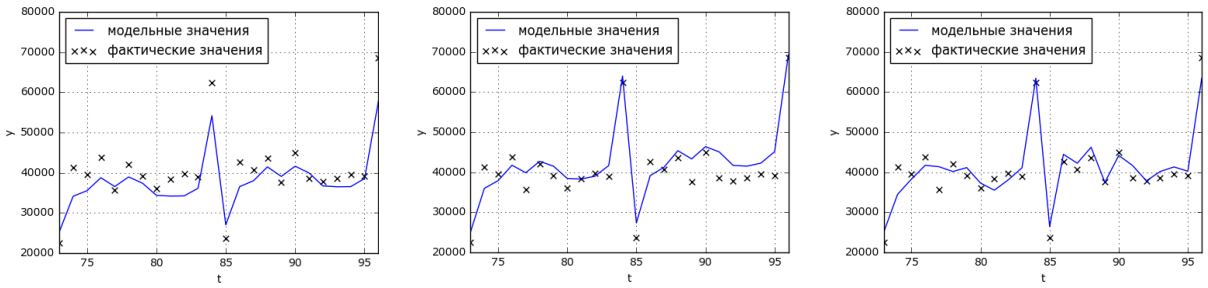
модель	значение AIC	значение BIC
SARIMA$(1,1,2)(0,1,0)_{12}$	18,08402	18,19059
SARIMA$(2,1,1)(0,1,0)_{12}$	18,09408	18,20161
SARIMA$(1,1,1)(1,1,0)_{12}$	18,22605	18,34531
SARIMA$(2,1,2)(0,1,0)_{12}$	18,12943	18,27280

Таблица 8: Сравнение моделей по информационным критериям

Итак, построена адекватная модель авторегрессии — скользящего среднего на основе временного ряда, который предварительно был приведён к станционарному виду с помощью взятия конечной и сезонной разностей.

6. Выбор математической модели с наилучшими прогностическими свойствами

Сравнение свойств моделей производится на контрольной выборке, которая в данном случае состоит из 24-х значений среднедушевого денежного дохода населения Санкт-Петербурга за период с января 2015 по декабрь 2016 года.



(a) Трендовая модель (b) Модель Хольта – Уинтерса (c) SARIMA$(1,1,2)(0,1,0)_{12}$

Рис. 12: Фактические значения контрольной выборки и рассчитанные значения с помощью моделей (a), (b), (c)

В качестве показателей ошибок прогнозирования временных рядов рассмотрим среднюю абсолютную ошибку в процентах (MAPE), квадратный корень из среднеквадратичной ошибки (RMSE), среднюю абсолютную ошибку (MAE) и медианную абсолютную ошибку (MdAE). Величина каждого из них показывает, насколько прогнозные значения близки к реальным.

Пусть $\widehat{y}_i$ — прогнозное значение в i-ый момент времени, y_i — фактическое значение, n — количество моментов времени, на которое был построен прогноз. Тогда

$$MAPE(y, \widehat{y}) = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \widehat{y}_i|}{y_i} \cdot 100\%, \tag{22}$$

$$RMSE(y, \widehat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}, \tag{23}$$

$$MAE(y, \widehat{y}) = \frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y}_i|, \tag{24}$$

$$MdAE(y, \widehat{y}) = median(|y_1 - \widehat{y}_1|...|y_n - \widehat{y}_n|). \quad (25)$$

Расчёты показателей для каждой построенной модели, проведённые с помощью библиотеки $scikit-learn$ языка Python, приведены в таблице 9. Очевидно, что наименьшие значения ошибок соответствуют модели с лучшими прогностическими свойствами. Заметим, что самую высокую точность прогнозирования на контрольной выборке продемонстрировала SARIMA$(1,1,2)(0,1,0)_{12} \approx 94\%$.

