The GMM Parameter Normalization Puzzle

Charles A. Fleischman*

Division of Research and Statistics
Federal Reserve Board

September 1997

Abstract:

A feature of GMM estimation--the use of a consistent estimate of the optimal weighting matrix rather than the joint estimation of the model parameters and the weighting matrix--can lead to the sensitivity of GMM estimation to the choice of parameter normalization. In many applications, including Euler equation estimation, a model parameter multiplies the equation error in some, but not all, normalizations. But, conventional GMM estimators that either hold the estimate of the weighting matrix fixed or allow some limited iteration on the weighting matrix fail to account for the dependence of the weighting matrix on the parameter vector implied by the multiplication of the error by the parameter. In finite samples, GMM effectively minimizes the square of the parameter times the objective function that obtains from an alternative normalization where no parameter multiplies the equation error, resulting in estimates that are smaller (in absolute value) than those from the alternative normalization. Of course, normalization is irrelevant asymptotically.

Journal of Economic Literature Classifications: C13, C51, E24, C81

---

*This paper is a revised version of Chapter 4 of my PhD dissertation at the University of Michigan. I would like to thank Bob Barsky, Susanto Basu, John Fernald, Jonathan Parker, Shinichi Sakata, Scott Schuh, Dan Sichel, Karl Whelan, and, especially, Matthew Shapiro and Spencer Krane for extremely helpful conversations and comments on earlier drafts, and Jan Kmenta for always asking, "What is the error?" I would also like to thank Matthew Shapiro for making data available to me and the Bureau of Labor Statistics for providing me with unpublished data from the establishment survey. All errors (structural or not) are my responsibility. The views expressed in this paper do not reflect those of the Federal Reserve Board or any of its staff. Contact information: Mailstop 80, Federal Reserve Board, Washington, DC 20551; (202) 452-6473. Email: cfleischman@frb.gov

I. Introduction

Generalized method of moments (GMM) techniques have been used to estimate the structural parameters of the Euler equations (first-order conditions) from a wide variety of dynamic rational expectations models.[1] However, the well-documented finite-sample sensitivity of GMM estimation to changes in parameter normalization has raised important issues about the usefulness of GMM estimators. Recent work by Krane and Braun (1991), West and Wilcox (1993), and Fuhrer, Moore, and Schuh (1995) has highlighted the economically important sensitivity of GMM estimates of the parameters of the linear-quadratic inventory model to the choice of the normalization of the Euler equation.[2] Yet, GMM estimation remains popular for its ability to provide asymptotically efficient estimates in nonlinear models with assumptions only concerning instrument orthogonality. And, of course, normalization is irrelevant asymptotically; GMM estimators that differ only in their choice of parameter normalization have identical asymptotic distributions.[3]

Because normalization is irrelevant asymptotically, the key to the normalization sensitivity must be related to finite-sample estimation procedures and not to the formation of the theoretical GMM objective function. Hausman (1975) shows that an important difference between full information maximum likelihood (FIML) and three-stage-least-squares, which is the GMM estimator under the assumptions that the residuals are homoskedastic and serially uncorrelated, is that FIML iterates simultaneously on the parameters and covariance matrix so that "the instruments used by FIML are mutually consistent with the parameter estimates *in the given sample*, while for other estimators the instruments are consistent with the parameter estimates only asymptotically. (p. 727)" Hansen, Heaton, and Yaron (HHY, 1996) differentiate three classes of GMM estimators based on the treatment of the weighting matrix: An estimator that holds fixed an initial consistent estimate of the weighting matrix as the selection of parameter values iterates towards convergence; an estimator that iterates on the

---

[1]Recent examples include estimates of the linear-quadratic inventory model (e.g. West (1986), Ramey (1991), Krane and Braun (1991), West and Wilcox (1993), and Fuhrer, Moore, and Schuh (1995)), models of convex adjustment costs for capital and labor (e.g. Shapiro (1986a, 1986b), Burgess and Dolado (1989), Burda (1991), Pfann and Palm (1993), Oliner, Rudebusch, and Sichel (1996), Sbordone (1996), Pfann (1996), Fleischman (1996), and Basu and Kimball (1997)), intertemporal consumption models (e.g. Hansen and Singleton (1982)), and price-adjustment models (e.g. Roberts, Stockton, and Struckmeyer (1994)).

[2]See Bartelsman (1995) for a discussion of the sensitivity to changes in equation normalization of Hall's (1988, 1990) two-stage-least-squares estimates of the price-markup ratio and returns-to scale.

[3]In finite samples, normalization is also irrelevant when the parameters are just identified.

weighting matrix but not at every change in parameter values; and an estimator that continuously updates its estimate of the weighting matrix to reflect the current parameter values. HHY show that the GMM estimator that continuously updates its estimate of the weighting matrix to reflect the current parameter values is immune from the normalization sensitivity because, like maximum likelihood estimators, this GMM estimator accounts explicitly for the dependence of the GMM weighting matrix on the model parameters.[4] However, the GMM implementations in popular econometric packages, such as RATS and TSP, do not include a continuously updating estimator.

In this paper I identify how a feature of most conventional GMM estimators--the use of a consistent estimate of the optimal weighting matrix rather than the joint estimation of the model parameters and the weighting matrix--can explain the GMM parameter normalization puzzle. Specifically, conventional GMM estimators that either hold the estimate of the weighting matrix fixed or allow some limited iteration on the weighting matrix fail to account for the dependence of the weighting matrix on the parameter vector. In finite samples, two GMM estimators that differ only in the parameter normalization of the estimating equation minimize two different objective functions. Using a simple, linear Euler equation as an example, I derive the orthogonality conditions, weighting matrices, objective functions, and first-order conditions for a pair of GMM estimators that differ only in the choice of parameter normalization. I show that in the more "natural" normalization plus an expectations error--the one that obtains when an unobserved expectation in the Euler equation is replaced by its *ex poste* realization--a model parameter multiplies the rational expectations error. This normalization implies restrictions on the weighting matrix that cannot be imposed by the GMM estimators typically used. For limited class of dynamic rational expectations models--those that include only one unobserved expectation--there is an alternative normalization where no model parameter multiplies the rational expectations error. This alternative normalization implies no restrictions on the weighting matrix.

In the first normalization, the failure to parameterize the weighting matrix implies that the GMM estimator effectively minimizes, in part, the square of the parameter that multiplies the rational expectations error. This introduces an extra term in the first-order conditions for which the GMM estimator is the solution; of course, this extra term has a plim of zero and thus vanishes

---

[4]HHY argue, however, that there is no particular advantage to using the continuously updating estimator to obtain point estimates; the elimination of the normalization sensitivity is, in their opinion, at least offset by the thickening of the tails of the distribution of the estimator. They state that "(i)n this sense continuous updating sometimes inherits the defects of maximum likelihood estimators relative to two-stage least squares estimators in the classical simultaneous-equations environment." (p. 278)

asymptotically, preserving the asymptotic irrelevance of the choice of normalization. But in finite samples, the estimate of the parameter that multiplies the rational expectations error will be smaller (in absolute value) than corresponding estimate from the alternative normalization. In the alternative normalization, the finite-sample GMM objective function is similar in structure to the theoretical GMM objective function because no parameters multiply the rational expectations error. I argue, therefore, that the estimator that implies no restrictions on the weighting matrix is preferable, at least on theoretical grounds.

Empirically, the predictions about the relationship between the estimators are borne out. I compare GMM estimates of the costs of adjusting production worker employment for more than 100 four-digit SIC manufacturing industries from two different normalizations of a cost minimization condition derived from the Euler equations for the rational expectations dynamic cost minimization model from Fleischman (1996).[5] The results are quite sensitive to the choice of normalization.[6] Estimates of adjustment costs from the more natural normalization are economically trivial and imply a median half-life of adjustment of 0.5 month. In contrast, GMM estimates of the alternative normalization where no parameter multiplies the expectations error are larger (often by more than an order of magnitude) and imply a median half-life of employment adjustment of 2.0 months, which is comparable to estimates previously found in the literature.[7]

I reestimate Fleischman's labor adjustment cost model using Quasi Maximum Likelihood (QML) and compare the QML estimates to the two sets of GMM estimates. The QML estimator is a useful benchmark for choosing between the GMM estimators because the QML estimates are consistent and use the same orthogonality conditions as the GMM estimates.[8] Most importantly, the QML estimator is invariant to renormalization because the likelihood function accounts for any changes to the score and covariance matrices caused by reparameterization of the model. The QML

---

[5]There is no closed-form solution to the adjustment cost model. As an alternative to Monte-Carlo simulations, I provide estimates of the model for a large number of four-digit industries.

[6]Similarly, I find normalization sensitivity in GMM estimates of the costs of adjusting nonproduction worker employment from a single equation implementation of Shapiro's (1986) model of dynamic factor demand.

[7]See Hamermesh (1993) for a summary and synthesis of the literature estimating the costs of adjusting employment and the speed of employment adjustments. Implied estimates (by the parameters estimated in papers using different methodologies) of the half-life of employment adjustment are clustered towards the lower end of the range from 1.5 months in Bernanke (1986) to 10 months in Kennan (1988).

