# Estimating and Testing VARs for Firm Employment and Wages

Manuel Arellano

- We discuss various aspects of inference with autoregressive models in the context of an empirical illustration.
- We consider autoregressive employment and wage equations estimated from the panel of firms used by Alonso-Borrego and Arellano (1999).
- This is a balanced panel of 738 Spanish manufacturing companies, for which there are available annual observations for the period 1983-1990.
- We consider various specializations of a bivariate VAR(2) model for the logs of employment and wages, denoted $n_{it}$ and $w_{it}$ respectively.
- Individual and time effects are included in both equations.
- The form of the model is

$$n_{it} = \delta_{1t} + \alpha_1 n_{i(t-1)} + \alpha_2 n_{i(t-2)} + \beta_1 w_{i(t-1)} + \beta_2 w_{i(t-2)} + \eta_{1i} + v_{1it} \quad (1)$$

$$w_{it} = \delta_{2t} + \gamma_1 w_{i(t-1)} + \gamma_2 w_{i(t-2)} + \lambda_1 n_{i(t-1)} + \lambda_2 n_{i(t-2)} + \eta_{2i} + v_{2it}. \quad (2)$$

**Univariate AR Estimates for Employment**

- We begin by obtaining alternative estimates of a univariate AR(1) model for employment (setting $\alpha_2 = \beta_1 = \beta_2 = 0$).
- Table 6.3 compares OLS estimates in levels, first-differences, and within-groups with those obtained by GMM using as instruments for the equation in first differences all lags of employment up to $t-2$. The results are broadly consistent with what would be expected for an AR data generation process with unobserved heterogeneity.
- Taking GMM estimates as a benchmark, OLS in levels is biased upwards, and WG and OLS in differences are biased downwards, with a much larger bias in the latter.
- The one- and two-step GMM estimates in the $4$-th and $5$-th columns, respectively, are based on the sample moments $b_N(\beta) = (b'_{3N}, ..., b'_{8N})'$, where $\beta$ is the $7 \times 1$ parameter vector $\beta = (\alpha, \Delta\delta_3, ..., \Delta\delta_8)'$ and
$$b_{tN} = \frac{1}{738} \sum_{i=1}^{738} \begin{pmatrix} 1 \\ n_i^{t-2} \end{pmatrix} \left( \Delta n_{it} - \Delta\delta_t - \alpha \Delta n_{i(t-1)} \right) \;\; (t = 3, ..., 8). \tag{3}$$
$b_N(\beta)$ contains $27$ orthogonality conditions in total, so that there are $20$ overidentifying restrictions.
- These are tested with the Sargan statistic. There is a contrast between the value of the one-step Sargan statistic (35.1), which is too high for a chi-square with 20 degrees of freedom, and the robust two-step statistic which is much smaller (15.5).

- This should not be taken as evidence against the overidentifying restrictions, but as an indication of the presence of conditional heteroskedasticity.
- Column 6 in Table 6.3 reports two-step GMM estimates of an AR(2) model. Since one cross-section is spent in constructing the second lag, the two orthogonality conditions in $b_{3N}$ are lost, so we are left with $25$ moments. There is a second autoregressive coefficient but $\Delta\delta_3$ is lost, so the total number of parameters is unchanged.
- Finally, the last column in Table 6.3 presents continuously updated GMM estimates of the AR(2) model. They use the same moments as GMM2, but the weight matrix is continuously updated.

Table 6.3
Univariate AR Estimates for Employment

| | OLS-levels | OLS-dif. | WG | GMM1 | GMM2 | GMM2 | C.U. GMM2 |
|---|---|---|---|---|---|---|---|
| $n_{i(t-1)}$ | $0.992$ | $0.054$ | $0.69$ | $0.86$ | $0.89$ | $0.75$ | $0.83$ |
| | $(0.001)$ | $(0.026)$ | $(0.025)$ | $(0.07)$ | $(0.06)$ | $(0.09)$ | $(0.09)$ |
| $n_{i(t-2)}$ | | | | | | $0.04$ | $0.03$ |
| | | | | | | $(0.02)$ | $(0.02)$ |
| Sargan | – | – | – | $35.1$ | $15.5$ | $14.4$ | $13.0$ |
| (d.f.) | | | | $(20)$ | $(20)$ | $(18)$ | $(18)$ |
| $m_1$ | $2.3$ | $-0.6$ | $-9.0$ | $-8.0$ | $-7.6$ | $-6.0$ | |
| $m_2$ | $2.2$ | $2.3$ | $0.6$ | $0.5$ | $0.5$ | $0.3$ | |

$N = 738, T = 8, 1983 - 1990$. Heteroskedasticity robust standard errors in parentheses. Time dummies included in all equations.

