SIMULATING JOLTS HIRES AND SEPARATIONS DATA USING HISTORICAL QCEW DATA October 2010

Mark Crankshaw U.S. Bureau of Labor Statistics
2 Massachusetts Avenue NE, Suite 4985, Washington, DC 20212

Abstract

Keywords: JOLTS, simulation, synthetic data

The Bureau of Labor Statistics produces the monthly JOLTS survey which is a 16,000 unit sample of business establishments drawn from a population frame of over 8 million establishments. While the JOLTS sample is allocated and selected with the goal of accurately reflecting the general composition of the population, the JOLTS sample does not reflect the general composition of the population with respect to the establishment age. If these very young establishments systematically differ from relatively older establishments with respect to hires and separations rates, then the JOLTS hires and separations rates may be biased. A simulation of JOLTS data using historical QCEW employment data was developed to determine the nature and extent of the bias. It was found that JOLTS respondents systematically underreport separations, the JOLTS sample is over-representative of older more stable firms, and that the JOLTS sample did not fully capture those hires and separations that accrue to firms entering and exiting the labor market. As a result of the simulation, remedial measures have been undertaken by JOLTS to correct these shortcomings.

Disclaimer: *The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the Bureau of Labor Statistics or the United States government.*

1. INTRODUCTION

The JOLTS survey is a 16,000 unit sample of business establishments drawn from a population frame (the Bureau of Labor Statistics' Quarterly Census of Employment and Wages Longitudinal Data Base or LDB) of over 8 million establishments that attempts to measure US labor market dynamism. The JOLTS sample is stratified by NAICS Industry Division, Census region, and establishment size and while the sample may accurately reflecting the general composition of the population in those respects the JOLTS sample may not reflect the general composition of the population in other important respects (namely, with respect to the age of establishment). Since it takes a considerable amount of time (8-12 months) to create the frame, allocate and select a sample, and to contact and enroll respondents into the survey, it is likely that the population frame from which the survey is drawn no longer reflects the current population of business establishments, particularly with respect to establishment age. Currently, the JOLTS survey has no way to sample those very young establishments (new businesses) that came into existence during the 8-12 month period of lag needed to enroll establishments into the survey. If

these very young establishments systematically differ from relatively older establishments with respect to hires and separations rates, then the JOLTS hires and separations rates may be biased.

JOLTS respondents report employment, hires and separations. It is expected that changes in employment should be reflected in the number of hires net separations. However, owing to the fact that employment has a different time reference (employment as of the 12th of the month) than either hires or separations (which are the reported number of hires or separations over the course of the reporting month), it is permissible for JOLTS respondents to report JOLTS data that is unbalanced. JOLTS data is unbalanced when there is a change of employment that is not accounted for by the net of hires and separations. It is possible that a JOLTS reporter may under or over report hires and separations and not account for employment change. The systematic under/over reporting of either hires and separations may lead to a divergence between employment trends and hires minus separations trends.

In addition to lacking extremely young establishments in the JOLTS sample, it is possible that firms exiting the labor force may not participate and report data even when sampled. This would bias the separations rate downward.

The points above lead to a number of important questions regarding the JOLTS survey: To what extent does the JOLTS sample reflect the general composition of the population of business establishments it attempts to estimate with respect to establishment age? Does the hires and separations rate of establishments vary with age? How so? Does JOLTS fully capture the hires and separations from firms entering and exiting the job market? Do JOLTS reporters under or over report hires and/or separations? If it were possible to plausibly simulate JOLTS hires and separations rates for all establishments on the population frame, treating all firms identically irrespective of age or responsiveness, it would be possible to address the above questions.