[8]The QML estimator is similar conceptually to the continuously updating GMM estimator that HHY discuss. In addition, if the residuals are normally distributed, the QML estimator is the limited information maximum-likelihood (LIML) estimator, and thus efficient among the class of limited-information estimators.

estimates are similar both quantitatively and economically to the larger of the two sets of GMM estimates, which I argue provides further empirical support in favor of GMM estimation of the alternative normalization.

Following this introduction, the rest of the paper is organized as follows. In section II, I derive two alternative GMM estimators of a simple, linear Euler equation and use this example to demonstrate how the failure to parameterize the weighting matrix explains the normalization sensitivity. In section III, I demonstrate this point empirically. I discuss the GMM estimation of two normalizations of the employment adjustment cost model from Fleischman (1996) and provide two sets of GMM estimates for individual four-digit SIC manufacturing industries. I show that the GMM estimates of adjustment costs from the alternative normalization are always larger than those from the natural normalization, and, thus, more economically plausible. In section IV, I reestimate the model using QML and find that the QML estimates of adjustment costs are economically similar to the GMM estimates from the alternative normalization. Finally, in Section V, I summarize the results and point out useful directions for future work. In Appendix A, I briefly describe the dynamic labor demand model from which I obtain the Euler equations estimated in Sections III and IV.

II. Finite-Sample Implications of the Failure to Parameterize the GMM Weighting Matrix

Econometric packages commonly used by applied macroeconometricians--such as TSP and RATS--implement GMM estimators that are subject to normalization sensitivity because these estimators either hold constant an initial estimate of the weighting matrix or allow iteration on the weighting matrix that does not fully account for its structure. Thus, to evaluate empirical work, it is important to try to better understand the source of the normalization sensitivity. In this section, I demonstrate how the failure of common GMM estimators to parameterize their weighting matrixes leads to the well-documented finite-sample sensitivity of GMM estimates to parameter normalization.[9]

I begin with a general specification of a linear Euler equation from a dynamic rational expectations model:

$$Y_t = E_t[Y_{t+1}]\,\beta$$

[9]I consider only the case of a single unobserved expectation, or, more generally, of a single dominant source of equation error, because a normalization where no parameter multiplies an expectation error exists only for this limited class of estimating equations. However, this simplification is not necessary to demonstrate that the failure to parameterize the weighting matrix is the source of the GMM normalization sensitivity.

where Y_t is an endogenous variable, β is a single unknown parameter, and $E_t[\ ]$ is the expectations operator conditional on information available in period t or earlier.

I will derive two GMM estimators of β that differ only in the choice of parameter normalization of the Euler equation. Note that $E_t[Y_{t+1}]$ is unobservable and that $Y_{t+1} = E_t[Y_{t+1}] + \epsilon_{t+1}$, where ϵ_{t+1} is a rational expectations error orthogonal to information available in period t or earlier. I obtain the first normalization by following the standard practice and replacing the unobserved conditional expectation, $E_t[Y_{t+1}]$, with the *ex poste* realization and the expectations error to obtain:

$$Y_t = Y_{t+1}\beta - \epsilon_{t+1}\beta \tag{2.1}$$

In this representation of the Euler equation, the model parameters can be consistently estimated by GMM if there are sufficient instruments, Z_t, such that

$$plim \frac{Z_t' \epsilon_{t+1}}{T} = 0 \tag{2.2}$$

A GMM estimator of β sets a weighted average of sample moments as close to their population values as possible, where the moments are based on the orthogonality of the instruments, Z_t, and the rational expectations error, ϵ_{t+1}, implied by equation (2.2). In this normalization--which I refer to as the level normalization because the level of β enters equation (2.1) directly--the finite sample orthogonality conditions are:[10]

$$h_L(Y_{t+1}, Y_t, Z_t; \beta) = \frac{1}{T} Z_t' \epsilon_{t+1} \beta \tag{2.3}$$

Note that β multiplies the rational expectations error in equation (2.1), and thus appears in the orthogonality conditions, equation (2.3). At this point, most practitioners drop the β from the equation error and the orthogonality conditions, replacing it with, for example, $\eta_{t+1} = \beta\epsilon_{t+1}$. This substitution is important because it obscures the fact that β multiplies the rational expectations error. In fact, this may help explain why the presence of a parameter multiplying the rational expectations error had not been identified previously as a source of the parameter normalization puzzle.

In the level normalization, the GMM estimator of β, $\hat{\beta}_L$, minimizes the GMM objective function:

---

[10]Including β in the orthogonality conditions raises a tricky methodological issue. The orthogonality conditions would obviously be satisfied exactly if $\beta = 0$, even if $plim \frac{Z_t' \epsilon_{t+1}}{T} \neq 0$. However, if $\beta = 0$, there is no dynamic model nor is there any equation error. Consequently, if $\beta = 0$, the parameters of the Euler equation are not estimable.

$$G_L(Y_{t+1}, Y_t, Z_t; \beta) = h_L^{\prime} W_L^{-1} h_L \tag{2.4}$$

which is the distance between the sample moments of the orthogonality conditions, h_L, and their population values, where the distance metric weights by the weighting matrix. The optimal weighting matrix is the asymptotic covariance matrix of the orthogonality conditions:

$$W_L = \lim_{T\to\infty} T h_L h_L^{\prime}$$
$$= \lim_{T\to\infty} \frac{1}{T} Z_t^{\prime}(\epsilon_{t+1}\beta)(\epsilon_{t+1}\beta)^{\prime} Z_t$$

Because β is a scalar, the optimal GMM weighting matrix for the level normalization can be rewritten as:

$$W_L = \lim_{T\to\infty} \frac{1}{T}\beta^2 Z_t^{\prime}(\epsilon_{t+1})(\epsilon_{t+1})^{\prime} Z_t \tag{2.5}$$

This normalization implies parameter restrictions on the GMM weighting matrix in equation (2.5). However, a conventional GMM estimator that does not parameterize its weighting matrix cannot impose these restrictions.

By dividing equation (2.1) through by β, I obtain an alternative normalization of the Euler equation:[11]

$$Y_{t+1} = \frac{1}{\beta} Y_t + \epsilon_{t+1} \tag{2.6}$$

where no model parameter multiplies the unobserved expectation. This second normalization--which I refer to as the inverse normalization because $1/\beta$ enters the Euler equation--can be consistently estimated under the same conditions required for the consistent estimation of the level normalization, namely that equation (2.2) holds. Renormalization, of course, does not change the economic interpretation of the model or the orthogonality conditions. The orthogonality conditions for the inverse normalization are:

---

[11]In this normalization, it is clear that β cannot equal 0. Consequently, the parameter vector, β does not have a compact support. However, as I noted above, if β equals 0, there is no equation error, and thus no estimable model. I follow the standard practice in this literature by not addressing this important methodological point here.

$$h_I(Y_{t+1}, Y_t, Z_t; \beta) = \frac{1}{T} Z_t' \epsilon_{t+1} \tag{2.7}$$

and the optimal weighting matrix is

$$W_I = \lim_{T\to\infty} \frac{1}{T} Z_t' \epsilon_{t+1} \epsilon_{t+1}' Z_t \tag{2.8}$$

Because no parameters multiply the rational expectations error in the Euler equation or the orthogonality conditions, the inverse normalization implies no parameter restrictions on the weighting matrix in equation (2.8).

Asymptotically, the GMM objective functions for the level and inverse normalizations are identical:

$$\begin{aligned}
& plim\left[\frac{\beta\epsilon_{t+1}' Z_t}{T} W_L^{-1} \frac{Z_t'\epsilon_{t+1}\beta}{T}\right] = \\
& plim\left[\frac{\beta\epsilon_{t+1}' Z_t}{T} plim\left(\frac{\beta^2 Z_t'\epsilon_{t+1} Z_t'\epsilon_{t+1}}{T}\right)^{-1} \frac{Z_t'\epsilon_{t+1}\beta}{T}\right] = \\
& plim\left[\frac{\epsilon_{t+1}' Z_t}{T} \left(\frac{Z_t'\epsilon_{t+1} Z_t'\epsilon_{t+1}}{T}\right)^{-1} \frac{Z_t'\epsilon_{t+1}}{T}\right] = \\
& plim\left[\frac{\epsilon_{t+1}' Z_t}{T} W_I^{-1} \frac{Z_t'\epsilon_{t+1}}{T}\right]
\end{aligned} \tag{2.9}$$

Renormalization changes the orthogonality conditions and the weighting matrix in such as way as the differences are offset completely in the construction of the population objective function: In the theoretical GMM objective function for the level normalization--shown in the top two lines of equation (2.9)--the β in the orthogonality conditions offset the β in the weighting matrix.

