Supplementary On-Line Appendix for
"Cointegrated Factor Augmenting Productivity"
Alwyn Young, London School of Economics, February 2025

A. Monte Carlos for Stationarity & Unit Root Tests

This appendix uses Monte Carlos to evaluate the accuracy of the panel stationarity and unit root tests available in Stata. Those that have rejection rates closer to nominal value are featured in Table 4 in the paper, which presents results for these tests on the KLEMS sample. The Stata panel tests are the Hadri (2000) Lagrange multiplier test for stationarity & the Breitung (2000), Harris & Tsavalis (1999), Im, Pesaran & Shin (2003), and Levin, Lin & Chu (2002) tests for a unit root, as well as Fisher-type tests for a unit root that combine the p-values from individual Dickey & Fuller (1979) or Phillips & Perron (1988) tests for each time series in the panel. The Fisher-type tests use four alternative summary statistics of the *N* p-values:

$$\text{(A1)}\quad P = -2\sum_{i=1}^{N}\ln(p_i);\;\; Z = \sum_{i=1}^{N}\frac{\Phi^{-1}(p_i)}{\sqrt{N}};\;\; L = \sum_{i=1}^{N}\ln\left(\frac{p_i}{1-p_i}\right);\;\; P_m = -\sum_{i=1}^{N}\frac{\ln(p_i)+1}{\sqrt{N}}.$$

I refer to these as Fisher-P, Fisher-Z, Fisher-L and Fisher-P_m. When the regression does not include lags, the Dickey-Fuller and Phillips-Perron p-values are identical. However, when lags are included, they differ, and are then distinguished by the initials DF and PP in the tables below.

The starting point for each simulation is a data generating process (*dgp*) for which the tested null is true, parameterized off of the KLEMS samples. When the test is for a unit root, I estimate the following equation for each KLEMS sample:

$$\text{(A2)}\quad g(y_{it}) = \beta_i + \beta_t + \gamma_i t + \rho g(y_{it-1}) + \varepsilon_{it},$$

where $g(y_{it}) = y_{it} - y_{it-1}$ (measured in lns), β_i and β_t are industry and year fixed effects, γ_i is the industry specific time trend, and ρ allows for first order autocorrelation. γ_i and ρ are only estimated and included in some *dgps*. The point estimates, plus iid shocks, are then used to generate new values for $g(y_{it})$, which are cumulated to create a y_{it} series with a unit root for the subsequent unit root tests. For normal shocks, I estimate (A2) using OLS and set the standard deviation of the iid Monte Carlo shocks equal to its root mean squared error. For t-distributed shocks, I estimate (A2) using maximum likelihood based upon the t-distribution and set the standard deviation and degrees of freedom of the iid Monte Carlo t-distributed shocks equal to their estimated values. For tests of stationarity, (A2) is estimated using y_{it} and y_{it-1}, rather than their first differences, on the left and right hand sides and estimated values and iid normal or t-distributed shocks are used to create the stationary series. Since Table 4 in the paper reports results for both the ln series and its growth rate (first difference), results below are differentiated by whether y_{it} in (A2) and the *dgp* is the ln series or its growth rate. In terms of simulation results, this matters most when autoregression is included in the *dgp*, as the estimated autoregressive parameters in (A2) vary greatly depending upon whether y_{it} is the ln series or its difference.

Table A1 below reports empirical rejection rates at the .05 level of the different tests in 500 Monte Carlo iterations for specifications that include industry and year fixed effects or the same plus industry specific time trends. Separate results are reported for the Hadri test with a heteroskedastic or homoskedastic covariance estimate and for the Im-Pesaran-Shin test using a finite sample (based upon Monte Carlos by the authors) or asymptotic distribution, as these are options provided by Stata. Table A2 below adds the lagged value of $g(y_{it})$ or y_{it} to the right-hand side of the baseline estimating equation (A2) and hence to the *dgp*. The Hadri, Harris-Tsavalis and Im-Pesaran-Shin with finite sample distribution tests of Table A2 do not contain a correction for autoregression as an option, as the distributions have not been worked out for this case. As these tests are featured in the paper, they are included to show how the accuracy of the tests deteriorates when the *dgp* is autoregressive. The remaining tests in the table contain a correction for first order autoregression. Table A3 summarizes the mean and standard deviation of the empirical rejection probabilities across 12 cells (6 series, with fixed effects and with fixed effects & time trends) differentiated by whether the shocks are normal or t-distributed and whether the *dgp* (and test when available) includes an AR1 process.

Focusing on Table A3 in particular, we see that without an AR1 in the *dgp*, the Hadri, HT and IPS (finite sample) tests provide rejection probabilities that are reasonably close to the .05 nominal value, but when the *dgp* includes an AR1 they are highly inaccurate. In that case, the IPS (asymptotic) and Fisher-Z and -L (Dickey-Fuller) tests provide the best performance. Based on these results, the presentation in the paper reports the results of those six tests on the KLEMS sample. Table A4

Table A1: Unit Root & Stationarity Tests:
Empirical Rejection Rates in 500 Iterations at .05 Level of True Nulls

	normal errors		t - errors		normal errors		t - errors		normal errors		t - errors	
	FE	trend	FE	trend	FE	trend	FE	trend	FE	trend	FE	trend
	Hadri (heteroskedastic cov) null = stationary				Hadri (homoskedastic cov) null = stationary				IPS (finite sample) null = has unit root			
ln(TFP)	.048	.036	.080	.064	.048	.046	.202	.166	.026	.018	.086	.048
ln(K/L)	.048	.048	.190	.042	.044	.056	.252	.142	.046	.014	.170	.028
ln(I/L)	.034	.028	.064	.066	.040	.034	.212	.136	.026	.008	.062	.040
g(TFP)	.036	.032	.044	.024	.052	.042	.114	.090	.030	.016	.032	.020
g(K/L)	.052	.048	.036	.022	.060	.054	.100	.088	.024	.034	.026	.022
g(I/L)	.044	.024	.036	.050	.044	.030	.140	.128	.028	.118	.038	.032
	IPS (asymptotic) null = has unit root				Harris-Tsavalis null = has unit root				Breitung null = has unit root			
ln(TFP)	.002	.996	.000	.864	.000	.986	.000	.988	.002	.002	.002	.008
ln(K/L)	.000	.932	.000	.846	.000	.954	.000	.908	.000	.004	.000	.012
ln(I/L)	.000	.994	.012	.844	.006	1.00	.010	.992	.006	.008	.018	.032
g(TFP)	.068	1.00	.026	.980	.044	1.00	.022	1.00	.084	.058	.064	.052
g(K/L)	.034	1.00	.036	.996	.044	1.00	.008	1.00	.098	.058	.074	.088
g(I/L)	.064	1.00	.030	.984	.026	1.00	.028	1.00	.070	.088	.048	.086
	Levin-Lin-Chu null = has unit root				Fisher-P DF & PP null = has unit root				Fisher-Z DF & PP null = has unit root			
ln(TFP)	.422	.862	.370	.786	.012	.024	.064	.122	.000	.000	.000	.004
ln(K/L)	.304	.764	.258	.764	.002	.020	.036	.140	.000	.002	.000	.004
ln(I/L)	.450	.870	.494	.790	.020	.046	.128	.176	.004	.006	.004	.006
g(TFP)	.632	.954	.574	.896	.080	.088	.206	.202	.052	.032	.030	.020
g(K/L)	.632	.934	.582	.930	.108	.106	.176	.268	.050	.038	.020	.040
g(I/L)	.612	.964	.560	.886	.096	.130	.190	.242	.044	.050	.022	.028
	Fisher-L DF & PP null = has unit root				Fisher-P_m DF & PP null = has unit root							
ln(TFP)	.000	.002	.014	.012	.012	.030	.074	.128				
ln(K/L)	.000	.004	.002	.022	.002	.022	.036	.148				
ln(I/L)	.006	.006	.028	.032	.026	.050	.136	.182				
g(TFP)	.064	.040	.082	.064	.086	.088	.210	.212				
g(K/L)	.050	.040	.084	.128	.112	.108	.182	.272				
g(I/L)	.046	.064	.084	.092	.102	.138	.200	.250				

Notes: Each cell represents the average rejection rate at the .05 level in 500 Monte Carlo iterations. FE = data generating process and estimating equations include industry & year fixed effects; trend = data generating process and estimating equations include industry specific time trends in addition to industry & year fixed effects. DF = Dickey Fuller, PP = Phillips-Perron, IPS = Im-Pesaran-Shin.

provides results for the KLEMS samples for all the tests considered in these Monte Carlos. With regards to the Hadri test of stationarity, the Hadri test with the homoskedastic covariance provides larger p-values for the growth rates of the series than the heteroskedastic covariance results reported in the paper, where I argued that the evidence suggested the growth series were stationary. Both tests agree in finding p-values of .000 for the stationarity of the ln series in all specifications. With regards

Table A2: Empirical Rejection Rates in 500 Iterations at .05 Level of True Nulls with Autoregressive (AR1) Data Generating Processes

	normal errors		t - errors		normal errors		t - errors		normal errors		t - errors	
	FE	trend	FE	trend	FE	trend	FE	trend	FE	trend	FE	trend
	Hadri (heteroskedastic cov) null = stationary				Hadri (homoskedastic cov) null = stationary				IPS (finite sample) null = has unit root			
ln(TFP)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.010	.060	.000	.004
ln(K/L)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.000	.000	.000	.000
ln(I/L)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.168	.490	.012	.000
g(TFP)	.000	.000	.018	.002	.002	.000	.086	.016	1.00	1.00	1.00	1.00
g(K/L)	.622	.384	.954	.794	.666	.426	.862	.722	1.00	1.00	1.00	1.00
g(I/L)	.000	.000	.030	.014	.000	.000	.086	.072	1.00	1.00	1.00	1.00
	IPS (asymptotic) null = has unit root				Harris-Tsavalis null = has unit root				Breitung null = has unit root			
ln(TFP)	.000	.000	.006	.008	.000	.004	.014	.034	.000	.000	.000	.002
ln(K/L)	.000	.000	.000	.000	.000	.000	.002	.000	.000	.000	.000	.000
ln(I/L)	.000	.000	.018	.026	.008	.280	.070	.042	.000	.000	.000	.000
g(TFP)	.036	.046	.036	.044	1.00	.000	1.00	.000	.058	.026	.000	.002
g(K/L)	.052	.026	.024	.034	1.00	.000	1.00	.000	.026	.020	.000	.002
g(I/L)	.052	.042	.028	.062	1.00	.000	1.00	.002	.036	.032	.000	.004
	Levin-Lin-Chu null = has unit root				Fisher-P DF null = has unit root				Fisher-Z DF null = has unit root			
ln(TFP)	.366	.574	.346	.542	.014	.020	.090	.152	.000	.000	.002	.006
ln(K/L)	.270	.614	.294	.544	.000	.030	.044	.108	.000	.002	.000	.000
ln(I/L)	.342	.522	.442	.532	.032	.028	.152	.176	.002	.000	.008	.010
g(TFP)	.220	.190	.312	.332	.080	.162	.182	.228	.042	.054	.028	.024
g(K/L)	.190	.076	.284	.234	.096	.120	.132	.184	.056	.030	.014	.028
g(I/L)	.172	.050	.228	.208	.100	.128	.168	.208	.054	.050	.014	.030
	Fisher-L DF null = has unit root				Fisher-P_m DF null = has unit root				Fisher-P PP null = has unit root			
ln(TFP)	.000	.002	.012	.028	.016	.020	.092	.152	.038	.214	.040	.056
ln(K/L)	.000	.002	.002	.006	.000	.032	.044	.112	.000	.000	.000	.000
ln(I/L)	.002	.000	.034	.034	.038	.032	.158	.180	.222	.656	.096	.034
g(TFP)	.050	.068	.060	.072	.086	.166	.192	.234	1.00	1.00	.998	1.00
g(K/L)	.064	.044	.064	.072	.114	.132	.140	.190	1.00	1.00	.992	1.00
g(I/L)	.064	.064	.060	.086	.112	.136	.168	.210	1.00	1.00	1.00	1.00
	Fisher-Z PP null = has unit root				Fisher-L PP null = has unit root				Fisher-P_m PP null = has unit root			
ln(TFP)	.004	.024	.000	.004	.004	.034	.002	.006	.044	.228	.046	.062
ln(K/L)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
ln(I/L)	.060	.324	.010	.002	.082	.342	.018	.002	.244	.674	.100	.036
g(TFP)	1.00	1.00	.986	1.00	1.00	1.00	.996	1.00	1.00	1.00	.998	1.00
g(K/L)	1.00	1.00	.972	1.00	1.00	1.00	.978	1.00	1.00	1.00	.992	1.00
g(I/L)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Notes: Each cell represents the average rejection rate at the .05 level in 500 Monte Carlo iterations. FE = data generating process and estimating equations include industry & year fixed effects; trend = data generating process and estimating equations include industry specific time trends in addition to industry & year fixed effects. DF = Dickey Fuller, PP = Phillips-Perron, IPS = Im-Pesaran-Shin.