The primary motivation for simulating JOLTS data was an attempt to better understand and perhaps help correct the long-standing divergence between Current Employment Statistics (CES) employment trends and the estimated hires minus separations trends published by JOLTS. It is expected that CES employment trends should over time be tracked closely by JOLTS hires minus separations estimates. However, it was found that, even over a substantial period of time, the employment trend in CES (to which JOLTS data is benchmarked) differed substantially in both direction and magnitude from JOLTS hires minus separations and, more worryingly, that the disparity was increasing over time. This issue was examined in detail by R. Jason Faberman in a 2005 paper entitled ―*Analyzing the JOLTS Hires and Separations Data*" in which Faberman found that *"The study finds that both differences [divergence between CES and JOLTS and the lower magnitude of JOLTS hires/separations estimates] stem from several sources. These include a potential underreporting of hires and separations in JOLTS by respondents (particularly those with large contractions), an over-representation in JOLTS of establishments with relatively stable employment, and an inability to adequately capture entry and exit (which is a handicap for any establishment sample)."*

The simulation that is described in this paper was an attempt to determine whether and to what extent the contentions detailed in the Faberman paper could be precisely identified or enumerated. The simulation attempted to determine whether the JOLTS sample was truly representative of the population of businesses it was attempting to estimate and whether there were any systematic differences between JOLTS respondents and non-respondents. In addition, the simulation hoped to demonstrate conclusively that JOLTS respondents report employment, hires and separations data in a way which systematically violates the expectation that change in CES employment may be tracked by JOLTS hires minus separations.

This paper will describe and explain the simulation methodology, briefly discuss the results of the simulation, and highlight several remedial steps taken as a result of the simulation.

2. SIMULATION

The Longitudinal Data Base, as mentioned above, serves as the frame for the JOLTS sample and is a census of all in-scope establishments in the country. The LDB contains the monthly employment for each establishment in the population over each establishment's lifespan—both those establishments sampled by or reported to JOLTS as well as all those could have been sampled. The comprehensive employment data contained on the LDB is therefore a useful input for a simulation since there is a logical relationship between employment change and the level of hires and separations. It can be supposed, for any given firm, that hires minus separations should *approximate* the net change in employment for that firm. This supposition can be extended further:

1. Establishments that experience no change in employment should, on average, have reported hires equal to reported separations rates.
2. Establishments that are expanding in employment should, on average, have reported hires greater than reported separations.
3. Establishments that are contracting in employment should, on average, have reported separations greater than reported hires.
4. New units (births) should have reported hires equal to at least to first reported employment.
5. Units falling off the frame (deaths) should have reported separations equal to at least to the last reported employment.

Let M_1 be the employment on the LDB for an establishment for a given month and let M_2 be the employment on the LDB for the same establishment in the subsequent month. M_1 and M_2 are known for all establishments on the LDB (the LDB contains employment and wage data for every establishment). A naïve simulation for hires and separations could therefore be made using only net employment change using LDB data. This naïve simulation would produce an estimate of hires equal to the net employment increase

when there was an increase in employment and an estimate of separations equal to the net employment decrease when there was a decrease in employment.

This naïve simulation can extend of the above suppositions and can be expressed in precise mathematical terms:

1. If $M_1 = M_2$, then $H_2=S_2=0$ where H_2 are the hires for the establishment in month 2 and S_2 are the separations for the establishment in month 2. There is not net change in employment and therefore there are neither simulated hires nor separations.
2. If $M_1 < M_2$, then $H_2= M_2 - M_1$ and $S_2 =0$. There is a net increase in employment and the net increase is equal to the number of simulated hires.
3. If $M_1 > M_2$, then $H_2= 0$ and $S_2 = M_1 – M_2$. There is a net decrease in employment and the net decrease is equal to the number of simulated separations.
4. If $M_1 = .$, $M_2 \neq .$, that is, the establishment is a birth unit, then $H_2= M_2$ and $S_2 =0$.
5. If $M_2 = .$, $M_1 \neq .$, that is, the establishment is a death unit, then $H_2=0$ and $S_2 = M_1$

However, this naïve estimate would underestimate JOLTS hires and separations substantially since there are likely to be an amount of offsetting hires and separations that do not contribute to net employment change.