In finite samples, however, GMM estimation begins with a consistent estimate of the weighting matrix obtained from a simpler, yet still consistent estimation procedure.[12] The most common GMM procedures hold the weighting matrix constant while the parameter values iterate to

[12]For example, in the empirical work below I use NL2SLS to obtain an initial estimate of the weighting matrix for NL3SLS, and then I use NL3SLS to obtain a consistent estimator of the weighting matrix for GMM. The difference between the GMM and NL3SLS estimators is that the GMM estimator used in this paper allows different instruments for the two equations. See Hamilton (1994) for a concise description of GMM estimation.

convergence; alternative GMM estimators allow both the parameters and the weighting matrix to be updated at each iteration. Neither GMM procedure, however, allows for parameterization of the weighting matrix.[13] This is not an issue in the inverse normalization where the weighing matrix does not include terms multiplied by the adjustment cost parameter. In this normalization, the GMM estimator minimizes:

$$\underset{\beta}{Min} \; h_I(Y_{t+1}, Y_t, X_t, Z_t; \beta)' \, \hat{W}_I^{-1} \, h_I(Y_{t+1}, Y_t, X_t, Z_t; \beta)$$

where $\hat{W}_I$ is a consistent estimate of W_I. The GMM estimator of β for the level normalization minimizes:

$$\underset{\beta}{Min} \; h_L(Y_{t+1}, Y_t, Z_t; \beta)' \, \hat{W}_L^{-1} \, h_L(Y_{t+1}, Y_t, Z_t; \beta)$$

which, because $h_L = \beta h_I$, is equivalent to minimizing

$$\underset{\beta}{Min} \; \beta^2 h_I(Y_{t+1}, Y_t, Z_t; \beta)' \, \hat{W}_L^{-1} \, h_I(Y_{t+1}, Y_t, Z_t; \beta)$$

where $\hat{W}_L$ is a consistent estimates of W_L. The failure to parameterize the weighting matrix implies that it is not possible to impose the restriction that the β that appear in the asymptotic optimal weighting matrix for the level normalization is the same β that appears in the orthogonality conditions.

Because the GMM estimators most commonly implemented in empirical work do not parameterize the weighting matrix, the estimates do not account for its dependence on β; specifically, unlike in the theoretical objective function, the β in the orthogonality conditions are not cancelled by the β in the weighting matrix. As shown above, in effect, the GMM estimator for the level normalization minimizes β^2 times the objective function that would obtain if the dependence of the weighting matrix on β were parameterized. In finite samples, therefore, the two GMM estimators solve different first-order conditions despite identification by the same orthogonality conditions. The GMM estimator of β in the level normalization, $\hat{\beta}_L$ solves the first-order conditions:

$$2\beta \, h_I' \hat{W}_L^{-1} h_I + \beta^2 \left(\frac{\partial h_I}{\partial \beta} \right)' \hat{W}_L^{-1} h_I \Bigg|_{\beta = \hat{\beta}_L} = 0 \qquad \textbf{(2.10)}$$

while the GMM estimator of β in the inverse normalization, $\hat{\beta}_I$ solves:

---

[13] I limit the discussion to GMM estimators that do not fully iterate on the parameters of the weighting matrix. The continuously updating GMM estimator discussed by Hansen, Heaton, and Yaron (1996) is similar in spirit to the quasi-maximum likelihood estimator that I discuss in section IV.

$$\left(\frac{\partial h_I}{\partial \beta}\right)' \hat{W}_I^{-1} h_I \Big|_{\beta = \hat{\beta}_I} = 0 \qquad \textbf{(2.11)}$$

The extra term in the first-order conditions for the GMM estimator of level normalization--equation (2.10)--arises from the failure to parameterize the weighting matrix. Of course, asymptotically the first-order conditions for the inverse and level normalizations are identical because $plim\ h_I = 0$ and $plim\ \beta^2 \hat{W}_L = plim\ \hat{W}_I$, so this extra term vanishes.

In finite samples, however, the differences between the two first-order conditions imply that GMM estimation of the inverse and level normalizations yield different parameter estimates even when the same instruments are used. To establish the relationship between the estimates of β from the inverse and level normalizations, I must examine two cases: (1) both estimates of β have the same sign and (2) the estimates of β have different signs. The extra term in equation (2.10) must have the same sign as the estimate of β from the level normalization because $h_I' \hat{W}_L^{-1} h_I$ is a weighted sum of squares--implying that it must be non-negative.

For case 1, assume without loss of generality that both estimates of β are positive. Then, the first term in equation (2.10) is positive. This implies that the second term in equation (2.10) is negative. So, with $\hat{W}_L^{-1}$ proportional to $\hat{W}_I^{-1}$, the slope of the objective function for the inverse normalization evaluated at $\hat{\beta}_L$ must be negative:

$$\left(\frac{\partial h_I}{\partial \beta}\right)' \hat{W}_I^{-1} h_I \Big|_{\beta = \hat{\beta}_L} < 0 \qquad \textbf{(2.12)}$$

Since GMM minimizes its objective function, if the objective function has a negative slope when evaluated at $\beta = \hat{\beta}_L$ (and a zero slope when $\beta = \hat{\beta}_I$), $\hat{\beta}_I$ must be larger than $\hat{\beta}_L$.

Now, consider case 2. Assume that $\hat{\beta}_L < 0$ and $\hat{\beta}_I > 0$. The, the first term in equation (2.10) must be negative, implying that the second term must be positive. Therefore, the slope of the GMM objective function for the inverse normalization evaluated at $\beta = \hat{\beta}_L$ must be positive. This, in turn, implies that $\hat{\beta}_L > \hat{\beta}_I$, which is impossible. Similarly, if $\hat{\beta}_L > 0$ and $\hat{\beta}_I < 0$, then the GMM objective function for the inverse normalization evaluated at $\beta = \hat{\beta}_L$ must be negative, so $\hat{\beta}_L < \hat{\beta}_I$, which is also impossible. This establishes that the two estimates must be of the same sign, so only case 1 is relevant (and $\hat{\beta}_I$ must be larger than $\hat{\beta}_L$).

The GMM estimator of the inverse normalization should be preferred on theoretical grounds.[14] Although the estimates have identical asymptotic properties, the failure to impose the parameter restriction in the weighting matrix of the level normalization leads to a finite-sample estimate of β that is "too low." Moreover, because the inverse normalization implies no similar restrictions, the use of a consistent estimate of the weighting matrix has fewer consequences; in particular, the finite-sample objective function and first-order conditions in the inverse normalization are identical in form to the theoretical GMM objective function and first-order conditions.[15]

III. An Application: GMM Estimation of a Model of Costly Employment Adjustment

In this section, I demonstrate the empirical relevance of the arguments offered above. I provide two sets of GMM estimates of the costs of adjusting employment from the dynamic labor demand model in Fleischman (1996).[16] I find that estimates of adjustment costs for four-digit SIC manufacturing industries from the level nomalization of the Euler equation are economically trivial, while those from the inverse normalization are significantly larger. Given that it is GMM estimation of the more natural level normalization that produces estimates of adjustment costs that are too small, it is important to recognize and address the implications of the normalization sensitivity.

The key parameter that I estimate is γ, the adjustment cost parameter for production worker employment. Following Bils (1987), the model identifies the costs of adjusting production worker employment based on the relative movements of average weekly hours, H_t, and production worker employment, L_t. Under this identification scheme, the estimates of adjustment costs will be large if an industry reacts to an increase in its demand for production worker hours by increasing overtime hours--and paying the overtime premium--rather than increasing production worker employment. In comparison, an industry with low adjustment costs will have a higher ratio of the variance of employment relative to the variance of weekly hours than an industry with large adjustment costs.

---

[14]Of course, other considerations may figure into the choice of a preferred normalization, especially instrument relevance. In Euler equation estimation, selecting a normalization on the basis of instrument relevance likely points in the same direction, as valid instruments are likely more highly correlated with the endogenous period t variables than with the period t+1 variable.

[15]In this section, I have limited discussion to the case where there is a single source of equation error. In most applications, however, there are multiple sources of error; in addition to the expectations error featured above, the residual in an Euler equation also likely includes a specification error and, possibly, measurement errors.

[16]See Appendix A for a description of the model and the derivation of the estimating equation.

The model has two equations that are estimated jointly:[17]

$$\Delta Log(O_t) = e_0 + \eta\, \Delta Log(H_t) + \zeta_t \tag{3.1}$$

and

$$\begin{aligned} &\gamma R_{t+1} \frac{C_{t+1}}{C_t} \frac{L_{t+1}}{L_t} \Delta Log(L_{t+1}) - \gamma [\frac{1}{2}(\Delta Log(L_t)^2 + \Delta Log(L_t)] \\ &+ \frac{1}{\beta} \frac{O_t W_t^b [\eta - 1] + H_t W_t(H_t)]}{C_t} [1 + \frac{\gamma}{2}(\Delta Log(L_t))^2] - 1 = \gamma \epsilon_{t+1} \end{aligned} \tag{3.2}$$

where the second equation is shown in its "level" normalization, and O_t is weekly overtime hours, R_{t+1} is the real interest rate between periods t and t+1, C_t is hourly compensation--the sum of hourly wages, $W_t(H_t)$, which depend of weekly hours, and nonwage benefits, B_t--and W_t^b is hourly wages excluding overtime; η is the elasticity of weekly overtime hours with respect to total average weekly hours and β is the ratio of the output elasticity of weekly hours relative to the output elasticity of production worker employment. ϵ_{t+1} is the rational expectations error, which is the equation error in the absence of measurement errors or specification errors.

