The Social Network of Money

Boyan Bejanov and Geoffrey R. Dunbar[1]

November 2015

Preliminary, Do Not Cite

Overview

We propose a method to quantify the usage of physical currency in unobserved and networked sub-groups. Our method identifies sets of parameters governing withdrawals, the movement of cash between sub-groups and between sub-groups and issuers when those parameters are time-invariant from currency flows. Our method may return the empty set if there is no time invariant set of parameters which rationalize the data. If the parameters are not time invariant, we show how to recover estimates of the parameter means from currency flows. In short, we show that money *has* memory. We use confidential data from the Bank of Canada's currency circulation trials which track the withdrawal and deposits of a known set of bank notes in circulation using optical scanners and the unique serial number of each note to identify the network parameters. Our methods may also be applicable to the study of the shadow (financial) economy and biological models of migration.

[1] Boyan Bejanov and Geoffrey Dunbar: Bank of Canada, 234 Laurier Ave W., Ottawa, Ontario, Canada. Our views do not necessarily represent those of the Bank of Canada or her employees. We thank Angelika Welte for valuable discussions throughout this project. We thank Rallye Shen for superb research assistance. We thank Marc Charron and Greg Kealey for making the circulation trial data available to us. All errors are our own. Corresponding author email:dunbarg@bankofcanada.ca.

1 Introduction

In this paper, we consider the identification and estimation of parameters governing exchanges of objects between (unobserved) groups and between these groups and an observer. We define a network topology that features directed paths and parameters which govern the probability of an object transiting along each path. We make two methodological contributions. Our first contribution is a method to set-identify network parameters via integer programming using data from the entry and exit points of the network when those parameters are time invariant (fixed). Although set-identification may appear an unusual approach, our focus is to identify fixed parameters given rational-number data. Indeed, as we show, recognizing that the objects are atomistic, and so the parameters are rational numbers, is central to our identification strategy. We argue, and demonstrate in our application, that a network structure with fixed parameters may not be an appropriate model for the data. We show that in such cases no sets are identified. If one is not, our second contribution is to provide a method to estimate the distribution of the network parameters using time series data when those parameters are not fixed but reside in the space of rationals.

As an application, we make use of a novel currency circulation trial dataset collected by the Bank of Canada. Currency circulation trials track the withdrawals and deposits of particular bank notes to determine aspects of their fitness: wear and tear, mutilation and graffiti (markings). To do so, circulation trials track bank notes by their serial numbers as they enter the operations centres of the Bank of Canada's Currency Department. These notes are introduced into the economy virtually simultaneously. Thus, there are no, regular, repeated injections of notes which, as we argue shortly, distinguishes our environment from those typically studied in machine learning. We define withdrawal and deposit probabilities for each note to characterize the passage of notes through unobserved groups in the economy. Because we do observe group membership, we cannot label them but possible candidate labels for the latent groups might be individuals and firms; cash users and non-cash users; legal economy and underground economy, or; day market and night market participants in money-search models, depending on the data.

In our application, there are no fixed parameters that are able to rationalize our data. We conclude that, at least, one parameter must be time-varying. Applying our second method to the data reveals two groups with differing probabilities of note circulation. In particular we find that notes essentially flow from one group, which is responsible for obtaining notes from financial institutions, to the second group which obtains notes much less often from financial institutions. The second group is also about half as likely to deposit notes with financial institutions. Thus, over time, the stock of circulating notes is primarily held by the second group.

Interestingly, we find no evidence in other disciplines that this model has been studied in the manner in which we formulate it. The literature of fluid dynamics appears to focus on models with explicit parameters, such as viscosity and temperature, and also appears to measure resistance from known obstructions such

as valves or directional changes. Similarly, the study of electrical mechanics appears to have focused on observed electrical circuits and has measured resistance directly. Another literature that may appear related to ours are epidemological models of population size such as the widely used Cormack-Jolly-Seber model (see Williams *et al.* 2002). The key difference between this literature and our paper is that we know precisely the number of bank notes in circulation (or at least that have been issued). What we are interested in learning is the composition of various sub-groups in the economy but since we cannot sample notes from these groups directly and notes bear no physical markings from passing through groups, these models are inapplicable to our data. Our research is similar in spirit to a recent literature identifying preference types in the consumption literature using revealed preference arguments, see Crawford and Pendakur (2013) and Cherche *et al.* (2012). The key difference is that we do not use preference restrictions to identify types but rather identify types as resistance in the flow of currency in circulation.

Perhaps the most related literature is the machine learning literature (*e.g.* Barber 2012) but this literature typically proceeds using iterative information to construct Bayesian updates from sequences of states (in our case, injections of notes). Our model is similar only in spirit to hidden (latent) Markov models used in this literature. Given the lack of a sequence of observations on the initial states of our system, we show how to proceed via integer programming rather than updating non-parametric density estimates. Thus, our methodological contribution is to derive an implementable method of sequential elimination using integer programming instead of statistical learning. Our key methodological insight is that the data in our system are discrete and that every data point must reside in the set of (positive) rational numbers. Thus we can without loss of generality restrict our system parameters to those that generate data that belong to rational numbers. We then show how the probability characterization of our system implies restrictions on the feasible parameter space and demonstrate set-identification of our system parameters. These sets, if they exist, are analogous to the non-parametric densities in machine-learning. We argue that our method works if these sets are non-smooth and by extension could fail to satisfy some regularity conditions associated with the estimation of non-parametric densities which would prove problematic for existing machine-learning methods. Perhaps more importantly, the set of rational numbers are Lebesque measure zero of the set of Real numbers and so any method which updates on a continuous distribution defined over the Real numbers will by necessity update using information that cannot represent our underlying data. Indeed, as we show, there is no point-identified system in our data so **any** method which yields estimates of one is mis-specified.

Our second contribution follows from our first. If our first method fails to find fixed parameters which can rationalize the data, then these parameters may not be fixed. We show that if no fixed parameters exist for a given set of groups, then no fixed parameters exist for a network with a larger number of groups. In such a case, we show that time series data potentially may be used to identify the mean, and potentially the distribution, of the network parameters.

Theoretically, our procedure could be used to find the minimal number of latent groups that rationalize the observed data. With infinite computing resources and time, our procedure could theoretically identify every individual as a latent group. Thus, similar in spirit to Kocherlakota (1998), we show that currency flows embed memory and can in theory be used to identify the transaction history and existence of every agent who uses it. The difference between Kocherlakota (1998) and our model is that the former shows money *is* memory while we show that money *has* memory.

2 A Simple Fixed Model

For clarity of exposition, we shall assume that the total population of a region, N, is comprised of two types (or groups), with sub-populations $N_1 + N_2 = N$. We are interested in obtaining estimates of the total number of units of currency held by the groups N_1 and N_2 and also parameters governing their usage of physical currency. In later sections, we relax the restriction to two groups.

As a working example, consider an environment with two physical locations, llabelled 1 and 2, for introducing a quantity of bank notes into an economy. Let L_1 and L_2 be the quantities of physical currency at each location in some initial period 0. Bank notes are not anonymous since their identity is known by serial number. For the discussion which follows, we refer to notes and cash interchangeably. We assume that there exists an entity, the Bank, which is able to identify the bank note if it receives it.[1]

Let members of group, N_1 obtain cash from each location $j = \{1, 2\}$, with associated probability p_j, only during the withdrawal period, $t = 0$. Note that these probabilities sum to 1 because this is the probability that a note from a location transit from location j to group i.[2] The notes that are withdrawn are held either by members of group N_1 or group N_2 at the end of the withdrawal period $t = 0$. Thus, for example, with two groups $p_1 L_1$ is the quantity of notes from location L_1 held by group N_1 at the end of the withdrawal period, $t = 0$. In this example, $(1 - p_1)L_1$ is then held by group N_2 at $t = 0$.

We assume that time is discrete with fixed intervals indexed $t = 1, 2, 3, ...$ In any given interval t there are three possible trajectories for cash held by a group:

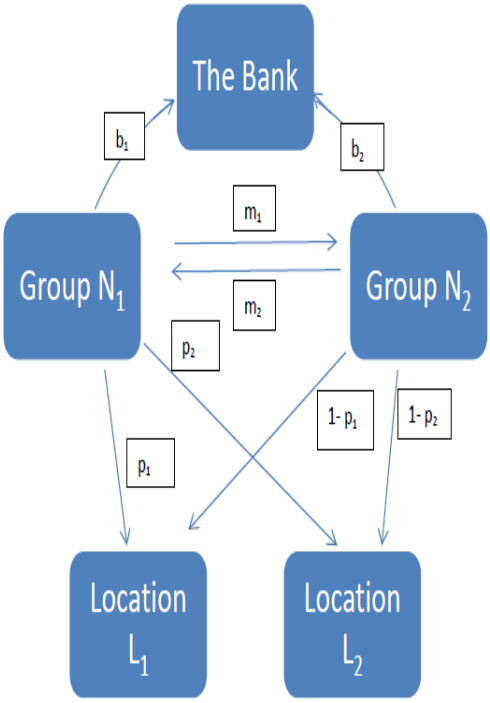
1. it can remain within a group,
2. it can be exchanged across groups, and
3. it can be deposited at the Bank.

Figure 1 illustrates the system.

[1] In practice, we assume that the Bank will be the Bank of Canada's Agency Operation Centre (AOC).

[2] Without loss of generality, we shall assume that all cash at either location is withdrawn although this assumption may be relaxed without much cost.

Figure 1: System Schematic



Beginning with the last trajectory, let the probability that a note from group i is captured by the bank at the end of a period t be b_i. Notice that we assume that the probability b_i is time-invariant in this example but in our application these probabilities may time-vary.[3] We interpret b_i as a compound probability of: using cash with a person or enterprise $\times$ the probability that the person or enterprise deposits the cash in a financial institution $\times$ the probability that the cash held by the financial institution is deposited at the Bank. We note that the last probability is effectively known, given the sampling probabilities of the Bank of Canada's AOCs. However, the remaining implicit probabilities are not separately identified in our experimental design.