For example, an establishment may have increased in employment from 20 employees to 25 employees. While there may indeed have been 5 hires to account for the net employment change, there may have been additional hires that were offset by separations—for example, there may have actually been 7 hires and 2 separations. In this example, in addition to the 5 hires that accounted for the employment change, there were 2 hires and 2 separations that offset each other. We will define this offsetting level of hires and separations as the *offsetting level of churn*.

A simulation of JOLTS hires and separations would have to account for the hires and separations that result from a net change in employment as well as the offsetting level of churn (which will be defined in this paper as Ω_h for hires and Ω_s for separations).

Taking into account offsetting churn we can express the above suppositions as follows :

1. If $M_1 = M_2$, then $H_2=S_2=\Omega_h = \Omega_s$, where H_2 are the hires for the establishment in month 2 and S_2 are the separations for the establishment in month 2 and Ω_h, Ω_s are the offsetting level of churning. There is no net increase in employment and this is reflected in the fact that simulated hires and separations offset.
2. If $M_2 > M_1$, then $H_2= M_2 - M_1 + \Omega_h$ and $S_2 = \Omega_s$, where Ω_h, Ω_s are the offsetting level of churning additional to the expansion in employment.
3. If $M_1 < M_2$, then $H_2= \Omega_h$ and $S_2 = M_1 – M_2 + \Omega_s$, where Ω_h, Ω_s are the offsetting level of churning additional to the contraction in employment.
4. If $M_1 = .$, $M_2 \neq .$, that is, the establishment is a birth unit, then $H_2= M_2$ and $S_2 = \Omega_s$.

5. If M_2 = ., $M_1 \neq$., that is, the establishment is a death unit, then $H_2 = \Omega_h$ and $S_2 = M_1$

Since the LDB is a census of business establishments in the US, the value of M_1 and M_2 are known for every establishment on the LDB. All that would be necessary to adequately simulate JOLTS hires and separations data would be the LDB employment data (M_1 , M_2)and the appropriate values of Ω_h and Ω_s . However, the values of Ω_h and Ω_s are not known for any establishment or any class of establishments. This paper will demonstrate how an estimate of the offsetting churns (Ω_h , Ω_s) can be derived using historical JOLTS data. Thus the LDB can be used to simulate JOLTS hires and separations data using the employment data contained on the LDB and an estimate of the offsetting level of churn.

3. ESTIMATING OFFSETTING LEVEL OF CHURN

As stated in the previous section, the LDB is a census of employment data (each record has a value for M_1 , M_2) and this employment data may be used to derive a naïve simulation of hires and separations. What this naïve simulation of hires and separations lacks to sufficiently approximate actual reported JOLTS data is an appropriate level of offsetting churn.

Like the LDB, historical JOLTS data also contains reported employment. This historical employment data for JOLTS can likewise be used to derive a naïve simulation—that is, a simulation of hires and separations that is based solely on net employment change and completely lacks offsetting churn. JOLTS reported hires and separations data, however, accounts for net employment change as well as the additional level of offsetting churn. Therefore, the difference between JOLTS reported data (which accounts for net employment change and offsetting churn) and the naïve simulation of JOLTS hires and separations data based on reported JOLTS employment (which accounts only for net employment change) would provide an estimate of offsetting churn.

An historical dataset of JOLTS reported data was constructed with the following restrictions: only those reporters which reported two consecutive months of employment (since the naïve simulation is driven by employment change) and well as both hires and separations data (to eliminate partial reporters) were included. This dataset was stratified by JOLTS industry-size class. Using this dataset, a naive simulation of hires and separations data was made such that:

1. For stable units (no employment change), the hires and separations were set to zero.
2. For expanding units, the hires were set equal to the increase in employment and the separations were set to zero.
3. For contracting units, the separations were set equal to the absolute decrease in employment and the hires were set to zero.