Note, that in equation (3.2), the adjustment cost parameter multiplies the rational expectations error. According to the arguments above, this will lead to GMM estimates of adjustment costs from this normalization that are "too low." To obtain the inverse normalization of this model, I divide equation (3.2) by the adjustment cost parameter, γ, to obtain:

$$\begin{aligned} &R_{t+1} \frac{C_{t+1}}{C_t} \frac{L_{t+1}}{L_t} \Delta Log(L_{t+1}) - [\frac{1}{2}(\Delta Log(L_t)^2 + \Delta Log(L_t)] \\ &+ \frac{1}{\gamma} \frac{1}{\beta} \frac{O_t W_t^b [\eta - 1] + H_t W_t(H_t)]}{C_t} [1 + \frac{\gamma}{2}(\Delta Log(L_t))^2] - \frac{1}{\gamma} = \epsilon_{t+1} \end{aligned} \tag{3.3}$$

If the model is correctly specified, the disturbance ϵ_{t+1} is, by definition, uncorrelated with information available in period t or earlier.

I estimate the model for more than 100 four-digit SIC industries using monthly data covering

[17] Fleischman (1996) obtains equation (3.2) by equating two measures of marginal cost--one based on the intratemporal first-order condition for weekly hours and the other based on the intertemporal first-order condition for production worker employment. See Appendix A for a more detailed discussion.

the period 1982:7 to 1993:11.[18, 19] Because the model has no closed form solution, Monte-Carlo simulations are not possible; instead, I provide estimates of adjustment costs using the two different normalizations for a large number of industries in order to show the robustness of the results. More formal testing is left for future worker based on simpler models.

The inverse and level normalizations imply the same orthogonality conditions and the same set of admissible instruments. For the cost minimization condition (either equation (3.2) or (3.3)), the instruments are monthly dummies, a quadratic time trend, and aggregate (two-digit SIC industry) values of the percentage change in production worker employment and its square, weekly hours per worker and its square, weekly overtime hours per worker and its square, the percentage change in real (product) average hourly earnings and real (product) average hourly earnings excluding overtime, and the squared percentage change in production worker employment interacted with the quadratic time trend. These instruments are all dated period t, and thus are orthogonal to the period t+1 expectations error.[20] Because the instruments are constructed using only information from the BLS establishment survey, they are plausibly orthogonal to the measurement errors in fringe benefits.[21]

In addition, I constrain adjustment costs to be positive by expressing the adjustment cost parameter as $g = \log(\gamma)$ and replacing γ in equations (3.2) and (3.3) with e^g. Imposing this constraint

---

[18]See Fleischman (1996) for a complete description of the data. Data on employment, weekly hours, and hourly earnings are not seasonally adjusted. Compensation is the sum of fringe benefits (that do not directly depend on wages), wages and salaries, and non-wage benefits tied to wages. Annual data on non-wage compensation are available in the National Income and Product Accounts (NIPA) for two-digit SIC manufacturing industries. I interpolate (error-ridden) measures of monthly fringe benefits per worker for the four digit industries by assuming that fringe benefit payments are constant through the year. Were I to address this issue directly by formally modeling the measurement error, I would modify the error terms in equations (3.2) and (3.3) to include four measurement errors, and each of these errors would be multiplied by model parameters. This would obscure the sharp distinction between the inverse and level normalizations that I exploit below. Furthermore, because the variance of the measurement errors is quite small relative to the variance of the expectation error, including the measurement errors greatly complicates the exposition and the algebra but contributes little to clarity.

[19]I exclude continuous process industries and report results only for industries for which I obtained plausible parameter estimates. See Table 2.1 in chapter 2 of Fleischman (1996) for details. These industry exclusions should not systematically affect the comparisons of the different GMM estimators.

[20]I use different instruments for the two equations in the model. For supplementary equation identifying the relationship between overtime and total weekly hours, the instruments are monthly dummies and a quadratic time trend.

[21]Under the assumption that there is no measurement error in the two-digit SIC values of employment, weekly hours, overtime hours, hourly earnings, and hourly earnings excluding overtime, the two-digit SIC instruments are valid. Fleischman (1996) examines the robustness of the estimates of adjustment costs to this assumption by reestimating the model using aggregate instruments that are once- and twice-lagged and finds that the results are qualitatively similar.

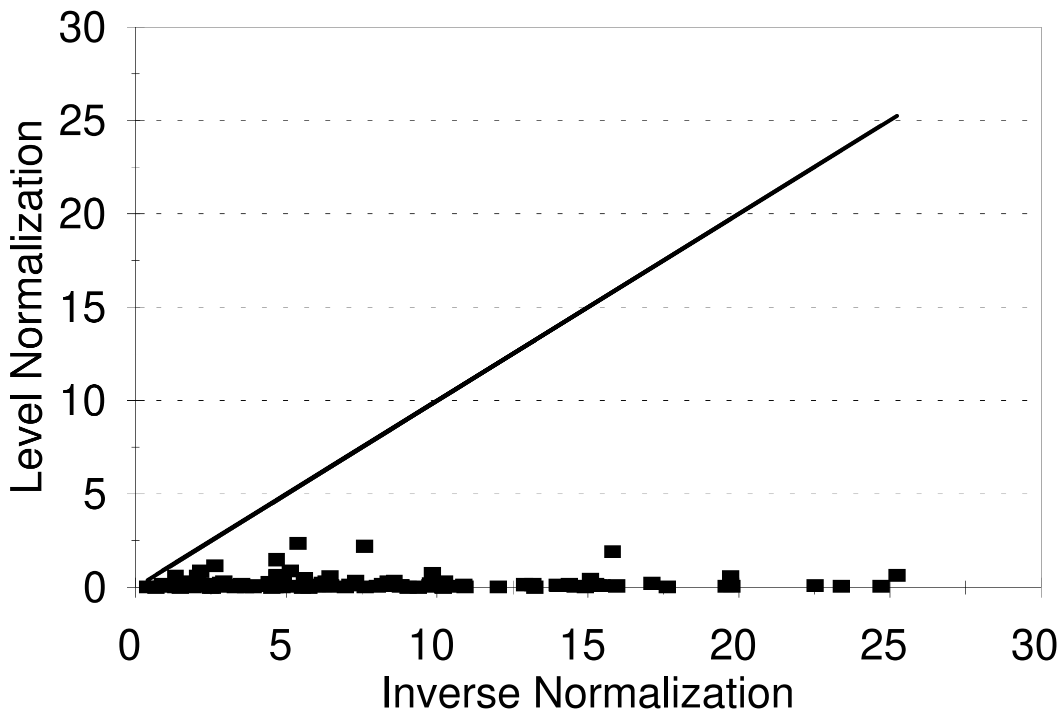

Figure 3.1
Comparison of GMM Estimates of Adjustment Costs: Level Normalization vs. Inverse Normalization

Notes for Figure 3.1:
GMM estimation of the level and inverse normalization is described in the text.

allows me to sidestep three methodological issues. First, dividing the level normalization by γ is only possible if γ is not equal to zero. Second, if γ is equal to zero, the model contains no expectations error. Third, if γ cannot equal zero but can take on negative values, which are themselves economically meaningless, then the parameter space for α is not compact. and it is therefore not possible to establish consistency of the GMM estimator.[22]

The estimates of adjustment costs and the implied speeds of employment adjustments from the two normalizations are strikingly different. In Figure 3.1, I plot the estimates of adjustment costs for

[22]See Amemiya (1985), p. 106. For this reason, I also estimate the parameter $1/\beta$ rather than β.

