Ekonomi-tek Volume / Cilt: 7 No: 1 Ocak / January 2018

Contents / İçindekiler

Contributors / Katkı Yapanlar

Aykut Attar
Recep Kök
Barış Soybilgen
Aysıt Tansel
Hüseyin Taştan
Ercan Uygur

Editor's Introduction

This issue starts the seventh volume of our *Ekonomi-tek* journal, and it contains two papers. One is a work on a machine-learning approach to inflation forecasting, while the other looks at the impact of minimum wages on the distribution of wages. Both subjects have often been debated recently.

The first paper is by Bige Küçükefe, of Namık Kemal University, and it outlines an interesting study of applied machine-learning models. More specifically, the author employs "supervised" machine-learning algorithm models for short-term inflation forecasting in Turkey by making use of inflation expectations surveys conducted by the Central Bank of Turkey (CBT).

The machine-learning models are implemented in Python software taken from the Scikit-learn library, an open-source software. Quarterly data go into the estimation procedures. Comparison of the forecast performances shows that the proposed machine-learning methods not only improve the accuracy of the surveyed forecasts but also outperform the official forecasts of the CBT and a univariate model.

The second paper in this issue is by Selin Pelek, of Galatasaray University. She investigates the effects of the minimum wage on wage distribution in Turkey, based on the micro data of the Household Labor Force Surveys (HLFS) for 2003 and 2005, provided by TURKSTAT. The sample used in the paper includes full-time wage earners in non-agricultural activities among the working-age population (those aged 15 to 65) who declare a net positive salary in the reference month. Part-time workers are excluded.

Her findings suggest that a minimum-wage hike in 2004 was accompanied by a significant reduction in wage inequality, especially among formal wage earners. Changes in the minimum wage compressed wage distribution in Turkey between 2003 and 2005, and wage inequality clearly improved over the period. The author argues that the driving force of this lessening of wage inequality is the rise of wages in the lower tail of wage distribution.

We look forward to presenting you with other interesting papers in our future issues.

Ercan Uygur
Editor
Ekonomi-tek

Editörün Sunuşu

Ekonomi-tek dergimizin bu sayısı yedinci cildi başlatmaktadır ve iki makale içermektedir. Birisi enflasyon öngörüsünde makina-öğrenimi (machine-learning) yaklaşımı konusunda bir çalışma, diğeri ise asgari ücretin ücret dağılımına etkisini inceleyen bir araştırmadır. Bunlar, yakın zamanda tartışılan konulardır.

Birinci makale, Namık Kemal Üniversitesi'nden Bige Küçükefe'nindir ve yazar bu makalede makina-öğrenimi modelleri konusunda ilginç bir çalışma sunmaktadır. Daha açık olarak, yazar, Türkiye Cumhuriyet Merkez Bankası (TCMB) tarafından yapılan enflasyon beklentileri anketlerinin sonuçlarından yararlanarak, kısa vadeli enflasyon öngörüleri için "denetlenmiş" ("supervised") makina-öğrenimi modelleri tahmin etmektedir.

Makalede makina-öğrenimi modelleri, bir açık-kaynak yazılımı olan Python paketinde yer alan Scikit-öğrenme yaklaşımı ile uygulanmaktadır. Tahmin yöntemlerinde üç-aylık veriler kullanılmıştır. Modellerin öngörüleri karşılaştırıldığında, önerilip uygulanan makina-öğrenimi modelleri yalnızca anket öngörülerini iyileştirmemiş, ayrıca TCMB'ninki de dahil olmak üzere diğer model öngörülerinden daha iyi performans göstermiştir.

Bu sayının ikinci makalesi Galatasaray Üniversitesi'nden Selin Pelek'indir. Bu makalede yazar, TÜİK'in 2003 ve 2005 Hane Halkı İşgücü Anketi (HHİA) verilerine dayanarak, Türkiye'de asgari ücretteki değişmelerin ücret dağılımına etkilerini araştırmaktadır. Çalışmada kullanılan örneklem, yalnızca tarım dışı faaliyetlerde tam-zamanlı ve çalışma çağındaki (15 ile 65 yaş aralığındaki) çalışanlardan ve anket ayında net bir ücret alanlardan oluşmaktadır. Yarı zamanlı çalışanlar kapsam dışında tutulmuştur.

Makalenin bulguları, asgari ücretin 2004'te sıçraması ile birlikte ücret eşitsizliğinde, özellikle kayıtlı ücret eşitsizliğinde, önemli bir azalma olduğunu ifade etmektedir. Asgari ücret değişmeleri, 2003 ile 2005 arasında ücret dağılımının yayılımını azaltmış ve bu dönemde eşitsizlikte önemli bir düzelme olmuştur. Yazara göre, ücret eşitsizliğinde azalmanın gerisindeki asıl etken, düşük ücretlerin görece daha hızlı yükselmesidir.

Sizlere gelecek sayılarda başka ilginç makaleler sunmayı diliyoruz.

Ercan Uygur
Editör
Ekonomi-tek

Ekonomi-tek Volume / Cilt: 7 No: 1 Ocak / January 2018, 1-16

Forecasting Inflation Using Summary Statistics of Survey Expectations: A Machine-Learning Approach

Bige Küçükefe[1]

Abstract

This paper aims to produce more accurate short-term inflation forecasts based on surveys of expectations by employing machine-learning algorithms. By treating inflation forecasting as an estimation problem consisting of a label (inflation) and features (summary statistics of surveys of expectations data), we train a suite of machine-learning models, namely, Linear Regression, Bayesian Ridge Regression, Kernel Ridge Regression, Random Forests Regression, and Support Vector Machines, to forecast the consumer-price inflation (CPI) in Turkey. We employ the Time Series Cross Validation Procedure to ensure that the training data exclude forecast horizon data. Our results indicate that these machine-learning algorithms outperform the official forecasts of the Central Bank of Turkey (CBT) and a univariate model.

JEL Codes: C82, E31

Keywords: Machine learning, forecast evaluation, inflation forecasting, surveys of expectations, summary statistics

[1] Namık Kemal University, bkucukefe@nku.edu.tr

1. Introduction

Applications of machine-learning algorithms in many fields have gained momentum in recent years. In this study, we explore the use of machine-learning methods to improve the accuracy of surveys of inflation expectations as a viable alternative to existing inflation-forecasting models.

Univariate models, expectation surveys, and Phillips curve models have been widely used to forecast future inflation. Comparison across different forecasting models is difficult due to differences in data, sample periods, and country-specific conditions. Debate over model performances, however, has attracted many researchers over the years. A comprehensive study by Öğünç et al. (2013) used a collection of econometric models that include univariate models, decomposition-based approaches, a Phillips curve motivated time-varying parameter model, a suite of VAR and Bayesian VAR models, and dynamic-factor models to forecast short-term inflation in Turkey. Their result revealed that a combination of these models leads to a reduction in forecast error.

In a similar approach, Kapetanios et al. (2008) argued that a single model is outperformed by combinations of various models. A milestone work by Atkeson and Ohanian (2001) compared the performance of a naïve moving-average model with a series of Phillips curve forecasting models and argued that the former performed better than the latter. However, a later work by Stock and Watson (2007) found that Phillips curve methods performed better for the period 1970-83, and the results of Atkeson and Ohanian were specific to the period 1984-99.

Surveys of inflation expectations offer an alternative approach to inflation forecasting. The superior forecasting performance of surveys has been highlighted by several researchers. In particular, Grothe and Meyler (2015) argued that short-term inflation expectations derived from survey and market data for the euro area and the United States were informative predictors of future inflation developments. Similarly, Ang et al. (2007) show that inflation expectations from survey data beat a wide variety of forecasting models that include time-series ARIMA models; regressions using real activity data motivated from the Phillips curve, and term structure models that include linear, non-linear, and arbitrage-free specifications.

A similar study by Gil-Alana et al. (2012) revealed that survey-based expectations outperform standard time-series models in US quarterly inflation out-of-sample predictions. Furthermore, Altuğ and Çakmaklı (2016) formulated a statistical model of inflation that combines data from survey expectations of inflation and argued that the model with survey expectations yields superior predictive performance to the model lacking them, as well as popular benchmarks, such as the backward-looking Phillips curves and the naive forecasting rule.

The appeal of machine learning stems from its ability to uncover complex structures hidden in large datasets without explicit programming. Classification and regression are two central applications of "supervised" machine learning. Both involve making estimations of an unknown target from a set of known features by applying various algorithms, such as support-vector machines, random forests, and deep neural networks.

We employ "supervised" machine-learning models for short-term inflation forecasting by using the Central Bank of Turkey (CBT)'s survey of inflation expectations. We treat summary statistics of survey data as features without time stamp for a supervised machine-learning problem for which the label to estimate is the future inflation. From the viewpoint of machine learning, this is a standard regression problem. We implemented the machine- learning models in Python software by using the Scikit-learn (Pedregosa et al., 2012) library, which is opensource software. The codes are available from the author upon request.

The remainder of this paper is organized as follows: Section 2 briefly explains the machine- learning methods we employed to improve the forecasting accuracy of inflation-expectation surveys. Section 3 presents the data and methodology by which we compared the forecasts. Section 4 discusses the forecast performance of the machine-learning algorithms and compares their accuracy with the survey data, the CBT official forecasts, a naive MA method, and a univariate model, and Section 5 contains the conclusion.

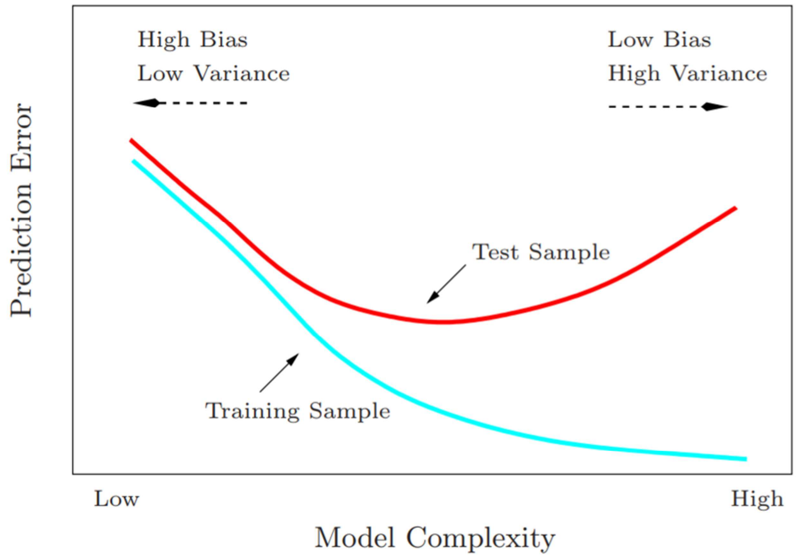
2. Machine-Learning Models

The success of machine learning primarily lies in its ability to discover unknown complex structures hidden in datasets. The common principle that underlies a supervised machine-learning model is to learn a target function (*f*) that maps input variables (*X*) to an output variable (*Y*). Without defining an explicit solution methodology that may not even exist, supervised machine-learning models "learn" from sample data and make estimations for out-of-sample data primarily for the purpose of binary classification, multiclass classification, and regression, among many other applications.

Dealing with under-fitting and over-fitting problems is important when selecting and tuning supervised machine models. The under-fitting problem occurs if the model doesn't represent the sample adequately. The over-fitting, on the other hand, occurs if a model fits best on sample data and fails on out-of-sample data.

Therefore, bias-variance trade-off in machine learning is closely related to model complexity. As the model complexity increases, the variance tends to go up, and the bias tends to decrease—and vice versa [Mullainathan and Spiess, 2017]. Fig.1 depicts the prediction error as a function of model complexity with bias-variance combinations for a machine-learning model [Hastie et al., 2017, p. 38)].

Figure 1. Test and Training Error as a Function of Model Complexity



The performance evaluation of a supervised machine-learning model involves dividing a dataset into test and training sets. The training set is used to train the model, which is being evaluated against the test dataset. In our study, the summary statistics of inflation-expectation surveys (the features) and the actual inflation (label), without time information, are divided into two groups: the training and forecast sets.

We use five types of machine-learning algorithms: Linear Regression, Random Forests, Support-Vector Machines, Bayesian Ridge Regression, and Kernel Ridge Regression. Before introducing each of these models, we first present some fundamental aspects of supervised machine learning. See Hastie et al. (2009) and James et al. (2013) for details about the models.

2.1. Linear Regression

The underlying assumption of linear regression is that there is approximately a linear relation between response (Y) and variables (X) in a dataset with only quantitative values. We can write this relationship as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon \tag{1}$$

Where $\beta_0,\ \beta_1, \dots, \beta_p$ are the regression coefficients and ϵ is the error term. For p = 1, (1) transforms into a simple linear regression. The regression coefficients in (1) are estimated by using the least-squares approach in the formula:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta} x_2 + \cdots + \hat{\beta}_p x_p + \epsilon \tag{2}$$

such that the sum of squared residuals (RSS) is minimum. RSS is defined as

$$RSS = \sum_{i=1}^{n} \left(y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_p x_{ip} \right) \right)^2 \tag{3}$$

2.2. Random Forests Regression

Random Forests (Breiman(2001)) are a collection of simple decision-tree predictors. Each decision tree in a Random Forest can produce a response for a set of input values. An algorithm determines the split points, splitting variables, and topology of a decision tree. The tree grows after solving for each split until a tuning parameter (tree size), which controls the model's complexity, is reached. Random Forests aim to improve the predictive performance of decision trees by aggregating many of them. Furthermore, Random Forests overcome the problem of the strong predictor estimation in the bagged-trees approach by allowing for a smaller number of randomly selected predictors for each split. The average of the predictions from all the trees is the ensemble estimation of the Random Forest model. Growing a regression tree requires an algorithm that automatically decides on the splitting variables and split points. The response of a model consisting of M regions $R_1, R_2, \dots, R_M$ is determined by

$$f(x) = \sum_{i=m}^{M} c_i I(x \in R_m) \tag{4}$$

for each region. Starting with all data, the algorithm first considers the pair of half planes in terms of the splitting variable $j$and the split point s as

$$R_1(j,s) = \{X | X_j \leq s\} \text{ and } R_2(j,s) = \{X | X_j > s\} \tag{5}$$

and seeks the splitting variable j and split point s that minimize

$$\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_1(j,s)} (y_i - c_2)^2 \tag{6}$$

For any value of j and s, the inner minimization is solved by

$$\widehat{c_1} = average\big(y_i \big| x_i \in R_1(j,s)\big),$$

$$\widehat{c_2} = average\big(y_i \big| x_i \in R_2(j,s)\big), \tag{7}$$

By scanning through all the inputs, we determine the best pair of (j,s) is determined. Then the same process is repeated on resulting regions of data until the tree grows an optimal size according to a tuning parameter that determines the model's complexity.

One common strategy is to grow a large tree with some minimum node size, such as 5. Then this large tree is pruned using a *cost-complexity pruning* procedure. First, define a sub-tree $T \subset T_0$ of any tree obtained by pruning T_0, that is, collapsing any number from its non-terminal nodes. Letting

$$N_m = \#\{x_i \in R_M\},$$

$$\widehat{c_m} = \frac{1}{N_m} \textstyle\sum_{x_i \in R_m} y_i, \tag{8}$$

$$Q_m(T) = \frac{1}{N_m} \sum_{x_i \in R_m} (y_i - \widehat{c_m})^2$$

where m is the index of terminal nodes in region R_M, we define the cost-complexity criterion

$$C_\alpha(T) = \textstyle\sum_{m=1}^{|T|} N_m Q_m(T) + \alpha|T| \tag{9}$$

where $|T|$ denotes the number of terminal nodes in T, and α is the tuning parameter ($\alpha \geq 0$), which governs the bias-variance tradeoff in the model. The idea is to find the sub-tree $T_\alpha \subseteq T_0$ to minimize $C_\alpha(T)$ in (9) for each α. Small values of α yield larger trees Tα and vice versa. The full tree T_0 is returned with α = 0.

For each α there is a unique smallest sub-treeT_α that minimizes (9). To find T_α, the algorithm collapses the internal mode that produces the smallest per-node increase in $\sum_m N_m Q_m(T)$ until it produces the single-node (root) tree. This method is called *weakest link pruning* and gives a finite sequence of sub-trees that contain T_α. The tuning parameter α s estimated by choosing the value $\widehat{\alpha}$ to minimize the cross-validated sum of squares, from which $T_{\widehat{\alpha}}$ is the final tree.

Decision trees yield high-variance, low-bias output. One way to reduce high variance is to use the bagging technique, which simply fits the same regression tree many times to bootstrap-sampled versions of the training dataset and compute the average. Since each tree generated in the bagging of B trees is identically distributed, the variance of the average is given by

$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2 \tag{10}$$

The second term disappears in (10) as B increases (more trees), and, hence, the benefits of averaging get weaker due to the size of the correlation of pairs in the bagged trees. The Random Forests aim to solve this problem by randomly selecting the input variables in the tree-growing process. Before each split, m variables are selected randomly from the input variables (p) as candidates for splitting. After growing B trees with this procedure, the Random Forest regression predictor is defined as

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T(x; \Theta_b) \tag{11}$$

where Θ_b characterizes the bth Random Forest tree in terms of split variables, cutpoints at each node, and terminal-node values.