- From the orthogonality conditions above only first-differences of time effects are directly estimated. The initial time effect can be estimated as

$$\widehat{\delta}_3 = \frac{1}{738} \sum_{i=1}^{738} (y_{i3} - \widehat{\alpha}_1 y_{i2} - \widehat{\alpha}_2 y_{i1}) \tag{4}$$

  and, given estimates of their changes, the rest can be estimated recursively from $\widehat{\delta}_t = \widehat{\Delta\delta}_t + \widehat{\delta}_{t-1}$ $(t = 4, ..., 8)$.

- Given the large cross-sectional sample size, the realizations of the time effects in the data can be accurately estimated, but with only $6$ time series observations we do not have enough information to consider a stochastic model for $\delta_t$.

- On the other hand, individual effects can be estimated as

$$\widehat{\eta}_i = \frac{1}{T-2} \sum_{s=3}^{T} \widehat{u}_{is} \tag{5}$$

  where $\widehat{u}_{is} = y_{is} - \widehat{\delta}_s - \widehat{\alpha}_1 y_{i(s-1)} - \widehat{\alpha}_2 y_{i(s-2)}$.

- Here the situation is the reverse. Since the $\widehat{\eta}_i$ are averages of just $T - 2 = 6$ observations, they will typically be very noisy estimates of realizations of the effects for particular firms.

- However, the variance of $\eta_i$ can still be consistently estimated for large $N$.

- Optimal estimation of $\sigma_\eta^2$ and the $\sigma_t^2$ requires consideration of the data covariance structure, but noting that the errors in levels $u_{it} \equiv \eta_i + v_{it}$ satisfy $Var(u_{it}) = \sigma_\eta^2 + \sigma_t^2$ and $Cov(u_{it}, u_{is}) = \sigma_\eta^2$, simple consistent estimates can be obtained as:

$$\widehat{\sigma}_\eta^2 = \frac{2}{T(T-1)} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \widehat{Cov}\left(\widehat{u}_{it}, \widehat{u}_{is}\right) \tag{6}$$

$$\widehat{\sigma}_t^2 = \widehat{Var}(\widehat{u}_{it}) - \widehat{\sigma}_\eta^2. \tag{7}$$

- For the AR(2) employment equation Alonso-Borrego and Arellano reported $\widehat{\sigma}_\eta^2 = .038$ and $T^{-1} \sum_{t=1}^{T} \widehat{\sigma}_t^2 = .01$. Thus, variation in firm specific intercepts was approximately $4$ times larger than the average random error variance.

- In this example time dummies are important for the model to be accepted by the data. Without them, GMM2 estimates of the AR(2) employment equation in first differences yielded a Sargan statistic of $59.0$ ($d.f.18$) without constant, and of $62.7$ ($d.f.18$) with constant. Thus, implying a sound rejection of the overidentifying restrictions.

- For the firms in our data set, average growth of employment during the $7$ year period 1984-90 is $1$ percent, but this is the result of almost no growth in the first two years, $1$ percent growth in 1986, $2$ percent in 1987-89 and zero or negative growth in 1990.

- Given such pattern, it is not surprising that we reject the restrictions imposed by the cross-sectional orthogonality conditions with a common intercept or a linear trend.

