MOMENT ESTIMATION
IN A CORRELATED RANDOM COEFFICIENTS LINEAR PANEL DATA MODEL: A FUNCTIONAL FIXED POINT APPROACH TO IDENTIFICATION AND ESTIMATION

Stefan Hoderlein, Emory University
Robert Sherman, Caltech
Akshay Srivastava, Caltech

Abstract

We develop a linear regression panel data model allowing random coefficients to be correlated with regressors not only within periods but also across periods. The random coefficients are modeled as sums of independent possibly nonidentically distributed past and current shocks. This structure allows feedback, in the sense that future regressors can be correlated with all past shocks, and also allows all lagged dependent variables to be regressors.

For each time period, we identify all marginal first moments of the random coefficients. These moments have causal interpretations. The identification results rest on a novel functional fixed point argument and lead to natural estimators of these moments. We provide simulation evidence of the usefulness of this approach.

Introduction

The Model and Assumptions

The model we present allows $k \geq 2$ random coefficients in each time period and requires $T \geq k$ periods of data to identify the marginal first moments of these coefficients. These moments are average partial effects and are the causal effects of interest in this paper.

The case $k = 2$ is very special. Identification and estimation for this case is fairly straightforward and has been treated by FHHPS (2019). The cases $k \geq 3$ are qualitatively different. These cases have also been treated by FHHPS (2019) without using functional fixed point arguments by requiring stronger assumptions. For example, the FHHPS approach does not allow lagged dependent variables of order one as regressors.

In this paper, we use functional fixed point arguments to identify marginal first moments in these more flexible models when $k \geq 3$. Throughout we view T as fixed, but arbitrary, and let the cross section dimension tend to infinity.

For ease of exposition, we initially consider the $k = 3$ case. As stated above, we require $T \geq 3$, where T is the number of periods in the panel data set. For $t = 1, 2, \ldots, T$, define

$$\begin{aligned} Y_t &= A_t + B_t X_t + C_t Z_t \\ A_t &= A_{t-1} + U_t \\ B_t &= B_{t-1} + V_t \\ C_t &= C_{t-1} + W_t \end{aligned}$$

where $A_0 \equiv U_0$ is an initializing random intercept, $B_0 \equiv V_0$ and $C_0 \equiv W_0$ are initializing scalar random slope coefficients, and X_t and Z_t are observed scalar regressors. The (U_t, V_t, W_t)s can be viewed as technology or preference shocks. Note that for $t = 1, 2, \ldots, T$,

$$A_t = \sum_{s=0}^{t} U_s, \quad B_t = \sum_{s=0}^{t} V_s, \quad \text{and} \quad C_t = \sum_{s=0}^{t} W_s\,. \tag{1}$$

That is, the random coefficients are sums of past and current shocks. Write S_t for shocks (U_t, V_t, W_t), $t = 0, 1, 2, \ldots, T$. Write R_t for all nonconstant tth period regressors (X_t, Z_t), $t = 1, 2, \ldots, T$.

Define the 3×3 matrix M_3 as follows:

$$M_3 = \begin{bmatrix} 1 & X_1 & Z_1 \\ 1 & X_2 & Z_2 \\ 1 & X_3 & Z_3 \end{bmatrix} = \begin{bmatrix} 1 & R_1 \\ 1 & R_2 \\ 1 & R_3 \end{bmatrix}$$

Define M_{3t} analogous to M_3 for successive periods t, $t+1$, and $t+2$. Thus, $M_3 \equiv M_{31}$. Finally, let S_{X_t} denote the support of X_t, $t = 1, 2, \ldots, T$.

We make the following assumptions for $t = 1, 2, \ldots, T$ where $T \geq 3$:

A0. The linear panel data model defined above holds.

A1. $S_t \perp (S_0, \ldots, S_{t-1}, R_1, \ldots, R_t)$.

A2. R_t is continuously distributed, $S_{R_1} = \cdots = S_{R_T}$, and $\boldsymbol{P}\{\det M_{3t} = 0\} = 0$.

A3. For $t = 2, \ldots, T$, $\boldsymbol{E}(1, R_t)(1, R_t)'$ is invertible.

Assumption **A0** defines the model and implies (1). Assumption **A1** says that current shocks are independent of past shocks as well as past and current regressors, but allows future regressors to be correlated with past shocks. This is desirable, since future regressors are often choice variables whose values depend on past shocks. The assumption that current shocks can be correlated with future regressors can be interpreted to mean that it takes at least one period for decision makers to react to current shocks when choosing levels of regressors. This assumption also allows lagged values of the dependent variables to be regressors for lags of any order. (1) and **A1** imply that current shocks are independent of past random coefficients: $S_t \perp (A_s, B_s, C_s)$ for $s < t$. In addition, (1) and **A1** imply that any regressor in any time period can be correlated with any random coefficient in any time period through past shocks. Thus, the model allows endogeneity, or unobserved heterogeneity, of a general type.

The continuous distribution, common support, and determinant assumptions in **A2** are used to streamline identification arguments.

Finally, the invertibility condition **A3** is a mild regularity condition used to identify shock means in a local regression.[1]

The rest of the paper is organized as follows. The next section presents the identification argument using the functional fixed point idea. Section 3 describes the estimation procedure. Section 4 presents simulation results illustrating estimator performance. Section 5 concludes.

Identification

This section presents the functional fixed point argument for identifying marginal first moments of the random coefficients of the model introduced in the last section.

We start by proving the result for the case $k = 3$, where k is the number of random coefficients in each time period. We then prove the result for $k = 4$, which generalizes in an obvious way to cover the general case $k \geq 3$.

[1]As in FHHPS (2019), any number of additional terms with nonrandom slope coefficients can be included in the response Y_t in model **A0**. In this case these coefficients are estimated along with the shock means in a local regression. For ease of exposition, we have not included these additional terms as part of the model.

We begin with the $k = 3$ case, where we require $T \geq 3$, where T is the number of periods in the panel data set. For simplicity, consider the case $T = 3$. Generalization is immediate to the case $k = 3$ and $T \geq 3$ by applying the arguments for the $k = 3$ and $T = 3$ case to all sets of three consecutive time periods of data.

Recall the model and assumptions from the previous section. It follows from assumption **A0** that

$$\begin{aligned} Y_1 &= A_1 + B_1 X_1 + C_1 Z_1 \\ Y_2 &= A_1 + B_1 X_2 + C_1 Z_2 + U_2 + V_2 X_2 + W_2 Z_2 \\ Y_3 &= A_1 + B_1 X_3 + C_1 Z_3 + (U_2 + U_3) + (V_2 + V_3) X_3 + (W_2 + W_3) Z_3 \,. \end{aligned}$$

Invoke the common support condition in **A2** to see that

$$\begin{aligned} \boldsymbol{E}(Y_2 - Y_1 \mid X_2 = X_1 = x_2, Z_2 = Z_1 = z_2) &= \boldsymbol{E}U_2 + \boldsymbol{E}V_2 x_2 + \boldsymbol{E}W_2 z_2 \\ \boldsymbol{E}(Y_3 - Y_2 \mid X_3 = X_2 = x_3, Z_3 = Z_2 = z_3) &= \boldsymbol{E}U_3 + \boldsymbol{E}V_3 x_3 + \boldsymbol{E}W_3 z_3 \,. \end{aligned}$$

Invoke **A3**, and use standard regression arguments with stayers (those sampling units whose regressor values do not change from period to period) to identify all 6 shock expectations in the last display.

Recall that $R_t = (X_t, Z_t)$. Define $R_{12} = (R_1, R_2)$. Abbreviate $\boldsymbol{E}(Y_t \mid R_{12} = r_{12}, R_3 = r_3)$ to $\boldsymbol{E}(Y_t \mid r_{12}, r_3)$. If we average the Y_ts conditional on $(R_{12}, R_3) = (r_{12}, r_3)$ and apply **A1** we get the following system of equations:

$$\begin{aligned} \boldsymbol{E}(Y_1 \mid r_{12}, r_3) &= \boldsymbol{E}(A_1 \mid r_{12}, r_3) + \boldsymbol{E}(B_1 \mid r_{12}, r_3) x_1 + \boldsymbol{E}(C_1 \mid r_{12}, r_3) z_1 \\ \boldsymbol{E}(Y_2 \mid r_{12}, r_3) &= \boldsymbol{E}(A_1 \mid r_{12}, r_3) + \boldsymbol{E}(B_1 \mid r_{12}, r_3) x_2 + \boldsymbol{E}(C_1 \mid r_{12}, r_3) z_2 \\ &+ \boldsymbol{E}(U_2 \mid r_{12}, r_3) + \boldsymbol{E}(V_2 \mid r_{12}, r_3) x_2 + \boldsymbol{E}(W_2 \mid r_{12}, r_3) z_2 \\ \boldsymbol{E}(Y_3 \mid r_{12}, r_3) &= \boldsymbol{E}(A_1 \mid r_{12}, r_3) + \boldsymbol{E}(B_1 \mid r_{12}, r_3) x_3 + \boldsymbol{E}(C_1 \mid r_{12}, r_3) z_3 \\ &+ \boldsymbol{E}(U_2 \mid r_{12}, r_3) + \boldsymbol{E}(V_2 \mid r_{12}, r_3) x_3 + \boldsymbol{E}(W_2 \mid r_{12}, r_3) z_3 \\ &+ \boldsymbol{E}U_3 + \boldsymbol{E}V_3 x_3 + \boldsymbol{E}W_3 z_3 \,. \end{aligned}$$

If we can identify $\boldsymbol{E}(A_1 \mid r_{12}, r_3)$, $\boldsymbol{E}(B_1 \mid r_{12}, r_3)$, and $\boldsymbol{E}(C_1 \mid r_{12}, r_3)$, then we can achieve the objective of identifying $\boldsymbol{E}A_1$, $\boldsymbol{E}B_1$, and $\boldsymbol{E}C_1$ by averaging over (R_{12}, R_3). However, above we have a system of 3 linear equations in the 6 unknowns $\boldsymbol{E}(A_1 \mid r_{12}, r_3)$, $\boldsymbol{E}(B_1 \mid r_{12}, r_3)$, $\boldsymbol{E}(C_1 \mid r_{12}, r_3)$, $\boldsymbol{E}(U_2 \mid r_{12}, r_3)$, $\boldsymbol{E}(V_2 \mid r_{12}, r_3)$, and $\boldsymbol{E}(W_2 \mid r_{12}, r_3)$, an underidentified system. The problem is that **A1** allows R_3 to be correlated with all the second period shocks, introducing 3 extra unknowns.

Alternatively, if we can identify, say, $\boldsymbol{E}(A_1 \mid r_{12})$, $\boldsymbol{E}(B_1 \mid r_{12})$, and $\boldsymbol{E}(C_1 \mid r_{12})$, then we can identify $\boldsymbol{E}A_1$, $\boldsymbol{E}B_1$, and $\boldsymbol{E}C_1$ by averaging over R_{12}. If we average the Y_ts conditional on $R_{12} = r_{12}$ and apply **A1** we get the following system of equations:

$$\begin{aligned} \boldsymbol{E}(Y_1 \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 \mid r_{12}) x_1 + \boldsymbol{E}(C_1 \mid r_{12}) z_1 \\ \boldsymbol{E}(Y_2 \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 \mid r_{12}) x_2 + \boldsymbol{E}(C_1 \mid r_{12}) z_2 \\ &+ \boldsymbol{E}U_2 + \boldsymbol{E}V_2 x_2 + \boldsymbol{E}W_2 z_2 \\ \boldsymbol{E}(Y_3 \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 X_3 \mid r_{12}) + \boldsymbol{E}(C_1 Z_3 \mid r_{12}) \\ &+ \boldsymbol{E}U_2 + \boldsymbol{E}(V_2 X_3 \mid r_{12}) + \boldsymbol{E}(W_2 Z_3 \mid r_{12}) \\ &+ \boldsymbol{E}U_3 + \boldsymbol{E}V_3 \boldsymbol{E}(X_3 \mid r_{12}) + \boldsymbol{E}W_3 \boldsymbol{E}(Z_3 \mid r_{12}) \,. \end{aligned}$$

This is a system of 3 linear equations in the 3 principle unknowns $\boldsymbol{E}(A_1 \mid r_{12})$, $\boldsymbol{E}(B_1 \mid r_{12})$, and $\boldsymbol{E}(C_1 \mid r_{12})$, as well as the 4 extra unknowns $\boldsymbol{E}(B_1 X_3 \mid r_{12})$, $\boldsymbol{E}(C_1 Z_3 \mid r_{12})$, $\boldsymbol{E}(V_2 X_3 \mid r_{12})$, and

$\boldsymbol{E}(W_2Z_3 \mid r_{12})$, again, an underidentified system. Extra unknowns arise because both factors in products of either coefficients and regressors or second period shocks and regressors are generally functions of R_3, and so are mixed together when taking expectations given $R_{12} = r_{12}$.

