MPRA
Munich Personal RePEc Archive

# US labor market conditions and migration: a reassessment of Bahar (2025)

RODRIGUEZ, FRANCISCO and Bravo, Giancarlo

Center for Economic and Policy Research, University of Denver, Oil for Venezuela

24 December 2025

Online at https://mpra.ub.uni-muenchen.de/129209/
MPRA Paper No. 129209, posted 05 Jun 2026 15:16 UTC

# US labor market conditions and migration: a reassesment of Bahar (2025)

Francisco Rodríguez*and Giancarlo Bravo†

December 19, 2025

**Abstract**

Bahar (2025) argues that there is a long-term cointegrating relationship between US job vacancies and Southwest border crossings. We show that his conclusion is based on a misspecified Engle–Granger test applied to first differences. Once the Engle–Granger test is correctly applied to levels, evidence for a cointegrating relationship vanishes, invalidating the paper's approach to estimating short- and long-run elasticities. Bahar's approach is therefore uninformative about the relationship between US labor market conditions and migration.

## 1 Introduction

Using monthly data for the United States, Bahar (2025a, 2025b) argues that there is a strong positive correlation between Southwest U.S. border crossings and labor market tightness, leading to the conclusion that "pull" factors play a significant role in driving migration patterns towards the U.S. (Bahar 2025a, 679). We show that this conclusion is premised on findings from a misspecified cointegration test. Correcting this mistake makes the finding of a long-term relationship disappear and undermines the justification for the methods used in the paper to estimate both short and long-term elasticities.[1]

Section 2 explains why Bahar's choice to apply a cointegration test to first differences of I(1) variables is incorrect. Section 3 shows how applying a correct cointegration test in levels fails to yield the conclusive evidence of cointegration claimed by Bahar. Section 4 documents additional new coding errors and misrepresentations which we have found in Bahar's paper. Section 5 discusses how our results invalidate Bahar's econometric approach and findings.

## 2 Methods

Bahar's paper studies the relationship between flows of irregular migrants into the United States and U.S. labor market conditions. His empirical analysis focuses on two indicators: encounters

*Center for Economic and Policy Research and University of Denver

†Oil for Venezuela

1. Note that Bahar (2025b) is a corrected version of Bahar (2025a), and that as of the time of this writing there are three different versions of replication code published for the paper. Where necessary, we will distinguish between the original (or uncorrected) and corrected versions of the paper, and will refer to the various replication codes in chronological order of publication.

at the Southwest United States border and job openings per unemployed person in the United States. Bahar claims that both of these series are I(1), and that "when the Engle-Granger cointegration test was applied to these two series, this produced a test statistic of -18.829, well above [sic] the critical value of −4.024 at the 1% significance level."(Bahar 2025a, 676–677).[2] Bahar's Appendix F makes a similar claim regarding the quarterly series for these variables. Neither of these claims was affected by the June 2025 correction of Bahar's paper.

Bahar's cointegration findings come from applying Engle-Granger cointegration tests to the first differences of the series. We conclude that he applied the test to the first differences of the series in question based on several pieces of information. First, although neither his corrected replication code (version 3.0) nor the replication code published alongside the original paper (version 2.0) implement Engle-Granger tests, a version of the code that was briefly available prior to publication (version 1.0) does contain a misspecified Engle-Granger test in first differences.[3] Second, Bahar and Hausmann (2025b, p.3), who claim to use the same methodology as Bahar (2025a), explicitly state that they apply the Engle–Granger test to the first differences of the variables [p.3], as can also be verified in the replication code for that paper.

Third, we are able to either nearly replicate or exactly replicate Bahar's cointegration results using a first-difference test, while applying the standard test in levels yields very different values. More concretely, applying the test in first differences from version 1.0 of the replication code to Bahar's full sample yields a test statistic of -19.208, while applying it to a shorter time window (which Bahar's uncorrected code used for some results) gives a test statistic of -18.834, very close to the -18.829 value that he reports. Applying the test in first differences to the shortened time window on the quarterly data exactly reproduces Bahar's test statistic of -8.656 reported in Appendix F.[4]

The Engle-Granger cointegration test is designed to be run on the levels of variables that are I(1). If the series are I(1) – and Bahar claims that they are [5] – then their first differences must be stationary, and the Engle-Granger test applied to them will trivially find evidence of cointegration. This is because linear combinations of stationary variables are, under general conditions, stationary, and the Engle-Granger test is a test for stationarity in a linear combination of the variables it is applied on (the OLS residuals from the first stage regression). Rodríguez and Bravo (2025) present Monte Carlo simulations showing that applying the Engle-Granger test to the first differences of two non-cointegrated unit root series results in false positive rates of 100 percent.