After the naïve simulation, for each establishment in the restricted historical dataset, there were now two sets of hires and separations estimates: the actual reported levels of hires and separations and the naïve simulation of hires and separations. The difference between the reported levels of hires and separations (which account for both net employment and offsetting churn) and the naïve simulation of hires and separations (which account solely for net employment change) would be the estimate of offsetting churn. Hence, to estimate the levels of offsetting churn (Ω_h, Ω_s) for any industry-size class (id-size) cell the following formulas were used:

$$(1)\quad \hat{\Omega}_h = \frac{\sum_{id-size} Hires_{reported} - \sum_{id-size} Hires_{naive}}{\sum_{id-size} Employment_{reported}}$$

$$(2)\quad \hat{\Omega}_s = \frac{\sum_{id-size} Separations_{reported} - \sum_{id-size} Separations_{naive}}{\sum_{id-size} Employment_{reported}}$$

The following table yields the results of the estimation of the level of offsetting churn for each JOLTS industry division:

JOLTS INDUSTRY DIVISION	Reported Employment	Reported Hires	Reported Separations	Naïve Hires	Naïve Separations	$\hat{\Omega}_h$	$\hat{\Omega}_s$	TOTAL
Natural Resources & Mining	3,190,611	68,548	68,036	44,464	41,636	0.75%	0.83%	1.58%
Construction	2,825,571	176,082	159,630	83,082	98,640	3.29%	2.16%	5.45%
Non-durable Manufacturing	9,388,851	210,340	237,151	96,089	126,512	1.22%	1.18%	2.40%
Durable Manufacturing	32,178,673	381,977	474,113	211,467	293,084	0.53%	0.56%	1.09%
Wholesale Trade	5,061,639	95,709	107,437	47,461	65,575	0.95%	0.83%	1.78%
Retail Trade	6,410,949	348,984	343,716	166,274	145,326	2.85%	3.09%	5.94%
Transport-Warehouse-Utility	22,590,523	399,003	386,679	175,298	189,780	0.99%	0.87%	1.86%
Information	4,984,113	100,520	111,134	40,276	48,978	1.21%	1.25%	2.46%
Finance & Insurance	13,321,405	229,736	244,456	94,422	90,321	1.02%	1.16%	2.17%
Real Estate- Rental & Leasing	1,099,972	48,422	49,467	25,406	26,648	2.09%	2.07%	4.17%
Professional & Business Services	21,876,210	537,554	545,379	268,918	291,850	1.23%	1.16%	2.39%
Temp Help	1,453,236	154,689	141,737	25,064	29,221	8.92%	7.74%	16.66%
Educational Services	14,435,257	250,601	198,000	242,303	225,400	0.06%	-0.19%	-0.13%
Health Care & Social Assistance	59,729,373	1,082,626	816,985	391,001	283,061	1.16%	0.89%	2.05%
Arts & Entertainment	4,702,542	361,934	338,517	234,835	240,774	2.70%	2.08%	4.78%
Accomodation & Food Services	8,634,535	335,063	319,773	153,996	168,092	2.10%	1.76%	3.85%
Other Services	2,389,212	111,700	88,198	55,231	55,473	2.36%	1.37%	3.73%
Federal Government	78,895,422	1,105,155	909,028	469,642	308,288	0.81%	0.76%	1.57%
State & Local Education	129,695,838	2,302,817	1,798,179	2,051,116	1,918,315	0.19%	-0.09%	0.10%
State & Local Non-Education	98,825,721	1,302,993	1,071,656	647,963	576,633	0.66%	0.50%	1.16%
TOTAL	521,689,653	9,604,453	8,409,271	5,524,308	5,223,607	0.78%	0.61%	1.39%

There are a number of substantive findings in the table above. First, the level hires and separations that result from offsetting churn are quite large (approximately 40% of hires and separations result from offsetting churn). Second, the level of hires and separations that result from offsetting churn varies considerably across industry—in high turnover industries such as Retail Trade, Construction, Arts & Entertainment, and Accommodation

& Food Services the level of offsetting churn is very high, while in contrast offsetting churn in Education (Private and Government) is virtually non-existent. The Temporary Help industry is particularly unusual in that very little of the reported hires and separations in that industry result from a change in reported employment.