**Bivariate VAR Estimates for Employment and Wages**

- For the rest of the tutorial we focus on the bivariate model (1)-(2) since it allows us to illustrate a richer class of problems.
- Table 6.4 presents OLS in levels and GMM2 in differences for employment (columns 1 and 2), and wages (columns 4 and 5).
- The table also contains GMM estimates that combine levels and differences, but these will be discussed below in conjunction with testing for mean stationarity.
- In line with the univariate results, the OLS estimates in levels for both equations are markedly different to GMM2 in differences, and imply a substantially higher degree of persistence, which is consistent with the presence of heterogeneous intercepts.
- The GMM estimates use as instruments for the equations in first differences all the available lags of employment and wages up to $t-2$. With $T = 8$, a second-order VAR and time dummies, there are 36 overidentifying restrictions for each equation. Neither of the Sargan test statistics provide evidence against these restrictions.
- It may be possible to improve the efficiency by jointly estimating the two equations. Optimal joint GMM estimates would use a weight matrix that takes into account the correlation between the moment conditions of the employment and wage equations.

Table 6.4
VAR Estimates

| | Employment | | | Wages | | |
|---|---|---|---|---|---|---|
| | OLS-levels | GMM2 dif. | GMM2 lev.&dif. | OLS-levels | GMM2 dif. | GMM2 lev.&dif. |
| $n_{i(t-1)}$ | 1.11 | 0.84 | 1.17 | 0.08 | $-0.04$ | 0.08 |
| | (0.03) | (0.09) | (0.03) | (0.03) | (0.10) | (0.03) |
| $n_{i(t-2)}$ | $-0.12$ | $-0.003$ | $-0.13$ | $-0.07$ | 0.05 | $-0.06$ |
| | (0.03) | (0.03) | (0.02) | (0.03) | (0.03) | (0.02) |
| $w_{i(t-1)}$ | 0.14 | 0.08 | 0.13 | 0.78 | 0.26 | 0.78 |
| | (0.03) | (0.08) | (0.02) | (0.03) | (0.11) | (0.02) |
| $w_{i(t-2)}$ | $-0.11$ | $-0.05$ | $-0.11$ | 0.18 | 0.02 | 0.08 |
| | (0.03) | (0.02) | (0.02) | (0.03) | (0.02) | (0.02) |
| $\chi^2_{ce}(2)$ | 41.7 | 7.2 | 43.7 | 26.1 | 3.3 | 10.4 |
| $p$-value | 0.00 | 0.03 | 0.00 | 0.00 | 0.19 | 0.006 |
| Sargan | – | 36.9 | 61.2 | – | 21.4 | 64.2 |
| (d.f.) | | (36) | (48) | | (36) | (48) |
| $p$-value | | 0.43 | 0.096 | | 0.97 | 0.06 |
| $m_1$ | $-0.6$ | $-6.8$ | $-8.0$ | 0.05 | $-5.7$ | $-9.5$ |
| $m_2$ | 1.6 | 0.2 | 1.3 | $-2.7$ | 0.5 | $-0.6$ |

$N = 738, T = 8, 1983 - 1990$. Heteroskedasticity robust standard errors in parentheses. Time dummies included in all equations.

$\chi^2_{ce}(2)$ is a Wald test statistic of the joint significance of cross effects.

**Testing for Residual Serial Correlation**

- If the errors in levels are serially independent, those in first differences will exhibit first- but not second-order serial correlation.
- Moreover, the first-order serial correlation coefficient should be equal to $-0.5$.
- In this regard, an informal but often useful diagnostic is provided by the inspection of the autocorrelation matrix for the errors in first differences.
- Serial correlation matrices for employment and wages based on GMM residuals in first-differences are shown in Table 6.5, broadly conforming to the expected pattern.

Table 6.5

(a) GMM1 (dif.) Residual Serial Correlation Matrix for Employment

$$\begin{pmatrix} 1. & & & & \\ -.53 & 1. & & & \\ .10 & -.49 & 1. & & \\ -.04 & -.015 & -.46 & 1. & \\ -.015 & .04 & -.08 & -.44 & 1. \end{pmatrix}$$

(b) GMM1 (dif.) Residual Serial Correlation Matrix for Wages

$$\begin{pmatrix} 1. & & & & \\ -.51 & 1. & & & \\ .03 & -.33 & 1. & & \\ .004 & -.035 & -.42 & 1. & \\ .009 & .00 & -.03 & -.39 & 1. \end{pmatrix}$$