Table 3.1.a -- Summary Statistics GMM Estimates of Half-life of Employment Adjustment -- Level Normalization					
Two Digit SIC Industry	Number of Four Digit Industries	Median Half-life Estimate	Mean Half-life Estimate	25th Percentile Estimate	75th Percentile Estimate
All Industries	104	0.50	0.54	0.37	0.68
Nondurables	39	0.50	0.54	0.37	0.68
Durables	65	0.48	0.54	0.36	0.68
20	20	0.57	0.58	0.42	0.73
22	5	0.48	0.48	0.29	0.68
23	7	0.45	0.45	0.34	0.50
24	6	0.72	0.69	0.47	0.86
25	5	0.48	0.53	0.41	0.68
26	1	0.61	0.61	N/A	N/A
27	1	0.90	0.90	N/A	N/A
28	3	0.49	0.59	0.41	0.87
31	2	0.22	0.22	N/A	N/A
32	9	0.62	0.61	0.48	0.73
33	4	0.33	0.33	0.29	0.36
34	10	0.55	0.61	0.42	0.86
35	14	0.49	0.57	0.40	0.69
36	14	0.39	0.42	0.30	0.49
37	2	0.44	0.44	N/A	N/A
38	1	0.84	0.84	N/A	N/A

See notes at end of table.

the four-digit industries for which estimates are available for both normalizations, with the estimates from the level normalization arrayed along the vertical axis and the estimates from the inverse normalization arrayed along the horizontal axis. The diagonal line has a slope of one. The estimates of adjustment costs from the inverse normalization are clearly larger than those from the inverse normalization, generally by more than an order of magnitude. Moreover, the figure shows that the estimated adjustment costs from the level normalization seem to be unrelated to the estimates from the inverse normalization. The disparity between the estimates of the model from the two normalizations are economically as well as statistically significant. The wide discrepancy between these sets of results underscores the importance of choosing an appropriate normalization for the estimating equation. From the estimates of adjustment costs using the level normalization, one could conclude that production worker employment is essentially costless to adjust. The estimates from the inverse

Table 3.1.b -- Summary Statistics GMM Estimates of Half-life of Employment Adjustment -- Inverse Normalization					
Two Digit SIC Industry	Number of Four Digit Industries	Median Half-life Estimate	Mean Half-life Estimate	25th Percentile Estimate	75th Percentile Estimate
All Industries	133	2.00	2.37	1.32	2.87
Nondurables	55	1.65	2.12	1.11	2.25
Durables	78	2.38	2.54	1.58	3.05
20	21	1.21	1.46	0.93	2.05
22	11	2.13	3.21	1.80	3.04
23	10	1.17	1.20	1.04	1.32
24	6	1.71	2.66	1.23	4.09
25	5	1.47	1.43	0.94	1.91
26	3	5.93	5.04	2.97	6.23
27	2	2.00	2.00	N/A	N/A
28	3	4.70	3.61	0.63	5.50
31	5	1.66	1.71	1.17	2.29
32	11	2.73	3.04	1.50	4.30
33	6	3.17	3.13	2.56	3.55
34	11	1.89	2.20	1.33	2.55
35	18	2.87	2.93	2.47	3.36
36	17	2.04	2.36	1.75	2.61
37	3	1.42	1.42	1.12	1.73
38	1	1.60	1.60	N/A	N/A

Notes for Table 3.1:

The median, 25th percentile, and 75th percentile values are estimates of the half-life of employment adjustment for a particular four-digit industry within the two-digit industry. The mean half-life is the unweighted average of half-lives estimated for the four-digit industries within the two-digit industry.

normalization, however, suggest that adjustment costs are significantly larger, but still relatively small.

In general, it is difficult to give a direct interpretation to the adjustment cost parameter. I focus on the implied estimates of the half-life of employment adjustment, the measure used by Hamermesh (1993) to compare the results of various studies of dynamic labor demand. In Table 3.1.a and 3.1.b, I present summary statistics for the estimates of the half-life of adjustment for 104 four-digit manufacturing industries using the level normalization and for 133 four-digit industries using the inverse normalization. I present the median, mean, 25th percentile, and 75th percentile estimate for both normalizations for each of the two-digit manufacturing industries and for the full sample, durable goods industries, and nondurable goods industries. The median implied half-life of employment adjustment from the level normalization is 0.5 month for the sample as a whole and for both durable and nondurable goods industries. This is four times as fast as the median half-life implied by the estimates of the inverse normalization. The estimates of the speed of adjustment from the inverse normalization are near the lower end of the range of estimates of the half-life of employment adjustments from studies summarized by Hamermesh (1993) that use monthly data. The estimates from the level normalization fall well outside this range.

The GMM estimates of adjustment costs described above show substantial normalization sensitivity. But, despite the theoretical arguments in section II, it is possible that this normalization sensitivity is driven by factors other than those described above. To address this concern, and to determine whether the theoretical implications of section II apply more generally, I examined the normalization sensitivity of Shapiro's (1986a) estimates of the costs of adjusting nonproduction workers.[23] There were five equations in Shapiro's (1986a) dynamic factor demand model: Euler equations for capital, production worker employment, nonproduction worker employment, and weekly production worker hours, as well as an equation for the wage bill. I limited my replication to the estimation of the equation for nonproduction workers, and specifically to the equation for nonproduction workers excluding cross adjustment terms:

$$\frac{a_n}{N_t} - g_{NN}(N_t - q_{t-1}N_{t-1})y_t + g_{NN}(N_{t+1} - q_t N_t)y_{t+1}q_t r_t - S_t^N = -g_{NN}\epsilon_{t+1}$$

In contrast to the Euler equation for nonproduction workers, the Euler equation for capital included two expectations errors, and, therefore, was too complicated as an example. In addition, the Euler equations for production worker employment and weekly hours were poorly estimated, with some specifications yielding negative adjustment costs. Using the Euler equation for nonproduction workers, I found normalization sensitivity and the predicted relationship between the estimates; the estimated adjustment cost parameter from the level normalization--the normalization reported by Shapiro--was 0.28 with a standard error of 0.07 while the estimated adjustment cost parameter from the inverse normalization was 0.43, also with a standard error of 0.07.[24]

---

[23]Matthew Shapiro kindly provided me data from his original study.

[24]The instruments included a constant and time trend, two lags of the number of nonproduction workers and the log of the number of nonproduction workers, one lag of the inverse of nonproduction workers, and one lag of nonproduction-worker compensation. Using the same instruments, the QML estimate of the adjustment cost parameter was 0.56 with a standard error of 0.24. Although the QML estimate is larger than either GMM estimate, it still suggests that the GMM estimate from the level normalization may somewhat understate adjustment costs.

IV. Quasi-Maximum Likelihood Estimates: Normalization is Irrelevant

Unlike conventional GMM estimation, estimates obtained through maximum-likelihood methods are invariant to changes in parameter normalizations (Hendry 1995, p. 382), both asymptotically and in finite samples. This property of ML estimators can be useful even for models where the distribution of the errors is not normal. Hausman (1975) shows that full information maximum likelihood (FIML) estimates of linear systems and systems that are non-linear in parameters but linear in variables have an instrumental variables (IV) interpretation, and are thus consistent, even when the error structure is misspecified as a multivariate normal distribution.[25] The consistency obtains because the maximization of the log likelihood function under the assumption of joint normality of the residuals is equivalent to a minimization of a weighted sum of squares when the Jacobian matrix--the derivatives of the residuals with respect to the endogenous variables--does not depend on the observations.

In this section, I briefly review Hausman's findings in the context of the adjustment cost model and use them to obtain an instrumental variables estimator that is invariant to changes in the normalizations of the equations. I derive a quasi-maximum likelihood (QML) estimator of the two equation structural model (equations (3.1) and (3.2)). I follow Hausman (1977) by treating each of the exogenous variables measured with error as endogenous by adding an equation to the model for each. I also treat $\Delta\log(H_t)$ as endogenous because it is correlated with the specification error in equation (3.1). Using the notation in Hausman (1975), the model can be written as:

$$Y\mathrm{B} + Z\Gamma = V$$

where T is the number of time periods; M is the number of equations; Y is the $T*M$ matrix of endogenous variables, including the four exogenous variables measured with error, dated no later than period t+1; Z_t is a $T*(M*(k_1 + k_2))$ matrix of the instruments; k_1 is the number of instruments in Z_1 and k_2 is the number of instruments in Z_2; V are the residuals; B is an $M*M$ nonsingular matrix of the coefficients on the endogenous variables; and Γ is an $(M*(k_1 + k_2))*M$ matrix of the coefficients on the instruments (exogenous variables). I suppress the nonlinearity of B and Γ for simplicity. The structural and measurement errors, V, are mutually independent and identically distributed M-variate normal:

[25]It is straightforward to show this for LIML estimation of single equation models. Because I estimate two equations that are both part of a larger system, the estimation is closer in spirit to LIML than to FIML.

$$V \sim N(0, \Sigma \otimes I_T)$$

Under this assumption, Hausman writes the log likelihood function for this as:

$$L(\mathrm{B}, \Gamma, \Sigma) = \frac{T}{2}\log\det(\Sigma)^{-1} + T|\log\det(B)| - \frac{T}{2}tr\left[\frac{1}{T}\Sigma^{-1}(Y\mathrm{B} + Z\Gamma)'(Y\mathrm{B} + Z\Gamma)\right]$$

Hausman simplifies the first order conditions for a maximum, $\partial L/\partial B = 0$, $\partial L/\partial \Gamma = 0$, and $\partial L/\partial(\Sigma^{-1}) = 0$ by solving for T in the condition, $\partial L/\partial(\Sigma^{-1}) = 0$, and substituting this expression into the $\partial L/\partial(B) = 0$ condition to obtain the necessary conditions for a maximum:

$$\begin{bmatrix} -Z' \\ (\mathrm{B}')^{-1}\Gamma' Z' \end{bmatrix} (Y\mathrm{B} + Z\Gamma)\,\Sigma^{-1} = 0 \tag{4.1}$$