Table A3: Mean and Standard Deviation of Empirical .05 Null Rejection Rates

	data generating processes without autoregression				data generating processes with 1st order autoregression			
	normal errors		t errors		normal errors		t errors	
	mean	sd	mean	sd	mean	sd	mean	sd
Hadri (heteroskedastic cov)	.040	.009	.060	.045	.584	.471	.651	.473
Hadri (homoskedastic cov)	.046	.009	.148	.052	.591	.470	.654	.443
Im-Pesaran-Shin (finite sample)	.040	.039	.040	.033	.561	.477	.501	.521
Im-Pesaran-Shin (asymptotic)	.507	.508	.482	.496	.021	.023	.024	.019
Harris-Tsavalis	.025	.045	.044	.049	.274	.445	.264	.445
Breitung	.014	.019	.002	.004	.016	.019	.001	.001
Levin-Lin-Chu	.700	.226	.658	.217	.299	.188	.358	.125
Fisher-P Dickey-Fuller	.061	.045	.163	.068	.068	.053	.152	.052
Fisher-Z Dickey-Fuller	.023	.023	.015	.014	.024	.025	.014	.011
Fisher-L Dickey-Fuller	.027	.026	.054	.040	.030	.031	.044	.028
Fisher-P_m Dickey-Fuller	.065	.046	.169	.069	.074	.057	.156	.053
Fisher-P Phillips-Perron	.061	.045	.163	.068	.594	.456	.518	.502
Fisher-Z Phillips-Perron	.023	.023	.015	.014	.534	.494	.498	.517
Fisher-L Phillips-Perron	.027	.026	.054	.040	.539	.490	.500	.518
Fisher-P_m Phillips-Perron	.065	.046	.169	.069	.599	.453	.520	.501

Note: Each mean and standard deviation calculated across the relevant 12 cells (six data series, with FE or FE plus trends) of Tables A1 and A2. When *dgp* is AR1, test includes AR1 control if the distribution with this control is available (not available for Hadri, HT & IPS finite sample).

to the unit root tests, all methods find p-values of .000 (or test statistics well above the critical value for the IPS finite sample test) in all specifications for the null that growth rates have unit roots (Table A4). For the ln levels of the different series, the results based upon the IPS and Fisher-L DF & -Z DF tests reported in the paper never reject the unit root null and were used to argue that the ln series contain a unit root. Tables A3 & A4 show that the Breitung test, which has negative size distortions in the Monte Carlos, similarly never rejects the unit root null, while tests with large positive size distortions in the Monte Carlos, such as Levin-Lin-Chu, the Fisher PP based tests, and the Fisher-P DF & Fisher-P_m DF, reject the null of a unit root in some instances.

Table A4: Stationarity (Hadri) and Unit Root (all others) Tests by KLEMS Series: test statistics relative to critical values (IPS finite sample) or p-values (all others)

		ln levels			growth rates		
		ln(TFP)	ln(K/L)	ln(I/L)	g(TFP)	g(K/L)	g(I/L)
		(A) without correction for autocorrelation:					
Hadri* (heteroskedastic cov)	FE	.000	.000	.000	.000	.000	.263
	trend	.000	.000	.000	.267	.000	.714
Hadri (homoskedastic cov)	FE	.000	.000	.000	.101	.000	.950
	trend	.000	.000	.000	.959	.000	.990
Harris-Tsavalis*	FE	.981	1.00	.023	.000	.000	.000
	trend	.830	1.00	.016	.000	.000	.000
Im-Pesaran-Shin* (finite sample)	FE	1.0	0.81	.96	3.6	3.0	3.7
	trend	.97	0.74	.98	2.6	2.3	2.7
Im-Pesaran-Shin (asymptotic distribution)	FE	.132	.960	.341	.000	.000	.000
	trend	.000	.089	.000	.000	.000	.000
Levin-Lin-Chu	FE	.001	.001	.118	.000	.000	.000
	trend	.000	.799	.006	.000	.000	.000
Breitung	FE	1.00	1.00	.479	.000	.000	.000
	trend	.984	1.00	.988	.000	.000	.000
Fisher-P DF & PP	FE	.001	.158	.005	.000	.000	.000
	trend	.005	.998	.000	.000	.000	.000
Fisher-Z DF & PP	FE	.039	.883	.168	.000	.000	.000
	trend	.243	1.00	.032	.000	.000	.000
Fisher-L DF & PP	FE	.044	.844	.163	.000	.000	.000
	trend	.173	1.00	.025	.000	.000	.000
Fisher-P_m DF & PP	FE	.000	.158	.002	.000	.000	.000
	trend	.002	.996	.000	.000	.000	.000
		(B) with correction for first order autocorrelation:					
Im-Pesaran-Shin* (asymptotic distribution)	FE	.280	.877	.920	.000	.000	.000
	trend	.773	.998	.912	.000	.000	.000
Levin-Lin-Chu	FE	.001	.001	.655	.000	.000	.000
	trend	.012	.262	.376	.000	.000	.000
Breitung	FE	.996	1.00	.254	.000	.000	.000
	trend	.906	.991	.927	.000	.000	.000
Fisher-P DF	FE	.029	.378	.732	.000	.000	.000
	trend	.369	.986	.326	.000	.000	.000
Fisher-Z DF*	FE	.216	.812	.929	.000	.000	.000
	trend	.778	.999	.826	.000	.000	.000
Fisher-L DF*	FE	.239	.865	.932	.000	.000	.000
	trend	.771	1.00	.810	.000	.000	.000
Fisher-P_m DF	FE	.023	.393	.740	.000	.000	.000
	trend	.383	.979	.339	.000	.000	.000
Fisher-P PP	FE	.001	.185	.008	.000	.000	.000
	trend	.003	.994	.000	.000	.000	.000

Table A4: continued

		ln levels			growth rates		
		ln(TFP)	ln(K/L)	ln(I/L)	g(TFP)	g(K/L)	g(I/L)
Fisher-Z PP	FE	.043	.862	.186	.000	.000	.000
	trend	.175	1.00	.020	.000	.000	.000
Fisher-L PP	FE	.046	.850	.176	.000	.000	.000
	trend	.117	1.00	.015	.000	.000	.000
Fisher-P_m PP	FE	.000	.188	.004	.000	.000	.000
	trend	.001	.988	.000	.000	.000	.000

Notes: (*) reported in paper based upon greater accuracy in Monte Carlos of Tables A1-A3. DF = Dickey Fuller & PP = Phillips-Perron, as described in text. above. FE = estimating equations include industry & year fixed effects as controls; trend = time specific trends in addition to industry & year fixed effects.

References

Breitung, J. (2000). The local power of some unit root tests for panel data. In Advances in Econometrics, Volume 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels. B. H. Baltagi, editor. Amsterdam: JAI Press.

Dickey, D. A., and W. A. Fuller (1979). "Distribution of the estimators for autoregressive time series with a unit root." *Journal of the American Statistical Association* 74: 427–431.

Hadri, K. (2000). "Testing for stationarity in heterogeneous panel data." *Econometrics Journal* 3: 148–161.

Harris, R. D. F., and E. Tzavalis (1999). "Inference for unit roots in dynamic panels where the time dimension is fixed." *Journal of Econometrics* 91: 201–226.

Im, K. S., M. H. Pesaran, and Y. Shin (2003). "Testing for unit roots in heterogeneous panels." *Journal of Econometrics* 115: 53–74.

Levin, A., C. F. Lin, and C. S. J. Chu (2002). "Unit root tests in panel data: Asymptotic and finite-sample properties." *Journal of Econometrics* 108: 1–24.

Phillips, P. C. B., and P. Perron (1988). "Testing for a unit root in time series regression." *Biometrika* 75: 335–346.

B. (III.5) & (III.7) with Average Factor Shares as Second Order Approximations

Equations (III.5) and (III.7) in the paper are based upon instantaneous rates of change. In the paper I claim that for changes across discrete time periods the same formulae are correct for second order approximations of the underlying functions if the average across the two periods of first order approximations of factor shares are used in place of the instantaneous $\boldsymbol{\theta}_{it}$. This motivates applying the formulae with average observed shares as a discrete time approximation. This appendix lays out the argument.

We may re-express the production function (III.1) as

$$\text{(B1)}\ \ \ln Q_{it} = \ln F^i(e^{\ln A_{1it}X_{1it}}, e^{\ln A_{2it}X_{2it}}, ..., e^{\ln A_{Jit}X_{Jit}}).$$

Its second order Taylor series (TS2) approximation around the point ($\ln(A_1X_1)$,..., $\ln(A_JX_J)$) is given by the translog production function:

$$\text{(B2)}\ \ \ln Q_{it}^{TS2} = \alpha + \sum_{j=1}^{J}\alpha_j \ln(\tilde{A}_{jit}\tilde{X}_{jit}) + ½\sum_{j=1}^{J}\sum_{k=1}^{J}\beta_{jk}\ln(\tilde{A}_{jit}\tilde{X}_{jit})\ln(\tilde{A}_{kit}\tilde{X}_{kit}),\ \text{where } \tilde{A}_{jit} = A_{jit}/A_j,$$

$$\tilde{X}_{jit} = X_{jit}/X_j,\ \ \alpha = \ln F^i(e^{\ln A_1X_1}, e^{\ln A_2X_2}, ..., e^{\ln A_JX_J}),\ \ \alpha_j = \frac{\partial \ln Q_{it}}{\partial \ln A_{jit}} = \frac{\partial \ln Q_{it}}{\partial \ln X_{jit}},$$

$$\beta_{jk} = \beta_{kj} = \frac{\partial^2 \ln Q_{it}}{\partial \ln X_{jit}\partial \ln X_{kit}} = \frac{\partial^2 \ln Q_{it}}{\partial \ln A_{jit}\partial \ln A_{kit}} = \frac{\partial^2 \ln Q_{it}}{\partial \ln X_{jit}\partial \ln A_{kit}} = \frac{\partial^2 \ln Q_{it}}{\partial \ln A_{jit}\partial \ln X_{kit}},$$

and all derivatives are evaluated at the point ($\ln(A_1X_1)$,..., $\ln(A_JX_J)$). I note that with perfect competition the partial of $\ln Q_{it}$ with respect to $\ln X_{jit}$ equals the factor income share θ_{jit}, the first order Taylor approximation of which at the point ($\ln(A_1X_1)$,..., $\ln(A_JX_J)$) is

$$\text{(B3)}\ \ \theta_{jit}^{TS1} = \frac{\partial \ln Q_{it}}{\partial \ln X_{jit}} + \sum_{k=1}^{J}\frac{\partial^2 \ln Q_{it}}{\partial \ln X_{jit}\partial \ln X_{kit}}\ln(\tilde{A}_{kit}\tilde{X}_{kit}) = \alpha_j + \sum_{k=1}^{J}\beta_{jk}\ln(\tilde{A}_{kit}\tilde{X}_{kit}),$$

where all derivatives are again evaluated at ($\ln(A_1X_1)$,..., $\ln(A_JX_J)$). We then see that

$$\text{(B4)}\ \theta_{jit}^{TS1} = \frac{\partial \ln Q_{it}^{TS2}}{\partial \ln X_{jit}}.$$

Finally, we note that

$$\text{(B5)}\ \ \ln\frac{Q_{it}^{TS2}}{Q_{it-1}^{TS2}} = \sum_{j=1}^{J}\left(\frac{\partial \ln Q_{it}^{TS2}}{\partial \ln X_{jit}} + \frac{\partial \ln Q_{it-1}^{TS2}}{\partial \ln X_{jit-1}}\right)\ln\frac{A_{jit}X_{jit}}{A_{jit-1}X_{jit-1}} = \sum_{j=1}^{J}\left(\frac{\theta_{jit}^{TS1}+\theta_{jit-1}^{TS1}}{2}\right)\ln\frac{A_{jit}X_{jit}}{A_{jit-1}X_{jit-1}},$$

$$\text{so that}\quad \ln\frac{Q_{it}^{TS2}}{Q_{it-1}^{TS2}} - \sum_{j=1}^{J}\left(\frac{\theta_{jit}^{TS1}+\theta_{jit-1}^{TS1}}{2}\right)\ln\frac{X_{jit}}{X_{jit-1}} = \sum_{j=1}^{J}\left(\frac{\theta_{jit}^{TS1}+\theta_{jit-1}^{TS1}}{2}\right)\ln\frac{A_{jit}}{A_{jit-1}}.$$

as was claimed in the paper.