[3]This assumption would seem unreasonable for some time intervals, such as daily time intervals, because of cyclical patterns in the data. It is unlikely that expenditures are similar between Wednesday and Saturday. In practice we expect that a time interval of 1 week would be consistent with this assumption.

We assume that the probability that cash transits from group N_1 to N_2 in a period t is m_1 and that m_2 is the converse transit probability. It follows that the probability that cash held by group N_1 is captured by the Bank is $b_1 + m_1 b_2$, with an analogous probability for cash held by group N_2. Note that this structure assumes that the probability of bank deposit for a note that has transited between groups is identical to the bank deposit probability for a note which did not transit. This assumption imposes that the probability of bank capture is unconditional on (independent of) the beginning ownership of that bill. An alternative structure might impose a sub-period structure in which bills first transit or not between groups and then are subsequently captured by the bank. The difference between this structure and ours is that the alternative assumes that deposits are always end of period events. Our structure, in contrast, is more general in that we permit deposits at any point.[4] We note the structure of our model implies that groups are defined by their parameters which govern the flow of notes. Thus an individual could be a convex combination of the behaviours of both groups. As an example, define a group, i, as being savers – individuals who do not spend or otherwise relinquish bank notes. In this case $m_i = b_i = 0$ for this group. Alternatively, a consumer group j could feature $m_j, b_j > 0$. Any particular individual could keep some notes for savings and also use some notes for consumption.

Define the quantity of notes from location L_1 that are captured by the Bank in period $t = 1$ as:

$$B_1^1 = L_1 p_1 (b_1 + m_1 b_2) + L_1 (1 - p_1)(b_2 + m_2 b_1).$$

Similarly, the quantity of notes from location L_2 is:

$$B_1^2 = L_2 p_2 (b_1 + m_1 b_2) + L_2 (1 - p_2)(b_2 + m_2 b_1).$$

Note that we are assuming in this description that the within group transit probabilities, m_1 and m_2, and the bank capture probabilities, b_1 and b_2, do not vary across the locations. This assumption may be reasonable for locations that are *close* in some sense so that the behaviour of the groups is invariant across these locations. Possible candidate locations that would seem likely to satisfy this assumption would be two ATMs in the same city or neighbourhood or perhaps two retailers with similar product lines. For our practical application which follows, our locations are the Bank of Canada's agency operations centres in the two largest cities in Canada, Toronto and Montreal. Whether these groups in these cities are sufficiently *close* is unknown. Given the highly sensitive nature of our data, we are unable to conduct our experiment with locations that might seem *a priori* to be more similar at this time.

As noted above, the assumption that $\{m_1, m_2, b_1, b_2\}$ are time invariant may not be reasonable. For the purposes of our discussion which follows, we define two classes of networks: fixed and stable. We define a

[4] Ultimately, the main difference is in the interpretation of the probabilities.

network as fixed if the parameters $\{m_1, m_2, b_1, b_2\}$ are time invariant. We define a network as stable if the parameters time-vary but the network structure (the direct graph) is time-invariant.

2.1 Fixed Networks

Note that the number of notes captured by the Bank reduces the number of notes that remain with the groups at the end of the period. At the beginning of period $t = 1$, group N_1 held $L_1 p_1$ notes from location L_1. During the period $t = 1$, these notes either transited to the Bank with probability b_1 or to group N_2 with probability m_1. At the same time, group N_1 received $L_1(1 - p_1)m_2$ notes from group N_2 and then retained a proportion $(1 - b_1)$ of these notes at the end of the period. The total holdings of L_1 notes by group N_1 at the end of period $t = 1$ (also the beginning of $t = 2$) is $L_1 p_1(1-b_1-m_1)+L_1(1-p_1)m_2(1-b_1)$. (Similar expressions can be derived for the remaining note holdings.) This system of note transitions can be represented generally as a matrix system. Define the $t = 0$ holdings of notes as $\mathcal{N}_0$ with element $\backslash_{i,j}$ where i indexes the group and j indexes the initial location. Then:

$$\mathcal{N}_0 = \begin{bmatrix} L_1 p_1 & L_2 p_2 \\ L_1(1-p_1) & L_2(1-p_2) \end{bmatrix} = \mathcal{P}\mathcal{L}. \tag{1}$$

Next define the transition matrix, $\mathcal{A}$ as:

$$\mathcal{A} = \begin{bmatrix} 1-b_1-m_1 & m_2(1-b_1) \\ m_1(1-b_2) & 1-b_2-m_2 \end{bmatrix}. \tag{2}$$

It is straightfoward to establish that the holdings of notes at the end of $t = 1$, $\mathcal{N}_1$ is:

$$\mathcal{N}_1 = \mathcal{A}\mathcal{N}_0, \tag{3}$$

and, more generally,

$$\mathcal{N}_t = \mathcal{A}\mathcal{N}_{t-1}. \tag{4}$$

The difference between note holdings in any two adjacent time periods must equal the number of notes captured by the Bank. The first period in which the Bank captures notes is $t = 1$ and we define $\mathcal{B}_1$ as:

$$\mathcal{B}_1 = \begin{bmatrix} B_1^1 \\ B_1^2 \end{bmatrix}, \tag{5}$$

using the definitions established above. We note that $\mathcal{B}_1$ may also be written generically as:

$$\mathcal{B}_t = (\mathcal{N}_{t-1} - \mathcal{N}_t)^T \iota \equiv ((\mathcal{I} - \mathcal{A})\mathcal{N}_{t-1})^T \iota \tag{6}$$

where ι is an $i \times 1$ vector of ones and $\mathcal{I}$ is a suitably dimensioned identity matrix. We assume that at least one element of $\mathcal{B}_t$ is non-zero. The system of Equations (4) and (6) jointly yield four constraints, two for

each equation independently. These four constraints can be used to recover four parameters from the system described in Figure 1 providing the equations are independent which generically requires $p_1 \neq p_2$. We note that we need to capture bank notes in at least two time periods to identify the parameters. If $p_1 = p_2$ then both groups, L_1 and L_2, yield identical information regarding the system parameters and our system is unidentified. Fortunately there is a simple practical solution to non-identification resulting from $p_1 = p_2$ since the sets L_1 and L_2 can be manipulated *ex-post* by researchers given that the sets of notes being tracked is arbitrary.

Nevertheless, as we argue shortly, we cannot point-identify the 6 parameters in the model. Unfortunately, a times series of bank note capture data do not help in identifying particular parameters because the characteristic polynomial of $\mathcal{A}$ does not provide additional information for Equation (4) and by extention Equation (6). Technically, we cannot identify the basis vector of our system. One easy method to verify non-identification is to note that the system can be rewritten as:

$$\mathcal{B}_t = M\mathcal{B}_{t-1} \tag{7}$$

where,

$$M = (\mathcal{L}\mathcal{P})^T \mathcal{A}^T (\mathcal{L}\mathcal{P})^{-T}.$$

This system of moments does not identify the system parameters because M is a 2×2 matrix and there are 6 parameters in M: $\{p_1, p_2\}$ and $\{m_1, m_2, b_1, b_2\}$. Thus, while time-series data on $\mathcal{B}$ may identify M, thethe system is under-identified in M.

2.2 Binning

Our method, while exact, may be computationally demanding depending on the number of elements in L. We also require that $p_i \neq p_j, \forall i \neq j$. These complications are inherent to the dataset under consideration. The main point we wish to make in this section is that, in at least some applications, the definition of location is potentially definable by the researcher. If we assume that our locations are, in fact, comprised of a number of different sub-locations then we can somewhat arbitrarily assign notes to sub-locations since the concept of a location is a draw from some initial probability distribution over note assignment to groups.

One possibility we consider, and choose to develop, is to construct bins from our circulation data, L, to reduce the size of the feasible set of $\{p_1, p_2, ..., p_j\}$. Suppose we construct m equal-sized bins from an initial sample of M notes and set $j = 2$ in each bin. Then the number of possible probability sets to consider in each bin is $(\frac{M}{jm})^2$. It is clear that we can arbitrarily choose m to reduce the computational burden. A second observation we make is that it is not essential that the distribution of notes across locations within bins are identical. In theory, for each bin, we could set $(\frac{M}{m} - 3)$ notes in one location and a single note in

the remaining location. Then the enumerable sets of probabilities are simply $(\frac{M}{m} - 3)$ in each bin. The cost of doing so is that we reduce the support for our estimates of the network parameters to $\{0, 1\}$ for some parameters. Such a step is obviously too reductionist, but our point is to illustrate how one can manipulate bins to improve computational speed. We next address how one can recover the parameters governing the network of the full data.

We define $\mathcal{B}_t(m)$ to be the number of notes from bin m received by the Bank at time t. Each $\mathcal{B}_t(m)$ is a $j \times 1$ vector. Let $N_0(m)$ be the initial distribution of notes in bin m. We assume for the moment (and demonstrate later) that we can recover $\mathcal{A}_t(m)$ given $\mathcal{B}_t(m)$. It follows that:

$$\mathcal{B}_1(m) = (\mathcal{N}_0(m)(\mathcal{I} - \mathcal{A}_1(m)))^T \iota. \tag{8}$$

We impose the additivity condition that $\mathcal{B}_t = \sum_m \mathcal{B}_t(m)$ to yield:

$$\mathcal{B}_1 = (\sum_m \mathcal{N}_0(m) + \sum_m \mathcal{N}_0(m)\mathcal{A}_1(m))^T \iota. \tag{9}$$

In comparison with equation (6), this implies that:

$$\mathcal{N}_0 = \sum_m \mathcal{N}_0(m) \tag{10}$$

and

$$\mathcal{A}_1\mathcal{N}_0 = \sum_m \mathcal{A}_1(m)\mathcal{N}_0(m). \tag{11}$$

Equation (10) implies that the $\{p_1, p_2, ..., p_j\}$ can be recovered as weighted sums of the bins. Calculation of $\mathcal{A}$ is less straightforward, unless $\mathcal{N}_0$ is invertible. If the inverse, or pseudo-inverse, exists then one can recover

$$\mathcal{A}_1 = \sum_m \mathcal{N}_0(m)^T \mathcal{A}_1(m) \mathcal{N}_0^{-1}.$$

If one cannot construct an inverse, one can simply use the constructed $\mathcal{N}_0$ to solve for $\mathcal{A}$.

