April 2011 Farm Labor Estimates Methodology

USDA NASS Research and Development Division
Stephen Busselberg, Mathematical Statistician
3251 Old Lee Highway, Room 305
Fairfax, Virginia 22030

1. Background

The Farm Labor Survey provides estimates of quarterly wage rates paid, number of workers hired, and average number of hours worked per week for field workers, livestock workers, supervisors, other workers, and all hired farm workers. These estimates are available for the U.S. and eighteen sub-regions comprised of mutually exclusive groups of states. In 2007, the January Farm Labor Survey was canceled due to budget shortfall. Official estimates for field workers, field and livestock workers, and all hired workers were provided by modeling historical data using time series techniques. In April 2011 the Farm Labor Survey was canceled due to a similar budget shortfall. This paper details the methodology used for modeling the April 2011 farm labor estimates. The April 2011 farm labor estimates include wage rates paid, hours per week worked, and number of workers hired at the U.S. and eighteen regional levels for field workers, field and livestock workers, and all hired workers.

2. Methodology

The class of models used for the April 2011 farm labor was Vector Autoregressive (VAR). The dependent variable is a vector of worker types (field, field and livestock, all hired) whose elements are indexed by *j=1,2,3.* The region is indexed by *i* and time is indexed by *t.* The base VAR form we will write as

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{bmatrix}_i = \sum_{k=1}^{P_i} \begin{bmatrix} \phi_{1,1,k} & \phi_{1,2,k} & \phi_{1,3,k} \\ \phi_{2,1,k} & \phi_{2,2,k} & \phi_{2,3,k} \\ \phi_{3,1,k} & \phi_{3,2,k} & \phi_{3,3,k} \end{bmatrix}_i \begin{bmatrix} x_{1,t-k} \\ x_{2,t-k} \\ x_{3,t-k} \end{bmatrix}_i + \begin{bmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \end{bmatrix}_i$$

For matrix notation the subscript *j* is omitted, and we write $\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{bmatrix}_i = \boldsymbol{x}_{it}$. $\boldsymbol{z}_{it}$ follows a Gaussian distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_i$. The regions are modeled independent of one another.

2.1 Wage Rates

The natural log transformation of the wage rate was modeled due to a time variant variance of the wage rate. A second order Taylor Series approximation was used to estimate the expected wage rate given the model estimate which equates to the expected natural log of the wage rate; i.e. if *w* is the wage rate, we want $E[w]$ but we have $E[\ln(w)]$ since we are modeling $\ln(w)$. A first order Taylor Series approximation was used to estimate the variance of the expected wage rate $\mathrm{var}(E[w])$. Estimates of the wage rate w for a region *i* and worker type *j* at time *t* were calculated as follows. Note that if the subscript *j* is omitted, the notation symbolizes a column vector of worker types.

2.1.1 The Expected Wage Rate

$$\hat{w}_{ijt} = E\left[w_{ijt}\right] = e^{\hat{\boldsymbol{y}}_{\boldsymbol{ijt}}}\left(1 + \frac{1}{2}\sigma^2_{\hat{\boldsymbol{y}}_{\boldsymbol{ijt}}}\right)$$

$$\hat{\boldsymbol{y}}_{it} = E[\ln(\boldsymbol{w}_{it})] = \ln(\boldsymbol{w}_{it-4}) + \boldsymbol{\alpha}_{it} + \sum_{k=1}^{P_i} \boldsymbol{\Phi}_{ik}[\ln(\boldsymbol{w}_{it-k}) - \ln(\boldsymbol{w}_{it-k-4}) - \boldsymbol{\alpha}_{it-k}]$$

$$\boldsymbol{\alpha}_{it} = \sum_{q=1}^{3} \boldsymbol{\beta}_{iq}[t^q - (t-4)^q]$$

The term $\sigma^2_{\hat{\boldsymbol{y}}_{ijt}}$ is the model prediction variance of the expected natural log transformation $\hat{\boldsymbol{y}}_{ijt}$. The first order Taylor Series approximation for the prediction variance of the expected wage rate $\hat{w}_{ijt}$ is

$$\sigma^2_{\hat{w}_{ijt}} = e^{2\hat{y}_{ijt}}\left(1 + \frac{1}{4}\sigma^2_{\hat{y}_{ijt}}\right)\sigma^2_{\hat{y}_{ijt}}$$

2.1.2 Derivation of the Expected Log Wage Rate

$\hat{\boldsymbol{y}}_{it}$ is derived as follows:

$$\boldsymbol{x}_{it} = \ln(\boldsymbol{w}_{it}) - \sum_{q=0}^{3} \boldsymbol{\beta}_{iq} t^q$$

$$\boldsymbol{u}_{it} = \boldsymbol{x}_{it} - \boldsymbol{x}_{it-4}$$

$$\boldsymbol{u}_{it} = \sum_{k=1}^{P_i} \boldsymbol{\Phi}_{ik}\boldsymbol{u}_{it-k} + \boldsymbol{z}_{it}$$

