# Cyclical Dynamics in Idiosyncratic Labor-Market Risks: Evidence From March CPS 1968-2011

zhao, bo

China Center for Economic Research (CCER), Peking University

10 June 2013

# Cyclical Dynamics in Idiosyncratic Labor-Market Risks: Evidence From March CPS 1968-2011[☆]

Bo Zhao[a,*]

[a] *China Center for Economic Research, National School of Development, Peking University, Beijing 100871, China*

**Abstract**

The paper estimates the household labor earning process using the March Current Population Survey 1968-2011. GMM estimates confirm that the results in Storesletten et al. (2004) still hold in a much larger data set over a longer period. The persistent idiosyncratic risk is strongly countercyclical, with an annual auto-correlation equal to .973 and an standard deviation that increases by 72.5 percent (from .090 to .156) as the macroeconomy moves from peak to trough.

*Keywords:* Countercyclical, Idiosyncratic risk, Incomplete market model

*JEL:* E32, J30, C51

## 1. Introduction

Understanding the interaction between the aggregate risks and idiosyncratic household labor-income risks is one of the most important topics in macroeconomics. Storesletten et al. (2004) develop a generalized method of moments estimator using household-level labor earnings data from the Panel

*Corresponding author. Tel: +86 010 62758375; Fax: +86 010 62750030
*Email address:* bo.robert.zhao@gmail.com (Bo Zhao)

Study of Income Dynamics 1968-1994. The results are that idiosyncratic risk is (i) highly persistent, with an annual autocorrelation coefficient of 0.95, and (ii) strongly countercyclical, with a conditional standard deviation that increases by 75 percent (from 0.12 to 0.21) as the macroeconomy moves from peak to trough.

This paper extends their research to a longer period (from 1968 to 2011) using data from the March Current Population Survey (CPS). It has many advantages over traditional studies using the PSID, such as larger sample size and longer sample period. The basic identification strategy is to use both the cohort and age variations in household labor-income risks. Suppose there are two groups of households aged 55. The first group members were born in 1925 and the second group members were born in 1955. The first group should exhibit a larger within-group variance of labor income after controlling for other individual characteristics. The reason is that the households born in 1925 have experienced more contractions years than households born in 1955. If income risks have systematic differences between recessions and booms, then the cohort variations should help to identify these differences. Similarly, one can compare the variance of labor income at different years for two groups of households of same cohort.

The paper matches the cross-sectional CPS data to form a two-year overlapping panel and estimates a time-series model for idiosyncratic labor-market risks. It confirms the results in Storesletten et al. (2004). The persistent idiosyncratic risk is strongly countercyclical, with an annual autocorrelation equal to .973 and an standard deviation that increases by 72.5 percent (from .090 to .156) as the macroeconomy moves from peak to trough.

## 2. Time-Series Model for Idiosyncratic Risks

Following Storesletten et al. (2004), I write down the same time-series model for idiosyncratic labor-earning risks. Denote $y_{it}^h$ as the logarithm of labor earning for households $i$ of age $h$ at time $t$. $y_{it}^h$ is specified as follows:

$$y_{i,t}^h = \theta_0 + \theta_1^T D(Y_t) + \theta_2^T x_{i,t}^h + u_{i,t}^h \tag{1}$$

where $D(Y_t)$ is a vector of year dummy variables, $t = 1968, \ldots, 2011$, and $x_{i,t}^h$ is a vector composed of age, age squared, age cubed, race dummies, educational attainment dummies, and state dummies. Educational attainment of household $i$ is defined by the education level of the household head. It is a category variable, including less than high school, high school, some college, college, and advanced. The component $u_{i,t}^h$ is the random component of a household's earnings that is idiosyncratic to them. The residual earnings $u_{i,t}^h$ obeys the following stochastic process:

$$u_{i,t}^h = \alpha_i + z_{i,t}^h + \varepsilon_{i,t}^h \tag{2}$$

$$z_{i,t}^h = \rho z_{i,t}^{h-1} + \eta_{i,t}^h \tag{3}$$

where $\alpha_i \sim \text{Niid}\left(0, \sigma_\alpha^2\right)$, $\varepsilon_{i,t}^h \sim \text{Niid}\left(0, \sigma_\varepsilon^2\right)$, $\eta_{i,t}^h \sim \text{Niid}\left(0, \sigma_t^2\right)$, $z_{i,t}^0 = 0, \forall i, t$. The conditional variance of $\eta_{i,t}^h$ has a regime-switching variance:

$$\sigma_t^2 = \begin{cases} \sigma_E^2 & \text{if } t \text{ is an expansion year} \\ \sigma_C^2 & \text{if } t \text{ is a contraction year} \end{cases} \tag{4}$$