	MAPE, %	RMSE	MAE	MdAE
трендовая модель с учётом сезонности	8,47	4315,68	3548,93	2992,84
модель Хольта – Уинтерса	7,28	3236,39	2719,00	2481,10
SARIMA$(1,1,2)(0,1,0)_{12}$	5,92	2926,50	2379,06	1899,16

Таблица 9: Сравнение прогностических свойств

Итак, проведено сравнение прогностических свойств представленных математических моделей и выявлена модель с наилучшими свойствами — SARIMA$(1,1,2)(0,1,0)_{12}$. С помощью неё далее построим прогноз на 3 следующих шага. Выбранной модели немного уступает по точности модель Хольта – Уинтерса, которая, тем не менее, входит в разновидность одних из лучших способов прогнозирования временных рядов [4]. Трендовая модель с учётом сезонности также показала неплохой результат на контрольной выборке — за счёт того, что исходные данные имеют ярко выраженный линейный тренд и предсказуемую динамику в течение года, но данная модель является довольно грубым приближением реального процесса.

7. Построение прогноза

Построим прогноз на 3 шага с помощью модели (21) и доверительный интервал с вероятностью 95%.

Прогноз для каждого шага, $h = 1, 2, 3$, будем искать по формуле

$$(1 - 0{,}7747L)(1 - L)(1 - L^{12})\widehat{y}_{t+h} = (1 + 1{,}4635L - 0{,}4943L^2)\widehat{\epsilon}_{t+h}. \quad (26)$$

Пусть в (21) $(1 - 0{,}7747L)(1 - L)(1 - L^{12}) = A(L)$, $(1 + 1{,}4635L - 0{,}4943L^2) = b(L)$. Дисперсия ошибки прогноза на h шагов равна

$$\sigma^2_{e_{t+h}} = \widehat{\sigma}^2_{\varepsilon} \sum_{i=0}^{h-1} C_i^2, \quad (27)$$

где C_i находятся из $A(z)C(z) = b(z)$, $C(z) = \sum_{i=0}^{\infty} C_i z^i$, $C_0 = 1$ [10].

Тогда, так как остатки модели удовлетворяют стандартному предположению о нормальности распределения, доверительный интервал с вероятностью 95% для y_{t+h} можно найти из следующего соотношения:

$$|\widehat{y}_{t+h} - y_{t+h}| \leq 1{,}96\sqrt{\widehat{\sigma}^2_{\varepsilon} \sum_{i=0}^{h-1} C_i^2}. \quad (28)$$

шаг	прогнозное значение	нижняя граница	верхняя граница
97	26649,83	22433,87	30865,79
98	45471,77	40309,13	50643,41
99	43534,18	37340,58	49727,78

Таблица 10: Прогноз по модели SARIMA$(1,1,2)(0,1,0)_{12}$ и значения границ доверительного интервала

Значения отличаются от прогноза, который был построен в [13] с помощью модели Хольта – Уинтерса. Прогностические свойства модели можно улучшить, осуществив корретировку параметров с учётом данных контрольной выборки.