We first define *x* as the natural log of the wage rate and de-trend the natural log of the wage rate by subtracting out a cubic trend as a function of time. Cubic is the highest order polynomial trend for any region. We then define *u* as a seasonal fourth difference to filter out quarterly seasonality (See note at the end of this section). Next we model *u* as VAR with $\boldsymbol{z}_{it} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$. Substituting the second line from above into the third yields

$$\boldsymbol{x}_{it} - \boldsymbol{x}_{it-4} = \sum_{k=1}^{P_i} \boldsymbol{\Phi}_{ik}(\boldsymbol{x}_{it-k} - \boldsymbol{x}_{it-k-4}) + \boldsymbol{z}_{it}$$

Substituting the definition of $\boldsymbol{x}_{it}$ into the above and consolidating terms gives

$$\begin{aligned} \ln(\boldsymbol{w}_{it}) - \sum_{q=0}^{3} \boldsymbol{\beta}_{iq} t^q - \ln(\boldsymbol{w}_{it-4}) + \sum_{q=0}^{3} \boldsymbol{\beta}_{iq}(t-4)^q \\ = \sum_{k=1}^{P_i} \boldsymbol{\Phi}_{ik} \left[\ln(\boldsymbol{w}_{it-k}) - \sum_{q=0}^{3} \boldsymbol{\beta}_{iq}(t-k)^q - \ln(\boldsymbol{w}_{it-k-4}) + \sum_{q=0}^{3} \boldsymbol{\beta}_{iq}(t-k-4)^q \right] + \boldsymbol{z}_{it} \end{aligned}$$

$$\begin{aligned} \ln(\boldsymbol{w}_{it}) - \ln(\boldsymbol{w}_{it-4}) - \sum_{q=1}^{3} \boldsymbol{\beta}_{iq}[t^q - (t-4)^q] \\ = \sum_{k=1}^{P_i} \boldsymbol{\Phi}_{ik} \left\{ \ln(\boldsymbol{w}_{it-k}) - \ln(\boldsymbol{w}_{it-k-4}) - \sum_{q=1}^{3} \boldsymbol{\beta}_{iq}[(t-k)^q - (t-k-4)^q] \right\} + \boldsymbol{z}_{it} \end{aligned}$$

$$\begin{aligned} \ln(\boldsymbol{w}_{it}) = \ln(\boldsymbol{w}_{it-4}) + \sum_{q=1}^{3} \boldsymbol{\beta}_{iq}[t^q - (t-4)^q] \\ + \sum_{k=1}^{P_i} \boldsymbol{\Phi}_{ik} \left\{ \ln(\boldsymbol{w}_{it-k}) - \ln(\boldsymbol{w}_{it-k-4}) - \sum_{q=1}^{3} \boldsymbol{\beta}_{iq}[(t-k)^q - (t-k-4)^q] \right\} + \boldsymbol{z}_{it} \end{aligned}$$

Last we take the expectation of both sides of the above equation

$$\begin{aligned} \widehat{\boldsymbol{y}}_{it} &= E[\ln(\boldsymbol{w}_{it})] \\ &= \ln(\boldsymbol{w}_{it-4}) + \sum_{q=1}^{3} \boldsymbol{\beta}_{iq}[t^q - (t-4)^q] \\ &\quad + \sum_{k=1}^{P_i} \boldsymbol{\Phi}_{ik} \left\{ \ln(\boldsymbol{w}_{it-k}) - \ln(\boldsymbol{w}_{it-k-4}) - \sum_{q=1}^{3} \boldsymbol{\beta}_{iq}[(t-k)^q - (t-k-4)^q] \right\} \end{aligned}$$

$$= \ln(\boldsymbol{w}_{it-4}) + \boldsymbol{\alpha}_{it} + \sum_{k=1}^{P_i} \boldsymbol{\Phi}_{ik}[\ln(\boldsymbol{w}_{it-k}) - \ln(\boldsymbol{w}_{it-k-4}) - \boldsymbol{\alpha}_{it-k}]$$

$$\boldsymbol{\alpha}_{it} = \sum_{q=1}^{3} \boldsymbol{\beta}_{iq}[t^q - (t-4)^q]$$

The autoregressive order P_i is determined by region using a minimum Akaike's adjusted Information Criterion (AICc).