Equation (4.1) provides a method-of-moments interpretation for the QML estimator. This estimator is invariant to changes in parameter normalization. The stacked first-order conditions in equation (4.1) are weighted versions of the orthogonality conditions for the GMM estimation, where the weighting matrix is the covariance matrix of the residuals. This characteristic is key to the invariance of the QML estimation. Renormalization of the model systematically changes the score vector, the first matrix on the left-hand side of equation (4.1), the model ($Y\mathrm{B} + Z\Gamma$), and Σ^{-1} so that equation (4.1) remains unchanged. In addition, QML uses a restricted projection of the endogenous variables (or exogenous variables measured with error) onto the exogenous variables, Z, as instruments by imposing the values of known parameters:

$$\hat{Y} = Z\hat{\Gamma}\hat{\mathrm{B}}^{-1}$$

In contrast, GMM does not use this information. Hausman then shows that it is possible to rewrite the model as:

$$y = X\delta + V \tag{4.2}$$

such that the equations are stacked, and X includes all of the endogenous and predetermined variables whose coefficients are not known a priori to be zero. Then, equation (4.2) can be substituted into equation (4.1) and the first order conditions can be rewritten as:

$$\begin{bmatrix} -Z' \\ (B')^{-1}\Gamma' Z' \end{bmatrix} (y - X\delta)\,\Sigma^{-1} = 0$$

The QML estimator of δ, which includes both B and Γ, can be written in what Hausman refers to as instrumental variables form as:

$$\hat{\delta} = (\bar{W}'X)^{-1}\bar{W}'y$$

where the instruments are

$$\bar{W} = \hat{X}'(S \otimes I_T)^{-1}$$

and the weighting matrix, S, is equal to:

$$\frac{1}{T}(Y\hat{B} + Z\hat{\Gamma})'(Y\hat{B} + Z\hat{\Gamma})$$

The invariance of QML estimates to renormalization obtains because both the score and covariance matrices are reparameterized, leaving the likelihood function unchanged. The GMM estimators, however, like all instrumental variables estimators, are not invariant to the normalization. In the level normalization, GMM does not impose the parameter restrictions on the orthogonality conditions and the weighting matrix; in the inverse normalization, there are no restrictions to impose. In contrast to GMM, the QML estimator does not hold its weighting matrix constant; the QML first-order conditions impose all of the parameter restrictions implied by the model. Because there are no parameter restrictions implied by the inverse normalization, GMM estimation of this normalization is able to impose trivially all of the restrictions.

Figure 4.1 plots the QML estimates of adjustment costs against the GMM estimates of the level normalization. There is little relationship between the two sets of estimates. While there is substantial variability in the QML estimates, the GMM estimates of the level normalization are universally smaller and show almost no variation. Figure 4.2 plots the QML estimates against the GMM estimates of the inverse normalization. The QML estimates of adjustment costs are similar to, though somewhat smaller than, the GMM estimates of the inverse normalization, with most GMM estimates of the inverse normalization lying just above the 45 degree line. I present the summary statistics for the QML estimates of the implied half-life of adjustment in Table 4.1. The QML estimate of the implied half-life of adjustment of 1.5 months falls between the two GMM estimates of 2.0 months for the inverse normalization and 0.5 month for the level normalization.

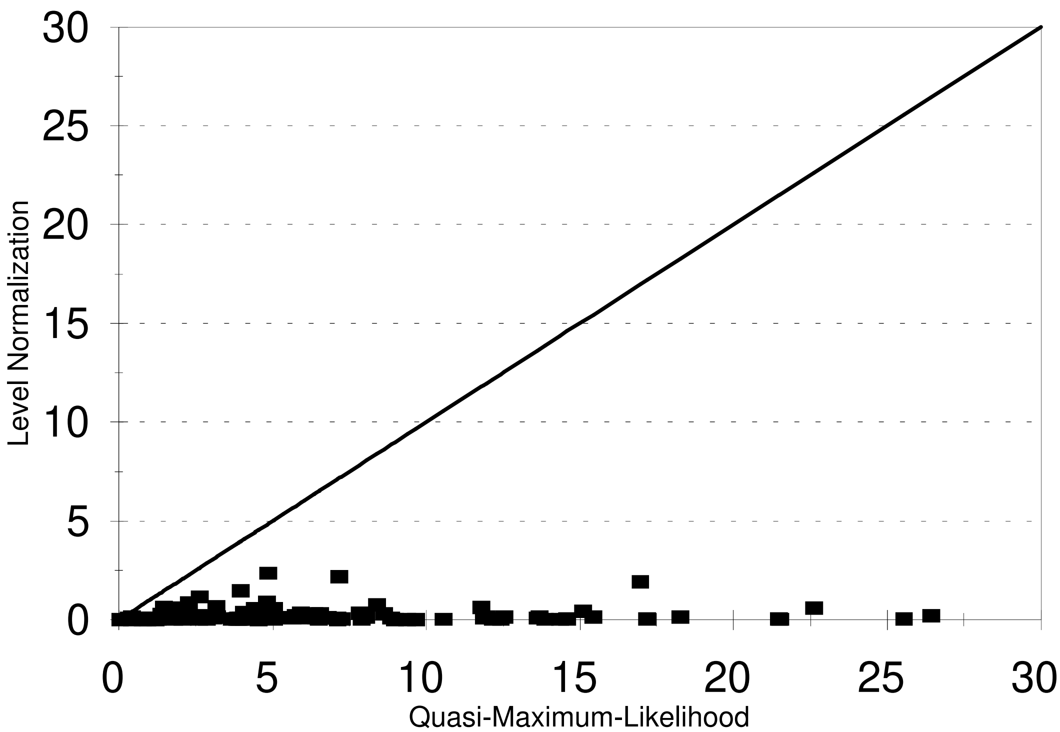

Figure 4.1
Comparison of Estimated Adjustment Costs: GMM (Level Normalization) vs. QML

Notes for Figure 4.1:
GMM results are from estimation of equations (3.1) and (3.2); QML results are from estimation of equations (3.1) and (3.2) jointly with equations for each of the period t variables measured with error.

V. Conclusions

In this paper, I identify how a feature of most conventional GMM estimators--the use of a consistent estimate of the weighting matrix rather than the joint estimation of the weighting matrix and the model parameters--can lead to the finite sample sensitivity of GMM estimation. I find that in Euler equation estimation, at least one of the model parameters multiplies the rational expectations error(s) in most normalizations, including the most natural normalization, which obtains when an unobserved conditional expectation is replaced with its realized value and an expectations error. When a model parameter multiplies the error, it implies restrictions on the structure of the optimal weighting

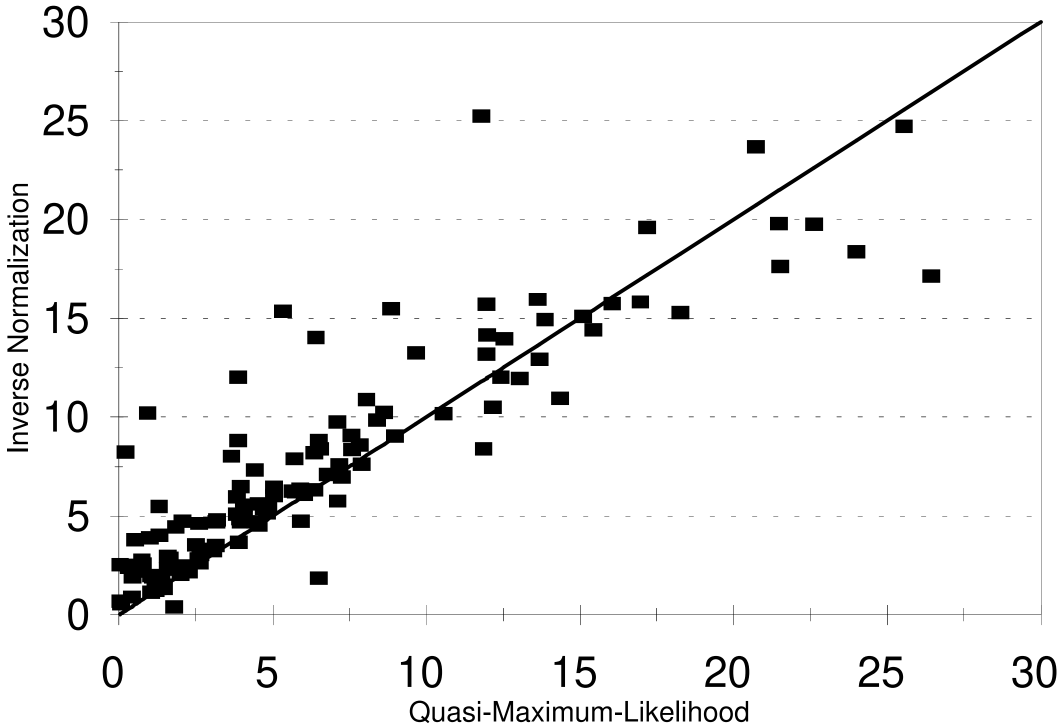

Figure 4.2
Comparison of Estimates of Adjustment Costs: GMM (Inverse Normalization) vs. QML

Notes for Figure 4.2:
GMM results are from estimation of equations (3.1) and (3.3); QML results are from estimation of equations (3.1) and (3.2) jointly with equations for each of the period t variables measured with error.

matrix. But, because GMM estimators that either hold the initial consistent estimate of the weighting matrix fixed or allow only limited iteration on the weighting matrix do not parameterize the weighting matrix, these estimators cannot impose these restrictions. For a limited class of models--those that are linear in variables and contain only a single dominant source of equation error--there is an alternative normalization where no parameter multiplies the rational expectations error. In this alternative normalization, there are no restrictions on the structure of the weighting matrix.