- Formal tests of serial correlation are provided by the $m_1$ and $m_2$ statistics reported in Table 6.4 for the VAR model (and also in Table 6.3 for the univariate results).
- They are asymptotically distributed as $\mathcal{N}(0,1)$ under the null of no autocorrelation, and have been calculated from residuals in first differences (except for OLS in levels).
- So if the errors in levels were uncorrelated, we would expect $m1$ to be significant, but not $m2$, as is the case for the GMM2-dif estimates for employment and wages.
- The $m_j$ statistics (Arellano and Bond, 1991) are moment tests of significance of the average $j$-th order autocovariance $r_j$:

$$r_j = \frac{1}{T-3-j} \sum_{t=4+j}^{T} r_{tj} \tag{8}$$

where $r_{tj} = E\left(\Delta v_{it} \Delta v_{i(t-j)}\right)$. Their null is $H_0 : r_j = 0$ and they are given by

$$m_j = \frac{\widehat{r}_j}{SE(\widehat{r}_j)} \tag{9}$$

where $\widehat{r}_j$ is the sample counterpart of $r_j$ based on first-difference residuals $\widehat{\Delta v}_{it}$ and $\widehat{r}_{tj} = N^{-1} \sum_{i=1}^{N} \widehat{\Delta v}_{it} \widehat{\Delta v}_{i(t-j)}$.

- The estimates in Table 6.4 are based on the assumption that given individual and time effects $n_{it}$ and $w_{it}$ only depend on the past two observations. Provided $T$ is sufficiently large, the $m_j$ statistics can be used to test assumptions on lag length.

## Testing for Stationarity in Mean of Initial Observations

- We turn to consider GMM estimates that combine levels and differences, as shown in columns 3 (employment) and 6 (wages) of Table 6.4.
- For the employment equation, estimates are based on the following $40$ moments for errors in differences:

$$b^d_{tN} = \sum_{i=1}^{738} \begin{pmatrix} n_i^{t-2} \\ w_i^{t-2} \end{pmatrix} \left(\Delta n_{it} - \Delta\delta_{1t} - \alpha_1 \Delta n_{i(t-1)} - \alpha_2 \Delta n_{i(t-2)} - \beta_1 \Delta w_{i(t-1)} - \beta_2 \Delta w_{i(t-2)}\right) \tag{10}$$

$$(t = 4, ..., 8),$$

together with $6$ moments for the period-specific constants:

$$b^c_{tN} = \sum_{i=1}^{738} \left(n_{it} - \delta_{1t} - \alpha_1 n_{i(t-1)} - \alpha_2 n_{i(t-2)} - \beta_1 w_{i(t-1)} - \beta_2 w_{i(t-2)}\right) \ (t = 3, ..., 8), \tag{11}$$

and $12$ additional moments for errors in levels:

$$b^\ell_{tN} = \sum_{i=1}^{738} \begin{pmatrix} \Delta n_{i(t-1)} \\ \Delta w_{i(t-1)} \end{pmatrix} \left(n_{it} - \delta_{1t} - \alpha_1 n_{i(t-1)} - \alpha_2 n_{i(t-2)} - \beta_1 w_{i(t-1)} - \beta_2 w_{i(t-2)}\right) \tag{12}$$

$$(t = 3, ..., 8).$$

The moments are functions of the $10\times 1$ parameter vector $\beta = (\delta_3, ..., \delta_8, \alpha_1, \alpha_2, \beta_1, \beta_2)$, so that there are $48$ overidentifying restrictions.