Таким образом, в январе 2017 года (97 шаг) среднедушевой доход

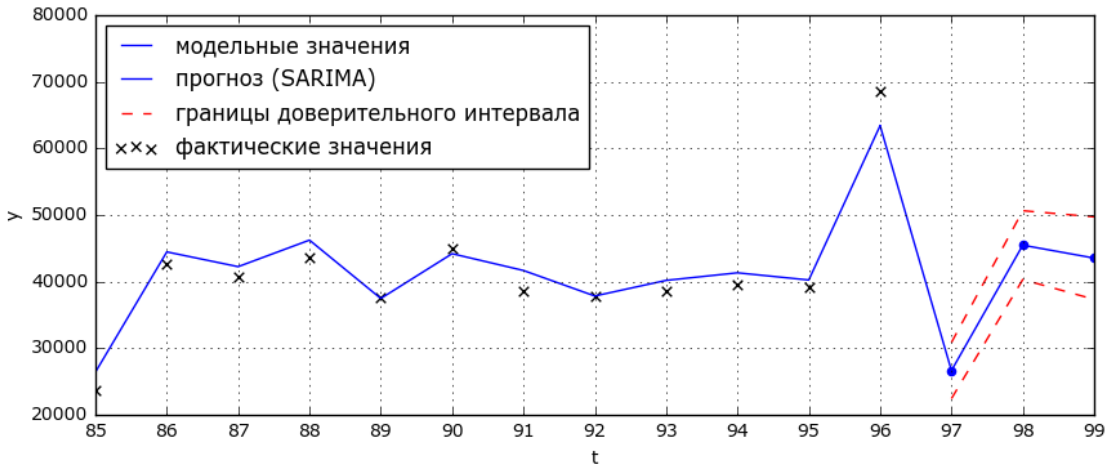


Рис. 13: Часть контрольной выборки с фактическими и модельными значениями, прогноз на 3 шага и границы доверительного интервала

населения Санкт-Петербурга составит 26649,83 руб. Интервальный прогноз с вероятностью 95% — от 22433,87 до 30865,79 руб. Низкое значение среднедушевого дохода для января по сравнению с другими объясняется учётом в модели сезонной компоненты исходного временного ряда.

8. Выводы

Итак, по результатам данной работы получена модель, которая наилучшим образом соответствует временному ряду $y_t, t = \overline{1,96}$, построенному на значениях среднедушевого денежного дохода населения Санкт-Петербурга. Модель SARIMA$(1,1,2)(0,1,0)_{12}$ адекватно описывает динамику данных, и с её помощью возможно прогнозирование будущих значений.

Прогнозные значения среднедушевого дохода на три первых месяца 2017 года представлены в таблице 11.

месяц 2017 г.	прогнозное значение	нижняя граница	верхняя граница
январь	26649,83	22433,87	30865,79
февраль	45471,77	40309,13	50643,41
март	43534,18	37340,58	49727,78

Таблица 11: Точечные и интервальные прогнозные значения

Как видно из рис. 14, тенденция к росту среднедушевого дохода сохранится.

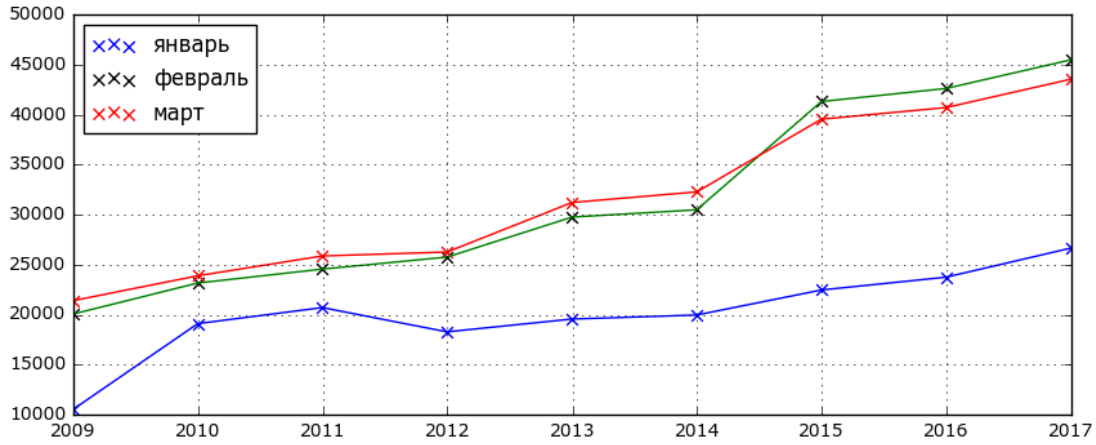


Рис. 14: Значения среднедушевого дохода в январе, феврале и марте 2009–2017 гг.