Our discussion thusfar in this section has implicitly assumed that there is a single $\mathcal{A}(m)$ for each m. There is nothing which guarantees such uniqueness of $\mathcal{A}(m)$. When $\mathcal{A}(m)$ is not unique for a given bin m, and by extension the cardinality of sets of associated probabilities is greater than one, then we propose a method to construct probabilities for the complete data. We note that equation (10) implies that:

$$\mathcal{N}_0 = \sum_m \mathcal{P}(m)\mathcal{L}(m)$$

where we index the matrices $\mathcal{P}(m)$ and $\mathcal{L}(m)$ by their bin, m. It is immediate that the number of possible combinations of $\mathcal{P}(m)$ may be large unless we impose known restrictions on $\mathcal{L}(m)$ which we can do because we arbitrarily choose the bins. If each $\mathcal{L}(m)$ is identical then it follows that the possible enumerable probabilities in $\mathcal{P}(m)$ are identical across m since the network parameters do not depend on the initial probabilities. This by extension implies that the possible parameters $\{b_1, b_2, m_1, m_2\}$ reside in the same subset of rational numbers for each m. Thus if a candidate $\mathcal{A}(m)$ in bin m is not a candidate $\mathcal{A}(n)$ in bin n, then $\mathcal{A}(m)$ cannot be a candidate $\mathcal{A}$. The proof is straightforward: $\mathcal{P}(m)$ is independent of the direct graph implicit in $\mathcal{A}$, therefore the parameters governing the directed graph cannot depend on the bin. Thus, we can without loss of generality restrict our attention to the set of matrices $\tilde{\mathcal{A}}_t$ such that $\tilde{\mathcal{A}}_t = \mathcal{A}_t(1) \cap \mathcal{A}_t(2) \cap ... \cap \mathcal{A}_t(m)$. Because we are only using a single period of data to calculate each $\mathcal{A}_t$, there must exist at least one $\mathcal{A}_t$ that rationalizes our data if our network is fixed.

Finally, we observe that one can readily assure that $p_i \neq p_j$ by choosing prime number samples for each location except the last (the jth). This implies that the probabilities cannot be rationally factored and so cannot be exactly identical.

3 Set Identification of the Fixed Model

As we have argued above we can identify 4 parameters using at least two periods of observations on $\mathcal{B}$. Thus, we can identify m_1, m_2, b_1 and b_2 for a given pair $\{p_1, p_2\}$ and so our identification problem for our system boils down to identification of $\{p_1, p_2\}$. In the absence of further identifying information we would have a continuum of feasible solutions. Fortunately, the structure of our problem implies a number of further restrictions.

First, because our dependant variable is a count variable, the number of notes deposited at the Bank, it is by construction a natural number whose set we denote $\mathbb{N}$.[5] Thus, our model can be reduced to sets of $\{p_1, p_2\}$ that yield non-negative integer counts of bank note deposits in the initial distribution of notes. Given a countable injection of notes, L_1, L_2, there are a countable and finite number of such sets. For each p_i, we denote the rational number set $\mathcal{P}_i = \{n_i/L_i\}, \forall n_i = 0, 1, 2, ..., L_i$ so that $p_i \in \mathcal{P}_i$. There are thus $L_i + 1$ elements in $\mathcal{P}_i$ and so the feasible sets of $\{p_1, p_2\}$ have dimension $(L_1 + 1)(L_2 + 1)$. We refer to these sets as feasible sets. In this section we characterize the feasible sets of parameters of our model. Then we show how additional deposit data from the system can be used to reduce the feasible sets.

A second restriction is that all parameters of our system must lie in the interval $[0, 1]$ since they are probabilities. Moreover because every observed deposit of banknotes, *i.e.* every element of $\mathcal{B}_t$ $\forall t$, must be an integer then all parameters of the model must imply a natural number for deposits. Now consider again

[5] For our purposes, we define the set of natural numbers to include 0.

the first observed notes:

$$B_1^1 = L_1 p_1(b_1 + m_1 b_2) + L_1(1 - p_1)(b_2 + m_2 b_1)$$

and,

$$B_1^2 = L_2 p_2(b_1 + m_1 b_2) + L_2(1 - p_2)(b_2 + m_2 b_1).$$

Straightforward algebraic manipulation yields:

$$b_2 + m_2 b_1 = [\frac{p_1 B_1^2}{L_2} - \frac{p_2 B_1^1}{L_1}]/[p_1 - p_2].$$

which implies that observing the first period notes in combination with a particular pair $\{p_1, p_2\}$ yields a constraint on $b_2 + m_2 b_1$. Additionally, we may use the fact that the maximum possible value of B_1^1 is L_1 and likewise L_2 is the maximum possible value for B_2^2. This immediately yields that the maximum value $b_2 + m_2 b_1$ may take is 1. A similar algebraic manipulation yields:

$$b_1 + m_1 b_2 = [\frac{(1 - p_2)B_1^1}{L_1} - \frac{(1 - p_1)B_1^2}{L_2}]/[p_1 - p_2].$$

This expression yields $\max(b_1 + m_1 b_2) = 1$. We also note that a corollary is that the linear relationship between p_1 and p_2 can be expressed in terms of the model parameters and data:

$$p_2 = \frac{\frac{B_1^2}{L_2} - (b_2 + m_2 b_1)}{\frac{B_1^1}{L_1} - (b_2 + m_2 b_1)} p_1.$$

This $1:1$ mapping implies $\exists$ a unqiue $\{p_1, p_2\}$ for each set of parameters of the model. Furthermore because $b_2 + m_2 b_1 \in [0, 1]$ we can bound the possible values of p_2 as,

$$p_2 \in \big[[0, \frac{B_1^2 L_1}{B_1^1 L_2}] \cup [\frac{(B_1^2 - L_2)L_1}{(B_1^1 - L_1)L_2}, \frac{1}{p_1}]\big] p_1 \tag{12}$$

The restrictions that only natural numbers of notes are deposited also implies a third – each element of $\mathcal{A}$ must also be a rational number in $[0, 1]$ which follows from the parameter restrictions above. Thus we can eliminate any solution, $\{m_1, m_2, b_1, b_2\}$, that does not satisfy these restrictions conditional on a pair $\{p_1, p_2\}$. By extension we can eliminate any pair $\{p_1, p_2\}$ that produces such solutions. The diagonal elements of $\mathcal{A}$ imply that $b_1 + m_1 \in [0, 1]$ and $b_2 + m_2 \in [0, 1]$. We can further restrict $b_1 \in [0, 1 - m_1]$ and $b_2 \in [0, 1 - m_2]$. The off-diagonal elements of $\mathcal{A}$ provide no restrictions by themselves.

A fourth restriction is that every element in a $\mathcal{N}_t$, $t = 0, 1, 2, ..., T$ must be an integer. This restriction can be used to further reduce the sets of feasible $\{p_1, p_2\}$ without restricting the solution space since we can solve every non-linear system for different values $\{m_1, m_2, b_1, b_2\}$. To see the bite of this restriction, consider the first period distribution of notes, N_1. The integer restriction implies:

$$\mathcal{A}\mathcal{N}_0 \in \mathbb{N}.$$

We can use this restriction to eliminate sets of $\{p_1, p_2\}$. This integer restriction has further application sequentially to our problem since it also must be that $\mathcal{A}^2\mathcal{N}_0 \in \mathbb{N}$. As we have noted, it must be that all notes deposited are integers and we can solve every non-linear system for different values $\{m_1, m_2, b_1, b_2\}$ so we can impose additional integer restrictions. Let $\tau = 1, 2, 3, ..., T$ be the set of linear systems we observe and let $\mathcal{A}_\tau$ be the matrix $\mathcal{A}$ associated with each (note that $\mathcal{A}_\tau$ potentially changes in every odd period). Then it must be that:

$$\prod_{\tau} \mathcal{A}_\tau \mathcal{N}_0 \in \mathbb{N}.$$

And so we can sequentially eliminate any solution that does not satisfy this integer restriction.

We first establish that sequential elimination may reduce our identified set of $\{p_1, p_2\}$. Returning to our network structure, we write $\mathcal{B}_3$ as:

$$\mathcal{B}_3 = ((\mathcal{I} - \mathcal{A}_3)\mathcal{A}_1^2\mathcal{N}_0)^T\iota. \tag{13}$$

We can similarly express

$$\mathcal{B}_4 = ((\mathcal{I} - \mathcal{A}_3)\mathcal{A}_3\mathcal{A}_1^2\mathcal{N}_0)^T\iota \tag{14}$$

and solve for $\mathcal{A}_3$. Notice that we have at this point by construction found a matrix (if it exists) $\mathcal{A}_1$ that rationalizes our data for a set of $\{p_1, p_2\}$. We are now asserting that for this set we can find a matrix $\mathcal{A}_3$ such that equations (13) and (14) hold and that the restrictions on $\mathcal{A}_3$ hold (i.e. that is rational and with each element $[0, 1]$, etc.) for a given initial $\mathcal{N}_0$. By induction, sequential observations $\mathcal{B}_t$ may reduce our sets of feasible $\{p_1, p_2\}$.

Any $\{p_1, p_2\}$ that yields feasible parameters is a potential solution to our model and, in the absence of any additional information or restrictions, each candidate solution is equally probable. Just as clearly, every non-solution has probability zero (including all irrationals). There is no likelihood mapping that we can apply to the feasible solutions apart from a uniform distribution over all feasible solutions because there is only one observation of the system (everything else is deterministic).

However, this raises interesting insights for any solution method utilizing real-dimensioned probability distributions for similar latent-group systems with finite support. It is well known that the set of rationals as a proportion of the real numbers is effectively Lebesque measure zero despite the fact that the rationals are dense in the Real numbers. By extension, any solution method using distributions defined over the real numbers without restriction, such as a non-parametric regression using Bayesian updating, would seem inevitably to update mass points using 'information' that with almost surely zero probability comes from the system itself, despite the fact that the rationals are dense. Indeed, any solution proposed that is not uniform

in probability over $\{p_1, p_2\}$ across feasible solutions cannot be a true solution of the model. In contrast, our method reveals (an) exactly identified set(s).