Functional fixed point ideas can be used to obtain a system of 3 linear equations in the 3 unknowns $\boldsymbol{E}(A_1 \mid r_{12})$, $\boldsymbol{E}(B_1 \mid r_{12})$, and $\boldsymbol{E}(C_1 \mid r_{12})$. Averaging over R_{12} then leads to identification of the marginal means $\boldsymbol{E}A_1$, $\boldsymbol{E}B_1$, $\boldsymbol{E}C_1$, as well as the marginal means in periods 2 and 3. The idea is to replace x_3 and z_3 with functions of r_{12} so that separation between various expectations and regressors can be achieved. However, not just any function of r_{12} will suffice. We must find an identifiable substitute for $\boldsymbol{E}(Y_3 \mid r_{12})$ corresponding to this special function of r_{12}.

To this end, recall that $Y_3 = A_3 + B_3X_3 + C_3Z_3$. Note that Y_3 depends on $R_3 = (X_3, Z_3)$ implicitly through (A_3, B_3, C_3) and explicitly through (X_3, Z_3). So, write $Y_3 \equiv Y_3(R_3, R_3)$ where the first R_3 refers to implicit dependence and the second R_3 to explicit dependence. Note that $Y_3(s,t)$, where both s and t are realizations of R_3, can be a realization of Y_3 from the model if and only if $s = t$. Also note that, conditional on $R_{12} = r_{12}$, Y_3 depends implicitly and explicitly on R_3.

Let $r_3^0 = (x_3^0, z_3^0) \equiv (x_3^0(r_{12}), z_3^0(r_{12}))$ denote a function of r_{12} that is also a possible realization of $R_3 \mid r_{12}$. For example, we might take $r_3^0 = \boldsymbol{E}(R_3 \mid r_{12}) = (\boldsymbol{E}(X_3 \mid r_{12}), \boldsymbol{E}(Z_3 \mid r_{12}))$. Or we might take, for 2×1 vectors α, β, γ, and δ, $r_3^0 = (r_1\alpha + r_2\beta, r_1\gamma + r_2\delta)$. And so on.

Consider the object

$$\boldsymbol{E}(Y_3(\cdot, r_3^0) \mid r_{12}) = \boldsymbol{E}(A_3 + B_3x_3^0 + C_3z_3^0 \mid r_{12})\,.$$

By the LIE (law of iterated expectations),

$$\boldsymbol{E}(Y_3(\cdot, r_3^0) \mid r_{12}) = \boldsymbol{E}\left[\boldsymbol{E}(Y_3(r_3, r_3^0) \mid r_{12}, r_3) \mid r_{12}\right]$$

where the outer expectation on the RHS is over R_3 given r_{12}. By the IMVT (integral mean value theorem), there exists a function $r_3^1 = (x_3^1, z_3^1) \equiv (x_3^1(r_{12}), z_3^1(r_{12}))$ such that

$$\boldsymbol{E}(Y_3(\cdot, r_3^0) \mid r_{12}) = \boldsymbol{E}(Y_3(r_3^1, r_3^0) \mid r_{12}, r_3^1)\,.$$

The IMVT step defines a mapping: $m(r_3^0) = r_3^1$. Consider the sequence of mappings

$$r_3^i = m(r_3^{i-1}) \quad i = 1, 2, \ldots$$

Does this sequence of functions have a fixed point?

Suppose for the moment that a fixed point function exists, and call it

$$r_3^* = (x_3^*, z_3^*) \equiv (x_3^*(r_{12}), z^*(r_{12}))\,.$$

This means that

$$\boldsymbol{E}(Y_3(\cdot, r_3^*) \mid r_{12}) = \boldsymbol{E}(Y_3(r_3^*, r_3^*) \mid r_{12}, r_3^*)\,.$$

If we knew r_3^*, we would be able to estimate the LHS using the RHS: average the observed Y_3s local to r_{12} and r_3^*.

Now return to the original system of equations but this time replace $Y_3 \equiv Y_3(R_3, R_3)$ with $Y_3(R_3, r_3^*)$ and take expectations of Y_1, Y_2, and $Y_3(R_3, r_3^*)$ conditional on $R_{12} = r_{12}$ to get

$$\begin{aligned}
\boldsymbol{E}(Y_1 \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 \mid r_{12})x_1 + \boldsymbol{E}(C_1 \mid r_{12})z_1 \\
\boldsymbol{E}(Y_2 \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 \mid r_{12})x_2 + \boldsymbol{E}(C_1 \mid r_{12})z_2 \\
&+ \boldsymbol{E}U_2 + \boldsymbol{E}V_2x_2 + \boldsymbol{E}W_2z_2 \\
\boldsymbol{E}(Y_3(\cdot, r_3^*) \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 \mid r_{12})x_3^* + \boldsymbol{E}(C_1 \mid r_{12})z_3^* \\
&+ \boldsymbol{E}U_2 + \boldsymbol{E}V_2x_3^* + \boldsymbol{E}W_2z_3^* \\
&+ \boldsymbol{E}U_3 + \boldsymbol{E}V_3x_3^* + \boldsymbol{E}W_3z_3^*\,.
\end{aligned}$$

We have achieved separation and can identify the special expectation $\boldsymbol{E}(Y_3(\cdot, r_3^*) \mid r_{12})$. Deduce that we have 3 linear equations in the 3 unknowns $\boldsymbol{E}(A_1 \mid r_{12})$, $\boldsymbol{E}(B_1 \mid r_{12})$, and $\boldsymbol{E}(C_1 \mid r_{12})$. By the determinant condition in **A2** we can solve this system on a set of probability one and so can identify $\boldsymbol{E}(A_1 \mid r_{12})$, $\boldsymbol{E}(B_1 \mid r_{12})$, and $\boldsymbol{E}(C_1 \mid r_{12})$. By averaging over R_{12} we can identify $\boldsymbol{E}A_1$, $\boldsymbol{E}B_1$, and $\boldsymbol{E}C_1$. Since all the shock moments are identified, we can identify $\boldsymbol{E}A_2$, $\boldsymbol{E}B_2$, and $\boldsymbol{E}C_2$ and $\boldsymbol{E}A_3$, $\boldsymbol{E}B_3$, and $\boldsymbol{E}C_3$ as well.

So, the questions is: can we find a fixed point function? It turns out that we can if we restrict the random coefficients to be linear regressions in R_{12} and R_3.

Consider the following model:

$$\begin{aligned} A_3 &= \alpha_0 + R_{12}\alpha + R_3 a + \epsilon_A \\ B_3 &= \beta_0 + R_{12}\beta + R_3 b + \epsilon_B \\ C_3 &= \gamma_0 + R_{12}\gamma + R_3 c + \epsilon_C \end{aligned}$$

for any constants α_0, β_0, γ_0, any 4×1 vectors α, β, and γ, and any 2×1 vectors a, b, and c where all the errors have mean zero conditional on R_{12} and R_3. Note that this model is consistent with the original model assumptions. For example, it follows from **A0** and **A1** that $A_3 = U_0 + U_1 + U_2 + U_3$ and that U_3 is independent of $(U_0, U_1, U_2, R_{12}, R_3)$. Thus, we can take $\epsilon_A = U_3 - \boldsymbol{E}U_3$ and write $U_3 = \boldsymbol{E}U_3 + \epsilon_A$. Since ϵ_A is independent of R_{12} and R_3, ϵ_A must have mean zero conditional on R_{12} and R_3. Moreover, $\boldsymbol{E}U_3$ can be absorbed into α_0. Assumption **A1** also allows A_3 to depend on R_{12} and R_3 through past shocks, and so a model linear in R_{12} and R_3 is also consistent with the original model. Argue similarly for B_3 and C_3.

Recall that $Y_3 = A_3 + B_3X_3 + C_3Z_3 = Y_3(R_3, R_3)$. Choose starting functions $r_3^0 = (x_3^0, z_3^0)$ where $x_3^0 = \boldsymbol{E}(X_3 \mid r_{12})$ and $z_3^0 = \boldsymbol{E}(Z_3 \mid r_{12})$.

Then, by direct calculation (and, for ease of notation, we take all the intercepts to be zero), we get

$$\begin{aligned} \boldsymbol{E}(Y_3(\cdot, r_3^0) \mid r_{12}) &= \boldsymbol{E}(A_3 + B_3x_3^0 + C_3z_3^0 \mid r_{12}) \\ &= \boldsymbol{E}(A_3 \mid r_{12}) + \boldsymbol{E}(B_3 \mid r_{12})x_3^0 + \boldsymbol{E}(C_3 \mid r_{12})z_3^0 \\ &= r_{12}\alpha + \boldsymbol{E}(R_3 \mid r_{12})a + [r_{12}\beta + \boldsymbol{E}(R_3 \mid r_{12})b]\, x_3^0 \\ &+ [r_{12}\gamma + \boldsymbol{E}(R_3 \mid r_{12})c]\, z_3^0 \\ &= r_{12}\alpha + r_3^0 a + \left[r_{12}\beta + r_3^0 b\right] x_3^0 + \left[r_{12}\gamma + r_3^0 c\right] z_3^0 . \end{aligned}$$

The LIE and the IMVT imply there exists $r_3^1 = (x_3^1(r_{12}), z_3^1(r_{12}))$ such that

$$\begin{aligned} \boldsymbol{E}(Y_3(\cdot, r_3^0) \mid r_{12}) &= \boldsymbol{E}(Y_3(r_3^1, r_3^0) \mid r_{12}, r_3^1) = \boldsymbol{E}(Y_3(r_3^1, r_3^0) \mid r_{12}) \\ &= r_{12}\alpha + r_3^1 a + \Big[r_{12}\beta + r_3^1 b\Big]\, x_3^0 + \Big[r_{12}\gamma + r_3^1 c\Big]\, z_3^0 . \end{aligned}$$

Therefore, we may take $r_3^1 = r_3^0$ and conclude that $r_3^0 = r_3^*$.

In other words, the natural starting functions turn out to be fixed point functions.

We now consider the case $k = 4$ and $T = 4$. The generalization to case $k = 4$ and $T \geq 4$ is immediate by considering all sets of 4 consecutive time periods of data. For simplicity, we consider the case where the number of nonrandom coefficients in the model is zero. For time periods $t = 1, 2, 3, 4$, define

$$\begin{aligned} Y_t &= A_t + B_tX_t + C_tZ_t + D_tL_t \\ A_t &= A_{t-1} + U_t \\ B_t &= B_{t-1} + V_t \\ C_t &= C_{t-1} + W_t \\ D_t &= D_{t-1} + M_t \end{aligned}$$

where $A_0 \equiv U_0$ is an initializing random intercept, $B_0 \equiv V_0$, $C_0 \equiv W_0$, and $D_0 \equiv M_0$ are initializing scalar random slope coefficients, and X_t, Z_t, and L_t are observed scalar regressors. The (U_t, V_t, W_t, M_t)s can be viewed as technology or preference shocks. As before, for $t = 1, 2, 3, 4$,

$$A_t = \sum_{s=0}^{t} U_s, \quad B_t = \sum_{s=0}^{t} V_s, \quad C_t = \sum_{s=0}^{t} W_s, \quad \text{and} \quad D_t = \sum_{s=0}^{t} M_s. \tag{2}$$

Again, the random coefficients are sums of past and current shocks.

Write S_t for shocks (U_t, V_t, W_t, M_t) and R_t for regressors (X_t, Z_t, L_t), $t = 1, 2, 3, 4$. Define the 4×4 matrix M_4 as follows:

$$M_4 = \begin{bmatrix} 1 & X_1 & Z_1 & L_1 \\ 1 & X_2 & Z_2 & L_2 \\ 1 & X_3 & Z_3 & L_3 \\ 1 & X_4 & Z_4 & L_4 \end{bmatrix} = \begin{bmatrix} 1 & R_1 \\ 1 & R_2 \\ 1 & R_3 \\ 1 & R_4 \end{bmatrix}$$

Finally, let S_{R_t} denote the support of R_t, $t = 1, 2, 3, 4$.

We make the following assumptions for $t = 1, 2, 3, 4$:

A0. The linear panel data model defined above holds.

A1. $S_t \perp (S_0, \ldots, S_{t-1}, R_1, \ldots, R_t)$.

A2. R_t is continuously distributed, $S_{R_1} = S_{R_2} = S_{R_3} = S_{R_4}$, and $\boldsymbol{P}\{\det M_4 = 0\} = 0$.

A3. For $t = 2, 3, 4$, $\boldsymbol{E}(1, R_t)'(1, R_t)$ is invertible.