Однако интересно посмотреть, какова покупательская способность спрогнозированных значений в рублях относительно тех же периодов

прошлых годов. Для этого разделим найденные значения на индексы инфляции за январь, февраль и март соответственно в 2009–2016 гг.

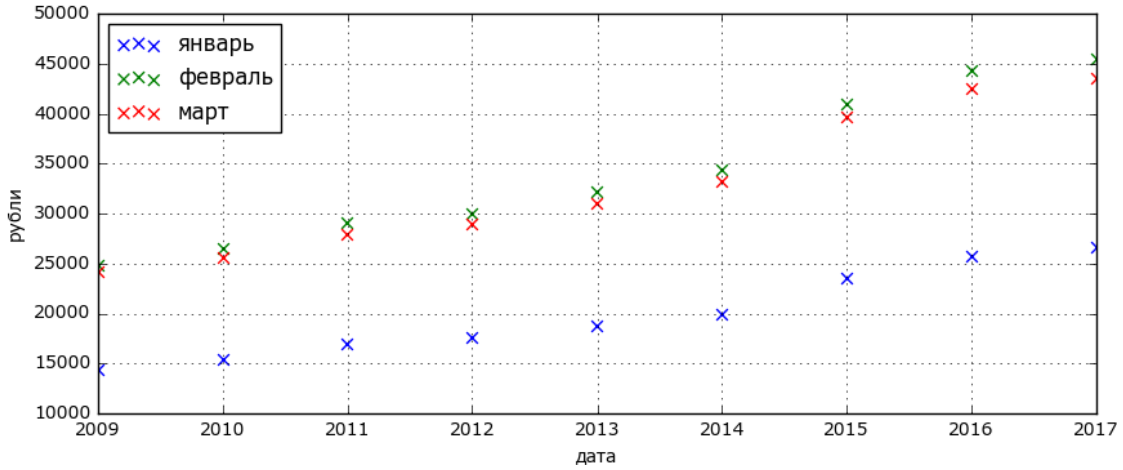


Рис. 15: Значения среднедушевого дохода в январе, феврале и марте 2017 г. в ценах января, февраля и марта 2009–2016 гг.

Согласно рис. 15, покупательская способность денег, которые составляет среднедушевой доход в феврале 2017 г. (45472 руб.), равна покупательской способности 35 тыс. руб. в 2014 г. и 25 тыс. руб. в 2009 г. Значит, за бóльший доход в 2017 году можно приобрести тот же набор товаров и услуг, что и за меньший несколькими годами ранее.

Сравним темп инфляции и изменение в процентах среднедушевого дохода в январе, феврале и марте 2010–2017 гг. относительно тех же периодов в 2009 году. Для расчёта темпа инфляции примем в качестве базового месяца январь 2009 г.

В 2017 году ожидается продолжение роста среднедушевого дохода, опережающего темп инфляции (рис. 16). Однако нужно учитывать, что ИПЦ, на основе которого рассчитывался темп инфляции по формуле (13), измеряет средний уровень цен на товары и услуги только потребительской корзины.

В разделе 3 было отмечено, что почти треть доходов жителей Санкт-Петербурга приходится на наиболее обеспеченную часть. Среднедушевой доход в размере меньше, чем 45 тыс. руб./мес, имеют 70% населения города, т.е. почти две трети жителей могут позволить себе менее 4-х наборов потребительской корзины в месяц[2]. Напряжённая ситуация

[2]Величина прожиточного минимума за 4 квартал 2016 г. в Санкт-Петербурге составила 10526,4