2.3. Support-Vector Regression

The idea behind Support-Vector Machines (SVM) (Vapnik (1995)) is to find hyperplanes that separate different classes in a training dataset. SVM Regression (SVR) is a form of SVM with a numerical dependent variable instead of a categorical one. SVR relies on kernel functions (linear, polynomial, radial basis, etc.) to construct optimal hyperplanes.

Kernel function transforms the training data from nonlinear space to linear space. This transformation allows SVR to find an optimum hyper plane. Mapping back to the original space completes the algorithm. For a linear-regression model, $f(x) = x^T\beta + \beta_0$, and estimation of β and β_0 are possible through minimization of

$$H(\beta, \beta_0) = \sum_{i=1}^{N} V\big(y_i - f(x_i)\big) + \frac{\lambda}{2}||\beta||^2 \tag{12}$$

where

$$V_t = \begin{cases} 0, & if|r| < \epsilon, \\ |r| - \epsilon, & otherwise \end{cases}$$

The solution functions for $\hat{\beta}$ and $\hat{\beta}_0$ that minimize H have the form

$$\hat{\beta} = \sum_{i=1}^{N}(\hat{\alpha}_i{}^* - \hat{\alpha}_i)\, x_i \tag{13}$$

$$\hat{f}(x) = \sum_{i=1}^{N}(\hat{\alpha}_i{}^* - \hat{\alpha}_i)\, \langle x, x_i\rangle + \beta_0 \tag{14}$$

Where $\widehat{\alpha}_i$ and $\widehat{\alpha}_i{}^*$ are called the support vectors, which solve the quadratic optimization problem

$$\min_{\alpha_i,\alpha_i{}^*} \epsilon \sum_{i=1}^{N} (\alpha_i{}^* + \alpha_{i,}) - \sum_{i=1}^{N} y_i(\alpha_i{}^* - \alpha_i) + \frac{1}{2} \sum_{i,i'=1}^{N} (\alpha_i{}^* - \alpha_i)$$

$$(\alpha_{i'}^* - \alpha_{i'})\langle x, x_i{}' \rangle \tag{15}$$

subject to

$$0 \le \alpha_\mathrm{i}, 0 \le \alpha_\mathrm{i}{}^* \le 1/\lambda,$$

$$\sum_{i=1}^{N} (\alpha_i{}^* - \alpha_i) = 0,$$

$$\alpha_i \alpha_i{}^* = 0$$

2.4. Kernel Ridge Regression

Kernel Ridge Regression (KRR) (Hsiang (1975) is a form of linear regression. KRR imposes a penalty on the size of the regression coefficients, which are estimated by obtaining $\hat{\beta}^{ridge}$ values that minimize

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta^2 = RSS + \sum_{j=1}^{p} \beta_j{}^2 \tag{16}$$

where $\lambda \ge 0$ is a tuning parameter to be calculated separately. Limiting the size of the regression coefficients alleviates the high variance problem caused by the large coefficients of the correlated variables in a model. Writing the criterion (16) in matrix form,

$$RSS(\lambda) = (y - \boldsymbol{X}\beta)^T (y - \mathbf{X}\beta) + \lambda \beta^T \beta \tag{17}$$

we estimate the regression parameters as

$$\widehat{\beta}^{ridge} = (\boldsymbol{X}^T \boldsymbol{X} + \lambda \mathbf{I})^{-1} \boldsymbol{X}^{\boldsymbol{T}} y \tag{18}$$

where I is the identity matrix with size $p \times p$. The choice of quadratic penalty in (17) makes the KRR solution a linear function of y. If the inputs are orthonormal, the KRR estimates are a scaled version of the least-squares estimates ($\hat{\beta}^{ridge} = \hat{\beta}/(1 + \lambda)$).

2.5. Bayesian Ridge Regression

Bayesian regression introduces uninformative priors over the hyper parameters of the model. The output y is assumed to be Gaussian distributed around X, w such that:

$$p(y|X, w, \alpha) = \mathcal{N}(w|X, w, \alpha) \tag{19}$$

Where α is treated as a random variable and estimated from data. In a probabilistic model estimated by Bayesian Ridge Regression, the prior for the parameter w is given by a spherical Gaussian:

$$p(w|\lambda) = \mathcal{N}(w|0, \lambda^{-1}\mathbf{I}_{\boldsymbol{p}}) \tag{20}$$

The priors over α and λ are gamma distributions (the conjugate prior for the precision of the Gaussian normal distribution). The parameters w, α, and λ are estimated jointly in the model.

3. Data and Methodology

The Central Bank of Turkey (CBT) conducts a survey each month to monitor the expectations of experts from the financial and real sectors. The questionnaire includes short-term inflation forecasts (current month, next month, and two months ahead) in addition to many other expectations of economic variables, such as exchange and interest rates. The survey reports also contain summary statistics that consist of mode, median, minimum, arithmetic mean, and maximum and minimum values. We obtained the survey of expectations data from the CBT and the Consumer-Price Inflation (CPI) data from TURKSTAT. The data cover the period from August 2001 to December 2017, for which we produced monthly inflation forecasts for three horizons: current month (h=1), next month (h=2), and two months ahead (h=3).

The Time-Series Cross-validation (TSCV) procedure uses the past data only for the training of the machine-learning models without any information about the forecast horizon or beyond, thereby producing out-of-sample forecasts. We used an expanding-window rather than the rolling-window estimation procedure to permit the use of more data for the learning process. Expanding the window-procedure estimates model on a sample running from $1, 2, \dots, t$ and we produced forecasts of variables at date $t + h : h > 0$.

The performance metric we employed to evaluate the forecasts by the machine-learning models is the Relative Root Mean Square Error (R-RMSE), which is calculated by dividing the model forecast RMSE by the survey forecast RMSE:

$$R - RMSE = \frac{\sqrt{\sum_{i=1}^{N}\left(\pi_i - f_i^m\right)^2}}{\sqrt{\sum_{i=1}^{N}\left(\pi_i - f_i^s\right)^2}} \quad (21)$$

where π is the actual inflation, f^m is the model forecast, and f_s is the arithmetic mean of survey expectations. A relative RMSE value less than unity indicates that the model improved on the surveys' forecasts and vice versa.

Furthermore, we computed the Empirical Cumulative Distribution Function (ECDF) for Bootstrap RMSEs of the forecasts. ECDF is defined as:

$$\text{ECDF}\ (RMSE) = \frac{\text{number of elements in the sample} \leq RMSE}{N} \quad (22)$$

where N is the total number of elements in the sample. ECDF used the data from the Bootstrap replicates of the forecasts and their RMSEs to draw statistical inferences. We used 10,000 bootstrap replicates with replacements.

We compare the forecast performance of the machine learning models with a suite of forecasts that consist of survey data, the CBT forecasts, a time-series forecasting model (TBATS), and a naive moving-average model. TBATS is a state space-modeling framework (de Livera et al. (2011)) for univariate time series forecasting with complex seasonal patterns. This trigonometric framework has both linear and nonlinear time-series modeling capacity with single seasonality, multiple seasonality, high period, and non-integer seasonality. TBATS incorporates Box-Cox transformations, Fourier representations with time-varying coefficients, and ARMA error correction. We used the R implementation of the TBATS model to compute the point forecasts for the next three months, starting from the first month of the respective quarter, from 2016-Q4 to 2017-Q4.

The CBT doesn't publish its monthly inflation forecasts. Hence, we use the quarterly inflation forecasts for comparison. Although quarterly inflation forecasts are not provided separately in the CBT's inflation reports, one can calculate them by solving the following equation analytically for f_Q.

$$(1 + \pi_{9m})\left(1 + f_Q\right) = 1 + f_{12m} \quad (23)$$

where π_{9m} is the actual past nine months' inflation and f_{12m} is the CBT's yearly inflation forecast at the end of the quarter that is available in the CBT's inflation report.

Calculating the quarterly inflation forecasts makes use of the next month and two months ahead forecasts from the models, such that the next quarter inflation forecast ($\hat{\pi}_{Qm}$), starting with the current month m , is given by

$$\hat{\pi}_{Qm} = \left(1 + f_{m-1}^{h(2)}\right)\left(1 + f_{m}^{h(2)}\right)\left(1 + f_{m}^{h(3)}\right) - 1 \tag{24}$$

where $f_m^{h(n)}$ is the model forecast in month m for horizon n.

As a benchmark, we employ a naive Moving Average (MA) model, in which the next quarter's inflation is equal to the arithmetic mean of the past four quarters' inflation rates (Atkeson and Ohanian, 2001).

$$\hat{\pi}_{t+Q} = \frac{1}{4}\left(\pi_{t-Q_1} + \pi_{t-Q_2} + \pi_{t-Q_3} + \pi_{t-Q_4}\right) \tag{25}$$

Even though it is well established that a naive MA model does not perform well for short time horizons for economies lacking stable inflation dynamics, the model nonetheless serves as a benchmark.

The results we obtained with the default parameters of the machine-learning models in the Scikit-learn library are satisfactory to the extent that they demonstrate the effectiveness of the proposed approach.

4. Forecasting Performance Evaluation

The forecast period covers five quarters, from 2016-Q4 to 2017-Q4. We evaluate the forecast performance of the machine-learning models, the surveys, the CBT, the TBATS model, and a naive MA model serving as a benchmark. The ECDF of Bootstrap mean of RMSEs offers useful information for comparing forecasting performances.

Table 1 shows the probability of fractional improvement by the machine-learning models at each time horizon, based on the Bootstrap mean of fractional improvements. At horizon 1, only the LR provides a fractional improvement, with a 68% probability, whereas the other models remain below the 50% critical level, which indicates no improvement.

On the other hand, RFR performs better than the survey, with an 87% probability at horizon 2, followed by the KRR and BRR, with 74% and 60%. Both the LR and SVR remain below 50% at horizon 2, meaning that they fail to improve the accuracy of survey forecasts.

Significant fractional improvement is achieved by the models at horizon 3, with 99% probability by RFR and 98% by KRR, BRR, and SVR. The accuracy gain by LR occurs with lower probability than the other models, but with 68%, it can still improve the survey forecast at horizon 3.

Table 1. Probability of Fractional Improvement by Machine-Learning Models

Forecast horizon	RFR	KRR	LR	BRR	SVR
h=1	0.33	0.36	**0.68**	0.32	0.39
h=2	**0.87**	0.74	0.49	0.60	0.30
h=3	**0.99**	0.98	0.68	0.98	0.98

Note: Forecast horizons are in months. Bold text indicates the most accurate model for the forecast horizon.
Abbreviations: RFR: Random Forests Regression,
KRR: Kernel Ridge Regression, LR: Linear Regression,
BRR: Bayesian Ridge Regression, SVR: Support Vector Regression

Table 2 gives RMSEs and relative RMSEs for the forecasts. Although the survey forecasts performed worse than the CBT and TBATS, all the models scored smaller RMSEs, in the range of 0.76 to 0.83. With 0.76 RMSE, LR outperformed the other machine-learning models, but the difference is negligible, except for KRR, which has an RMSE of 0.83. TBATS produced better forecasts than the CBT.

MA yielded the worst forecast performance due to the rapidly changing inflation dynamics during the forecast period. The CBT scored only 7% less than the RMSE. By comparison, the TBATS scored 18% better than the survey forecasts. The machine-learning models improved on the survey forecasts by up to 38%.

Table 2. Quarterly CPI Forecast Performances

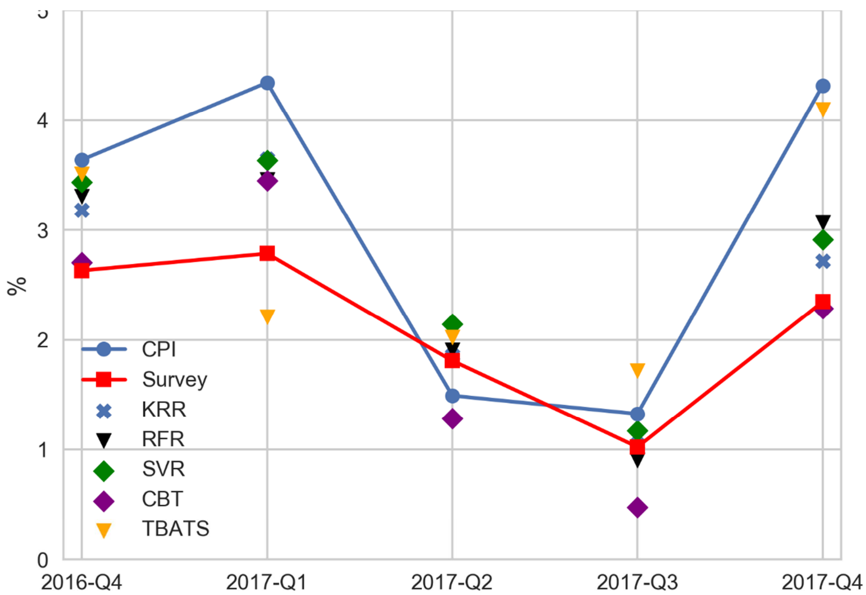
Forecast by	RMSE	Relative RMSE (MA)	Relative RMSE (Survey)
MA	1.70	1.00	1.39
CBT	1.14	0.67	0.93
TBATS	1.01	0.59	0.82
Survey	1.23	0.72	1.00
LR	0.76	0.45	0.62
BRR	0.78	0.46	0.64
RFR	0.77	0.45	0.63
KRR	0.83	0.49	0.68
SVR	0.77	0.45	0.63

Note: Forecast period from 2016-Q4 to 2017-Q5.
Abbreviations: CPI: Consumer-Price Inflation, MA: Naïve Moving Average forecast, CBT: Central Bank of Turkey forecasts, TBATS: A univariate model by de Livera et al. (2011). See note to Table 1 for other abbreviations.

The quarterly inflation forecasts of the machine-learning models, surveys, CBT and TBATS are shown in Figure 2. Due to the limited number of forecasting points and changing performances of the models for different periods, it is difficult to judge the forecast accuracies from the figure. On the other hand, Figure 3 shows the ECDF of the Bootstrap means of forecast absolute errors, which reveals the forecasting performances.

It is apparent from the figure that the machine-learning models outperform the CBT forecasts, which are slightly more accurate than the surveys. Even though the TBATS produced the best forecasts, its performance was not consistent, and large forecasting errors moved the TBATS to a ranking between the CBT and the machine-learning models.

Figure 2. Quarterly CPI Forecasts

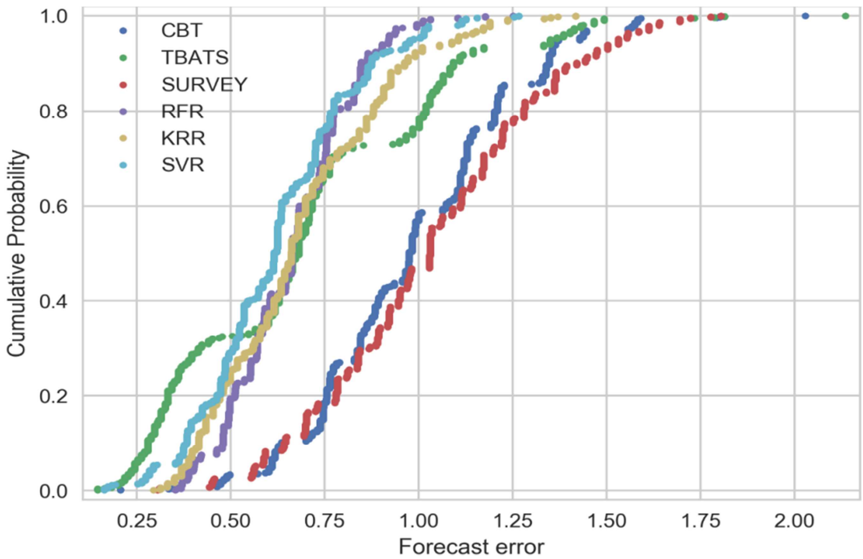


Note: See the note to Table 2 for abbreviations.

To summarize, even though the surveys performed worse than the CBT and the TBATS forecasts, the machine-learning models relying only on the summary statistics of the survey data produced significant levels of accuracy, which led to better forecasts than those of the CBT and the TBATS, as measured by the RMSE. In particular, the RFR yielded the best performance among the selected machine-learning models.

Further accuracy improvement of the models is possible through parameter optimization and an increase in the size of the training data. In addition, extending the forecast period with new data as they become available will contribute to better evaluation of the machine-learning models and improve the forecasting capability of inflation-expectations surveys.

Figure 3. Bootstrap Means of Forecast Errors.



Note: See the note to Table 2 for abbreviations.

5. Conclusion

We have employed a suite of machine-learning models to improve the accuracy of surveys of inflation expectations, conducted by the Central Bank of Turkey (CBT). A training set consisting of only the summary statistics of survey data and actual inflation in Turkey was used. The comparison of forecast performances vis-a-vis the forecasts by the CBT and a univariate model (TBATS) shows that the proposed method not only improves the accuracy of the surveyed forecasts but also outperforms the CBT and TBATS, which are themselves more accurate than the surveys.

We treat the inflation forecasting as an estimation problem in machine learning. The summary statistics of survey data form the features set and the actual inflation is used as labels. The time-series cross validation procedure ensures that the forecast horizon data are not included in the training set for the machine-learning model. Among the models, the RFR yielded the best fractional improvement.