Assumption **A1** says that current shocks are independent of past shocks as well as current and past regressors, but allows regressors in future periods to be correlated with past shocks. Assumption **A1** can be interpreted to mean that it takes at least one period for decision makers to react to current shocks when choosing levels of regressors. This assumption also allows lagged values of the dependent variable of any order to be regressors. The other assumptions are used as in the case $T = 3$.

It follows from assumption **A0** that

$$\begin{aligned} Y_1 &= A_1 + B_1 X_1 + C_1 Z_1 + D_1 L_1 \\ Y_2 &= A_1 + B_1 X_2 + C_1 Z_2 + D_1 L_2 \\ &+ U_2 + V_2 X_2 + W_2 Z_2 + M_2 L_2 \\ Y_3 &= A_1 + B_1 X_3 + C_1 Z_3 + D_1 L_3 \\ &+ (U_2 + U_3) + (V_2 + V_3) X_3 + (W_2 + W_3) Z_3 + (M_2 + M_3) L_3 \\ Y_4 &= A_1 + B_1 X_4 + C_1 Z_4 + D_1 L_4 \\ &+ (U_2 + U_3 + U_4) + (V_2 + V_3 + V_4) X_4 + (W_2 + W_3 + W_4) Z_4 + (M_2 + M_3 + M_4) L_4 . \end{aligned}$$

Invoke the common support condition in **A2** to see that

$$\begin{aligned} \boldsymbol{E}(Y_2 - Y_1 \mid X_2 = X_1 = x_2, Z_2 = Z_1 = z_2, L_2 - L_1 = l_2) &= \boldsymbol{E}U_2 + \boldsymbol{E}V_2 x_2 + \boldsymbol{E}W_2 z_2 + \boldsymbol{E}M_2 l_2 \\ \boldsymbol{E}(Y_3 - Y_2 \mid X_3 = X_2 = x_3, Z_3 = Z_2 = z_3, L_3 - L_2 = l_3) &= \boldsymbol{E}U_3 + \boldsymbol{E}V_3 x_3 + \boldsymbol{E}W_3 z_3 + \boldsymbol{E}M_3 l_3 \\ \boldsymbol{E}(Y_4 - Y_3 \mid X_4 = X_3 = x_4, Z_4 = Z_3 = z_4, L_4 - L_3 = l_4) &= \boldsymbol{E}U_4 + \boldsymbol{E}V_4 x_4 + \boldsymbol{E}W_4 z_4 + \boldsymbol{E}M_4 l_4 . \end{aligned}$$

Invoke **A3**, and use standard regression arguments with stayers (those sampling units whose regressor values do not change from period to period) to identify all 12 shock expectations in the last display.

Recall that $R_t = (X_t, Z_t, L_t)$. Define $R_{12} = (R_1, R_2)$ and $R_{34} = (R_3, R_4)$. Abbreviate $\boldsymbol{E}(Y_t \mid R_{12} = r_{12}, R_{34} = r_{34})$ to $\boldsymbol{E}(Y_t \mid r_{12}, r_{34})$. Averaging the Y_ts conditional on $(R_{12}, R_{34}) = (r_{12}, r_{34})$ and applying **A1** yields the following system of equations:

$$
\begin{aligned}
\boldsymbol{E}(Y_1 \mid r_{12}, r_{34}) &= \boldsymbol{E}(A_1 \mid r_{12}, r_{34}) + \boldsymbol{E}(B_1 \mid r_{12}, r_{34})x_1 + \boldsymbol{E}(C_1 \mid r_{12}, r_{34})z_1 + \boldsymbol{E}(D_1 \mid r_{12}, r_{34})l_1 \\
\boldsymbol{E}(Y_2 \mid r_{12}, r_{34}) &= \boldsymbol{E}(A_1 \mid r_{12}, r_{34}) + \boldsymbol{E}(B_1 \mid r_{12}, r_{34})x_2 + \boldsymbol{E}(C_1 \mid r_{12}, r_{34})z_2 + \boldsymbol{E}(D_1 \mid r_{12}, r_{34})l_2 \\
&+ \boldsymbol{E}(U_2 \mid r_{12}, r_{34}) + \boldsymbol{E}(V_2 \mid r_{12}, r_{34})x_2 + \boldsymbol{E}(W_2 \mid r_{12}, r_{34})z_2 + \boldsymbol{E}(M_2 \mid r_{12}, r_{34})l_2 \\
\boldsymbol{E}(Y_3 \mid r_{12}, r_{34}) &= \boldsymbol{E}(A_1 \mid r_{12}, r_{34}) + \boldsymbol{E}(B_1 \mid r_{12}, r_{34})x_3 + \boldsymbol{E}(C_1 \mid r_{12}, r_{34})z_3 + \boldsymbol{E}(D_1 \mid r_{12}, r_{34})l_3 \\
&+ \boldsymbol{E}(U_2 \mid r_{12}, r_{34}) + \boldsymbol{E}(V_2 \mid r_{12}, r_{34})x_3 + \boldsymbol{E}(W_2 \mid r_{12}, r_{34})z_3 + \boldsymbol{E}(M_2 \mid r_{12}, r_{34})l_3 \\
&+ \boldsymbol{E}(U_3 \mid r_{12}, r_{34}) + \boldsymbol{E}(V_3 \mid r_{12}, r_{34})x_3 + \boldsymbol{E}(W_3 \mid r_{12}, r_{34})z_3 + \boldsymbol{E}(M_3 \mid r_{12}, r_{34})l_3 \\
\boldsymbol{E}(Y_4 \mid r_{12}, r_{34}) &= \boldsymbol{E}(A_1 \mid r_{12}, r_{34}) + \boldsymbol{E}(B_1 \mid r_{12}, r_{34})x_4 + \boldsymbol{E}(C_1 \mid r_{12}, r_{34})z_4 + \boldsymbol{E}(D_1 \mid r_{12}, r_{34})l_4 \\
&+ \boldsymbol{E}(U_2 \mid r_{12}, r_{34}) + \boldsymbol{E}(V_2 \mid r_{12}, r_{34})x_4 + \boldsymbol{E}(W_2 \mid r_{12}, r_{34})z_4 + \boldsymbol{E}(M_2 \mid r_{12}, r_{34})l_4 \\
&+ \boldsymbol{E}(U_3 \mid r_{12}, r_{34}) + \boldsymbol{E}(V_3 \mid r_{12}, r_{34})x_4 + \boldsymbol{E}(W_3 \mid r_{12}, r_{34})z_4 + \boldsymbol{E}(M_3 \mid r_{12}, r_{34})l_4 \\
&+ \boldsymbol{E}U_4 + \boldsymbol{E}V_4x_4 + \boldsymbol{E}W_4z_4 + \boldsymbol{E}M_4l_4
\end{aligned}
$$

If we can identify $\boldsymbol{E}(A_1 \mid r_{12}, r_{34})$, $\boldsymbol{E}(B_1 \mid r_{12}, r_{34})$, $\boldsymbol{E}(C_1 \mid r_{12}, r_{34})$, and $\boldsymbol{E}(D_1 \mid r_{12}, r_{34})$, then we can achieve the objective of identifying $\boldsymbol{E}A_1$, $\boldsymbol{E}B_1$, $\boldsymbol{E}C_1$, and $\boldsymbol{E}D_1$ by averaging over (R_{12}, R_{34}). However, above we have a system of 4 linear equations in the 12 unknowns $\boldsymbol{E}(A_1 \mid r_{12}, r_{34})$, $\boldsymbol{E}(B_1 \mid r_{12}, r_{34})$, $\boldsymbol{E}(C_1 \mid r_{12}, r_{34})$, $\boldsymbol{E}(D_1 \mid r_{12}, r_{34})$, $\boldsymbol{E}(U_2 \mid r_{12}, r_{34})$, $\boldsymbol{E}(V_2 \mid r_{12}, r_{34})$, $\boldsymbol{E}(W_2 \mid r_{12}, r_{34})$, $\boldsymbol{E}(M_2 \mid r_{12}, r_{34})$, $\boldsymbol{E}(U_3 \mid r_{12}, r_{34})$, $\boldsymbol{E}(V_3 \mid r_{12}, r_{34})$, $\boldsymbol{E}(W_3 \mid r_{12}, r_{34})$, and $\boldsymbol{E}(M_3 \mid r_{12}, r_{34})$, an underidentified system. Extra unknowns arise because **A1** allows R_{34} to be correlated with all the second period shocks, and allows R_4 to be correlated with all the third period shocks.

Alternatively, if we can identify, say, $\boldsymbol{E}(A_1 \mid r_{12})$, $\boldsymbol{E}(B_1 \mid r_{12})$, $\boldsymbol{E}(C_1 \mid r_{12})$, and $\boldsymbol{E}(D_1 \mid r_{12})$, then we can identify $\boldsymbol{E}A_1$, $\boldsymbol{E}B_1$, $\boldsymbol{E}C_1$, and $\boldsymbol{E}D_1$ by averaging over R_{12}. If we average the Y_ts conditional on $R_{12} = r_{12}$ and apply **A1** we get the following system of equations:

$$
\begin{aligned}
\boldsymbol{E}(Y_1 \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 \mid r_{12})x_1 + \boldsymbol{E}(C_1 \mid r_{12})z_1 + \boldsymbol{E}(D_1 \mid r_{12})l_1 \\
\boldsymbol{E}(Y_2 \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 \mid r_{12})x_2 + \boldsymbol{E}(C_1 \mid r_{12})z_2 + \boldsymbol{E}(D_1 \mid r_{12})l_2 \\
&+ \boldsymbol{E}U_2 + \boldsymbol{E}V_2x_2 + \boldsymbol{E}W_2z_2 + \boldsymbol{E}M_2l_2 \\
\boldsymbol{E}(Y_3 \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1X_3 \mid r_{12}) + \boldsymbol{E}(C_1Z_3 \mid r_{12}) + \boldsymbol{E}(D_1L_3 \mid r_{12}) \\
&+ \boldsymbol{E}U_2 + \boldsymbol{E}(V_2X_3 \mid r_{12}) + \boldsymbol{E}(W_2Z_3 \mid r_{12}) + \boldsymbol{E}(M_2L_3 \mid r_{12}) \\
&+ \boldsymbol{E}U_3 + \boldsymbol{E}(V_3X_3 \mid r_{12}) + \boldsymbol{E}(W_3Z_3 \mid r_{12}) + \boldsymbol{E}(M_3L_3 \mid r_{12}) \\
\boldsymbol{E}(Y_4 \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1X_4 \mid r_{12}) + \boldsymbol{E}(C_1Z_4 \mid r_{12}) + \boldsymbol{E}(D_1L_4 \mid r_{12}) \\
&+ \boldsymbol{E}U_2 + \boldsymbol{E}(V_2X_4 \mid r_{12}) + \boldsymbol{E}(W_2Z_4 \mid r_{12}) + \boldsymbol{E}(M_2L_4 \mid r_{12}) \\
&+ \boldsymbol{E}U_3 + \boldsymbol{E}(V_3X_4 \mid r_{12}) + \boldsymbol{E}(W_3Z_4 \mid r_{12}) + \boldsymbol{E}(M_3L_4 \mid r_{12}) \\
&+ \boldsymbol{E}U_4 + \boldsymbol{E}V_4\boldsymbol{E}(X_4 \mid r_{12}) + \boldsymbol{E}W_4\boldsymbol{E}(Z_4 \mid r_{12}) + \boldsymbol{E}M_4\boldsymbol{E}(L_4 \mid r_{12}) .
\end{aligned}
$$

This operation yields a system of 4 linear equations in 22 unknowns: the four principal unknowns $\boldsymbol{E}(A_1 \mid r_{12})$, $\boldsymbol{E}(B_1 \mid r_{12})$, $\boldsymbol{E}(C_1 \mid r_{12})$, and $\boldsymbol{E}(D_1 \mid r_{12})$, as well as the other 18 unidentified conditional expectations of products given above. This is a highly underidentified system! Here the 18 extra unknowns arise when both factors in a product are functions of R_3 or R_4 and so are mixed together when taking expectations given $R_{12} = r_{12}$.

As before, functional fixed point methods will enable us to obtain a system of 4 equations in the 4 unknowns $\boldsymbol{E}(A_1 \mid r_{12})$, $\boldsymbol{E}(B_1 \mid r_{12})$, $\boldsymbol{E}(C_1 \mid r_{12})$, and $\boldsymbol{E}(D_1 \mid r_{12})$. The idea is to replace the components of r_3 and r_4 with functions of r_{12} so that separation between factors of products can be achieved when taking expectations given r_{12}. Simultaneously, we must find identifiable substitutes for $\boldsymbol{E}(Y_3 \mid r_{12})$ and $\boldsymbol{E}(Y_4 \mid r_{12})$ corresponding to these special functions of r_{12}.