GMM estimation of a normalization of the Euler equation where a model parameter multiplies the expectations error effectively minimizes the square of this parameter times a term that converges in probability to the asymptotic objective function divided by the square of the parameter. In contrast, GMM estimation of the alternative normalization minimizes a finite sample objective function that

Table 4.1 -- Summary Statistics Quasi-Maximum-Likelihood Estimates of Half-life of Employment Adjustment					
Two Digit SIC Industry	Number of Four Digit Industries	Median Half-life Estimate	Mean Half-life Estimate	25th Percentile Estimate	75th Percentile Estimate
All Industries	126	1.48	1.95	0.99	2.61
Non-Durables	52	1.08	1.58	0.70	1.68
Durables	74	1.90	2.21	1.29	2.86
20	21	1.04	1.27	0.78	1.67
22	11	1.58	2.82	1.08	2.61
23	9	0.69	0.67	0.41	0.85
24	6	2.08	2.76	1.17	4.07
25	5	1.27	1.12	0.60	1.56
26	1	1.53	1.53	N/A	N/A
27	2	1.75	1.75	N/A	N/A
28	3	4.53	3.48	0.41	5.51
31	5	0.43	0.62	0.37	0.96
32	10	2.13	2.53	1.29	3.21
33	6	3.14	3.24	2.64	3.74
34	10	1.40	1.53	1.12	1.97
35	18	2.85	2.81	2.29	3.42
36	15	1.43	1.74	1.00	2.54
37	3	1.17	1.10	0.80	1.34
38	1	1.40	1.40	N/A	N/A

Notes for Table 4.1:

The median, 25th percentile, and 75th percentile values are estimates of the half-life of employment adjustment for a particular four-digit industry within the two-digit industry. The mean half-life is the unweighted average of half-lifes estimated for the four-digit industries within the two-digit industry.

converges to the asymptotic objective function. Consequently, the parameter estimate from the first normalization will be smaller in absolute value than the estimate from the alternative normalization. And, because when no parameters multiply the rational expectations error there are no unimposed restrictions on the weighting matrix, GMM estimates from the alternative normalization should be preferred on theoretical grounds.

The findings here are significant because they point to the source of the GMM normalization puzzle. For a limited class of models, I can explain why estimates from one normalization will be larger (in absolute value) than the estimates from a different normalization. However, I leave it to future work to explain why the two sets of estimates can be sufficiently different as to lead to different economic interpretations. One area for future exploration that likely offers substantial payoffs is the interaction between the failure to parameterize the weighting matrix and the degree of instrument relevance. In particular, the instruments used to estimate the dynamic labor demand model were more correlated with the period t endogenous variable that is multiplied by the adjustment cost parameter in the inverse normalization than with the realization of the period t+1 variable multiplied by the adjustment cost parameter in the level normalization; poor instrument relevance may explain why the GMM estimates of adjustment costs from the level normalization were so small.

Another important limitation of the results in this paper is that they apply only to models with a single dominant source of equation error. For many linear models, including the linear-quadratic inventory model, there is no normalization that is analogous to the inverse normalization. Unlike the model studied here, there are two rational expectations errors in the linear-quadratic inventory model, one dated period t+1 and one dated period t+2. These disturbances are multiplied by different parameters. Consequently, none of the possible normalizations of the model have no parameters multiplying both disturbances.

I show that for models with a single expectations error, QML estimates can be useful when choosing between two competing sets of GMM estimates from different normalizations of the same estimating equation. But, QML estimation may not lead to consistent parameter estimates for models that are nonlinear in variables because the method of moments interpretation of the QML estimator depends on the constancy of the Jacobian. Examples of models for which this concern may be important include asymmetric adjustment cost models (e.g., Burda 1991) or time-varying adjustment costs models (e.g., Burgess and Dolado 1989). GMM estimation of these models may not be feasible because these are subject to the same sensitivity to the choice of parameter normalization but, like the linear-quadratic inventory model, offer no clear choice of parameter normalization.

Appendix A. A Model of Dynamic Labor Demand

The dynamic labor demand model I describe here is taken from Fleischman (1996). A representative firm has variable costs that include both production worker compensation and the costs of adjusting production worker employment. The firm minimizes the expected discounted value of current and future variable costs by choosing production worker employment (L_t) and hours per production worker (H_t):[26]

$$\underset{\{L_{t+s},\, H_{t+s}\}}{Min} \quad E_t[\sum_{s=0}^{\infty}(\prod_{j=0}^{s}(\frac{1}{1+r_{t+s}}))L_{t+s}[B_{t+s} + W_{t+s}(H_{t+s})H_{t+s}][1 + \tfrac{\gamma}{2}(\Delta Log(L_{t+s})^2]] \;, \qquad s{=}0,1,2,..... \tag{A.1}$$

$$\begin{aligned} s.t. \quad Y_{t+s} &= F(G(L_{t+s}, H_{t+s}), N_{t+s}, K_{t+s}, M_{t+s}, E_{t+s})A_{t+s} \\ G(L_{t+s}, H_{t+s}) &= L_{t+s} H_{t+s}^{\beta} \end{aligned}$$

where s and t both index time; B_t is non-wage benefits per worker; $W_t(H_t)$ is hourly wages that depend on the weekly hours of production workers; r_t is the real interest rate from period t to period t+1; Y_t is gross output; N_t is non-production workers; K_t is the capital stock; M_t is materials input; E_t is energy input; A_t is the level of productivity; $G(L_t, H_t)$ is production worker labor; γ is the adjustment cost parameter; and β is the elasticity of production worker labor with respect to hours per production worker. Total production worker hours is the product of one flexible factor (hours per worker) and one quasi-fixed factor (production worker employment).

The key features of the model are adjustment costs that are symmetric and quadratic in the size of the (net) percentage change in production worker employment and proportional to production worker compensation, separability of the production function, output elasticities of hours per worker and production worker employment that can differ, elastic supply of production workers at the market rate of compensation per worker, and a schedule of hourly wage rates that is increasing in the number

[26] I treat the capital stock and non-production workers as fixed factors in the short run, and materials and energy inputs as fully flexible factors. As I explain below, separability of the production function allows me to focus on the allocation of production worker hours into hours per worker and the number of production workers, conditional on the firm using the cost-minimizing levels of the other inputs. Thus dropping the other factors from the variable costs shown in equation (A.1) has no effect on the first-order conditions for production worker employment and hours per worker. The assumptions of separability of the production function and no cross-adjustment terms are driven by the lack of monthly output and capital stock data for the four-digit manufacturing industries. See Fleischman (1996) for more detail.

of hours per production worker and is internalized by the firm. The separability of the production function allows me to focus on the firm's allocation of production worker hours into hours per worker and the number of production workers. Changes in the level of output or the prices of other factors affect only the level of total production worker hours, not its composition.

The firm's first order conditions for employment and hours per worker are:

$$C_t[1 + \gamma[\tfrac{1}{2}(\Delta Log(L_t)^2 + \Delta Log(L_t)]] - \gamma E_t[R_{t+1} \frac{L_{t+1}}{L_t} C_{t+1} \Delta Log(L_{t+1})] = \mu_t \frac{\partial F_t}{\partial G_t} G(L_t, H_t) L_t^{-1} \qquad \textbf{(A.2)}$$

and

$$L_t[H_t \frac{dW_t(H_t)}{dH_t} + W_t(H_t)][1 + \frac{\gamma}{2}(\Delta Log(L_t))^2] = \mu_t \frac{\partial F_t}{\partial G_t} G(L_t, H_t) H_t^{-1} \beta \qquad \textbf{(A.3)}$$

where $\partial F_t/\partial G_t$ is the marginal product of production worker labor, $G(L_t, H_t)$, and μ_t, the Lagrange multiplier on the production function constraint, is marginal cost in period t. I equate marginal cost from equations (A.2) and (A.3), and cancel common terms to obtain:

$$\gamma E_t[R_{t+1} \frac{C_{t+1}}{C_t} \frac{L_{t+1}}{L_t} \Delta Log(L_{t+1})] - \gamma[\frac{1}{2}(\Delta Log(L_t)^2 + \Delta Log(L_t)] + \frac{1}{\beta} \frac{[H_t^2 \frac{dW_t(H_t)}{dH_t} + H_t W_t(H_t)]}{C_t}[1 + \frac{\gamma}{2}(\Delta Log(L_t))^2] = 1 \qquad \textbf{(A.4)}$$

where C_t is total compensation per production worker, which is equal to non-wage benefits per worker plus average weekly wages.

