

## Appendix: Labor Data

The labor data are of poor quality. First, provincial employment data according to the three sectors are available in the National Bureau of Statistics (NBS) publication *Fifty Years* for the years 1978-98. Since 1993, these provincial data are also available in the *Statistical Yearbook*, with not necessarily the same data for the overlapping years 1993-98, or the here reported overlapping years 1995-98. This causes the appearance of two values for these years in the time series charts. For some provinces, the difference is striking; for example, total employment in Gansu province in 1998 stood at, depending on source, 15.40 vs. 11.76m. For others, it is mainly a sectoral re-shuffle; for example, according to one source, employment in the secondary and tertiary sectors in Sichuan in 1998 were 8.24 and 8.85m (together, 17.09m), but according to another source 7.17 and 10.28m (together, 17.45m).

Second, even within one source, the data are not necessarily consistent over time. For example, in Heilongjiang, between 1997 and 1998 (in one and the same source), employment in the primary sector increased from 5.82 to 8.27m, while it fell in the secondary sector from 5.12m to 3.87m, and in the tertiary sector from 5.54 to 4.87m.

Third, the sum across provinces of provincial total employment (either as published, or as sum of the three sectors, with identical results) yields a national employment value that is always very close to the national employment data published by the NBS for an exhaustive approximately 15 sectors; however, these data are based on regular report forms and do not capture all employment.

Starting with 1990, the NBS publishes a new total based on report form data *and* adjustments in accordance with population censuses and surveys, which in 1990 exceeded the report form total (previously economy-wide total) by 15%. The provincial employment data do not adopt adjustments, and thus constitute underestimates of varying degrees in all years. While in 1978-1989, the sum provincial data came close to the published national total (the report form total), since 1990 it falls 10-15% short of the (new, adjusted) national total (see Figure 1). I.e., the provincial economy-wide employment data are likely to fall about 10-15% short of (unknown) provincial economy-wide employment in any one year. What this means for *individual sectors* (whose employment, within each province, add up to provincial economy-wide employment) is unclear.

Fourth, provincial primary sector employment data always fall less short of the national data than do provincial secondary sector employment data (Figure 1).<sup>1</sup> In other words, if the national data were correct, provincial primary sector employment data tend to be overestimates in comparison to provincial secondary sector employment data. This matters for AY's analysis if the degree of overestimation changes over time.<sup>2</sup>

---

<sup>1</sup> In 1978-1989, when the national data are the (low) report form data, the sum provincial primary sector employment actually *exceeds* national primary sector employment.

<sup>2</sup> A further change in employment statistics is that starting with the 1998 report form employment data, the *Statistical Yearbook* only reports those laborers who are actually "on their post" (*zai gang*), i.e., no longer includes

Fifth, Thomas Rawski and Robert Mead (1998) find that official employment data “massively overestimate the number of Chinese farm workers” (on the order of 100m, p. 767) because they tend to represent residuals; as/if the number of migrant workers increases, and Thomas Rawski and Robert Mead think the over-count was minimal in 1979, this is likely to be ever more so the case.<sup>3</sup>

The effect of these data shortcomings on AY’s conclusions on *variances* across provinces, in his third quantitative analysis, is unclear.

In his fourth quantitative analysis, the primary sector employment data enter directly into the regressions; to the extent that provinces adopted different compilation methods at different points of time, and the extent to which provincial idiosyncrasies are correlated with other explanatory variables, the regression results are affected. But it is, again, not clear to me in what way.

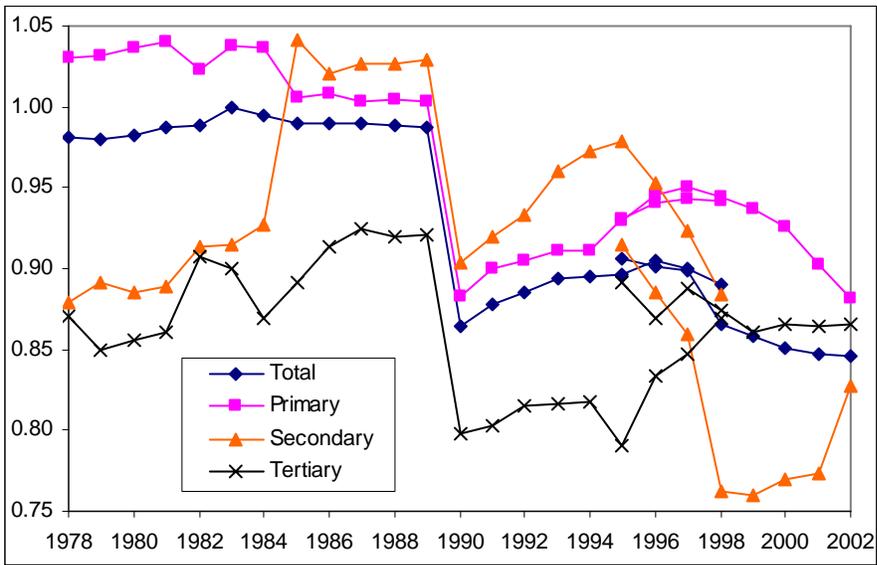
### ***References only used in this appendix***

Rawski, Thomas G., and Robert W. Mead. “On the Trail of China’s Phantom Farmers.” *World Development* 26, no. 5 (May 1998): 767-81.

---

those who are still part of a state-owned unit but no longer actively work in it; nationwide, secondary sector employment stagnated in 1998 (compared to 1997; *Statistical Yearbook 1999*, p. 134). Provincial data tend to show a similar stagnation in 1998.

<sup>3</sup> Albert Park and Yang Du (2003) suspect that AY’s findings in his third section might altogether be driven by the overestimation of primary sector employment data. The official primary sector employment data reported by Thomas Rawski and Robert Mead (1998) in 1990 through 1995 (their final year) falls approximately 30m short of the official data reported in the *Statistical Yearbook 2005*, p. 118.



National employment data are from the *Statistical Yearbook 2005*, p. 118. Provincial employment data through 1998 are from *Fifty Years*, with occasional holes filled using *Seventeen Years*. Provincial *sectoral* data exclude Tianjin in the years 1978-84, for which no data are available; total employment in Tianjin in 1978 and 1984 was 0.92% of sum provincial employment in both years. Provincial employment data in the second series, from 1995-2002, are from the corresponding individual issues of the *Statistical Yearbook*.

**Figure 1. Sum Provincial Relative to National Employment Data**