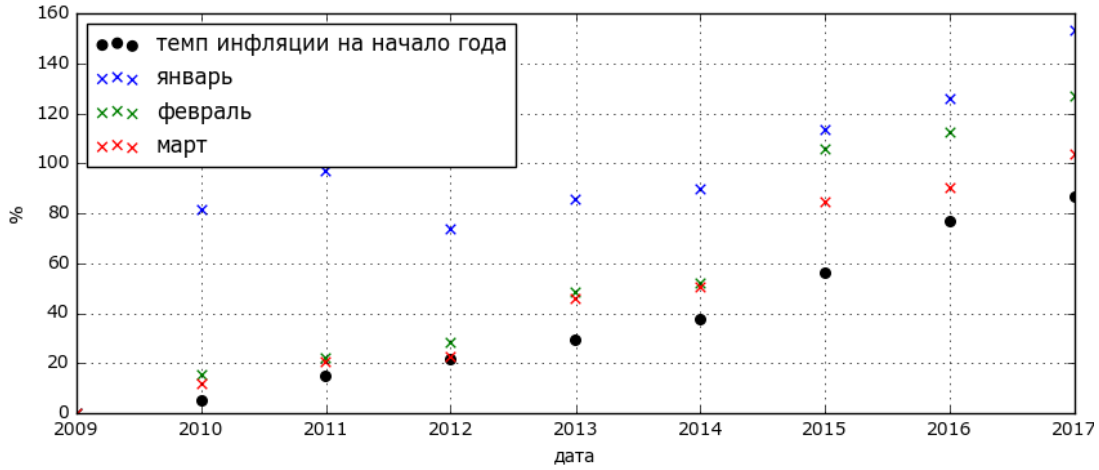


Рис. 16: Темп инфляции и процентное изменение среднедушевого дохода в 2009–2017 гг.

с дифференциацией доходов в стране, отдельных городах и регионах требует привлечения государственного участия в повышенном объёме путём налоговой политики и программ для малоимущих. Однако вместе с тем социальная политика государства должна быть гибким и тонким инструментом, чтобы перераспределительные процессы не привели к снижению деловой активности и эффективности наёмного труда [23].

Согласно построенной модели 16, среднедушевой доход зависит от значений среднемесячной заработной платы, прожиточного минимума и минимального размера оплаты труда. Недавно премьер-министр РФ поручил увеличить МРОТ до уровня прожиточного минимума в течение ближайших нескольких лет [18]. В соответствии с результатами, полученными в данной работе, реализация этого указа должна спровоцировать рост значения среднедушевого дохода и оказать благоприятное воздействие на уровень жизни населения.

руб./мес. на душу населения

Заключение

В данной работе были достигнуты следующие результаты:

- изучены три метода прогнозирования временных рядов: регрессионные модели, метод Хольта – Уинтерса, модели авторегрессии — скользящего среднего;
- построен временной ряд по значениям имеющихся данных и проведён его предварительный анализ;
- предложена спецификация регрессионной модели, включающая значимые экономические показатели, от которых зависит среднедушевой доход, на основе корреляционного анализа;
- построены три математические модели на основе исследуемого временного ряда: трендовая модель с учётом сезонности, модель Хольта – Уинтерса, сезонная интегрированная модель авторегрессии — скользящего среднего;
- выбрана лучшая модель на контрольной выборке по ошибкам прогноза — SARIMA$(1,1,2)(0,1,0)_{12}$;
- с помощью выбранной модели построен точечный и интервальный прогнозы на январь, февраль, март 2017 года.

По материалам работы был сделан доклад “Прогнозирование среднедушевого дохода населения Санкт-Петербурга” на конференции CPS’17.

Список литературы

[1] Alfares H.K. Nazeeruddin M. Electric load forecasting: literature survey and classifcation of methods // International Journal of Systems Science. — 2002. — Vol. 33. — P. 23–34.

[2] The CIA library: The World Factbook. Distribution of family income - Gini index [Электронный ресурс]. — URL: https://www.cia.gov/library/publications/the-world-factbook/fields/2172.html (online; accessed: 07.02.2017).

[3] Ceriani Lidia, Verme Paolo. The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini // The Journal of Economic Inequality. — 2012. — Vol. 10, no. 3. — P. 421–443. — URL: http://dx.doi.org/10.1007/s10888-011-9188-x.