To this end, recall that

$$\begin{aligned} Y_3 &= A_3 + B_3 X_3 + C_3 Z_3 + D_3 L_3 \\ Y_4 &= A_4 + B_4 X_4 + C_4 Z_4 + D_4 L_4 \,. \end{aligned}$$

Start with Y_3 and note that Y_3 depends on R_3 implicitly through (A_3, B_3, C_3, D_3) and explicitly through (X_3, Z_3, L_3). So, we may write $Y_3 \equiv Y_3(R_3, R_3)$ where the first R_3 refers to implicit dependence and the second R_3 to explicit dependence. Similarly, we may write $Y_4 \equiv Y_4(R_4, R_4)$. Note that $Y_3(s,t)$ and $Y_4(s,t)$, where both s and t are realizations of R_3 and R_4, respectively, can be realizations from the model if and only if $s = t$.

Let $r_3^0 = (x_3^0, z_3^0, l_3^0) \equiv (x_3^0(r_{12}), z_3^0(r_{12}), l_3^0(r_{12}))$ denote a function of r_{12} that is also a possible realization of $R_3 \mid r_{12}$. Similarly, let $r_4^0 = (x_4^0, z_4^0, l_4^0) \equiv (x_4^0(r_{12}), z_4^0(r_{12}), l_4^0(r_{12}))$ denote a function of r_{12} that is also a possible realization of $R_4 \mid r_{12}$. For example, we might take

$$\begin{aligned} r_3^0 &= \boldsymbol{E}(R_3 \mid r_{12}) = (\boldsymbol{E}(X_3 \mid r_{12}), \boldsymbol{E}(Z_3 \mid r_{12}), \boldsymbol{E}(L_3 \mid r_{12})) \\ r_4^0 &= \boldsymbol{E}(R_4 \mid r_{12}) = (\boldsymbol{E}(X_4 \mid r_{12}), \boldsymbol{E}(Z_4 \mid r_{12}), \boldsymbol{E}(L_4 \mid r_{12})) \,. \end{aligned}$$

Consider the objects

$$\begin{aligned} \boldsymbol{E}(Y_3(\cdot, r_3^0) \mid r_{12}) &= \boldsymbol{E}(A_3 + B_3 x_3^0 + C_3 z_3^0 + D_3 l_3^0 \mid r_{12}) \\ \boldsymbol{E}(Y_4(\cdot, r_4^0) \mid r_{12}) &= \boldsymbol{E}(A_4 + B_4 x_4^0 + C_4 z_4^0 + D_4 l_4^0 \mid r_{12}) \,. \end{aligned}$$

By the LIE (law of iterated expectations),

$$\begin{aligned} \boldsymbol{E}(Y_3(\cdot, r_3^0) \mid r_{12}) &= \boldsymbol{E}\left[\boldsymbol{E}(Y_3(r_3, r_3^0) \mid r_{12}, r_3) \mid r_{12}\right] \\ \boldsymbol{E}(Y_4(\cdot, r_4^0) \mid r_{12}) &= \boldsymbol{E}\left[\boldsymbol{E}(Y_4(r_4, r_4^0) \mid r_{12}, r_4) \mid r_{12}\right] \end{aligned}$$

where the outer expectations on the right are over r_3 given r_{12} and r_4 given r_{12}, respectively.

By the IMVT (integral mean value theorem), there exist functions $r_3^1 \equiv r_3^1(r_{12})$ and $r_4^1 \equiv r_4^1(r_{12})$ such that

$$\begin{aligned} \boldsymbol{E}(Y_3(\cdot, r_3^0) \mid r_{12}) &= \boldsymbol{E}(Y_3(r_3^1, r_3^0) \mid r_{12}, r_3^1) \\ \boldsymbol{E}(Y_4(\cdot, r_4^0) \mid r_{12}) &= \boldsymbol{E}(Y_4(r_4^1, r_4^0) \mid r_{12}, r_4^1) \,. \end{aligned}$$

The IMVT steps defines mappings:

$$\begin{aligned} m(r_3^0) &= r_3^1 \\ \mu(r_4^0) &= r_4^1 \,. \end{aligned}$$

Consider the sequences of mappings

$$\begin{aligned} r_3^i &= m(r_3^{i-1}) \quad i = 1, 2, \ldots \\ r_4^i &= \mu(r_4^{i-1}) \quad i = 1, 2, \ldots \end{aligned}$$

Do these sequences of functions have fixed points?

Suppose for the moment that such fixed point functions exist, and call them

$$\begin{aligned} r_3^* &= (x_3^*, z_3^*, l_3^*) \equiv (x_3^*(r_{12}), z_3^*(r_{12}), l_3^*(r_{12})) \\ r_4^* &= (x_4^*, z_4^*, l_4^*) \equiv (x_4^*(r_{12}), z_4^*(r_{12}), l_4^*(r_{12})) \,. \end{aligned}$$

This means that

$$\begin{aligned} \boldsymbol{E}(Y_3(\cdot, r_3^*) \mid r_{12}) &= \boldsymbol{E}(Y_3(r_3^*, r_3^*) \mid r_{12}, r_3^*) \\ \boldsymbol{E}(Y_4(\cdot, r_4^*) \mid r_{12}) &= \boldsymbol{E}(Y_4(r_4^*, r_4^*) \mid r_{12}, r_4^*) \,. \end{aligned}$$

If we knew r_3^* and r_4^* we would be able to estimate the LHS quantities using the RHS quantities: average the observed Y_3s local to r_{12} and r_3^* and average the observed Y_4s local to r_{12} and r_4^*.

Now return to the original system of equations but this time replace $Y_3 \equiv Y_3(R_3, R_3)$ with $Y_3(R_3, r_3^*)$ and $Y_4 \equiv Y_4(R_4, R_4)$ with $Y_4(R_4, r_4^*)$. Then take expectations of Y_1, Y_2, $Y_3(R_3, r_3^*)$ and $Y_4(R_4, r_4^*)$ conditional on $R_{12} = r_{12}$ to get

$$\begin{aligned} \boldsymbol{E}(Y_1 \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 \mid r_{12})x_1 + \boldsymbol{E}(C_1 \mid r_{12})z_1 + \boldsymbol{E}(D_1 \mid r_{12}l_1 \\ \boldsymbol{E}(Y_2 \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 \mid r_{12})x_2 + \boldsymbol{E}(C_1 \mid r_{12})z_2 + \boldsymbol{E}(D_1 \mid r_{12}l_2 \\ &+ \boldsymbol{E}U_2 + \boldsymbol{E}V_2x_2 + \boldsymbol{E}W_2z_2 + \boldsymbol{E}M_2l_2 \\ \boldsymbol{E}(Y_3(\cdot, r_3^*) \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 \mid r_{12})x_3^* + \boldsymbol{E}(C_1 \mid r_{12})z_3^* + \boldsymbol{E}(D_1 \mid r_{12})l_3^* \\ &+ \boldsymbol{E}U_2 + \boldsymbol{E}V_2x_3^* + \boldsymbol{E}W_2z_3^* + \boldsymbol{E}M_2l_3^* \\ &+ \boldsymbol{E}U_3 + \boldsymbol{E}V_3x_3^* + \boldsymbol{E}W_3z_3^* + \boldsymbol{E}M_3l_3^* \\ \boldsymbol{E}(Y_4(\cdot, r_4^*) \mid r_{12}) &= \boldsymbol{E}(A_1 \mid r_{12}) + \boldsymbol{E}(B_1 \mid r_{12})x_4^* + \boldsymbol{E}(C_1 \mid r_{12})z_4^* + \boldsymbol{E}(D_1 \mid r_{12})l_4^* \\ &+ \boldsymbol{E}U_2 + \boldsymbol{E}V_2x_4^* + \boldsymbol{E}W_2z_4^* + \boldsymbol{E}M_2l_4^* \\ &+ \boldsymbol{E}U_3 + \boldsymbol{E}V_3x_4^* + \boldsymbol{E}W_3z_4^* + \boldsymbol{E}M_3l_4^* \\ &+ \boldsymbol{E}U_4 + \boldsymbol{E}V_4x_4^* + \boldsymbol{E}W_3z_4^* + \boldsymbol{E}M_3l_4^* \,. \end{aligned}$$

We have achieved separation and can identify the special expectations $\boldsymbol{E}(Y_3(\cdot, r_3^*) \mid r_{12})$ and $\boldsymbol{E}(Y_4(\cdot, r_4^*) \mid r_{12})$. Deduce that we have 4 linear equations in the 4 unknowns $\boldsymbol{E}(A_1 \mid r_{12})$, $\boldsymbol{E}(B_1 \mid r_{12})$, $\boldsymbol{E}(C_1 \mid r_{12})$, and $\boldsymbol{E}(D_1 \mid r_{12})$. By the determinant condition in **A2** we can solve this system on a set of probability one and so can identify $\boldsymbol{E}(A_1 \mid r_{12})$, $\boldsymbol{E}(B_1 \mid r_{12})$, $\boldsymbol{E}(C_1 \mid r_{12})$ and $\boldsymbol{E}(D_1 \mid r_{12})$. By averaging over R_{12} we can identify $\boldsymbol{E}A_1$, $\boldsymbol{E}B_1$, $\boldsymbol{E}C_1$ and $\boldsymbol{E}D_1$. Since all the shock moments are identified, we can identify $\boldsymbol{E}A_2$, $\boldsymbol{E}B_2$, $\boldsymbol{E}C_2$ and $\boldsymbol{E}D_2$, $\boldsymbol{E}A_3$, $\boldsymbol{E}B_3$, $\boldsymbol{E}C_3$, and $\boldsymbol{E}D_3$, and $\boldsymbol{E}A_4$, $\boldsymbol{E}B_4$, $\boldsymbol{E}C_4$, and $\boldsymbol{E}D_4$ as well.

So, the questions is: can we find the fixed point functions?

Consider the following model:

$$\begin{aligned} A_3 &= \alpha_0 + R_{12}\alpha + R_{34}a + \epsilon_A \\ B_3 &= \beta_0 + R_{12}\beta + R_{34}b + \epsilon_B \\ C_3 &= \gamma_0 + R_{12}\gamma + R_{34}c + \epsilon_C \\ D_3 &= \delta_0 + R_{12}\delta + R_{34}d + \epsilon_D \\ A_4 &= \tilde{\alpha}_0 + R_{12}\tilde{\alpha} + R_{34}\tilde{a} + \epsilon_{\tilde{A}} \\ B_4 &= \tilde{\beta}_0 + R_{12}\tilde{\beta} + R_{34}\tilde{b} + \epsilon_{\tilde{B}} \\ C_4 &= \tilde{\gamma}_0 + R_{12}\tilde{\gamma} + R_{34}\tilde{c} + \epsilon_{\tilde{C}} \\ D_4 &= \tilde{\delta}_0 + R_{12}\tilde{\delta} + R_{34}\tilde{d} + \epsilon_{\tilde{D}} \end{aligned}$$

for any constants α_0, β_0, γ_0, δ_0, $\tilde{\alpha}_0$, $\tilde{\beta}_0$, $\tilde{\gamma}_0$, $\tilde{\delta}_0$, any 6×1 vectors α, β, γ, δ, $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\gamma}$, $\tilde{\delta}$, and any 6×1 vectors a, b, c, d, $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, and $\tilde{d}$ where all the errors have mean zero conditional on R_{12} and R_{34}. Note that this model is consistent with the original model assumptions. For example, it follows from **A0** and **A1** that $A_4 = U_0 + U_1 + U_2 + U_3 + U_4$ and that U_4 is independent of $(U_0, U_1, U_2, U_3, R_{12}, R_{34})$. Thus, we can write $U_4 = \boldsymbol{E}U_4 + \epsilon_{\tilde{A}}$ where $\epsilon_{\tilde{A}} = U_4 - \boldsymbol{E}U_4$. Since $\epsilon_{\tilde{A}}$ is independent of R_{12} and R_{34}, $\epsilon_{\tilde{A}}$ must have mean zero conditional on R_{12} and R_{34}. Moreover, $\boldsymbol{E}U_4$ can be absorbed into $\tilde{\alpha}_0$. Assumption **A1** also allows A_4 to depend on R_{12} and R_{34} through past shocks, and so a model linear in R_{12} and R_{34} is also consistent with the original model. Similarly for B_4, C_4, and D_4. What about the third period random coefficients? For example, take $A_3 = U_0 + U_1 + U_2 + U_3$. Since the model allows R_4 to be correlated with all past shocks, the argument above for A_4 does not apply to A_3 without further restrictions. For example, suppose that one of the shocks that comprise A_3, say, U_0, can be decomposed into a sum $P_0 + Q_0$, where P_0 is allowed to be correlated with (R_{12}, R_{34}), but $\boldsymbol{E}(Q_0 \mid R_{12}, R_{34}) = 0$. Then Q_0 plays the role of $U_4 - \boldsymbol{E}U_4$ in the argument above for A_4 and the argument goes through for A_3. Similarly for B_3, C_3, and D_3. As another example, if all parameters and regressors are jointly normally distributed, then the conditional mean of each parameter given regressors is normal and can be written as its conditional mean plus an error that is independent of all regressors and thus is consistent with the original model.