Turning to the upper rows of (III.5), we focus on the case of the nested three factor production function in (III.6) with $\mathbf{E}_{it}^{-1}$ given by (III.7). In this case, the upper rows of (III.5) can be re-expressed as

$$\text{(B6)}\ \ \mathbf{E}_{it}\begin{bmatrix} g(A_{1it}X_{1it}/A_{3it}X_{3it}) \\ g(A_{2it}X_{2it}/A_{3it}X_{3it}) \end{bmatrix} = \begin{bmatrix} g(p_{1it}A_{3it}/p_{3it}A_{1it}) \\ g(p_{2it}A_{3it}/p_{3it}A_{2it}) \end{bmatrix}$$

$$\text{with } \mathbf{E}_{it} = \begin{bmatrix} -\frac{s_{2it}}{\sigma} - \frac{s_{1it}}{\eta} & \left(\frac{1}{\sigma}-\frac{1}{\eta}\right)s_{2it} \\ \left(\frac{1}{\sigma}-\frac{1}{\eta}\right)s_{1it} & -\frac{s_{1it}}{\sigma} - \frac{s_{2it}}{\eta} \end{bmatrix},\ \text{where } s_{jit} = \frac{\theta_{jit}}{\theta_{1it}+\theta_{2it}} \text{ for } j = 1 \text{ or } 2.$$

We wish to show that for discrete time periods (B6) with the average of first order Taylor series approximations of the s_{jit} is correct for a second order approximation of the profit maximizing condition (III.3).

We begin by deriving some useful relations. From Euler's theorem and homogeneity of degree 1 of G^i we have

$$\text{(B7)}\quad G_1^i\left(A_{1it}X_{1it},A_{2it}X_{2it}\right)A_{1it}X_{1it}+G_2^i\left(A_{1it}X_{1it},A_{2it}X_{2it}\right)A_{2it}X_{2it}=G^i\left(A_{1it}X_{1it},A_{2it}X_{2it}\right)$$

$$\frac{d}{d(A_{1it}X_{1it})}\Rightarrow\; G_{11}^iA_{1it}X_{1it}=-G_{21}^iA_{2it}X_{2it},\text{ with } G_j^i=\frac{\partial G^i}{\partial A_{jit}X_{jit}}\;\&\;G_{jk}^i=\frac{\partial^2G^i}{\partial A_{jit}X_{jit}\partial A_{kit}X_{kit}}.$$

Profit maximizing firms set the ratio of marginal products equal to the ratio of factor prices. For the first two factors this implies

$$\text{(B8)}\quad \frac{G_1^i(A_{1it}X_{1it}/A_{2it}X_{2it},1)}{G_2^i(A_{1it}X_{1it}/A_{2it}X_{2it},1)}=\frac{p_{1t}/A_{1it}}{p_{2t}/A_{2it}}$$

where we use the homogeneity of degree zero of derivatives of G^i. Differentiating:

$$\text{(B9)}\quad \frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}\frac{G_{11}^iG_2^i-G_{21}^iG_1^i}{G_2^iG_1^i}\frac{d\left(\frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}\right)}{\frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}}=\frac{d\left(\frac{p_{1t}A_{2it}}{p_{2t}A_{1it}}\right)}{\frac{p_{1t}A_{2it}}{p_{2t}A_{1it}}}$$

$$\Rightarrow\frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}\frac{G_{11}^iG_2^i+\frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}G_{11}^iG_1^i}{G_2^iG_1^i}=\frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}\frac{G_{11}^i}{G_1^i}\frac{\theta_{2it}+\theta_{1it}}{\theta_{2it}}=\frac{d\left(\frac{p_{1t}A_{2it}}{p_{2t}A_{1it}}\right)}{\frac{p_{1t}A_{2it}}{p_{2t}A_{1it}}}\frac{\frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}}{d\left(\frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}\right)}=-\frac{1}{\sigma}$$

$$\Rightarrow\frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}\frac{G_{11}^i}{G_1^i}=-\frac{s_{2it}}{\sigma}\quad\Rightarrow\frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}\frac{G_{21}^i}{G_2^i}=-\left(\frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}\right)^2\frac{G_1^i}{G_2^i}\frac{G_{11}^i}{G_1^i}=\frac{s_{1it}}{\sigma},$$

where we define σ as the elasticity of substitution, i.e. the proportional change in the ratio of effective input 1 ($A_{1it}X_{1it}$) to effective input 2 ($A_{2it}X_{2it}$) for a proportional change in the relative cost per unit of effective input. (B8) also allows us to solve (at least implicitly) for the effective input 1 to input 2 ratio as a function of the relative costs per effective factor:

$$\text{(B10)}\quad \frac{A_{1it}X_{1it}}{A_{2it}X_{2it}}=f^i\left(\frac{p_{1t}}{A_{1it}},\frac{p_{2t}}{A_{2it}}\right).$$

As G^i is homogeneous of degree one, this implies the following cost per unit of G^i

$$\text{(B11)}\quad p_{G^it}=\frac{1}{G^i(f^i,1)}\left[\frac{p_{1t}}{A_{1it}}f^i\left(\frac{p_{1t}}{A_{1it}},\frac{p_{2t}}{A_{2it}}\right)+\frac{p_{2t}}{A_{2it}}\right].$$

In a two stage budgeting procedure, firms first set the ratio of marginal products of G^i to X_{3it} equal to the ratio of their relative prices

(B12) $\dfrac{F_G^i(G^i, A_{3it}X_{3it})}{A_{3it}F_3^i(G^i, A_{3it}X_{3it})} = \dfrac{F_G^i(g_{it},1)}{A_{3it}F_I^i(g_{it},1)} = \dfrac{p_{G^it}}{P_{3it}}$, with $F_G^i = \dfrac{\partial F^i}{\partial G^i}, F_3^i = \dfrac{\partial F^i}{\partial A_{3it}X_{3it}}$, & $g_{it} = \dfrac{G^i}{A_{3it}X_{3it}}$,

and totally differentiating then allows us to define the elasticity of substitution between the G^i aggregate and effective input 3:

$$\text{(B13)}\quad \frac{d\left(\dfrac{F_G^i(g_{it},1)}{F_3^i(g_{it},1)}\right)}{dg_{it}}dg_{it} = d\left(\frac{p_{G^it}}{P_{3it}/A_{3it}}\right)$$

$$\Rightarrow \frac{d\ln\left(\dfrac{F_G^i(e^{\ln g_{it}},1)}{F_3^i(e^{\ln g_{it}},1)}\right)}{d\ln g_{it}} = \frac{d\left(\dfrac{F_G^i(g_{it},1)}{F_3^i(g_{it},1)}\right)}{\dfrac{F_G^i(g_{it},1)}{F_3^i(g_{it},1)}}\frac{g_{it}}{dg_{it}} = \frac{d\left(\dfrac{p_{G^it}}{P_{3it}/A_{3it}}\right)}{\dfrac{p_{G^it}}{P_{3it}/A_{3it}}}\frac{g_{it}}{dg_{it}} = -\frac{1}{\eta}.$$

With the above results, we turn finally to the second order approximation of (III.3). For j = 1 or 2 and X_{3it} as the "numeraire" factor we can express the j^{th} row of (III.3) as

$$\text{(B14)}\quad h_{it}^j\left(\frac{A_{1it}X_{1it}}{A_{3it}X_{3it}}, \frac{A_{2it}X_{2it}}{A_{3it}X_{3it}}\right) = \frac{G_j^i\left(\dfrac{A_{1it}X_{1it}}{A_{3it}X_{3it}}\dfrac{A_{3it}X_{3it}}{A_{2it}X_{2it}}, 1\right)F_G^i(g_{it}, 1)}{F_3^i(g_{it}, 1)} = \frac{p_{jt}/A_{jit}}{p_{3t}/X_{3it}},$$

$$\text{where } g_{it} = G^i\left(\frac{A_{1it}X_{1it}}{A_{3it}X_{3it}}, \frac{A_{2it}X_{2it}}{A_{3it}X_{3it}}\right)$$

and where we make use of the homogeneity of degree zero of first derivatives and the homogeneity of degree one of G^i. Define

$$\text{(B15)}\quad x_{1it} = \frac{A_{1it}X_{1it}}{A_{3it}X_{3it}},\ x_{2it} = \frac{A_{2it}X_{2it}}{A_{3it}X_{3it}},\ \&\ g_{it} = \frac{G^i}{A_{3it}X_{3it}} = G^i(x_{1it}, x_{2it}),$$

so that we may write

$$\text{(B16)}\quad \ln h_{it}^j(e^{\ln x_{1it}}, e^{\ln x_{2it}}) = \ln G_j^i(e^{\ln x_{1it}-\ln x_{2it}}, 1) + \ln\left(\frac{F_G^i(e^{\ln g_{it}}, 1)}{F_3^i(e^{\ln g_{it}}, 1)}\right).$$

We now define and note

$$\text{(B17)}\quad \gamma_1^{1i} = \frac{\partial \ln h_{it}^1(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial \ln x_{1it}} = \frac{G_{11}^i}{G_1^i}\frac{x_{1it}}{x_{2it}} + \frac{d\ln\left(\dfrac{F_G^i(e^{\ln g_{it}},1)}{F_3^i(e^{\ln g_{it}},1)}\right)}{d\ln g_{it}}\frac{G_1^i x_{1it}}{g_{it}} = -\frac{1}{\sigma}s_{2it} - \frac{1}{\eta}s_{1it},$$

$$\gamma_2^{1i} = \frac{\partial \ln h_{it}^1(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial \ln x_{2it}} = -\frac{G_{11}^i}{G_1^i}\frac{x_{1it}}{x_{2it}} + \frac{d\ln\left(\dfrac{F_G^i(e^{\ln g_{it}},1)}{F_3^i(e^{\ln g_{it}},1)}\right)}{d\ln g_{it}}\frac{G_2^i x_{2it}}{g_{it}} = \frac{1}{\sigma}s_{2it} - \frac{1}{\eta}s_{2it},$$

$$\gamma_1^{2i} = \frac{\partial \ln h_{it}^2(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial \ln x_{1it}} = \frac{G_{21}^i}{G_2^i}\frac{x_{1it}}{x_{2it}} + \frac{d\ln\left(\dfrac{F_G^i(e^{\ln g_{it}},1)}{F_3^i(e^{\ln g_{it}},1)}\right)}{d\ln g_{it}}\frac{G_1^i x_{1it}}{g_{it}} = \frac{1}{\sigma}s_{1it} - \frac{1}{\eta}s_{1it},$$

$$\gamma_2^{2i} = \frac{\partial \ln h_{it}^2(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial \ln x_{2it}} = -\frac{G_{21}^i}{G_2^i}\frac{x_{1it}}{x_{2it}} + \frac{d\ln\left(\dfrac{F_G^i(e^{\ln g_{it}},1)}{F_3^i(e^{\ln g_{it}},1)}\right)}{d\ln g_{it}}\frac{G_2^i x_{2it}}{g_{it}} = -\frac{1}{\sigma}s_{1it} - \frac{1}{\eta}s_{2it},$$

where we make use of the results in (B9) and (B13). As the elasticities are constant across the range of the data (i.e. region of approximation), we have

$$\text{(B18)}\quad \gamma_{11}^{1i} = \frac{\partial^2 \ln h_{it}^1(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial^2 \ln x_{1it}} = -\frac{1}{\sigma}\frac{\partial s_{2it}}{\partial \ln x_{1it}} - \frac{1}{\eta}\frac{\partial s_{1it}}{\partial \ln x_{1it}}$$

$$\gamma_{22}^{1i} = \frac{\partial^2 \ln h_{it}^1(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial^2 \ln x_{2it}} = \frac{1}{\sigma}\frac{\partial s_{2it}}{\partial \ln x_{2it}} - \frac{1}{\eta}\frac{\partial s_{2it}}{\partial \ln x_{2it}},$$