[4] Makridakis S. Andersen A. Carbone R. Fildes R. Hibon M. Lewandowski R. Newton J. Parzen E. Winkler R. The accuracy of extrapolation (time series) methods: Results of a forecasting competition // Journal of Forecasting. — 1982. — Vol. 1. — P. 111–153.

[5] Makridakis S. Wheelwright S. Hyndman R. J. Forecasting: Methods and Applications. Third Edition. — New York: Wiley, 1998.

[6] P. Ebbes. A non-technical guide to instrumental variables and regressor–error dependencies // Quantile. — 2007. — no. 2. — P. 3–20.

[7] Paul F. Velleman David C. Hoaglin. Aplications, Basics, and Computing of Exploratory Data Analysis. — The Internet-First University Press, 2004.

[8] S.K. Prajakta. Time series Forecasting using Holt-Winters Exponential Smoothing // Kanwal Rekhi School of Information Technology Journal. — 2004. — P. 13.

[9] Spyros G. Makridakis Steven C. Wheelwright Victor E. McGee. Forecasting: Methods and Applications (Wiley Series in Management). — John Wiley and Sons Inc, 1983.

[10] В.П. Носко. Эконометрика. Книга 1. — Дело, Москва, 2014.

[11] Ведомости. Обратный результат продуктового эмбарго [Электронный ресурс]. — URL: https://www.vedomosti.ru/opinion/articles/2015/11/03/615335-obratnii-rezultat-produktovogo-embargo (дата обращения: 05.05.2017).

[12] Е.В. Такмакова. Модель денежных доходов населения России и определение направлений совершенствования политики регулирования доходов на этой основе // Национальные интересы: приоритеты и безопасность. — 2016. — С. 14–21. — URL: http://www.fin-izdat.ru/journal/national/detail.php?ID=69089.

[13] Е.В. Бабшукова. Прогнозирование среднедушевого дохода населения Санкт-Петербурга // Процессы управления и устойчивость. — 2017.

[14] Е.М. Четыркин. Финансовая математика. — Дело, Москва, 2000.

[15] ИТАР-ТАСС. Путин: о реализации майских указов нужно судить по улучшению качества жизни россиян [Электронный ресурс]. — URL: http://tass.ru/politika/1954107 (дата обращения: 05.05.2017).

[16] Магнус Я.Р. Катышев П.К. Пересецкий А.А. Эконометрика. Начальный курс: Учеб. — Дело, Москва, 2004.

[17] Официальный сайт Администрации Санкт-Петербурга. Статистика и аналитика [Электронный ресурс]. — URL: http://gov.spb.ru/helper/new_stat/ (дата обращения: 15.04.2017).

[18] РБК. Медведев поручил уравнять МРОТ и прожиточный минимум [Электронный ресурс]. — URL: http://www.rbc.ru/economics/02/05/2017/59082cc19a7947a75c42f126 (дата обращения: 09.05.2017).

[19] РИА Новости. Введение санкций против России и ответные шаги [Электронный ресурс]. — URL: https://ria.ru/infografika/20170317/1490100910.html (дата обращения: 05.05.2017).

[20] (Росстат) Федеральная служба государственной статистики. Регионы России. Социально-экономические показатели 2016. Статистический сборник. — 2016. — URL: http://www.gks.ru/bgd/regl/b16_14p/Main.htm.

[21] Социально-экономическое положение Санкт-Петербурга в январе-декабре 2016 года (экономический доклад в таблицах). — Петростат, 2017. — URL: http://petrostat.gks.ru/.

[22] Стратегия экономического и социального развития Санкт-Петербурга на период до 2030 года [Электронный ресурс]. — URL: http://spbstrategy2030.ru/ (дата обращения: 12.05.2017).

[23] Титов В.А. Климашина В.В. Статистический анализ социальной дифференциации населения РФ по величине среднедушевых доходов в современных условиях // Фундаментальные исследования. — 2016. — no. 3-1.