Recall that

$$\begin{aligned} Y_3 &= A_3 + B_3X_3 + C_3Z_3 + D_3L_3 = Y_3(R_3, R_3) \\ Y_4 &= A_4 + B_4X_4 + C_4Z_4 + D_4L_4 = Y_4(R_4, R_4)\,. \end{aligned}$$

Choose starting functions $r_3^0 = (x_3^0, z_3^0, l_3^0)$ where $x_3^0 = \boldsymbol{E}(X_3 \mid r_{12})$, $z_3^0 = \boldsymbol{E}(Z_3 \mid r_{12})$, and $l_3^0 = \boldsymbol{E}(L_3 \mid r_{12})$, and $r_4^0 = (x_4^0, z_4^0, l_4^0)$ where $x_4^0 = \boldsymbol{E}(X_4 \mid r_{12})$, $z_4^0 = \boldsymbol{E}(Z_4 \mid r_{12})$, and $l_4^0 = \boldsymbol{E}(L_4 \mid r_{12})$. Then, by direct calculation (and, for ease of notation, we take all the intercepts to be zero), we get

$$\begin{aligned} \boldsymbol{E}(Y_3(\cdot, r_3^0) \mid r_{12}) &= \boldsymbol{E}(A_3 + B_3x_3^0 + C_3z_3^0 + D_3l_3^0 \mid r_{12}) \\ &= \boldsymbol{E}(A_3 \mid r_{12}) + \boldsymbol{E}(B_3 \mid r_{12})x_3^0 + \boldsymbol{E}(C_3 \mid r_{12})z_3^0 + \boldsymbol{E}(D_3 \mid r_{12})l_3^0 \\ &= r_{12}\alpha + \boldsymbol{E}(R_{34} \mid r_{12})a + [r_{12}\beta + \boldsymbol{E}(R_{34} \mid r_{12})b]\, x_3^0 \\ &+ [r_{12}\gamma + \boldsymbol{E}(R_{34} \mid r_{12})c]\, z_3^0 + [r_{12}\delta + \boldsymbol{E}(R_{34} \mid r_{12})d]\, l_3^0 \\ &= r_{12}\alpha + (r_3^0, r_4^0)a + [r_{12}\beta + (r_3^0, r_4^0)b]\, x_3^0 + [r_{12}\gamma + (r_3^0, r_4^0)c]\, z_3^0 + [r_{12}\delta + (r_3^0, r_4^0)d]\, l_3^0 \\ \boldsymbol{E}(Y_4(\cdot, r_4^0) \mid r_{12}) &= \boldsymbol{E}(A_4 + B_4x_4^0 + C_4z_4^0 + D_4l_4^0 \mid r_{12}) \\ &= \boldsymbol{E}(A_4 \mid r_{12}) + \boldsymbol{E}(B_4 \mid r_{12})x_4^0 + \boldsymbol{E}(C_4 \mid r_{12})z_4^0 + \boldsymbol{E}(D_4 \mid r_{12})l_4^0 \\ &= r_{12}\tilde{\alpha} + \boldsymbol{E}(R_{34} \mid r_{12})\tilde{a} + [r_{12}\tilde{\beta} + \boldsymbol{E}(R_{34} \mid r_{12})\tilde{b}]\, x_4^0 \\ &+ [r_{12}\tilde{\gamma} + \boldsymbol{E}(R_{34} \mid r_{12})\tilde{c}]\, z_4^0 + [r_{12}\tilde{\delta} + \boldsymbol{E}(R_{34} \mid r_{12})\tilde{d}]\, l_4^0 \\ &= r_{12}\tilde{\alpha} + (r_3^0, r_4^0)\tilde{a} + [r_{12}\tilde{\beta} + (r_3^0, r_4^0)\tilde{b}]\, x_4^0 + [r_{12}\tilde{\gamma} + (r_3^0, r_4^0)\tilde{c}]\, z_4^0 + [r_{12}\tilde{\delta} + (r_3^0, r_4^0)\tilde{d}]\, l_4^0\,. \end{aligned}$$

The LIE and IMVT imply $r_3^1 = (x_3^1(r_{12}), z_3^1(r_{12}), l_3^1(r_{12}))$ and $r_4^1 = (x_4^1(r_{12}), z_4^1(r_{12}), l_4^1(r_{12}))$ exist such that

$$\begin{aligned} \boldsymbol{E}(Y_3(\cdot, r_3^0) \mid r_{12}) &= \boldsymbol{E}(Y_3(r_3^1, r_3^0) \mid r_{12}, r_3^1) = \boldsymbol{E}(Y_3(r_3^1, r_3^0) \mid r_{12}) \\ &= r_{12}\alpha + (r_3^1, r_4^0)a + \left[r_{12}\beta + (r_3^1, r_4^0)b\right] x_3^0 + \left[r_{12}\gamma + (r_3^1, r_4^0)c\right] z_3^0 + \left[r_{12}\delta + (r_3^1, r_4^0)d\right] l_3^0 \\ \boldsymbol{E}(Y_4(\cdot, r_4^0) \mid r_{12}) &= \boldsymbol{E}(Y_4(r_4^1, r_4^0) \mid r_{12}, r_4^1) = \boldsymbol{E}(Y_4(r_4^1, r_4^0) \mid r_{12}) \\ &= r_{12}\tilde{\alpha} + (r_3^0, r_4^1)\tilde{a} + \left[r_{12}\tilde{\beta} + (r_3^0, r_4^1)\tilde{b}\right] x_4^0 + \left[r_{12}\tilde{\gamma} + (r_3^0, r_4^1)\tilde{c}\right] z_4^0 + \left[r_{12}\tilde{\delta} + (r_3^0, r_4^1)d\right] l_4^0\,. \end{aligned}$$

Therefore, we may take $r_3^1 = r_3^0$ in the expression for $\boldsymbol{E}(Y_3(\cdot, r_3^0) \mid r_{12})$ and $r_4^1 = r_4^0$ in the expression for $\boldsymbol{E}(Y_4(\cdot, r_4^0) \mid r_{12})$ and conclude that $r_3^0 = r_3^*$ and $r_4^0 = r_4^*$.

In other words, the natural starting functions turn out to be fixed point functions.

In the previous arguments we have considered the cases $k = 3$ and $k = 4$. What about $k \geq 5$? The model and assumptions scale immediately in obvious ways. The functional fixed point argument scales in a straightforward manner as well: replace R_3 in the $k = 3$ case and R_{34} in the $k = 4$ case with $R_{3:k}$ in the general case, where $R_{3:k} = (R_3, \ldots, R_k)$ with R_t a $k-1$ vector of nonconstant tth period regressors. Conditioning on $(R_1, R_2, R_{3:k})$ and then averaging leads to an underidentified system of k linear equations in $k(k-1)$ unknowns. It is straightforward (but tedious) to show that for $k \geq 5$, conditioning on R_{12} and then averaging leads to an underidentified system of k linear equations in $(k^3 - 7k + 8)/2$ unknowns. The functional fixed point argument leads to a system of k linear equations in k unknowns which is identified with probability one. This is achieved by taking the starting functions to be $r_t^0 = \boldsymbol{E}(R_t \mid R_{12})$, $t = 3, \ldots, k$ and modeling each random coefficient in each time period as a linear combination of all regressors in all periods plus an error that has conditional mean zero conditional on all regressors. By analogy with $k = 4$ case, this model is consistent with the original model under the following sufficient restrictions: for $t = 3, \ldots, k-1$, at least one of the shocks that make up each of the random coefficients in the tth time period can be decomposed into a sum of two terms, one of which can be correlated with all the regressors and the other which is uncorrelated with all the regressors. As in the $k = 4$ case, these sufficient restrictions are trivially satisfied if all parameters and regressors are jointly normally distributed. Under these restrictions, the starting functions are fixed point functions.

Moment Estimation

In this section, we show how to estimate the moments of the random coefficients identified in the last section. We do so by replacing expectations with sample analogues. As before, for ease of notation, we take the number of nonrandom coefficients in the model to be zero. We consider the special case $T = 3$. Recall that $R_{12} = (R_1, R_2)$ where $R_t = (X_t, Z_t)$, $t = 1, 2, 3$. Realizations of these random vectors are denoted r_{12} and r_t, respectively. The objects of interest are $\boldsymbol{E}A_t$, $\boldsymbol{E}B_t$, and $\boldsymbol{E}C_t$, $t = 1, 2, 3$. This special case captures the essential features of the general case.

Recall the definition of the 3×3 matrix $M_3 \equiv M_3(r_{12}, r_3)$ and define $M_3^* \equiv M_3(r_{12}, r_3^*)$ where $r_3^* = (x_3^*, z_3^*) = (\boldsymbol{E}(X_3 \mid r_{12}), \boldsymbol{E}(Z_3 \mid r_{12}))$, the fixed point function. Suppose that (r_{12}, r_3^*) is a realization of (R_{12}, R_3) for which $\det M_3^* \neq 0$. By **A2** the set of such points has probability one. From the identification argument we get

$$\begin{bmatrix} \boldsymbol{E}(A_1 \mid r_{12}) \\ \boldsymbol{E}(B_1 \mid r_{12}) \\ \boldsymbol{E}(C_1 \mid r_{12}) \end{bmatrix} = M_3^{*-1} \begin{bmatrix} \boldsymbol{E}(Y_1 \mid r_{12}) \\ \boldsymbol{E}(Y_2 \mid r_{12}) - c_1(r_2) \\ \boldsymbol{E}(Y_3(\cdot, r_3^*) \mid r_{12}) - c_2(r_3^*) \end{bmatrix}.$$

where

$$\begin{aligned} c_1(r_2) &= \boldsymbol{E}U_2 + \boldsymbol{E}V_2 x_2 + \boldsymbol{E}W_2 z_2 \\ c_2(r_3^*) &= \boldsymbol{E}U_2 + \boldsymbol{E}U_3 + (\boldsymbol{E}V_2 + \boldsymbol{E}V_3)x_3^* + (\boldsymbol{E}W_2 + \boldsymbol{E}W_3)z_3^*. \end{aligned}$$

We estimate $(\boldsymbol{E}U_2, \boldsymbol{E}V_2, \boldsymbol{E}W_2)$ and $(\boldsymbol{E}U_3, \boldsymbol{E}V_3, \boldsymbol{E}W_3)$ with the local regression estimators:

$$\begin{aligned} (\widehat{\boldsymbol{E}}U_2, \widehat{\boldsymbol{E}}V_2, \widehat{\boldsymbol{E}}W_2) &= \underset{u,v,w}{\operatorname{argmin}} \sum_{j=1}^{n} \left(Y_{2j} - Y_{1j} - u - vX_{2j} - wZ_{2j}\right)^2 \{|R_{2j} - R_{1j}| \leq h_n\} \\ (\widehat{\boldsymbol{E}}U_3, \widehat{\boldsymbol{E}}V_3, \widehat{\boldsymbol{E}}W_3) &= \underset{u,v,w}{\operatorname{argmin}} \sum_{j=1}^{n} \left(Y_{3j} - Y_{2j} - u - vX_{3j} - wZ_{3j}\right)^2 \{|R_{3j} - R_{2j}| \leq h_n\} \end{aligned}$$

where h_n is a bandwidth satisfying $h_n \to 0$ as $n \to \infty$. For example, $h_n \propto n^{-1/6}$ is the approximately optimal bandwidth in the sense of minimizing mean squared error for two-dimensional nonparametric regression estimators.