Applying the Engle-Granger test to first differences of I(1) series is therefore incorrect. The original presentation of the test by its creators, standard econometrics textbooks and statistical software manuals specify that the method is designed to be applied on the levels of the corresponding variables (see, e.g., Engle and Granger (1987, p. 268), Greene (2012, p.

2. Bahar provides no results from unit root tests—either in the paper, online appendix, or replication code—to support his claim that the series are I(1). Our application of standard unit root tests to his data gives conflicting results regarding whether these series can be characterized as I(1).

3. The code in that version reads "egranger d.openingsrate_nsa d.encounters_ nsa". The correct code to run a conventional Engle-Granger test in levels would be "egranger openingsrate_nsa encounters_ nsa".

4. Note that Bahar misreports the 1% critical values as -4.024 in both cases (the correct values are -3.936 and -4.020, respectively).

5. See Bahar (2025a, p. 676): "Both series are integrated of order 1."

1001), Schaffer (2010)). An exhaustive literature search has found no literature supporting the use of the test on first-differenced data nor any implementations by paper others than Bahar (2025a) and Bahar and Hausmann (2025a).

Table 1: Engle-Granger Test Statistics Across Specifications

| | **Logarithms** | | **Untransformed** | |
|---|---|---|---|---|
| | Encounters | Openings | Encounters | Openings |
| Engle-Granger | -3.153* | -2.183 | -4.061*** | -3.156* |
| Augmented Engle-Granger, 12 lags | -2.614 | -1.922 | -2.182 | -1.851 |
| Augmented Engle-Granger, 12 lags and trend | -2.968 | -3.052 | -2.860 | -3.790* |
| Bahar's reported Engle-Granger test statistic | | -18.829*** | | |

Results represent the Engle-Granger test statistic for various specifications. Columns 1 and 3 normalize the coefficient of the cointegrating vector on encounters to one, while Columns 2 and 4 normalize the coefficient on openings to one. The last row reports the result presented by Bahar (2025a, 677). Significance levels: *** 1 percent, ** 5 percent, * 10 percent.

# 3 Results

Table 1 reports the results of twelve specifications of Engle-Granger cointegration tests applied to Bahar's data and representing combinations of choice of variable (logarithms or untransformed), lag and trend controls, and normalization assumption. We report this range of specifications because the text of Bahar's paper is itself unclear regarding whether the test was run on the untransformed values of the variables or their logarithms, and which variable was normalized as the dependent variable.[6] It is also customary to assess the sensitivity of the Engle-Granger test to the inclusion of lags and a trend term. The set of specifications reported in Table 1 thus gives a reasonable sense of the range of results that would emerge from running conventional Engle-Granger tests on Bahar's data.

The range of test statistics obtained in these estimates spans from -1.851 to -4.061, with a median value of -2.914 — less than one-sixth the magnitude of the -18.829 statistic reported by Bahar. Of the twelve specifications, only one rejects the null of no cointegration at a significance level of 5% or lower, in sharp contrast with Bahar's claim of having comfortably rejected the null of no cointegration at a 1% significance level.