[24] Федеральная служба государственной статистики. Индексы потребительских цен по Российской Федерации в 1991 - 2017 гг. [Электронный ресурс]. — URL: http://www.gks.ru/free_doc/new_site/prices/potr/tab-potr1.htm/ (дата обращения: 08.03.2017).

[25] Федеральная служба государственной статистики. Коэффициент Джини (индекс концентрации доходов) [Электронный ресурс]. — URL: http://www.gks.ru/dbscripts/cbsd/dbinet.cgi?pl=2340003 (дата обращения: 06.02.2017).

[26] Федеральная служба государственной статистики. Уровень жизни населения [Электронный ресурс]. — URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/level/ (дата обращения: 07.02.2017).

[27] Ханк Д. Э. Уичери Д. У. Райте А. Дж. Бизнес-прогнозирование: пер. с англ. — Вильямс, 2003.

[28] Чистик О.Ф. Статистические методы инфляционных процессов в экономике России. — Вестник СГЭУ, 2009.

[29] Эксперт Online. Два года санкций: как изменилась жизнь в России [Электронный ресурс]. — URL: http://expert.ru/ural/2016/13/dva-goda-sanktsij-kak-izmenilas-zhizn-v-rossii/ (дата обращения: 05.05.2017).

A. Результаты оценивания моделей

variable	coefficient	std. error	t-statistic	prob.
X1(-1)	0,368721	0,097213	3,792913	0,0003
X2(-1)	1,600756	0,358810	4,461292	0,0000
X3(-1)	1,218756	0,451424	2,699803	0,0083

R-squared	0,925033	Akaike info criterion	17,93992
Adjusted R-squared	0,923403	Schwarz criterion	18,02057
S.E. of regression	1873,359	Durbin-Watson stat	1,966716
F-statistic	9130,000	Jarque-Bera stat	3,651320
Prob(F-statistic)	0,000000	Prob(JB)	0,161111
F-statistic (LM test)	4,488600	F-statistic (GQ test)	1,879489
Prob(F-stat LM test)	0,013900	Prob(F-stat GQ test)	0,018792

Таблица 12: Модель множественной регрессии

variable	coefficient	std. error	t-statistic	prob.
c	20491,12	417,2497	49,10996	0,0000
trend	207,3847	10,14319	20,44571	0,0000

R-squared	0,856565	Akaike info criterion	17,84377
Adjusted R-squared	0,854516	Schwarz criterion	17,90701
S.E. of regression	1788,713	Durbin-Watson stat	1,732227
F-statistic	418,0272	Jarque-Bera stat	11,70245
Prob(F-statistic)	0,000000	Prob(JB)	0,002876
F-statistic (LM test)	3,728468	F-statistic (GQ test)	0,917244
Prob(F-stat LM test)	0,029100	Prob(F-stat GQ test)	0,598691

Таблица 13: Трендовая модель

variable	coefficient	std. error	t-statistic	prob.
ar.L1	0,774703	0,164130	4,720067	0,0000
ma.L1	-1,463520	0,221317	-6,612765	0,0000
ma.L2	0,494272	0,209512	2,359161	0,0219

R-squared	0,304656	Akaike info criterion	18,08402
Adjusted R-squared	0,279370	Schwarz criterion	18,19059
S.E. of regression	1994,007	Jarque-Bera stat	6,195120
Durbin-Watson stat	1,922051	Prob(JB)	0,045159
F-statistic (LM test)	0,378111	F-statistic (White test)	1,126040
Prob(F-stat LM test)	0,687000	Prob(F-stat White test)	0,346700