Perhaps the most important finding is that the initial assumption, namely, that hires minus separations should approximate the net change in employment does not hold even over time in many industries. The extent to which the estimates of Ω_h and Ω_s differ is an indication of the extent to which the initial assumption does not hold.

Recall that the naïve simulation of hires and separations in the table above accounts only for the reported changes in the employment. One would expect that the churn in excess of the net employment change should offset—as is the case in the majority of industries. However, there are a number of industries in which the churn in excess of the net employment change does not offset. This is not problematic for any given establishment for a given month. There are indeed times when the respondent may reasonably report unbalanced data. However, this imbalance is not expected to remain intact over many respondents and over a substantial length of time—over time the churn in excess of net employment change should become balanced. A systemic imbalance over a substantial period time for an industry may indicate that JOLTS reporters have interpreted the definition of employment, hires, or separations such that the initial assumption of balance does not hold. For example, a reporter may report an increase of employment and failed to ever report the hires that accounted for the employment increase. Conversely, a reporter may report a decrease of employment and failed to ever report the separations that accounted for the employment increase.

In the Construction, Temp Help, Arts & Entertainment, and Other Services industries the levels of churn in excess of the net employment change are not offsetting but rather are systematically biased in the direction of hires. Such a disparity could occur when reporters in those industries under-report separations. This is hypothetically possible that since there may be a time lag between a change in employment (i.e., employee dropped from payroll) and a subsequent separation, and it is possible that the separation, when it later occurs, may not ever be reported to JOLTS. Temp Help is a particularly problematic industry because determining whether a temporary worker is paid (i.e., employed) is straight forward while determining whether the same worker is no longer available to the temp agency (i.e., separated) may not be.

The results from the table above confirm the contentions of Faberman that there is substantial under-reporting (primarily of separations) by JOLTS reporters and that the reported imbalance of hires and separations by JOLTS reporters is a significant contributor to the divergence between JOLTS and CES trends.

Subsequent to this research, the JOLTS program has implemented an alignment procedure that helps to greatly mitigate the divergence between JOLTS and CES trends

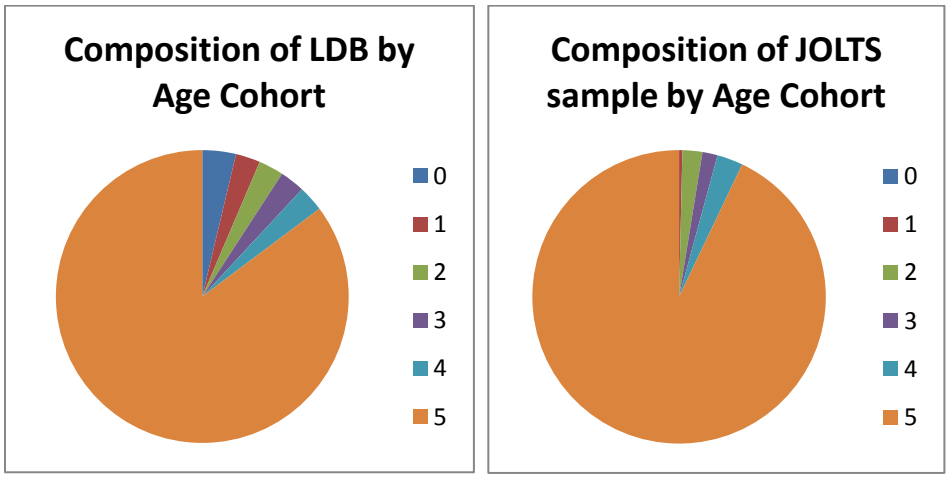
(see –The CES/JOLTS Divergence: How to Apply the Monthly Alignment Method to Help Close the Gap" by Cheng, Hudson *et al* for further details).