The variable $\eta_{i,t}^h$ and $\varepsilon_{i,t}^h$ are the persistent labor-earning risks and transitory labor-earning risks, respectively. The variable $\alpha_i$ is the individual random effect. The iid assumption implies that individual characteristic $\alpha_i$ is not correlated with residual earnings, either $\eta_{i,t}^h$ or $\varepsilon_{i,t}^h$. This is a strong assumption. However, this enables us to use the short panel dimension to estimate the time-series model.

## 3. Data

The data are from the Annual Demographic File and Income Supplement of Current Population Survey 1968-2011, known as the March CPS. Our point of interest is the family labor earnings. The components of family labor earnings vary over time due to the changes in the variable definitions. See Appendix for the definition of labor earnings.

Although the monthly Current Population Survey is designed mainly for cross-sectional analysis, it is in fact a rotating panel. Households are asked about labor-force related questions for four consecutive months and excluded from the survey for eight months before they are interviewed again for four months. In addition to the usual monthly survey, households are asked about detailed questions on their income last year in March every year. Theoretically speaking, half of sample at year $t$ (those in their first, second, third, and fourth month in the sample) can be matched to the half of sample at year $t + 1$ (those in their fifth, sixth, seventh, and eighth month in the sample).

The Bureau of Labor Statistics gives recommendations on the matching of March CPS files. The March CPS 1968-1971 can be matched according to the random cluster code, the serial number, and month in the sample. The March CPS 1973-1975 can be matched according to the random cluster code, the segment number, the serial number, and month in the sample. For the period 1977-1985, 1986-1994, and 1996-2011, the same individual can be matched according to the household identification number, month in the sample, and personal line number. For the rest of the sample period, the redesign of March CPS survey prevents the matching of records. See the code book of March CPS for further details.

However, the actual matching rate based on the official matching criteria is considerably less than the theoretical upper limit of fifty percent because of the non-response, mortality, migration and reading errors. Moreover, relying on these identifiers alone can include false merges (the matched households are actually not the same one) and reject true merges (excluding the same household). This paper adopts the methodology as recommended in Madrian and Lefgren (1999) by adding more criteria. It only keeps those matched households whose gender and race information do not change in the two consecutive surveys. To reduce the probability of rejecting true merges, it keeps those matched households with age difference that is larger than -1 and smaller than 3 in the two consecutive years.

The sample includes family records with male households head aged between 25 and 60. It excludes female householders to avoid the severe sample selection issue related to female participation. It also excludes the households' record with annual growth rate of residual family labor earnings that is higher than 15 or less than 1/15.

Following Storesletten et al. (2004), I define contractions and expansions based on three statistics: the NBER definition of business cycles, the aggregate unemployment rate, and the growth rate of Gross National Product. According to the first criterion, the recessions are defined as those years for which the majority of the months are contractionary by the definition of NBER. For case in which contractions spanned less than 12 months but more than six months over two calendar years, I define the first year as a recession. According to the second criterion, the recessions are defined as years for which the unemployment rate was greater than 7.5 percent. If the rate was greater than 7.5 percent but fell more than 3 percent relative to previous year, it was not counted as a recession year. In addition, years for which the unemployment rate has risen by more than 3 percent were defined as recession years. According to the third criterion, the recessions are defined as those years for which the growth rate of real Gross National Product is below 2 percent.