References

Altuğ, S., & C. Çakmaklı, (2016), "Forecasting Inflation Using Survey Expectations and Target Inflation: Evidence for Brazil and Turkey," *International Journal of Forecasting*, *32* (1), pp. 138-153. https://doi.org/10.1016/j.ijforecast.2015.03.010

Ang, A., Bekaert, G., & M. Wei, (2007), "Do Macro-variables, Asset Markets, or Surveys Forecast Inflation Better?" *Journal of Monetary Economics*, *54* (4), pp. 1163-1212. https://doi.org/10.1016/j.jmoneco.2006.04.006

Atkeson, A., & L. E. Ohanian, (2001), "Are Phillips Curves Useful for Forecasting Inflation?" *Federal Reserve Bank of Minneapolis Quarterly Review*, *25* (1), pp. 2-11.

Breiman, L., (2001), "Random Forests," *Machine Learning*, *45* (1), pp. 5-32. https://doi.org/10.1023/A:1010933404324

Gil-Alana, L., A. Moreno, & F. de Gracia, (2012), "Exploring Survey-Based Inflation Forecasts," *Journal of Forecasting*, 31 (6), pp. 524-539. https://doi.org/10.1002/for.1235

Grothe, M., & A. Meyler, (2015), "Inflation Forecasts: Are Market-based and Survey-based Measures Informative?," *European Central Bank, Working Paper No. 1865*.

Hastie, T., R. Tibshirani, & J. Friedman, (2017), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer.

Hsiang, T. C., (1975), "A Bayesian View on Ridge Regression," *Journal of the Royal Statistical Society, Series D (The Statistician)*, 24 (4), pp. 267-268.

James, G., D. Witten, T. Hastie, & R. Tibshirani, (2013), *An Introduction to Statistical Learning*, Vol. 103, ISBN 1461471389, doi:10.1007/978-1-4614-7138-7

Kapetanios, G., V. Labhard, & S. Price, (2008), "Forecast Combination and the Bank of England's Suite of Statistical Forecasting Models," *Economic Modeling*, *25* (4), pp. 772-792. https://doi.org/10.1016/j.econmod.2007.11.004

Livera, A. M., R. J. Hyndman, & R. D. Snyder, (2011), "Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing," *Journal of the American Statistical Association*, 106 (496), pp. 1513-1527. https://doi.org/10.1198/jasa.2011.tm09771

Mullainathan, S., & J. Spiess, (2017), "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives,* 31 (2), pp. 87-106. https://doi.org/10.1257/jep.31.2.87

Öğünç, F., K. Akdoğan, S. Başer, M. G. Chadwick, D. Ertuğ, T. Hülagü, & N. Tekatlı, (2013), "Short-term Inflation Forecasting Models for Turkey and A Forecast Combination Analysis," *Economic Modeling, 33,* pp. 312-325.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, & E. Duchesnay, (2012), "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, 12, pp. 2825-2830. https://doi.org/10.1007/s13398-014-0173-7.2

Stock, J. H., & M. W. Watson, (2007), "Why Has US Inflation Become Harder to Forecast?" *Journal of Money, Credit and Banking*, 39, pp. 3-33. https://doi.org/10.1111/j.1538-4616.2007.00014.x

Vapnik, V. N., (1995), *The Nature of Statistical Learning: Theory*, Springer. https://doi.org/10.1109/TNN.1997.641482

Ekonomi-tek Volume / Cilt: 7 No: 1 Ocak / January 2018, 17-59

The Impact of the Minimum Wage on Wage Distribution: The Evidence from Turkey

Selin Pelek[1]

Abstract

In this paper, we investigate the effect of the minimum wage on the entire system of wage distribution. More specifically, we address the issue of wage inequality by taking into account the potential distributional outcomes of minimum-wage legislation. We decompose the wage differences and the variations in the wage inequality before and after the sizable minimum-wage boost in 2004, following the methodology introduced by DiNardo, Fortin, and Lemieux (1996). We use a non-parametric reweighting approach to decompose the effects of the minimum-wage hike as well as other factors that may have affected the wage distribution. Our main findings confirm that the minimum wage played a pivotal role in reducing wage inequality for both Turkish male and female wage earners between the years 2003 and 2005.

JEL Codes: J31, J38

Keywords: Minimum wage, wage inequality, counterfactual distributions

[1] Galatasaray University, Department of Economics, spelek@gsu.edu.tr

1. Introduction

Turkey's stellar economic performance over the last decade has been accompanied by a shift in its labor market—a rise in the share of salaried workers and a considerable decline in the agricultural population (Ben Salem et al., 2011). However, in tandem with the remarkable growth rates that were experienced after two severe economic crises, in 2000 and 2001, the inequality issue has arisen to occupy center stage in Turkey—both in political debate and as a subject of economic research. An OECD report entitled *Divided We Stand: Why Inequality Keeps Rising* emphasizes that the gap between the rich and the poor widened after the global economic crisis, and the social contract has started to unravel even in OECD countries (OECD, 2011). According to the inequality indicators of the OECD, Turkey, Chile, and Mexico are the emerging countries with the highest rates of income inequality.

Many economists argue that wage evolution is central to examining inequality, claiming that the main reason for changes in inequality is the movement in the rate of wage dispersion (Houller et al., 2012). Given that employment earnings constitute the greatest share of total household income among the working-age population in most OECD countries, the correlation between wage dispersion and rising inequality is not surprising (OECD, 2011). Consequently, more economists are becoming interested in the dynamics of the changes in wage distribution, especially in those countries where income inequality is relatively higher, such as the US or Latin American economies. In this study, we focus on the wage inequality in Turkey, where the share of wage earners in total employment exceeds 67% as of 2017.

The economic literature on wage inequality in developed countries has mostly concentrated on the role of increasing demand for skilled labor due to technological advances, international trade, and job-search frictions (Juhn et al. 1993; Acemoğlu, 2002; Attanasio et al. 2004; Moore and Ranjan, 2005; Kumar and Mishra, 2008; Mortensen, 2005). These studies usually ignore the potential effects of institutional factors on the labor market. However, Bourguignon et al. (2007) highlight the importance of institutional changes for gaining an understanding of inequality trends, especially in developing countries. According to their results, the direction for research into inequality needs to focus on the costs and benefits of public policies such as taxation, the social-security system, subventions, or the minimum wage.

In their influential study, DiNardo et al. (1996) emphasize that labor-market institutions, especially the minimum wage, are as important as market forces in explaining the changes in wage distribution in the US in the early 1980s. Another important study, by Lee (1999), argues that the erosion of the

US federal minimum wage in real terms during the 1980s accounts for much of the worsening in wage inequality in the lower tail of the distribution, particularly for women. Autor, Katz, and Kearney also claim that the decline in the real minimum wage is the primary source of the rising wage inequality seen in recent decades in the United States (Autor et al., 2005).

In their comprehensive paper on the effects of the minimum wage in the UK, Butcher et al. (2012) have developed a model in which the minimum wage has an impact on wage inequality but no significant effect on employment. Moreover, they suggest that the introduction of the UK minimum wage in 1999 explains a key part of the evolution of wage inequality in that country between 1998 and 2010. In sum, there is growing evidence that, under the influence of an efficient minimum-wage policy, the difference between high and low wages becomes smaller, in favor of the latter.

The research on the effect of the minimum wage on wage distribution in developing countries is scarcer than it is in developed ones (Gindling and Terrell, 2005). However, the limited evidence from emerging markets indicates that the wage-compression side effect of the minimum wage is stronger there than it is in developed countries (Lemos, 2009). The labor market in these nations is mainly characterized by a large proportion of informal employment. In this framework, the commonly used theoretical model for testing the distributional effect of the minimum wage is the Welch-Gramlich-Mincer Two-Sector Model (Welch, 1976; Gramlich 1976; Mincer, 1976).

Under the assumption that workers are perfectly mobile, this model suggests that a higher minimum wage could push down the wages in the uncovered sector (meaning that the minimum-wage legislation is not applied to all sectors) due to the movement of workers from the covered sector to the uncovered sector. Thus, the expected effects of the minimum wage on other wages in both the covered and uncovered sectors go in the opposite direction. However, contrary to the predictions of the Two-Sector Model, the evidence from (mostly) Latin American economies indicates that the minimum wage has a positive distributional effect not only in the formal sector, but also in the informal one (Lemos, 2009; Cunningham, 2007; Maloney and Mendez, 2004; Neumark et al., 2006; Fajnzylber, 2001; Khamis, 2008). Fajnzylber (2001) highlights the seeming presence of minimum-wage effects across the whole distribution, including informal salaried workers in Brazil.

Maloney and Mendez (2004) point out the redistributional impact of the minimum wage on the wage distribution of formal and informal workers in Latin American countries. Furthermore, in their theoretical paper, Fugazza and Jacques (2003) develop a model in which labor-market institutions, including the minimum wage, are efficient for reducing the informal sector, and, under certain circumstances, the labor earnings in the formal and informal sectors move in the same direction.[2] Especially in an emerging economy, where there is substantial wage inequality, it is worth investigating the bindingness of the minimum wage. If a minimum wage is binding, one could get a preliminary idea of its enforcement or coverage. Theoretically, enforced minimum-wage legislation with high compliance would generate a censored distribution at the level of the minimum wage, with no workers earning below that level. Nevertheless, noncompliance is widespread, particularly in developing countries (Maloney and Mendez, 2004); thus, the truncation at the minimum-wage level may not be obvious. However, if a spike appears around the minimum wage in wage distribution, one can assume that the minimum wage is somewhat binding (Cunningham, 2007).

This study investigates the effects of the minimum wage on wage distribution in Turkey, based on the micro data of the Household Labor Force Surveys (HLFS) provided by TURKSTAT. The Turkish labor market is known for its late but rapid adaptation to urbanization over the last several decades. This urbanization process implied a major labor reallocation from agriculture to industry and services. As mentioned above, the share of wage-earners in total employment jumped, from 50% to 67%, while the share of unpaid family workers plunged between 2003 and 2016. However, as in many other parts of the developing world, e.g., in Latin America, this typical process of sectoral reallocation has been followed by persistently high unemployment in urban areas and substantial levels of informal employment among salaried workers.

Although we observe a slight decline in the share of informal employment in recent years, this fact is due to the ongoing process of economic restructuring from agriculture towards urban-based employment sectors, rather than the result of a successful public policy to combat informality (Ben Salem et al., 2011). A noteworthy share of salaried employees, around 26% according to the Labor Force Survey in 2010, is still outside of labor-market legislation, i.e., have informal jobs. The evidence of labor-income differentials between the formal and informal segments in the Turkish labor market confirms the existence of an informal penalty. This is in line with the traditional theory of the formal salaried workers being paid more than the informal ones (Tansel and Kan, 2012; Baltagi et al., 2012). In a recent study, Tansel et al. (2019) identify the rising tide of wage inequality for the years 2005 through 2011 in

[2] It is a common practice in the literature on developing countries to use the terms uncovered and informal interchangeably (Gindling and Terrell, 2005); we use them in the same way in this paper.

Turkey, theorizing that this phenomenon could be due to weak labor-market institutions, as well as weak enforcement, and widespread informality.

Over the past decade, Turkish wage earners have benefited from two hefty raises in the real minimum wage. One of the highest occurred in 2004, when the minimum-wage commission decided to raise it by 26.6% in real terms. The second one was implemented more recently, in January 2016: the net minimum wage was upped from 1,000 TL to 1,300 TL. Other increases that were granted between 2004 and 2016 were minor. In this paper, we investigate the effects of the big raise of 2004 on the entire profile of Turkish wage distribution. More specifically, we address the issue of wage inequality by taking into account the potential distributional outcomes of the minimum-wage legislation.

With the methodology introduced by DiNardo, Fortin, and Lemieux (1996—DFL hereafter), we decompose the wage differences and the variations in wage inequality before and after the minimum-wage increase in 2004. We use a non-parametric reweighting approach to decompose the impact of the raise as well as other factors that may have influenced the wage distribution. Our main findings confirm that the minimum wage played a pivotal role in reducing wage inequality for both male and female Turkish wage earners between 2003 and 2005. We control for changes in the individual characteristics over two years and show that they do not have significantly affect wage distribution. This result seems reasonable, since a two-year period is short for a robust change in individual attributes.

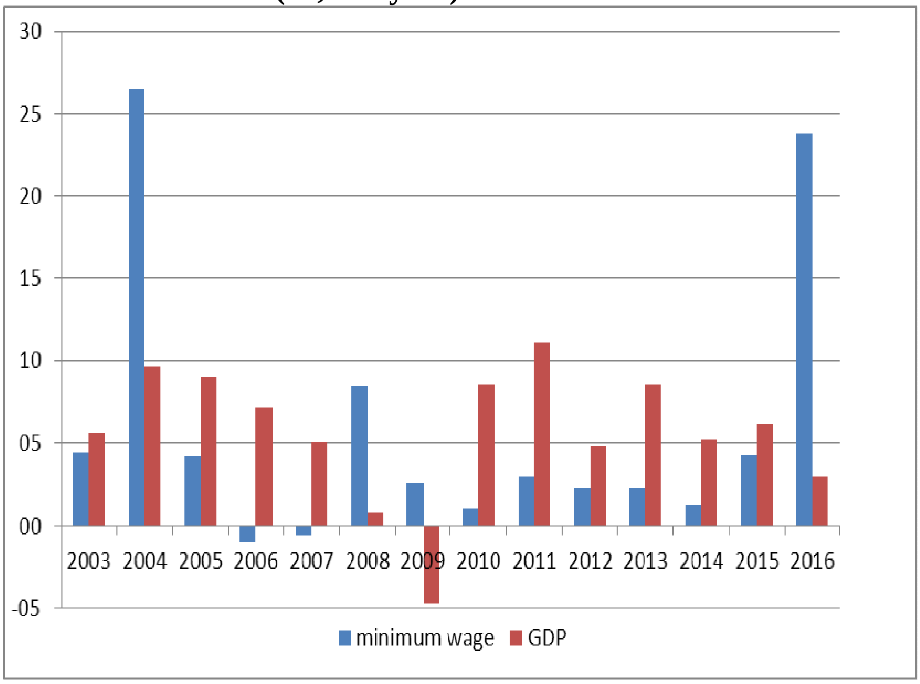
The rest of the paper is organized as follows. Section 2 discusses the evolution of the minimum wage in the Turkish labor market over recent years. Section 3 describes the data set and discusses related issues. Section 4 presents a detailed explanation of the methodology used, and Section 5 reports our empirical results. Finally, Section 6 contains the conclusion and offers suggestions for further research.

2. The minimum wage in Turkey over the past decade

After a severe economic crisis in 2001, Turkey enjoyed a speedy recovery ushered in by a single-party government that has stayed in power since the end of 2002. The economic growth rates reached an average of about 6% a year between 2003 and 2016, even including 2009, when GDP actually contracted. We observed a similar recovery after 2009 as well. The minimum-wage increases also averaged about 6% during the same period.

Figure 1 below presents the annual growth rates in GDP and the minimum wage in real terms during a period when the Justice and Development Party (abbreviated as AKP in Turkish) was in power (it still is), between 2003 and 2016.[3] Note that except for 2006 and 2007, the real minimum wage was raised consistently. As mentioned above, the biggest jumps in the mandatory minimum wage in the wake of the AKP coming to power in 2002 were put through in January 2004, just before local elections, and in January 2016.[4] Although the total magnitude of the raising of the real minimum wage equaled GDP growth between 2003 and 2016, this was more a reflection of the big boosts given in 2004 and in 2016. The main purpose of this study is to investigate whether and to what extent the 2004 substantial increase in the minimum wage affected wage distribution and countered wage inequality.

Figure 1. GDP Growth Rates and the Real Minimum Wage Increases (%, Per year)



Source: TURKSTAT and Ministry of Labor and Social Security

---

3 We exclude the economic crisis years 2000-2001 and the first year of the recovery period, 2002. The single-party government formed by AKP came to power at the end of 2002 and has provided a more stable macroeconomic environment since 2003.

4 These are the monthly net minimum wages for workers aged 16 and older. The average of the minimum wages was taken into account for the years having more than one adjustment, and all wage levels were deflated by 2003 prices using the Consumer Price Index.

For an international comparison of minimum wages, we use the minimum wage/mean and median-wage ratio. This ratio, also known as the Kaitz (1970) index, is the most preferred indicator for cross-country studies, since it provides a basis for the relative level of the minimum wage (Burkhauser et al., 2000). Even though we have used both the mean and median wages as denominators, opting for only the median wage as the denominator is superior, as it omits extreme earnings (Maloney and Mendez, 2004; OECD, 1998).

Table 1 below provides the ratio of the monthly minimum wage to both the mean and median wage for full-time wage earners, between 2003 and 2016 in OECD countries. According to the previous literature, a lower Kaitz index indicates that the minimum wage is relatively weak and probably does not affect a large number of employees, while a higher Kaitz index is generally associated with a bigger share of minimum-wage earners, i.e., a higher minimum wage relative to other wages, which, in turn, could have large ramifications on the labor market (Rycx and Kampelmann, 2012).