3.1 The Two-Group Multiverse

We now address the question of whether repeated observation of the system can yield useful information regarding the system parameters. For example, one may observe additional injections of notes from different initial locations. Can anything be learned from repeated observation of the system?

The structure of the system described above suggests that repeated observation can bring additional identifying information. While repeated observations of note capture from the initial injections can yield additional information regarding the $\{p_1, p_2\}$, it does not yield additional identifying information for a given $\mathcal{A}$ because the parameters are permitted (expected) to time vary. However, we can apply a similar intuition to bring additional identifying power for the $\mathcal{A}$ matrix if we observe additional locations. In our discussion which follows, we define *exact* solutions A to be those that satisfy all observers.

Suppose, for example, that we observe note injections from an additional number of locations (these could, in principle, be the same locations at a different point in time). With a slight abuse of notation, we define location sets $\mathcal{L}(k)$ for k sets of $\{L_1, L_2\}$ locations and associate a initial location-group probability matrix $\mathcal{P}(k)$. For each location we can write:

$$\mathcal{B}_t(k) = (\mathcal{N}_{t-1}(k) - \mathcal{N}_t(k))^T \iota \equiv ((\mathcal{I} - \mathcal{A}_t)\mathcal{N}_{t-1}(k))^T \iota.$$

Each location set k is an observer of the system $\mathcal{A}_t$ and differs in its perspective according to the initial probabilities $\mathcal{P}(k)$ and the initial note injections $\mathcal{L}(k)$. These initial probabilities and injections are unlikely to be identical.

To illustrate, let $k = 1$ be the initial observer of the system and assume that this observer proposes candidate solutions $\mathcal{A}(1)$ and associated $\{p_1, p_2\}$. The second observer, taking the $\mathcal{A}(1)$ as given, then solves for its associated probabilities $\mathcal{P}(2)$ using $\mathcal{B}_1(2)$ and the candidate solutions $\mathcal{A}(1)$:

$$\mathcal{B}_1(2) = ((\mathcal{I} - \mathcal{A}(1))\mathcal{P}(2)\mathcal{L}(2))^T \iota.$$

These equations yield the unknown $\mathcal{P}(2)$. We note that if $\mathcal{P}(2) \notin [0, 1]$, then we can discard the associated $\mathcal{A}(1)$ as it is not consistent with the second observer's observations. Furthermore, we can also check whether the $\mathcal{A}(1)$ and $\mathcal{P}(2)$ are consistent with the second period observation by the second observer:

$$\mathcal{B}_2(2) = ((\mathcal{I} - \mathcal{A}(1))\mathcal{A}(1)\mathcal{P}(2)\mathcal{L}(2))^T \iota.$$

If the candidate $\mathcal{A}(1)$ and $\mathcal{P}(2)$ do not satisfy the second period observations then these cannot be exact

solutions. Thus we may eliminate the associated $\mathcal{A}(1)$ as an exact solution. We may proceed similarly for all k.

Naturally, this procedure is problematic if different observers observe any randomness in the model parameters. Suppose, for example, that observer k observes $\mathcal{A}(k) = E(k)\mathcal{A}$, where $E(k)$ is a matrix of random perturbations of the system parameters. It follows that such randomness implies that $B_2(k)$ will almost certainly fail to hold for any $P(k)$ found to satisfy $B_1(k)$ given $\mathcal{A}$. So this method raises the question of observing randomness in the elements and how such randomness may inform the likelihood of particular solutions.

We now argue that, unlike the case with a single observer, multiple observers can provide information over the likelihood of feasible solutions. Suppose, for example, that we observe k location sets and each location set proposes a $\mathcal{A}(k)$. It follows using straightforward logic that solutions observed in many $\mathcal{A}(k)$ are more likely to be observed than solutions in a single $\mathcal{A}(k)$ if we were to examine an $k+1$th observer. If we view each observer as a random sample of the population, it follows that the sample likelihood of some observation $\mathcal{A}(i)$ is $\sum_k (1/k)\mathbb{I}(\mathcal{A}(i) \in A(k))$, where $\mathbb{I}$ is an indicator function indicating that $\mathcal{A}(i)$ is found in $\mathcal{A}(k)$. Thus, in the absence of additional information, repeated observers are necessary and sufficient to describe a likelihood mapping over the sets $\{p_1, p_2\}$.

4 Two Group Algorithm

In this section, we further specialize our model to the two group case and show how to construct exact solutions, given $\{p_1, p_2\}$. Notice that Equation (6) can be manipulated to solve:

$$q \equiv (\mathcal{I} - \mathcal{A})^T \iota = (\mathcal{N}_0^T)^{-1}\mathcal{B}_1 = \begin{bmatrix} \frac{L_2(1-p_2)B_1^1 - L_1(1-p_1)B_1^2}{L_1L_2(p_1-p_2)} \\ \frac{L_1p_1B_1^2 - L_2p_2B_1^1}{L_1L_2(p_1-p_2)} \end{bmatrix}.$$

Where B_1^i is the number of notes from location L_i captured by the bank in period 1. Using $\mathcal{B}_2$ and Equation (6) we can solve similar expressions $r \equiv (\mathcal{I} - \mathcal{A})^T A^T \iota = (\mathcal{N}_0^T)^{-1}\mathcal{B}_2$. We note that q and r are simply functions of data and $\{p_1, p_2\}$.

Given q and r we can solve explicitly for the matrix $\mathcal{A}$, and b_1, b_2, m_1 and m_2. In particular,

$$\mathcal{A} = \begin{bmatrix} \frac{q_2(q_1-1)+r_1}{q_1-q_2} & \frac{q_2(q_2-1)+r_2}{q_1-q_2} \\ \frac{q_1(1-q_1)-r_1}{q_1-q_2} & \frac{q_1(1-q_2)-r_2}{q_1-q_2} \end{bmatrix}.$$

where q_j is the jth element of the vector q and r_k is the kth element of the vector r. Each element of $\mathcal{A}$ has an explicit solution and so we can quickly evaluate if a given $\{p_1, p_2\}$ yields elements of $\mathcal{A}$ that are bounded $[0, 1]$.

If a candidate $\{p_1, p_2\}$ yields feasible values for $\mathcal{A}$ we can proceed to solve exactly for m_1, m_2, b_1 and b_2. W.l.o.g we find it convenient to express each parameter as a function of m_2 and then solve m_2 given the

elements of $\mathcal{A}$. Using our equations for q and r and the quadratic formula, we can write:

$$m_2 = \frac{-(a_{2,1} - a_{1,2} + a_{2,2}a_{1,1}) + \sqrt{(a_{2,1} - a_{1,2} + a_{2,2}a_{1,1})^2 + 4a_{1,1}a_{2,2}a_{1,2}}}{2a_{1,1}}$$

where the + in the quadratic root is necessary and sufficient for $m_2 > 0$. We can then express:

$$m_1 = \frac{a_{2,1}}{a_{2,2} + m_2},$$

$$b_1 = 1 - \frac{a_{1,2}}{m_2}$$

and,

$$b_2 = 1 - \frac{a_{2,1}}{m_1}.$$

Given $\{p_1, p_2\}$, these solutions exactly identify the parameters of the model given the data.

At this point we can sketch our solution algorithm:

1. Choose a candidate p_1 sequentially (in either order) from $n_1 = 0, 1, ..., L_1$ and compute $p_1 = n_1/L_1$.
2. Choose a candidate p_2 such that $p_2 \in \left[[0, \frac{B_1^2 L_1}{B_1^1 L_2}] \cup [\frac{(B_1^2 - L_2)L_1}{(B_1^1 - L_1)L_2}, \frac{1}{p_1}]\right] p_1$ and p_2 is rational denominated by L_2.
3. Solve for $\mathcal{A}$ and verify.
4. Eliminate candidate $\{p_1, p_2\}$ if
$$\mathcal{A}\mathcal{N}_0 \notin \mathbb{N}.$$
5. Solve for $\{m_1, m_2, b_1, b_2\}$ given $\mathcal{A}$.
6. Eliminate candidate $\{p_1, p_2\}$ if any parameter < 0 or > 1.
7. Next proceeding iteratively, we take any remaining candidate $\{p_1, p_2\}$ and construct $\mathcal{N}_2$. We then return to step 3 and re-solve for $\{m_1, m_2, b_1, b_2\}$ using B_3, B_4, etc. and continue until the sets $\{p_1, p_2\}$ can be reduced no further.

4.1 Two Group Application

We use data from the circulation trials of $50 polymer notes conducted by the Bank of Canada. The circulation trials follow roughly 7 million notes released from one location and 1.4 million notes released from the second location. We use the solution algorithm presented in Section 4 to evaluate all feasible candidate solutions for a fixed network. Our solution algorithm fails to find a single feasible set of parameters,

$\{m_1, m_2, b_1, b_2\}$, and probabilities, $\{p_1, p_2\}$, that rationalize our data. Our algorithm took roughly two days to search over every feasible parameter. Given our argument above, we can conclude that there does not exist any fixed network structure underlying the circulation trial data. Thus, we can conclude that at least one parameter must be time-varying.

We also conclude that any method that estimates a fixed network structure for such data is incorrect since no fixed network exists. This would extend to any proposed method utilizing continuous distributions.

We conjecture that our results may be applicable to models that examine micro-founded monetary frictions, depending on the nature of the data. In particular, it appears that any matching model that features fixed matching between our groups is generally unable to rationalize the circulation trial data.

4.2 Extension to More Than Two Groups

The lack of a fixed network structure underlying the data can be generalized to rule out any fixed network structure underlying our currency data. If the network structure is fixed, we show that additional nodes, N_i, can be represented as convex combinations of the two initial nodes. We illustrate the argument using our two location representation but our results straightforwardly generalize to a more location case.