## 4. Estimation

The time-series model specifications imply the following moment conditions:

$$E_t \mathbf{g}\left(i, h, t; \boldsymbol{\Theta}\right) = 0 \tag{5}$$

$\boldsymbol{\Theta} \equiv \{\rho, \sigma_\alpha, \sigma_\varepsilon, \sigma_E, \sigma_C\}$ is a $5 \times 1$ vector to be estimated. $h$ is the age of households head $i$ at year $t$. The components of $\mathbf{g}$ can be written as:

$$g^1\left(i, h, t; \boldsymbol{\Theta}\right) = u_{it}^h u_{it}^h - \left(\sigma_\alpha^2 + \sigma_\varepsilon^2\right) - \sum_{j=0}^{h-1} \rho^{2j}\left(I_{t-j}\sigma_E^2 + \left(1 - I_{t-j}\right)\sigma_C^2\right) \tag{6}$$

$$g^2\left(i, h, t; \boldsymbol{\Theta}\right) = u_{it}^h u_{it+1}^{h+1} - \sigma_\alpha^2 - \rho \sum_{j=0}^{h-1} \rho^{2j}\left(I_{t-j}\sigma_E^2 + \left(1 - I_{t-j}\right)\sigma_C^2\right) \tag{7}$$

where $h = 1, \ldots, H, t = 1, \ldots, T$. Therefore, there are $2HT$ moment conditions in all. Obviously, the total number of moment conditions available is larger than then total number of unknown parameters. The General Method of Moment (GMM) estimator solves the following minimization problem

$$\hat{\mathbf{\Theta}} = \arg_{\mathbf{\Theta}\in\Xi} \min \hat{\mathbf{g}}\left(\mathbf{\Theta}\right)' \hat{\mathbf{W}} \hat{\mathbf{g}}\left(\mathbf{\Theta}\right) \tag{8}$$

where $\hat{\mathbf{g}}\left(\mathbf{\Theta}\right)$ is a $2HT \times 1$ vector, which is the sample counterparts of $\mathbf{g}$

$$\hat{g}^1\left(h,t,N_{ht};\mathbf{\Theta}\right) = \frac{1}{N_{ht}} \sum_{i=1}^{N_{ht}} g^1\left(i,h,t;\mathbf{\Theta}\right) \tag{9}$$

$$\hat{g}^2\left(h,t,N_{ht};\mathbf{\Theta}\right) = \frac{1}{N_{ht}} \sum_{i=1}^{N_{ht}} g^2\left(i,h,t;\mathbf{\Theta}\right) \tag{10}$$

where $N_{ht}$ is the sample size of each $(h,t)$ cell. Let $N \equiv \min\{N_{ht}\}$ be smallest panel size in the sample. $\hat{\mathbf{W}}$ is a $2HT \times 2HT$ symmetric positive definite weighting matrix, which is to be specified later. Because the data used is an unbalanced panel, the covariance matrix must be adjusted accordingly. Denote $\kappa_{ht} \equiv \frac{N_{ht}}{N}$ and

$$\Lambda \equiv diag\left[\kappa_{ht},\kappa_{ht}\right]_{2HT\times 2HT} \tag{11}$$

which is a diagonal matrix. The block diagonal matrix for asymptotic covariance will be scaled by the matrix $\Lambda$.

Under the standard assumption for GMM, the estimator satisfies

$$\sqrt{N}\left(\hat{\mathbf{\Theta}} - \mathbf{\Theta}\right) \to N\left(0, Avar\left(\hat{\mathbf{\Theta}}\right)\right) \tag{12}$$

where the estimator of $Avar\left(\hat{\mathbf{\Theta}}\right)$ is

$$\frac{1}{N}\left(\hat{\mathbf{G}}'\hat{\mathbf{W}}\hat{\mathbf{G}}\right)^{-1} \hat{\mathbf{G}}'\hat{\mathbf{W}}\left(\hat{\mathbf{S}}\mathbf{\Lambda}\right)\hat{\mathbf{W}}\hat{\mathbf{G}}\left(\hat{\mathbf{G}}'\hat{\mathbf{W}}\hat{\mathbf{G}}\right)^{-1} \tag{13}$$

where

$$\hat{\mathbf{G}} = \left[ \frac{1}{N_{ht}} \sum_{i=1}^{N_{ht}} \frac{\mathbf{g}\left(i,h,t;\hat{\mathbf{\Theta}}\right)}{\partial \mathbf{\Theta}'} \right]_{2HT\times 5} \tag{14}$$

$$\hat{\mathbf{S}} = \left[ \frac{1}{N_{ht}} \sum_{i=1}^{N_{ht}} \mathbf{g}\left(i,h,t;\hat{\mathbf{\Theta}}\right) \mathbf{g}\left(i,h,t;\hat{\mathbf{\Theta}}\right)' \right]_{2HT\times 2HT} \tag{15}$$