4. SIMULATION RESULTS

All establishments on the LDB back to the inception of the JOLTS series (Dec 2000) were simulated using change in LDB employment and the industry-size class estimates of offsetting churn. This produced simulated hires and separations for every establishment on the LDB. Each establishment was classified by age based on the initial month that the establishment reported positive employment to the QCEW as follows:

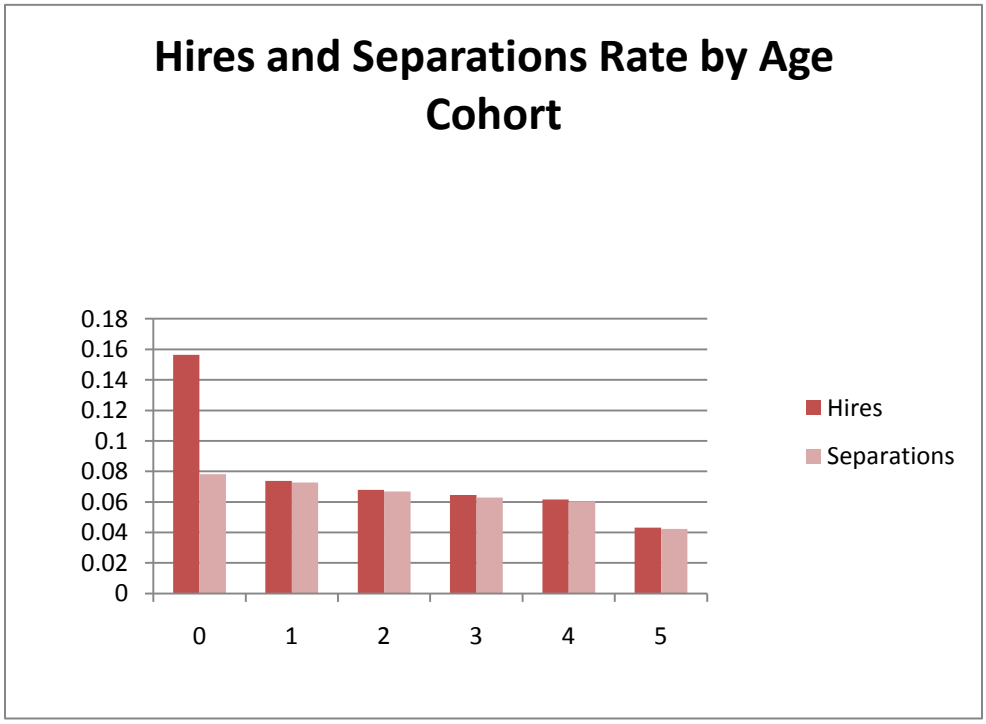
- Those establishments whose initial month of positive employment was less than 12 months prior to the simulation were classified as Age 0
- Those establishments whose initial month of positive employment was more than 12 months prior but less than 24 month prior to the simulation were classified as Age 1
- Those establishments whose initial month of positive employment was more than 24 months prior but less than 36 month prior to the simulation were classified as Age 2
- Those establishments whose initial month of positive employment was more than 36 months prior but less than 48 month prior to the simulation were classified as Age 3
- Those establishments whose initial month of positive employment was more than 48 months prior but less than 60 month prior to the simulation were classified as Age 4
- And those establishments whose initial month of positive employment was more than 60 months prior to the simulation were classified as Age 5

It was found that the composition of the LDB (that is, the population of establishments that JOLTS samples) has an age distribution that differs considerably from the JOLTS sample. The graphs below illustrate the distribution by each of all establishments in the population over the period of analysis and the distribution by each sampled establishment in JOLTS over the same time period:



The LDB population contained far more establishments (14.77%) than the JOLTS sample (7.04%) in the younger age classes (Age 0 to Age 4). This disparity is even more dramatic in the youngest age classes (Age 0 and Age 1, that is, those establishments that had been on the frame for less than 24 months)—the LDB was comprised by 6.42% of such establishments while the JOLTS sample was comprised of a scant 0.36% of such establishments (with the JOLTS sample having none in Age 0).