Suppose there exist $i > 2$ possible groups. The matrix representation of the system remains unchanged. In particular, it remains that:

$$\mathcal{N}_x = \mathcal{A}\mathcal{N}_{x-1},$$

where $\mathcal{N}_x$ is a $i \times 2$ matrix of allocations of the bank notes from locations 1 and 2 in period x. Next define $\hat{\mathcal{N}}_x$ as a 2×2 matrix representing a two group representation of this system. We then define a matrix M such that:

$$\hat{\mathcal{N}}_x = M\mathcal{N}_x. \tag{15}$$

It follows that:

$$\hat{\mathcal{N}}_x = M\mathcal{A}\mathcal{N}_{x-1}. \tag{16}$$

If we define a 2×2 matrix $\hat{\mathcal{A}}$ such that:

$$\hat{\mathcal{N}}_x = \hat{\mathcal{A}}\hat{\mathcal{N}}_{x-1}, \tag{17}$$

then it follows that:

$$M\mathcal{A}\mathcal{N}_{x-1} = \hat{\mathcal{A}}M\mathcal{N}_{x-1} \tag{18}$$

If we define some matrix M^* as the Moore-Penrose pseudoinverse of M' then we can write:

$$\hat{\mathcal{A}} = M\mathcal{A}(M^*)' \tag{19}$$

Since our system is linear, the pseudoinverse exists. Thus, if there does not exist a $\hat{\mathcal{A}}$ that rationalizes our data for a fixed two group case then there does not exist any fixed network representation of our data.

5 Stable Networks

The lack of fixed network structure may be potentially convenient. Recall our two group system:

$$\mathcal{B}_t = (\mathcal{N}_{t-1} - \mathcal{N}_t)^T\iota \equiv ((\mathcal{I} - \mathcal{A})\mathcal{N}_{t-1})^T\iota,$$

with

$$\mathcal{N}_{t-1} = \mathcal{A}^{t-1}\mathcal{N}_0$$

The identification problem inherent in the system is that powers of $\mathcal{A}$ do not yield identifying information about $\mathcal{A}$ in a deterministic system. However, the same is not true if $\mathcal{A}_t$ is time-varying. As we show next, time variation in the system parameters may help to identify $\{p_1, p_2\}$ and so we can establish identification of our model parameters and the system probabilities via observation of B_t. We define such networks as stable networks because we maintain the assumption that the direct graph underlying the system is unchanging.

We note that assuming $\mathcal{A}$ is time-varying is isomorphic to assuming sampling variability in $\mathcal{B}$. We note that one might be tempted to write a model with measurement error in the number of notes returned to the Bank. Yet doing so would, in our view, be a mistake as it could *only* represent arithmetic error in adding up the number of returned notes. In any other interpretation, regarding $\mathcal{B}_t$ as a sample of notes at time t produced by the network implies that any variation over time in the sample produced by the network must result from variation imposed by the network.

5.1 Estimation of Stable Network Structure

We rewrite our model with time-varying $\mathcal{A}_t$ such that:

$$\mathcal{N}_t = \mathcal{A}_t\mathcal{N}_{t-1} = (\prod_{j=1}^{t}\mathcal{A}_j)\mathcal{N}_0. \tag{20}$$

It is straightforward to show that:

$$L - \sum_{j=1}^{t}\mathcal{B}_j = (\mathcal{N}_0)^T(\prod_{j=1}^{t}\mathcal{A}_j)^T\iota, \tag{21}$$

where $L = diag(\mathcal{L})$. It remains that $\mathcal{A}_t$ is rational in every period and that the constraints outlined in Section 3 must hold in every period. We assume that the data generating process underlying $\mathcal{A}_t$ is *i.i.d.*

such that the difference in the realization of $\mathcal{A}_t$ from it's average value does not depend on the difference in prior time periods.[6] We require a feasible probability space that is discrete and bounded $[0, 1]$.

Both sides of Equation 21 are $N \times 1$ vectors, however the term $\prod_{j=1}^{t} \mathcal{A}_j$ is problematic to characterize especially as t grows.[7] We propose a simple method-of-moments approach to solve for the average parameters of the network. Our method is straightforward but relies on two sets of assumptions, one of which is innocuous. The first assumption is that the note sensors are error-free and that there is no measurement error in our dependent variables of interest. This assumption while particularly strict in some settings, is not strict in ours as the error rates of the optical scanners used to identify the notes are very, very, small. This assumption may also be relaxed at the cost of weaker identification of the model parameters. The second assumption that we require are that the time variation in the parameters is distributed *iid*. This assumption has real bite as it implies no correlation in the time-variation between parameters and no correlation within parameters. It is possible, however, to extend our method to deal with across and within variation.

We note that equation (21) can be written as:

$$L - \sum_{j=1}^{t} \mathcal{B}_j = (\mathcal{N}_0)^T(\Psi_t), \tag{22}$$

where,

$$\Psi_t = (\prod_{j=1}^{t} \mathcal{A}_j)^T \iota. \tag{23}$$

Ψ_t is a vector.

For ease of exposition, we focus on the 2 group case. We define a vector $V = [1 - 1]$ and, normalizing by the number of initial number of notes $\mathcal{L}$, we write

$$V\mathcal{L}^{-T}(L - \sum_{j=1}^{t} \mathcal{B}_j) = V(\mathcal{P})^T(\Psi_t) = (p_1 - p_2)V(\Psi_t), \tag{24}$$

where it is clear that V takes within vector differences so that both sides of equation (24) are scalars of the within element differences. Since these are scalars, we can take a ratio of any period difference, denoted k, to yield equations isomorphic to:

$$\frac{V\mathcal{L}^{-T}(L - \sum_{j=1}^{t} \mathcal{B}_j)}{V\mathcal{L}^{-T}(\mathcal{L} - \sum_{j=1}^{t-k} \mathcal{B}_j)} = \frac{(p_1 - p_2)V(\Psi_t)}{(p_1 - p_2)V(\Psi_{t-k})}. \tag{25}$$

[6] We conjecture that we may be able to relax this assumption slightly by assuming a Markov process.

[7] One option to estimate the parameters in A would be to proceed via some method of indirect inference. This method is potentially computationally burdensome, particularly given that our fixed network structure algorithm took roughly 36 hours to evaluate the feasible parameter space on a 32 node cluster.

It suffices to note that simple arithmetic leads to the cancelation of the unknown probabilities in $\mathcal{P}$ (which also highlights the idenification issue). We also note that because $\{p_1, p_2\}$ are probabilities, then their difference is bounded $[-1, 1]$ which implies an additional restriction to the moments.

To find the mean parameters of the system, we write:

$$\mathcal{A}_j = \mathcal{A} + \mathcal{E}_j$$

where $\mathcal{E}_j$ is the time variation from the average $\mathcal{A}$ in period j. By our assumptions, the expectation, E, of the realizations of $\mathcal{A}_j$ are:

$$\mathsf{E}[\mathcal{A}_j] = \mathcal{A} + \mathsf{E}[\mathcal{E}_j] = \mathcal{A}$$

Using our assumption that the $\mathcal{E}_j$ are distributed *i.i.d.*, $\mathsf{E}[\Psi_j] = (\mathcal{A}^j)^T \iota$.

To solve for the mean values of our system parameters, we can construct moments based on equation (25). We note that the average $\mathcal{A}$ is rational but not restricted to our initial set because the set of interest is discrete. For example, we consider moments

$$\mathsf{E}\left[\frac{V\mathcal{L}^{-T}(L - \sum_{j=1}^{t} \mathcal{B}_j)}{V\mathcal{L}^{-T}(L - \sum_{j=1}^{t-k} \mathcal{B}_j)}\right] - \mathsf{E}\left[\frac{V(\Psi_t)}{V(\Psi_{t-k})}\right] = 0. \tag{26}$$

Other moments which eliminate the $\{p_1, p_2\}$ do exist but are not as numerically appealing as the difference. The reason is subtle and stems directly from the network structure imposed which has implications for the support of the random noise in each period. The network structure imposes that the random element in any period t is equal to the current draw of $\mathcal{E}_t$ and lags of previous $\mathcal{E}_j$, for $j < t$, multiplied by matrix powers of $\mathcal{A}$. Thus the support of the random element changes throughout the sample period and depends upon the model parameters. The difference moments we propose impose the same support between the data 'part', $\frac{V\mathcal{L}^{-T}(L-\sum_{j=1}^{t} \mathcal{B}_j)}{V\mathcal{L}^{-T}(L-\sum_{j=1}^{t-k} \mathcal{B}_j)}$, and the model part, $\frac{V(\Psi_t)}{V(\Psi_{t-k})}$ precisely because we are find model parameters that rationalize the data. Other moments typically impose a different support between the first part of the moment and the second and consequently lead to numerical issues as any optimization routine may overweight early period moments whose support for the random elements is larger than for later periods.[8]

One obvious difficulty with Equation (26) is that the second expectation term is biased because of the time-variation in the stochastic part of the denominator renders the expectation biased by Jensen's Inequality. One approach to deal with the potential bias is to fix the denominator at a given period, z, and write the moments for period t as

$$\mathsf{E}\left[\frac{V\mathcal{L}^{-T}(L - \sum_{j=z+1}^{t} \mathcal{B}_j)}{V\mathcal{L}^{-T}(L - \sum_{j=1}^{z} \mathcal{B}_j)}\right] - \mathsf{E}\left[\frac{V(\Psi_t)}{V(\Psi_z)}\right] = 0. \tag{27}$$

[8]This issue is admittedly nuanced and we are happy to provide more details but opt not to do so here. Our numerical simulations strongly support our argument and these results are available upon request.

One might think that a natural candidate is $z = 1$. We remind that the normalization is simply to eliminate the unknown nuisance parameters $\{p_1, p_2\}$ and so any period normalization suffices. In this case, the 'measurement error' is a scalar constant across all moments (*i.e.* it is not time-varying). Thus, the proper correction factor in this case is a scalar. Unfortunately, this correction factor is unknown.

This observation, in fact, motivates a possible robustness exercise since the bias introduced by the expectation is constant across all moments and can be characterized by introducing sample draws from the distribution of the $\mathcal{E}_j$s to estimate $\mathcal{E}_1$. If the distribution of $\mathcal{E}_j$ were known then one could effectively sample from the distribution. While one could calculate a probability-weighted estimate of the system parameters, doing so also requires constructing inverses of the sampling pdf which may be computationally burdensome. In practice, the distribution of the $\mathcal{E}_j$ is unlikely to be known and so sampling for $\mathcal{E}_1$ may be infeasible.