$$\gamma_{21}^{1i} = \frac{\partial^2 \ln h_{it}^1(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial \ln x_{2it}\partial \ln x_{1it}} = \frac{1}{\sigma}\frac{\partial s_{2it}}{\partial \ln x_{1it}} - \frac{1}{\eta}\frac{\partial s_{2it}}{\partial \ln x_{1it}} = -\frac{1}{\sigma}\frac{\partial s_{2it}}{\partial \ln x_{2it}} - \frac{1}{\eta}\frac{\partial s_{1it}}{\partial \ln x_{2it}} = \frac{\partial^2 \ln h_{it}^1(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial \ln x_{1it}\partial \ln x_{2it}} = \gamma_{12}^{1i}$$

$$\gamma_{11}^{2i} = \frac{\partial^2 \ln h_{it}^2(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial^2 \ln x_{1it}} = \frac{1}{\sigma}\frac{\partial s_{1it}}{\partial \ln x_{1it}} - \frac{1}{\eta}\frac{\partial s_{1it}}{\partial \ln x_{1it}}$$

$$\gamma_{22}^{2i} = \frac{\partial^2 \ln h_{it}^2(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial^2 \ln x_{2it}} = -\frac{1}{\sigma}\frac{\partial s_{1it}}{\partial \ln x_{2it}} - \frac{1}{\eta}\frac{\partial s_{2it}}{\partial \ln x_{2it}},$$

$$\gamma_{21}^{Li} = \frac{\partial^2 \ln h_{it}^2(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial \ln x_{2it}\partial \ln x_{1it}} = -\frac{1}{\sigma}\frac{\partial s_{1it}}{\partial \ln x_{1it}} - \frac{1}{\eta}\frac{\partial s_{2it}}{\partial \ln x_{1it}} = \frac{1}{\sigma}\frac{\partial s_{1it}}{\partial \ln x_{2it}} - \frac{1}{\eta}\frac{\partial s_{1it}}{\partial \ln x_{2it}} = \frac{\partial^2 \ln h_{it}^2(e^{\ln x_{1it}}, e^{\ln x_{2it}})}{\partial \ln x_{1it}\partial \ln x_{2it}} = \gamma_{12}^{2i}.$$

where we have kept in mind that as s_{jit} is a function of $\ln x_{1it}/x_{2it}$, and $s_{1it}+s_{2it} = 1$

$$\text{(B19)}\quad \frac{\partial s_{jit}}{\partial \ln x_{1it}} = -\frac{\partial s_{jit}}{\partial \ln x_{2it}},\ \frac{\partial s_{1it}}{\partial \ln x_{1it}} = -\frac{\partial s_{2it}}{\partial \ln x_{1it}},\ \&\ \frac{\partial s_{1it}}{\partial \ln x_{2it}} = -\frac{\partial s_{2it}}{\partial \ln x_{2it}}.$$

The second order Taylor series approximation of (III.3) is then given by:

$$\text{(B20)}\quad \ln h_{it}^{jTS2} = \ln h^{j*} + \gamma_1^{j*}\ln\frac{x_{1it}}{x_1^*} + \gamma_2^{j*}\ln\frac{x_{2it}}{x_2^*}$$

$$+ ½\gamma_{11}^{j*}\ln\frac{x_{1it}}{x_1^*}\ln\frac{x_{1it}}{x_1^*} + ½\gamma_{22}^{j*}\ln\frac{x_{2it}}{x_2^*}\ln\frac{x_{2it}}{x_2^*} + \gamma_{12}^{j*}\ln\frac{x_{1it}}{x_1^*}\ln\frac{x_{2it}}{x_2^*}$$

where the superscripted * denote values at the point of expansion (x_1^*, x_2^*). Similarly, the first order Taylor expansions of s_{1it} & s_{2it} are given by

$$\text{(B21)}\quad s_{1it}^{TS1} = s_{1i}^* + \frac{\partial s_{1it}^*}{\partial \ln x_{1it}}\ln\frac{x_{1it}}{x_1^*} + \frac{\partial s_{1it}^*}{\partial \ln x_{2it}}\ln\frac{x_{2it}}{x_2^*}\ \ \&\ \ s_{2it}^{TS1} = s_{2i}^* + \frac{\partial s_{2it}^*}{\partial \ln x_{1it}}\ln\frac{x_{1it}}{x_1^*} + \frac{\partial s_{2it}^*}{\partial \ln x_{2it}}\ln\frac{x_{2it}}{x_2^*}.$$

Taking the ln difference across two time periods of (B20)

$$\text{(B22)}\quad \ln\frac{h_{it}^{jTS2}}{h_{it-1}^{jTS2}} = \frac{1}{2}\left(\frac{\partial \ln h_{it}^{jTS2}}{\partial \ln(x_{1it}/x_1^*)} + \frac{\partial \ln h_{it-1}^{jTS2}}{\partial \ln(x_{1it-1}/x_1^*)}\right)\ln\frac{x_{1it}}{x_{1it-1}} + \frac{1}{2}\left(\frac{\partial \ln h_{it}^{jTS2}}{\partial \ln(x_{2it}/x_2^*)} + \frac{\partial \ln h_{it-1}^{jTS2}}{\partial \ln(x_{2it-1}/x_2^*)}\right)\ln\frac{x_{2it}}{x_{2it-1}}$$

where

$$\text{(B23)}\quad \frac{\partial \ln h_{it}^{1TS2}}{\partial \ln(x_{1it}/x_1^*)} = \gamma_1^{1i*} + \gamma_{11}^{1i*} \ln\frac{x_{1it}}{x_1^*} + \gamma_{21}^{1i*} \ln\frac{x_{2it}}{x_2^*} = -\frac{1}{\sigma}s_{2i}^* - \frac{1}{\eta}s_{1i}^* +$$

$$\left(-\frac{1}{\sigma}\frac{\partial s_{2it}^*}{\partial \ln x_{1it}} - \frac{1}{\eta}\frac{\partial s_{1it}^*}{\partial \ln x_{1it}}\right)\ln\frac{x_{1it}}{x_1^*} + \left(-\frac{1}{\sigma}\frac{\partial s_{2it}}{\partial \ln x_{2it}} - \frac{1}{\eta}\frac{\partial s_{1it}}{\partial \ln x_{2it}}\right)\ln\frac{x_{2it}}{x_2^*} = -\frac{s_{2it}^{TS1}}{\sigma} - \frac{s_{1it}^{TS1}}{\eta}$$

and it is similarly easily seen that

$$\text{(B24)}\quad \frac{\partial \ln h_{it}^{1TS2}}{\partial \ln(x_{2it}/x_2^*)} = \frac{s_{2it}^{TS1}}{\sigma} - \frac{s_{2it}^{TS1}}{\eta},\quad \frac{\partial \ln h_{it}^{2TS2}}{\partial \ln(x_{1it}/x_1^*)} = \frac{s_{1it}^{TS1}}{\sigma} - \frac{s_{1it}^{TS1}}{\eta},\ \ \&\ \ \frac{\partial \ln h_{it}^{2TS2}}{\partial \ln(x_{2it}/x_2^*)} = -\frac{s_{1it}^{TS1}}{\sigma} - \frac{s_{2it}^{TS1}}{\eta}.$$

so that across discrete time periods the upper part of (III.5) can be approximated as:

$$\text{(B25)}\quad \mathbf{E}_{it}\begin{bmatrix}\ln(x_{1it}/x_{1it-1})\\ \ln(x_{2it}/x_{2it-1})\end{bmatrix} = \begin{bmatrix}\ln(p_{1it}A_{3it}/p_{3it}A_{1it}) - \ln(p_{1it-1}A_{3it-1}/p_{3it-1}A_{1it-1})\\ \ln(p_{2it}A_{3it}/p_{3it}A_{2it}) - \ln(p_{2it-1}A_{3it-1}/p_{3it-1}A_{2it-1})\end{bmatrix}$$

$$\text{with } \mathbf{E}_{it} = \begin{bmatrix} -\frac{½(s_{2it}^{TS1}+s_{2it-1}^{TS1})}{\sigma} - \frac{½(s_{1it}^{TS1}+s_{1it-1}^{TS1})}{\eta} & \left(\frac{1}{\sigma}-\frac{1}{\eta}\right)\left(\frac{s_{2it}^{TS1}+s_{2it-1}^{TS1}}{2}\right) \\ \left(\frac{1}{\sigma}-\frac{1}{\eta}\right)\left(\frac{s_{1it}^{TS1}+s_{1it-1}^{TS1}}{2}\right) & -\frac{½(s_{1it}^{TS1}+s_{1it-1}^{TS1})}{\sigma} - \frac{½(s_{2it}^{TS1}+s_{2it-1}^{TS1})}{\eta} \end{bmatrix},$$

as was our objective to show.

C. Consistency with Factor Price Growth and Factor Shares as Regressors

As noted in the paper, the endogeneity of relative factor price growth does not inhibit consistency of the estimation procedure as long as year fixed effects are included, even though with non-constant factor shares factor price growth still affects (albeit in practice minimally) elasticity estimates in finite samples. This appendix uses a simplified version of the model to establish this principle theoretically with the minimum algebra possible, showing that independence of the factor augmenting shocks from the levels of factor shares is sufficient to ensure that the derivatives of a normal likelihood asymptotically equal zero at the true parameter values. The growth rate of factor shares, however, is affected by the factor augmenting shocks, and this induces a very slight correlation between the level of factor shares and the shocks, raising the possibility of inconsistency due to this relation. To alleviate such concerns, this appendix uses Monte Carlos to establish that for the elasticities estimated in the paper these correlations do not inhibit root-N convergence of parameters (& N super convergence of cointegration parameters) to materially trivial levels of mean squared error. The factor prices in these Monte Carlos are endogenous, and hence they illustrate (more generally than the specific example in the theoretical proof) that this does not inhibit consistency. I begin with the Monte Carlos, as these are the easiest to absorb.

(i) Monte Carlos

I use the point estimates of models 1 & 2 of Section IV of the paper to establish the data generating process. From these models we get estimates of the industry x factor fixed effects, year x factor fixed effects, diagonal variance of year x industry factor augmenting shocks, elasticities, coefficients on lagged values of shocks, and, where applicable, degrees of freedom of the t-distribution governing the factor augmenting shocks and the cointegration parameters **β** and **α**. I use all of these as is, except for the fixed effects, for which I calculate the (unrestricted) covariance across factors and then use independent draws from the multivariate normal[1] to create new industry x factor and time x factor fixed effects for the expanding industry x time samples below. Thus, the data generating process for factor augmenting technical change is, as in the models of the paper:

$$\text{(C1a)}\quad \mathbf{g}(A_{jit}) = \hat{\mathbf{\Gamma}}\mathbf{g}(A_{jit-1}) + \boldsymbol{\eta}_t + \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{it}$$

$$\text{(C1b)}\quad \mathbf{g}(A_{jit}) = \hat{\mathbf{\Gamma}}\mathbf{g}(A_{jit-1}) + \hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}}'\mathbf{ln}(A_{jit-1}) + \boldsymbol{\eta}_t + \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{it}$$

where $\boldsymbol{\eta}_t \sim N(\mathbf{0},\mathbf{V}(\hat{\boldsymbol{\eta}}_t))$, $\boldsymbol{\eta}_i \sim N(\mathbf{0},\mathbf{V}(\hat{\boldsymbol{\eta}}_i))$, and $\boldsymbol{\varepsilon}_{it} \sim N(\mathbf{0},\mathbf{V}(\hat{\boldsymbol{\varepsilon}}_{it}))$ or $\sim t_v(\mathbf{0},\mathbf{V}(\hat{\boldsymbol{\varepsilon}}_{it}))$, and where hats indicate values estimated in the paper, V(**x**) indicates the covariance matrix of vector **x**, and while the covariance matrix of the fixed effects $\boldsymbol{\eta}_t$ & $\boldsymbol{\eta}_i$ is unrestricted, the covariance matrix of the factor augmenting shocks $\boldsymbol{\varepsilon}_{it}$ is made diagonal. The means of the fixed effects are set equal to zero (contrary to their estimated means) so as to avoid a tendency of factor shares to gravitate to corners (0 or 1) as the number of time periods grows.