More precisely,

$$\hat{\mathbf{G}} = \begin{bmatrix} \frac{\partial \hat{g}^1(h,t)}{\partial \rho} & \frac{\partial \hat{g}^1(h,t)}{\partial \sigma_\alpha} & \frac{\partial \hat{g}^1(h,t)}{\partial \sigma_\varepsilon} & \frac{\partial \hat{g}^1(h,t)}{\partial \sigma_E} & \frac{\partial \hat{g}^1(h,t)}{\partial \sigma_C} \\ \frac{\partial \hat{g}^2(h,t)}{\partial \rho} & \frac{\partial \hat{g}^2(h,t)}{\partial \sigma_\alpha} & \frac{\partial \hat{g}^2(h,t)}{\partial \sigma_\varepsilon} & \frac{\partial \hat{g}^2(h,t)}{\partial \sigma_E} & \frac{\partial \hat{g}^2(h,t)}{\partial \sigma_C} \end{bmatrix}_{2HT\times 5} \tag{16}$$

and

$$\hat{\mathbf{S}} = diag \begin{bmatrix} \frac{1}{N_{ht}} \sum_{i=1}^{N_{ht}} \left[g^1(i,h,t)\right]^2 & \frac{1}{N_{ht}} \sum_{i=1}^{N_{ht}} g^1(i,h,t)\, g^2(i,h,t) \\ \frac{1}{N_{ht}} \sum_{i=1}^{N_{ht}} g^1(i,h,t)\, g^2(i,h,t) & \frac{1}{N_{ht}} \sum_{i=1}^{N_{ht}} \left[g^2(i,h,t)\right]^2 \end{bmatrix} \tag{17}$$

which is a $2 \times 2$ block diagonal matrix. This is due to the fact that each households only stay in the sample for two periods. Hence, $g^1(i,h,t)$ and $g^2(i,h,t)$ are not correlated with other moment conditions.

It is standard in the literature to use the optimal weighing matrix $\hat{\mathbf{S}}^{-1}\left(\hat{\mathbf{\Theta}}\right)$. Following I use a two-step GMM procedure as described in Ogaki (1993) and name the results as the optimal weight minimum distance (OWMD) estimator. In the first step, set $\hat{\mathbf{W}} = \mathbf{I}$ and compute the equal weight minimum distance estimator $\hat{\mathbf{\Theta}}_0$. The second step is to Set $\hat{W} = \hat{\mathbf{S}}^{-1}\left(\hat{\mathbf{\Theta}}_0\right)$, compute $\hat{\mathbf{\Theta}}$.[1]

[1] I also use the identity matrix as the weighting matrix, known as equal weight minimum distance (EWMD) estimator. Altonji and Segal (1996) shows that it has many good properties when the sample size is small. The results of EWMD estimator is only slightly different from the OWMD estimator. Therefore, I stick to the OWMD estimator.

Table 1: Regression Result: The First-stage OLS

| | Constant | Age | $\frac{Age^2}{100}$ | $\frac{Age^3}{1000}$ | Family Size | Educ./Race/Year Dummies |
|---|---|---|---|---|---|---|
| Coef. | 10.1 | -.091 | .304 | -.029 | .029 | Yes |
| Std Err | .061 | .004 | .011 | .001 | .001 | — |

The overidentifying restriction (OIR) test statistic

$$OIR = N * \hat{\mathbf{g}}' \left(\hat{S}\Lambda\right)^{-1} \hat{\mathbf{g}} \tag{18}$$

is asymptotically distributed as $\chi^2 (2HT - 5)$ under $H_0 : E_t \mathbf{g} (i, h, t; \mathbf{\Theta}) = 0$.

Table 1 reports the first-stage regression result with R squared equal to .15. Table 2 reports the results from two-stage optimal GMM. There are 2520 moment conditions(from age 25 to 60 in 35 years). The overidentifying restriction statistics for three definitions of business cycles are 344.7, 344.3, and 348.1 respectively. Therefore, we can not reject the model specifications.