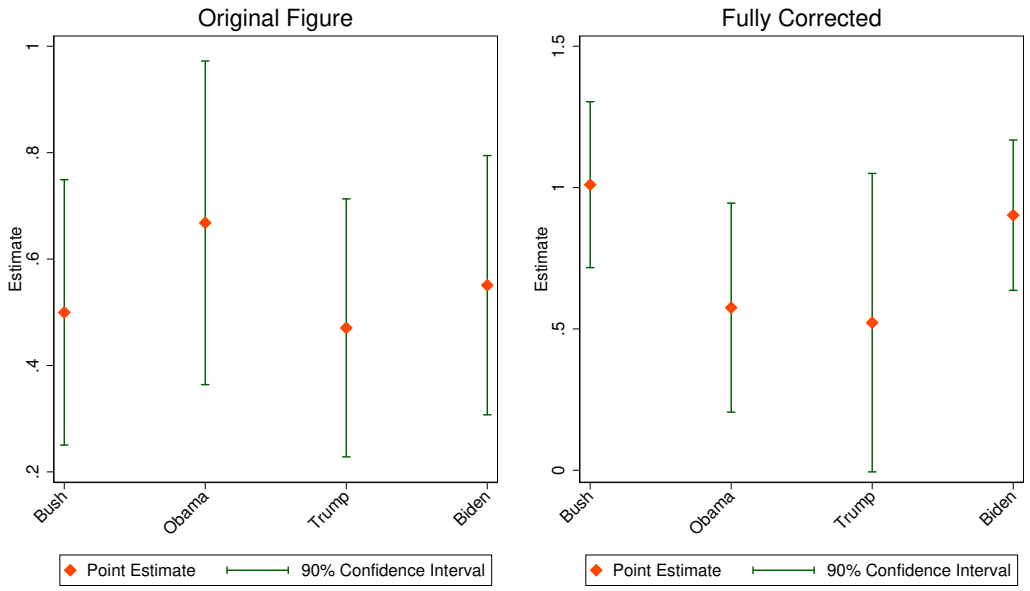
# 4 Additional issues

## 4.1 Miscoding of equation specifications

There are two additional instances in which the results reported by Bahar in his corrected article are different from those that emerge from the specifications that he claims to estimate. In both cases, the discrepancies stem from a miscoding that leads to the inclusion of different sets of controls than those reported in the text.

6. Since the author uses an OLS estimate of the cointegrating vector in logarithms for his ARDL specification, it would be natural to conclude that the test had been carried out in logarithms; however, the miscoded version of the code in version 1.0 applies the test to the untransformed variables.

Figure 1: Bahar's Figure 3(a), original and corrected




*Note*: This figure compares the original Figure 3(a), which incorrectly included the year-on-year controls from the ARDL specification in the levels equation, with a corrected version that includes only the controls that correspond to the levels regression.

According to Bahar, Figure 3(a) (which was unaffected by his June 2025 correction) "plots estimated elasticities based on levels, from specification (B1)" (Bahar 2025a, p.679). However, Bahar incorrectly included in his estimation of this equation a set of controls – the lagged residuals from the error correction term and the lagged 12-th differences in log encounters and openings – that do not belong in specification (B1) but rather come from specification (B2), the ARDL model in seasonal differences. As shown in Figure 1, this makes a substantial difference to his estimates. For example, the true estimated elasticity during the Bush administration is 1.01, not the 0.50 displayed in Figure 3(a) of the paper.

The results shown in Bahar's Table E1, Panel B, are also incorrect for a similar reason. Bahar's paper states (in both the original and uncorrected version) that these specifications include "lagged values (by one quarter) of the year-over-year crossings and openings (in logs) as well as the Error Correction Term explained in the main body of the text." (Bahar 2025b, p.4; 2024, p. OA-10) However, we have verified that Bahar included none of these controls but instead mistakenly included the level of total crossings for the corresponding crossing point. In Table 2 we show Bahar's original results as well as the results that obtain from correcting his code.

## 4.2 Reproducibility and other issues

The replication code provided by Bahar to accompany the correction is incomplete. In particular, it invokes two .do files—AnalysisQuarterlyData.do and AnalysisByNationality.do—which are not included in the replication package. Without these files, it is impossible to reproduce Appendices D and F.