Table 1 highlights Turkey's position in first place: it has the highest Kaitz index among the countries listed. Other countries having a relatively elevated Kaitz index are France, Belgium, Ireland, New Zealand, Australia, Slovenia, and Latvia. According to OECD statistics, another significant point is that the ratio of the minimum wage to the median wage is almost double that of the minimum wage to the mean wage. This may be due to the existence of extreme high wages and/or the compression of wages at the bottom of the distribution. Nevertheless, it should be kept in mind that the OECD bases its estimates of the mean and median wages on the Structure of Earnings Survey.

These data, which are provided by TURKSTAT, cover employees who are registered wage earners in all establishments employing 10 and more employees. Thus, the estimated wages, especially mean wages, might be upwardly biased, given that wage earners in the SMEs and informal employees are not covered in this data base. With the notable run-up in the Turkish minimum wage in 2004, the Kaitz index for the country changed dramatically. The ratio of the minimum wage to the median wage soared, from 58% to 75%, and it has not deviated much from that since then. Even the more recent increase of 2016 did not alter the minimum wage to mean/median wage ratios.

Therefore, it is worth examining the distributional effects of the minimum wage in the Turkish labor market, where the bite of this economic factor is significantly deeper than in the other countries. We focus on the effects of the hike of 2004 by measuring the changes in wage distribution in the country between 2003 and 2005.

Table 1. Minimum Wages Relative to Mean and Median Wages, i.e., Kaitz Index of Full-time Wage Earners

		2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Turkey*	**Mw to Mean W**	0,31	0,39	0,39	0,38	0,38	0,38	0,38	0,39	0,39	0,40	0,40	0,39	0,40	0,42	0,42
	Mw to Median W	0,59	0,74	0,74	0,73	0,72	0,71	0,71	0,70	0,71	0,73	0,72	0,69	0,70	0,74	0,74
United States**	Mw to Mean W	0,26	0,25	0,24	0,24	0,23	0,25	0,27	0,28	0,28	0,27	0,27	0,27	0,25	0,25	0,24
	Mw to Median W	0,33	0,32	0,32	0,31	0,31	0,34	0,37	0,39	0,38	0,38	0,37	0,37	0,36	0,35	0,34
United Kingdom**	Mw to Mean W	0,35	0,36	0,37	0,37	0,38	0,38	0,38	0,38	0,38	0,39	0,39	0,40	0,41	0,41	0,44
	Mw to Median W	0,42	0,43	0,45	0,45	0,47	0,46	0,46	0,46	0,47	0,47	0,47	0,48	0,49	0,49	0,54
France**	Mw to Mean W	0,52	0,53	0,54	0,51	0,51	0,51	0,51	0,50	0,50	0,51	0,51	0,51	0,50	0,50	0,50
	Mw to Median W	0,64	0,66	0,67	0,63	0,63	0,63	0,63	0,62	0,62	0,63	0,63	0,63	0,62	0,62	0,62
Korea*	Mw to Mean W	0,27	0,28	0,30	0,31	0,33	0,34	0,36	0,36	0,36	0,34	0,35	0,36	0,38	0,40	0,41
	Mw to Median W	0,34	0,35	0,37	0,39	0,43	0,44	0,45	0,45	0,45	0,43	0,44	0,46	0,49	0,50	0,53
Spain*	Mw to Mean W	0,28	0,28	0,30	0,31	0,32	0,32	0,32	0,32	0,32	0,32	0,32	0,31	0,31	0,31	0,34
	Mw to Median W	0,35	0,35	0,37	0,39	0,39	0,39	0,39	0,38	0,38	0,38	0,38	0,37	0,37	0,37	0,40
Portugal*	Mw to Mean W	0,33	0,33	0,33	0,33	0,33	0,33	0,34	0,36	0,36	0,36	0,36	0,39	0,40	0,42	0,43
	Mw to Median W	0,47	0,47	0,46	0,47	0,48	0,49	0,50	0,53	0,53	0,52	0,52	0,55	0,56	0,59	0,61
Greece*	Mw to Mean W	0,34	0,32	0,31	0,31	0,31	0,33	0,33	0,38	0,36	0,30	0,31	0,32	0,33	0,33	0,33
	Mw to Median W	0,45	0,44	0,45	0,45	0,46	0,48	0,48	0,48	0,52	0,44	0,46	0,47	0,48	0,48	0,48
Poland*	Mw to Mean W	0,35	0,35	0,34	0,34	0,32	0,35	0,37	0,37	0,37	0,39	0,40	0,41	0,41	0,43	0,44
	Mw to Median W	0,43	0,43	0,42	0,42	0,40	0,43	0,46	0,45	0,45	0,48	0,50	0,51	0,51	0,53	0,54

		2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Hungary*	Mw to Mean W	0,36	0,36	0,35	0,36	0,34	0,34	0,34	0,35	0,36	0,40	0,40	0,40	0,40	0,39	0,40
	Mw to Median W	0,47	0,47	0,46	0,48	0,47	0,46	0,47	0,47	0,49	0,54	0,54	0,54	0,53	0,51	0,53
Czech Republic*	Mw to Mean W	0,34	0,35	0,36	0,36	0,35	0,33	0,32	0,32	0,31	0,31	0,31	0,32	0,33	0,34	0,35
	Mw to Median W	0,40	0,40	0,41	0,42	0,41	0,38	0,38	0,38	0,37	0,36	0,37	0,37	0,39	0,40	0,41
Romania*	Mw to Mean W	0,35	0,33	0,32	0,28	0,26	0,30	0,32	0,32	0,33	0,33	0,35	0,38	0,40	0,41	0,44
	Mw to Median W	0,46	0,44	0,42	0,37	0,38	0,43	0,44	0,43	0,45	0,45	0,48	0,51	0,55	0,56	0,60
Estonia	Mw to Mean W	0,32	0,34	0,32	0,30	0,30	0,32	0,34	0,34	0,33	0,32	0,33	0,34	0,35	0,35	0,35
	Mw to Median W	0,39	0,42	0,40	0,37	0,36	0,38	0,40	0,40	0,39	0,38	0,40	0,40	0,41	0,41	0,41
Slovak Republic*	Mw to Mean W	0,36	0,35	0,35	0,35	0,35	0,34	0,36	0,37	0,36	0,36	0,36	0,37	0,37	0,38	0,38
	Mw to Median W	0,45	0,44	0,43	0,45	0,44	0,43	0,45	0,46	0,46	0,45	0,46	0,45	0,47	0,48	0,48
Slovenia	Mw to Mean W	NA	NA	0,43	0,43	0,41	0,41	0,41	0,48	0,49	0,50	0,52	0,49	0,49	0,48	0,48
	Mw to Median W	NA	NA	0,51	0,52	0,51	0,51	0,51	0,59	0,61	0,62	0,64	0,60	0,60	0,59	0,58
Latvia	Mw to Mean W	0,34	0,35	0,30	0,28	0,28	0,32	0,37	0,38	0,41	0,39	0,37	0,39	0,41	0,41	0,39
	Mw to Median W	0,44	0,46	0,40	0,36	0,37	0,40	0,47	0,49	0,51	0,49	0,47	0,49	0,52	0,51	0,48
Lithuania	Mw to Mean W	0,36	0,37	0,37	0,35	0,33	0,34	0,35	0,40	0,39	0,38	0,44	0,41	0,40	0,45	0,43
	Mw to Median W	0,46	0,47	0,46	0,44	0,41	0,42	0,44	0,50	0,48	0,48	0,55	0,51	0,50	0,56	0,54
Mexico**	Mw to Mean W	0,30	0,30	0,30	0,29	0,28	0,29	0,28	0,27	0,27	0,27	0,27	0,29	0,29	0,29	0,31
	Mw to Median W	NA	NA	0,38	0,38	0,37	0,36	0,37	0,35	0,36	0,36	0,37	0,37	0,37	0,37	0,40

Source : OECD. Available at:https://stats.oecd.org/Index.aspx?DataSetCode=MIN2AVE
Notes: *Monthly earnings, **Hourly earnings, ***Weekly earnings of full-time wage earners.

3. Data

We use the 2003 and 2005 HLFS annual micro data provided by TURKSTAT. In Turkey, the HLFS is the main data source for labor-market statistics as it collects detailed information from the labor-supply perspective and assembles a broad range of facts on the socio-economic conditions of both formal and informal workers. The definitions and classifications of the variables in the HLFS have been harmonized with international standards, as determined by Eurostat and the ILO. Economic activities and occupations are coded at four-digit levels, according to the NACE and ISCO-88 classifications, and results are given within nine main groups. These data regularly survey the main demographic and socio-economic characteristics of households' members, such as age, gender, marital status, labor-market status, tenure, hours worked, income from paid employment, informal employment, and unemployment duration.

Thus, the HLFS enables us to control for certain relevant individual characteristics that may affect wages. Being the product of standardized sampling and weighting methods, these data are designed to be representative of the whole non-institutional population of Turkey. The annual results are published as a cross-sectional design. Of course, we realize that the lack of longitudinal data structure over this period somewhat limits the empirical research; nevertheless, we make up for this by taking advantage of the large sample size of the HLFS and using appropriate estimation methods for repeated cross-sectional data.

A question about earnings from paid employment was added to the HLFS in 2003. However, the date of inclusion of this additional information does not pose a problem since our period of interest centers on 2004, when the massive boost was enacted. By taking into account the potential time-lagged effects of this increase, we investigate changes in wage distributions from 2003 to 2005.

Note that our sample includes full-time wage earners in non-agricultural activities among the working-age population (those aged 15 to 65) who declare a net positive salary in the reference month. We simply exclude those wage earners who work less than 30 hours per week, i.e., part-time workers. This restriction is completely conventional for research into wage structure (Katz and Murphy, 1992; Verdugo et al., 2012).

Furthermore, the percentage of part-time employees among all wage earners is miniscule (about 0.8% in 2004 and 1.4% in 2005, according to the labor surveys provided by TURKSTAT; there was no specific question about employment type in 2003), contrary to industrialized countries. Since the minimum wage and the reported employment earnings in the HLFS are on a monthly basis, we prefer to work with monthly wages. Finally, we exclude observations of the lowest 1% as well as the highest 1% of the wage distribution in order to avoid the effect of outliers on the estimation. Consequently, our sample comprises 33,023 men and 8,821 women in 2003, and 53,978 men and 13,476 women in 2005. Table 2 reports the summary statistics of the sample.

Table 2. Characteristics of Full-time Wage Earners in Turkey (%)

	Men		**Women**	
	2003	**2005**	**2003**	**2005**
Average age	33.5	33.8	30	30.5
Years of schooling	8.5	8.5	10.5	10.3
Education				
Illiterate	1	1	1	2
Literate, but not completed any school	1	2	1	2
Primary school	41	39	23	22
Secondary school	15	17	9	10
High school, vocational or technical high school	27	27	32	31
University, faculty or upper	15	14	34	33
Married	76	75	48	46
Urban population	78	84	85	89
Tenure (year)	19	19.3	13.5	14.2
Sector				
Industry	31	33	31	30
Construction	9	9	1	1
Services	60	58	68	69
Unskilled	13	13	10	12
Informal wage earners	27	28	22	25
Below minimum wage	13	14	17	16
At or near the min. wage	9	16	13	20
Number of obs.	**33,023**	**53,978**	**8,821**	**13,476**

Source: The HLFS, 2003 and 2005; own calculations

We do not observe any significant changes in the characteristics of full-time wage earners from 2003 to 2005 for either men or women. It is not surprising given the fact that a two-year period of time is not long enough for any structural changes in a labor market to show themselves. However, the workforce has become more educated. The share of primary-school graduates has declined slightly, while the average years of schooling have remained unchanged.[5] The most remarkable shift has occurred in the share of urban workers. The urbanization rate has gone up throughout the country, and the urban population among wage earners has expanded correspondingly while the sectoral decomposition has remained stable.

The share of unskilled wage earners has increased among women, from 10% to 12%. Another key indicator is the informal employment rate, which has remained almost stable among male wage earners, while it has moved upward among women, from 22% to 25%, over two years.

The proportion of workers who earn at or near the minimum wage[6] has jumped from 9% to 16% and from 13% to 20% among male and female wage earners, respectively. In light of the minimum-wage hike in 2004, this growth in the minimum-wage population seems plausible. Even so, an unusual feature is puzzling: the proportion of workers paid below the minimum wage has stayed almost unchanged.

Furthermore, according to the data provided by TURKSTAT, the unemployment rate did not worsen; on the contrary, it dipped slightly between 2003 and 2005. The total unemployment rate was 10.5% (13.8% nonagricultural) in 2003 and 10.6% (13.5% nonagricultural) in 2005. By gender, it was 10.7% (12.6% nonagricultural) in 2003 and 10.5% (12.2% nonagricultural) in 2005 for men, while it was 10.1% (18.9% nonagricultural) in 2003 and 11.2% (18.7% nonagricultural) in 2005 for women.

In order to get a more detailed picture of workers, we divide our sample into two sub-groups: formal wage earners who are covered by a social-security program due to their primary jobs and informal wage earners who are not covered. Those having social-security coverage numbered 23,857 males and 6,811 females in the 2003 sample; and 38,848 males and 10,055 females in the 2005 sample. The informal wage earners' sample comprises 9,166 men and 2,010 women in 2003; and 15,130 men and 3,421 women in 2005.

Tables 3 and 4 provide the individual and job characteristics of these workers separately.

[5] We do not go into detail on the comparison between male and female workers within our framework. However, we would like to highlight that female wage earners are younger, more urbanized, and more educated than male wage earners: 64% of female full-time wage earners have completed high school or above, compared to 41% of males.

[6] Following the previous literature, we define at or near minimum wage those workers whose monthly salaries are between 0.95 and 1.05 of the minimum wage (Lemos, 2004b).

Table 3. Characteristics of Formal Full-time Wage Earners in Turkey (%)

	Men		Women	
	2003	**2005**	**2003**	**2005**
Average age	34.5	34.7	30.5	31.1
Years of schooling	9.3	9.2	11.4	11.4
Education				
Illiterate	0	0	0	0
Literate, but not completed any school	0	1	0	1
Primary school	33	33	16	16
Secondary school	14	15	7	8
High school, vocational or technical high school	33	32	35	33
University, faculty or upper	20	19	42	42
Married	81	80	51	50
Urban population	79	84	86	90
Tenure (year)	19.2	19.5	13,1	13.7
Sector				
Industry	33	35	27	26
Construction	4	4	1	1
Services	63	61	72	73
Unskilled	12	12	8	9
Below minimum wage	4	3	6	4
At or near the min. wage	9	16	14	20
Number of obs.	**23,857**	**38,848**	**6,811**	**10,055**

Source: The HLFS, 2003 and 2005, own calculations

Similar to the entire population of Turkish wage earners (and ignoring the growth in the urban population), the characteristics of formal and informal wage earners did not change markedly over the 2003-2005 period. However, the minimum-wage variables display a noteworthy variation over the same period. Note that a non-negligible segment of the informal wage earners are paid near the minimum-wage level. In fact, fully half of informal female wage earners and around 40% of informal male wage earners are earning below the minimum wage. In addition, among formal full-time wage earners, 3-4% of men and 4-6 % of women declared that their salary was less than the minimum wage. This could reflect a measurement error.

On the other hand, keeping in mind that a lower minimum wage (around 85% of the adult minimum wage) is typically given to those under the age of 16, one can assume that certain percentage of these workers are between 15 and 16.[7] Another explanation could be over-reporting distortions due to the other advantages of being registered in the social security system, such as retirement or health insurance. After all, it is clearly seen that the 2004 windfall widened the proportion of minimum-wage earners by 7% and 8% among formal and informal wage earners, respectively.

Table 4. Characteristics of Informal Full-time Wage Earners in Turkey (%)

	Men		Women	
	2003	**2005**	**2003**	**2005**
Average age	30.8	31.5	28	28.7
Years of schooling	6.4	6.7	7	7.1
Education				
Illiterate	2	2	5	6
Literate, but not completed any school	2	5	4	7
Primary school	62	52	48	40
Secondary school	17	21	15	19
High school, vocational or technical high school	13	17	23	23
University, faculty or upper	2	3	5	5
Married	62	61	37	35
Urban population	76	83	82	86
Tenure (year)	18.4	18.9	15	15.6
Sector				
Industry	28	28	42	41
Construction	22	21	1	1
Services	50	51	57	58
Unskilled	15	17	19	21
Below minimum wage	37	39	56	51
At or near the min. wage	7	15	11	19
Number of obs.	**9,166**	**15,130**	**2,010**	**3,421**

Source: The HLFS, 2003 and 2005, own calculations

The other way to measure bindingness of the minimum wage is to examine the distribution of wages. In order to see if the mandatory minimum wage is binding, and how the wages are distributed, we take a commonly used graphical approach. Kernel density plots provide a clearer representation of

7 Unfortunately, we could not exclude them because these age groups are determined as 15-19 in the LFS. However, the share of the 15-19 age group among formal wage earners who are paid less than the minimum wage is only about 14% among men and 19% among women. Thus, the wage earners aged between 15 and 16 do not seem to be overrepresented in this group.

wage levels and spotlight where the minimum-wage hikes. Kernel density estimators are essentially a continuous version of discrete histograms and approximate the density f(w) based on observations w. They smooth a line between each observation w_i along the x-axis within a selected bandwidth. More formally, Kernel density estimation can be expressed as:

$$\widehat{f_h}(w) = \sum_{i=1}^{n} \frac{\theta_i}{h} K(\frac{w - w_i}{h})$$

where n is the size of the classes, θ_i is the sample weight of observation i, h is the bandwidth, K(.) is the kernel function, and x some point along the x-axis. Kernel function simply estimates the density $\widehat{f_h}(x)$ from the fraction of the sample that is close to x, *i.e,.* the fraction that falls into the bandwidth, h.