Factor augmenting productivity growth is, of course, treated as unobserved. The observed data consist of vectors of relative factor input growth, total factor productivity growth, relative factor price changes and factor shares. As in the paper:

$$\text{(C2)}\quad \begin{bmatrix} g(X_{1it}/X_{3it}) \\ g(X_{2it}/X_{3it}) \\ g(TFP_{it}) \end{bmatrix} = \begin{bmatrix} \mathbf{E}_{it}^{-1} \\ \mathbf{0}_2' \end{bmatrix} \begin{bmatrix} g(p_{1t}/p_{3t}) \\ g(p_{2t}/p_{3t}) \end{bmatrix} + \begin{bmatrix} -\mathbf{E}_{it}^{-1} - \mathbf{I}_2, \mathbf{E}_{it}^{-1}\mathbf{i}_2 + \mathbf{i}_2 \\ \overline{\boldsymbol{\theta}}_{it}' \end{bmatrix} \begin{bmatrix} g(A_{1it}) \\ g(A_{2it}) \\ g(A_{3it}) \end{bmatrix}$$

where

$$\text{(C3)}\quad \hat{\mathbf{E}}_{it}^{-1} = \frac{1}{\overline{\theta}_{1it} + \overline{\theta}_{2it}} \begin{bmatrix} -\hat{\sigma}\overline{\theta}_{2it} - \hat{\eta}\overline{\theta}_{1it} & (\hat{\sigma} - \hat{\eta})\overline{\theta}_{2it} \\ (\hat{\sigma} - \hat{\eta})\overline{\theta}_{1it} & -\hat{\sigma}\overline{\theta}_{1it} - \hat{\eta}\overline{\theta}_{2it} \end{bmatrix}$$

and where superscripted bars denote averages between periods t-1 and t of factor shares and hats again indicate elasticities estimated in the paper. Relative factor prices are determined by matching supply to changes in average relative factor demand with a relative supply growth curve with slope 1. Average relative factor

[1]For most factors the kurtosis of these estimated fixed effects is not exceptionally different from the normal, and hence I do not assume that they are distributed multivariate t.

demand growth is either determined by the average of all industries, a "small industry" model, or by one single dominant industry, a "large industry" model, so:

$$\text{(C4)}\quad \begin{matrix}\text{small}\\ \text{industries}\end{matrix}:\begin{bmatrix} g(p_{1t}/p_{3t}) \\ g(p_{2t}/p_{3t}) \end{bmatrix} = \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N} g(X_{1it}/X_{3it}) \\ \sum_{i=1}^{N} g(X_{2it}/X_{3it}) \end{bmatrix} \quad \begin{matrix}\text{one large}\\ \text{industry}\end{matrix}:\begin{bmatrix} g(p_{1t}/p_{3t}) \\ g(p_{2t}/p_{3t}) \end{bmatrix} = \begin{bmatrix} g(X_{11t}/X_{31t}) \\ g(X_{21t}/X_{31t}) \end{bmatrix},$$

where N denotes the number of industries. In the small industry model, the influence of individual industries on the growth of relative prices goes to zero, which asymptotically are driven solely by year fixed effects common to all industries. In the large industry model, although there are N industries which contribute data, factor augmenting productivity growth in industry #1, which is large, determines factor price changes, so that relative price growth is always influenced by both the year fixed effects (common to all industries) and the industry fixed effects and iid shocks specific to industry 1.

Factor shares in period 1 are set by taking three independent draws from the (0,1) uniform distribution for each industry, multiplying them by 2, 3 and 5, and dividing by their total, so that the initial shares for capital, labour and intermediates have an expected value of .2, .3 & .5 (as is roughly found in the data). Factor shares then evolve according to the formula:

$$\text{(C5)}\quad \begin{bmatrix} g(\theta_{1it}/\theta_{3it}) \\ g(\theta_{2it}/\theta_{3it}) \end{bmatrix} = \begin{bmatrix} g(p_{1t}/p_{3t}) \\ g(p_{2t}/p_{3t}) \end{bmatrix} + \begin{bmatrix} g(X_{1it}/X_{3it}) \\ g(X_{2it}/X_{3it}) \end{bmatrix}.$$

Given the complex dependence of factor augmenting productivity growth $\mathbf{g}(A_{jit})$ on past values and cumulated past values in (C1) above, 500 initial periods (which are then discarded) are used to initiate the system and ensure that the distribution of initial rates of factor augmenting productivity growth is close to its ergodic distribution.

Table C1 reports the root mean squared error of key parameter estimates across 50 data generating process realizations for the VAR, VEC1, normal and t versions of the top two models. The ratio of years to industries is set at 3:6, reflecting the ratio found in the 61 industry 1987-2021 BEA US KLEMS data in the paper, with each multiplied by 10 as the sample increases. Reported in the table are the mean squared error for the elasticities, the t-degrees of freedom and the cointegrating factors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. As expected of a consistent estimator, root mean squared error for most parameters is inversely proportional to the root number of observations, i.e. falls by an order of magnitude with a 100 fold increase in the sample size. The exceptions are the estimates of $\boldsymbol{\beta}$ which, by virtue of being multiplied by the non-stationary levels of factor augmenting productivity, are known to be superconsistent (Johansen 1995) with root mean square error inversely proportional to the number of observations, as shown

Table C1: Root Mean Squared Error Across 50 DGP Iterations of Key Parameter Estimates
(Monte Carlos based on point estimates of models 1 & 2)

model	VAR			VEC1							
years x industries	σ	η	*dof*	σ	η	*dof*	β_L	β_I	α_K	α_L	α_I
(A) small industry model: price growth determined by average industry demand growth											
model #1, normal distribution											
30 x 60	.0147	.0145		.0144	.0167		.1250	.2082	.0441	.0150	.0338
300 x 600	.0012	.0011		.0010	.0011		.0021	.0034	.0030	.0010	.0027
model #2, normal distribution											
30 x 60	.0123	.0285		.0120	.0298		.1319	.5702	.0339	.0102	.0211
300 x 600	.0012	.0023		.0010	.0022		.0020	.0079	.0026	.0005	.0017
model #1, t distribution											
30 x 60	.0167	.0162	.1579	.0163	.0175	.1610	.1586	.1551	.0076	.0050	.0055
300 x 600	.0012	.0018	.0141	.0009	.0029	.0146	.0039	.0060	.0004	.0002	.0002
model #2, t distribution											
30 x 60	.0127	.0377	.1525	.0133	.0384	.1524	.1531	.4503	.0063	.0029	.0045
300 x 600	.0010	.0025	.0164	.0013	.0030	.0186	.0037	.0189	.0003	.0001	.0002
(B) large industry model: price growth determined by demand growth in one industry											
model #1, normal distribution											
30 x 60	.0143	.0135		.0137	.0155		.1251	.2068	.0441	.0150	.0338
300 x 600	.0011	.0014		.0009	.0010		.0021	.0035	.0030	.0010	.0027
model #2, normal distribution											
30 x 60	.0117	.0256		.0116	.0271		.1329	.5665	.0339	.0102	.0211
300 x 600	.0013	.0025		.0010	.0019		.0020	.0084	.0026	.0005	.0017
model #1, t distribution											
30 x 60	.0152	.0124	.1579	.0149	.0143	.1610	.1579	.1557	.0076	.0049	.0054
300 x 600	.0011	.0012	.0144	.0011	.0018	.0146	.0043	.0072	.0003	.0002	.0002
model #2, t distribution											
30 x 60	.0120	.0300	.1526	.0124	.0309	.1526	.1455	.4514	.0063	.0029	.0044
300 x 600	.0012	.0022	.0190	.0017	.0026	.0182	.0033	.0308	.0003	.0002	.0004

Notes: Root mean squared error calculated across 50 *dgp* iterations, in some t cases dropping one or two instances where the likelihood did not converge. *dof* = multivariate t-distribution degrees of freedom.

in the table. In sum, as claimed earlier, despite the endogeneity of factor price movements and the levels of factor shares, with year and & industry fixed effects the estimates appear to be consistent in exactly the fashion expected of VAR & VEC models. This is true for both the "small" and "large" industry models.

(b) Formal Proof

For a formal proof that the endogeneity of relative price changes does not prevent consistency, I simplify so as to make the point with minimal algebraic complexity. I examine a two factor model estimated with the normal distribution, no cointegration, and only year fixed effects, in which the number of industries goes to infinity but the number of years is constant. For this framework, $\mathbf{E}_{it}^{-1}$ equals σ, the elasticity of substitution, and the model is

$$\text{(C6)}\quad \begin{bmatrix} g(X_{1it}/X_{2it}) \\ g(TFP_{it}) \end{bmatrix} = \begin{bmatrix} -\sigma \\ 0 \end{bmatrix} g(p_{1t}/p_{2t}) + \underbrace{\begin{bmatrix} \sigma-1 & 1-\sigma \\ \theta_{1it} & \theta_{2it} \end{bmatrix}}_{\mathbf{C}_{it}} \begin{bmatrix} g(A_{1it}) \\ g(A_{2it}) \end{bmatrix},$$

$$\text{with } \mathbf{y}_{it} = \begin{bmatrix} g(X_{1it}/X_{2it}) + \sigma g(p_{1t}/p_{2t}) \\ g(TFP_{it}) \end{bmatrix} \ \& \ \mathbf{g}(A_{jit}) = \mathbf{C}_{it}^{-1}\mathbf{y}_{it},$$

where we define the matrix $\mathbf{C}_{it}$ and $\mathbf{y}_{it}$ to reduce clutter below. Factor augmenting productivity growth depends only on year fixed effects and iid shocks:

$$\text{(C7)}\quad \mathbf{g}(A_{jit}) = \boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_{it},$$

We assume that the first moment of $E(\boldsymbol{\varepsilon}_{it}) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}_{it}\boldsymbol{\varepsilon}_{it}') = \mathbf{V}$, where $\mathbf{V}$ is diagonal, there exists a $\gamma > 1$ such that $E((\varepsilon_{jit}^2)^\gamma)$ exists, and that $\boldsymbol{\varepsilon}_{it}$ is independent of $\boldsymbol{\theta}_{it}$. For a given value of $\hat{\sigma}$, $\mathbf{g}(A_{jit})$ can be calculated from (C6). We use the notation $\mathbf{g}(A_{jit}(\hat{\sigma}))$ to distinguish between these estimated values and the true values $\mathbf{g}(A_{jit})$.

With a normal likelihood, the ln likelihood of the model is given by:

$$\text{(C8)}\quad \sum_{it}\ln L_{it} = -NT\ln(2\pi) - ½\sum_{it}\ln(\det(\hat{\mathbf{C}}_{it}\hat{\mathbf{V}}\hat{\mathbf{C}}_{it}')) - ½\sum_{it}(\hat{\mathbf{y}}_{it} - \hat{\mathbf{C}}_{it}\hat{\boldsymbol{\eta}}_t)'\hat{\mathbf{C}}_{it}'^{-1}\hat{\mathbf{V}}^{-1}\hat{\mathbf{C}}_{it}^{-1}(\hat{\mathbf{y}}_{it} - \hat{\mathbf{C}}_{it}\hat{\boldsymbol{\eta}}_t)$$

$$\text{or } \overline{L} = \sum_{it}\frac{\ln L_{it}}{NT} = -\ln(2\pi) - ½(\hat{\sigma}-1)^2 - ½\ln(\hat{v}_1\hat{v}_2) - ½\sum_{it}\frac{(\mathbf{g}(A_{jit}(\hat{\sigma})) - \hat{\boldsymbol{\eta}}_t)'\hat{\mathbf{V}}^{-1}(\mathbf{g}(A_{jit}(\hat{\sigma})) - \hat{\boldsymbol{\eta}}_t)}{NT},$$

where $\hat{\mathbf{V}}$ is diagonal with entries $(\hat{v}_1, \hat{v}_2)$ estimating the diagonal covariance matrix of the shocks, $\hat{\mathbf{C}}_{it}$ is the estimate of $\mathbf{C}_{it}$ based upon the estimate $\hat{\sigma}$ of σ, $\hat{\mathbf{y}}_{it}$ and $\mathbf{g}(A_{jit}(\hat{\sigma}))$ are based upon $\hat{\mathbf{C}}_{it}$ and (C6) above, $\sum_{it}$ denotes summation across both industry and time indices, N the number of industries, & T the number of time periods, we define the average likelihood $\overline{L}$ and use the fact that determinant$(\hat{\mathbf{C}}_{it}\hat{\mathbf{C}}_{it}') = (\hat{\sigma}-1)^2$.