Table 2: Original and Corrected Coefficients in Table E1.B

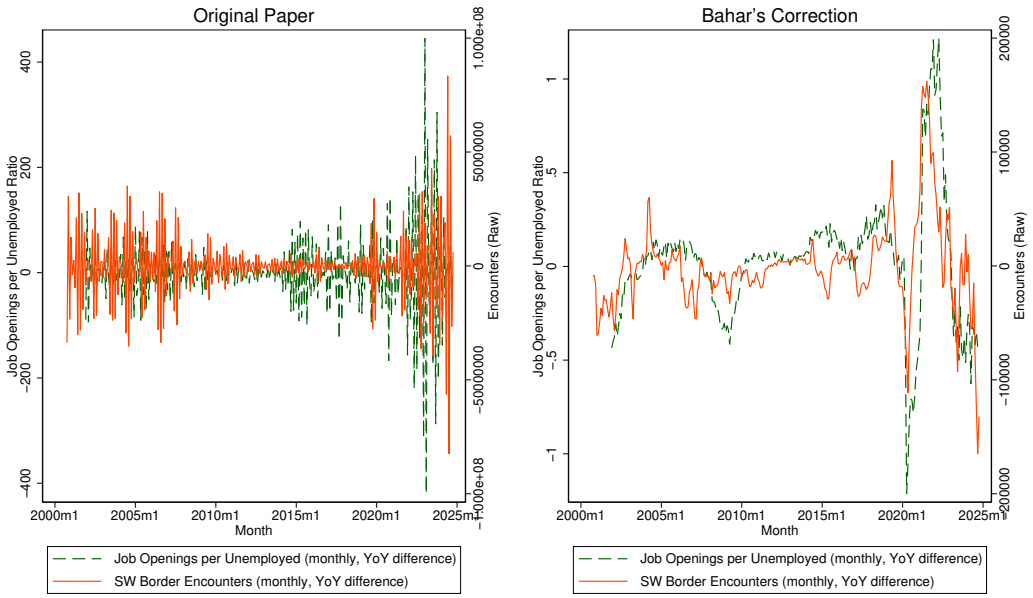
| | Big Bend | Del Rio | El Centro | El Paso | Laredo | Rio Grande | San Diego | Tucson | Yuma |
|---|---|---|---|---|---|---|---|---|---|
| Bahar's correction | 0.047 | 0.587 | 0.360 | 0.799 | 0.075 | 0.451 | 0.288 | 0.290 | 1.158 |
| | (0.07) | (0.09)*** | (0.07)*** | (0.09)*** | (0.06) | (0.12)*** | (0.07)*** | (0.09)*** | (0.16)*** |
| Fully Corrected | 0.449 | 0.203 | 0.486 | 0.721 | 0.438 | 0.579 | 0.304 | 0.315 | 1.270 |
| | (0.20)** | (0.14) | (0.14)*** | (0.18)*** | (0.14)*** | (0.10)*** | (0.12)** | (0.16)** | (0.20)*** |

Notes: Row 1 shows the results reported in Panel B of Table E1 in Bahar (2025b, p. 4). Row 2 corrects Bahar's specification by including the full set of controls corresponding to his equation (B2), and by dropping a superfluous control introduced by a coding error. Significance levels: *** 1 percent, ** 5 percent, * 10 percent.

Bahar's correction states that some variables measuring year-over-year differences were incorrectly computed in the original article "due to a coding error in the statistical software." (2025, p.1) This is incorrect. The statistical software used for Bahar's analysis, Stata, correctly calculates year-over-year differences using the `S12` command. The error in Bahar's paper stems not from a software error, but from his mistaken use of Stata's `D12` command instead, which calculates 12-th differences, instead of `S12`.

Bahar also characterizes this miscoding of year-over-year differences as a "minor coding error". This characterization is also incorrect. As documented in Rodríguez, Rosnick, and Bravo (2025, p. 10), the mistake appears in 23 separate lines of Bahar's code. Figure 2 reproduces Bahar's own correction to his Figure 1(c), juxtaposing the original series (based on 12-th differences) with the corrected year-over-year differences. The original and corrected figures bear little resemblance to each other. The stark visual contrast between the two panels underscores the severity of the coding error.

Figure 2: Bahar's Figure 1(a), original paper and his correction




*Note*: This figure compares the original Figure 1(c), which incorrectly used the 12-th difference operator, with Bahar's own correction, which corrects the mistake in the calculation of year-over-year differences.