Thus, the choice of the bandwidth is critical since Kernel estimation is sensitive to the bandwidth chosen to smooth. In this paper, we use 2,000 point estimates and the Gaussian[8] Kernel estimator. The optimal bandwidth is specified with Sheather and Jones' selector based on Silverman's method (Silverman, 1986).[9]

Figure 2 and Figure 3 display Kernel estimates of the real monthly wages of full-time workers by gender in 2003 and 2005.

It is clearly seen that the minimum wage is somewhat binding in Turkey; however, it is not necessarily enforced as a wage floor. A considerable number of full-time workers are subminimum earners, which is similar to the situation in other developing countries. It is worth noting that the minimum wage produces a sharper spike in the wage distribution of women than of men. This difference indicates that the wages of female workers are more concentrated around the minimum-wage level, which accords with the results presented by Calavrezo and Pelek (2011) in their research into low-wage workers in Turkey.

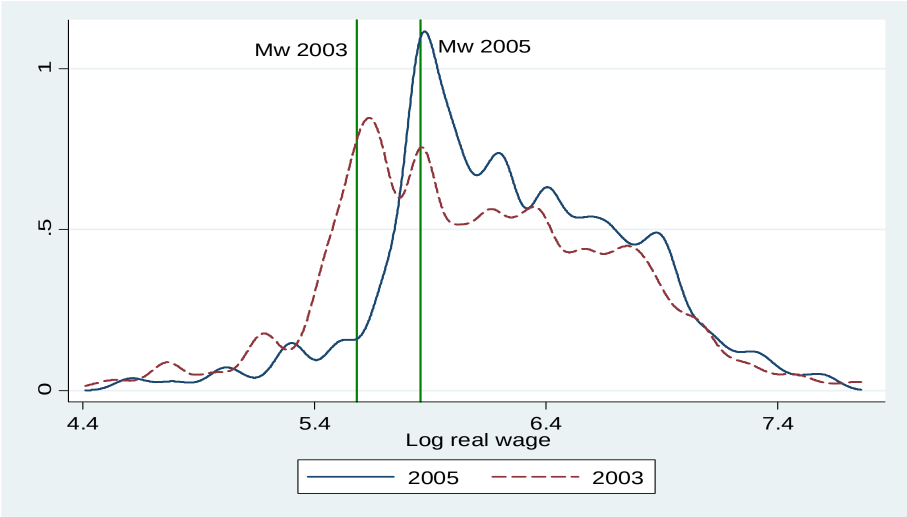
The most significant change over the two years is that the left side of the wage distribution has shifted to the right while the right side has remained almost stable.

---

8 The Gaussian Kernel function is a conventional choice in literature. However, the use of other functions does not change the results dramatically.

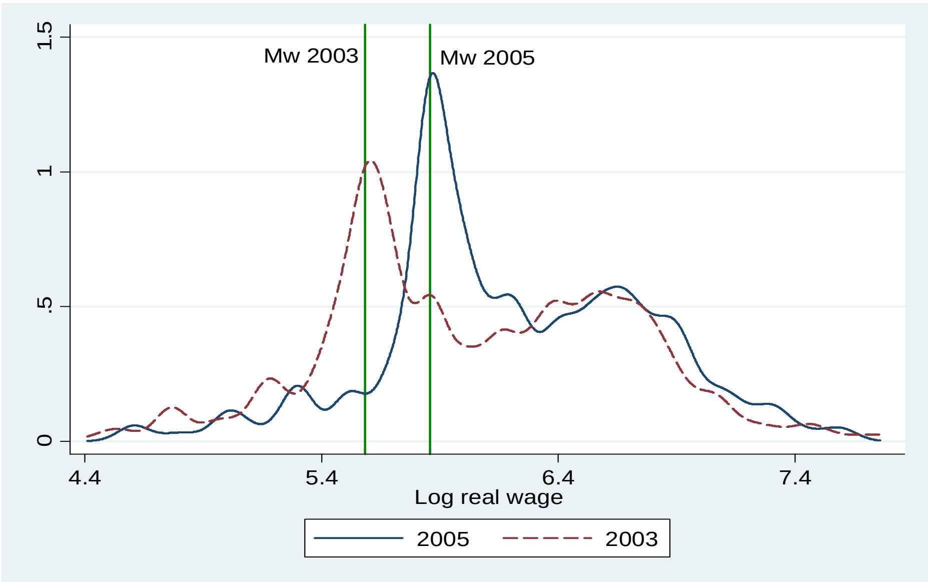
9 For a more detailed explanation of Kernel estimation, see Deaton (1997), Maloney and Mendez (2004), and Cunningham (2007).

Figure 2. Kernel Density Plots of Full-time Male Wage Earners



Source: The HLFS, 2003 and 2005, own calculations

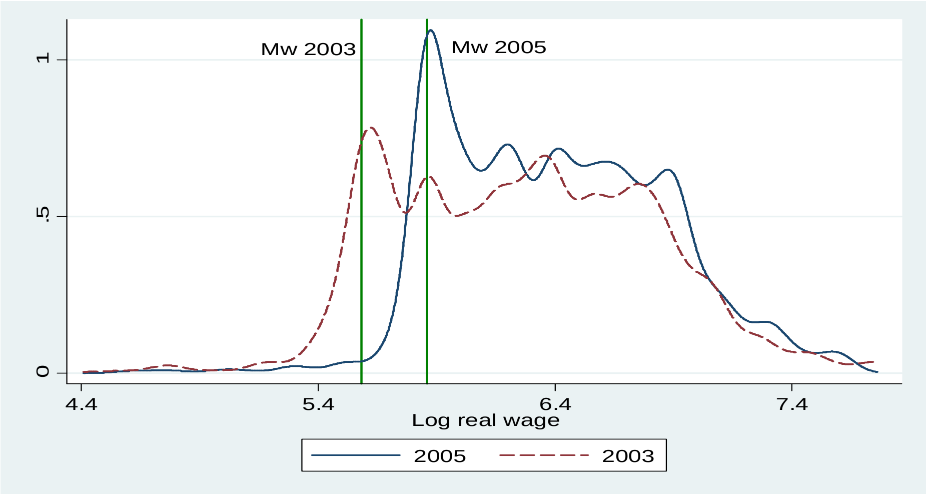
Figure 3. Kernel Density Plots of Full-time Female Wage Earners



Source: The HLFS, 2003 and 2005, own calculations

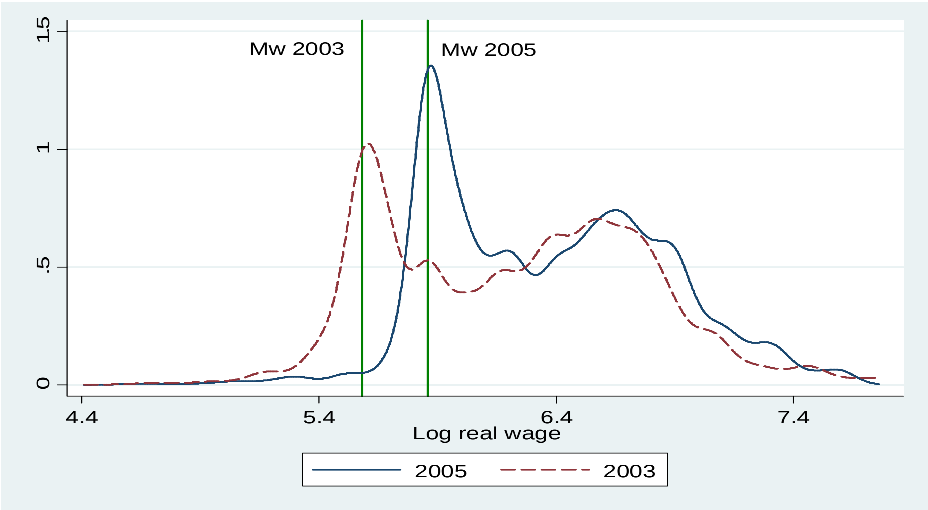
Figures 4-7 display the wage distributions of the formal and informal wage earners by gender.

Figure 4. Kernel Density Plots of Full-time Formal Male Wage Earners



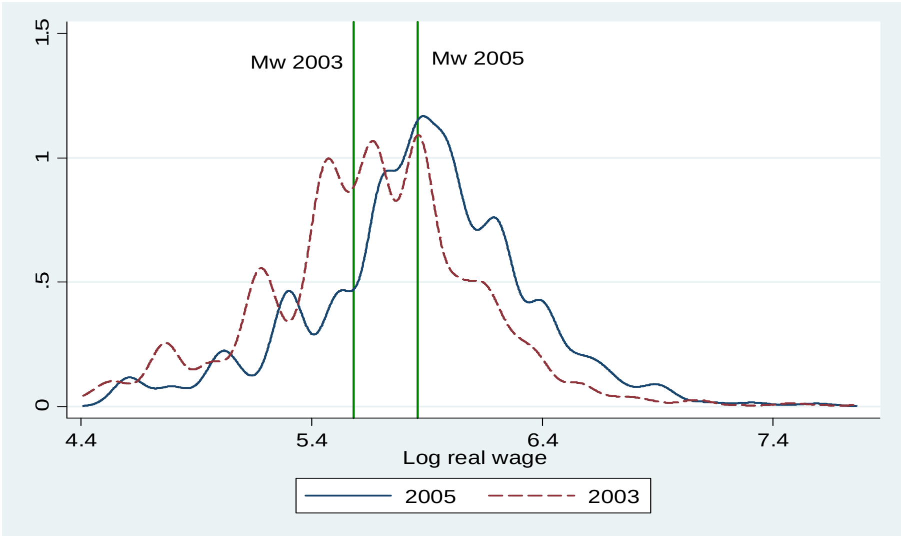
Source: The HLFS, 2003 and 2005, own calculations

Figure 5. Kernel Density Plots of Full-time Formal Female Wage Earners



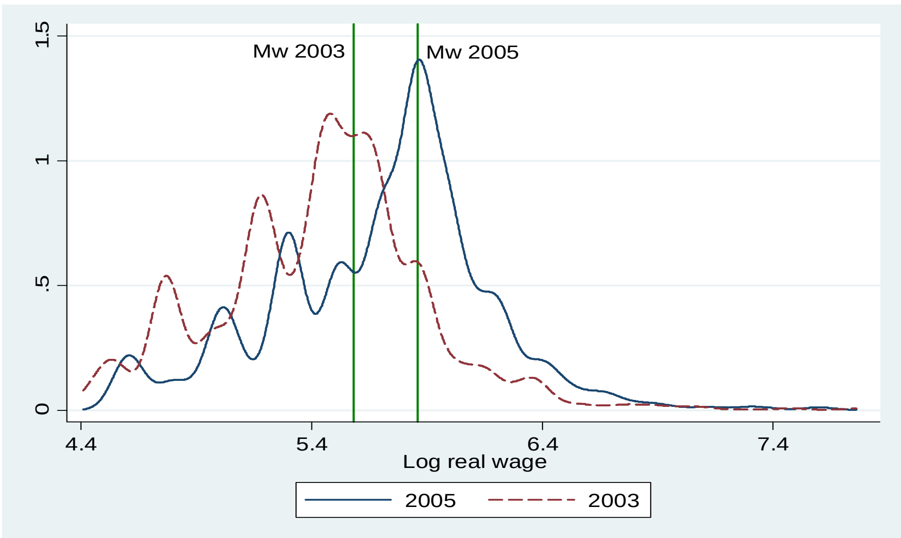
Source: HLFS 2003 and 2005, own calculations

Figure 6. Kernel Density Plots of Full-time Informal Male Wage Earners



Source: The HLFS, 2003 and 2005, own calculations

Figure 7. Kernel Density Plots of Full-time Informal Female Wage Earners



Source: The HLFS,2003 and 2005, own calculations

The minimum wage clearly truncates the wage distribution of the formal wage earners. The spikes at the minimum-wage level occur both for men and women. A significant wage increase is observed at the bottom of the wage distribution of the formal wage earners, while those earning high wages did not vary notably from 2003 to 2005. The shift is marked only on the left side of the wage distribution. Therefore, the minimum-wage hike in 2004 seems particularly important for the distribution of wages among formal workers. At the same time, the minimum wage is not well enforced as a wage floor in Turkey, given that a great number of wage earners are not registered with the social-security system and earn below the minimum wage, as mentioned above. However, although informal workers are not covered by labor legislation, the spikes are observed around the minimum wage. The wage curve of the informal wage earners as a whole shifted to the right between 2003 and 2005, unlike formal ones.

Cumulative density plots provide an alternative illustration of wage distribution. Bear in mind that no assumption about bandwidth is required for plotting cumulative density distribution. If a visible vertical “cliff” appears around the minimum-wage level, one can assume that the distribution of wages is not continuous, the minimum wage truncates (or probably multiplies) the wage distribution, and, thus, it is binding. If all employees are paid at least the minimum wage, this suggests that the minimum wage is enforced perfectly.

In the Appendix, we plot the cumulative density functions of the real monthly wages of full-time workers by gender in 2003 and 2005.

The vertical cliffs around the minimum wage become clearer in 2005. Both for male and female wage earners, the vertical cliffs around 2003’s minimum wage are not remarkable. Nevertheless, the observed *numeraire* (ripple) effects are very small, and so are negligible in the wage distribution. Cumulative density functions do not indicate that the wage distribution in Turkey has cliffs at three times the minimum wage, while only a barely visible vertical line appears around two times the minimum wage. This evidence is in line with the assumption that minimum wages mainly affect the earnings of those who are paid at or below that level (Brown, 1999; DiNardo et al., 1996).

As for the wage inequality trend in the Turkish labor market over the period under study, we observe that wage inequality decreased substantially between 2003 and 2005 according to the standard inequality indicators. Table 5 summarizes the inequality measures for full-time wage earners.

The standard deviation of log wages; the differences in the 95th and 5th percentiles, between log wages at the 90th and 10th percentiles, the 90th and 50th percentiles, the 75th and 25th percentiles, the 75th and 50th percentiles, the 50th

and 5th percentiles, the 50th and 10th percentiles, the 50th and 25th percentiles; the Gini, Theil, and Atkinson coefficients of real wages indicate that the wage inequality decreased over the period both for men and women. It should be noted that inequality decreases are sharper for the lower tail of the distribution.

Table 5. Inequality Measures of Full-time Wage Earners

Men	**2003**	**2005**	**Difference**	**2010**
Standard Deviation*	0.583	0.527	-0.055	0.519
p95-p5**	1.877	1.723	-0.154	1.691
p90-p10**	1.437	1.240	-0.196	1.258
p90-p50**	0.826	0.729	-0.097	0.759
p75-p25**	0.865	0.731	-0.134	0.763
p75-p50**	0.476	0.421	-0.055	0.435
p50-p5**	0.860	0.811	-0.049	0.803
p50-p10**	0.610	0.511	-0.099	0.497
p50-p25**	0.389	0.310	-0.079	0.302
Gini***	0.326	0.287	-0.039	0.294
Theil***	0.175	0.135	-0.04	0.142
Atkinson***	0.084	0.066	-0.018	0.069
Women	**2003**	**2005**	**Difference**	**2010**
Standard Deviation*	0.617	0.565	-0.051	0.553
p95-p5**	1.948	1.871	-0.077	1.883
p90-p10**	1.500	1.369	-0.131	1.345
p90-p50**	0.858	0.810	-0.049	0.867
p75-p25**	0.957	0.781	-0.176	0.837
p75-p50**	0.565	0.508	-0.057	0.548
p50-p5**	0.890	0.862	-0.028	0.729
p50-p10**	0.642	0.560	-0.082	0.550
p50-p25**	0.392	0.273	-0.119	0.223
Gini***	0.341	0.306	-0.035	0.312
Theil***	0.190	0.153	-0.037	0.159
Atkinson***	0.092	0.074	-0.017	0.076

Source: The HLFS, own calculations
* Standard deviation of log wages;
**Difference between the 90th and the 10th percentiles of the log wage distribution. Similar for the other measures.
*** Gini, Theil and Atkinson coefficients of real wages.

For instance, the wage gap between the log wages at the 90th and 10th percentiles decreased considerably, while the wage gap between the 90th and 50th percentiles did not change notably. The Gini, Theil and Atkinson coefficients went down almost by the same amount for male and female wage earners. In sum, all inequality measures suggest that the wages were compressed from 2003 to 2005, both for men and women. Moreover, the inequality measures of the year 2010 show that this equalizing trend held for the following years. Although the inequality measures are slightly higher in 2010, it should be noted that wage inequality lessened sustainably thereafter. Thus, we suggest that the change in wage inequality occurring between 2003 and 2005 was not illusory. The compression of the wage distribution may arise from a relative increase in the real wages in the lower tails, whereas no remarkable change appears in the upper tails of the wage distributions.

In order to refine the descriptive analysis, we report the inequality measures for formal and informal workers in Table 6 and Table 7, respectively. As these tables indicate, the wage inequality trends have gone in the opposite direction for formal and informal wage earners during the 2003-2005 period. The wage gap plummets vis-à-vis all inequality measures among formal wage earners, while this evolution is not observed among informal wage earners.