Note: Sinusoidal seasonal de-trending is also an alternative method for detrending and including a seasonal adjustment; i.e. $\widehat{\boldsymbol{y}}_{it} = \boldsymbol{\alpha}_{it} + \sum_{k=1}^{P_i} \boldsymbol{\Phi}_{ik}[\ln(\boldsymbol{w}_{it-k}) - \boldsymbol{\alpha}_{it-k}]$ where $\boldsymbol{\alpha}_{it} = \sum_{q=0}^{3} \boldsymbol{\beta}_{iq} t^q + \boldsymbol{\beta}_{i4} \sin\frac{1}{2}\pi t + \boldsymbol{\beta}_{i5} \cos\frac{1}{2}\pi t$. This option was not pursued as it was not favorable for all regions.

2.1.3 Derivation of the Expected Wage Rate

We can approximate the expected wage rate $\widehat{w} = E[w] = E\left[e^{\ln w}\right] = E[e^y] = E[f(y)]$. The second order Taylor Series expansion about y is

$$f(y) = f(\hat{y}) + f'(\hat{y})(y-\hat{y}) + \frac{1}{2}f''(\hat{y})(y-\hat{y})^2 + R$$
$$f(y) \approx f(\hat{y}) + f'(\hat{y})(y-\hat{y}) + \frac{1}{2}f''(\hat{y})(y-\hat{y})^2$$

where R is the Taylor Series remainder. If we take the expectation of both sides

$$E[f(y)] \approx f(\hat{y}) + f'(\hat{y})E[y-\hat{y}] + \frac{1}{2}f''(\hat{y})E[(y-\hat{y})^2]$$
$$\approx f(\hat{y}) + \frac{1}{2}f''(\hat{y})\mathrm{VAR}[y]$$

The above relationship yields

$$E[w] \approx e^{\hat{y}} + \frac{1}{2}e^{\hat{y}}\sigma_y^2$$
$$\approx e^{\hat{y}}\left(1 + \frac{1}{2}\sigma_{\hat{y}}^2\right)$$

2.1.4 Derivation of the Variance of the Expected Wage Rate Approximation

For the first order Taylor Series approximation of the variance, we first establish the first order expansion. It is the first two terms in the second order expansion in section 2.1.3.

$$f(y) \approx f(\hat{y}) + f'(\hat{y})(y - \hat{y})$$

We can then subtract the second order approximation for $E[f(y)]$ from both sides.

$$\begin{aligned} f(y) - E[f(y)] &\approx f(\hat{y}) + f'(\hat{y})(y - \hat{y}) - f(\hat{y}) - \frac{1}{2} f''(\hat{y})\mathrm{VAR}[y] \\ &\approx f'(\hat{y})(y - \hat{y}) - \frac{1}{2} f''(\hat{y})\mathrm{VAR}[y] \end{aligned}$$

Squaring both sides and taking the expectation yields

$$\begin{aligned} E[\{f(y) - E[f(y)]\}^2] &\approx f'(\hat{y})^2 E[(y - \hat{y})^2] + \frac{1}{4} f''(\hat{y})^2 \mathrm{VAR}[y]^2 \\ &\approx f'(\hat{y})^2 \mathrm{VAR}[y] + \frac{1}{4} f''(\hat{y})^2 \mathrm{VAR}[y]^2 \end{aligned}$$

Plugging in the definition of the terms above yields

$$\begin{aligned} \sigma_{\hat{w}}^2 &\approx e^{2\hat{y}} \sigma_{\hat{y}}^2 + \frac{1}{4} e^{2\hat{y}} \sigma_{\hat{y}}^4 \\ &\approx e^{2\hat{y}} \left(1 + \frac{1}{4} \sigma_{\hat{y}}^2\right) \sigma_{\hat{y}}^2 \end{aligned}$$

2.2 Hours and Number of Workers

Hours and number of workers were modeled using the same seasonal differencing transformation and are therefore similar in structure. Letting $\boldsymbol{x}_{it}$ represent the vector of worker types for hours per week or number of workers hired, the April 2011 estimate for hours per week or number of workers is

$$\hat{\boldsymbol{x}}_{it} = E[\boldsymbol{x}_{it}] = \boldsymbol{x}_{it-4} + \sum_{k=1}^{P_i} \Phi_{ik} (\boldsymbol{x}_{it-k} - \boldsymbol{x}_{it-k-4})$$

This is the vector equivalent of an ARIMA$(P_i, 0,0) \times (0,1,0)_4$.

References

Casella, George and Berger, Roger L. Statistical Inference. Pacific Grove, CA: Thomas Learning Inc., 2002.

Shumway, R.H. and Stoffer, D.S. Time Series Analysis and its Applications. New York: Springer Science+Business Media, LLC., 2006.