## 5 Conclusions

Correcting the misspecified tests identified in this note severely undermines Bahar's empirical strategy and sheds doubt on the validity of his findings. This is because the claim of cointegration between border encounters and job openings is central to the paper's empirical approach.

Bahar's empirical strategy is completely based on estimation of his equations (B1) and (B2). Equation (B1) is a levels regression between two I(1) series which, in the absence of cointegration, is well-known to be prone to spurious estimates (Hamilton 1994, p.557). Equation (B2) is an autoregressive distributed lag specification which includes an error-correction term (ECT) that is only defined if there is a cointegrating relationship between the variables. In the absence of such a relationship, the estimate of the ECT (i.e., the residual from the levels regression) will be I(1) and its inclusion in the regression will risk producing spurious results. Therefore, both the estimates of short- and long-run elasticities presented by Bahar are derived from misspecified regressions and cannot be relied upon.

Bahar's claim that there is a positive correlation between labor market tightness and the frequency of border crossings is an interesting hypothesis worthy of systematic investigation. Given that the methods applied in his paper to reach that conclusion are based on misspecified regressions premised on a cointegrating relationship for which there is no evidence, Bahar's article is uninformative on what the evidence tells us about this key question.

## References


Bahar, Dany. 2024. *Online Appendix for *Not a Border Crisis, but a Labor Market Crisis: The Often Overlooked "Pull" Factor of US Border Crossings**. Supplementary material. Published December 6, 2024. https://doi.org/10.1111/pam.22665/suppl_file/pam22665-sup-0001-onlineappendix.pdf.

———. 2025a. "Not a border crisis, but a labor market crisis: The often overlooked "pull" factor of U.S. border crossings." *Journal of Policy Analysis and Management* 44, no. 2 (April): 674–680. https://doi.org/10.1002/pam.22665.

———. 2025b. "Correction to "Not a border crisis, but a labor market crisis: The often overlooked "pull" factor of U.S. border crossings"." *Journal of Policy Analysis and Management* (June). https://doi.org/10.1002/pam.70015.

Bahar, Dany, and Ricardo Hausmann. 2025a. *Sanctions on Venezuela Are Not Driving Migration to the US Southwest Border: An Empirical Assessment.* CGD Note 382. Center for Global Development. Accessed April 9, 2025. https://www.cgdev.org/sites/default/files/sanctions-venezuela-not-driving-migration-southwest-border.pdf.

———. 2025b. "Stopping Venezuela's Exodus Hinges on Restoring Hope." *Americas Quarterly* (March). Accessed May 8, 2025. https://americasquarterly.org/article/stopping-venezuelas-exodus-hinges-on-restoring-hope/.

Engle, Robert F., and C. W. J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55 (2): 251–276. ISSN: 00129682, 14680262, accessed May 8, 2025. http://www.jstor.org/stable/1913236.

Greene, William H. 2012. *Econometric Analysis.* 7th ed. Upper Saddle River, NJ: Prentice Hall. ISBN: 9780131395381.

Hamilton, James D. 1994. *Time Series Analysis.* Princeton University Press. ISBN: 9780691042893. https://press.princeton.edu/books/hardcover/9780691042893/time-series-analysis.

Rodríguez, Francisco, and Giancarlo Bravo. 2025. *Why Bahar and Hausmann Tell Us Nothing About Venezuelan Migration Flows to the United States.* https://cepr.net/publications/why-bahar-and-hausmann-tell-us-nothing-about-venezuelan-migration-flows-to-the-united-states/. Center for Economic and Policy Research.

Rodríguez, Francisco, David Rosnick, and Giancarlo Bravo. 2025. *Did Sanctions Relief Drive Venezuelan Migration to the US?* https://cepr.net/publications/did-sanctions-relief-drive-venezuelan-migration-to-the-us/. Center for Economic and Policy Research.

Schaffer, Mark E. 2010. *EGRANGER: Stata module to perform Engle-Granger cointegration tests and 2-step ECM estimation.* https://ideas.repec.org/c/boc/bocode/s457210.html. Statistical Software Components S457210, Boston College Department of Economics, revised 03 Feb 2022.