We could estimate the components of the fixed point function $r_3^* = (\boldsymbol{E}(X_3 \mid r_{12}), \boldsymbol{E}(Z_3 \mid r_{12}))$ and the conditional means $\boldsymbol{E}(Y_1 \mid r_{12})$ and $\boldsymbol{E}(Y_2 \mid r_{12})$ with the nonparametric regression estimators

$$\begin{aligned}
\widehat{\boldsymbol{E}}(X_3 \mid r_{12}) &= \frac{\sum_{j=1}^n X_{3j}\{|R_{12j} - r_{12}| \le \gamma_n\}}{\sum_{j=1}^n \{|R_{12j} - r_{12}| \le \gamma_n\}} \\
\widehat{\boldsymbol{E}}(Z_3 \mid r_{12}) &= \frac{\sum_{j=1}^n Z_{3j}\{|R_{12j} - r_{12}| \le \gamma_n\}}{\sum_{j=1}^n \{|R_{12j} - r_{12}| \le \gamma_n\}} \\
\widehat{\boldsymbol{E}}(Y_1 \mid r_{12}) &= \frac{\sum_{j=1}^n Y_{1j}\{|R_{12j} - r_{12}| \le \gamma_n\}}{\sum_{j=1}^n \{|R_{12j} - r_{12}| \le \gamma_n\}} \\
\widehat{\boldsymbol{E}}(Y_2 \mid r_{12}) &= \frac{\sum_{j=1}^n Y_{2j}\{|R_{12j} - r_{12}| \le \gamma_n\}}{\sum_{j=1}^n \{|R_{12j} - r_{12}| \le \gamma_n\}}
\end{aligned}$$

where γ_n is a bandwidth satisfying $\gamma_n \to 0$ as $n \to \infty$. For example, $\gamma_n \propto n^{-1/8}$ is the approximately optimal bandwidth in the sense of minimizing mean squared error for four-dimensional nonparametric regression estimators.

Define $\hat{r}_3^* \equiv (\widehat{\boldsymbol{E}}(X_3 \mid r_{12}), \widehat{\boldsymbol{E}}(Z_3 \mid r_{12}))$ and $\hat{R}_3^* \equiv (\widehat{\boldsymbol{E}}(X_3 \mid R_{12}), \widehat{\boldsymbol{E}}(Z_3 \mid R_{12}))$. We could estimate $\boldsymbol{E}(Y_3(\cdot, r_3^*) \mid r_{12}) = \boldsymbol{E}(Y_3(r_3^*, r_3^*) \mid r_{12}, r_3^*)$ with the nonparametric regression estimator

$$\widehat{\boldsymbol{E}}(Y_3(\cdot, r_3^*) \mid r_{12}) = \frac{\sum_{j=1}^n Y_{3j}\{|(R_{12j}, \hat{R}_{3j}^*) - (r_{12}, \hat{r}_3^*)| \le \lambda_n\}}{\sum_{j=1}^n \{|(R_{12j}, \hat{R}_{3j}^*) - (r_{12}, \hat{r}_3^*)| \le \lambda_n\}}$$

where λ_n is a bandwidth satisfying $\lambda_n \to 0$ as $n \to \infty$. For example, $\lambda_n \propto n^{-1/10}$ is the approximately optimal bandwidth in the sense of minimizing mean squared error for six-dimensional nonparametric regression estimators.

We note that in estimating all conditional means above, to achieve better finite sample performance we replace the local constant regression estimates with fixed bandwidths with local linear regression estimates with corresponding nearest neighbor bandwidths. For example, the four-dimensional conditional expectations use a number of neighbors proportional to $n * (n^{-1/8})^4 = n^{1/2}$ while the six-dimensional conditional expectations use a number of neighbors proportional to $n * (n^{-1/10})^6 = n^{2/5}$.

Let $\widehat{c_1}(r_2)$ and $\widehat{c_2}(r_3^*)$ denote the constants defined above with both the population shock means and the population fixed point functions replaced by their estimates. Let $\widehat{M}_3^*$ denote M_3^* with the population fixed point functions replaced by their estimates. We then estimate the vector of conditional means $(\boldsymbol{E}(A_1 \mid r_{12}), \boldsymbol{E}(B_1 \mid r_{12}), \boldsymbol{E}(C_1 \mid r_{12}))'$ with the following sample analogue:

$$\begin{bmatrix} \widehat{\boldsymbol{E}}(A_1 \mid r_{12}) \\ \widehat{\boldsymbol{E}}(B_1 \mid r_{12}) \\ \widehat{\boldsymbol{E}}(C_1 \mid r_{12}) \end{bmatrix} = \widehat{M}_3^{*-1} \begin{bmatrix} \widehat{\boldsymbol{E}}(Y_1 \mid r_{12}) \\ \widehat{\boldsymbol{E}}(Y_2 \mid r_{12}) - \widehat{c_1}(r_2) \\ \widehat{\boldsymbol{E}}(Y_3(\cdot, r_3^*) \mid r_{12}) - \widehat{c_2}(r_3^*) \end{bmatrix}. \tag{3}$$

This leads to the following estimators of first period quantities of interest:

$$\begin{aligned}
\widehat{\boldsymbol{E}}A_1 &= \frac{\sum_{i=1}^n \widehat{\boldsymbol{E}}(A_1 \mid R_{12i})\{|\det \widehat{M}_{3i}^*| > \delta_n\}}{\sum_{i=1}^n \{|\det \widehat{M}_{3i}^*| > \delta_n\}} \\
\widehat{\boldsymbol{E}}B_1 &= \frac{\sum_{i=1}^n \widehat{\boldsymbol{E}}(B_1 \mid R_{12i})\{|\det \widehat{M}_{3i}^*| > \delta_n\}}{\sum_{i=1}^n \{|\det \widehat{M}_{3i}^*| > \delta_n\}} \\
\widehat{\boldsymbol{E}}C_1 &= \frac{\sum_{i=1}^n \widehat{\boldsymbol{E}}(C_1 \mid R_{12i})\{|\det \widehat{M}_{3i}^*| > \delta_n\}}{\sum_{i=1}^n \{|\det \widehat{M}_{3i}^*| > \delta_n\}}
\end{aligned}$$

where δ_n converges to zero as $n \to \infty$. For example, $\delta_n \propto n^{-1/5}$ is the approximately optimal bandwidth in the sense of minimizing mean square error for a one-dimensional nonparametric regression estimator. We get similar estimators for corresponding second and third period quantities.

Finally, we note that a simpler estimator of $(\boldsymbol{E}A_1, \boldsymbol{E}B_1, \boldsymbol{E}C_1)'$ is possible. To see this, recall that

$$\begin{bmatrix} \boldsymbol{E}(A_1 \mid r_{12}) \\ \boldsymbol{E}(B_1 \mid r_{12}) \\ \boldsymbol{E}(C_1 \mid r_{12}) \end{bmatrix} = M_3^{*-1} \begin{bmatrix} \boldsymbol{E}(Y_1 \mid r_{12}) \\ \boldsymbol{E}(Y_2 \mid r_{12}) - c_1(r_2) \\ \boldsymbol{E}(Y_3(\cdot, r_3^*) \mid r_{12}) - c_2(r_3^*) \end{bmatrix}.$$

For convenience, define

$$\boldsymbol{E}(Y \mid r_{12}) \equiv \begin{bmatrix} \boldsymbol{E}(Y_1 \mid r_{12}) \\ \boldsymbol{E}(Y_2 \mid r_{12}) - c_1(r_2) \\ \boldsymbol{E}(Y_3(\cdot, r_3^*) \mid r_{12}) - c_2(r_3^*) \end{bmatrix} \quad \text{and} \quad Y_0 \equiv \begin{bmatrix} Y_1 \\ Y_2 - c_1(r_2) \\ Y_3 - c_2(r_3^*) \end{bmatrix}.$$

Note that

$$\boldsymbol{E}(Y_0 \mid r_{12}) = \boldsymbol{E}(Y \mid r_{12}).$$

This is the key property that makes the identification argument work and obviates the need to estimate $\boldsymbol{E}(Y_1 \mid r_{12})$, $\boldsymbol{E}(Y_2 \mid r_{12})$, and $\boldsymbol{E}(Y_3(\cdot, r_3^*) \mid r_{12})$ with the usual nonparametric local average estimators. This leads to a consistent estimator of $(\boldsymbol{E}A_1, \boldsymbol{E}B_1, \boldsymbol{E}C_1)'$ that is simpler and faster to compute. This simpler estimator is a special case of the more general local average estimator where each component of $\boldsymbol{E}(Y \mid r_{12})$ is estimated with a single inconsistent but unbiased observation, namely, the "own" observation. This is analogous to White's (1980) heteroscedasticity-consistent estimator of the least squares variance-covariance matrix, where a single inconsistent but (asymptotically) unbiased squared residual is used to estimate a variance. As with White's estimator, the inconsistency of Y_0 as an estimator of $\boldsymbol{E}(Y \mid r_{12})$ does not affect the consistency of the estimator of $(\boldsymbol{E}A_1, \boldsymbol{E}B_1, \boldsymbol{E}C_1)'$, since the estimator of the latter quantity is obtained by averaging in the last stage of estimation. We illustrate the performance of both the general and simpler estimators in the next section, where we denote the general estimator as HSS and the simpler estimator as HSS_0.

Simulations

Previously, we showed how to identify and estimate parameter moments in a linear panel data model with correlated random coefficients, subject to certain restrictions on the processes that these coefficients follow. In this section, we use simulated data to both demonstrate the viability of this approach and comment on several features of its use in applied settings. First, we discuss parameter selection and its potential impacts within our estimation procedure. Next, we discuss the performance of our estimator on two different data generating processes: a feedback model and a lagged response model in which one of the regressors is a one-period lag of the response variable. For each of these models we also compare our estimator to natural competitors, and use large-sample simulations to informally suggest the rate of convergence obtained by our methodology.

Parameter Selection

To implement our procedure, there are generally four tuning parameters that must be chosen: a proportionality constant of the bandwidth for stayer selection (γ), two bandwidths used for the non-parametric estimation of conditional expectations (h_1, h_2), and the threshold for censoring observations associated with small determinants (δ). In the case of HSS_0, only γ, h_1, and δ are needed. To achieve consistency, these parameters must converge to zero as the sample size increases. In the last column of Table 1 we indicate the rates of convergence that are optimal in the mean-squared error sense. We now discuss the role that each of these parameters plays in estimation, as well as heuristics for assigning them values.

The parameter γ helps determine the number of stayers used to calculate shock moments. An implementation could use a single γ applied to both X and Z across all time periods, however in practice this is suboptimal for cases in which the two regressors have different distributions. To this end, we propose an alternative procedure that takes regressor distributions into account. Define the quantities $d_t^X = X_{t+1} - X_t$ and $d_t^Z = Z_{t+1} - Z_t$ and let

$$r_t^X = \sqrt{\widehat{E[(d_t^X)^2]}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{i,t}^X)^2} \qquad r_t^Z = \sqrt{\widehat{E[(d_t^Z)^2]}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{i,t}^Z)^2}$$

Notice that these quantities measure the deviation of d_t^X and d_t^Z from zero and have the same units as X and Z respectively. With these definitions, we classify stayers as observations satisfying

$$|d_{k,t}^X| < \gamma \cdot r_t^X \cdot n^{-1/6} \text{ and } |d_{k,t}^Z| < \gamma \cdot r_t^Z \cdot n^{-1/6} \tag{4}$$

These bandwidths are increasing in the variance of the regressors as well as the distance between regressors in successive periods. For our simulations, we choose $\gamma = 0.74$, which is calibrated to classify as stayers roughly 2.5 percent of a sample of $n = 2500$ in a model where d_t^X and d_t^Z are independent, standard normal random variables.

A larger γ tends to increase bias and reduce variance in the estimators of the shock means. A smaller γ tends to have the opposite effect. Notice that γ is used only in the estimation of shock means, and thus the effects of its selection can be seen in the parameter estimates for $t > 1$ but not for $t = 1$.

The bandwidth h_1 is used to estimate expectations that condition on R_{12} while h_2 is used for expectations that are conditional on (R_{12}, R_3^*). Two different parameters are required here since R_{12} contains four variables while (R_{12}, R_3^*) is a six-dimensional space. In our simulations, we estimate all of these expectations using k-nearest neighbors averaging performed with Gaussian kernel weighting, instead of the fixed bandwidth, uniform kernel averaging adopted in our theoretical discussions. This methodological choice makes our estimator more numerically robust and efficient in practice, especially in limited samples with non-uniform regressor support. In unreported results, we find that our procedure is robust to the choice of h_1 and h_2, meaning that they can be varied considerably without significantly impacting the resulting estimates. Notice that these parameters affect estimation of period $t = 1$ conditional expectations and so their influence carries through to the estimates for all t.

The final parameter required by our procedure is the threshold δ, which determines which observations are excluded from the averaging used to estimate $t = 1$ marginal means. Specifically, with the matrix A defined as

$$A = \begin{bmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ 1 & x_3^* & z_3^* \end{bmatrix} \tag{5}$$

our procedure rejects observations for which $|\det A| < \delta$. We choose δ as follows:

$$\delta = 0.5 \cdot s_D \cdot n^{-1/5} \tag{6}$$

where s_D is the sample standard deviation of the determinant of A defined in equation (5).

Feedback Data Generating Process

We now describe a feedback model covered by our methods but not by methods in the previous literature.