Compared with Storesletten et al. (2004), the GMM estimates using March CPS 1968-2011 show the following results. First, the auto-correlation is .973 in the benchmark case (NBER definition), which is more persistent than Storesletten et al. (2004). Second, the absolute value of idiosyncratic labor earning risk is smaller. Third, the idiosyncratic labor earnings is still strongly countercyclical. The standard deviation of persistent shocks increases by 72.5 percent in the recessions, compared with 75 percent in Storesletten et al. (2004) using NBER Indicators.

## 5. Conclusion

The paper estimates the household labor income process using March Current Population Survey 1968-2011. It exploits the rotating panel struc-

Table 2: Idiosyncratic Earnings Process: OWMD-GMM

| Business Cycle Indicators | $\rho$ | $\sigma_C$ | $\sigma_E$ | $\sigma_\varepsilon$ | $\sigma_\alpha$ | p-Value |
|---|---|---|---|---|---|---|
| NBER dates | .973 | .156 | .090 | .335 | .374 | .999 |
| | (.010) | (.026) | (.018) | (.003) | (.019) | |
| Unemployment Rate | .963 | .157 | .115 | .332 | .353 | .999 |
| | (.011) | (.025) | (.018) | (.003) | (.024) | |
| GNP Growth Rate | .978 | .145 | .085 | .336 | .378 | .999 |
| | (.010) | (.024) | (.018) | (.003) | (.018) | |

ture of the survey and matches the same individuals in the two consecutive years. There are two main findings from the GMM estimates. First, The auto-correlation of persistent shocks is very high, which is .973 in the benchmark specification. Second, the persistent shocks exhibit great countercyclicality. The standard deviation of persistent shocks increases by 72.5 percent from expansion period to contraction period (from .090 to .156). Therefore, the paper confirms that the results in Storesletten et al. (2004) still hold in a much larger data set over a longer period.

**Acknowledgement**

I am greatly indebted to my supervisor Kjetil Storesletten for his guidance, patience, and support. I acknowledge financial support from Leiv Eiriksson mobilitetsprogram 2010 and Den norske Bank's fund for social sciences 2011. I should thank Li Chao for excellent research assistance. All errors are my own.

## Appendix A. Data Definition

From 1968 to 1975, the family labor earnings is calculated as the sum of labor earnings from all family members. It consists of labor earnings from (1) wage or salary income, (2) social security or railroad retirement, (3) welfare or public assistance, (4) unemployment compensation, workmen's compensation, and government employee pension, (5) alimony, contribution, and anything else.

From 1976 to 1988, the family labor earnings are defined as the sum of family income from (1) wage and salary, (2) social security and railroad retirement, including money received from US, state, and local government, (3) income supplementary security, (4) public assistance and welfare, including aid to families with dependent children and other assistance, (5) veteran's payments, including veteran's payments, unemployment compensation, workmen 's compensation , (6) retirement income, including private pensions and annuities, military retirement, federal government's employee pensions, state or local government's pensions, and (7) child support.

After 1989, the family labor earnings are defined as the sum of family income from (1) wage and salary, (2) unemployment compensation, (3) worker's compensation, (4) social security, (5) income supplementary security, (6) public assistance and welfare, (7) veteran's benefits, (8) survivor's income, (9) disability payments, (10) retirement income, (11) education benefits, (12) child support, (13) alimony payments , and (14) Financial assistance income.

## References

Altonji, J.G., Segal, L.M., 1996. Small-sample bias in gmm estimation of covariance structures. Journal of Business & Economic Statistics 14, 353–66.

Madrian, B.C., Lefgren, L.J., 1999. A Note on Longitudinally Matching

Current Population Survey (CPS) Respondents. NBER Technical Working Papers 0247. National Bureau of Economic Research, Inc.

Ogaki, M., 1993. 17 generalized method of moments: Econometric applications, in: Econometrics. Elsevier. volume 11 of *Handbook of Statistics*, pp. 455 – 488.

Storesletten, K., Telmer, C.I., Yaron, A., 2004. Cyclical dynamics in idiosyncratic labor market risk. Journal of Political Economy 112, 695–717.