Таблица 14: Модель SARIMA$(1,1,2)(0,1,0)_{12}$

B. Программная реализация

Процедура **removing_seasonality** устраняет мультипликативную сезонную составляющую с периодом **s_months** из временного ряда **data**. Переменная **data** имеет тип *pd.DataFrame*, **s_months** — целочисленный.

```
def removing_seasonality(data, s_months):
    roll_mean_s = data.rolling(window = s_months).mean()
    roll_mean_s.dropna(inplace = 1)

    if s_months % 2 == 0:
        roll_mean_s = roll_mean_s.rolling(window = 2).mean()
        roll_mean_s.dropna(inplace = 1)

    ts_seas = data.copy()
    ts_seas.values[:] = np.nan

    for i in range(len(roll_mean_s.values)):
        ts_seas.values[i + s_months % 2] = \
            data.values[i + s_months % 2] / roll_mean_s.values[i]

    ts_av_seas = np.ones([s_months, 1])
    k = len(data.values) - s_months
    for i in range(s_months):
        s = 0
        for j in range(int(k / s_months)):
            s += ts_seas.values[i + s_months % 2 + s_months * j]
```

```
        ts_av_seas[i] = s / (int(k / s_months) + 1)

    ts_norm_av_seas = np.ones([s_months, 1])
    for i in range(s_months):
        ts_norm_av_seas[i] = ts_av_seas[i] / sum(ts_av_seas) * \
            s_months

    c = -s_months%2
    for i in range(len(data.values)):
        data.values[i] = data.values[i] / ts_norm_av_seas[c]
        if c == s_months - 1:
            c = 0
        else:
            c += 1
```

Функция **holt_winters_smoothing** возвращает массив из **n_preds** прогнозных значений по модели Хольта – Уинтерса с параметрами **alpha**, **beta**, **gamma**, построенной на основе временного ряда **data** с периодом сезонности **s_months**. **s_months** и **n_preds** — переменные целочисленного типа, **alpha**, **beta**, **gamma** — типа *float*; **data** — *np.array*.

```
def holt_winters_smoothing(data, s_months, alpha, beta, gamma, n_preds):
    result = []

    seas = {}
    seas_av = []
    n_seas = int(len(data) / s_months)
    for i in range(n_seas):
        seas_av.append(sum(data[s_months * i : s_months * i + \
            s_months])/float(s_months))
    for j in range(s_months):
        sum_of_vals = 0.0
        for k in range(n_seas):
            sum_of_vals += data[s_months * k + j] / seas_av[k]
        seas[j] = sum_of_vals / n_seas

    for i in range(len(data) + n_preds):
        if i == 0:
            smooth = data[0]

            s = 0.0
```

```
            for j in range(s_months):
                s += float(series[j + s_months] - data[j]) / s_months
            trend = s / s_months

            result.append(data[0])
            continue

        if i >= len(data):
            c = i - len(data) + 1
            result.append((smooth + c * trend) * seas[i % s_months])
        else:
            val = data[i]
            last_smooth, smooth = smooth, alpha * (val / seas[i % \
                s_months]) + (1 - alpha) * (smooth + trend)
            trend = beta * (smooth - last_smooth) + (1 - beta) * trend
            seas[i % s_months] = gamma * (val / smooth) + (1 - gamma) * \
                seas[i % months]
            result.append((smooth + trend) * seas[i % s_months])

    return result
```

С. Глоссарий

Среднедушевой денежный доход (СДД) — отношение общей суммы денежного дохода к *численности населения* (ЧЗ).

Среднемесячная номинальная заработная плата (СНЗП) — отношение фонда заработной платы, начисленного за месяц, к среднесписочной численности работников.

Прожиточный минимум (ПМ) — стоимостная оценка потребительской корзины, а также обязательные платежи и сборы.

Минимальный размер оплаты труда (МРОТ) — законодательно установленный минимум оплаты труда в месяц.

Число занятых (ЧЗ) — количество людей, выполнявших в рассматриваемый периодоплачиваемую наёмную работу.

Индекс потребительских цен (ИПЦ) — показатель инфляции вэкономике, отражающий соотношение цены стандартной корзины товаров и услугс ценой той же корзины в предыдущем периоде.

Численность населения (ЧН).