This fact strengthens our contention that the increase in the minimum wage played a key role in easing wage inequality between 2003 and 2005, even though the minimum-wage laws only cover registered workers. Among informal wage earners, only the differences between the log wages around the middle of the wage distributions are slightly lower, while the gap between the top and bottom of the wage distribution is somewhat wider. This result is in line with the Kernel density estimations, indicating that the minimum wage is located somewhere in the middle of the wage distribution in the informal sector. However, we must not lose sight of the three inequality parameters, the Gini, Theil, and Atkinson coefficients, which are a little lower in 2005 than in 2003, both for male and female informal wage earners.

In sum, these results suggest that the minimum-wage bonus of 2004 was accompanied by a reduction in wage inequality, especially among formal wage earners. However, a part of this equalizing trend could be attributable to changes in the individual characteristics of workers. In the next section, we estimate a hypothetical density that assumes that the individual characteristics of workers remain at the 2003 level in order to investigate the potential effects of this compression of the wage distribution. DFL (1996) methodology allows us to decompose the effects of institutional factors, such as the minimum wage or unions, and the individual characteristics on wage distribution under specific assumptions. We present the methodology in detail and discuss the assumptions of the model.

Table 6. Inequality Measures of Full-time Formal Wage Earners

Men	**2003**	**2005**	***Difference***
Standard Deviation*	0.536	0.468	-0.067
p95-p5**	1.631	1.411	-0.220
p90-p10**	1.373	1.158	-0.215
p90-p50**	0.697	0.630	-0.067
p75-p25**	0.853	0.743	-0.111
p75-p50**	0.411	0.372	-0.039
p50-p5**	0.759	0.588	-0.171
p50-p10**	0.676	0.528	-0.148
p50-p25**	0.443	0.370	-0.072
Gini***	0.301	0.263	-0.038
Theil***	0.146	0.111	-0.035
Atkinson***	0.071	0.054	-0.017
Women	**2003**	**2005**	**Difference**
Standard Deviation*	0.546	0.486	-0.060
p95-p5**	1.632	1.444	-0.188
p90-p10**	1.354	1.185	-0.169
p90-p50**	0.657	0.655	-0.002
p75-p25**	0.929	0.795	-0.134
p75-p50**	0.398	0.386	-0.011
p50-p5**	0.779	0.589	-0.190
p50-p10**	0.697	0.530	-0.167
p50-p25**	0.531	0.409	-0.122
Gini***	0.306	0.273	-0.033
Theil***	0.152	0.120	-0.032
Atkinson***	0.073	0.058	-0.015

Source: LFS, own calculations;
* Standard deviation of log wages;
**Difference between the 90th and the 10th percentiles of the log wage distribution. Similar for the other measures.
*** The Gini, Theil and Atkinson coefficients of real wages

4. Methodology

We follow the decomposition method developed by DiNardo, Fortin, and Lemieux (1996), as mentioned above. The DFL is a semi-parametric decomposition approach, which is an extended version of the standard Oaxaca Blinder method (OB hereafter). The OB analyzes only the counterfactual differences in mean wages, while the DFL generalizes the method to the whole distribution.

Table 7. Inequality Measures of Full-time Informal Wage Earners

Men	2003	2005	Difference
Standard Deviation*	0.462	0.471	0.009
p95-p5**	1.586	1.601	0.015
p90-p10**	1.138	1.160	0.022
p90-p50**	0.540	0.531	-0.008
p75-p25**	0.526	0.530	0.003
p75-p50**	0.260	0.265	0.005
p50-p5**	0.873	0.882	0.008
p50-p10**	0.598	0.629	0.030
p50-p25**	0.267	0.265	-0.002
Gini***	0.253	0.249	-0.004
Theil***	0.116	0.108	-0.008
Atkinson***	0.055	0.052	-0.003
Women	**2003**	**2005**	**Difference**
Standard Deviation*	0.452	0.476	0.024
p95-p5**	1.515	1.581	0.065
p90-p10**	1.155	1.192	0.037
p90-p50**	0.478	0.448	-0.030
p75-p25**	0.541	0.597	0.055
p75-p50**	0.226	0.201	-0.025
p50-p5**	0.783	0.947	0.164
p50-p10**	0.677	0.744	0.067
p50-p25**	0.315	0.396	0.080
Gini***	0.256	0.251	-0.005
Theil***	0.127	0.114	-0.013
Atkinson***	0.058	0.055	-0.003

Source: LFS, own calculations;
* Standard deviation of log wages;
**Difference between the 90th and the 10th percentiles of the log wage distribution. Similar for the other measures.
*** The Gini, Theil and Atkinson coefficients of real wages.

The estimated counterfactual distributions should be called "the density that would have prevailed if individual attributes had remained at their level and workers had been paid according to the wage schedule observed in." (DiNardo, Fortin, and Lemieux, 1996). In our research, we obtain the counterfactual distributions that give the density of wages in 2005 (assuming the characteristics of workers are the same as those observed in 2003. Therefore, the difference between the actual density of wages in 2005 and the counterfactual density estimated by DFL methodology reflects the potential effect of any factor, e.g., the minimum wage, the unionization rate, etc.. Before getting into the details of the methodology, we will give a basic explanation of wage decomposition.

The standard assumption in the OB decomposition is that the outcome variable *Y* of two groups *A* and *B* is linearly related to the covariates, *X*, and the error term *v* is independent of *X*:

$$Y_{gi} = \beta_{go} + \sum_{k=1}^{K} X_{ik}\beta_{gk} + v_{gi} \quad g=A,B \tag{1}$$

where $E\left(v_{gi}\middle|X_i\right) = 0$ and *X* is the vector of covariates for each observation i. Thus, the overall difference in average outcomes between two groups can be written as:

$$\widehat{\Delta_O} = \overline{Y_B} - \overline{Y_A}$$

$$\widehat{\Delta_O} = \left(\widehat{\beta_{BO}} - \widehat{\beta_{AO}}\right) + \sum_{k=1}^{K} \bar{X}_{Bk} \left(\hat{\beta}_{Bk} - \hat{\beta}_{Ak}\right) + \sum_{k=1}^{K} (\bar{X}_{Bk} - \bar{X}_{Ak})\hat{\beta}_{Ak}$$

$$\hat{\Delta}_o = \hat{\Delta}_S + \hat{\Delta}_X \tag{2}$$

where $\hat{\beta}_{gO}$ and $\hat{\beta}_{gk}$ are estimated intercept and slope coefficients, respectively. The first term in Equation (2) is typically referred to as the *wage structure* effect $(\widehat{\Delta_S})$, and the second term is the *composition effect* $(\widehat{\Delta_X})$, which is is also called the *explained effect* in the OB decomposition.

In their comprehensive review, Firpo et al. (2010) suggest that the wage-structure effect could be interpreted as a treatment effect that captures observed changes in a policy over time, such as unionization status or a minimum-wage hike. In this study, we attempt to decompose the changes in wage distribution into two components: the composition effect based on individual attributes and the wage-structure effect linked to the minimum-wage boost of 2004.

An important limitation of OB decomposition is that it sometimes estimates the wage structure and composition effect on the average outcome, which is linear. However, going beyond the mean is urged by many economists in order to get a more detailed idea of the effects of a treatment on overall distribution. DFL methodology serves this purpose via a reweighting procedure, which will be summarized below.

We begin with the illustration of each observation as a joint density function *f* over (w, z, mw_t, t); wages, individual attributes, minimum wages, and dates. In this study, our groups are determined in terms of date, *t* and *t-1*. The density of wages $f_t(w)$ at a given date t, can be expressed as the integral of the density of wages at date t_w conditional on a set of individual attributes z, and the minimum wage mw_t, over the distribution of individual attributes z, at date t_z.

$$f_t(w) = \int_{z\in\Omega_z} f(w|z, mw_t, t_w = t)dF(z|t_z = t)$$
$$\equiv f(w; mw_t, t_w = t, t_z = t) \tag{3}$$

where Ω_z is the domain of definition of the individual characteristics. Under the assumption that the distribution of individual characteristics does not depend on the level of the minimum wage, the hypothetical density of wages that would have prevailed if the individual attributes had remained as they were at time $t-1$ can be expressed as:

$$f_t^{z_{t-1}}(w) = \int_{z\in\Omega_z} f(w|z, t_w = t; mw_t)dF(z|t_z = t-1)$$
$$\equiv \int_{z\in\Omega_z} f(w|z, t_w = t; mw_t)\psi_z(z)dF(z|t_z = t) \tag{4}$$

where the reweighting function $\psi_z(z)$ in Equation (4) is defined as:

$$\psi_z(z) \equiv dF(z|t_z = t-1)/dF(z|t_z = t). \tag{5}$$

One can see that the unobservable counterfactual density is identical to the actual density at *t* except for the reweighting function, $\psi_z(z)$. Therefore, the critical point is the estimation of this reweighting function, $\hat{\psi}(z)$.

Applying Bayes's rule, this reweighting function can be specified as in Equation (6):

$$\psi_z(z) = \frac{\Pr(t_z=t-1|z)}{\Pr(t_z=t|z)}\frac{\Pr(t_z=t)}{\Pr(t_z=t-1)} \tag{6}$$

The probability of being in period *t*, given individual attributes *z*, could be estimated using a simple probit model:

$$\Pr(t_z = t|z) = \Pr\big(\epsilon > -\beta' H(z)\big) = 1 - \phi(-\beta' H(z)) \tag{7}$$

where in Equation (7) is the cumulative normal distribution, and *H(z)* is a vector of covariates that is a function of *z*.

Consider the actual density function for a group belonging to date *t*, $f_t(w)$ and the counterfactual density $f_t^{z_{t\text{-}1}}(w)$. We can decompose the overall changes into the composition effect and the wage-structure effect by the following specification:

$$\hat{\Delta}_o = f_t(w) - f_{t-1}(w) \tag{8}$$

$$\hat{\Delta}_o = \left(f_t(w) - f_t^{z_{t-1}}(w)\right) + (f_t^{z_{t-1}}(w) - f_{t-1}(w))$$

where the first term in Equation (8) is the composition effect and the second term is the wage-structure effect, referring to the minimum wage in our case. The obtained results are presented in the next section.

5. Results

To decompose the effects of the changes in the wage distribution, we obtain a counterfactual distribution by keeping the individual characteristics constant, as of 2003. The individual attributes used in the probit regressions are educational level, marital status, living area (urban or rural), experience, experience squared, activity (industry, construction, and services), occupation, and being registered with the social-security system. Figures 8 and 9 plot actual Kernel density estimations of full-time wage earners in 2003 and counterfactual Kernel density estimations in 2005, assuming that the individual characteristics remained constant from 2003.

The figures below show that the bottom part of the wage distribution has shifted to the right even if the individual characteristics kept constant to their 2003 level. Thus, the wage-structure effect seems to be the driving force in this equalizing period, both for male and female full-time wage earners. We suggest that if the measurable characteristics of full-time wage earners in 2005 had been the same as in 2003, we would observe again a remarkable shift to the right of wages located at the bottom part of the wage distribution.

In order to clarify the counterfactual analysis, one can estimate the inequality measures by using the hypothetical density of wages. Table 8 reports the inequality measures in 2005, which were obtained by keeping constant the individual attributes in 2003.

Figure 8. Kernel Density Plots of Male Full-time Workers in 2003 and 2005, with 2003's Individual Attributes

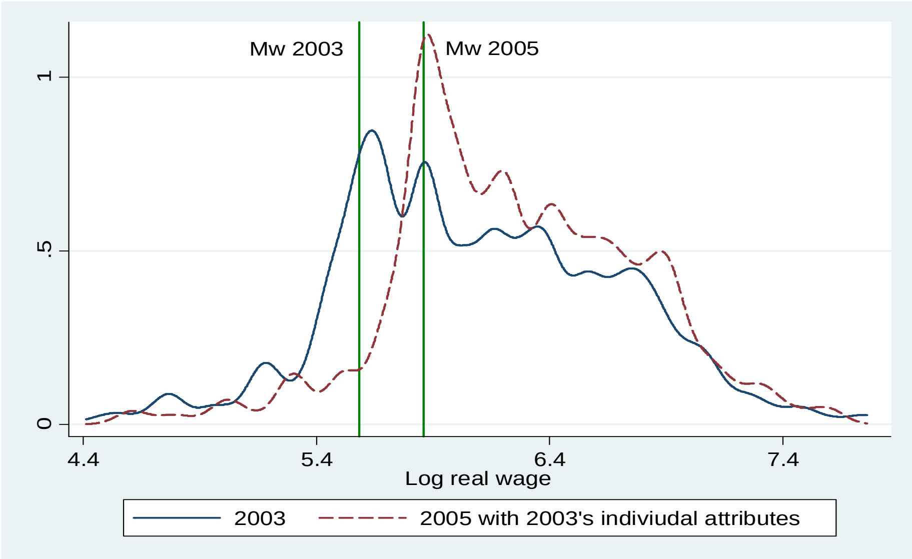


Figure 9. Kernel Density Plots of Female Full-time Workers in 2003 and 2005, with 2003's Individual Attributes

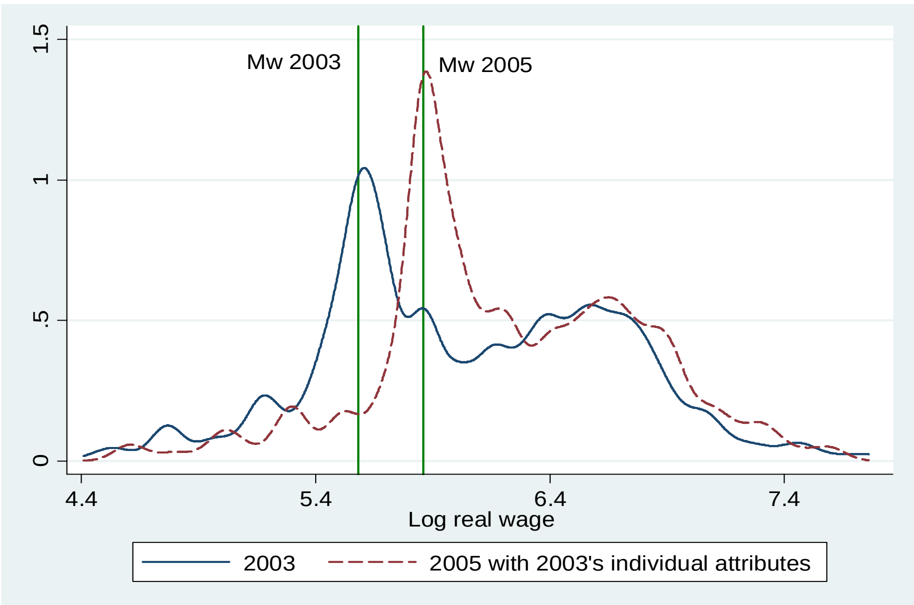


Table 8. Estimated Inequality Measures of Full-time Wage Earners Using Counterfactual Density in 2005

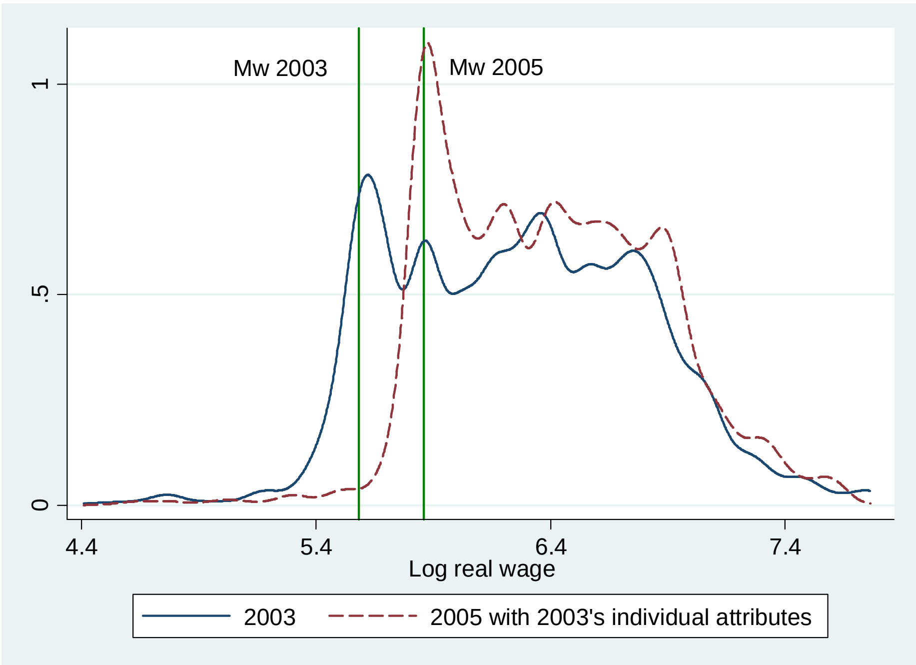
Men	2003	2005CF	Difference
Standard Deviation*	0.583	0.526	-0.056
p95-p5**	1.877	1.713	-0.164
p90-p10**	1.437	1.237	-0.199
p90-p50**	0.826	0.726	-0.101
p75-p25**	0.865	0.733	-0.132
p75-p50**	0.476	0.421	-0.055
p50-p5**	0.860	0.810	-0.050
p50-p10**	0.610	0.511	-0.099
p50-p25**	0.389	0.312	-0.077
Gini***	0.326	0.286	-0.04
Theil***	0.175	0.133	-0.042
Atkinson***	0.084	0.065	-0.019
Women	**2003**	**2005CF**	**Difference**
Standard Deviation*	0.617	0.562	-0.055
p95-p5**	1.948	1.861	-0.087
p90-p10**	1.500	1.341	-0.159
p90-p50**	0.858	0.800	-0.059
p75-p25**	0.957	0.781	-0.176
p75-p50**	0.565	0.503	-0.062
p50-p5**	0.890	0.862	-0.028
p50-p10**	0.642	0.541	-0.101
p50-p25**	0.392	0.278	-0.114
Gini***	0.341	0.303	-0.038
Theil***	0.190	0.150	-0.040
Atkinson***	0.092	0.073	-0.019

Note: 2005 is weighted to individual characteristics in 2003.
* Standard deviation of log wages.
**Difference between the 90th and the 10th percentiles of the log wage distribution. Similar for the other measures.
*** The Gini, Theil, and Atkinson coefficients of real wages.