I will establish conditions under which $\partial\overline{L}/\partial\hat{\sigma} = 0$ when $\hat{\sigma} = \sigma$, satisfying the first order condition for a maximum. This does not guarantee convergence of the estimate of $\hat{\sigma}$ to σ, which depends upon global (within the parameter range specified by the model) concavity of the objective function, but it does establish that the growth of relative prices does not prevent consistency of the estimator. In the paper I concentrate the likelihoods of the models as a function of the elasticities of substitution and (where applicable) degrees of freedom of the t-distribution and conduct systematic grid searches across these spaces. As noted therein, in all cases (i.e. twelve models each in their VAR, VEC1, VEC2, normal & t forms) I find the likelihood is single peaked within the specified elasticity of substitution restrictions.

Using the envelope theorem and the fact that we are maximizing with respect to the other parameters, we take the derivative of the average likelihood with respect to $\hat{\sigma}$:

$$\text{(C9)}\quad \frac{\partial \overline{L}(\hat{\sigma})}{\partial \hat{\sigma}} = -\frac{(\hat{\sigma}\text{-}1)}{(\hat{\sigma}\text{-}1)^2} - \sum_{it} \frac{\frac{\partial \mathbf{g}(A_{jit}(\hat{\sigma}))}{\partial \hat{\sigma}} \hat{\mathbf{V}}^{-1}(\mathbf{g}(A_{jit}(\hat{\sigma})) - \hat{\boldsymbol{\eta}}_t)}{NT}, \text{ with}$$

$$\frac{\partial \mathbf{g}(A_{jit}(\hat{\sigma}))}{\partial \hat{\sigma}} = \underbrace{\begin{bmatrix} -\theta_{2it}/(\hat{\sigma}-1)^2 & 0 \\ \theta_{1it}/(\hat{\sigma}-1)^2 & 0 \end{bmatrix}}_{\partial \mathbf{C}_{it}^{-1}/\partial \hat{\sigma}} \begin{bmatrix} g(X_{1it}/X_{2it}) + \hat{\sigma} g(p_{1t}/p_{2t}) \\ g(TFP_{it}) \end{bmatrix} + \underbrace{\begin{bmatrix} \theta_{2it}/(\hat{\sigma}-1) & 1 \\ \theta_{1it}/(1-\hat{\sigma}) & 1 \end{bmatrix}}_{\mathbf{C}_{it}^{-1}} \begin{bmatrix} g(p_{1t}/p_{2t}) \\ 0 \end{bmatrix},$$

$$\text{so that } \frac{\partial \mathbf{g}(A_{jit}(\sigma))}{\partial \hat{\sigma}} = \begin{bmatrix} -\theta_{2it}/(\sigma-1)^2 & 0 \\ \theta_{1it}/(\sigma-1)^2 & 0 \end{bmatrix} \begin{bmatrix} \sigma-1 & 1-\sigma \\ \theta_{1it} & \theta_{2it} \end{bmatrix} \begin{bmatrix} g(A_{1it}) \\ g(A_{2it}) \end{bmatrix} + \begin{bmatrix} \theta_{2it}/(\sigma-1) \\ \theta_{1it}/(1-\sigma) \end{bmatrix} g(p_{1t}/p_{2t})$$

$$= \frac{g(A_{1it}) - g(A_{2it})}{\sigma-1} \begin{bmatrix} -\theta_{2it} \\ \theta_{1it} \end{bmatrix} + \frac{g(p_{1t}/p_{2t})}{\sigma-1} \begin{bmatrix} \theta_{2it} \\ -\theta_{1it} \end{bmatrix}.$$

Substituting for $\mathbf{g}(A_{jit})$ and rearranging, we have:

$$\text{(C10)}\quad \frac{\partial \overline{L}(\sigma)}{\partial \hat{\sigma}} = \frac{1}{\sigma\text{-}1} \left\{ \begin{aligned} -1 + \sum_{it} \frac{\theta_{2it}[\eta_{1t} - \eta_{2t} + \varepsilon_{1it} - \varepsilon_{2it}]\hat{\varepsilon}_{1it}}{NT\hat{v}_1} - \sum_{it} \frac{\theta_{1it}[\eta_{1t} - \eta_{2t} + \varepsilon_{1it} - \varepsilon_{2it}]\hat{\varepsilon}_{2it}}{NT\hat{v}_2} \\ - \sum_{it} \frac{g(p_{1t}/p_{2t})\theta_{2it}\hat{\varepsilon}_{1it}}{NT\hat{v}_1} + \sum_{it} \frac{g(p_{1t}/p_{2t})\theta_{1it}\hat{\varepsilon}_{2it}}{NT\hat{v}_2} \end{aligned} \right\}$$

$$= \frac{1}{\sigma\text{-}1} \left\{ \begin{aligned} -1 + \sum_t \frac{\eta_{1t} - \eta_{2t} - g(p_{1t}/p_{2t})}{T} \left(\sum_i \frac{\theta_{2it}\hat{\varepsilon}_{1it}}{N\hat{v}_1} - \sum_i \frac{\theta_{1it}\hat{\varepsilon}_{2it}}{N\hat{v}_2} \right) \\ + \sum_t \frac{1}{T} \left(\sum_i \frac{[\varepsilon_{1it} - \varepsilon_{2it}]\hat{\varepsilon}_{1it}}{N\hat{v}_1} - \sum_i \frac{\theta_{1it}[\varepsilon_{1it} - \varepsilon_{2it}]}{N} \left(\frac{\hat{\varepsilon}_{2it}}{\hat{v}_2} + \frac{\hat{\varepsilon}_{1it}}{\hat{v}_1} \right) \right) \end{aligned} \right\}$$

$$\text{or}\quad = \frac{1}{\sigma\text{-}1} \left\{ \begin{aligned} -1 + \sum_t \frac{\eta_{1t} - \eta_{2t} - g(p_{1t}/p_{2t})}{T} \left(\sum_i \frac{\theta_{2it}\hat{\varepsilon}_{1it}}{N\hat{v}_1} - \sum_i \frac{\theta_{1it}\hat{\varepsilon}_{2it}}{N\hat{v}_2} \right) \\ + \sum_t \frac{1}{T} \left(\sum_i \frac{[\varepsilon_{2it} - \varepsilon_{1it}]\hat{\varepsilon}_{2it}}{N\hat{v}_2} + \sum_i \frac{\theta_{2it}[\varepsilon_{1it} - \varepsilon_{2it}]}{N} \left(\frac{\hat{\varepsilon}_{2it}}{\hat{v}_2} + \frac{\hat{\varepsilon}_{1it}}{\hat{v}_1} \right) \right) \end{aligned} \right\}$$

The two (alternative) formulations cover two possibilities discussed further below. Using the maximum likelihood solutions for $\hat{\boldsymbol{\eta}}_t$ and $\hat{v}_j$, we have

$$\text{(C11)}\quad \hat{\boldsymbol{\varepsilon}}_{it} = \mathbf{g}(A_{jit}(\hat{\sigma})) - \hat{\boldsymbol{\eta}}_t = \boldsymbol{\varepsilon}_{it} - \frac{1}{N}\sum_i \boldsymbol{\varepsilon}_{it} \;\;\&\;\; \hat{v}_j = \sum_{it} \frac{\hat{\boldsymbol{\varepsilon}}_{it}^2}{NT} = \sum_{it} \frac{\boldsymbol{\varepsilon}_{it}^2}{NT} - \sum_t \left(\frac{1}{N} \sum_i \boldsymbol{\varepsilon}_{it} \right)^2,$$

where as before $\sum_{it}$ denotes summation across both industry and time indices, while $\sum_i$ & $\sum_t$ denote summation across the industry or time indices alone. From the strong law of large numbers, as $N \to \infty$

$$\text{(C12)}\quad \sum_i \frac{\boldsymbol{\varepsilon}_{it}}{N} \xrightarrow{a.s.} \mathbf{0},\; \sum_i \frac{\varepsilon_{1it}\varepsilon_{2it}}{N} \xrightarrow{a.s.} \mathbf{0},\; \sum_i \frac{\varepsilon_{ait}^2}{N} \xrightarrow{a.s.} v_a \;\&\; \hat{v}_a \xrightarrow{a.s.} v_a \;\text{ for } a = 1 \text{ or } 2.$$

In addition, for a and b equal to 1 or 2

$$\text{(C13)}\quad E\left(\sum_i \frac{\theta_{ait}\varepsilon_{bit}}{N}\right) = \sum_i \frac{E(\theta_{ait})E(\varepsilon_{bit})}{N} = \sum_i \frac{E(\theta_{ait})*0}{N} = 0$$

$$E\left(\sum_i \frac{\theta_{ait}\varepsilon_{bit}}{N}\sum_j \frac{\theta_{ajt}\varepsilon_{bjt}}{N}\right) = \sum_{ij} \frac{E(\theta_{ait}\theta_{bjt}\varepsilon_{ait}\varepsilon_{bjt})}{N^2} = \sum_i \frac{E(\theta_{ait}^2)E(\varepsilon_{bit}^2)}{N^2} \le \frac{v_b}{N},$$

where the last line follows from the fact that for $i \neq j$ $E(\theta_{ait}\theta_{ajt}\varepsilon_{bit}\varepsilon_{bjt}) = E(\theta_{ait}\theta_{ajt})E(\varepsilon_{bit}\varepsilon_{bjt}) = E(\theta_{ait}\theta_{ajt})*0$, and throughout we make use of the fact that $0 \le E(\theta_{ait}) \le 1$ & $0 \le E(\theta_{ait}\theta_{ajt}) \le 1$. From (C13), we see that $\sum_i\theta_{ait}\varepsilon_{bit}/N$ converges in mean square, and hence in probability, to 0. Similarly, for $b \neq c$

$$\text{(C14)}\quad E\left(\sum_i \frac{\theta_{ait}\varepsilon_{bit}\varepsilon_{cit}}{N}\right) = \sum_i \frac{E(\theta_{ait})E(\varepsilon_{bit})E(\varepsilon_{cit})}{N} = \sum_i \frac{E(\theta_{ait})*0*0}{N} = 0$$

$$E\left(\sum_i \frac{\theta_{ait}\varepsilon_{bit}\varepsilon_{cit}}{N}\sum_j \frac{\theta_{ajt}\varepsilon_{bjt}\varepsilon_{cjt}}{N}\right) = \sum_{ij} \frac{E(\theta_{ait}\theta_{ajt}\varepsilon_{bit}\varepsilon_{bjt}\varepsilon_{cit}\varepsilon_{bjt})}{N^2} = \sum_i \frac{E(\theta_{ait}^2)E(\varepsilon_{bit}^2)E(\varepsilon_{cit}^2)}{N^2} \le \frac{v_b v_c}{N},$$

and hence $\sum_i\theta_{ait}\varepsilon_{bit}\varepsilon_{cit}/N$ also converges in probability to 0.

The only remaining term in (C10) is that appearing in the lower-right hand corner, where for a and b equal to 1 or 2, we have

$$\text{(C15)}\quad \sum_i \frac{\theta_{ait}\varepsilon_{bit}^2}{N} = \sum_i \frac{\theta_{ait}}{N}\sum_i \frac{w_i\varepsilon_{bit}^2}{W_N},\ \text{ where } w_i = \theta_{ait}\ \&\ W_N = \sum_i \theta_{ait}.$$

From Jamison, Orey & Pruitt's (1965) Theorem 1 on the convergence of weighted averages of independent variables, we know that

$$\text{(C16)}\quad \text{If and only if (a)} \lim_{N\to\infty} W_N = \sum_i \theta_{ait} = \infty,\ \text{(b)} \lim_{N\to\infty} w_N/W_N = 0,$$

$$\text{(c)} \lim_{T\to\infty} T*\text{Prob}[\varepsilon_{bit}^2 \ge T] = 0,\ \&\ \text{(d)} \lim_{T\to\infty} \int_{\varepsilon_{bit}^2<T} \varepsilon_{bit}^2 dF \text{ exists, then } \sum_i \frac{w_{it}\varepsilon_{bit}^2}{W_N} \xrightarrow{p} E(\varepsilon_{bit}^2),$$

where F is the cumulative distribution function of ε_{bit}^2. Condition (a) is assured, as if it does not hold (say) for $a = 1$, as $\theta_{2it} = 1 - \theta_{1it}$ we can simply switch to the "or" version of (C.10) and ensure that it does. As $\theta_{ait} \le 1$, condition (b) follows. If for some $\gamma > 0$ $E((\varepsilon_{bit}^2)^\gamma)$ exists, as is true for the normal, then (c) follows from Markov's Inequality and (d) as well.[2] Consequently, we have:

$$\text{(C17)}\quad \sum_i \frac{\theta_{ait}\varepsilon_{bit}^2}{N\hat{v}_b} - \sum_i \frac{\theta_{ait}\varepsilon_{ait}^2}{N\hat{v}_a} = \underbrace{\sum_i \frac{\theta_{ait}}{N}}_{\le 1}\left(\frac{1}{\hat{v}_b}\sum_i \frac{w_i\varepsilon_{bit}^2}{W_N} - \frac{1}{\hat{v}_a}\sum_i \frac{w_i\varepsilon_{ait}^2}{W_N}\right) \xrightarrow{p} 0.$$

From the above results, we see that depending upon whether we follow the second or third equality in (C10):

[2]Jamison, Orey & Pruitt treat w_i as non-stochastic, but their proof follows through provided it is independent of the weighted variable and conditions (a) and (b) are always satisfied, as is true here.

$$\text{(C18)}\quad \frac{\partial \bar{L}(\sigma)}{\partial \hat{\sigma}} \overset{p}{\to} \frac{1}{\sigma\text{-}1}\left\{-1+\sum_t \frac{1}{T}\frac{v_1}{v_1}\right\}=0 \text{ or } \frac{\partial \bar{L}(\sigma)}{\partial \hat{\sigma}} \overset{p}{\to} \frac{1}{\sigma\text{-}1}\left\{-1+\sum_t \frac{1}{T}\frac{v_2}{v_2}\right\}=0,$$

as was claimed earlier.