This disparity in composition by age is significant because the simulated hires and separations rates varied considerably across age cohorts. The chart below details the hires and separations rate by age cohort:



The graph above confirms two contentions stated in the Faberman paper: first, the JOLTS sample is over-represented by establishments in the oldest age class and that oldest age class is the most stable (in terms of levels of hires and separations); second, the JOLTS sample completely lacks those firms that have recently entered the labor force and it is this cohort whose hires and separations rates are far greater than their older counterparts.

5. BIRTH MODEL

To rectify the lack of reported data from establishments entering (and also exiting—since it was found that indeed exiting firms sampled by JOLTS do not tend to respond to JOLTS) the labor force, the JOLTS program implemented a birth/death model. As mentioned in the introduction, the firms that just enter the labor force cannot be sampled by JOLTS. It takes a number of months to develop the frame upon which the JOLTS sample is drawn. It takes a number of months to contact and enroll the establishments that are sampled. Finally, it takes time to rotate sample into JOLTS estimation. In the time it takes to develop a frame, select a sample, contact and enroll sampled units, and rotate those units into estimation, those units have aged and do not have the same dynamic qualities they possessed earlier. The birth/death model adopted by JOLTS uses the simulated hires and separations data from the youngest size classes (Ages 0 and Ages 1) to account for the churn (hires and separations) of those young establishments which the

JOLTS program simply cannot sample. The birth-death model is reflected in current JOLTS estimation.

However, since the simulation itself is based on the JOLTS frame (the LDB), it too has a lag period (the frame must be developed and the simulation run). To cope with this lag between JOLTS estimation and the availability of LDB upon which to simulate JOLTS data, the birth-model has been forecasted using CES year on year employment trends. The JOLTS program is currently exploring other and perhaps better forecasting techniques.

6. OTHER CORRECTIVE ACTIONS

While the birth-death model allows the JOLTS program to better account for those establishments entering and exiting the labor force, the JOLTS sample was still over-represented by older more stable firms. To mitigate this fact, the JOLTS program has updated sampling procedures ensuring that younger establishments are introduced into the JOLTS sample on a timelier basis. Subsequent to the implementation of the JOLTS birth-death model, JOLTS has introduced a quarterly update of younger establishments (those age 1). In addition, JOLTS has post-stratified sampling weights with respect to the age of establishment to ensure that the JOLTS sample reflects the population with respect to age.

7. CONCLUSION

The 2005 paper by Faberman contended that there were three sources for the divergence between CES employment and JOLTS hires minus separations trends: 1) underreporting of JOLTS hires and separations by JOLTS respondents 2) the overrepresentation of stable firms in the JOLTS sample and 3) the inability of the JOLTS program to sample and collect data from establishments entering and exiting the labor market. A simulation of JOLTS data using the QCEW LDB was conducted to determine whether these contentions were correct. The simulations produced based on QCEW data did in fact verify each of Faberman's contentions and the simulation helped point to several remedial measures to mitigate the impact of those contentions. To mitigate the fact that JOLTS respondents underreport JOLTS hires and separations, the JOLTS program has introduced an alignment procedure. To mitigate the fact that the JOLTS sample was over representative of older more stable firms, the JOLTS program has introduced a quarterly birth-sample and post-stratified estimation such that the JOLTS sample reflects the age structure of the population. Finally, to mitigate the fact that JOLTS has been unable to account for firms entering and exiting the labor force, the JOLTS program has introduced a birth-death model.

REFERENCES

Cheng, Edmund, Hudson, Nicole, Kropf, Jurgen, and Mercurio, Jeannine 2009. ―The CES/JOLTS Divergence: How to Apply the Monthly Alignment Method to Help Close the Gap‖. *2009 Proceedings of the Annual Statistical Association.* Alexandria, VA: American Statistical Association.

Faberman, R. Jason, 2005. ―Analyzing the JOLTS Hires and Separations Data.*" 2005 Proceedings of the Annual Statistical Association.* Alexandria, VA: American Statistical Association.