The estimated inequality measures confirm that the wage-structure effect has played a key role in this equalizing period, not the composition effect. For both women and men, the differences between the actual and hypothetical inequality measures are very small, even close to zero. These results suggest that the changes in the structure of wages had a much larger impact on wage distribution than shifts in individual attributes for all wage earners. Keeping in mind that a period of two years is insufficient for a robust evolution of individual or demographic attributes in a labor market, we are not surprised to find that the wage-structure effect lies behind almost the total change in wage distribution. Another interesting point is that the wage differentials in the lower percentiles were mostly reduced among men, while a similar shrinkage for women occurred in the middle percentiles.

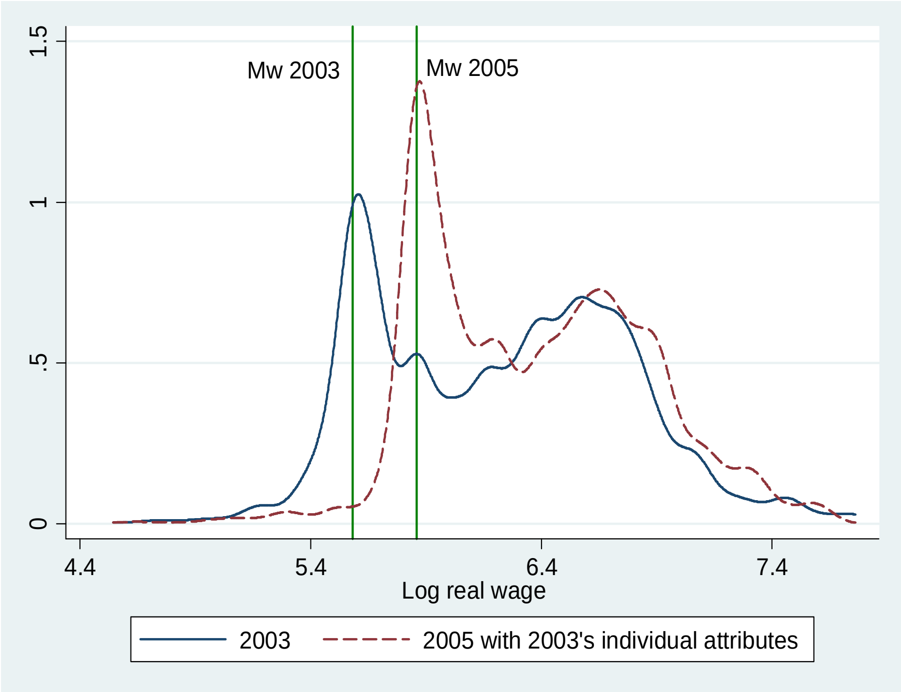
Similar to the descriptive part, we prefer to repeat our analysis for two sub-groups of full-time wage earners. Figures 10-11 and Table 9 report the results for full-time formal workers. Figures 12-13 and Table 10 report those for full-time informal workers. We keep the same variables to control for the individual attributes in probit regressions.

Figure 10. Kernel Density Plots of Male Full-time Formal Workers in 2003 and 2005, with 2003's Individual Attributes



The results for full-time formal wage earners confirm that the wage-structure effect played a key role in bringing about the changes in wage distribution between 2003 and 2005. For both males and females, the Kernel plots do not evince any notable change, while the individual characteristics are kept constant with their 2003 levels. The estimated inequality measures suggest that the change in the measurable individual characteristics explains only a tiny part of the changes in wage distribution. On the other side, we argue that low wage earners in formal jobs benefited from the minimum-wage hike in 2004.

Figure 11. Kernel Density Plots of Female Full-time Formal Workers in 2003 and 2005, with 2003's Individual Attributes



The wage differentials between the upper and lower tails of the wage distribution were reduced substantially for both men and women. Minor changes show up on the upper side of the wage distribution when we control for the individual attributes, designating them as remaining as they were in 2003. For instance, the wage differentials in the 90th and 50th percentiles become positive among females, albeit not much above zero. For both male and female wage earners, the major declines were seen in the lower percentiles of the wage distribution.

As mentioned above, the changes in the wage distributions of informal wage earners are small in comparison with the formal ones. Differences in individual attributes explain a part of this variation. Note that this result is plausible, given that the wage-structure effect does not extend to the informal sector.

If wage differentials are taken as a measure of inequality, one could expect wage inequality among female informal workers to be slightly higher—if individual characteristics had remained unchanged at their 2003 levels.

The estimated measures for men have mostly negative signs, though they approach zero, indicating that the wage differentials decreased over the period under study. However, the shift in the wage distribution of informal workers needs to be explained by other labor-market factors, such as low-high productivity or supply-side changes.

Table 9. Estimated Inequality Measures of Formal Full-time Wage Earners Using Counterfactual Density in 2005

Men	**2003**	**2005CF**	**Difference**
Standard Deviation*	0.536	0.469	-0.066
p95-p5**	1.631	1.408	-0.223
p90-p10**	1.373	1.158	-0.215
p90-p50**	0.697	0.625	-0.072
p75-p25**	0.853	0.748	-0.106
p75-p50**	0.411	0.372	-0.039
p50-p5**	0.759	0.593	-0.166
p50-p10**	0.676	0.533	-0.142
p50-p25**	0.443	0.375	-0.067
Gini***	0.301	0.263	-0.038
Theil***	0.146	0.111	-0.035
Atkinson***	0.071	0.054	-0.017
Women	**2003**	**2005CF**	**Difference**
Standard Deviation*	0.546	0.486	-0.060
p95-p5**	1.632	1.440	-0.192
p90-p10**	1.354	1.175	-0.179
p90-p50**	0.657	0.660	0.003
p75-p25**	0.929	0.792	-0.137
p75-p50**	0.398	0.394	-0.003
p50-p5**	0.779	0.578	-0.201
p50-p10**	0.697	0.515	-0.182
p50-p25**	0.531	0.398	-0.134
Gini***	0.306	0.272	-0.034
Theil***	0.152	0.120	-0.032
Atkinson***	0.073	0.058	-0.015

Note: 2005 is weighted to individual characteristics in 2003.
* Standard deviation of log wages.
**Difference between the 90th and the 10th percentiles of the log wage distribution. Similar for the other measures.
*** The Gini, Theil, and Atkinson coefficients of real wages.

Figure 12. Kernel Density Plots of Male Full-time Informal Workers in 2003 and 2005, with 2003's Individual Attributes

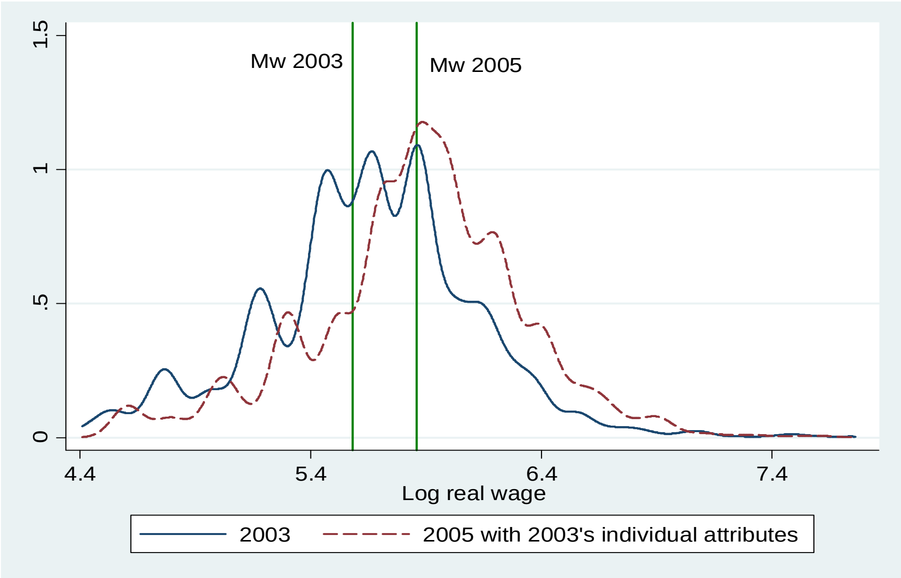


Figure 13. Kernel Density Plots of Female Full-time Informal Workers in 2003 and 2005, with 2003's Individual Attributes

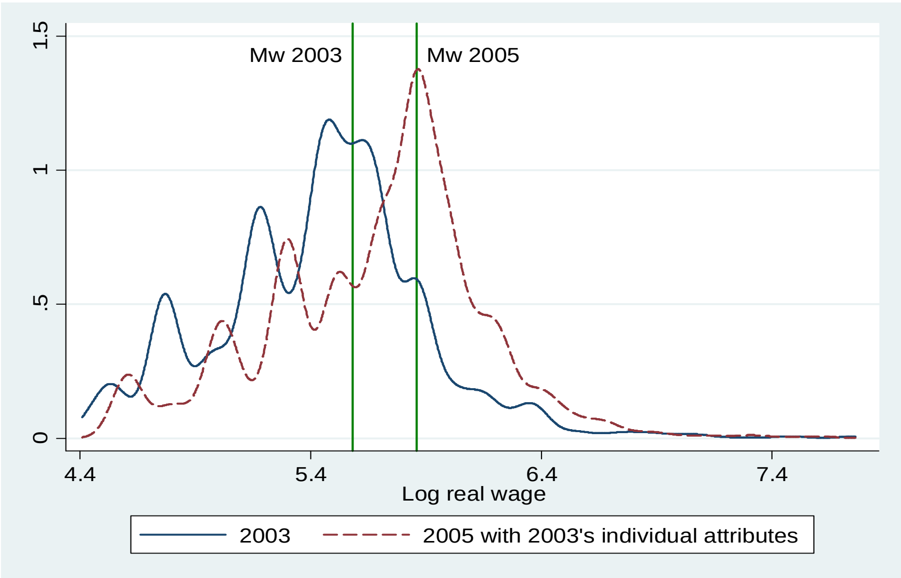


Table 10. Estimated Inequality Measures of Informal Full-time Wage Earners Using Counterfactual Density in 2005

Men	**2003**	**2005CF**	**Difference**
Standard Deviation*	0.462	0.464	0.002
p95-p5**	1.586	1.576	-0.010
p90-p10**	1.138	1.147	0.008
p90-p50**	0.540	0.518	-0.022
p75-p25**	0.526	0.523	-0.003
p75-p50**	0.260	0.258	-0.002
p50-p5**	0.873	0.878	0.005
p50-p10**	0.598	0.629	0.030
p50-p25**	0.267	0.265	-0.002
Gini***	0.253	0.243	-0.010
Theil***	0.116	0.102	-0.014
Atkinson***	0.055	0.049	-0.006
Women	**2003**	**2005CF**	**Difference**
Standard Deviation*	0.452	0.473	0.021
p95-p5**	1.515	1.579	0.064
p90-p10**	1.155	1.190	0.035
p90-p50**	0.478	0.448	-0.030
p75-p25**	0.541	0.612	0.070
p75-p50**	0.226	0.206	-0.020
p50-p5**	0.783	0.957	0.174
p50-p10**	0.677	0.743	0.065
p50-p25**	0.315	0.406	0.091
Gini***	0.256	0.248	-0.008
Theil***	0.127	0.109	-0.018
Atkinson***	0.058	0.053	-0.005

Note: 2005 is weighted to individual characteristics in 2003.
* Standard deviation of log wages.
**Difference between the 90th and the 10th percentiles of the log wage distribution. Similar for the other measures.
*** The Gini, Theil, and Atkinson coefficients of real wages.

These results verify the findings of previous research into the wage-compression effect of the minimum wage in other countries. Autor et al. (2010) point out that the decline in the real value of the minimum wage in the UK is responsible for 30-50% of the growth of lower-tail inequality there. Butcher et al. (2012) propose that the fall in wage inequality in the bottom half of the wage distribution has been most marked in the lowest segments of the labor market, which is consistent with the rise in the national minimum wage in Turkey. Lemos (2004a) indicates that an increase in the minimum wage strongly compresses wage distribution in Brazil. Our results are in line with these studies.

6. Conclusion

In this paper, we focus on wage distribution in Turkey, a developing country with a dynamic labor market. We assess the changes in wage distribution between 2003 and 2005. We find this relatively short period interesting as an area to investigate due to the remarkable minimum-wage uptick of 2004. The wage distribution indicates that the minimum wage is somewhat binding in Turkey.

However, a significant part of full-time wage earners are paid less than the minimum wage due to the informality issue. Furthermore, Turkey has the highest Kaitz index among all OECD countries, indicating that wages are clustered around the minimum wage to some extent. The results obtained by using the HLFS data suggest that the minimum wage compressed wage distribution in Turkey between 2003 and 2005. Wage inequality clearly improved over the period.

We argue that the driving force of this lessening of wage inequality is the rise of wages in the lower tail of wage distribution, caused by the minimum-wage hike. The results also signal that higher wages have not varied notably. We estimate a counterfactual distribution by keeping the measurable individual attributes constant at their 2003 level. The econometric results confirm the influence of the 2004 generosity on easing wage inequality in the country, with this wage-structure effect being especially visible in the formal sector. The changes in the individual attributes do not appear to have any impact on the wage-distribution trend over the period under study—not surprising considering the two-year period was insufficient for such an effect to manifest itself.

However, the lighthouse effect of the minimum wage on the informal sector seems to be small. Also, the distributional effect of the minimum wage has not been reflected on the informal side of the labor market. As for the gender issue, the results indicate that the equalizing trend is observed almost to the same degree among male and female wage earners.

Nevertheless, we would like to emphasize the need for additional and better research into wage inequality in Turkey, perhaps with different databases and methodologies. Since empirical studies are complicated by the limited availability of panel data, the way forward appears to be investigations undertaken conjointly into the employment and distributional effects of the minimum wage. At the same time, future researchers could seek out the impact of the minimum wage on inequality in a broader sense, in areas like poverty or income inequality. In this paper, we argue that the minimum wage is an effective tool for reducing wage inequality despite our not having found any evidence of its ameliorating income inequality in Turkey.

References

Acemoğlu, D., (2002), "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, pp. 7-72.

Attanasio, O., P. K. Goldberg, and N. Pavcnik, (2004), "Trade Reforms and Wage Inequality in Colombia," *Journal of Development Economics,74*, pp. 331-366.

Autor, D. H., L. F. Katz, and M. S. Kearney, (2005), "Trends in US Wage Inequality: Re-Assessing the Revisionists," *NBER Working Papers No: 11627*.

Autor, D. H., A. Manning, and C. L. Smith, (2010), "The Contribution of the Minimum Wage to US Wage Inequality over Three Decades: A Reassessment," *National Bureau of Economic Research Working Paper, No: 16533* .

Baltagi, B. H., Y. S. Başkaya, and T. Hulagu, (2012), "How Different are the Wage Curves for Formal and Informal Workers: Evidence from Turkey," *Central Bank of the Republic of Turkey, Working Papers No: 12/16* .

Ben Salem, M., L. Bensidoun, and S. Pelek, (2011), "Informal Employment in Turkey: An Overview," *Region et développement, 34*, pp.57-84.

Bhorat, H., R. Kanbur, and N. Mayet, (2012), "Minimum Wage Violation in South Africa," *International Labor Review, 151 (3)*, pp. 277-287.

Bourguignon, F., F. H. Ferreira, and M. Menéndez, (2007), "Inequality of Opportunity in Brazil," *Review of Income and Wealth*, 53(4), pp. 585-618.

Brown, C., (1999), "Minimum Wages, Employment, and the Distribution of Income," In O. Ashenfelter, & D. Card, *The Handbook of Labor Economics*,Vol. 3b , pp. 2101-63, Elsevier.

Burkhauser, R. V., K. A. Couch, and D. Wittenburg, (2000), "A Reassessment of the New Economics of the Minimum Wage Literature, with Monthly Data from the Current Population Survey," *Journal of Labor Economics, 18* (4), pp.653-680.

Butcher, T., R. Dickens, and A. Manning, (2012), "Minimum Wages and Wage Inequality: Some Theory and an Application to the UK," *CEP Discussion Paper No: 1177*.

Calavrezo, O., and S. Pelek, (2011), Les Emplois du Bas de L'échelle Salariale en Turquie: Une Description des Salariés des Secteurs Formels et Informels, Hors Agriculture," *Travail et Emploi, 126,* pp. 45-60.