- The estimates for the wage equation were obtained in exactly the same manner.
- Employment and wage changes lagged two periods or more are not used as instruments for the equations in levels because they are redundant given those already included.
- We report two-step robust estimates whose weight matrix is based on the kind of one-step residuals described above.
- Note that, contrary to what we would expect under mean stationarity, the combined levels & differences GMM estimates in both equations are closer to the OLS-levels estimates than to GMM in differences.
- A test of the moment restrictions (12) is a test of whether, given an aggregate time effect, the mean of the distribution of initial observations and the mean of the steady state distribution coincide.
- This can be done by computing incremental Sargan test statistics. Specifically, under the null of mean stationarity, the difference between the $lev.\&dif.$ and the $dif.$ Sargan statistics would be asymptotically distributed as a $\chi^2$ with $12$ degrees of freedom.
- Since we obtain $\Delta S_n = 24.3$ ($p$-val. $0.0185$) for employment, and $\Delta S_w = 42.8$ ($p$-val. $0.00$) for wages, the null is rejected for the two equations, although somewhat more marginally so in the case of employment.

**Testing for the Presence of Unobserved Heterogeneity**

- In the absence of unobserved heterogeneity OLS in levels are consistent estimates, but more generally estimation (eg. of the employment equation) could be based on the following $60$ sample moments

$$b_{tN}^{*} = \sum_{i=1}^{738} \begin{pmatrix} 1 \\ n_i^{t-1} \\ w_i^{t-1} \end{pmatrix} \left(n_{it} - \delta_{1t} - \alpha_1 n_{i(t-1)} - \alpha_2 n_{i(t-2)} - \beta_1 w_{i(t-1)} - \beta_2 w_{i(t-2)}\right) \tag{13}$$

$$(t = 3, ..., 8).$$

- Given the $46$ moments in (10) and (11), (13) adds the following $14$ moments:

$$b_{3N}^{h} = \sum_{i=1}^{738} \begin{pmatrix} n_{i1} \\ n_{i2} \\ w_{i1} \\ w_{i2} \end{pmatrix} (n_{i3} - \delta_{13} - \alpha_1 n_2 - \alpha_2 n_{i1} - \beta_1 w_{i2} - \beta_2 w_{i1}) \tag{14}$$

$$b_{tN}^{h} = \sum_{i=1}^{738} \begin{pmatrix} n_{i(t-1)} \\ w_{i(t-1)} \end{pmatrix} \left(n_{it} - \delta_{1t} - \alpha_1 n_{i(t-1)} - \alpha_2 n_{i(t-2)} - \beta_1 w_{i(t-1)} - \beta_2 w_{i(t-2)}\right) \tag{15}$$

$$(t = 4, ..., 8).$$

- Thus a test for the validity of the moments (14) and (15) can be regarded as testing for the presence of unobserved heterogeneity.

- This can be done by calculating combined GMM estimates based on (10), (11), (14) and (15) -or equivalently levels-GMM estimates based on (13)- and obtaining the corresponding incremental Sargan tests relative to GMM in differences.

- The resulting estimates for employment and wages are very close to OLS, and both incremental tests reject the absence of unobserved heterogeneity. The incremental Sargan statistics ($d.f. = 14$) take the values $\Delta S_n^h = 36.0$ ($p$-val. $0.001$) for employment, and $\Delta S_w^h = 47.2$ ($p$-val. $0.00$) for wages.

**Testing for Granger Non-Causality with and without Heterogeneity**

- The hypothesis that employment does not Granger-cause wages conditional on individual and time effects imposes the restrictions $\lambda_1 = \lambda_2 = 0$. Conversely, to test whether wages Granger-cause employment we examine the validity of $\beta_1 = \beta_2 = 0$.
- The testing of these restrictions is of some interest in our example because a version of model (1)-(2) in which the wage equation only includes its own lags can be regarded as the reduced form of an intertemporal labour demand model under rational expectations (as in Sargent, 1978).
- Wald test statistics of the joint significance of cross-effects are reported in Table 6.4 for the two equations. For the GMM2 estimates in first-differences we find that wages Granger-cause employment, but employment does not Granger-cause wages.
- An interesting point is that conditioning on individual effects is crucial for this result. As shown in Table 6.4, if the tests were based upon the OLS estimates in levels, the hypothesis that employment does not Granger-cause wages would be clearly rejected. This illustrates how lack of control of individual heterogeneity could result in a spurious rejection of non causality.
- Moreover, Granger non-causality would also be rejected using the estimates that impose mean stationarity of the initial observations. Thus, in short panels assumptions about initial conditions also matter for the assessment of non causality.