References

Jamison, Benton, Steven Orey & William Pruitt (1965). "Convergence of Weighted Averages of Independent Random Variables." Z. Wahrscheinlichkeitstheorie 4: 40-44.

D. Monte Carlos for Johansen Type Asymptotically Valid Cointegration Tests

This appendix shows that Johansen's (1995) cointegration (trace) test and Larsson, Lyhagen & Löthgren's (2001) extension to panel data have very large size distortions when evaluated using their asymptotic distributions. I apply the tests to data generating processes based upon my industry x year panels for K/L, I/L and TFP growth. For the test of whether any co-integration exists at all (in a single industry *i*), Johansen compares the likelihoods of vector auto-regression and full rank cointegration models:

$$\text{(D1)}\quad \mathbf{g}(y_{jit}) = \mathbf{\Gamma}_i \mathbf{g}(y_{jit-1}) + \mathbf{z}_{it}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_{it} \quad \text{(VAR)}$$

$$\mathbf{g}(y_{jit}) = \mathbf{\Gamma}_i \mathbf{g}(y_{jit-1}) + \mathbf{\Pi}_i \,\mathbf{ln}(y_{jit-1}) + \mathbf{z}_{it}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_{it} \quad \text{(VEC)}$$

where *j* denotes any of the 3 measures, $\mathbf{\Gamma}$ and $\mathbf{\Pi}$ are 3x3 matrices, $\mathbf{z}_{it}$ either constants for each *j* or constants plus time trends, and $\boldsymbol{\varepsilon}_{it}$ is a Jx1 vector of errors. Johansen's test is designed for a single time series, so below I test each time series separately and report the average rejection rate. Larsson et al (2001) extend the test to panel data by using the normalized mean of the individual time series likelihood ratios.[3]

I begin by estimating the VAR in (D1), i.e. running a specification that imposes the null. The VAR is run assuming the errors are multivariate normal or (using maximum likelihood techniques) distributed multivariate t. The covariance and (in the case of the t-) degrees of freedom estimates are then used to create new multivariate iid (across *it*) errors, which are added to the point estimates to create new data. Table D1 below reports the average empirical rejection rates of the true null of no cointegration across 500 draws of new data. As shown, size distortions in Johansen's test are large, with the empirical rejection probability of the true null at the nominal .01 and .05 levels varying from .060 to .165 & .192 to .294, respectively, with rejection rates systematically higher when both the data generating process and

[3]Both methods also allow for testing rank > 0 levels of cointegration, but I concentrate here on results for the test of rank = 0, i.e. is there any cointegration at all.

Table D1: Monte Carlo Rejection Probabilities of True Null
of No Cointegration among ln(TFP), ln(K/L), ln(I/L)

	(a) Johansen's (1995) trace test conducted industry by industry				(b) Larsson et al (2001) panel normalized mean trace test			
	constant		+ trend		constant		+ trend	
	.01	.05	.01	.05	.01	.05	.01	.05
normal errors	.060	.192	.096	.252	.670	.728	.830	.860
t-distributed errors	.143	.252	.165	.294	.604	.644	.752	.780

Notes: Reported Johansen rejection rates are the average of the 61 industry level rejection rates for 500 Monte Carlos each, while Larsson et al rejection rates are based on 500 Monte Carlos for all industries together. Test statistics evaluated using asymptotic critical values, means & variances calculated by Osterwald-Lenum (1992).

test include time trends. Rejection rates for Larsson et al's panel normalized mean trace test are remarkable, ranging from .604 to .830 and .644 to .860 at the .01 and .05 levels, respectively

Neither of these tests is suitable for testing cointegration in my panel model, where the dependent variables are transformations of implicit variables which may be cointegrated and the model includes additional non-standard parameters such as elasticities of substitution. The Monte Carlos in Table D1 suggest, however, that asymptotic theory for my panel model is unlikely to be of much use in the sample sizes encountered in the paper. For that reason, I make use of the wild bootstrap for inference, showing in the next appendix that when estimation is done using the multivariate-t distribution it consistently delivers reasonably accurate null rejection probabilities.

References

Johansen, Soren (1995). Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press.

Larrson, Rolf, Johan Lyhagen, & Mickael Löthgren (2001). "Likelihood-based cointegration tests in heterogeneous panels". *Econometrics Journal* 4: 109-142.

Osterwald-Lenum, Michael (1992). "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics." *Oxford Bulletin of Economics and Statistics* 54: 461-472.

E. Structural Model Cointegration Tests using the Wild Bootstrap

This appendix reports Monte Carlos for the wild bootstrap-based cointegration tests I implement for the structural models in the paper. Panel cointegration tests based upon asymptotic distributions, examined in the appendix above, have nominal .01 level empirical rejection rates in my sample sizes of over .6 (Table D1). Moreover, they do not cover the models used in the paper, where we are testing the level of cointegration of an underlying latent variable that is unobserved but related to observables through matrices determined by observed factor input shares and estimated elasticities of substitution. For these reasons, I implement a wild bootstrap based upon transformations of the residuals of the null hypothesis. Although by no means perfect, when implemented using a t-distributed likelihood, which underweights outliers, these tests afford rejection probabilities that are much closer to nominal value than found using asymptotic theory. As in the cointegration tests described in appendix D, I follow Johansen (1995) and use the difference in the ln likelihood of the model of full cointegration (i.e. the number of cointegrating relations equaling the number of dependent variables) and that of cointegration of level r as the test statistic.

I begin each test by estimating the model of the null hypothsis. As in the paper, we have:

$$\text{(E1)}\ \mathbf{g}(y_{jit}) = \begin{bmatrix} \mathbf{g}(X_{jit}/X_{\mathrm{J}it}) \\ g(TFP_{it}) \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{E}_{it}^{-1} \\ \mathbf{0}'_{\mathrm{J}-1} \end{bmatrix}}_{\mathbf{B}_{it}} \mathbf{g}(p_{jt}/p_{\mathrm{J}t}) + \underbrace{\begin{bmatrix} -\mathbf{E}_{it}^{-1} - \mathbf{I}_{\mathrm{J}-1}, \mathbf{E}_{it}^{-1}\mathbf{i}_{\mathrm{J}-1} + \mathbf{i}_{\mathrm{J}-1} \\ \boldsymbol{\theta}'_{it} \end{bmatrix}}_{\mathbf{C}_{it}} \mathbf{g}(A_{jit}),$$

so that the unobserved latent variables are related to the observables by $\mathbf{g}(A_{jit}) = \mathbf{C}_{it}^{-1}[\mathbf{g}(y_{jit}) - \mathbf{B}_{it}\mathbf{g}(p_{jt}/p_{\mathrm{J}t})]$. The matrix $\mathbf{E}_{it}^{-1}$ depends upon factor shares and elasticities of substitution. In the case of the three factor model of the paper, for instance, it is given by

$$\text{(E2)}\ \ \mathbf{E}_{it}^{-1} = \frac{1}{\theta_{1it} + \theta_{2it}} \begin{bmatrix} -\sigma\theta_{2it} - \eta\theta_{1it} & (\sigma - \eta)\theta_{2it} \\ (\sigma - \eta)\theta_{1it} & -\sigma\theta_{1it} - \eta\theta_{2it} \end{bmatrix}.$$

Estimated elasticities of substitution uncover the latent factor augmenting change, which is modelled as having either a VAR or VEC form with factor x year & factor x time fixed effects and mutually orthogonal iid shocks:

$$\text{(E3a)}\ \ \text{VAR}: \mathbf{g}(A_{jit}) = \boldsymbol{\Gamma}\mathbf{g}(A_{jit-1}) + \boldsymbol{\eta}_t + \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{it}$$

$$\text{(E3b)}\ \ \text{VEC}: \mathbf{g}(A_{jit}) = \boldsymbol{\Gamma}\mathbf{g}(A_{jit-1}) + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{ln}(A_{jit-1}) + \boldsymbol{\eta}_t + \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{it}.$$

The baseline null hypothesis is the VAR or a VEC model with r cointegrating equations. I then either multiply the individual estimated residuals $\hat{\varepsilon}_{jit}$ by an iid variable that is ±1 with 50/50 probabilities, generating mutually orthogonal iid shocks, or I multiply the observation specific vector of residuals $\hat{\boldsymbol{\varepsilon}}_{it}$ by a common ±1, retaining the empirical off-diagonal covariance of the iid vectors of shocks. These are then added to estimated fixed effects $\hat{\boldsymbol{\eta}}_t$ & $\hat{\boldsymbol{\eta}}_i$ and cumulated forward using the estimated lag & cointegration parameters $\hat{\boldsymbol{\Gamma}}$, $\hat{\boldsymbol{\alpha}}$ & $\hat{\boldsymbol{\beta}}$. "Pre-sample" values of $\mathbf{ln}(A_{ji0})$ and $\mathbf{g}(A_{ji0})$ do not change and are set at those used in the original estimation, i.e. **0** and the value uncovered by $\hat{\mathbf{C}}_{i0}^{-1}[\mathbf{g}(y_{ji0}) - \mathbf{B}_{i0}\mathbf{g}(p_{j0}/p_{\mathrm{J}0})]$, respectively. The bootstrapped latent variables are then transformed into observed data $\mathbf{g}(y_{jit})$ using (E1) above and the estimated values (based on estimated elasticities of substitution) of $\hat{\mathbf{E}}_{it}^{-1}$. Original factor input shares are kept as fixed data/regressors. The ln likelihoods for this new data of a model of full cointegration and the restricted model of the null hypothesis are then calculated and their difference compared to that found in the original data. In the paper I use 200 bootstrap draws to calculate the probability of a test statistic greater than that observed under the null.

The above describes the wild bootstrap test procedure. To evaluate the accuracy of the procedure, I need artificial data that is similar to my sample. To that end, I similarly begin by taking point estimates for each model and generating artificial data. Instead of using actual residuals, however, I use normally or t-distributed errors with either an orthogonal covariance matrix or unrestricted covariance matrix and degrees of freedom equal to that estimated on the residuals of the original sample. I then estimate the model on this artificially generated data and use estimated residuals to conduct the wild bootstrap using 200 data bootstrap iterations. Thus, for a test of r = 0 cointegration (i.e. a VAR) the Monte Carlo procedure is:

(1) Estimate VAR model parameters on original data.
(2) Use those VAR parameters to generate data that has normally or t-distributed iid mutually orthogonal or correlated error vectors.
(3) Estimate the model on the data in (2) and calculate the VAR vs full cointegration test statistic for the data.
(4) Use the newly estimated VAR parameters and residuals to implement the wild bootstrap, using 200 iterations to calculate the p-value of the test statistic calculated in (3).