Under this model, shocks are distributed multivariate normally according to

$$\begin{bmatrix} U_t \\ V_t \\ W_t \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{U_t} \\ \mu_{V_t} \\ \mu_{W_t} \end{bmatrix}, \begin{bmatrix} 4.0 & 1.3 & -0.7 \\ 1.3 & 1.0 & -0.7 \\ -0.7 & -0.7 & 1.0 \end{bmatrix} \right) \tag{7}$$

with the means in each period being

$$(\mu_{U_t}, \mu_{V_t}, \mu_{W_t}) = \begin{cases} (1.5, 2.5, -3.5) & \text{for } t = 0, 1 \\ (10, 8.5, -5) & \text{for } t = 2 \\ (15, 7.0, -7.0) & \text{for } t = 3 \end{cases}$$

Coefficients are then defined as the cumulative sums of shocks with

$$A_t = \sum_{s=0}^{t} U_s, \quad B_t = \sum_{s=0}^{t} V_s, \quad C_t = \sum_{s=0}^{t} W_s \tag{8}$$

Notice that shocks are correlated within each period, but independent across periods. Further, the shock means in periods 2 and 3 have relatively large magnitudes, so coefficients deviate substantially from their $t = 1$ values.

Regressors in this model respond directly to the shocks from previous periods and are correlated with coefficients from all periods. Define the intermediate variables

$$\begin{aligned} \widetilde{X}_1 &= V_0 + 0.25W_0 \\ \widetilde{X}_2 &= V_1 + 0.25W_1 + 0.15(V_0 + 0.25W_0) \\ \widetilde{X}_3 &= V_2 + 0.25W_2 + 0.15(V_1 + 0.25W_1) + 0.15^2(V_0 + 0.25W_0) \\ \widetilde{Z}_1 &= -0.25V_0 + W_0 \\ \widetilde{Z}_2 &= -0.25V_1 + W_1 + 0.15(-0.25V_0 + W_0) \\ \widetilde{Z}_3 &= -0.25V_2 + W_2 + 0.15(-0.25V_1 + W_1) + 0.15^2(-0.25V_0 + W_0) \end{aligned} \tag{9}$$

Then, the actual regressors are computed by standardizing these variables within each time period and adding an arbitrary offset. That is,

$$X_t = \frac{\widetilde{X}_t - E[\widetilde{X}_t]}{\sqrt{\text{Var}[\widetilde{X}_t]}} + 2.5, \quad Z_t = \frac{\widetilde{Z}_t - E[\widetilde{Z}_t]}{\sqrt{\text{Var}[\widetilde{Z}_t]}} - 0.5 \tag{10}$$

Here, we standardize within periods, instead of setting our regressors exactly equal to the intermediate variables, to introduce significant overlap between regressor distributions in successive time periods without changing the correlation structure of our model. This overlap ensures enough stayers to adequately estimate shock means. The arbitrary offsets simply demonstrate that our approach does not require demeaned regressors.

Given these definitions, the response variable is then given by $Y_t = A_t + B_t X_t + C_t Z_t$ for periods $t = 1, 2, 3$.

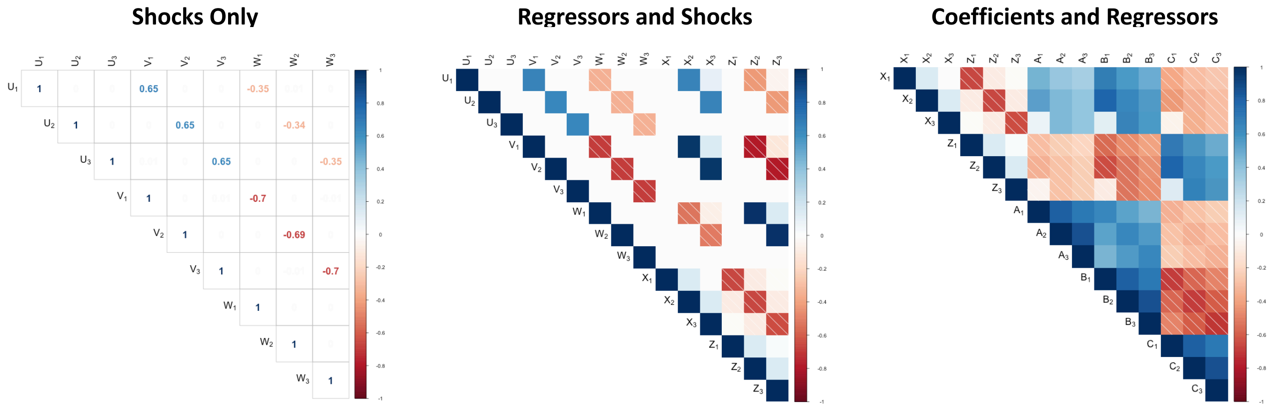

Figure 1: *Correlation matrices for various components of the model we use for simulations are diagrammed. Each cell represents the correlation between one pair of indicated variables with the coloration of that cell indicating the sign and magnitude of the associated relationship. Notice that shocks are correlated within the same time period, but not across different time periods. Further, regressors are correlated with shocks from previous time periods and with coefficients from all time periods.*

The correlation structure of our model is summarized in Figure 1, which plots correlation matrices for three different groups of variables from our model. It should be noted that this model does contain several simplifications not required by our theoretical treatment. For instance, shocks can be non-normal and heteroskedastic, while regressors can have more complicated relations to the shocks and include any amount of independent random noise.

We generated data from this process for a panel of $n = 2500$ individuals over $r = 1000$ replications. For each realization of simulated data, we used both variants of the procedure – HSS and HSS_0 – to estimate all shock and parameter moments with the parameter values given in the first column of Table 1 (h_2 is not needed for HSS_0). Instead of reporting the multiplier γ, we report the quantity $\bar{b}$, which is the average bandwidth, over replications, used to estimate shock means.

Figure 2 plots the distributions of the estimates obtained using the HSS_0 estimator along with the true values of the quantities being estimated. It shows that the estimates produced by our procedure center around the correct values and exhibit a relatively normal distribution. The corresponding distributions for the HSS estimator, which we do not reproduce here, have a very similar shape.

In Table 2 we compare our two estimators (given in rows 1 and 3) with two other methods

	$n = 2500$		$n = 10000$		
	Feedback	*Lag*	*Feedback*	*Lag*	Scale
$\bar{b}$	0.261	0.365	0.207	0.290	$n^{-1/6}$
h_1	20	20	40	40	$n^{1/2}$
h_2	18	18	32^*	32^*	$n^{2/5}$
$\bar{\delta}$	0.172	0.511	0.130	0.388	$n^{-1/5}$

Table 1: *The parameters used by our procedure for the two sample sizes considered are tabulated. $\bar{b}$ is the average bandwidth used to estimate shock means. h_1 and h_2 correspond to bandwidths used for non-parametric regressions conditional on four and six variables respectively and are given in terms of the number of points used in nearest-neighbors averaging. δ is the threshold for rejecting observations based on the determinant criterion. We report $\bar{\delta}$, its average value over replications. The h_2 parameter with stars was not used in any reported simulations but represents the value of h_2 that would be used by the HSS estimator on a sample of size $n = 10000$. It is provided as a reference value.*

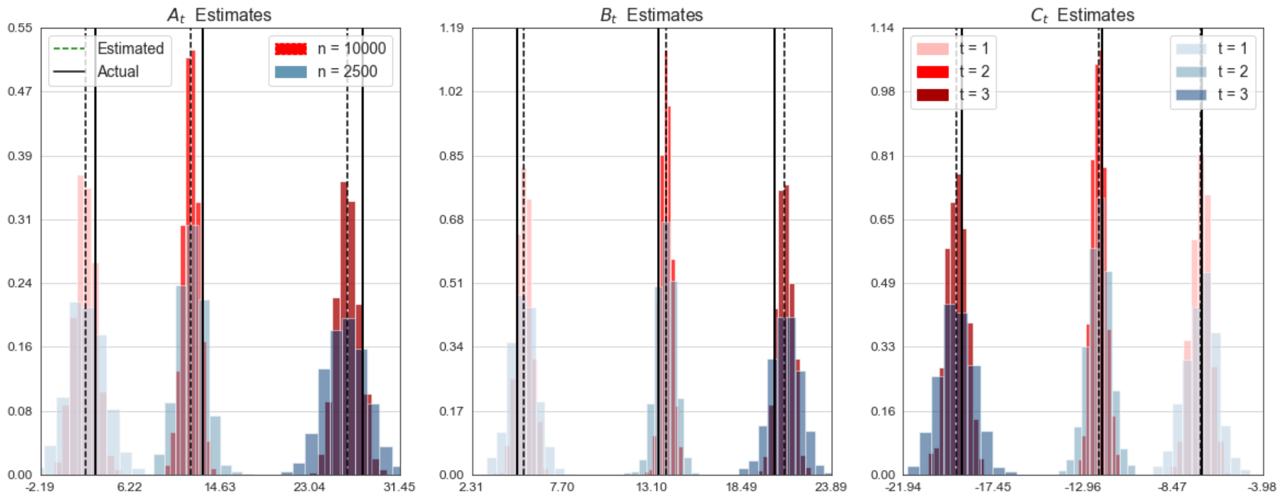

Figure 2: *The distributions of parameter estimates obtained by the HSS_0 estimator for the feedback data generating process with two different sample sizes are plotted. The histograms in blue correspond to $n = 2500$ and the ones in red represent estimates for $n = 10,000$. Solid black lines near the center of each distribution indicate the true value of the underlying parameter moments and the dotted lines indicate the mean of the distribution of estimates.*

available from previous literature. The second row summarizes the performance of the estimator proposed by Graham and Powell (2012), which was intended to estimate parameter moments in a similar but more restrictive panel data model. Note that their procedure requires the selection of one parameter – a bandwidth for the selection of stayers – which we do by generalizing the heuristic used in their applied example. Specifically, we estimate the standard deviation c_D of the determinant of the matrix A as defined in (5) and then set the bandwidth to be $h = 0.5c_D \times n^{-1/3}$. This results in approximately four percent of the population being classified as stayers in each replication. For reference, with our choice of parameters we classify on average 2.5 percent of the population as stayers and reject about four percent of the sample as ill-conditioned.

The fourth row of Table 2 describes the results obtained by ignoring the structure of the data and simply applying OLS within each period. That is, we consider the data for each period separately and use OLS to estimate (A_t, B_t, C_t) independently for $t = 1, 2, 3$. We then derive the shock estimates as $(U_t, V_t, W_t) = (A_t, B_t, C_t) - (A_{t-1}, B_{t-1}, C_{t-1})$.

Each entry of these three rows provides a different performance metric. The first entry of each row gives the relative bias of the associated estimates. For an arbitrary parameter P, with $\widehat{E[P]}_j$ representing the estimate of $E[P]$ obtained on replication j, this quantity is defined as

$$\text{Rel. Bias} = \frac{1}{E[P]} \left| \frac{1}{r} \sum_{j=1}^{r} \widehat{E[P]}_j - E[P] \right| \tag{11}$$

The second entry of each row gives the root Mean Square Error (rMSE) of the associated estimates:

$$\text{rMSE} = \sqrt{\frac{1}{r} \sum_{j=0}^{r} \left(\widehat{E[P]}_j - E[P] \right)^2} \tag{12}$$

The third entry of each row provides a comparison between methods by calculating the ratio of the rMSE obtained using the indicated estimator to that achieved by our own baseline procedure, HSS. This quantity is thus given by

Feedback Data Generating Process

		U_t			V_t			W_t		
		$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$
HSS	*Rel. Bias*	–	0.02	0.04	–	0.004	0.03	–	0.02	0.010
	rMSE	–	2.10	2.42	–	0.96	1.11	–	0.95	1.08
	Rel. rMSE	–	1	1	–	1	1	–	1	1
GP	*Rel. Bias*	–	0.15	0.12	–	0.10	0.10	–	0.16	0.12
	rMSE	–	5.24	5.07	–	2.45	2.18	–	2.21	2.21
	Rel. rMSE	–	2.50	2.10	–	2.54	1.97	–	2.31	2.06
$\mathbf{HSS_0}$	*Rel. Bias*	–	0.001	0.03	–	0.002	0.03	–	0.03	0.02
	rMSE	–	2.03	2.40	–	0.94	1.09	–	0.95	1.05
	Rel. rMSE	–	0.97	0.99	–	0.97	0.98	–	1.00	0.98
OLS	*Rel. Bias*	–	0.09	0.009	–	0.05	0.009	–	0.03	0.004
	rMSE	–	1.03	0.60	–	0.49	0.28	–	0.30	0.29
	Rel. rMSE	–	0.49	0.25	–	0.51	0.25	–	0.31	0.27