An alternative method to reduce bias stemming from $\mathcal{E}$ is available. By assumption, the $\mathcal{E}_j$ are distributed mean zero and *i.i.d.* and so the median of a sample draw from the $\mathcal{E}_j$ is, in expectation, zero. Since our normalization can use any period as the base period, we propose using a period z which implicitly contains a sample of z draws from the distribution of the $\mathcal{E}_j$, and then to use the remaining $t - 1$ periods to solve for the system parameters. Since each of the $\mathcal{E}_1 ... \mathcal{E}_z$ will multiply $\mathcal{A}$ then, in expectation, the sum of these terms should be zero (the remaining terms depend on higher order powers of the random deviations whose effect we ignore as a first-order approximation). We can, in fact, bound the probability that the sum $S_z = V(\mathcal{E}_1 + ... + \mathcal{E}_z)\iota$ is greater than some amount $\tau > 0$ using Hoeffding's (1963) inequality for bounded distributions. This, by extension, provides a bound on the bias which can be incorporated into any solution method to assess the sensitivity of the results. Hoeffding's inequality states

$$\mathbb{P}[S_z - \mathsf{E}[S_z] \geq \tau] \leq \exp(-\frac{2z^2\tau^2}{\sum_{i=1}^{z}(c_i - a_i)^2})$$

where $\mathbb{P}[V\mathcal{E}_i\iota \in [a_i, c_i]] = 1$. The specification of the bounds is not entirely clear. Since none of the elements in A can be negative, these provide implicit bounds on a_i and c_i. In our empirical work, we assume $a_i = -0.01$ and $c_i = 0.01$.[9] If one feels these are too tight, these bounds can be relaxed. What remains is the optimal choice of the number of time periods, z. Ideally one would like to set z as large as possible to reduce the probability that the bias exceeds τ. Thus, the optimal choice of z is simply the number of time periods, T, if the errors are not time dependent, *e.g.* by exhibiting seasonality. For our empirical application to the circulation trial data we set $z = 50$, so the probability that the errors exceed 0.0001, $\tau > 0.0001$, is less than 0.0000001.

[9] We assume these bounds because simple algebra implies that the $V\mathcal{E}_i\iota$ equate to a difference in the errors for the b terms and a difference in the errors for products of the m and b terms. Equally clearly, the second terms subtract from the first and so that maximum difference that can be obtained is the difference between the errors for b_2 minus the errors for b_1. Since the errors cannot bet larger than b_2 or b_1, and we know that for our system that the bs are small, our bounds for a_i and c_i seem very conservative.

5.2 Computational Fixed Point

We note that, depending on the data, the appropriate mininimizer for the moment conditions is not trivial to choose. In particular, it remains that all solutions to the model must be rational, including the random variation. As a result, for systems with sufficiently small numbers of objects, continuous solution methods may not be appropriate. However, as long as these systems are sufficiently small, one can straightforwardly adapt the solution algorithm proposed in Section 4 to search over the permissible solutions. Alternatively, one could assume that the elements of $\mathcal{A}_t$ are drawn from a known distribution. For example, one could assume that the parameters are described by a triangle distribution over the feasible rationals with location specific pivot $h_{i,j}$, for locations $i = \{1, 2\}$ and groups $j = \{1, 2\}$ and then solve for the pivots that provide the best fit to the data.[10]

Our reductionist method inherent in $\mathcal{V}$ is not without computational cost – in particular it yields an identification problem resulting from the operator V. Specifically, $y = b_1(1 - m_2) - b_2(1 - m_1) = 0$ implies $V(\Psi_m) = 0$ for all m because the eigenvector associated with y returns a single eigenvalue. Thus any direct solution method applied to Equation (29) is not well-defined without some type of adaptation. One possibility would be to use a lasso-type penalty term applied to y, however such an approach is *ad-hoc*.

We construct a computational method which applies a fixed-point logic based on Equation (24). We note that one can conveniently write:

$$V\mathcal{L}^{-T}(L - \sum_{j=1}^{t} \mathcal{B}_j) + K_1 = V(\mathcal{P})^T(\Psi_t) = (p_1 - p_2)(V(\Psi_t) + K_2), \tag{28}$$

where $K_1 = (p_1 - p_2)K_2$. By extension, we can re-write our model moments as:

$$\mathsf{R}(K_2) = \sum_t \left| \frac{V\mathcal{L}^{-T}(L - \sum_{j=z+1}^{t} \mathcal{B}_j) + K_1}{V\mathcal{L}^{-T}(L - \sum_{j=1}^{z} \mathcal{B}_j) + K_1} - \frac{V(\Psi_t) + K_2}{V(\Psi_z) + K_2} \right|. \tag{29}$$

We let $K_1 = 1$ which implies that $K_2 = 1/(p_1 - p_2)$. By our assumptions K_2 is bounded away from 0. Thus, even if $y = 0$, neither the LHS or RHS denominator go to zero and standard computational approaches are feasbile.

This approach replaces y with a second unknown parameter K_2 which may, at first glance, appear to offer little gain. However, $1/(p_1 - p_2)$ is a scalar since, in our model, p_1 and p_2 do not change over time. Thus K_2 is a constant. Moreover, the parameters of our network do not depend on the initial probabilities so neither do they depend on K_2. Our computational approach is therefore to guess K_2, and solve for $\{b_1, b_2, m_1, m_2\}$, given this choice of K_2. We then update our guess of K_2 using a standard Newton step:

$$K_2^{new} = K_2^{old} - \sum_t \frac{\mathsf{R}(K_2)}{\mathsf{R}'(K_2)},$$

[10]This procedure would, in fact, also identify the variance associated with each parameter. We return to this issue below.

where $\mathsf{R}'(K_2)$ is the derivative of $\mathsf{R}(K_2)$ with respect to K_2. We note that one can analytically express:

$$\mathsf{R}'(K_2) = \frac{V(\Psi_z) - V(\Psi_t)}{(V(\Psi_z) + K_2)^2},$$

where we impose the equilibrium restriction that our parameters do not depend on K_2 along the solution path. This approach is potentially problematic so to ensure robustness, we propose using a wide-mesh of starting points to ensure robustness. The analytical expression for $\mathsf{R}'(K_2)$ is convenient because if $y = 0$ then the numerator is likewise 0. Thus, we can evaluate whether the true solution is $y = 0$. If $y = 0$ then our model implies that the groups are identical up to a scalar translation. We impose a stopping criterion $K_2^{new} - K_2^{old} < \delta$, for δ small, to yield our solutions in terms of K_2 and $\{b_1, b_2, m_1, m_2\}$.

While we have written the above function, $\mathsf{R}(K_2)$, as a scalar, in practice any vector of appropriate moments suffices. Indeed, it is immediate that the solutions obtained should be identical whether we average over the $T-1$ observations or whether we construct a single vector of $T-1$ moments since the mean of the expectation should be equal to the expectation of the mean. However, we note that one could, in theory, solve for a collection of sets of $\{m_1, m_2, b_1, b_2\}$ from a collection of fixed sub-blocks of observations. For example, suppose one believes that seasonal or trend factors lead to variation in the system parameters. One could construct moments based on at least 4 periods of data and solve for $\{m_1, m_2, b_1, b_2\}$ which satisfy each sub-block separately. Although we do not wish to push the analogy too far, this procedure mimics a non-parametric procedure for recovering a distribution of $\{m_1, m_2, b_1, b_2\}$.

The choice of the L1 norm as the objective function is not innocuous. Alternatives to the L1 norm, such as the L2 norm, are common in numerical work and there are well-known advantages and disadvantages to each. However, the L1 is preferred because it imposes a constant relationship between p_1 and p_2. Our 'trick' to solve the system requires the cancelation of the unknown probabilities and, as we discussed in the fixed model setting, the labelling of groups requires imposing a symmetry condition. The L2 norm confounds the symmetry relationship between two possible moments, *e.g.* one moment where $p_1 > p_2$ and the second where $p_2 > p_1$. Because our procedure only identifies groups and not labels, the L2 norm will only by chance find solutions where both moments label the groups identically. If the moments do not label the groups identically, then the L2 norm is not identified because it will find solutions where the implied probabilities are identical $\tilde{p}_1 = (p_1 + p_2)/2 = \tilde{p}_2$. Because the L1 norm only uses one moment to solve for the system parameters, the labelling of groups is by necessity unique.

Once we have solutions for $\{m_1, m_2, b_1, b_2\}$ and K_2, we can recover the individual $\{p_1, p_2\}$. We recall that one can write the system using equation (7) as:

$$\mathcal{B}_t = M\mathcal{B}_{t-1}$$

where,

$$M = (\mathcal{L}\mathcal{P})^T \mathcal{A}^T (\mathcal{L}\mathcal{P})^{-T}.$$

Although this system of moments does not identify the 6 system parameters because M is a 2×2 matrix, given solutions for $\{m_1, m_2, b_1, b_2\}$, and the restriction implied by K_2, we can use equation (7) to find the $\{p_1\}$ that maximizes the model's fit to the data. Then $p_2 = (1 - p_1 K_2)/K_2$.

As a final observation, we note that recovering the variance in the parameters $\{m_1, m_2, b_1, b_2\}$ appears difficult without knowing the distribution of the $\mathcal{E}_j$. Given a candidate distribution, one could, perhaps, solve parametrically for variances that produce the best fit to the data, perhaps using higher-order moments. The main difficulty is that there are four unobserved variances and, without imposing further assumptions, only one observation of the variance of $V(\Psi)$. As we noted above, one possible approach is to assume a known distribution for the parameters, such as a triangle distribution. One could then parametrically identify the variances. We leave this topic for future research.