For the VEC1 model, following Johansen's (1995) sequential testing procedure, I:

(1) Estimate VEC1 model parameters on original data.
(2) Use those parameters to generate cointegrated data that has normally or t-distributed iid mutually orthogonal or correlated error vectors.
(3) Estimate the VAR model on the data in (2) and calculate the VAR vs full cointegration test statistic for the data.
(4) Use the newly estimated VAR parameters and residuals to implement the wild bootstrap, using 200 iterations to calculate the p-value of the test statistic calculated in (3).
(5) If rejecting at the nominal level in (4), estimate the VEC1 model on the data in (2 and calculate the VEC1 vs full cointegration test statistic for the data.
(6) Use the newly estimated VEC1 parameters and residuals to implement the wild bootstrap, using 200 iterations to calculate the p-value of the test statistic calculated in (5).

I refer to each run through of (1)-(4) or (1)-(6) as a "sample".

Table E1 reports the Monte Carlo rejection probabilities of true null at nominal levels .01, .02 & .05. I run 10 samples for each of the 12 models examined in Section IV of the paper. In a few samples the likelihood fails to converge in steps (3) or (5) above, and those are excluded from the calculations, as are wild bootstrap runs which fail to converge in steps (4) or (6) above. In testing accuracy for a true VAR,

Table E1: Monte Carlo Empirical Rejection Probabilities of True Nulls Regarding Cointegration among Factor Augmenting Productivities by Level of Test

cointegration in *dgp*	data generating process	likelihood model	nominal level .01	.02	.05
0 (VAR)	t-distributed	t-distributed	.025	.042	.119
1 (VEC1)	diagonal shocks	diagonal covariance	.026	.026	.111
0 (VAR)	t-distributed	t-distributed	.050	.092	.158
1 (VEC1)	correlated shocks	diagonal covariance	.017	.043	.111
0 (VAR)	t-distributed	t-distributed	.025	.068	.136
1 (VEC1)	diagonal shocks	unrestricted covariance	.026	.034	.077
0 (VAR)	t-distributed	t-distributed	.050	.075	.133
1 (VEC1)	correlated shocks	unrestricted covariance	.051	.060	.085
0 (VAR)	normally distributed	normally distributed	.188	.265	.444
1 (VEC1)	diagonal shocks	diagonal covariance	.017	.034	.076
0 (VAR)	t-distributed	normally distributed	.283	.317	.492
1 (VEC1)	diagonal shocks	diagonal covariance	.235	.193	.151

Notes: Test rejects any given null when the fraction of wild bootstrap iterations with a test statistic greater than that of the sample is less than or equal to the nominal level. Test for VEC1 *dgp* follows Johansen's sequential procedure, testing VAR null and then testing VEC1 null if VAR is rejected. Reported rejection probability is 1 – probability of accepting the true null in the sequential procedure.

the table reports the probability the wild bootstrap selected in favour of r > 0 cointegration. In testing accuracy for a true VEC1, the table reports the probability the wild bootstrap accepted the null of no cointegration (VAR) or rejected that null but subsequently also rejected the VEC1 null in favour of r > 1 cointegration, as this represents the error rate of Johansen's sequential testing procedure. It is possible for empirical rejection rates at a higher nominal level to be lower in testing a true VEC1 because of a reduced failure to reject in the first step test of a VAR.

As can be seen, with a normal likelihood and either normal or heavy tailed t - distributed data, the wild bootstrap performs poorly and inconsistently. However, results using a t-likelihood (which underweights outliers) on heavy-tailed data, as encountered in the paper, are much more reliable. In particular, a .01 nominal level ensures a rejection probability not greater than .051 even when the wild bootstrap *dgp* is misspecified, imposing a diagonal covariance when the *dgp* is non-diagonal or allowing off-diagonal covariance when the *dgp* is diagonal. For this reason, I use a nominal level of .01 to evaluate the t-distribution Johansen sequential test for the actual sample in the paper.

F. Supplementary Table for Section IV

Standard errors in panel (A) of Table 7 of the paper are based upon a wild bootstrap that imposes a diagonal covariance matrix by multiplying each factor

Table F1: Ratio of Wild Bootstrap Standard Error Estimates for Table 7(A):
Empirical Off-Diagonal Covariance of Shocks/Imposing Diagonal Covariance Matrix

	β_L	β_I	α_K	α_L	α_I	ρ_{KL}	ρ_{KI}	ρ_{LI}
1:	1.19	1.23	1.03	0.96	0.99	0.96	0.90	0.84
2:	0.86	0.88	1.01	0.96	0.78	0.98	1.06	0.97
3:	0.89	1.10	0.95	0.91	0.84	0.98	0.96	1.03
4:	2.00	1.51	0.86	0.76	0.86	1.02	1.02	1.19
5:	1.64	1.50	1.04	1.13	0.95	0.99	0.94	0.94
6:	4.70	4.42	1.28	1.35	1.23	0.91	0.96	0.98
7:	1.10	1.25	1.02	1.34	1.24	1.23	0.96	1.20
8:	0.58	0.55	1.25	1.11	1.10	0.99	0.95	1.22
9:	1.64	1.33	1.07	1.26	1.06	0.93	0.93	0.95
10:	1.17	1.26	1.15	1.24	1.14	0.90	0.93	1.12
11:	1.43	1.26	1.03	1.04	0.97	0.89	0.87	0.84
12:	0.78	0.70	1.05	1.98	1.19	0.91	0.98	1,12

Notes: Variables as defined in the paper. 1-12 denotes models described in Table 5 of the paper.

augmenting shock by an independent ±1 variable. The table notes indicates thatstandard error estimates based upon a bootstrap that retains the off-diagonal covariance of the shocks by multiplying each *it* three-tuple of shocks by a common ±1 are very similar. Table F1 reports the ratio of the standard error estimate using the wild bootstrap that retains the off-diagonal covariance to that of the wild bootstrap that imposes a diagonal covariance matrix. As shown, most of the ratios are near one.

G Estimation using the Multivariate t as Equivalent to Weighted Estimation

This appendix shows that estimation using the multivariate-t produces the same parameter estimates as a weighted version of the multivariate normal, where the weights are determined by the degrees of freedom and the inverse covariance matrix weighted deviations of the observation residuals from predicted values. I illustrate this with a seemingly unrelated system of equations with the same regressors used for each dependent variable, although the same results apply to the non-linear VEC models estimated in the paper.

The ln likelihood for the multivariate normal is:

$$\text{(G1)}\quad \ln L = \sum_{it}\left[-\frac{J}{2}\ln(2\pi) - \frac{1}{2}\ln(\det(\mathbf{V})) - \frac{1}{2}(\mathbf{y}'_{it} - \mathbf{x}'_{it}\boldsymbol{\beta})\mathbf{V}^{-1}(\mathbf{y}_{it} - \boldsymbol{\beta}'\mathbf{x}_{it})\right],$$

where subscript *it* denotes the industry *i* x year *t* observation, Σ_{it} indicates summation across all such observations, $\mathbf{y}_{it}$ is a Jx1 vector of dependent variables, $\mathbf{x}_{it}$ a Kx1 vector of regressors, and $\mathbf{V}$ JxJ (symmetric) & $\boldsymbol{\beta}$ KxJ matrices of parameters, respectively. The first order conditions for maximizing the ln likelihood are:

$$\text{(G2)}\quad \frac{\partial \ln L}{\partial v_{jk}} = 0 \Rightarrow \sum_{it} [-(\mathbf{V}^{-1})_{jk} + (\mathbf{y}'_{it} - \mathbf{x}'_{it}\boldsymbol{\beta})(\mathbf{V}^{-1})_j(\mathbf{V}^{-1})'_k(\mathbf{y}_{it} - \boldsymbol{\beta}'\mathbf{x}_{it})] = 0$$

$$\frac{\partial \ln L}{\partial \beta_{jk}} = 0 \Rightarrow \sum_{it} x_{jit}(\mathbf{V}^{-1})_k(\mathbf{y}_{it} - \boldsymbol{\beta}'\mathbf{x}_{it}) = 0$$

where $(\mathbf{V}^{-1})_k$ and $(\mathbf{V}^{-1})_{jk}$ denote the k^{th} column and jk^{th} element of $\mathbf{V}^{-1}$. The solutions to these equations are given by:

$$\text{(G3)}\quad \hat{\boldsymbol{\beta}} = \left(\sum_{it} \mathbf{x}_{it}\mathbf{x}'_{it}\right)^{-1} \sum_{it} \mathbf{x}_{it}\mathbf{y}'_{it} \;\;\&\;\; \hat{\mathbf{V}} = \frac{1}{NT}\sum_{it} (\mathbf{y}_{it} - \hat{\boldsymbol{\beta}}'\mathbf{x}_{it})(\mathbf{y}'_{it} - \mathbf{x}'_{it}\hat{\boldsymbol{\beta}})$$

When $\mathbf{V}$ is constrained to be diagonal, the solution is the same except that we set the off-diagonal elements of $\hat{\mathbf{V}}$ to 0.

The corresponding ln likelihood for the multivariate t is

$$\text{(G4)}\quad \ln L = \sum_{it} \left[\begin{array}{c} -\frac{J}{2}\ln(2\pi) + \frac{\tau}{2}\ln(\tau) - .5\ln(\det(\mathbf{V})) + \ln\Gamma\left(\frac{\tau+J}{2}\right) - \ln\Gamma\left(\frac{\tau}{2}\right) \\ -\left(\frac{\tau+J}{2}\right)\ln[\tau + (\mathbf{y}'_{it} - \mathbf{x}'_{it}\boldsymbol{\beta})\mathbf{V}^{-1}(\mathbf{y}_{it} - \boldsymbol{\beta}'\mathbf{x}_{it})] \end{array} \right].$$

where τ is the degrees of freedom & Γ is the gamma function. Taking τ momentarily as given, we have the first order conditions

$$\text{(G5)}\quad \frac{\partial \ln L}{\partial v_{jk}} = 0 \Rightarrow \sum_{it} \left[-(\mathbf{V}^{-1})_{jk} + (\tau + J)\frac{(\mathbf{y}'_{it} - \mathbf{x}'_{it}\boldsymbol{\beta})(\mathbf{V}^{-1})_j(\mathbf{V}^{-1})'_k(\mathbf{y}_{it} - \boldsymbol{\beta}'\mathbf{x}_{it})}{\tau + (\mathbf{y}'_{it} - \mathbf{x}'_{it}\boldsymbol{\beta})\mathbf{V}^{-1}(\mathbf{y}_{it} - \boldsymbol{\beta}'\mathbf{x}_{it})} \right] = 0$$

$$\frac{\partial \ln L}{\partial \beta_{jk}} = 0 \Rightarrow \sum_{it} \frac{x_{jit}(\mathbf{V}^{-1})_k(\mathbf{y}_{it} - \boldsymbol{\beta}'\mathbf{x}_{it})}{\tau + (\mathbf{y}'_{it} - \mathbf{x}'_{it}\boldsymbol{\beta})\mathbf{V}^{-1}(\mathbf{y}_{it} - \boldsymbol{\beta}'\mathbf{x}_{it})} = 0$$

and solutions

$$\text{(G6)}\quad \hat{\boldsymbol{\beta}} = \left(\sum_{it} w_{it}\mathbf{x}_{it}\mathbf{x}'_{it}\right)^{-1} \sum_{it} w_{it}\mathbf{x}_{it}\mathbf{y}'_{it} \;\;\&\;\; \hat{\mathbf{V}} = \frac{1}{NT}\sum_{it} w_{it}(\mathbf{y}_{it} - \hat{\boldsymbol{\beta}}'\mathbf{x}_{it})(\mathbf{y}'_{it} - \mathbf{x}'_{it}\hat{\boldsymbol{\beta}}),$$

$$\text{where}\;\; w_{it} = \frac{(\tau + J)}{\tau + (\mathbf{y}'_{it} - \mathbf{x}'_{it}\hat{\boldsymbol{\beta}})\hat{\mathbf{V}}^{-1}(\mathbf{y}_{it} - \hat{\boldsymbol{\beta}}'\mathbf{x}_{it})} \;\;\&\;\; \sum_{it} w_{it} = NT,$$

and where again if $\mathbf{V}$ is constrained to be diagonal we set the off-diagonal terms of $\hat{\mathbf{V}}$ to 0. Thus, for a given τ the solution for the other parameters looks like a weighted version of the standard solution for multivariate normal errors, where the weights $\tau + J$ divided by the degrees of freedom plus the inverse covariance matrix weighted deviation from means. As τ goes to infinity and the distribution converges to the multivariate normal, these weights converge to 1. Given these solutions, the likelihood can then be maximized with respect to τ.