		A_t			B_t			C_t		
		$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$
HSS	*Rel. Bias*	0.29	0.08	0.06	0.08	0.03	0.03	0.02	0.02	0.02
	rMSE	1.99	1.63	2.45	0.90	0.71	1.05	0.81	0.60	0.93
	Rel. rMSE	1	1	1	1	1	1	1	1	1
GP	*Rel. Bias*	0.12	0.09	0.10	0.06	0.04	0.06	0.06	0.03	0.03
	rMSE	3.40	3.12	4.33	1.56	1.41	1.88	1.47	1.32	1.43
	Rel. rMSE	1.71	1.92	1.77	1.73	1.99	1.78	1.80	2.19	1.53
$\mathbf{HSS_0}$	*Rel. Bias*	0.34	0.08	0.06	0.09	0.03	0.03	0.004	0.02	0.02
	rMSE	2.04	1.62	2.43	0.91	0.70	1.05	0.78	0.60	0.91
	Rel. rMSE	1.02	0.99	0.99	1.01	0.99	0.99	0.95	0.99	0.98
OLS	*Rel. Bias*	2.24	0.59	0.28	0.65	0.27	0.18	0.18	0.12	0.08
	rMSE	6.74	7.66	7.79	3.23	3.67	3.74	1.25	1.43	1.46
	Rel. rMSE	3.39	4.70	3.18	3.59	5.19	3.55	1.54	2.37	1.57

Table 2: *Summary statistics are tabulated for the performance of several estimators in fitting the parameters of our feedback data generating process. The first and second entries of each row give the relative bias and rMSE respectively, while the third computes the ratio of rMSE for each estimator to that obtained using the HSS benchmark. Values of this last metric that are greater than one indicate better performance by HSS.*

$$\text{Rel. rMSE} = \frac{\text{rMSE Alternate Estimates}}{\text{rMSE } HSS} \tag{13}$$

Notice that the relative rMSE for HSS is exactly one by construction and values greater than one indicate that HSS outperforms the listed estimator, in terms of rMSE, by the given factor.

Table 2 demonstrates that there is practically no difference in the performance of HSS and HSS_0, though the latter estimator is significantly more computationally efficient and requires choosing three rather than four parameters. Additionally, this table provides strong evidence that either of our estimators is able to offer significant improvements over both the Graham Powell methodology and naive application of OLS. Both procedures outperform both the Graham Powell procedure

		U_t			V_t			W_t		
		$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$
n = 2500	*Rel. Bias*	–	0.00	0.03	–	0.00	0.03	–	0.03	0.02
	Std. Err.	–	(2.03)	(2.34)	–	(0.94)	(1.08)	–	(0.94)	(1.05)
n = 10,000	*Rel. Bias*	–	0.01	0.02	–	0.00	0.02	–	0.01	0.01
	Std. Err.	–	(1.26)	(1.37)	–	(0.59)	(0.63)	–	(0.59)	(0.63)

		A_t			B_t			C_t		
		$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$
n = 2500	*Rel. Bias*	0.34	0.08	0.06	0.09	0.03	0.03	0.00	0.02	0.02
	Std. Err.	(1.76)	(1.25)	(1.88)	(0.79)	(0.56)	(0.86)	(0.78)	(0.57)	(0.85)
n = 10,000	*Rel. Bias*	0.34	0.09	0.05	0.08	0.03	0.03	0.02	0.02	0.01
	Std. Err.	(1.01)	(0.73)	(1.12)	(0.47)	(0.33)	(0.51)	(0.48)	(0.32)	(0.51)

Table 3: *Parameter estimates for our feedback process are compared on sample sizes of $n = 2500$ and $n = 10,000$. The table shows that increasing the sample size by a factor of four does decrease the resulting standard error by roughly a factor of two, as theory suggests.*

and OLS in terms of rMSE for all coefficient moments, and offer improvements in relative bias in many instances as well. For the shock moments, our procedure again dominates the Graham Powell methodology but it is interesting to note that the naive application of OLS performs better than all other estimators.

We also analyzed the asymptotic behavior of our estimator empirically by again simulating data from the model described above but this time for a panel of $n = 4 \cdot 2500 = 10000$ individuals. For each of $r = 1000$ replications, we fit the HSS_0 estimator to the simulated data using the parameters given in the second column of Table 1. The resulting estimates are plotted in Figure 2 and tabulated in Table 3. The histograms suggest that these estimates are again normally distributed and the table shows that this increase of the sample size by a factor of four roughly halves the standard errors of our estimates, suggestive of root-n convergence.

Lagged Response Data Generating Process

One theoretical advantage of our estimator is that it allows for models that include lags of the response variable as regressors – something that to our knowledge has not yet been accomplished by previous literature. Given the importance of models of this form in many applied settings, we now use simulated data to analyze the performance of our estimator in regards to the lagged response setting in particular.

We again generate multivariate normal shocks, except now the means are smaller in magnitude. These changes simplify the model and result in regressor magnitudes that are more stable over time. Specifically, shocks are generated for $t = 0, \ldots, 3$ according to

$$\begin{bmatrix} U_t \\ V_t \\ W_t \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{U_t} \\ \mu_{V_t} \\ \mu_{W_t} \end{bmatrix}, \begin{bmatrix} 1.00 & 0.33 & -0.18 \\ 0.33 & 0.25 & -0.18 \\ -0.18 & -0.18 & 0.25 \end{bmatrix} \right) \tag{14}$$

with the means in each period defined as

Lagged Response Model

		U_t			V_t			W_t		
		$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$
HSS	*Rel. Bias*	–	0.10	0.006	–	0.05	0.20	–	0.17	0.24
	rMSE	–	0.13	0.13	–	0.22	0.22	–	0.18	0.14
	Rel. rMSE	–	1	1	–	1	1	–	1	1
GP	*Rel. Bias*	–	0.97	2.66	–	0.05	0.33	–	0.06	0.31
	rMSE	–	0.38	0.43	–	0.63	0.55	–	0.50	0.25
	Rel. rMSE	–	2.82	3.38	–	2.90	2.50	–	2.72	1.76
HSS_0	*Rel. Bias*	–	0.24	0.02	–	0.04	0.25	–	0.19	0.25
	rMSE	–	0.14	0.12	–	0.22	0.21	–	0.20	0.15
	Rel. rMSE	–	1.03	0.97	–	1.00	0.98	–	1.07	1.01
OLS	*Rel. Bias*	–	4.69	5.14	–	3.13	0.73	–	2.17	1.91
	rMSE	–	0.47	0.52	–	0.79	0.23	–	0.66	0.58
	Rel. rMSE	–	3.53	4.11	–	3.64	1.04	–	3.59	4.02

		A_t				B_t				C_t		
		$t=1$	$t=2$	$t=3$		$t=1$	$t=2$	$t=3$		$t=1$	$t=2$	$t=3$
HSS	*Rel. Bias*	0.03	0.02	0.03		0.13	0.10	0.03		0.22	0.09	0.006
	rMSE	0.17	0.17	0.21		0.21	0.15	0.19		0.23	0.13	0.06
	Rel. rMSE	1	1	1		1	1	1		1	1	1
GP	*Rel. Bias*	0.07	0.06	0.51		0.02	0.03	0.10		0.06	0.02	0.06
	rMSE	0.30	0.31	0.55		0.52	0.28	0.43		0.51	0.18	0.14
	Rel. rMSE	1.80	1.82	2.58		2.45	1.87	2.24		2.23	1.45	2.16
HSS_0	*Rel. Bias*	0.04	0.009	0.01		0.14	0.10	0.02		0.22	0.09	0.003
	rMSE	0.18	0.18		0.21	0.15	0.19		0.24	0.13	0.07	
	Rel. rMSE	1.07	1.02	0.98		1.00	1.00	1.00		1.06	1.00	1.06
OLS	*Rel. Bias*	1.23	0.74	1.72		1.35	0.14	0.29		0.02	0.74	1.03
	rMSE	0.98	0.52	1.03		0.68	0.13	0.33		0.06	0.67	1.24
	Rel. rMSE	5.95	2.98	4.87		3.20	0.88	1.70		0.26	5.23	19.23

Table 4: *Summary statistics are tabulated for the performance of several estimators in fitting the parameters of our lagged response process. The first and second entries of each row give the relative bias and rMSE respectively, while the third computes the ratio of the rMSE of each estimator to that obtained by the HSS benchmark. Values of this last metric that are greater than one indicate better performance by HSS.*

$$(\mu_{U_t}, \mu_{V_t}, \mu_{W_t}) = \begin{cases} (-0.4, -0.25, 0.30) & \text{for } t = 0, 1 \\ (0.10, -0.25, 0.30) & \text{for } t = 2, 3 \end{cases}$$

Notice that the correlation between shocks within each period has not changed from (14), the different covariance matrix results from the reduced variances only.

Coefficients are exactly the cumulative sums of shocks as in equation (8). Regressors are simulated similarly as to before except now Z_t is the one-period lagged response variable and we no longer offset regressors away from having zero mean. That is, we generate the intermediate variables given by

		U_t			V_t			W_t		
		$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$
n = 2500	*Rel. Bias*	–	0.24	0.02	–	0.04	0.25	–	0.19	0.25
	Std. Err.	–	(0.14)	(0.12)	–	(0.22)	(0.20)	–	(0.19)	(0.13)
n = 10,000	*Rel. Bias*	–	0.11	0.05	–	0.04	0.12	–	0.11	0.17
	Std. Err.	–	(0.08)	(0.08)	–	(0.13)	(0.13)	–	(0.11)	(0.08)

		A_t			B_t			C_t		
		$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$
n = 2500	*Rel. Bias*	0.04	0.01	0.01	0.14	0.10	0.02	0.22	0.09	0.00
	Std. Err.	(0.17)	(0.18)	(0.21)	(0.20)	(0.13)	(0.19)	(0.20)	(0.10)	(0.07)
n = 10,000	*Rel. Bias*	0.03	0.02	0.02	0.12	0.10	0.04	0.18	0.08	0.02
	Std. Err.	(0.09)	(0.09)	(0.12)	(0.12)	(0.07)	(0.11)	(0.12)	(0.06)	(0.04)

Table 5: *Parameter estimates for our lagged response process are compared on sample sizes of $n = 2500$ and $n = 10,000$. The table shows that increasing the sample size by a factor of four does decrease the resulting standard error by roughly a factor of two, suggestive of $\sqrt{n}$ convergence.*

$$\begin{aligned}
\widetilde{X}_1 &= V_0 + 0.25W_0 \\
\widetilde{X}_2 &= V_1 + 0.25W_1 + 0.15(V_0 + 0.25W_0) \\
\widetilde{X}_3 &= V_2 + 0.25W_2 + 0.15(V_1 + 0.25W_1) + 0.15^2(V_0 + 0.25W_0) \\
\widetilde{Z}_1 &= -0.25V_0 + W_0 \\
Z_2 &= Y_1 \\
Z_3 &= Y_2
\end{aligned}$$

and then standardize the intermediate variables to produce the associated regressors. That is,

$$X_t = \frac{\widetilde{X}_t - E[\widetilde{X}_t]}{\sqrt{\mathrm{Var}[\widetilde{X}_t]}}, \quad Z_t = \frac{\widetilde{Z}_t - E[\widetilde{Z}_t]}{\sqrt{\mathrm{Var}[\widetilde{Z}_t]}}$$

Notice that the lagged response variables Z_2 and Z_3 are not standardized.

Again, we simulated data from this model for a cross-section of $n = 2500$ individuals and fit each estimator over $r = 1000$ replications. The parameters used by both versions of our estimator are summarized in the first column of Table 1, again HSS_0 does not use h_1. Summary statistics for the resulting parameter distributions, as described in equations (11) - (13), are given in Table 4. This table shows that either variant of our estimator dominates both the Graham Powell procedure and naive OLS in terms of rMSE for all quantities and performs better in terms of bias in many cases as well. Notice that the OLS estimates of the shock moments now appear to be inconsistent, unlike in the feedback case.

We also considered a larger panel of $n = 10000$ for the lagged response model, again to analyze the asymptotic behavior of our estimator empirically. The results of applying the HSS_0 estimator to data simulated from this model, with parameters described in Table 1, are given in Table 5 and again suggest root-n convergence of the estimator.

Conclusion

References

Fox, J., Hadad, V., Hoderlein, S., Petrin, A., and R. Sherman (2019) "Identification and estimation in a correlated random coefficients linear panel data model," Working paper.
Graham, B, and J. Powell (2012) "Identification and estimation of partial effects in "irregular" correlated random coefficient panel data models," *Econometrica*, 80, 2105–2152.
White, H. (1980) "A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity," *Econometrica*, 48, 817–838.