We also note that our method can be equally applied to sampled data. While we have taken the view that our data represent a population, nothing in particular hinges on this assumption. We note that one could use straightforward GMM estimation to recover system parameters by minimizing

$$\left[\frac{V\mathcal{L}^{-T}(L - \sum_{j=m+1}^{t} \mathcal{B}_j)}{V\mathcal{L}^{-T}(L - \sum_{j=1}^{m} \mathcal{B}_j)} - \frac{V(\Psi_t)}{V(\Psi_m)}\right]' W^{-1} \mathbb{F}\left(\left[\frac{V\mathcal{L}^{-T}(L - \sum_{j=m+1}^{t} \mathcal{B}_j)}{V\mathcal{L}^{-T}(L - \sum_{j=1}^{m} \mathcal{B}_j)} - \frac{V(\Psi_t)}{V(\Psi_m)}\right]\right) = 0. \tag{30}$$

where W is a covariance (weighting) matrix and $\mathbb{F}$ is a suitable transformation of the moment that preserves symmetry of the labelled groups. We leave investigation of such suitable transformations for future research.

5.3 Multiple Groups

Finally, we note that it is straightforward to extend our method to deal with larger groups (and/or locations) using the same principle by an appropriate characterization of the operator used to eliminate the unknown probabilities.

To analyse systems with multiple groups requires a more careful discussion of the transition probabilities. In a two group system, the probabilities represent the probability that a note switches between groups in a given time period. With more than two groups, there is the possibility that a note moves between, say, group 1 and group 2 and then to group 3 in a given time period. Essentially, one must take a position on the compounding of probabilities.

For our purposes, we assume that a note can make at most one transition in a given period between groups and that deposits are realized at the period end. Thus, for a three group system, our $\mathcal{A}$ becomes

$$\mathcal{A} = \begin{bmatrix} 1 - b_1 - m_{1,2} - m_{1,3} & m_{2,1}(1 - b_1) & m_{3,1}(1 - b_1) \\ m_{1,2}(1 - b_2) & 1 - b_2 - m_{2,1} - m_{2,3} & m_{3,2}(1 - b_2) \\ m_{1,3}(1 - b_3) & m_{2,3}(1 - b_3) & 1 - b_3 - m_{3,1} - m_{3,2} \end{bmatrix}, \tag{31}$$

where $m_{i,j}$ is the probability that a note transits from group i to group j in a period. (The general group case with m groups is analogous.) Since there are three groups, our initial distribution of notes, $\mathcal{N}_0$ from the two locations is a 3×2 matrix

$$\mathcal{N}_0 = \begin{bmatrix} L_1 p_{1,1} & L_2 p_{2,1} \\ L_1 p_{1,2} & L_2 p_{2,2} \\ L_1(1 - p_{1,1} - p1, 2) & L_2(1 - p_{2,1} - p_{2,2}) \end{bmatrix} = \mathcal{P}\mathcal{L}, \tag{32}$$

where $p_{i,j}$ is the probability that a note from location i was initially withdrawn by group j.

To implement our solution method described above, we need to find moments that cancel the unknown probabilities. Clearly, a single period difference is unable to do so. We next present one method to eliminate the unknown probabilities (there may be a number of possible methods). For ease of notation we write the elements of $\Psi_t = [\psi_t^1, \psi_t^2, \psi_t^3]^T$ and note that we can express for any period t

$$\frac{VL_{t-1}(\psi_t^2 - \psi_t^3) - VL_{t-1}(\psi_{t-1}^1 - \psi_{t-1}^2)}{((\psi_{t-1}^1 - \psi_{t-1}^2)(\psi_t^2 - \psi_t^3)) - ((\psi_t^1 - \psi_t^2)(\psi_{t-1}^2 - \psi_{t-1}^3))} = K \tag{33}$$

where K is a constant (and equal to $p_{1,1} - p_{2,1}$). Thus we can take any period $t > 1$ as our normalization and solve a system of moments for the unknown parameters.[11] One can deduce that the operator necessary to eliminate the unknown probabilities, given two locations, uses Gaussian elimination on a system of VL_t equations to reduce the system to one unknown combination of probabilities. Since these probabilities are constant, they can form the moments used to solve for the system parameters. In cases with more than two initial locations the procedure is similar and involves solving for a fixed set of probabilities which can be used to construct moments over time.

To illustrate in general notation, consider a system with N groups and M locations so that $\mathcal{P}$ is $N \times M$. Also note that Ψ is $N \times 1$. Let V remain a $1 \times M$ vector difference operator as before. We can generally write our system as before (see equation (24)):

$$V\mathcal{L}^{-T}(L - \sum_{j=1}^{t} \mathcal{B}_j) = V(\mathcal{P})^T(\Psi_t).$$

We note that $\mathcal{P}$ has the desireable property that each column sums to 1 since the elements are probabilities. As a result, $V\mathcal{P}^T$ (a $1 \times N$ vector) has the likewise desireable property that one element (we normalize to the last element) implicitly sums the elements of V. The 'trick' to solving the moments is to choose V such that the sum of its elements is 0 (notice that in our two-group example we defined V in this way).

Next, note that Ψ_t is a $N \times 1$ so that $V\mathcal{P}^T\Psi_t$ is a scalar which can be expressed as:

[11] The operator to eliminate the unknown probabilities in this case is: (i) to take the within period difference using V as defined above, then (ii) rewrite the expression to solve for $p_{1,2} - p_{2,2}$ and (iii) use the solution for $p_{1,2} - p_{2,2}$ in a different period to solve for $p_{1,1} - p_{2,1}$.

$$\sum_{i=1}^{N-1}(\sum_{j=1}^{M} v_j p_{ij}(\Delta_{iN}\Psi_t)), \tag{34}$$

where v_j is the jth element of the vector V, p_{ij} is the ijth element of the matrix $\mathcal{P}$ and $\Delta_{iN}\Psi_t = \psi_{i,t} - \psi_{N,t}$ is the difference of the ith and Nth element of Ψ_t. (It is trivial that in our two group case this is just a scalar in $\Delta_{iN}\Psi_t$). These summations are over $N-1$ unknown constants (the $v_j p_{ij}$ which are constant over t). Our procedure involves eliminating the unknowns using successive periods of Ψ_t and the data. Thus, for example, in a model with three groups we require two time periods (as shown above). Generally, we require $N-1$ time periods to eliminate the unknown probabilities since the v_j are constant and p_{ij} is constant in dimension i.

We define σ as the $1 \times N-1$ vector of the inner summation term and Δ as an operator which gives the $N-1 \times 1$ vector of the $\Delta_{iN}\Psi_t$ elements. We then define a matrix Ψ^t which is a $N-1 \times N-1$ matrix in which each column is a vector Ψ_t for columns of t from $t = \tau$ to $t = \tau + N - 1$, for some period τ. Finally, we define $\mathcal{Y}_t$ as an $1 \times N-1$ vector in which each element is $V\mathcal{L}^{-T}(L - \sum_{j=1}^{t}\mathcal{B}_j)$ for $t = \tau$ to $t = \tau + N - 1$. We can then rewrite equation (34) as:[12]

$$\sigma(\Delta\Psi^t) = \mathcal{Y}_t. \tag{35}$$

Since each element of σ is invariant to the choice of τ it is then straightforward to elimnate σ using the difference moments we proposed in the two group case,

$$\frac{\sigma(\Delta\Psi^t)}{\sigma(\Delta\Psi^m)} = \frac{\mathcal{Y}_t}{\mathcal{Y}_m}. \tag{36}$$

where we again normalize by some period m. One can then construct an objective function with a lasso-type penalty

$$\tilde{\mathsf{R}} = \sum_t |\frac{\sigma(\Delta\Psi^t)}{\sigma(\Delta\Psi^m)} - \frac{\mathcal{Y}_t}{\mathcal{Y}_m}| + \lambda g \tag{37}$$

where λg is a penalty function corresponding to the eigenvalue identification problem described above for the two group case.

We caution that simply eliminating the unknown probabilities does not identify the system parameters. Our method still requires sufficient data to solve for the unknown system parameters. In general we will require $N^2 + M - 1$ data points to exactly identify the system parameters.

[12] Δ is therefore an operator which subtracts the last element of Ψ_t from the remaining rows and then deletes the last row.

5.4 Minimal Group Number

As discussed above, if $y = 0$ we cannot distinguish captured notes by groups since the proportion of notes held by each group is identical over time. Clearly, there are (potentially) a continuum of possible values of the model parameters such that $y = 0$ (and not simply that $b_1 = b_2$ and $m_1 = m_2$). As we noted above, it would be useful to test whether a model with K groups is preferred to a model with $K - 1$ groups, without necessarily specifying which group is not supported by the data. We note that we cannot conduct statistical tests on the solved parameters without specifying a distribution underlying the $\mathcal{E}$ because the observed residuals are an otherwise unknown combination of perturbations of the parameters. Nevertheless, it is possible to statistically test, sequentially, the significance of additional groups. Returning to our two group case, we note that if the true number of groups is 1, then our model implies onlyone parameter, a b, since our model does not restrict flows within groups. Thus, we can straightfowardly obtain a value of b from the data. The question addressed here is whether a two-group model is statistically preferred to a single group model. We denote the observed residuals from the single group model as e^1 and the observed residuals from the two group model as e^2. (Generically we denote e^j as the residuals from the jth group model.) We let $F_{e^1}(x)$ represent the empirical distribution function (edf) of the residuals from the single group model and $F_{e^2}(x)$ represent the edf of the two group model. By our assumptions on $\mathcal{E}$, these edf's are independent. Thus, following Kolmogorov (1933), we define the supremum statistic:

$$D_{e^1,e^2} = sup_x |F_{e^1}(x) - F_{e^2}(x)| \tag{38}$$

where D_{e^1,e^2} is the well-known Kolmogorov-Smirnoff statistic. D_{e^1,e^2} permits testing the equivalence of the distributions of e^1 and e^2 at a given level of significance α. Thus, we can test whether the distribution of the residuals across groups are identical. If they are, there is no statistical rationale to choose the model with a larger number of groups. If the distributions are different, then the model with the lowest mean-squared error provides the closest 'fit' to the data. An open question is whether this testing approach is valid sequentially, *i.e.* whether choosing $K - 1$ groups instead of K implies choosing $K - 1$ groups against $K + n$ groups, $\forall n > 0$. Is the minimal group number also the optimal group number?