I follow Bils (1987) in using movements in hours per worker, overtime hours per worker, and production worker employment to identify movements in marginal cost, and hence, the costs of adjusting employment. Because workers receive an overtime premium and weekly overtime hours are positively related to total weekly hours, the firm perceives that hourly wages are an increasing function of the number of hours per worker.[27] Under the assumptions of a constant elasticity of overtime hours with respect to weekly hours (η) and after imposing the relationship between the hourly wage paid, W_t^p, and the base wage (average hourly earnings excluding overtime), W_t^b, implicit in the Bureau of

---

[27] I assume that the overtime premium is equal to one-half of the straight-time (base) wage. This assumption is the only one consistent with the construction of the data by the Bureau of Labor Statistics.

Labor Statistics (BLS) estimates of hourly wage and weekly hours-- $W_t = W_t^b(1 + \frac{1}{2}\frac{O_t}{H_t})$--equation (A.4) can be rewritten:

$$\gamma E_t[R_{t+1}\frac{C_{t+1}}{C_t}\frac{L_{t+1}}{L_t}\Delta Log(L_{t+1})] - \gamma[\frac{1}{2}(\Delta Log(L_t))^2 + \Delta Log(L_t)] + \frac{1}{\beta}\frac{O_t W_t^b[\eta - 1] + H_t W_t(H_t)]}{C_t}[1+\frac{\gamma}{2}(\Delta Log(L_t))^2] - 1 = 0 \qquad \textbf{(A.5)}$$

where O_t is overtime hours per worker I follow Shapiro (1986a, 1986b) by jointly estimating equation (A.5) along with an equation that identifies the parameters of the marginal wage schedule:

$$\Delta Log(O_t) = e_0 + \eta\,\Delta Log(H_t) + \zeta_t \qquad \textbf{(A.6)}$$

where e_0 is the average rate of change in overtime hours and ζ_t is a specification error representing the determinants of the relationship between weekly hours and weekly overtime hours left out of equation (A.6).

Equation (A.5) contains the unobservable period t expectation of the discounted percentage change in period t+1 employment (where the discount factor includes the interest rate as well as the ratio of the period t+1 to the period t values of total production worker compensation). The realized value is equal to the time t expectation plus the rational expectations forecast error, ϵ_{t+1}:

$$R_{t+1}\frac{C_{t+1}}{C_t}\frac{L_{t+1}}{L_t}\Delta Log(L_{t+1}) = E_t[R_{t+1}\frac{C_{t+1}}{C_t}\frac{L_{t+1}}{L_t}\Delta Log(L_{t+1})] + \epsilon_{t+1}$$

I substitute the realized value into equation (A.5) to obtain:

$$\gamma R_{t+1}\frac{C_{t+1}}{C_t}\frac{L_{t+1}}{L_t}\Delta Log(L_{t+1}) - \gamma[\frac{1}{2}(\Delta Log(L_t))^2 + \Delta Log(L_t)] + \frac{1}{\beta}\frac{O_t W_t^b[\eta - 1] + H_t W_t(H_t)]}{C_t}[1+\frac{\gamma}{2}(\Delta Log(L_t))^2] - 1 = \gamma\epsilon_{t+1} \qquad \textbf{(A.7)}$$

In this normalization, which I refer to as the level normalization, the adjustment cost parameter, γ, multiplies the rational expectations error, ϵ_{t+1}. I obtain an alternative normalization of the marginal cost condition, which I refer to as the inverse normalization, by dividing equation (A.7) by γ:

$$R_{t+1}\frac{C_{t+1}}{C_t}\frac{L_{t+1}}{L_t}\Delta Log(L_{t+1}) - [\frac{1}{2}(\Delta Log(L_t)^2 + \Delta Log(L_t)]$$

$$+\frac{1}{\gamma}\frac{1}{\beta}\frac{O_t W_t^b[\eta - 1] + H_t W_t(H_t)]}{C_t}[1+\frac{\gamma}{2}(\Delta Log(L_t))^2] - \frac{1}{\gamma} = \epsilon_{t+1} \tag{A.8}$$

In the text I describe the joint estimation of equations (A.6) with either equation (A.7) or equation (A.8)

References

Amemiya, Takeshi. 1985. *Advanced Econometrics.* Cambridge, MA: Harvard University Press.

Bartelsman, Eric. J. 1995. "Of Empty Boxes: Returns to Scale Revisited." *Economics Letters.* 49:59-67.

Basu, Susanto and Miles S. Kimball. 1997. "Cyclical Productivity with Unobserved Input Variation." Unpublished manuscript, University of Michigan.

Bernanke, Ben. 1986. "Employment, Hours and Earnings in the Depression." *American Economic Review.* 76: 82-109.

Bils, Mark. 1987. "The Cyclical Behavior of Marginal Cost and Price." *American Economic Review.* 77:838-855.

Burda, Michael C. 1991. "Monopolistic Competition, Costs of Adjustment, and the Behavior of European Manufacturing Employment." *European Economic Review.* 35: 61-79.

Burgess, Simon and Juan Dolado. 1989. "Intertemporal Rules with Variable Speed of Adjustment: An Application to UK Manufacturing Employment." *Economic Journal.* 99:347-65.

Fleischman, Charles A. 1996. *Heterogeneous Employment Adjustment Costs in U.S. Manufacturing Industires.* Doctoral Dissertation. University of Michigan.

Fuhrer, Jeffrey C., George R. Moore, and Scott D. Schuh. 1995. "Estimating the Linear Quadratic Inventory Model: Maximum Likelihood versus Generalized Method of Moments." *Journal of Monetary Economics.* 35:115-157.

Hall, Robert E. 1988. "The Relation Between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy.* 96:921-47.

____. 1990. "Invariance Properties of Solow's Productivity Residual." in Peter Diamond, ed. *Growth/Productivity/Unemployment.* Cambridge, MA: MIT Press.

Hamermesh, Daniel S. 1993. *Labor Demand.* Princeton, NJ: Princeton University Press.

Hamilton, James D. 1994. *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Hansen, Lars Peter, John Heaton, and Amir Yaron. 1996. "Finite-Sample Properties of Some Alternative GMM Estimators." *Journal of Business and Economic Statistics.* 14:262-280.

Hansen, Lars Peter and Kenneth J. Singleton. 1982. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica.* 50:1269-1286.

Hausman, Jerry A. 1975. "An Instrumental Variable Approach to Full Information Estimators for

Linear and Certain Nonlinear Econometric Models." *Econometrica.* 43:727-739.

____. 1977. "Errors in Variables in Simultaneous Equations Models." *Journal of Econometrics.* 5:389-401.

Hendry, David F. 1995. *Dynamic Econometrics.* New York: Oxford University Press. p. 382.

Kennan, John. 1988. "An Econometric Analysis of Fluctuations in Aggregate Labor Supply and Demand." *Econometrica.* 47:1433-41.

Krane, Spencer D. and Steven N. Braun. 1991. "Production Smoothing Evidence from Physical Product Data." *Journal of Political Economy.* 99:558-581.

Roberts, John M., David J. Stockton, and Charles S. Struckmeyer. 1994. "Evidence on the Flexibility of Prices." *Review of Economics and Statistics.* 76:142-150.

Oliner, Stephen, Glenn Rudebusch, and Daniel Sichel. 1996. "The Lucas Critique Revisited: Assessing the Stability of Empirical Euler Equations for Investment." *Journal of Econometrics.* 70:291-316.

Pfann, Gerard A. 1996. "Factor Demand Models with Nonlinear Short-Run Fluctuations." *Journal of Economic Dynamics and Control.* 20:315-331.

____ and Franz C. Palm. 1993. "Asymmetric Adjustment Costs in Non-Linear Labour Demand Models for the Netherlands and U.K. Manufacturing Sectors." *Review of Economic Studies.* 60:397-412.

Ramey, Valerie A. 1991. "Nonconvex Costs and the Behavior of Inventories." *Journal of Political Economy.* 99:306-334.

Sbordone, Argia M. 1996. "Cyclical Productivity in a Model of Labor Hoarding." *Journal of Monetary Economics.* 38:331-361.

Shapiro, Matthew D. 1986a. "The Dynamic Demand for Capital and Labor." *The Quarterly Journal of Economics.* 101:513-542.

____. 1986b. "Capital Utilization and Capital Accumulation: Theory and Evidence." *Journal of Applied Econometrics.* 1:211-234.

West, Kenneth D. 1986. "A Variance Bounds Test of the Linear Quadratic Inventory Model." *Journal of Political Economy.* 94:374-401.

____ and David W. Wilcox. 1993. "Some Evidence on the Finite Sample Behavior of an Instrumental Variables Estimator of the Linear Quadratic Inventory Model." NBER Technical Working Paper No. 139.