Cunningham, W., (2007), *Minimum Wages and Social Policy: Lessons from Developing Countries.* Washington: World Bank.

Deaton, A, (1997), *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Washington DC: Johns Hopkins University Press.

DiNardo, J., N. M. Fortin, and T. Lemieux, (1996), "Labor Market Institutions and the Distributions of Wages, 1973-1992: A Semiparametric Approach," *Econometrica, 64* (5), pp. 1001-1044.

Fajnzylber, P., (2001), "Minimum Wage Effects Throughout the Wage Distribution: Evidence from Brazil's Formal and Informal Sectors," *Anais do XXIX Encontro Nacional de Economia*.

Firpo, S., N. Fortin, and T. Lemieux, (2010), "Decomposition Methods in Economics," in O. Ashenfelter, & D. Card, Eds., *Handbook of Labor Economics,* Vol. 4, pp 1-102. Elsevier.

Fugazza, M., and J. F. Jacques, (2003), "Labor Market Institutions, Taxation, and the Underground Economy," *Journal of Public Economics, 88,* pp. 395-418.

Gindling, T. H., and K. Terrell, (2005), "The Effect of Minimum Wages on Actual Wages in Formal and Informal Sectors in Costa Rica," *World Development, 33* (11), pp. 1905-1921.

Gramlich, E. M., (1976), "Impact of Minimum Wages on Other Wages, Employment, and Family Incomes," *Brookings Papers on Economic Activity, 2*, pp. 409-451.

Houller, P., I. Joumard, M. Pissu, and D. Bloch, (2012), "Less Income Inequality and More Growth – Are They Compatible? Part 1. Mapping Income Inequality Across the OECD," *OECD Economics Department Working Papers*.

Juhn, C., K. M. Murphy, and B. Pierce, (1993), "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy, 101*, pp. 410-442.

Kaitz, H., (1970), "Experience of the Past: The National Minimum, Youth Unemployment, and Minimum Wages," *US Bureau of Labor Statistics Bulletin, 1657,* pp. 30-54.

Katz, L. F., and K. M. Murphy, (1992), “Changes in Relative Wages, 1963-1987: Supply and Demand Factors,” *Quarterly Journal of Economics, 107* (1), pp. 35-78.

Khamis, M., (2008), “Does the Minimum Wage Have a Higher Impact on the Informal than on the Formal Labor Market? Evidence from Quasi-Experiments,” *IZA Discussion Paper No. 3911* .

Kumar, U., and P. Mishra, (2008), “Trade Liberalization and Wage Inequality: Evidence From India,” *Review of Development Economics,12*, pp. 291-311.

Lee, D. S., (1999), “Wage Inequality in the United States During the 1980s: Rising Dispersion or Falling Minimum Wage?,” *Quarterly Journal of Economics, 114* (3), pp. 977-1023.

Lemos, S., (2004a), “A Menu of Minimum Wage Variables for Evaluating Wages and Employment Effects: Evidence from Brazil,” *IZA Discussion Paper No.1069*.

Lemos, S., (2004b), “Are Wage and Employment Effects Robust to Alternative Minimum Wage Variables?,” *IZA Dicussion Paper No. 1070.*

Lemos, S., (2009), “Minimum Wage Effects in a Developing Country,” *Labor Economics, 16* (2), pp. 224-237.

Maloney, W., and J. Mendez, (2004), “Measuring the Impact of Minimum Wages. Evidence from Latin America,” in N. Books, *Law and Employment: Lessons from Latin America and the Caribbean,* pp. 109-130. NBER.

Mincer, J., (1976), “Unemployment Effects of Minimum Wages,” *Journal of Political Economy, 84*, pp. 87-104.

Moore, M., and P. Ranjan, (2005), “Globalization vs Skill-Based Technological Change: Implications for Unemployment and Wage Inequality,” *Economic Journal, 115* (503), pp. 391-422.

Mortensen, D. T., (2005), *Wage Dispersion: Why Are Similar Workers Paid Differently?,* The MIT Press.

Neumark, D., W. Cunningham, and L. Siga, (2006), “The Effects of the Minimum Wage in Brazil on the Distribution of Family Incomes: 1996-2001,” *Journal of Development Economics, 80* (1), pp. 136-159.

OECD, (2011), *Divided We Stand: Why Inequality Keeps Rising*, OECD Publishing.

OECD, (1998), "Making the Most Out of the Minimum: Statutory Minimum Wages, Employment, and Poverty," in *Employment Outlook*, pp. 31-79. Paris: OECD.

Rycx, F., and S. Kampelmann, (2012), *Who Earns Minimum Wages in Europe? New Empirical Evidence Based on Household Surveys*, Brussels: European Trade Union Institute report (ETUI).

Silverman, B. W., (1986), *Density Estimation for Statistics and Data Analysis,* London : Chapman and Hall.

Tansel, A., and E. O. Kan, (2012), "The Formal/Informal Employment Earnings Gap: Evidence from Turkey," *IZA Discussion Paper No. 6556*.

Tansel, A., B. Çakar and A. Güven, (2019), "Wage Inequality and Wage Mobility in Turkey," *Social Indicators Research*, *142* (1), pp. 107-129.

Verdugo, G., H. F. Horny, and G. Horny, (2012), "Évolution des Inégalités Salariales en France: le Rôle des Effets de Composition," *Banque de France No: 370.*

Welch, F., (1976), "Minimum Wage Legislation in the United States," in O. Ashenfelter & J. Blum, Ed., *Evaluating the Labor Market Effects of Social Programs,* Princetion University Press.

Appendix

Figure A1. Cumulative Density Function of Log Real Wages of Men in 2003

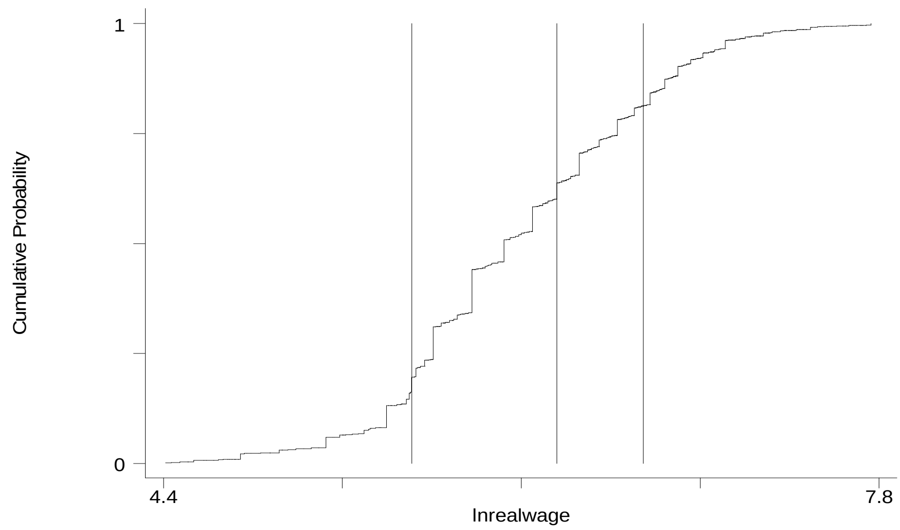


Figure A2. Cumulative Density Functions of Log Real Wages of Formal and Informal Male Workers in 2003

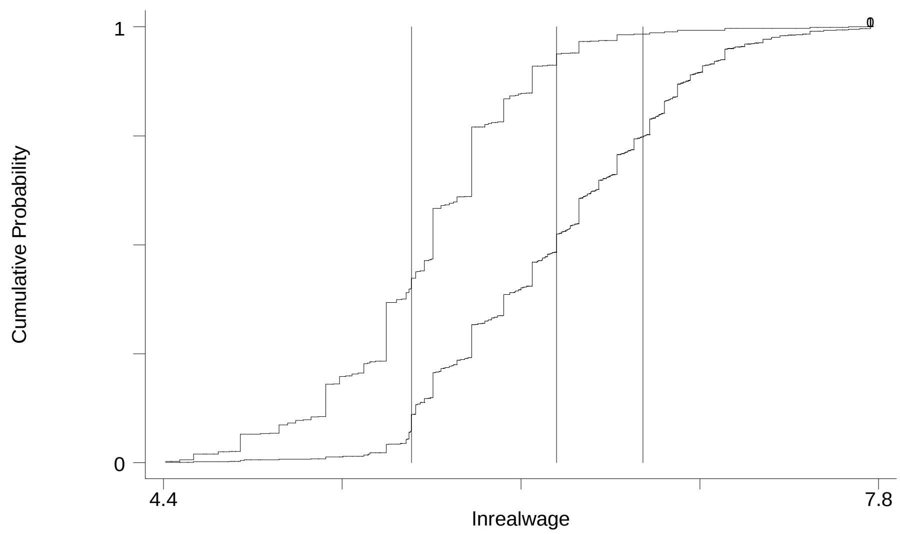


Figure A3. Cumulative Density Function of Log Real Wages of Men in 2005

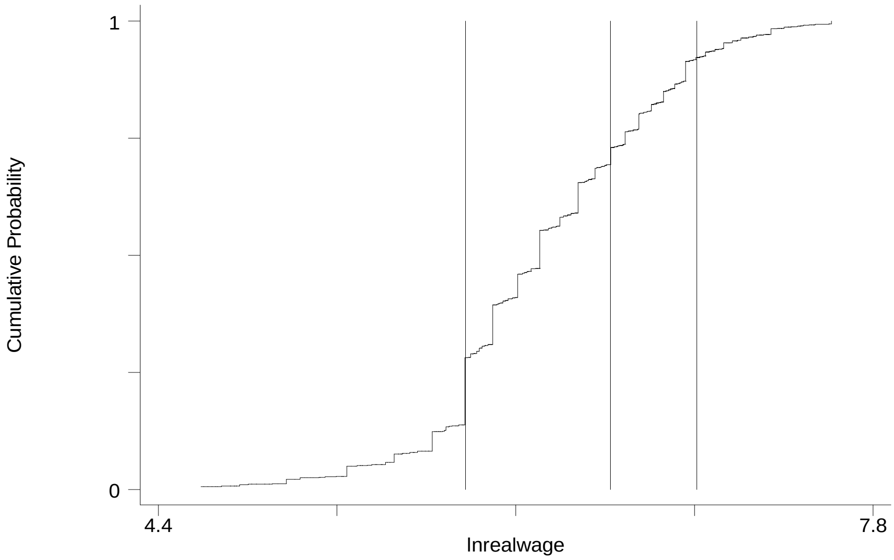


Figure A4. Cumulative Density Functions of Log Real Wages of Formal and Informal Male Workers in 2005

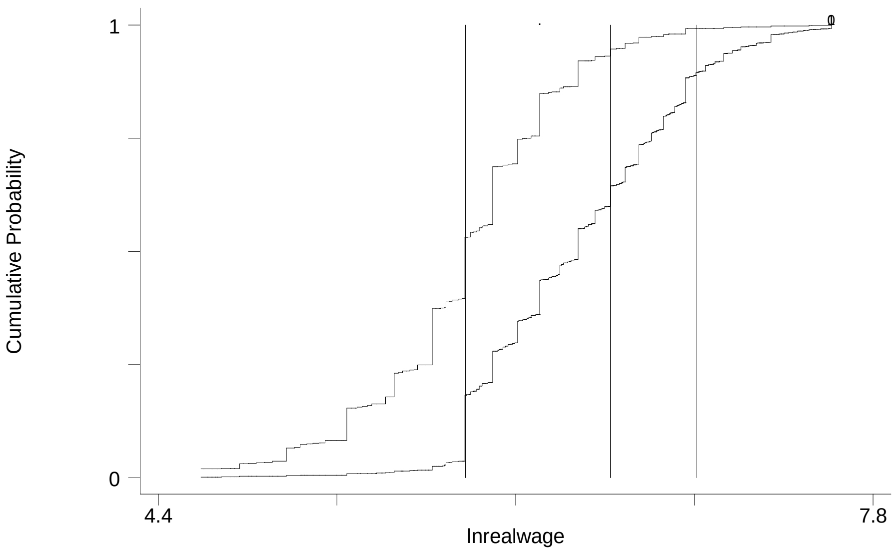


Figure A5. Cumulative Density Function of Log Real Wages of Women in 2003

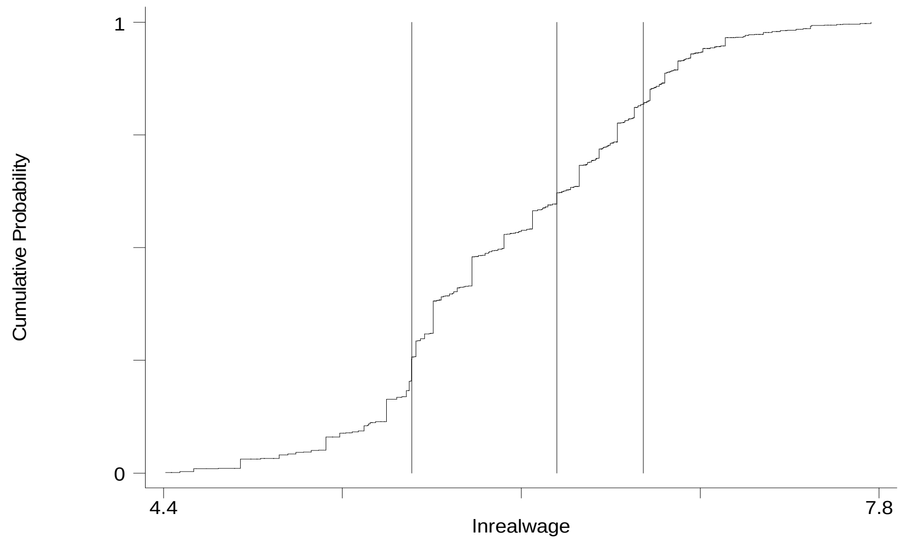


Figure A6. Cumulative Density Functions of Log Real Wages of Formal and Informal Female Workers in 2003

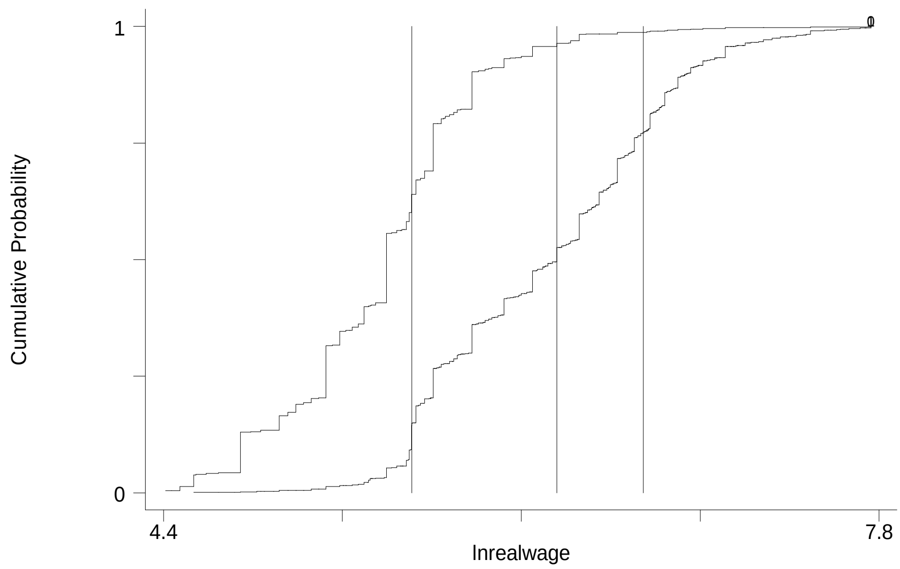


Figure A7. Cumulative Density Function of Log Real Wages of Women in 2005

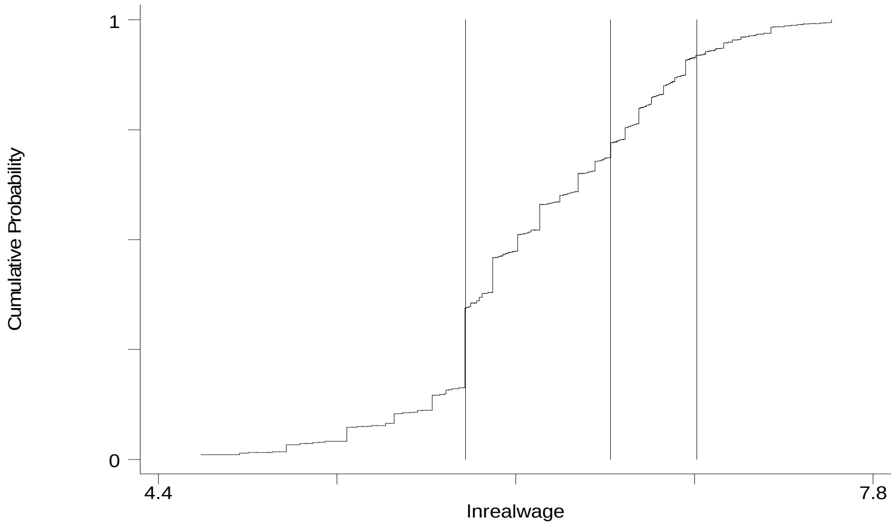


Figure A8. Cumulative Density Functions of Log Real Wages of Formal and Informal Female Workers in 2005