Defining an optimal solution for our problem is non-trivial as there is no obvious measure of closeness or an obvious measure of an optimum. One candidate definition might be that the optimal number of groups minimizes a distance metric based on the observed residuals (such as mean squared error we suggest above). However, it still remains to define the optimal distribution of the residuals. It cannot be that the residuals are uniformly zero as this would imply a fixed network and thus the solution method of the stable network inapplicable. Yet, clearly a model with residuals that are uniformly zero would be optimal if such a model existed. We assert from our fixed model whether such a model exists. Thus, as a point of logic, it would seem that if we can apply the stable network solution then there is no model which is optimal even as the

limit of an asymptotic sequence in the number of groups. Yet equally clearly, because the data is discrete, the maximal number of groups is the number of initial objects issued (*e.g.* note injections). Hence, our model suggests an upper bound on the number of possible groups, $\bar{n}$. In theory, with sufficient data, one could test down from $\bar{n}$ using the Kolmogorov-Smirnoff statistic until a difference in the distributions was detected at some specified level of significance. This would then suggest a 'no-better-fit' number of groups. In practice, researchers are unlikely to have a sufficiently long series of data to do so. One possibility is to start at the largest n the data will support and test downwards from there. We leave this open question of the proper stopping point for the number of groups for further research.

6 Application to Bank Notes

In this section we first present some simulation evidence that our method provides reliable estimates of the network parameters. We then turn to our confidential circulation trial data and estimate the network structure underlying the usage of Canadian currency.

6.1 Simulation Evidence

[In progress / To be Completed]

6.2 Bank Note Data

We use our confidential circulation trial data, amounting to 64 weeks of data, and normalizing to period 50. We solve for our system parameters and find $\{b_1 = 0.0403, b_2 = 0.0275, m_1 = 0.2162, m_2 = 0.0015\}$ and $\{p_1 = 1, p_2 = 0.56\}$. Our results suggest that there are distinct sub-groups of cash-users in the population. In particular, group N_2 is more likely to hold notes for a longer period of time and also more likely to receive notes from group N_1 than to give notes to group N_1. Thus, welfare effects from inflation, for instance, would be more prononced for group N_2 than N_1.

7 Conclusion

This paper has proposed a novel methodology to model the usage of currency in the Canadian economy. Using confidential data from the Bank of Canada's currency circulation trials, we demonstrate a procedure to set-identify model parameters and describe a solution algorithm that describes these sets. We also show how additional periods of data may reduce the identified set and how to generalize our results. We are unable to find a fixed network structure to currency usage in Canada. We find evidence of time-varying systems parameters and evidence of distinct sub-groups of cash-users in the population. While it would, of course, be interesting to identify the sub-groups, at present we lack sufficient auxiliary data to do so.

8 Appendix: General Solution for Fixed Networks

We now consider a general environment with $I > 2$ groups. As we will show, such environments are challenging to solve because of the increasing interactions between groups. Adding a subgroup entails an addition of unknowns b and p (2 parameters) and additional cash transit probabilities that depend on the number of groups. The number of cash transit probabilities between groups, $\mathcal{M}_I$, for $I > 1$ groups are

$$\mathcal{M}_I = \frac{I!}{(I-2)!}.$$

For example, if $I = 3$ there are 6 inter-group transit probabilities in total. Thus a third sub-group adds $2 + 4 = 6$ parameters to the system to be estimated. An environment with $I > 2$ groups faces the same lack of identification of the initial probabilities and so requires the same integer programming solution method. We restrict our attention to square systems in which the number of locations equals the number of groups.

The general model can be represented similar to the two group case. Define for j locations and I groups:

$$\tilde{\mathcal{N}}_0 = \begin{bmatrix} L_1 p_{1,1} & ... & L_j p_{1,j} \\ \vdots & \ddots & \vdots \\ L_1(1 - \sum_{i=1}^{I-1} p_{i,1}) & \cdots & L_j(1 - \sum_{i=1}^{I-1} p_{i,j}) \end{bmatrix}. \tag{39}$$

We can similarly define $\tilde{\mathcal{A}}$ as a $I \times I$ matrix of parameters (note that $\tilde{\mathcal{A}}$ is always square). We note that $I > 2$ implies that $j = n$ is no longer sufficient to identify our parameters. In general we will require $\mathcal{M}_I + I \equiv I \times I$ equations to identify the parameters for any given candidate matrix of probabilities, $\tilde{\mathcal{P}}$. If we have fewer locations, and thus fewer elements in $\tilde{\mathcal{P}}$, we will require more time observations. Conversely, if we have more locations, and thus more elements in $\tilde{\mathcal{P}}$, then we will require searching over a large set of feasible probabilities. In general, for I sufficiently large the problem becomes numerically intractable because there are $(L_1+1)(L_2+1)...(L_j+1)$ probabilities to search over. Nevertheless, in general the steps similar to those outlined above for the two group case work.

One key simplification to our method is to restrict the domain of the probabilities using the implicit restrictions on the elements of $\mathcal{A}$. We now demonstrate how to restrict the probabilities for generic numbers of groups. Returning to system equation, Equation (6), reproduced below,

$$\mathcal{B}_t = (\mathcal{N}_{t-1} - \mathcal{N}_t)^T \iota \equiv ((\mathcal{I} - \mathcal{A})\mathcal{N}_{t-1})^T \iota$$

we note that one can, arbitrarily and without loss of generality, express the $\mathcal{B}_t$ as a diagonal matrix Ξ_t^a with $\mathcal{B}_t$ along the diagonal. Restricting attention to square systems in which the number of groups equals the number of locations, it follows immediately that $\Xi_t^a \iota = \mathcal{B}_t$. This is one (extreme) possibility for a square matrix that yields $\mathcal{B}_t$ since every element of $\mathcal{B}_t$ must be in $\mathbb{N}$. From Equation (6), we can then express:

$$\Xi_1^a = ((\mathcal{I} - \mathcal{A})\mathcal{P}\mathcal{L})^T$$

so for a given $\mathcal{A}$ we can solve for restrictions on the probability matrix $\mathcal{P}$ as:

$$\mathcal{P}^a = (\mathcal{L}^{-T}\Xi_1^a(\mathcal{I} - \mathcal{A})^{-T})^T$$

We can also obtain additional restrictions using a matrix Ξ_t^b with $\mathcal{B}_t$ along its anti-diagonal such that:

$$\mathcal{P}^b = (\mathcal{L}^{-T}\Xi_1^b(\mathcal{I} - \mathcal{A})^{-T})^T.$$

$\mathcal{P}^a$ and $\mathcal{P}^b$ describe the restrictions on the feasible probabilities, given that each element of $\mathcal{A} \in [0, 1]$, analogous to the Equation (12).

Equation (6) yields I equations for each t and so together can identify $2I$ parameters. This is problematic because when $I > 2$ then $\mathcal{M}_I > I$ and so these equations are insufficient to identify our individual parameters. However, as we demonstrated in Section 3.1, additional observations on the system can yield additional identifying information. Unlike our discussion in that section, additional 'observers' here do not only provide over-identifying information but also set-identifying information for the remaining unidentified parameters. Using analogous arguments we can show that each additional observers add I identifying equations and so with sufficient additional sets of locations we can solve the I^2 parameters. In particular, for $I > 2$ we require $k = I - 2$ additional observations to solve our model.

The solution algorithm is:

1. Given the initial quantity of notes deposited in period 1 denoted $\tilde{\mathcal{B}}_1$ solve for expressions for $p_1, p_2, ..., p_j$ as functions of the parameters. Since each combination of parameters in $\mathcal{A}$ remains bounded $[0, 1]$ then construct bounds on each p_j, $j > 1$, as a function of p_1 given initial $\mathcal{B}_1$.

2. Fix the last probability vector p_j.

3. Choose a candidate p_1 sequentially (in either order) from $n_1 = 0, 1, ..., L_1$ and compute $p_1 = n_1/L_1$.

4. Choose the remaining candidate p_i for $i = 2, ..., j - 1$ given the bounds established in step 1 and p_i is rational denominated by L_i.

5. For the first observer, solve for as many candidate elements of $\mathcal{A}$ as is feasible given number of equations. Evaluate over remaining candidate elements that are rational and bounded.

6. For the second observer, evaluate each of the remaining candidate elements and p_j. If the evaluated p_j does not equal the chosen p_j then eliminate the associated candidate p_is.

7. Solve for the parameters and eliminate any candidate $\tilde{\mathcal{P}}$ if any parameter < 0 or > 1.

8. Eliminate candidate $\tilde{\mathcal{P}}$ if

$$\mathcal{AN}_0 \notin \mathbb{N}.$$

9. Next proceeding iteratively, we take any remaining candidate $\{p_1, ..., p_j\}$ and construct $\mathcal{N}_2$. We then return to step 3 and re-solve for parameters using B_3, B_4, etc. and continue until the sets $\{p_1, ..., p_j\}$ can be reduced no further.

One may wonder the extent to which our approach can be used to 'test' for the appropriate number of groups needed to rationalize the data. Since our method does not require statistical inference, then likelihood procedures are not applicable. What we instead seek to do is to find the minimum number of groups required to exactly fit our data. Our method may fail to reach a solution if the true number of groups is sufficiently high because we may suffer from underidentification if the number of groups grows too high in relation to our data. In such cases all we can say is that at least X groups are required to explain the data.

References

[1] Barber, Donald 2012. "Bayesian Reasoning and Machine Learning," manuscript.

[2] Cherchye, Laurens, Ian Crawford, Bram De Rock and Frederic Vermeulen 2012, "Gorman Revisited: Nonparametric Conditions for Exact Linear Aggregation," KU Leuven Discussion Paper.

[3] Crawford, Ian and Krishna Pendakur, 2013. "How Many Types Are There?", *The Economic Journal*, 123(567), 7795.

[4] Hoeffding, W. 1963. "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, 58(301), 13-30.

[5] Kleibergen, F. and Paap, R. 2006. "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 133, 97-126.

[6] Kocherlakota, Narayana 1998. "Money is Memory," *Journal of Economic Theory*, 81(2), 232251.

[7] Rubin, Donald 1976. "Inference and missing data," *Biometrika*,1976, 63 (3): 581-592.

[8] Williams, B.K., Nichols, J.D. and Conroy, M.J. 2002. *Analysis and Management of Animal Populations*. Academic Press, San Diego.